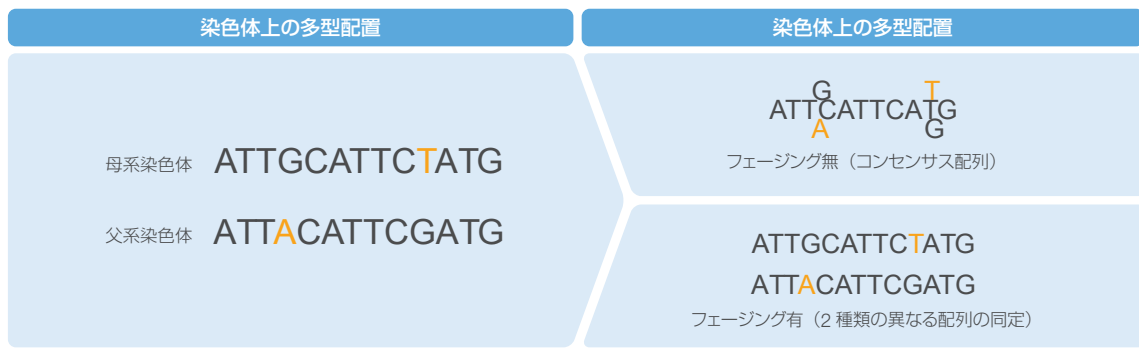


図 2: フェージングによるハプロタイプ情報の同定



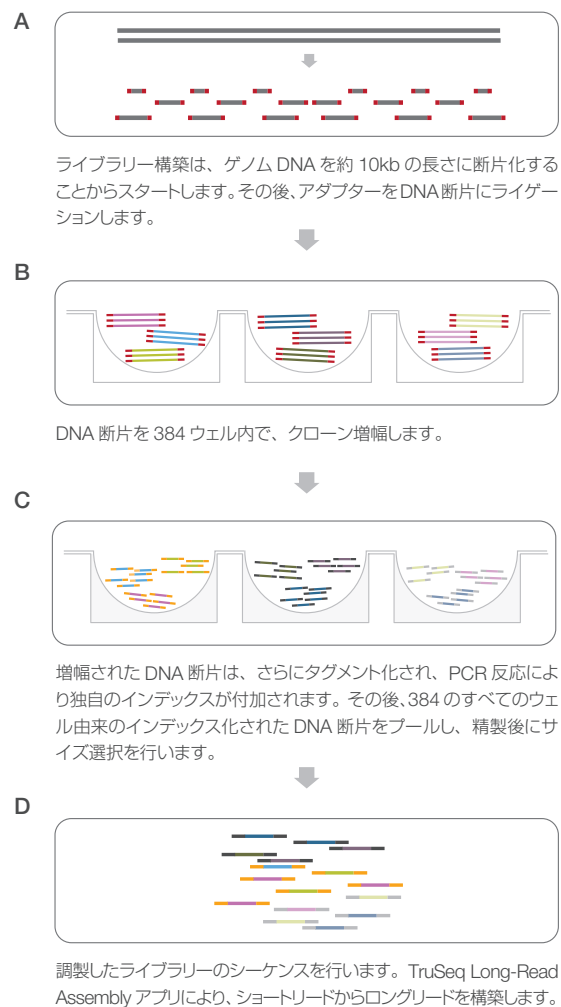
ゲノムフェージングにより、相同染色体上の多型の識別が行われ、リンクパターンおよびアレル特異的な発現についての研究が可能となります。

独自のライブラリー調製ケミストリー

TruSeq Synthetic Long-Read DNA ライブラリー調製キットおよびバーコードキットは、TruSeq および Nextera® ケミストリーと合成ロングリードテクノロジーを組み合わせ、シーケンス用の DNA ライブラリーを調製します (図 3)。バーコードキットには、各ウェル中のサンプルをラベル化するための 384 のインデックスが含まれています。これらのインデックスは、シーケンス後、合成ロングリードを正確に構築するために使用されます。キットには、マスターミックス試薬およびインデックスプライマーが予め入った 384 ウェルプレートが含まれています。このようなデザインにより、マニュアルでのピペット操作のステップを最小限に抑え、簡単な遠心ステップを使用して PCR によるバーコード化を可能にします。別売りの TruSeq Synthetic Long-Read DNA アクセサリーキットには、プレートのアレンジおよび遠心ステップのためのアラメントリングが含まれています。

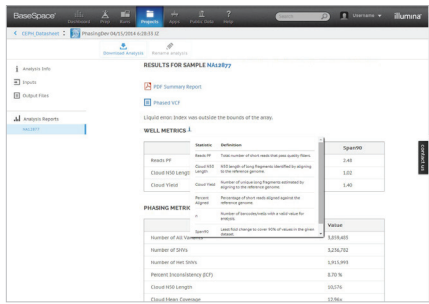
ライブラリー調製は、500ng のゲノム DNA を約 10kb に断片化することから始まります。アダプターをライゲーションした後、10kb の断片をサイズ選択します (図 3A)。次に、Nextera タグメンテーション反応により、各ウェル内の DNA 断片をさらに短く断片化し、同時にタグ化を行います。(図 3B)。その後 PCR 増幅ステップで、サンプルに固有のインデックスを付加します。384 のすべてのウェル由来のインデックス化された DNA 断片をプールのし、精製後にサイズ選択します (図 3C)。このライブラリー調製ステップは、6 時間のハンズオン時間を含め 3 日間で完了することができます。続いてライブラリーのシーケンスをイリミナ次世代シーケンサーシステムで行います (図 3D)。シーケンス後、TruSeq Phasing Analysis アプリによりショートリードから合成ロングリードが生成された後に、ゲノムフェージングが行われ、ハプロタイプ情報および変異リンクを同定します。

図 3: TruSeq Synthetic Long-Read DNA ライブラリー調製ワークフロー



TruSeq Synthetic Long-Read DNA ライブラリー調製キットは、シーケンスおよびフェージング用に DNA ライブラリーを調製します。TruSeq Phasing Analysis アプリでは、変異リンクおよび染色体特異的な変異を同定し、ヒトゲノムの包括的な情報を提供します。

図 4: シンプルなフェージング解析



直観的なユーザーインターフェースを特長とする TruSeq Phasing Analysis アプリにより、バイオインフォマティクスに関する専門的知見に関わらず、あらゆる生物学研究者のゲノムフェージングを簡素化します。結果は分かりやすい表およびグラフにより表示されます。

シンプルなフェージング解析

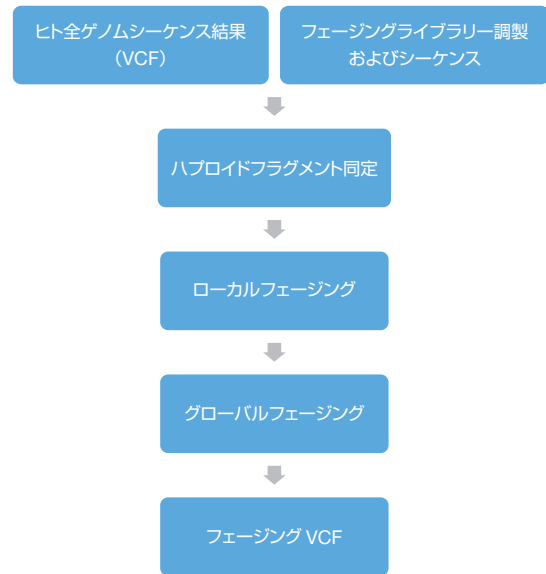
ライブラリー調製に加えて、TruSeq のフェージング法には BaseSpace 環境におけるプッシュボタン解析が含まれます。TruSeq Synthetic Long-Read DNA ライブラリー調製キットおよびバーコードキットとともに使用するようデザインされた TruSeq Phasing Analysis アプリは、全ゲノムシーケンス結果をロングリードのデータと組み合わせることにより、ハプロタイプコンティグを生成します。直観的なユーザーインターフェース(図 4)は、データ解析を簡素化します。全ゲノムシーケンス結果を変異コールフォーマット (VCF) でアップロードしてアプリを立ち上げるだけで全ゲノムのフェージングが行えます。

まず、アプリは TruSeq Synthetic Long-Read DNA バーコードキットからのインデックスに関連したショートリードを使用して、各ウェルから隣接したハプロイドフラグメントを同定します。次に、オーバーラップするヘテロ接合性 SNP を有するフラグメントをつなぎ合わせることで、ローカルフェージングが行われます。最後に、グローバルフェージング過程において、統計学的補完を利用してハプロタイプフラグメントがフェージングされた全ゲノム配列にアセンブルされます(図 5)。アプリはハプロタイプコンティグおよび信頼性スコアを、フェージングされた VCF ファイルで提供します。この業界標準のフォーマットは、詳しい解析および視覚化を行うための様々な下流解析ツールと互換性があります。TruSeq Phasing Analysis アプリは、複雑な統計解析を必要としない、ワンクリックのゲノムフェージングを可能とします。

高精度なデータ

TruSeq 法は、96% 以上の SNP のフェージングのために、独自のライブラリー調製ケミストリーと統計解析を組み合せます。TruSeq Phasing Analysis アプリは、全遺伝子⁵内の 94% 以上において 70% 以上の変異をフェージングし(表 1)、99% を超える精度を達成しています(図 6)。高品質のデータを新しいアルゴリズムと組み合わせることにより、TruSeq Synthetic Long-Read 技術は、実験的または統計的なアプローチと比較して、より包括的なハプロタイプ情報を提供します。

図 5: フェージング解析ワークフロー



TruSeq Phasing Analysis アプリは、複雑な解析も統計計算も必要とせず、ゲノムワイドのフェージングを可能とします。VCF ファイルをアップロードして、適切なシーケンスランを BaseSpace 内で選択するだけです。アプリはハプロタイプフラグメントをアセンブルしてゲノム全体において遺伝子および変異をフェージングし、フェージング VCF ファイルを提供します。

表 1: TruSeq Phasing Analysis アプリの結果

パラメーター	個人 NA12877	個人 NA12878	個人 NA12891	個人 NA12892
SNP				
グローバル N50 (bp)	846,345	581,317	700,672	597,340
フェージング (ローカル)	92.2%	90.7%	92.5%	90.7%
フェージング (グローバル)	98.7%	98.2%	98.5%	98.1%
SNP および Indel				
グローバル N50 (bp)	860,793	587,396	712,926	606,037
フェージング (ローカル)	86.3%	85.4%	87.9%	85.5%
フェージング (グローバル)	92.8%	92.8%	93.8%	92.8%
遺伝子フェージング (SNP およびインデル)				
100% の変異フェージング	45%	44%	47%	44%
70% の変異フェージング	95%	95%	95%	94%

個人 NA12877、NA12878 および NA12891 は、血縁関係であり、個人 NA12892 は血縁関係ではありません。TruSeq Synthetic Long-Read DNA ライブラリー調製キットを使用して DNA ライブラリーを調製し、解析は TruSeq Phasing Analysis アプリを使用して行いました。TruSeq Phasing Analysis アプリからの結果はフェージングされた多型の割合およびハプロタイプフラグメントの N50 長を示し、アプリによりヒトゲノム内の大部分の変異と遺伝子がフェージングされたことが示されています。

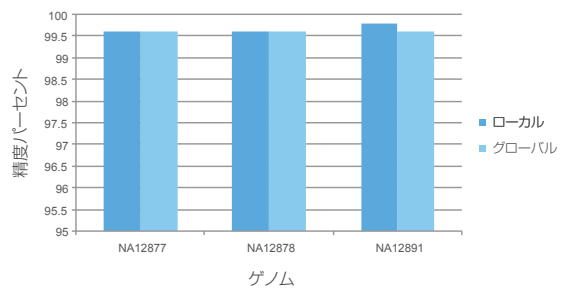
まとめ

TruSeq Synthetic Long-Read DNA ライブラリー調製キットおよびバーコードキットは、ヒト全ゲノムのシーケンスおよびフェーシング用のライブラリー調製のためにデザインされています。ショートリードから合成ロングリードを構築することにより、従来のトリオ解析および統計的推定と比較してより包括的で正確なフェーシングが提供されます。BaseSpace 内でのシンプルな解析により、最短のハンズオン時間でゲノム全体のフェーシングを行うことが可能となります。TruSeq Synthetic Long-Read DNA ライブラリー調製キット、バーコードキット、および専用アプリの組合せにより、シンプルで正確かつコスト効率の高い全ゲノムフェーシングが実現します。

参考文献

1. Tewhey R, Bansal V, Torkamani A, Topol EJ, Schork NJ (2011) The importance of phase information for human genomics. Nat Rev Genet 12:215–23.
2. Kuleshov V, Xie D, Chen R, Pushkarev D, Ma Z, et al. (2014) Whole-genome haplotyping using long reads and statistical methods. Nat Biotechnol 32:261–6.
3. Kleinjan DJ, Coutino P (2009) Cis-ruption mechanisms: disruption of cis-regulatory control as a cause of human genetic disease. Brief Funct Genomic Proteomic 8: 317–32.
4. BaseSpace (basespace.illumina.com) Accessed 04 June 2014.
5. Public Data - BaseSpace (basespace.illumina.com/datacentral) Accessed 04 June 2014.
6. Platinum Genomes | Illumina (www.illumina.com/platinumgenomes) Accessed 04 June 2014.
7. Complete Genomics > LFR (www.completegenomics.com/LFR) Accessed 04 June 2014.
8. Browning SR, Browning BL (2011) Haplotype phasing: existing methods and new developments. Nat Rev Genet 12: 703–14.

図 6：高いフェーシング精度



フェーシング精度は、ローカルハプロタイプの編集率およびグローバルのスイッチエラー率に基づくものです。TruSeq Phasing Analysis アプリは Browning, SRら (2011年)⁸の方法を改良して、Platinum Genomes Project⁶ (NA12877 および NA12878) および Complete Genomics⁷ (NA12891) 由来のリファレンスデータとの比較において、99%を超える精度を達成しています。これらの結果は、CEPH ファミリー由来の高品質 DNA を使用して生成されました。

製品情報

製品名	カタログ番号
TruSeq Synthetic Long-Read DNA Library Prep Kit (4 サンプル)	FC-126-1001
TruSeq Synthetic Long-Read DNA Barcode Kit (1 サンプル)	FC-126-1002
TruSeq Synthetic Long-Read DNA Barcode Kit (4 サンプル)	FC-126-1003
TruSeq Synthetic Long-Read DNA Accessory Kit	FC-126-1004

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

www.illumina.co.jp

代理店

本製品の使用目的は研究に限定されます。

© 2014 Illumina, Inc. All rights reserved.

Illumina, BaseSpace, Nextera, TruSeq, the pumpkin orange color, and the Genetic Energy streaming bases design are trademarks or registered trademarks of Illumina, Inc. その他の会社名や商品名は、各社の商標または登録商標です。予告なしに仕様を変更する場合があります。

Pub. No. 770-2014-J010 25SEP2014

illumina®