

NextSeq™ 1000、NextSeq 2000、NextSeq 550 シーケンスシステム間でのデータ相関性

イルミナの最新NGSシステムは革新的なデザインを提供する一方で、一般的なシーケンスアプリケーションに対してお客様が求める同一のデータ品質を維持します。

はじめに

過去20年にわたり、イルミナは次世代シーケンサー(NGS)性能の改善を率先して行い、ゲノム、トランスクリプトーム、およびエピゲノムを探索するために次世代シーケンサーをより簡単に経済的なものにしてきました。イルミナの最新プラットフォームである、NextSeq 1000およびNextSeq 2000シーケンスシステムは画期的なシステムデザイン、ケミストリーイノベーション、さまざまなライブラリー調製オプションとの互換性、そして装置に内蔵された統合型インフォマティクスを提供します。

NextSeq 1000およびNextSeq 2000シーケンスシステムは、NextSeq 550システムなどと同様にすべてのイルミナプラットフォームで使用されているSequencing by synthesis(SBS)ケミストリーを使用しています。イルミナのSBSケミストリーは世界の90%以上のシーケンスデータ生成を担い¹、最も高いエラーのないリードを産生すること²で、ゲノム全体にわたる頑健なベースコール³を可能にしています。

NextSeq 1000およびNextSeq 2000システムはクラスターの光強度を高め、チャンネル間のクロストークを減らし、シグナルノイズ比を改善するために最適化されています。光学的な最新の進歩と装置デザインを合わせて、NextSeq 1000およびNextSeq 2000シーケンスシステムはデータ出力の増加をサポートし、ランあたりのコストを削減する一方で、NextSeq 550システムと同等の高品質なデータを維持しています。

本アプリケーションノートでは、エクソーム、mRNAおよびシングルセルRNAの遺伝子発現などの重要なアプリケーションについて、NextSeq 1000およびNextSeq 2000シーケンスシステムがNextSeq 550システムと同等のデータ品質をもたらすことを示します。

方法

エクソームシーケンス

エクソームライブラリーは、NA12878ゲノムDNA(gDNA)(Coriell Institute for Medical Research)をNextera™ DNA Flex Pre-Enrichment Library PrepおよびEnrichment Reagents Kit(イルミナ、カタログ番号20025524)を用いて調製し、Illumina Exome Panel(イルミナ、カタログ番号20020183)でターゲットゲノム領域をキャプチャーしました。

シーケンスは、NextSeq 1000/2000 P2 Reagent Kit(200 cycles)(イルミナ、カタログ番号20040557)を用いて、NextSeq 2000システムで100 bp × 2のラン設定で実施しました。比較のため、同一ライブラリーをNextSeq 500/550 High-Output

Kit v2.5(300 cycles)(カタログ番号20024908)を用いて、NextSeq 550システムでも100 bp × 2のラン設定で実施しました。各装置につき8サンプルを1回のランで解析しました。

二次データ解析は、NextSeq 2000システムに内蔵されており、またBaseSpace™ Sequence Hubでも使用可能なDRAGEN™ Enrichment Pipeline v3.4で実施しました。バリエーション精度はPlatinum Genomes 2016 v1.0データセットに対して評価しました。⁴

mRNAシーケンス

メッセンジャーRNA(mRNA)ライブラリーは、Universal Human Reference RNA(Coriell Institute for Medical Research)をTruSeq™ Stranded mRNA Library Prep Kit(イルミナ、カタログ番号20020594)を用いて調製しました。

シーケンスは、NextSeq 1000/2000 P2 Reagent Kit(200 cycles)を用いて、NextSeq 2000システムで76 bp × 2のラン設定で実施しました。比較のため、同一ライブラリーをNextSeq 500/550 High-Output Kit v2.5(300 cycles)を用いて、NextSeq 550システムで76 bp × 2のラン設定でも実施しました。1ランにつき、24サンプルをプールのし、各装置で解析しました。

二次データ解析は、NextSeq 2000システムに内蔵されており、BaseSpace Sequencing Hubでも使用可能なDRAGEN RNA Pipeline v3.5.112を使用して実施しました。データはGenome Reference ConsortiumのHuman GRCh38(h38アセンブリ)に対してアライメントを行いました。

シングルセルRNAシーケンス

シングルセルRNAシーケンス(scRNA-Seq)のサンプルは、単一ドナーの全血から分離した4プレックスの末梢血単核細胞(PBMC)から調製しました。ライブラリーは、Chromium Controller(10x Genomics、カタログ番号120223)のChromium Single Cell Gene Expression v3 Solution(10x Genomics、カタログ番号1000092)を用いて約1,000個のPBMCから調製しました。

シーケンスは、NextSeq 1000/2000 P2 Reagent Kit(200 cycles)を用いて、NextSeq 2000システムで実施しました。比較のため、同一ライブラリーをNextSeq 500/550 High-Output Kit v2.5(150 cycles)(イルミナ、カタログ番号20024907)を用いて、NextSeq 550システムでも実施しました。ラン設定は10x Genomicsから供給された次のパラメーターに従って設定しました: 28サイクルのリード1、8サイクルのインデックスリード、91サイクルのリード2。

データ解析はCell Ranger scRNA-Seq解析パイプライン(10x Genomics, Inc.)を使用し、リードのアライメント、クラスタリングの実施、遺伝子発現の解析を行いました。

* NextSeq 2000システムはNextSeq 1000システムと比べてよりハイスループットなシステムですが、性能およびデータ品質は同等です。そのため、本アプリケーションノートではNextSeq 2000システムのみをデータ比較試験のために使用しました。

結果

エクソームシーケンス

一次解析および二次解析メトリクスは、一塩基多型 (SNV) および挿入と欠失 (indel) 双方に対する精度とリコール率、ラン出力 (産出量)、エラー率、アライメントしたリード%、ターゲット領域平均カバレッジ深度、およびカバレッジ均一性について評価しました。NextSeq 550およびNextSeq 2000システムのいずれもシーケンス出力およびデータ品質について公表されている仕様を超えており、いずれのシーケンスプラットフォームでも優れた性能を認めました (表1)。これらのデータより、NextSeq 2000システムでのエクソームシーケンスの結果はNextSeq 550システムの結果と同等であり、いずれのプラットフォームも高品質なデータと高い精度のあるバリアントコールを産生したことが示されます。

表1: エクソームシーケンスの同等性能

	NextSeq 550 システム	NextSeq 2000 システムの ラン1	NextSeq 2000 システムの ラン2
一次ランメトリクス			
パスフィルター クラスター数	514.2M	573.0M	565.1M
産生量	103.9 Gb	114.6 Gb	113.1 Gb
Q30%	92.20%	94.14%	94.15%
エラー率	0.34%	0.16%	0.16%
二次解析メトリクス^a			
常染色体コール率	95.61%	96.90%	96.90%
アライメントしたリード%	99.15%	99.53%	99.50%
ターゲット領域平均 カバレッジ深度	78.05	81.87	82.42
カバレッジ均一性	93.20%	96.92%	96.94%
SNV精度	99.3%	99.56%	99.59%
SNVリコール率	94.1%	94.88%	94.94%
Indel精度	88.0%	95.34%	95.71%
Indelリコール率	83.3%	84.90%	84.97%

a. 12サンプルの平均

mRNAシーケンス

NextSeq 550およびNextSeq 2000システムのいずれもシーケンス出力およびデータ品質について公表されている仕様を超えました (表2)。mRNA-Seqによる特異的RNAターゲットの定量では、4つのレプリケートすべてにわたり、両プラットフォーム間で優れた一致を示しました ($R^2 > 0.99$) (図1)。これらのデータは、NextSeq 2000システムがmRNA発現量についてNextSeq 550システムと同等のデータ品質を産生するという結果を強く示しています。

表2: mRNA-Seqの同等性能

メトリクス	NextSeq 550 システム	NextSeq 2000 システムの ラン1	NextSeq 2000 システムの ラン2
パスフィルター クラスター数	422.4M	537.8M	531.4M
産生量	63.6 Gb	80.2 Gb	79.75 Gb
Q30%	92.85%	92.93%	92.28%
エラー率	0.27%	0.14%	0.16%
検出した遺伝子の 平均数 (n=24)	24,355	26,507	24,739
検出した転写物の 平均数 (n=24)	74,387	88,125	81,948

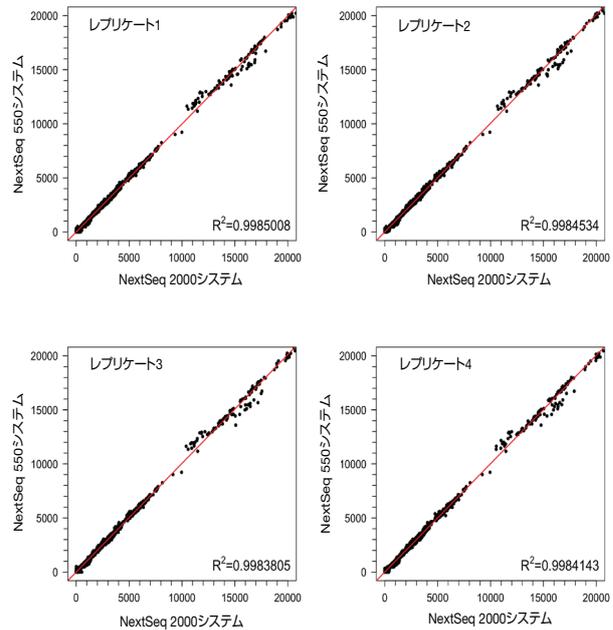


図1: mRNA-Seq結果の高い一致: NextSeq 550システムを用いたmRNA-Seqの特異的RNAターゲットの定量結果は、y軸にプロットされています。NextSeq 2000システムの結果はx軸にプロットされています。y = xのトレンドラインに沿ってデータの一致が認められます (4つのレプリケートすべてで $R^2 > 0.99$)。

シングルセル全トランスクリプトームシーケンス

NextSeq 550およびNextSeq 2000システムはシーケンス出力およびデータ品質について公表されている仕様を超えました (表3)。scRNA-Seqで分子バーコード (UMI) を用いた個々の細胞の定量では、4つのレプリケートすべてにわたり、両プラットフォーム間で優れた一致を示しました ($R^2 > 0.99$) (図2)。これらのデータは、NextSeq 2000システムがscRNA-Seq研究に対してNextSeq 550システムと同等のデータ品質を産生することをさらに示しています。

表3: scRNA-Seqの同等性能

メトリクス	NextSeq 550 システム	NextSeq 2000 システムの ラン1	NextSeq 2000 システムの ラン2
パスフィルター クラスター数	454.8M	498.7M	530.1M
産生量	53.2 Gb	58.3 Gb	62.0 Gb
Q30% R1	96.28%	94.05%	93.91%
Q30% R2	87.73%	91.35%	91.51%
エラー率	0.29%	0.14%	0.14%
検出した遺伝子数	18,529	18,541	18,438
細胞あたりのUMI数 中央値	5,220	5,175	5,132
細胞のリード割合	95.05%	94.95%	94.93%
推定細胞数	1,181	1,185	1,184

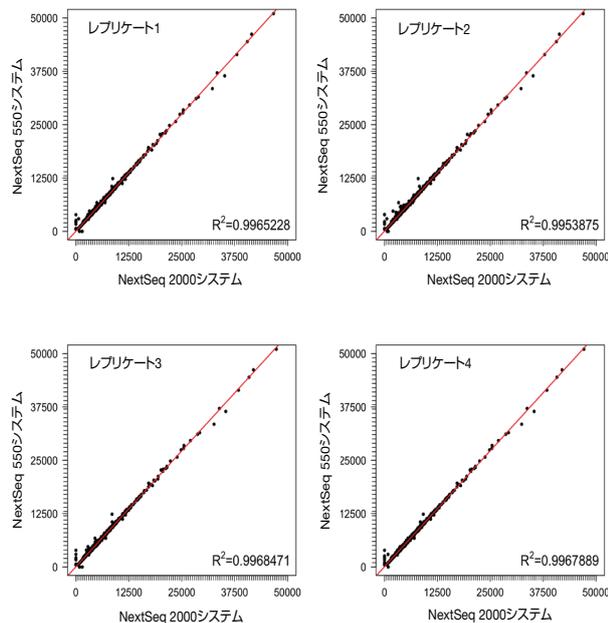


図2: scRNA-Seq結果の高い一致: NextSeq 550システムを用いたscRNA-Seqの個々の細胞(UMI)の定量結果は、y軸にプロットされています。NextSeq 2000システムの結果はx軸にプロットされています。y = x のトレンドラインに沿ってデータの一致が認められます(4つのレプリケートすべてで $R^2 > 0.99$)。

まとめ

NextSeq 1000およびNextSeq 2000シーケンスシステムは画期的なシステムデザイン、ケミストリーイノベーション、そして装置に内蔵されたインフォーマティクス解析ツールを備えており、ベンチトップ型シーケンスシステムで実現可能な研究領域を大きく広げます。これらのプラットフォームはシーケンスコストおよびラン性能に改善をもたらす一方で、ユーザーがイルミナに期待する高品質なデータを維持しています。NextSeq 550システムで一般的に実施されるエクソームシーケンス、mRNAシーケンス、シングルセルRNAシーケンスなどの重要なアプリケーションのデータについて、NextSeq 2000システムで発生されたデータと直接比較しました。結果として、NextSeq 550システムとNextSeq 2000システムの性能は高い相関性を示し、一貫性のある高品質なシーケンスデータを確実にお届けするというイルミナのコミットメントを反映することが示されました。

詳細はこちら

NextSeq 1000およびNextSeq 2000シーケンスシステムについての詳細は、jp.illumina.com/systems/sequencing-platforms/nextseq-1000-2000.htmlをご覧ください。

本ノートで参照したデータセットをダウンロードするには、basespace.illumina.com/datacentralをご覧ください。

参考文献

1. 社内資料。Illumina, Inc. 2017.
2. Based on a comparison of the top three industry-leading NGS platforms.社内資料。Illumina, Inc. 2016.
3. Ross MG, Russ C, Costello M, et al. *Characterizing and measuring bias in sequence data. Genome Biol.* 2013;14:R51.
4. Eberle MA, Fritzilas E, Krusche P, et al. *A reference data set of 5.4 million phased human variants validated by genetic inheritance from sequencing a three-generation 17-member pedigree. Genome Res.* 2017;27:157-164.

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

© 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。

Pub. No. 970-2020-001-B-JPN QB10338 09NOV2020

illumina[®]