

トランスクリプトームアセンブリーにおいて、 三分の一のコストで同等の高品質を達成

NovaSeq 6000 SP Reagent Kit (500 Cycles) と HiSeq Rapid Kit v2 (500 Cycles) の比較

はじめに

次世代シーケンサーは、生物種を問わずゲノム、トランスクリプトーム解析を実施するために必須のツールとなっている。イルミナ次世代シーケンサーに使われている SBS テクノロジーは、今日に至るまで様々な改良がなされ、最新の NovaSeq 6000 システム (カタログ番号 20012850) は、1 ランで最大 60 億塩基 (6 Tb) にもおよぶデータ出力を可能にし、飛躍的に Gb あたりのコストを削減した。またデータ量、リード長の両方で豊富な試薬ラインナップで目的にあったスループットを実現する。

今回、理化学研究所 生命機能科学研究センター 分子配列比較解析チーム (工樂樹洋チームリーダー) の西村理先生らのご協力のもと、NovaSeq 6000 システム 対応試薬として新たにリリースされた NovaSeq 6000 SP Reagent Kit (500 Cycles) (カタログ番号 20029137) の、多様な生物種での *de novo* トランスクリプトームアセンブリーにおける有用性の検証を行った。また本キットの性能が実績ある HiSeq システムより遜色ないことを確認するため、2 機種を用いて同一ライブラリーから塩基配列情報を取得し、比較した。

実験方法

4 種の脊椎動物 (図 1) から抽出された Total RNA を、TruSeq Stranded mRNA Library Prep kit (カタログ番号 20020594 および 20020595) を用いて調製した。*de novo* アセンブリーに有利な長いインサートを持つライブラリーを作製するために最適化した条件を用いて断片化を実施した¹。また西村先生が独自に開発した iMaster、iLike ツールを用いて、マルチプレックスにおけるインデックス配列の塩基組成、及びユニークさが確認された²。ライブラリーのサイズ分布を確認した後、HiSeq 1500 システムおよび NovaSeq 6000 システムの両方で、251 bp × 2 の条件でシーケンスを実施した。得られた NovaSeq からのリードを Trinity プログラムを用いてアセンブルし³、Corset を用いてクラスタリングを行った後、網羅性評価のためのウェブサーバ gVolante を用いて一般に種間で保存されているオノログのカバー率を比較した (図 2)。最後に NovaSeq と HiSeq それぞれの 10 M リードペアを *de novo* アセンブリーによって得られたコンティグにマッピングし、結果を比較した。

結果

- ライブラリーのインサート長とリード長
今回、TruSeq Stranded mRNA Library Prep kit の標準プロトコールに基づいて調製した場合より長い約 400 bp のインサートを持つ、トータル DNA 分子長約 500 bp のライブラリーを調製した (図 3)。ペアエンドリードのピークの配列長は HiSeq、NovaSeq 共に 318 bp であり、長いインサートを持つライブラリーの特徴を反映していること、HiSeq と NovaSeq で相違がないことが確認された。



図 1: 生物種の一例。左: ソメワケササクレヤモリ、右: イヌザメ。

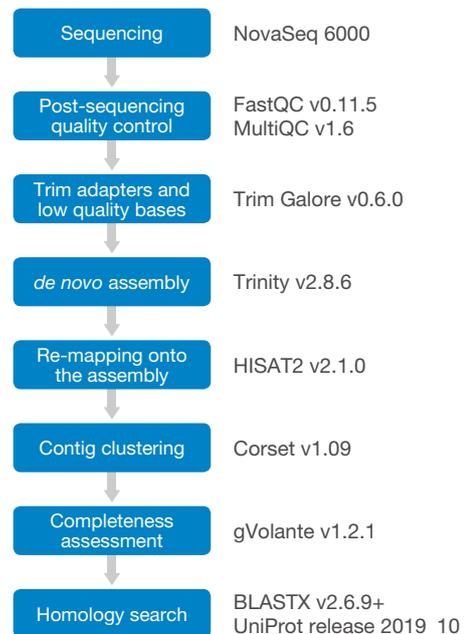


図 2: 解析ワークフロー

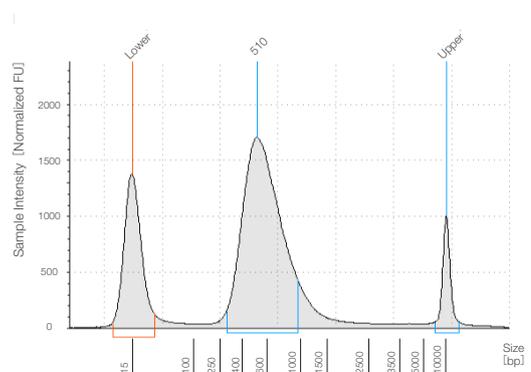


図 3: Agilent TapeStation によるライブラリーのサイズ分布の確認結果例 (サンプルB)

- NovaSeq の塩基ごとの Quality Value (QV 値) FastQC による Quality Value (QV 値) の評価では、長いリード長条件下でのシーケンスでも、NovaSeq のデータはシーケンス後半のクオリティの低下が少なく、4 サンプル共通して良好な結果を示した(図 4)。
- アセンブリ結果と遺伝子網羅性の確認 *de novo* アセンブリ後クラスタリングを行った結果、期待される平均コンティグ長、および N50 長を持つコンティグ配列群が得られた。またどの脊椎動物にも単一コピー存在すると期待されるオーソログの網羅性を基準としてこれらのアセンブリの完成度を評価した結果³、いずれのライブラリーでも検出されなかった遺伝子の数が非常に少なく、網羅性の高いアセンブリが得られている事が確認できた(表 1)。
- HiSeq と NovaSeq の比較 低クオリティ領域とアダプター配列を除去した 10 M リードペアを Corset プログラムで処理した後のコンティグにマッピングした。両者共、高いマップ率を示したが、いずれのライブラリーでも HiSeq のマップ率がわずかに高く(表 2)、機種によるシーケンスのクオリティ算出方法の差⁴が反映されていると考えられる。コストについては、NovaSeq 6000 SP Reagent Kit (500 Cycles) を用いた場合 Gb あたりの費用が約 2300 円、HiSeq を用いた場合約 7400 円であり、約三分の一のコストで同量のデータ量が得られる。

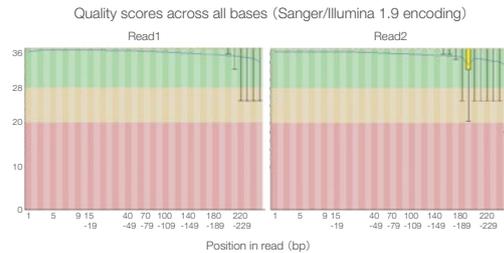


図 4: FastQC による塩基ごとの Quality Value (QV 値) の確認結果例

表 1: gVolante による 4 サンプルの遺伝子網羅性確認結果

サンプル	パイプライン	オーソログセット	ほぼ完全長が検出されたオーソログの割合 (%)	全体の一部のみが検出されたオーソログの割合 (%)	検出されなかったオーソログの割合 (%)
A	CEGMA	CVG	97.0	99.1	0.9
	BUSCO2	BUSCO Vertebrate	96.3	97.8	2.2
B	CEGMA	CVG	95.3	97.9	2.1
	BUSCO2	BUSCO Vertebrate	95.9	97.8	2.2
C	CEGMA	CVG	98.7	99.6	0.4
	BUSCO2	BUSCO Vertebrate	97.3	98.6	1.4
D	CEGMA	CVG	94.0	100.0	0.0
	BUSCO2	BUSCO Vertebrate	95.8	99.4	0.6

表 2: NovaSeq と HiSeq でのマップ率の比較

	マップ率 (%)			
	サンプル A	サンプル B	サンプル C	サンプル D
NovaSeq	91.52	92.84	93.68	90.85
HiSeq	92.74	93.97	94.40	91.39

お客様の声

今回、リード長とライブラリーのインサート長との調和、マルチプレックスにおける堅牢性、そして、アセンブリ結果の多角的な評価を意識することにより、期待されるクオリティのトランスクリプトームアセンブリを得ることができた。トランスクリプトームの *de novo* アセンブリの手段は既に定着し、改善の余地はほとんどないように思われているかもしれないが、すでに実績のあるハイスループットなショートリード取得を、従来のシーケンサーを用いた場合よりも、さらに格段に安価で実施できることは大きな魅力である。このアドバンテージは、多検体のメタバーコーディングや高速 SNP 取得、そして、他のプラットフォームと組み合わせることにより全ゲノムシーケンス等の多様な NGS 解析の敷居を下げることになるだろう。

謝辞：製品性能検証にあたり理化学研究所 工樂樹洋先生、西村理先生に多大なるご協力を賜りました。ここにあらためて感謝の意を表します。

参考文献

1. Hara Y, Tatsumi K, Yoshida M, Kajikawa E, Kiyonari H, & Kuraku S. (2015). Optimizing and benchmarking *de novo* transcriptome sequencing: from library preparation to assembly evaluation. *BMC Genomics*, 16, 977.
2. Nishimura O (2019). iLike and iMaster: tools for sequencing index combination. <https://transcriptome.riken.jp/iLike/>
3. Nishimura O, Hara Y, Kuraku S. (2017). gVolante for standardizing completeness assessment of genome and transcriptome assemblies. *Bioinformatics*. 2017; 33(22):3635- 3637.
4. NovaSeq™ 6000 System Quality Scores and RTA3 Software

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

Pub. No. 5026-200508-01

© 2020 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。

