

The PacBio logo is positioned in the top left corner. The background of the entire page is a vibrant gradient from teal to magenta, overlaid with a complex pattern of overlapping circles and dots, resembling a DNA double helix or a molecular structure.

# ヒトゲノミクス

HiFi シークエンシングで次の偉大な発見を

サンプルから、かつてないほどの情報を引き出す。  
HiFi シークエンシングをお選びください。

PacBio® HiFi シークエンシングでは、DNA、RNA それぞれの解析に、高精度な一分子ロングリードを使用します。正確なベースコールとリード長の長さという、HiFi シークエンシングにおけるユニークな組み合わせが、ハプロタイプフェージング、高 GC 領域のカバレッジ、バイサルファイトを用いない 5-メチルシトシンの検出を可能とします。

こうした特徴により、ゲノム、トランスクリプトーム、エピゲノムをこれまで以上に包括的に捉え、ヒトの遺伝的変異の全容を探ることができます。HiFi シークエンシングで得られる恩恵、そしてヒトゲノミクスにおける数々のアプリケーションによって、次の偉大な発見がもたらされるでしょう。

The PacBio logo is located in the bottom right corner, rendered in a magenta color.



# ターゲットシーケンシング

ゲノム解析が困難な領域でも、容易かつコスト効率の良い、大規模なシーケンシングが可能です。

## HiFi ターゲットシーケンシングの利点

HiFi リードを用いたターゲットシーケンシングでは、関心のあるゲノム領域のみを、容易に、コスト効率よく、大規模に解析することが可能です。ハイブリッドキャプチャおよびアンプリコンのワークフローがお客様のニーズに合わせて利用でき、正確なハプロタイプ解析やあらゆるバリエーション型の包括的検出など、HiFi リードに期待されるすべての恩恵を受けることができます。HiFi リードの長さと精度により、高相同性、高 GC 含量、反復配列などゲノム中の難読領域にアクセスすることや、HLA などの複雑な遺伝子ファミリーや CYP2D6 などの薬理遺伝子の説明を可能とします。

## カスタムハイブリッドキャプチャおよびアンプリコンのワークフロー



シーケンシングやマッピングの難しい領域でも安定したカバレッジ



ダイレクトフェージングによるハプロタイプの同定



従来の複数のワークフローを1つのアッセイに統合



コスト効率よく、柔軟にスケールアップ可能

### 遺伝子例

### HiFi シーケンシングにより解決される、これまでの NGS における課題

GBA	偽遺伝子、コピー数多型 (CNV)、遺伝子融合、遺伝子置換
CYP21A2	偽遺伝子、CNV、遺伝子融合、遺伝子置換
CYP2D6	偽遺伝子、CNV、遺伝子融合、遺伝子置換
FMR1	トリヌクレオチドリピート、メチル化
HBA1/2	高度に類似した遺伝子ファミリーメンバー、CNV、遺伝子融合
PMS2	偽遺伝子、遺伝子置換
SMN1/2	高度に類似した遺伝子ファミリーメンバー、CNV、遺伝子置換
LPA	コピー数が多く、かつ可変な、長いリピート

図 1. ショートリードを用いたシーケンシングでは配列決定が困難な遺伝子の一例。より包括的なリストで、HiFi シーケンシングではそれらが完全に解析可能であることが示されている<sup>1</sup>。  
注：一覧中の遺伝子にはターゲット・インフォーマティクスを必要とするものが含まれる。



## HiFi ターゲットエンリッチメント

ロングリードに最適化された二つのプロトコルで、より多くのゲノム領域にアクセスし、エンドツーエンドのソリューションを手にすることができます。

Twist Bioscience 社のカスタム DNA プローブを用いた、フレキシブルかつスケーラブルなワークフローにより、大小の遺伝子パネルや個々の遺伝子座をターゲットにすることができます。HiFi ターゲットエンリッチメントは、他の技術ではアクセスできないゲノム領域をカバーし、微小なバリエーションでも構造多型でも包括的にコールすることが可能です。HiFi リードを用いることで可能になる包括的解析で、ロングレンジフェージングや正確なハプロタイプ解析を実現します。

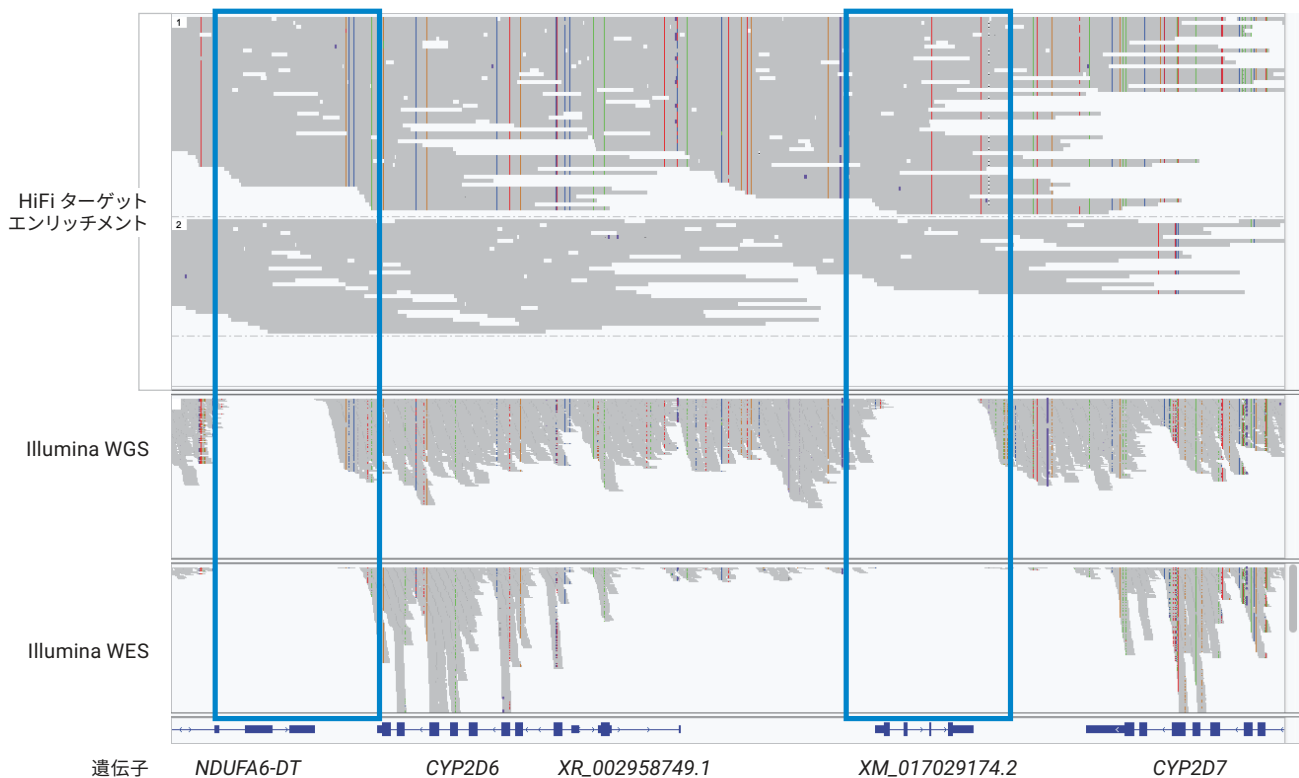


図 2. HiFi ターゲットエンリッチメントでキャプチャした CYP2D6 領域と、100 × ショートリード全エクソーム、50 × ショートリード全ゲノムとの比較。HiFi リードが形成する大きなフェーズブロックは CYP2D6-CYP2D7 領域にまたがっており、相同性の高い CYP2D7 遺伝子による偽遺伝子の曖昧性が解消され、ショートリードで見られるようなオフターゲットリードのマッピングを回避することができる。  
図中の青枠部分のカバレッジに違いが見られることからわかるように、HiFi リードは、ショートリード NGS ではアクセスできない偽遺伝子間のセグメント重複領域も配列決定している。

**「本当に精密なロングリードの能力のおかげで、偽遺伝子を完全に回避し、重要な薬理遺伝子を正確に調べることができます。」**

スチュアート・スコット (スタンフォード大学)、PacBio Roadshow 2022、サンフランシスコ



## HiFi アンプリコン

一つずつのプライマーペアとアンプリコンプール、容易なアッセイ最適化、ロングレンジフェージング、シンプルなデータ解析を特徴とするシンプルなワークフローをぜひお使いください。

PCR を用いたアプローチで、医学的に関連性の高い、個々の遺伝子または小規模遺伝子パネルをターゲットにします。増幅が可能であれば、HiFi シークエンシングは 1 本のリード中のアンプリコンを網羅し、ダイレクトフェージングによる明確なハプロタイプ解析を実現します。

HiFi アンプリコンシーケンシングにより、複数のアッセイを一つの迅速で容易、かつスケラブルで経済的なワークフローに置き換えることができます。

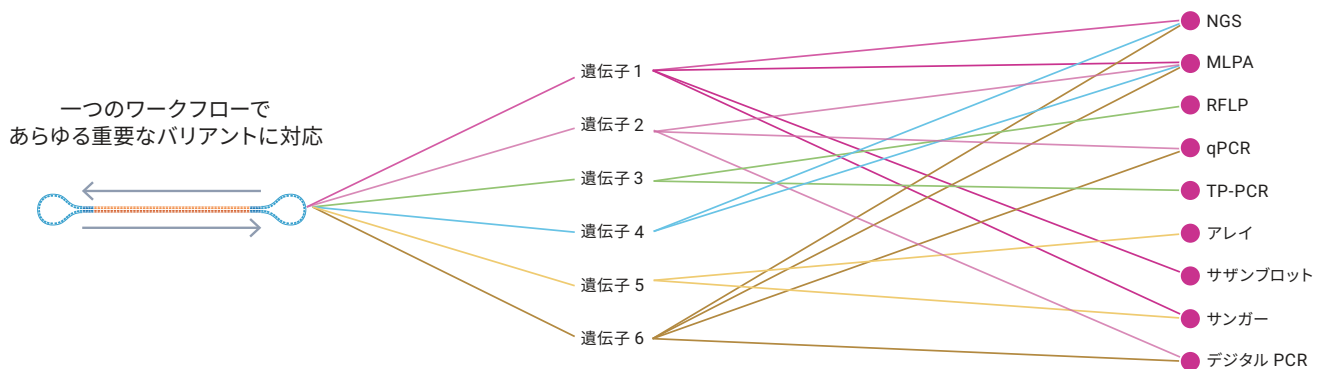


図 3. SMRTbell® ライブラリーは、多くのラボがさまざまな手法を用いて取り組まなければならない典型的な難読遺伝子のすべてに対して、構築と配列決定を可能とする。このアッセイ統合はいくつかの疾患研究分野で研究され、従来の方法と一致して、主要なバリエーションを検出することが示された<sup>1,2</sup>。いくつかのケースでは、この統合された手法の方が高い陽性検出率を示した<sup>3</sup>。

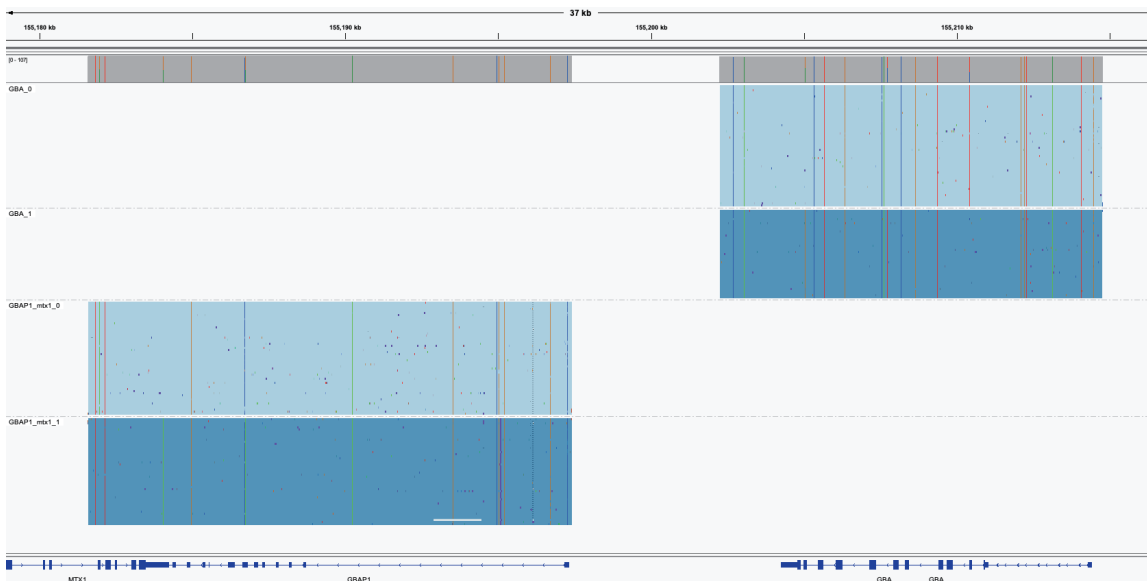


図 4. リソソーム酵素であるグルコセレブロシダーゼをコードする遺伝子 GBAP における変異はゴーシェ病の原因となる。GBAP は約 8 kb の領域に 11 個のエクソンを持ち、近傍に偽遺伝子 GBAP1 (エクソン類似度 96%) を持ち、350 以上のミスセンス、ナンセンス、スプライシングバリエーションを含み、バリエーションコールを行うには複雑な領域となっている。図では、GBAP (上側) と偽遺伝子 GBAP1 (下側) に特異的アンプリコンを用いたアッセイによって、完全なフェージングや偽遺伝子の曖昧さ解消、ハイブリッドアレルのキャプチャ、さらに両遺伝子座のコピー数の把握が可能であることが示されている。



## バリエント検出

ゲノム内のすべてのバリエントを包括的に検出することで、  
遺伝子と表現型を結びつけ、  
原因遺伝子やバリエントを新たに発見するための能力が向上します。

一塩基多型 (SNV) は最も多いバリエント型 (一人あたり 400 ~ 500 万) であり、次いで多いのは Indel です。構造多型やタンデムリピートの数ははるかに少ないので、それらの重要性は小さいと思われるかもしれませんが、構造多型はその大きさ (>50 bp) から、個人間に見られるゲノム変異としては SNV と Indel を合わせたものよりも多くを占めています (図 5)。

HiFi シークエンシングだけが正確で小規模なバリエントコールを可能とし、同時に構造多型 (SV) の検出を大幅に向上させることができます<sup>4</sup>。これらのバリエント型を完全に解読することで、ヒトの遺伝的多様性や疾患との関連性をより深く研究することが可能になります。

バリエント型における既知の病原性は、疾患との関連を考える上で重要です。ClinVar を分析すると、SNV よりも、SV、Indel、タンデムリピートの病原性バリエントの割合がずっと高いことがわかります (図 6)。HiFi を用いることで、研究者はすべてのタイプの変異を包括的に研究し、機能をよりよく解明し、ショートリードを用いた全ゲノムシークエンシング (WGS) や全エクソームシークエンシング (WES) では未解決な課題を改善できます。

バリエント型ごとの疾患関連の分類  
2022 年 7 月

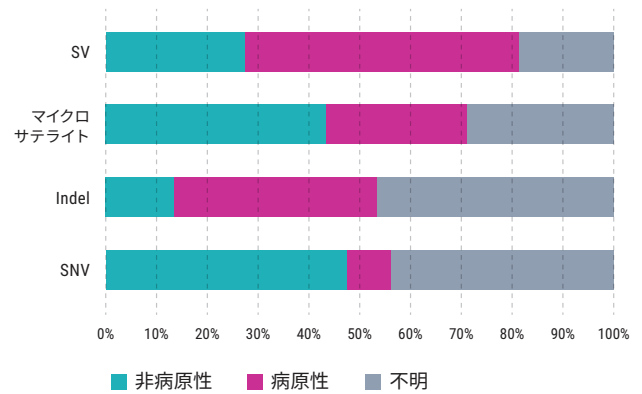


図 6. ClinVar におけるバリエントの分析では、SNV の約 9% に対して、SV の 50% 超、タンデムリピートの 25% 超、Indel の 45% 超が疾患関連であることが明らかになり、疾患の起源を研究する際に SNV だけでなく、それ以外も調べることの重要性を示している。

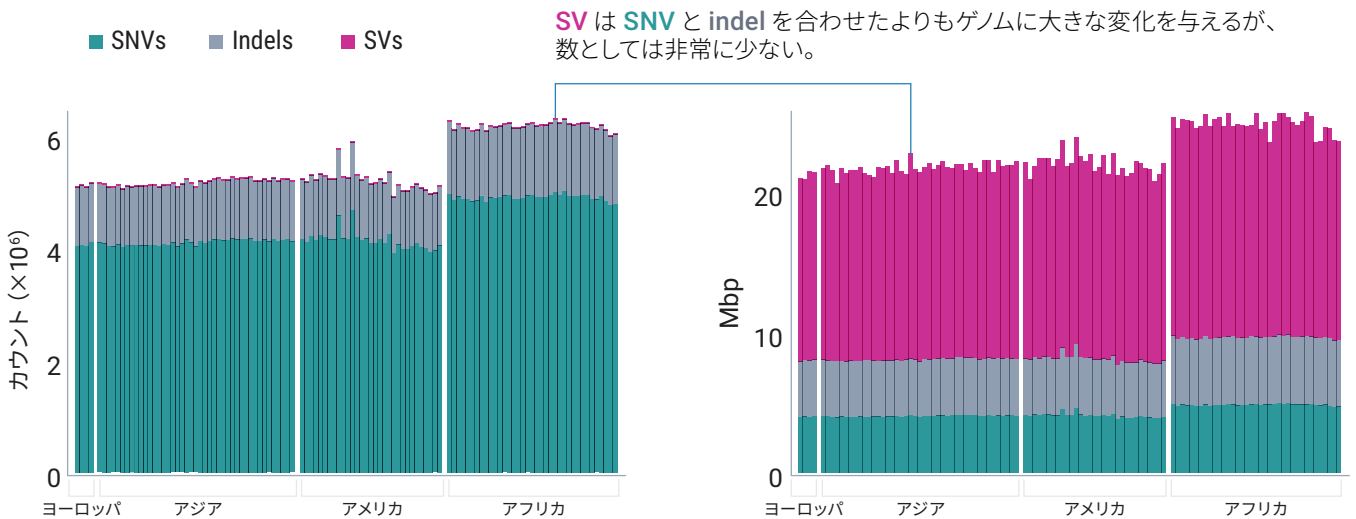


図 5. ゲノム全体で SV (約 25,000) と SNV (400 万以上) に比べて頻度が低い。しかし右図から、SNV が 4 Mb 程度であるのに対して SV は 10 Mb 以上であり、低頻度でも塩基対数としてはゲノム全体に影響を及ぼすことがわかる。このことは、SV がヒトの健康や疾患に大きな影響を与える可能性を示している。



## 全ゲノムシーケンシング

HiFi シーケンシングの長さ、正確さ、偏りのないカバレッジを利用して、完全なヒトゲノムアセンブリを作製できます。

HiFi シーケンシングにより、全ゲノム配列情報が驚くほどの精度と詳細さで入手できます。PacBio SMRT® (Single Molecule, Real-Time) テクノロジーに基づく HiFi リードは、典型的には約 15 kb の長さを持ち、99.9% の精度で、高 GC 領域や、重複領域、リピート領域やセントロメアでも配列決定可能です。同時に 5mC コールによってゲノムを俯瞰することができます。これらの最先端の利点から、*Telomere-to-Telomere Consortium* や *Human Pangenome Reference Consortium*

は、最初に正確なヒトゲノム<sup>6</sup>や二倍体ゲノム<sup>7</sup>を作製するのに適したシーケンシング技術として、HiFi シーケンシングを選択しました。

HiFi シーケンシングにより、あらゆるタイプのバリエーション（一塩基多型 (SNV)、微小な挿入・欠失 (indel)、SV、タンDEM リピート、異常伸長など）を最高の精度と再現性でコールできます。HiFi を用いた WGS は、研究目的に応じてさまざまな深度で実施可能です。

	10 倍カバレッジ	20 倍カバレッジ	30 倍カバレッジ
構造多型 F1	✓ 93.3%	✓ 94.9%	✓ 95.0%
SNV F1	✓ 98.2%	✓ 99.9%	✓ 99.9%
Indel F1	92.1%	✓ 98.2%	✓ 99.3%
フェージング フェーズブロック N50	✓ 64 kb	✓ 247 kb	✓ 252 kb
メチル化 WGBS との相関	✓ 0.89	✓ 0.92	✓ 0.93

### 10 倍デプス

8 ~ 10 倍の深度の HiFi ゲノムは、既存のショートリードを用いたデータセットを有意義に補強・補完するコスト効率の高い手法です。この深度では、構造多型、フェーズド 5mC プロファイル、リピート伸長を確実に検出できます。

### 20 ~ 30 倍デプス

この深度で行う HiFi WGS は、すべてのバリエーション型を検出し、ショートリードを用いたシーケンシングでは見逃されるゲノム領域を明らかにする比類のない能力を備えています。希少疾病研究などのアプリケーションでは、微小なバリエーション以外にも調べることができるため、発見や解決できる割合が向上することが示されています<sup>8</sup>。



## エピジェネティクス

最も正確なシーケンシングリードと DNA メチル化により、  
1 回の実験でヒトゲノムの新しい次元を見ることができます。

HiFi シーケンシングでは、追加のライブラリー調製なしで、正確な DNA ベースコールと同時に、CpG コンテキスト中の 5-メチルシトシン (5mC) の検出が可能です。この特徴により、フェージングによるハプロタイプ解析で、メチル化プロファイルを明らかにすることが可能です。ヒトゲノム研究者はこの特徴を利用して、タンデムリピートに関連するインプリンティング疾患やメチル化異常を調べることもできる可能性があります。

下図に示した例では、ある対立遺伝子上でのリピート伸長が、隣接するメチル化状態に影響を及ぼすことがわかります。

PacBio の HiFi シーケンシングは、特別な処理をせずにゲノム DNA から 4 つの DNA 塩基と 5mC を同時にコールします。長く正確なリードを用いた 1 回の標準的な HiFi ライブラリー調製で、ゲノム全体の遺伝子およびエピジェネティックなバリエーションの検出とフェージングが実現します。



図 7. 筋緊張性ジストロフィーのサンプル (Children's Mercy Kansas City) における、モザイク状の 5 kb DMPK 伸長に隣接する領域の高メチル化が、HiFi シーケンシングによるフェージングで同定された。マゼンタはメチル化塩基、青は非メチル化塩基を示す。



## タンデムリピート

一般的な全ゲノム解析パイプラインでは明確に解読されないヒトゲノム中のタンデムリピートについて、新たな知見が得られます。

タンデムリピートは、ヒトゲノムに最も多く存在するバリエーションの一つです。その繰り返しの性質上、ゲノム中で最も突然変異率が高く、そのゲノムの不安定性が疾患を引き起こす大きな要因となります。

ショートタンデムリピート (STR) の伸長によって引き起こされる疾患は 50 以上あり、いくつかの変異タンデムリピート (VNTR) は、アルツハイマー病、自閉症、てんかん、ALS など、複合疾患に関与することが示されています。

タンデムリピートはヒトの健康や疾患において鍵となる役割を持ちますが、一般に全ゲノム解析パイプラインの一部として詳細が明らかにされているわけではありません。これは、それが重要

ではないと考えられているからではなく、ショートリードを用いたシーケンシングでは配列決定が不可能であることが、長い間、解析方法の開発やその欠如に直接影響を及ぼしてきたからです。近年、ヒトゲノムを正確に配列決定する取り組みの成果の一つとして、構造多型の 50 ~ 80% が実際にはタンデムリピートであるとの推定がなされました<sup>8</sup>。しかし、現在の解析パイプラインはタンデムリピートに最適化されていないため、小さいものは Indel として、大きいものはバリエーションのクラスターとして報告されることが多いのが実情です。

HiFi シーケンシングにより、メチル化プロファイルとともに、リピート伸長や医学的に重要な分断配列の完全かつ正確なコールが可能になります。

### ロングリードシーケンシングによる革命的なリピート伸長疾患の同定

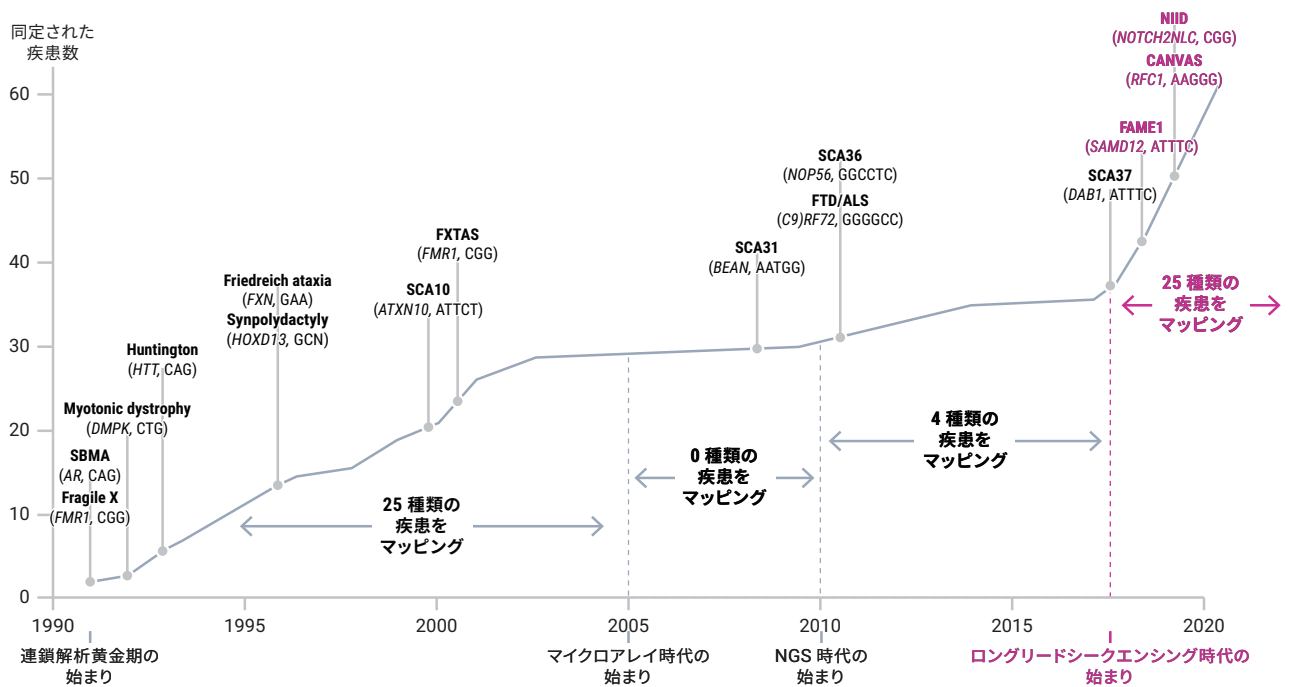


図 8. ヒト疾患におけるリピート伸長に関する発見の年表<sup>9</sup>。





## タンデムリピートジェノタイピングツールおよび タンデムリピートビジュアライザー

新しい HiFi タンデムリピート解析ツール一式を用いた  
包括的なリピート伸長のコールにより、ゲノムの高多型領域における  
遺伝的およびエピジェネティックな変異の評価が可能です。

タンデムリピートの正確な解析（リピート伸長の検出を含む）の  
ためには、リピート両端にまたがるシーケンスという特有の要件  
が求められます。HiFi シーケンシングでは、1本のリード  
中にリピート伸長全体が含まれた状態で配列解析できるため、  
リピートサイズ、分断配列、メチル化を同定することができます。

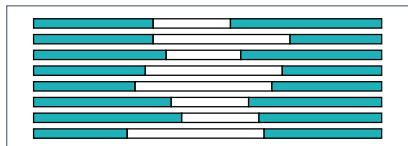
HiFi とタンデムリピートジェノタイピングツール（TRGT）、タン  
デムリピートビジュアライザー（TRVZ）、その他のタンデムリピ  
ート解析ソフトウェアなど一式を組み合わせることで、新規リピ  
ートアレルやメチル化パターンの発見など、タンデムリピート領  
域における遺伝子やエピゲノムの変化をゲノム全体で解析する  
ことが可能になります。

リピートを定義

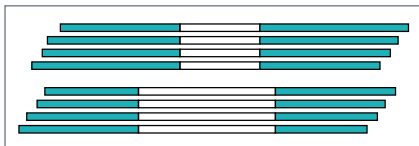
chr4      3074876      3074966      ID=HTT,STRUC=(CAG)<sub>n</sub>CAACAG(CCG)<sub>n</sub>



TR 配列と隣接領域を検出



リードをハプロタイプに割り当て



リピート配列のセグメント化

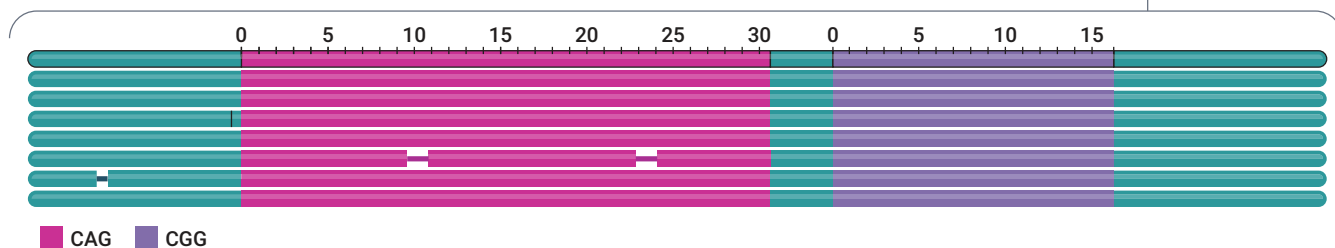
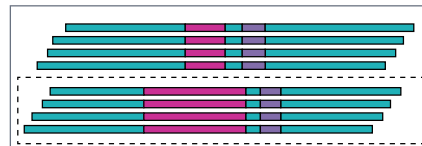


図 9. TRGT からの出力を TRVZ で簡略化して表示すると、フェージング解析されたハプロタイプの両方について DNA モチーフ、リピート数、モザイク配列、分断配列、フランキング配列が示される。TRGT の出力は、必要に応じて IGV で DNA 配列として見ることもできる。



## RNA シークエンシング

関心のある遺伝子やトランスクリプトーム全体の完全な転写産物アイソフォームを容易かつコスト効率よくシークエンシングすることができます。

### RNA から全長転写産物へ

転写産物のスプライシングと量の変化は、ゲノム変異によって健康状態と疾病の間の表現型の違いがどのように引き起こされるかを解読するための重要な手がかりとなります。サンプルに存在するすべてのアイソフォームを偏りなく直接検出できるのは、HiFi シークエンシングだけです。PacBio Iso-Seq<sup>®</sup> 法により、アセンブリや複雑なアルゴリズムなしで全長転写産物のアイソフォーム情報が得られます。

### Iso-Seq 法は以下のことに用いることができます。



組織やシングルセルの RNA サンプルに対して、アイソフォームレベルでトランスクリプトーム全体を網羅的にプロファイリング



選択的転写開始部位、ポリアデニル化部位、エクソン・スキッピングに関する解析評価



アレル特異的なスプライシングの同定



オープンリーディングフレーム (ORF) を直接予測することで、スプライシングの変化による発現タンパク質を推定



サンプル中で関心のある遺伝子をターゲット化

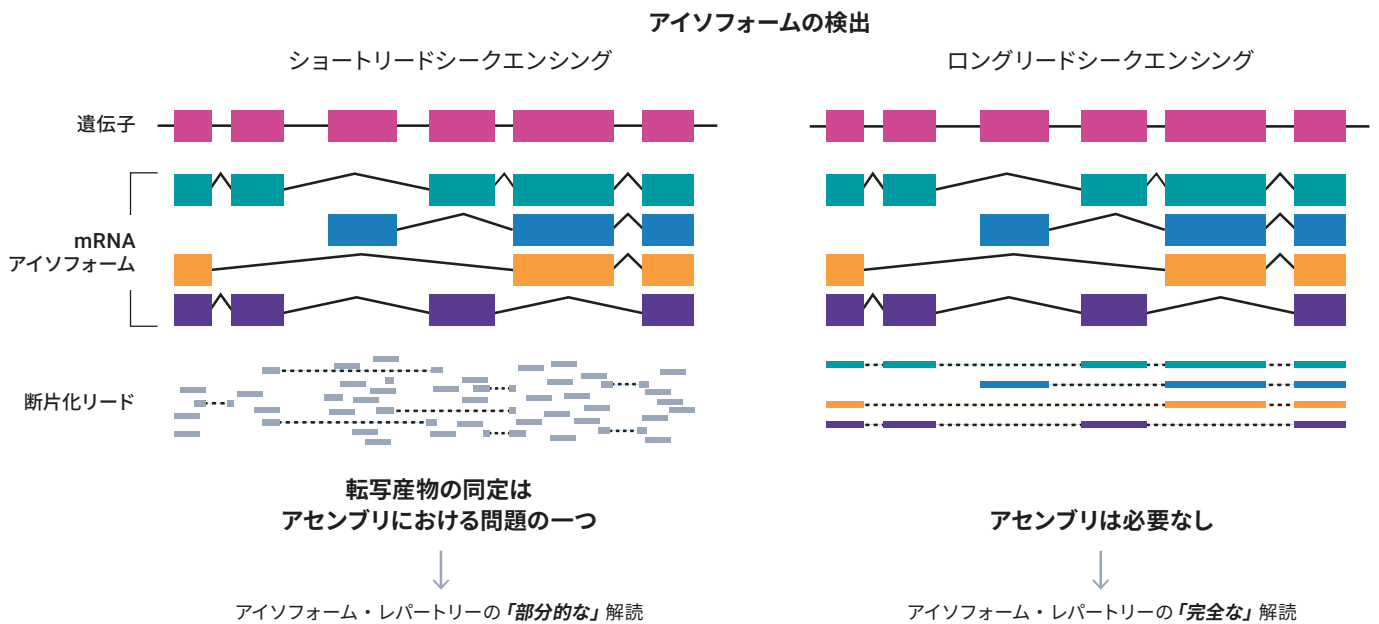


図 10. RNA シークエンシングにおける Iso-Seq 法の優位性。



## MAS-Seq

シングルセル RNA シークエンシングの分解能により、全長アイソフォーム情報を用いて、未知のアイソフォームの多様性が明らかになります。

### MAS-Seq によるシングルセル・アイソフォームシークエンシング

MAS-Seq 法<sup>10</sup>は、cDNA 分子を連結してより長い断片にすることでスループットを高める手法です。連結した分子のシークエンシングで生成される HiFi リードは、バイオインフォマティクスに基づいて分割され、元の cDNA 配列が復元されます。その結果、スループットが向上し、シークエンシングの必要回数が減少するため、コスト効率の高いシングルセルアイソフォーム解析が実現します。ショートリードによる直交シークエンシングデータを必要としません。

PacBio シングルセル Iso-Seq ワークフロー<sup>11</sup>では、全長 cDNA 配列を処理し、参照アノテーション（例：GENCODE）に照らして分類し、新規遺伝子やアイソフォームを同定します。出力は、Seurat<sup>12</sup>、Scanpy<sup>13</sup>、Kana<sup>14</sup> など三次解析ソフトウェアで使用可能な、遺伝子レベルおよびアイソフォームレベルのカウントマトリクスから構成されます。

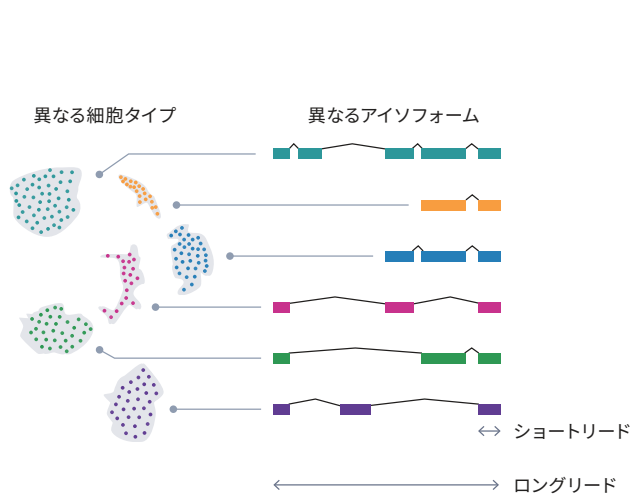


図 11. PacBio ロングリードを用いたシングルセル・アイソフォームシークエンシング。  
ショートリードは遺伝子末端のみを捕捉し、結果的にアイソフォームの多様性は見過ごされる。  
PacBio HiFi リードは、シングルセルバーコードおよび UMI の情報でアイソフォーム全体を高精度にカバーする。

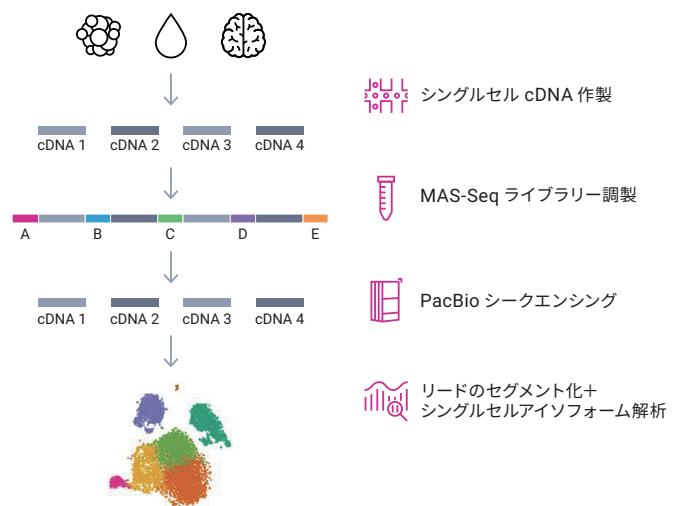


図 12. MAS-Seq によるシングルセル・アイソフォームシークエンシング。  
シングルセル由来の cDNA 分子を連結して、より大きなインサートライブラリーとしてシークエンシングを行う。PacBio HiFi リードはインサート全体をカバーするため、シークエンシング後にバイオインフォマティクスによって、最初の cDNA 分子を表すリードにセグメント化することができる。  
シングルセル Iso-Seq ワークフローでは、遺伝子およびアイソフォームレベルのカウントマトリクスを作成し、三次解析ソフトウェアでセルタイプの解析に使用することができる。

## Headquarters

1305 O' Brien Drive Menlo Park, CA  
94025 United States

## 日本支社

〒 220-0012  
神奈川県横浜市西区みなとみらい 3-7-1  
オーシャンゲートみなとみらい 8F  
パックバイオジャパン合同会社  
Info-JP@pacb.com

## 輸入販売元

トミーデジタルバイオロジー株式会社  
〒 112-0002  
東京都文京区小石川 1-1-17  
日本生命春日駅前ビル 3 階  
phone: 03-6240-0843 fax: 03-6240-0461  
info\_pac@digital-biology.co.jp

## READY TO GET STARTED WITH HIFI SEQUENCING?

 製品・サービス  
[pacb.com/products](https://pacb.com/products)

### KEY REFERENCES

1. Wenger, A.M., Peluso, P., Rowell, W.J. et al. (2019) Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat Biotechnol* 37, 1155–1162
2. Peng, C., Zhang, H., Ren, J. et al. (2022) Analysis of rare thalassemia genetic variants based on third-generation sequencing. *Sci Rep* 12, 9907
3. Tantirukdham N, Sahakitrungruang T, Chaisiwamongkol R, et al. (2002) Long-read amplicon sequencing of the CYP21A2 in 48 Thai patients with steroid 21-hydroxylase deficiency. *J Clin Endocrinol Metab.* 2022;107(7):1939–1947.
4. Ebert, P. et al. (2021) Haplotype-resolved diverse human genomes and integrated analysis of structural variation. *Science*.
5. Nurk, S. et al. (2022) The complete sequence of a human genome.
6. Wang, T. et al. (2022) The Human Pangenome Project: A global resource to map genomic diversity.
7. Cohen, Ana S. A., et al. 2019. Genomic answers for children: dynamic analyses of >1000 pediatric rare disease genomes. *Genetics in Medicine* 1336–48.
8. Estimated by Adam English (Baylor College of Medicine) with Truvari software; SV from 36 haplotype-resolved assemblies spanning at least 50 bp; TR definitions from the UCSC *SimpleRepeats* track
9. Depienne et. al (2021) 30 years of repeat expansion disorders: What have we learned and what are the remaining challenges? *AJHG*,108(5): 764-785
10. Procedure & checklist – Preparing MAS-Seq libraries using MAS-Seq for 10x single cell 3' kit
11. SMRT Link v11.1 user guide
12. Seurat: <https://satijalab.org/seurat/>
13. Scanpy: <https://scanpy.readthedocs.io/>
14. Kana: <https://www.jkanche.com/kana/>

Research use only. Not for use in diagnostic procedures. © 2022 Pacific Biosciences of California, Inc. ("PacBio"). All rights reserved. Information in this document is subject to change without notice. PacBio assumes no responsibility for any errors or omissions in this document. Certain notices, terms, conditions and/or use restrictions may pertain to your use of PacBio products and/or third-party products. Refer to the applicable PacBio terms and conditions of sale and to the applicable license terms at [pacb.com/license](https://pacb.com/license). Pacific Biosciences, the PacBio logo, PacBio, Circulomics, Omniome, SMRT, SMRTbell, Iso-Seq, Sequel, Nanobind, and SBB are trademarks of PacBio. All other trademarks are the sole property of their respective owners.