

illumina Complete Long Read Prep with Enrichment, Humanのための カスタムパネル デザイン

ヒトゲノムに対応した
柔軟性の高い、ターゲット
ロングリード濃縮

illumina®

はじめに

ヒト全ゲノムシーケンス (WGS) を実施する場合、ショートリードだけではマッピングが難しい一部の遺伝子領域がある場合があります。ロングリードシーケンスは、これらの解析が困難な領域に対処するために、標準的なショートリードWGSデータを補完することができます。Illumina Complete Long Readテクノロジーは、標準の次世代シーケンス (NGS) ワークフローを使用して、イルミナシーケンスシステムで連続したロングリードシーケンスを生成し、単一の解析パイプラインでデータを解析します (図1)。¹⁻³ Illumina Complete Long Read Prep with Enrichment, Humanによりターゲットアプローチを追加し、よりコスト効率の良いロングリードシーケンスを実現します。* Illumina Complete Long Read濃縮ケミストリーは、ターゲットとプローブデザインに高い柔軟性を提供し、マッピングが難しい領域の解析を支援し、フェージングしたシーケンスから追加の洞察を提供します。

ロングリード用の濃縮プローブパネルのデザイン

Illumina Complete Long Read Prep with Enrichment, Humanは、短い断片 (約200~500 bp) をキャプチャーするために通常使用されるアプローチとは異なるプローブデザイン方針を使用して、長い断片 (約7~10 kb) をキャプチャーします。イルミナ DesignStudio™ソフトウェアは、濃縮プローブパネルをデザインす

* 解析には同一サンプルからの30×カバレッジ以上の標準的なショートリードWGSデータが必要です。以前にランを実施したサンプルのFASTQファイルを使用できます。

るための無料で使いやすいツールです。DesignStudioアルゴリズムでは、GC含有量、ターゲット特異性、およびプローブ間隔 (ターゲット領域にプローブがいくつあるか) が考慮されます。120merのショートリード濃縮パネルの標準間隔は250~350 bpのプローブ幅です。ロングリード濃縮パネルのデザインについて、プローブ間隔を複数の長さでテストしたところ、コスト効率が良く、高性能のキャプチャーには1キロベースの幅が最適であることがわかりました。

ハイブリダイゼーション濃縮の有効性は、プローブの特異性に大きく依存します。オンターゲット濃縮の割合は、ターゲットのカバレッジ深度を達成するために必要なシーケンスの量に直接影響します。繰り返し領域では高い特異性を達成するのが難しくなります。ただし、より大きなプローブ幅を使用すると、パフォーマンスの悪いプローブを除外し、繰り返し領域 (最大幅1 kb) を回避し、より少ないプローブで濃縮効率を維持するための柔軟性が高まります (図2)。DesignStudioアルゴリズムは、これらの考慮事項を基にプローブの配置を推奨します。サードパーティー社製のパネルは、最高の性能とコスト効率を実現するために、同様のガイドラインを使用しているはずで、標準の濃縮プローブの間隔も完全に互換性があります。

プローブデザインとターゲット方針の柔軟性

Illumina Complete Long Read Prep with Enrichment, Humanは、研究目的に合わせてカスタムプローブパネルを選択およびデザインする高い柔軟性があります。個々のターゲット領域は、1塩基から最大数百キロベースに及び可能性があります。パネルの合計サイズは、わずか2.5 Mbのカスタムパネルから95 Mbを超えるカスタムパネルまでの範囲まで対応できます。研究者は、ターゲットロングリードを使用して、ショートリードデータのマッピングの可能性が低いこ



図1: 統合ワークフローの一部: 濃縮プロトコル、実績のあるイルミナシーケンスケミストリー、およびDRAGEN二次解析を用いたスケラブルで最適化されたライブラリー調製を使用して、コスト効率の良いターゲットロングリードWGSデータが得られます。解析には同一サンプルからの30×カバレッジ以上の標準的なショートリードWGSデータが必要です。以前にランを実施したサンプルのFASTQファイルを使用できます。

とが知られている特定の領域全体のカバレッジを向上させることができます。あるいは、ロングリードを遺伝子全体から長い多重遺伝子領域までをカバーするようにターゲットにし、バリエーションのフェージングとハプロタイプコールを実施することができます。

DesignStudioソフトウェアではいくつかのプレデザインパネルがご利用いただけます(表1)。これらのパネルは、医学的に重要な難読遺伝子 (CMRG)⁴、ファーマコゲノミクス (PGx) 検査アッセイで一般的にターゲットとされる遺伝子⁵⁻⁷、米国臨床遺伝・ゲノム学会 (ACMG) の二次的所見リスト (ACMG SF v3.1) に掲載されている遺伝子⁸、また主要組織適合性複合体 (MHC) 領域全体をターゲットとしています。⁹ タンパク質コーディング遺伝子内の個別の低カバレッジ領域を主にターゲットとするIllumina Human Comprehensive Panelは、プレデザインパネルまたはすぐに出荷できる既製のパネルとしてもご利用いただけます(イルミナカタログ番号: 20113836)。^{10,11} DesignStudioソフトウェアは、BEDファイル[†]からのカスタムパネルまたは既存のプレデザインを変更したカスタムパネルの設計に対応します。

カスタムプローブパネルの推奨シーケンス深度

Illumina Complete Long Read Prep with Enrichment, Human は、一貫性の高いロバストな性能を提供します。テストしたプレデザインパネルでは、ターゲットパネルサイズ1 Mbあたり約1.5 Gbのシーケンスデータ(約500万ペアエンドリード)で最適な性能が達成されました(図3)。性能が未知の新しくデザインされたパネルの場合、ターゲットパネルサイズ1 Mbあたり3 Gbのシーケンスデータ(約1,000万ペアエンドリード)が推奨される開始点ですが、さらなる最適化により削減することができる可能性があります。

解析が困難な領域の高精度のカバレッジとフェージング

Illumina Human Comprehensive PanelやCMRGパネルなど、特定の低カバレッジ領域の向上に重点を置いたロングリード濃縮プローブパネルは、ターゲット濃縮した解析が困難な領域でのバリエーションコールの精度を向上させます(図4)。CMRGパネルを使用したロングリード濃縮は、タンパク質コーディング領域全体のカバレッジとバリエーション検出の完全性を向上させることにも役立ちます(図5、図6)。

† BED=ブラウザの拡張可能なデータ形式。

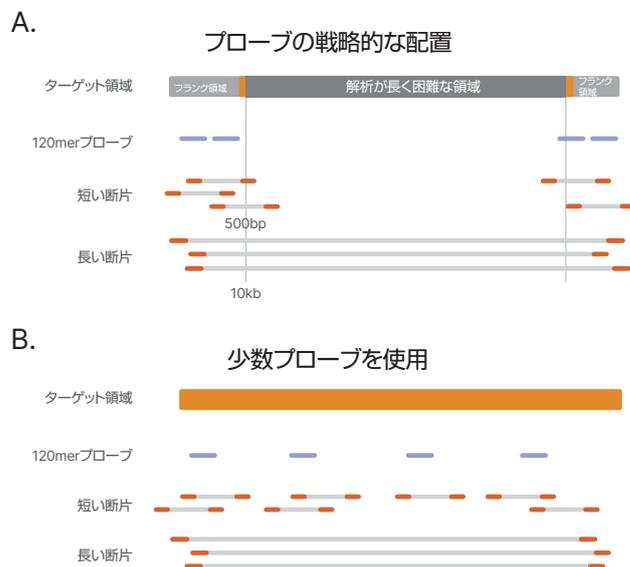


図2: 長い断片のハイブリダイゼーションによる濃縮効率の向上: 長い断片のハイブリダイゼーションには、(A) 極端なGC含有量、低複雑性、繰り返しなどのプローブデザインが難しい領域の外側にプローブを方針的に配置する、(B) 各ターゲット領域をキャプチャーするために必要なプローブの数を少なくするなど、短い断片のキャプチャーに比べて利点があります。DesignStudioアルゴリズムは、1 kb区域内のターゲット領域全体を検索して、プローブを配置するための最適なGC含有量と最も特異性の高い領域を探します。

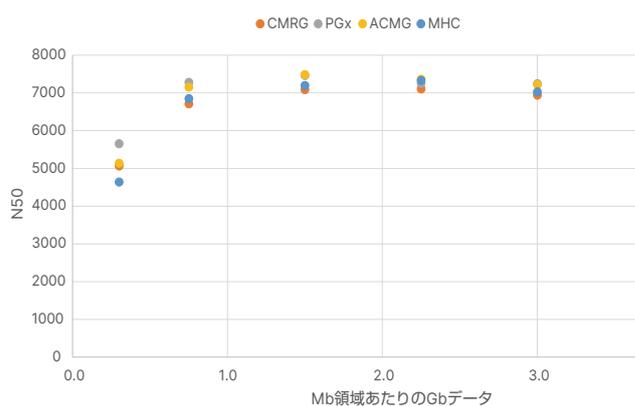


図3: カスタムプローブパネルのシーケンス要件: 最大N50に必要なシーケンスデータの検討より、Mbターゲット領域あたり1.5 Gb(約500万ペアエンドリード)が、Illumina Complete Long Readのデータ生成のターゲット領域を効率的に解析することが示されています。

表1: Illumina Complete Long Read Prep with Enrichment, Human用にプレデザインされた濃縮プローブパネル

パネル ^a	CMRGパネル	PGxパネル	ACMGパネル	MHCパネル
ターゲット遺伝子	ショートリードでは解析が困難であることが知られている391の医学的に関連する遺伝子 ⁵	ファーマコゲノミクス検査アッセイの一般的なターゲットである98遺伝子 ⁶⁻⁸	ACMG二次的所見リスト (ACMG SF v3.1) からの78の固有の遺伝子 ⁹	GRCh38.p14アセンブル内の全MHC領域の140以上の遺伝子 ¹⁰
ターゲット領域のサイズ ^b	22.5 Mb	8.1 Mb	7 Mb	4.9 Mb
サンプルあたりのシーケンス出力 ^c	~67.5 Gb	~24.3 Gb	~21 Gb	~14.7 Gb
プローブ数	~22.5K	~8.2K	~6.9K	~5.0K
N50 ^d	6.1 kb	7.3 kb	7.3 kb	7.3 kb
フェーズブロックN50 ^{d,e}	82.8 kb	94.4 kb	94.4 kb	357 kb
平均ターゲット領域のサイズ ^e	58 kb	83 kb	88 kb	5,000 kb
均一性 ^{d,f}	97.9%	99.0%	99.5%	97.8%
パディングしたリード濃縮 (PRE) ^{d,f}	80.1%	79.3%	66.3%	67.5%
フェージングしたヘテロ接合性SNV% ^d	98.9%	98.9%	99.6%	98.6%

- a. CMRG=医学的に重要な難読遺伝子, PGx=ファーマコゲノミクス, ACMG=米国臨床遺伝・ゲノム学会, MHC=主要組織適合性複合体。
b. ターゲット領域のサイズは、パディングしたプローブ位置の長さの合計であり、重複する部分で結合されます。
c. Illumina Complete Long Readの約30×最終カバレッジを生成するための、150 bp × 2でのシーケンスランおよびMbターゲット領域あたり500万~1,000万のペアエンドリード (約1.5~3 Gbデータ) が必要です。サンプルあたりのカスタムパネルデータ要件は、推奨される出発点にすぎません。ユーザーはパネルの性能に基づいて割り当てられたデータを最適化できません。
d. 50 ng HG002ゲノムDNA (Coriell, カタログ番号: NA24385) を使用して生成されたデータ。性能はDNAインプットとサンプル品質によって異なる場合があります。
e. フェーズブロックのサイズは、個々の連続するターゲット領域のサイズに制限されます。
f. 平均カバレッジに対して0.2以上の時カバレッジの均一性が割合 (%) として計算されます。PREは100* (パディングしたターゲットのアライメントされたリード数/アライメントされた合計リード数) として計算されます。

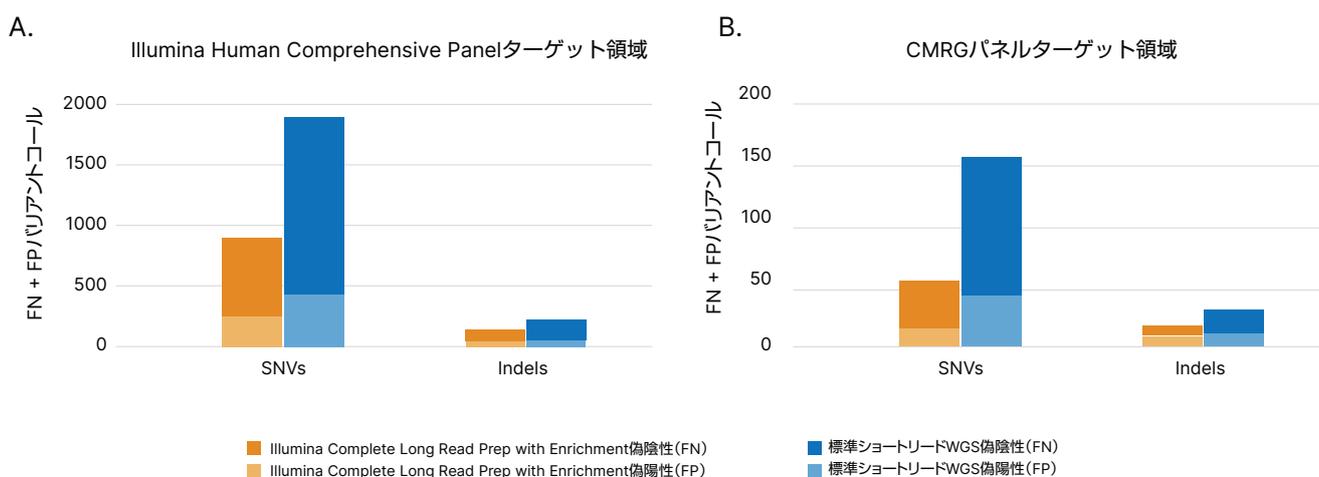


図4: ターゲットロングリードにより、解析が困難な領域でのバリエーションコール精度が向上: (A) Illumina Human Comprehensive Panelまたは (B) CMRGパネルのターゲットであるHG002遺伝子領域の1塩基変異 (SNV) および挿入/欠失 (Indel) に対する偽陰性 (FN) と偽陽性 (FP) のバリエーションコール。Illumina Complete Long Read Prep with Enrichment (オレンジ) と標準的なショートリードWGS (青) 使用時の比較。



図5：ターゲットロングリードによる低カバレッジの領域の拡張：HBG1に対するIntegrative Genomics Viewer (IGV) プロット。Illumina Complete Long Read Prep, Human WGS (上) およびIllumina Complete Long Read Prep with Enrichment, HumanとCMRGパネル (中) を使ったロングリードシーケンスと標準的なショートリードWGS (下) との比較。

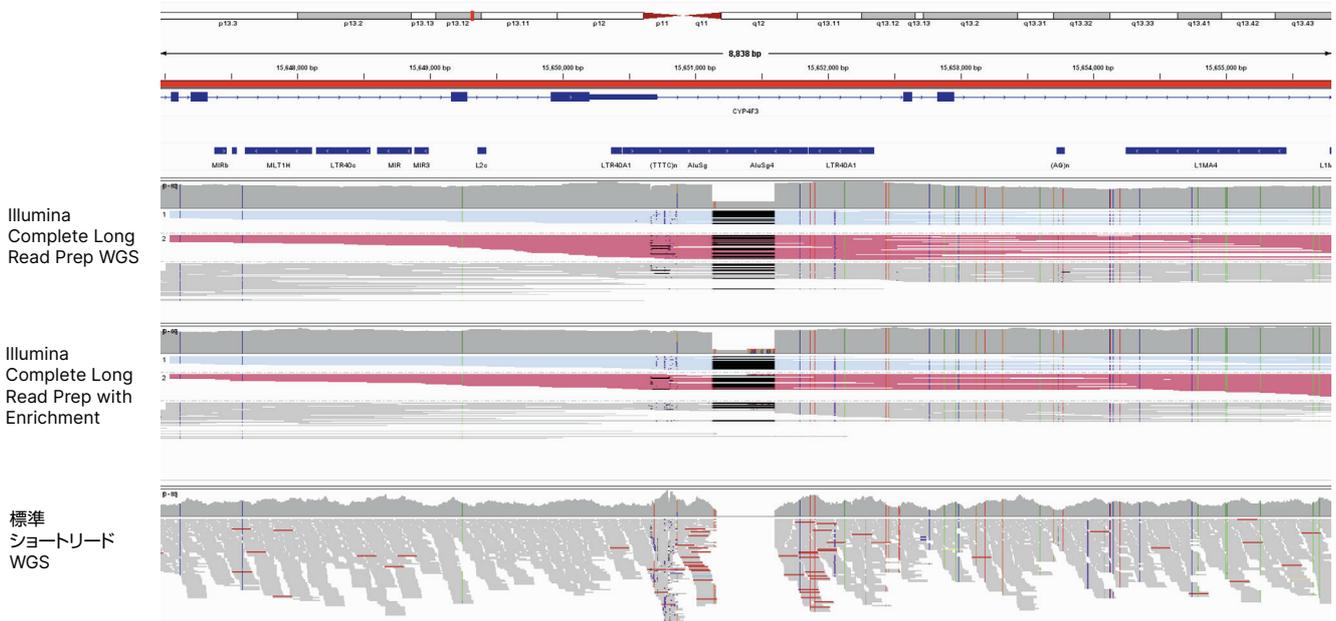
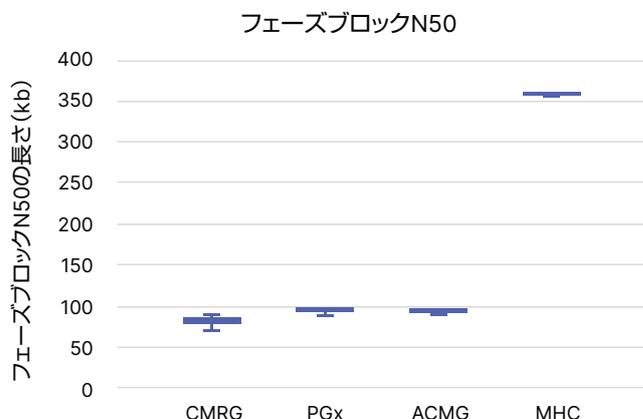


図6：ターゲットロングリードによる欠失境界の解析の向上：CYP4F3に対するIGVプロット。Illumina Complete Long Read Prep, Human WGS (上) およびIllumina Complete Long Read Prep with Enrichment, HumanとCMRGパネル (中) を使ったロングリードシーケンスおよびフェーシングと標準的なショートリードWGS (下) との比較。アリル1は青、アリル2はピンクで示しています。

ハプロタイプを解析するためのロングフェーズブロック

各パネルのフェーズブロックN50[†]は、ターゲット領域の連続した長さに関係します(図7, 表1)。CMRG、PGx、およびACMGパネルは、目的の全長遺伝子をターゲットとするように設計されており、約80~95 kbの平均フェーズブロックN50を生成し、ヘテロ接合性アリルの完全なフェージングを実現します(図8)。MHCパネルは、単一の約4.9 Mbの連続領域をターゲットにし、350 kbを超える平均フェーズブロックN50を生成して、全長遺伝子領域の解析を実現します(図9)。



† フェーズブロックN50は、ターゲット領域の合計アセンブリ長の50%にある連続シーケンスの最も短いブロックの長さを反映します。

図7: フェーズブロックN50は、連続するターゲット領域の長さに依存: CMRG、PGx、およびACMGパネルは目的の全長遺伝子をターゲットにし、平均フェーズブロックN50は約80~95 kbになります。MHCパネルは、主要組織適合性複合体領域全体をターゲットにし、350 kbを超える平均フェーズブロックN50を生成します。平均ターゲット領域のサイズは、CMRGパネルで58 kb、PGxパネルで83 kb、ACMGパネルで88 kb、MHCパネルで5,000 kbです。

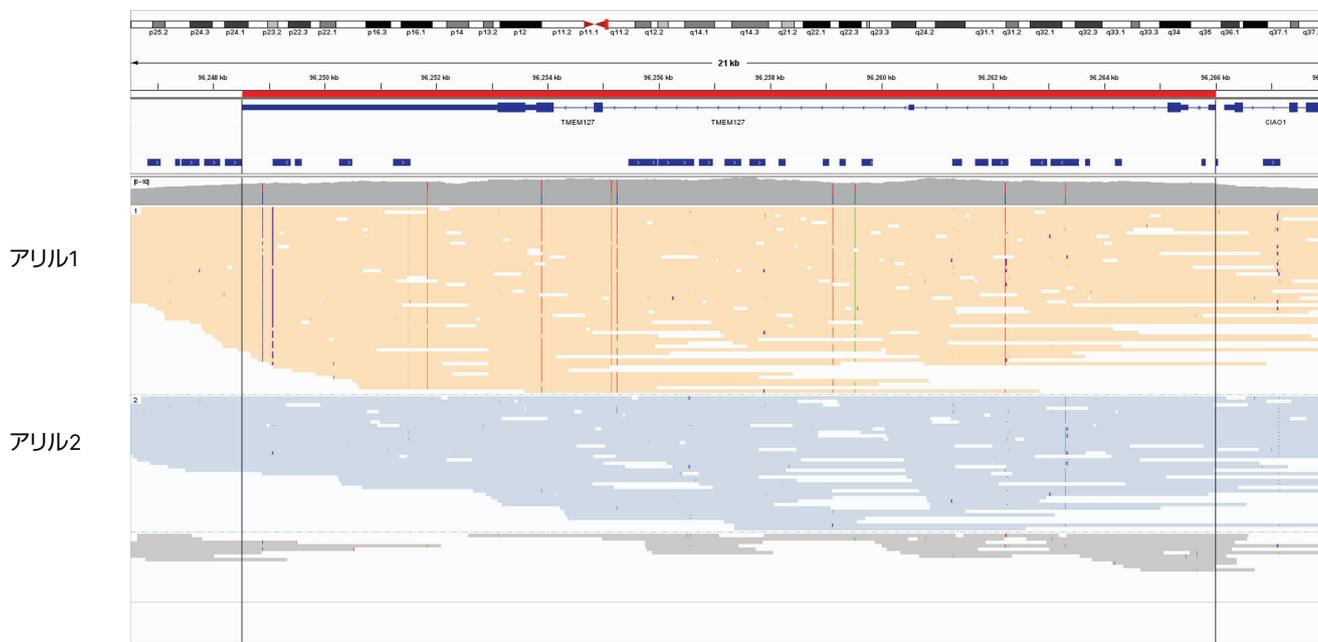


図8: ターゲットロングリードは、ヘテロ接合性SNVを含む領域のフェージングが可能: ロングリードシーケンスからのIGVプロット。Illumina Complete Long Read Prep with Enrichment, HumanとACMGパネルを使用したTMEM127 (21 kbの遺伝子) に対する1つのフェーズブロックに対する完全なフェージングを示しています。黄はアリル1。青はアリル2。

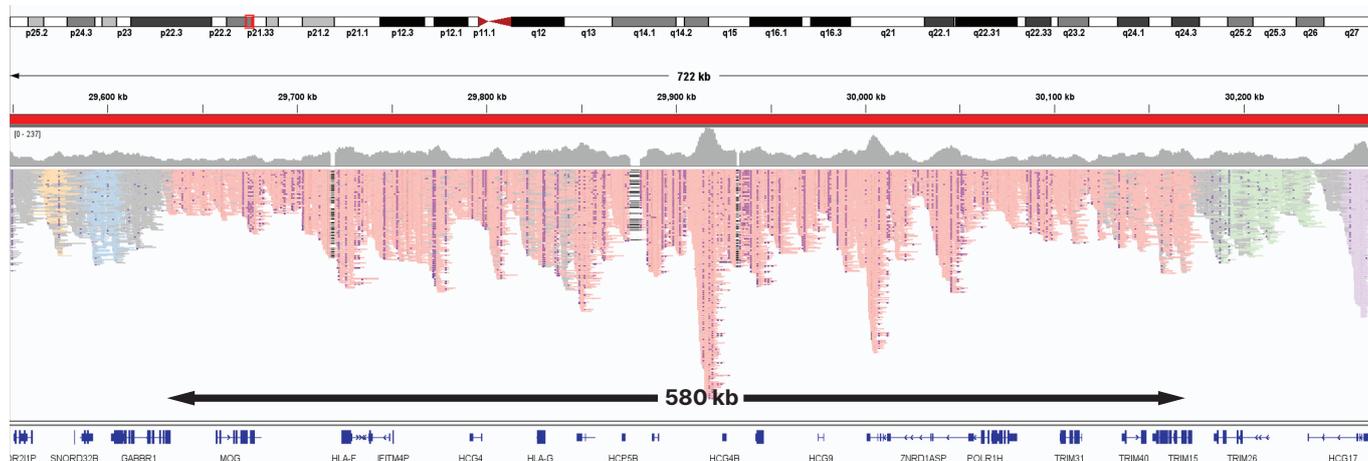


図9：ターゲットロングリードは、多型遺伝子のハプロタイプの解析に役立ちます。Illumina Complete Long Read Prep with Enrichment, Humanを使用したロングリードシーケンスからのIGVプロット。MHC遺伝子座位の722 kb領域にわたるフェージング。580 kbは1つのフェーズブロックに含まれる領域（ピンク）です。

まとめ

Illumina Complete Long Read Prep with Enrichment, Humanは、実績のあるイルミナショートリードWGSを補完し、最大の価値を提供するロングリードシーケンスに焦点を当てます。高い柔軟性があるため、プレデザインパネルを選択したり、DesignStudioアルゴリズムを使用してロングリードターゲット濃縮用のカスタムパネルを設計したりすることができます。濃縮プローブパネルは、コスト効率の高い、高精度なWGSを実現するための、ターゲット濃縮によりカバレッジを向上させる、または遺伝子全体をフェージングして洞察を追加する完全なワークフローソリューションを提供します。

詳細はこちら

[Illumina Complete Long Read Prep with Enrichment, Human DesignStudioアッセイデザインツール](#)
[ロングリードシーケンステクノロジー](#)

参考文献

1. Illumina. Illumina Complete Long Read Prep, Human data sheet. [illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/illumina-long-read-prep-human-data-sheet-m-gl-01420/illumina-long-read-prep-data-sheet-m-gl-01420.pdf](https://www.illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/illumina-long-read-prep-human-data-sheet-m-gl-01420/illumina-long-read-prep-data-sheet-m-gl-01420.pdf). Published 2022. Accessed September 22, 2023.
2. Illumina. Comprehensive whole-genome sequencing with Illumina Complete Long Read Prep, Human technical note. [illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/illumina-long-read-prep-human-tech-note-m-gl-01421/ilmn-long-read-hu-tech-note-m-gl-01421.pdf](https://www.illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/illumina-long-read-prep-human-tech-note-m-gl-01421/ilmn-long-read-hu-tech-note-m-gl-01421.pdf). Published 2022. Accessed September 22, 2023.
3. Roessler K. Illumina Complete Long Reads software analysis workflow for human WGS. <https://www.illumina.com/science/genomics-research/articles/complete-long-read-software-analysis.html>. Published 2023. Accessed September 22, 2023.
4. Wagner J, Olson ND, Harris L, et al. Curated variation benchmarks for challenging medically relevant autosomal genes. *Nat Biotechnol.* 2022;40(5):672-680. doi:10.1038/s41587-021-01158-1
5. PharmGKB. VIPs: Very Important Pharmacogenes. [pharmgkb.org/vips](https://www.pharmgkb.org/vips). Accessed September 22, 2023.
6. National Library of Medicine. GTR: Genetic Testing Registry. Precision HealthPGx Panel (25 Genes). ncbi.nlm.nih.gov/gtr/tests/593428/. Updated November 29, 2022. Accessed September 22, 2023.

7. Pratt VM, Everts RE, Aggarwal P, et al. Characterization of 137 Genomic DNA Reference Materials for 28 Pharmacogenetic Genes: A GeT-RM Collaborative Project. *J Mol Diagn.* 2016;18(1):109-123. doi:10.1016/j.jmoldx.2015.08.005
8. Miller DT, Lee K, Abul-Husn NS, et al. ACMG SF v3.1 list for reporting of secondary findings in clinical exome and genome sequencing: A policy statement of the American College of Medical Genetics and Genomics (ACMG). *Genet Med.* 2022;24(7):1407-1414. doi:10.1016/j.gim.2022.04.006
9. Kulski JK, Suzuki S, Shiina T. Human leukocyte antigen super-locus: nexus of genomic supergenes, SNPs, indels, transcripts, and haplotypes. *Hum Genome Var.* 2022;9(1):49. doi:10.1038/s41439-022-00226-5
10. Bekritsky MA, Colombo C, Eberle MA. Identifying genomic regions with high quality single nucleotide variant calling. Published 2021. Accessed August 30, 2023.
11. Illumina. Illumina Human Comprehensive Panel data sheet. [illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/illumina-long-read-enrich-hu-comp-panel-data-sheet-m-gl-02191/long-read-hu-comp-panel-data-sheet-m-gl-02191.pdf](https://www.illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/illumina-long-read-enrich-hu-comp-panel-data-sheet-m-gl-02191/long-read-hu-comp-panel-data-sheet-m-gl-02191.pdf). Published 2024. Accessed January 26, 2024.

イルミナ株式会社

〒108-0014 東京都港区芝 5-36-7 三田ベルジュビル 22階
Tel (03) 4578-2800 Fax (03) 4578-2810
jp.illumina.com

 www.facebook.com/illuminakk

販売店

本製品の使用目的は研究に限定されます。診断での使用はできません。 販売条件 : jp.illumina.com/tc

© 2024 Illumina, Inc. All rights reserved.

すべての商標および登録商標は、Illumina, Inc. または各所有者に帰属します。

商標および登録商標の詳細は jp.illumina.com/company/legal.html をご覧ください。

予告なしに仕様および希望販売価格を変更する場合があります。

illumina[®]