

SureSelect XT HS2

ターゲットエンリッチメントシステムを用いた 効率的なロングリードシーケンス解析

著者

生井聡史
アジレント・テクノロジー株式会社

石井善幸
アジレント・テクノロジー株式会社

吉崎史子
アジレント・テクノロジー株式会社

関真秀
東京大学大学院
新領域創成科学研究科

鈴木稔
東京大学大学院
新領域創成科学研究科

アブストラクト

ロングリードシーケンシングは近年精度やスループットが向上し、Telomere-to-Telomere (T2T) コンソーシアムによる完全ヒトゲノムの解読にも貢献し、Method of the year 2022 に選ばれるなど注目を集めています。この技術においても、解析を行いたい領域のみをシーケンシング前に濃縮するターゲットエンリッチメントの手法は、コストパフォーマンスを改善し、スループットを大幅に向上できるものと考えられます。本アプリケーションノートでは、アジレント・テクノロジー社の SureSelect XT HS2 target enrichment system のプロトコルを一部改変することで、ロングリードシーケンシング向けのターゲットエンリッチメントが可能なお示しします。この改変されたメソッドにより、ターゲットエンリッチメントを行った 5 kb 程度のインサート長のライブラリを得ることが可能で、ナノポアシーケンシングの結果、良好なオンターゲット率、ターゲットカバレッジが得られました。

はじめに

ロングリードシーケンシングはひとつつながりのリードで十数 kb にわたり塩基配列を読むことが可能なテクノロジーで、従来のショートリードシーケンシングでは解析の難しい、リピート配列を含む領域やファミリー間あるいは偽遺伝子と配列が類似している遺伝子、複雑な構造変化などの解析に強みを発揮します。ヒトゲノムの難読領域の解読に貢献したこともあり、Nature Methods 誌の 2022 年 Method of the year にも選出されました¹。しかしながら、ショートリードシーケンシングに比較するとスループットが低い、コストが高いという課題も指摘されており、より効率の良い方法の開発が求められています。

ターゲットキャプチャシーケンシングは、特定のゲノム配列から興味ある領域・遺伝子のみをあらかじめ物理的に濃縮してからシーケンシングを行うことで効率の良い解析が可能なテクノロジーです。これまでにおもにショートリードシーケンシングを対象に開発、実装されています。

アジレントの SureSelect ターゲットエンリッチメントシステムは RNA プローブによるハイブリダイゼーションを用いて対象配列をキャプチャするテクノロジーで、ショートリードシーケンシング向けに最適化されています。これまでに、断片化など若干の条件を変更して Pacific Biosciences 社、Oxford Nanopore Technologies 社のシーケンサ向けにターゲットエンリッチメントを行うプロトコルを公表していますが、キャプチャされるインサート長が 2 kb 程度と比較的短く、スタート DNA 量が多いという課題も残っています²。

本アプリケーションノートでは、追加でいくつかの条件を変更することで、500 ng のインプット DNA を用いて N50 で 5.3 kb のインサート長を実現することができましたので、プロトコル詳細と結果を紹介します。

材料と方法

ワークフロー全体を Fig 1 に示します。プロセスの全体を通して、サンプルの混合はピペティングで行い、ボルテックスによる長鎖 DNA へのダメージを避けるよう配慮しました。さらに詳細なステップバイステップのプロトコルについては、ご請求いただければご提供可能です。

1) DNA 断片化

OneSeq Human reference DNA、female (型番 5190-8850) を g-TUBE (Covaris、型番 52009) を用いてプロトコルに従い 10 kb をターゲットに断片化しました。15 µg の DNA を 150 µL に希釈して g-TUBE に移し、g-TUBE を Eppendorf 社の 5415R 遠心機を用いて 7200 rpm で 1 分遠心しました。続けて g-TUBE を逆向きにして再度同じ条件で遠心しました。g-TUBE の取り扱い説明書によると、インプット量が異なる場合、同じ 5415R 遠心機により 8 µg の DNA を 150 µL に希釈して 6000 rpm で 1 分、あるいは 4 µg 未満の DNA を 150 µL に希釈して 6000 rpm 1 分の遠心、それぞれチューブを逆向きにして同条件の遠心を行うことで 10 kb に断片化することができると記載されています。また、g-TUBE は最小 100 ng/150 µL での断片化まで検証されているとのことです。

断片化した DNA は 5200 Fragment Analyzer Systems (型番: M5310AA) でショートキャピラリー (型番: A2300-1250-3355) と HS Large Fragment 50kb Kit (型番: DNF-464-0500) を用いて分析し、サイズを確認しました。

2) ライブラリ調製

末端修復、dA 付加、アダプターライゲーションは SureSelect XT HS2 試薬キットを用いプロトコルに従い行いました。断片化した DNA 500 ng を用い、アダプターライゲーション後は 0.8x 容量の AMPure XP ビーズで SureSelect XT HS2 のプロトコルに従い精製しました。

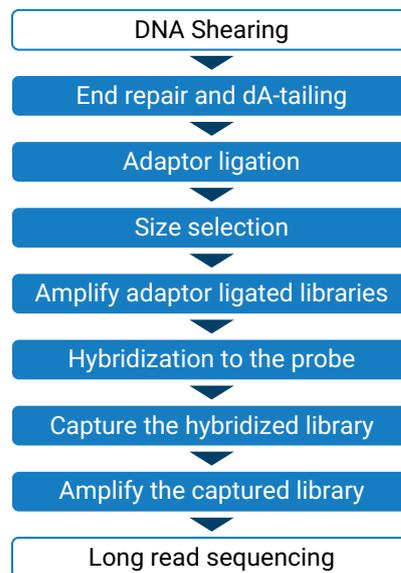


Fig.1 Workflow of target captured long-read sequencing library preparation

Table 1. SureSelect XT HS2 のワークフローに追加に必要な試薬

Materials	Vender	型番	Notes
g-TUBE	Covaris (エムエス機器)	52009	
10N NaOH	Sigma Aldrich	72068	同等品も可
1M Tris HCl pH 8.0	Sigma Aldrich	T3038	同等品も可
KOD FX Neo	Toyobo	KFX201	
Dynabeads M-270 Streptavidin	Thermo Fisher Scientific (ベリタス)	DB65305	
Native Barcoding Kit 24 V14	Oxford Nanopore Technologies	SQK-NBD114.24	

*SureSelect XT HS2 試薬も含む、必要な試薬・機器リストは Reference をご参照ください。

3) サイズセレクション

サイズセレクションは PacBio 社のサイズセレクションプロトコル³に従って行いました。AMPure XP ビーズを 10 mM Tris HCl pH 8.0 により 35% v/v に希釈し、その希釈したビーズ 129.5 µL をサンプルに加えました。ビーズに結合させるためのインキュベーションは室温で 15 分行い、上清を除去したのちに 80% エタノールで 2 回洗浄し、19 µL の 10 mM Tris HCl pH8.0 で室温で 2 分インキュベーションして溶出しました。上清を 18 µL、新しいチューブに回収して増幅に使用しました。

4) アダプター付きライブラリの増幅

PCR 増幅には、長鎖 DNA の増幅に定評のある KOD FX Neo (TOYOBO) を使用しました。試薬の調製とサーマルプロファイルを Table 2、3 に示します。

増幅したライブラリは 50 µL (対サンプル 0.5x ボリューム) の AMPure XP ビーズと混合し、室温で 15 分インキュベーションしました。ビーズを 80% エタノールで 2 回洗浄し、22 µL の 10 mM Tris HCl pH8.0 で懸濁し室温で 2 分インキュベーションしました。上清 20 µL を新しいチューブに回収しました。

プレキャプチャライブラリは 5200 Fragment Analyzer Systems で HS Large Fragment DNA Kit を使用して分析し、サイズ分布と収量を確認しました。

5) ハイブリダイゼーション

増幅したライブラリ 500 ng をハイブリダイゼーションに使用しました。キャプチャプローブは、ショートリード向けに設計された CGP Assay probe (2.671Mbp、Design ID : A3416642) を使用しました。

SureSelect XT HS2 試薬では、Fast ハイブリダイゼーション⁴と、Overnight ハイブリダイゼーション⁵の 2 通りの方法が選択可能です。このアプリケーションノートでは、両方のハイブリダイゼーション条件を試行しました。また、SureSelect XT HS2 の通常のプロトコルではハイブリダイゼーションの最初のステップで 95°C で 5 分インキュベーションして 2 本鎖 DNA を変性させていますが、長鎖 DNA にダメージを与える可能性を考慮して、95°Cでのインキュベーションを 1 分または 30 秒としました (Table 4、5)。そのほかの条件は SureSelect XT HS2 のプロトコルに従いました。

Table 2. プレキャプチャ PCR 反応液の組成

Reagent	Volume/1 Rxn
2x PCR buffer for KOD FX Neo	50 µL
2mM dNTPs	20 µL
SureSelect XT HS2 index primer	10 µL
KOD FX Neo (1.0U/µL)	2 µL
Adaptor ligated DNA samples	18 µL

Table 3. プレキャプチャ PCR のサーマルプロファイル

Step	# of cycles	Temp.	Time
1	1	94°C	2 min.
2	10*	94°C	15 sec.
		60°C	30 sec.
		68°C	15 min.
3	1	68°C	10 min.
4	1	4°C	Hold

*インプット DNA 量に応じて調整が必要

Table 4. Fast ハイブリダイゼーションのサーマルプロファイル

Step	# of cycles	Temp.	Time
1	1	95°C	1 min. / 30 sec.
2	1	65°C	10 min.
3*	1	65°C	1 min.
4	60	65°C	1 min.
		37°C	3 sec.
5	1	21°C	Hold (up to 16h)

*Pause to add probe-hybridization mix

Table 5. Overnight ハイブリダイゼーションのサーマルプロファイル

Step	# of cycles	Temp.	Time
1	1	95°C	1 min. / 30 sec.
2	1	65°C	Hold (for 16-19h)

6) ハイブリダイゼーションしたライブラリのキャプチャ

ストレプトアビジンビーズの調製

ハイブリダイゼーション後のライブラリのキャプチャには Dynabeads M-270 Streptavidin (Thermo Fisher Scientific) を用いました。ビーズ 50 μ L に 200 μ L の Binding buffer を加え、計 3 回洗浄し、最終的に 200 μ L の Binding buffer に懸濁しました。懸濁したビーズは次のキャプチャステップの直前に 68°C で 10 分予熱しました。また、Wash Buffer 2 の 70°C での予熱を開始しました。

Capture

ハイブリダイゼーションのサーマルプロファイルの終了後に、サンプルは 68°C で予熱した Streptavidin ビーズと混合し、68°C で 5 分インキュベートしました。マグネットでビーズを集め、上清を除去しました。

Wash 1

DNA が結合したビーズを 200 μ L の Wash Buffer 1 にピペティングにより懸濁しました。マグネットでビーズを集め、上清を除去しました。

Wash 2

Wash 1 で洗浄したビーズを 500 μ L の 70°C に予熱した Wash Buffer 2 に懸濁し、70°C で 5 分間インキュベーションしました。マグネットでビーズを集めて上清を除去し、再度 70°C の Wash Buffer 2 に懸濁し、合計で 3 回繰り返しました。

3 度目の Wash ステップで、まず 400 μ L の上清を除去したあとで、マグネットからはずしのこりの Wash 2 buffer にビーズを再懸濁し、全量を 0.2 mL の PCR チューブに移しました。再度マグネットでビーズを集め、上清を除去しました。

Elution

DNA が結合したビーズを 18 μ L の用事調製した 0.1M NaOH に懸濁し、室温で 10 分インキュベーションしました。次に 6 μ L の 600 mM Tris HCl (pH8.0) を加えて中和し、マグネットでビーズを集めました。上清を 24 μ L 回収し、新しい PCR チューブに移しました。

Table 6. ポストキャプチャ PCR 反応液の組成

Reagent	Volume/1 Rxn
2x PCR buffer for KOD FX Neo	50 μ L
2mM dNTPs	20 μ L
SureSelect Post-Capture Primer Mix	4 μ L
KOD FX Neo (1.0U/ μ L)	2 μ L
Captured libraries	24 μ L

7) キャプチャしたライブラリの増幅

キャプチャしたライブラリの増幅には KOD FX Neo (TOYOBO) を使用しました。PCR 反応の組成とサーマルプロファイルを表 6、7 に示します。

増幅後にサンプルを 50 μ L (サンプルの 0.5x 容量) の AMPure XP ビーズと混合し、室温で 15 分間インキュベーションしました。ビーズを 80% エタノールで 2 回洗浄し、42 μ L の 10 mM Tris HCl pH8.0 で懸濁して 2 分間インキュベーションして溶出しました。上清を 40 μ L 回収して新しいチューブに移しました。

増幅後のライブラリは 5200 Fragment Analyzer Systems で HS Large Fragment DNA Kit を用いて分析しました。Nanopore のアダプターライゲーション前に、Qubit dsDNA BR Assay Kit (Thermo Fisher Scientific) により定量しました。

8) ロングリードシーケンシングとデータ解析

ロングリードシーケンシング用のライブラリは Native Barcoding Kit 24 V14 (Oxford Nanopore Technologies、型番: SQK-NBD114.24) を用いてプロトコルに従いマルチプレックス用に調製しました。8 サンプルを混合して PromethION Flow Cell (FLO-PRO114M) にロードしました。シーケンスデータは Dorado 7.2.13 を用いて、ベースコールモデル dna_r10.4.1_e8.2_400bps_sup@v4.2.0 によりベースコールしました。重み付きヒストグラムの描画には Nanoplot 1.42.0 を用いました。

シーケンスデータはサンプル当たり 2.9M リードにノーマライズして以降の解析に使用しました。Nanopore と Illumina のアダプター配列は Porechop 0.2.4 と cutadapt 2.6 を使用して除去し、トリミング後のリードを Minimap 2-2.26 を用いて hg38 レファレンスゲノムにマッピングしました。% on-target、mean-coverage、fold enrichment、% coverage、AT/GC dropout を Picard 3.1.1 HsMetrics (Broad Institute) により計算しました。データ解析のワークフローについて、ステップバイステップのガイドラインもご提供可能です。

Table 7. ポストキャプチャ PCR のサーマルプロファイル

Step	# of cycles	Temp.	Time
1	1	94°C	2 min.
2	16*	94°C	15 sec.
		60°C	30 sec.
		68°C	15 min.
3	1	68°C	10 min.
4	1	4°C	Hold

* キャプチャプローブのサイズに応じて変更の必要あり

結果

断片化DNA、プレキャプチャ、ポストキャプチャライブラリのサイズ分布と収量

g-TUBE により断片化した DNA のサイズ分布は 5200 Fragment Analyzer System で HS Large Fragment Kit を用いて分析しました (Fig 2A)。3000 bp 未満のフラグメントはこのあとのサイズセレクションのステップで除去されるので、このステップでサイズ分布を確認することを推奨します。3000 bp 未満のフラグメントを対象にする場合やこの時点で 3000 bp 未満のフラグメントが多い場合には、インプット量や PCR サイクル、サイズセレクションの方法などについての最適化が必要になるものと考えられます。

アダプターライゲーションを行って増幅したサンプルのサイズ分布は、断片化 DNA と比較すると 1-1.5 kb 短くなっていました (Fig 2B and Table 7)。サイズセレクションをしなかった場合、おそらく短いフラグメントの増幅が優先されたためにサイズ分布が短いかわりにシフトしました (データ未掲載)。Pre-Capture Library #1 を以降のハイブリダイゼーションに使用しました。ハイブリダイゼーション後のサイズ分布は、プレキャプチャライブラリの分布よりもさらに 1 kb 短くなりました (Table 8)。

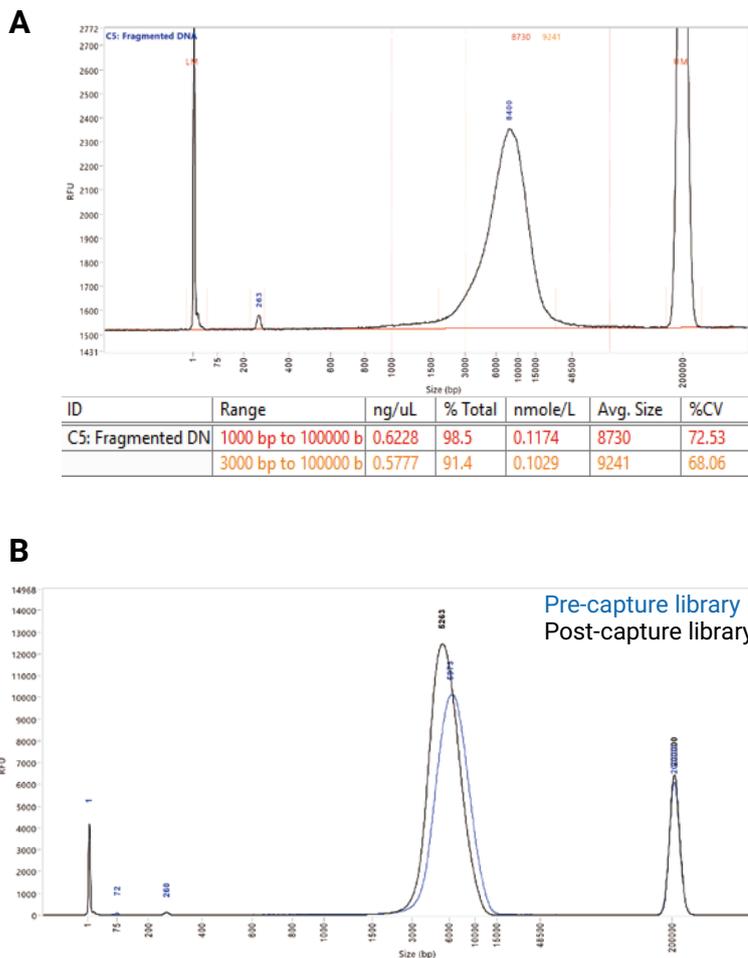


Fig.2 サイズ分布の例
断片化 DNA (A)、プレキャプチャおよびポストキャプチャライブラリ (B)

Table 8. プレキャプチャ、ポストキャプチャライブラリ QC のサマリー

Pre-Capture Library				
Sample #			Average size (bp)	Yield (ng)
1			6,499	4,815.2
2			6,680	6,502.4
3			7,712	5,856.4
4			6,241	7,130.6
Post-Capture Library				
Sample #	Hybridization option	Denaturing duration	Average size (bp)	Yield (ng)
1	Overnight	1 min.	5,455	2,425.5
2	Overnight	30 sec.	5,388	2,779.8
3	Fast	1 min.	5,483	1,939.2
4	Fast	30 sec.	5,623	2,637.0

シーケンス結果

キャプチャして増幅したライブラリは Nanopore アダプターとライゲーションし、PromethION でシーケンスを行いました。シーケンスメトリクスを Table 9 に示します。シーケンス長の N50 はおよそ

4.9–5.4 kb で、ポストキャプチャライブラリの電気泳動結果とも一致する結果でした (Fig 3)。ハイブリダイゼーション方式間の比較では、Fast ハイブリダイゼーションでより長いフラグメントが保存される傾向がありました。また変性時間の比較では

30 秒の方が長いフラグメントが保存される傾向がありましたが、いずれの実験条件も 1 サンプルずつのみの実施であるため、統計的な情報は得られていません。

Table 9. シーケンス結果のサマリー

	Overnight		Fast	
	1 min.	30 sec.	1 min.	30 sec.
No. of reads	3,176,795	3,850,233	3,340,509	3,351,175
Total bp	15,027,621,729	18,619,540,009	16,650,843,463	17,133,017,839
N50	4,947	5,064	5,229	5,356
1d max length (bp)	408,325	442,635	439,222	574,014
1d average length (bp)	4,730	4,836	4,985	5,113
1d median length (bp)	4,575	4,676	4,827	4,941
average qscore	15.2	15.1	14.8	15.0
1d pass reads (mean_qscore >= 10)	2,909,038	3,523,069	2,984,778	3,060,786
1d pass max length	408,325	401,776	270,356	331,457
1d pass average length	4,733	4,840	4,995	5,119
1d pass average qscore	15.9	15.8	15.7	15.7

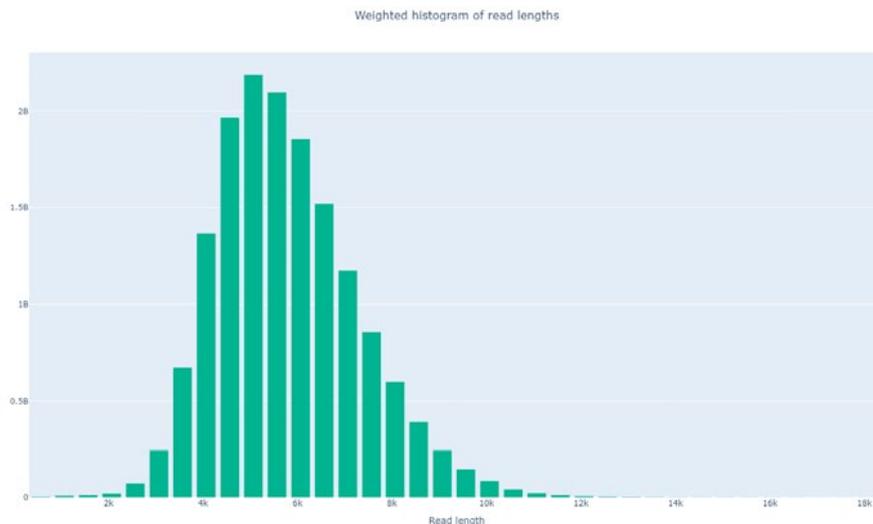


Fig.3 リード長の重みづけヒストグラム
Fast ハイブリダイゼーション、30 秒の変性条件で得られた 1d pass リードのリード長分布

ターゲットエンリッチメントの効率

それぞれのシーケンスデータは2.9Mリードにノーマライズし、Picard HsMetricsでシーケンスメトリクスを評価しました。いずれの実験条件でも On-target パーセントは70%を超えており、Fast ハイブリダイゼーションの方が高い値を示しました (Fig 4A)。平均カバレッジ、10x、20x、50x、100x カバレッジ % についても Fast ハイブリダイゼーションが高い値となりました (Fig 4B、4C)。

次に AT/GC Dropout については、overnight ハイブリダイゼーションが AT

dropout について良好な値となりましたが、GC dropout については Fast ハイブリダイゼーションの方が良好な値となり、両者のバランスを考えると Fast ハイブリダイゼーションのほうがバランスが取れたパフォーマンスであるといえます (Fig 4D)。今回使用した CGP プローブはショートリードで Fast ハイブリダイゼーションに最適化されたプローブなので、ハイブリダイゼーション間のパフォーマンスの差はここに起因する可能性もあります。プローブの設計変更などで改善の余地があるのではないかと考えています。

また、ハイブリダイゼーションの変性ステップ (30 秒もしくは1分) については、パフォーマンス上顕著な差は見られませんでした。30 秒の変性では短すぎて、GC 含量の高い領域のリードが減少するのではないかと懸念を抱いていましたが、実際には GC dropout の値には顕著な差はなく、変性時間として十分であると考えられます。インサート長に関しては30秒での変性条件のほうが若干長かったため、これらをあわせて本アプリケーションにおいては30秒の変性条件が適していると考えました。

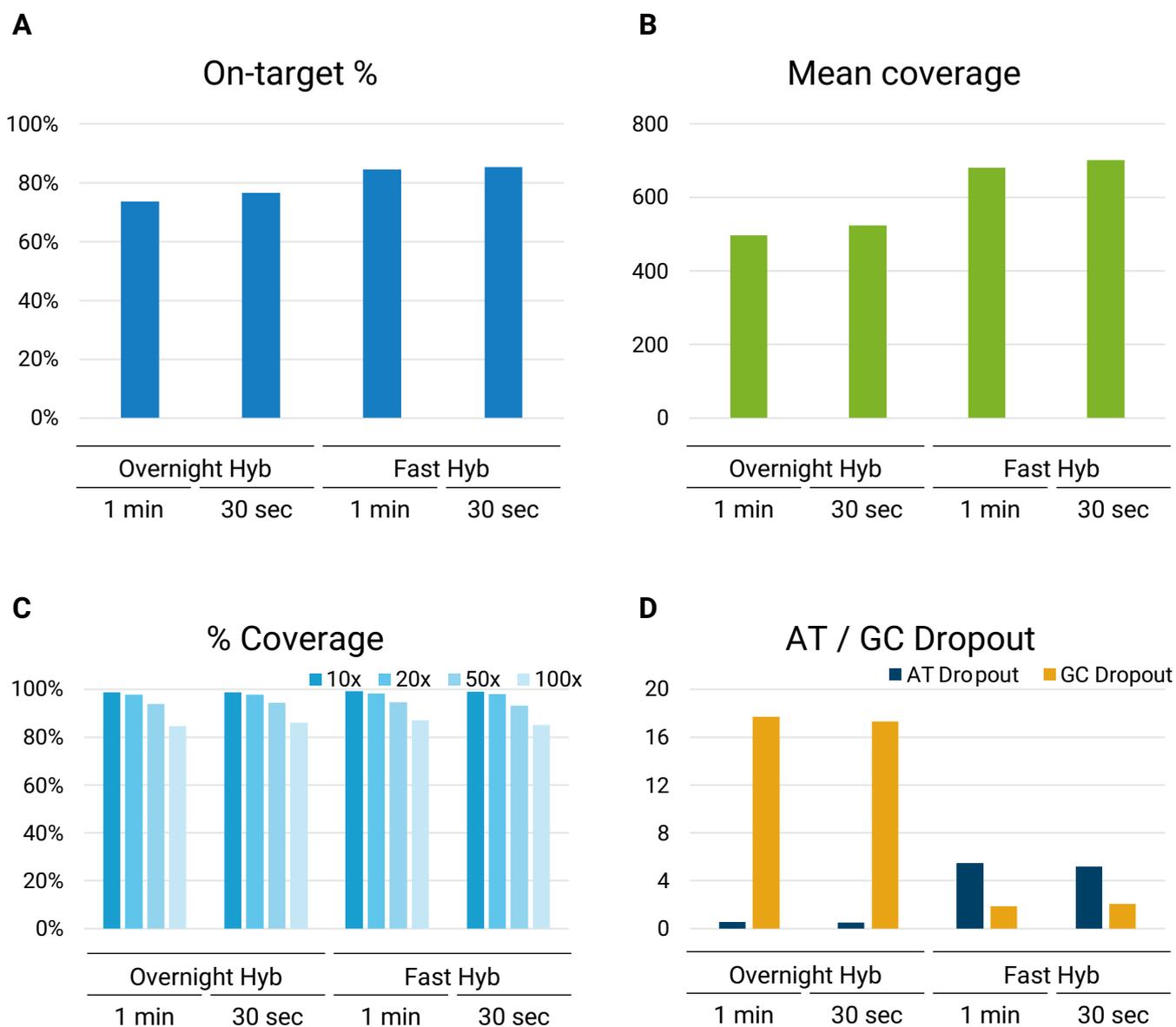


Fig.4 ターゲットエンリッチメントのメトリクスのサマリー

ターゲット領域の例

ターゲット領域のカバレッジの例を Fig 5 に示します。シーケンスメトリクスの中で on-target % が 70% 以上であったことから期待されるとおりに、各ターゲット領域とその周辺にリードが集積しており、ターゲットエンリッチメントが効率よくできていると考えられます。

ヒトゲノムの中には、GC リッチな領域、偽遺伝子との区別、リピート配列などショートリードシーケンスでは解析が難しい領域が知られています。その中でも、TERT プロモーター領域は C228T、C250T など特定のがん種でドライバー変異が知られていますが、この領域が GC リッチであるためショートリードシーケンスでのターゲットエンリッチメントではカバレッジが低くなりがちです。今回のロングリードでのターゲットエンリッチメントのデータでは、該当領域で十分なリード数によるカバレッジが得られました (Fig 5 and 6)。また、別の例として RB1 遺伝子のエクソン 15 は poly-A、poly-T に領域がはさまれているため、ショートリードシーケンスでの難読領域として知られていますが、こちらも今回のデータで十分なカバレッジが得られました (Fig 5 and 6)。

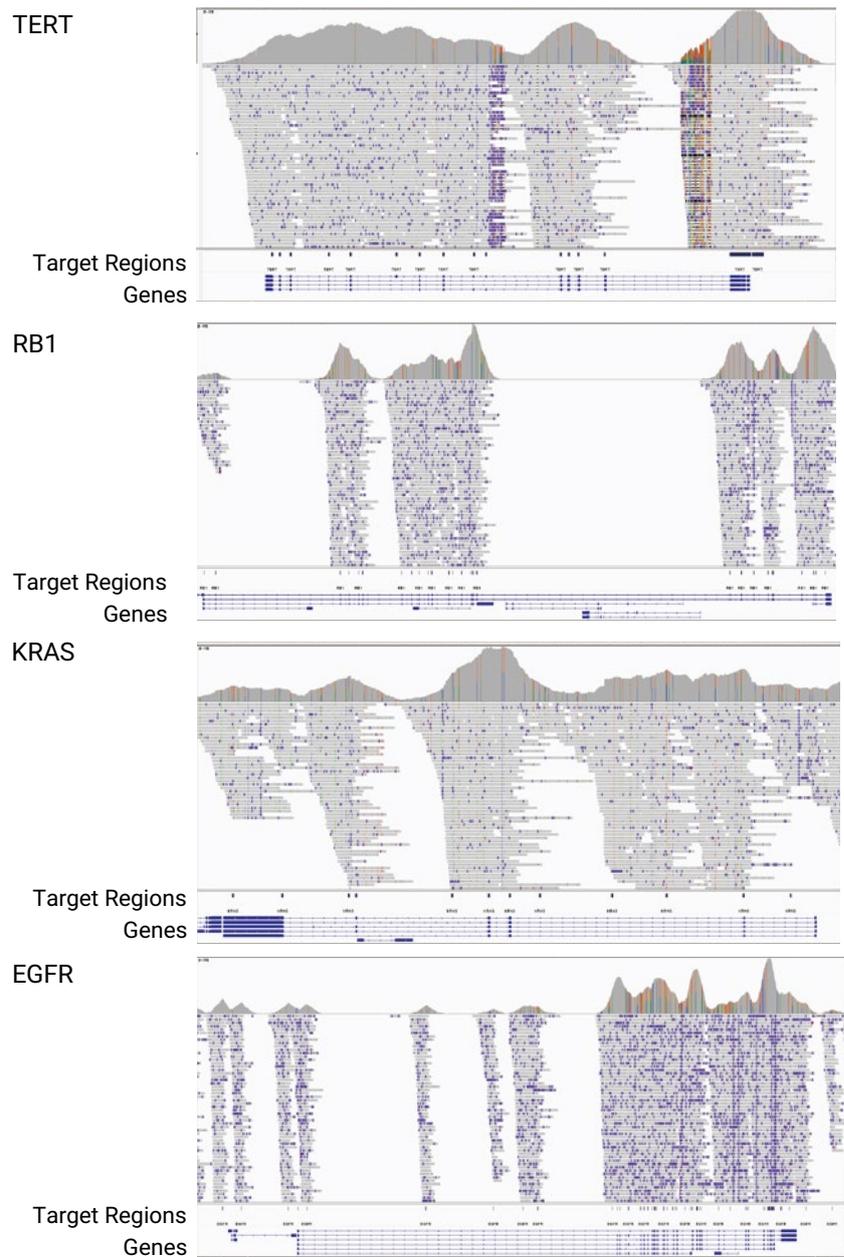


Fig.5 ターゲット領域の例

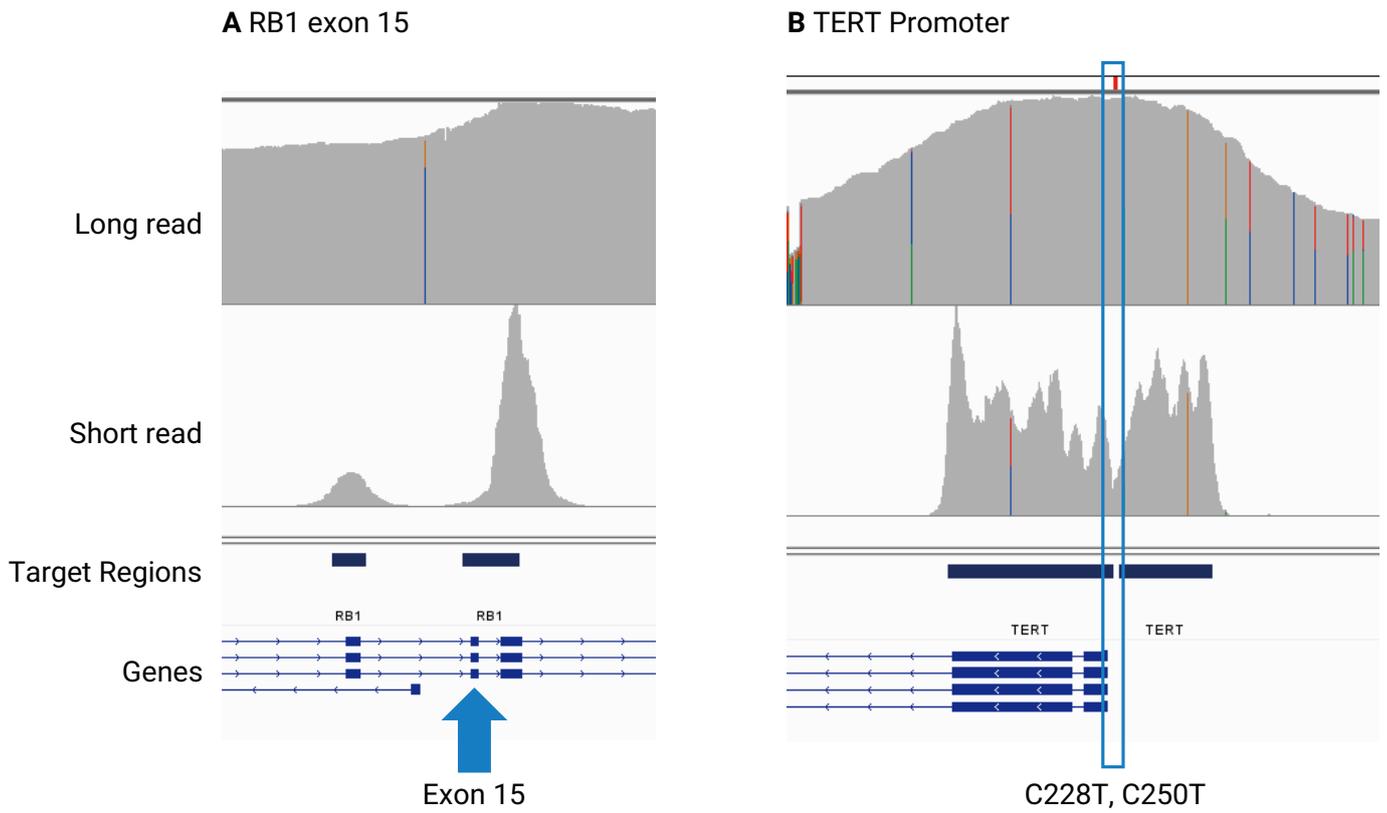


Fig.6 難読領域のカバレッジの例
 ショートリードのライブラリは、Agilent female reference DNA を用いて、SureSelect Cancer CGP Assay プローブにより SureSelect XT HS2 の標準プロトコルでターゲットエンリッチメントを行い、HiSeq4000 の 150 bp ペアエンドでシーケンスしました。

結論

本アプリケーションノートでは、SureSelect XT HS2 プロトコルを KOD FX Neo (DNA ポリメラーゼ) と組み合わせて変更を加えることにより、長いフラグメント (5-6 kb) のターゲットエンリッチメントを高い on-target %, 良好な均一性で実施できることを示しました。また、ショートリード向けにデザインされたキャプチャプローブを変更なしにそのまま使用して、長いフラグメントのターゲットエンリッチメントが可能であることも示しています。今後、キャプチャのパフォーマンスを向上させる、あるいはコスト効率のよい設計など、ロングリードに最適化したプローブデザインのさらなる検討が望まれます。本メソッドは難読領域のいわゆる“Dark gene”や HLA 領域などのハプロタイピング、ファーマコゲノミクス関連遺伝子、完全長 cDNA、メチル化解析、病原体ゲノムなどのターゲットエンリッチメントにも適用できる可能性があります。さらに、最終段階のアダプターを変更することで、PacBio シーケンシングにも適用できると考えています。この方法により、ロングリードシーケンスでもターゲットエンリッチメントを行い、1 ランあたりマルチプレックスでの解析やスループットの低いフローセルを用いることで、よりコスト効率の良い解析が可能になるものと期待しています。

Reference

1. Method of the Year 2022:long-read sequencing. Nat Methods 20, 1(2023). <https://doi.org/10.1038/s41592-022-01759-x>
2. Application Note:Use of Agilent SureSelect to perform targeted long-read nanopore sequencing. Agilent technologies application note, publication number 5991-8056EN, 2017
3. Procedure & Checklist–Using AMPure® PB Beads for Size-Selection Pacific Biosciences Part Number 101-854-900 Version 02(<https://www.pacb.com/wp-content/uploads/Procedure-Checklist-%E2%80%93Using-AMPure-PB-Beads-for-Size-Selection.pdf>)
4. SureSelect XT HS2 DNA Kits with Library Preparation(+/-MBC)/ Fast-Hyb Target Enrichment /Post-capture Pooling Workflow, Agilent Technologies user manual, publication number G9983-90000, 2023.
5. SureSelect XT HS2 DNA Kits with Library Preparation(+/-MBC)/ Overnight-Hyb Target Enrichment /Post-capture Pooling Workflow, Agilent Technologies user manual, publication number G9956-90000, 2023.

[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。
※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。

<http://www.agilent.com/chem/genomics:jp>

G240631

© Agilent Technologies, Inc. 2024

本書の一部または全部を書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, April 25, 2024
5994-7365JAJP