

iSeq™ 100システムを用いたシングルセル 遺伝子発現ライブラリーのQCおよびリバランスング

NextSeq™ 550システムやNovaSeq™ 6000システムで高い深度のランを行う前に、マルチプレックスしたシングルセルライブラリーの重要なメトリクスを評価することで品質を評価し精度の高いリバランスングが可能になります。

はじめに

複雑な多細胞組織の成り立ちや、その分化過程についての一つひとつのシングルセルの生物学的寄与を理解することは、次世代シーケンス(NGS)テクノロジーにおいて主要なアプリケーションになりつつあります。シングルセルシーケンス実験のデータは、オンコロジー、免疫学、発生生物学などの多くのさまざまな分野で新しい発見を促しています。NGSは、1回のハイスループットのシーケンスランで同時に数千から数万の個別の細胞をプロファイルできる高い検出力を備えています。使用するシングルセルアッセイ、サンプルタイプ、および実験デザインにより、特定のシングルセルライブラリーのシーケンス要件が大幅に変わることがあります。¹

マルチプレックスしたライブラリー、特にシングルセル遺伝子発現プロファイリング(sc-GEX)ライブラリーについての構成の正確な評価は、ライブラリーのローディングを最適化する際に非常に重要なステップとなります。推定細胞数、遺伝子間/イントロン/エクソンコンテンツ、細胞のリード割合などのsc-GEXライブラリーの重要なメトリクスは、QubitやqPCRのような従来の定量法では、シーケンス実施前に評価することはできません。ハイスループットのsc-GEX実験の有効性を最大にするために、高い深度のNGSランを行う前にまず初めに浅い深度でこれらのライブラリーをシーケンスすることにより重要なメトリクスの評価とその後のリバランスングが可能になります。ライブラリーのクオリティコントロール(QC)は時間と費用を節約するだけでなく、より一貫性のあるシーケンス結果につながり、これによりデータ解析と解釈を簡略化することができます。

本アプリケーションノートは、NextSeq 550システム(カタログ番号SY-415-1002)またはNovaSeq 6000システム(カタログ番号20012850)のいずれかでフルスケールの高い深度のシーケンスランを行う前にsc-GEXライブラリーのQCおよびリバランスングを実施するためのiSeq 100システム(カタログ番号20021532)の使用例を示しています(図1)。本アプリ

ケーションノートの一部では、10x Genomics Inc.と協力して細胞分離およびライブラリー調製を行いました。

iSeq 100システムによるライブラリーQC

iSeq 100システムとシンプルかつ簡略化されたワークフローを使用して、複数のsc-GEXライブラリーをプールし、浅い深度でシーケンスを行い、重要なsc-GEXライブラリーメトリクスについて解析を行うことが可能です。これらのメトリクスからプール内のライブラリーに関する重要な情報が得られます。初めに、これらのメトリクスを用いてライブラリー生成がうまくいったかのスクリーニングができるため、低い細胞収量または低いコンテンツのライブラリーを除外することができます。次に、これらのメトリクスはライブラリーのリバランスングを行うための情報として、複数のsc-GEXライブラリー内でインデックスあたり、または細胞あたりのシーケンス深度にライブラリーをノーマライズするために使用することができます。プールしたsc-GEXライブラリーのバランスを適切に取ることにより、インデックスあたりのリード数、または細胞あたりのリード数をより均一にすることができます。最終的に、NextSeq 550システムまたはNovaSeq 6000システムでシーケンスを行う際に、iSeq 100システムのシーケンスの結果をsc-GEXライブラリーの特定のメトリクスの予測として使用することができます。

手法

ライブラリー調製およびプール

10x Genomics Chromium Controller(カタログ番号120223)上でChromium Single Cell Gene Expression v3 Solutionを用いて、個別の16個のsc-GEXライブラリーを約1,000個の末梢血単核細胞(PBMC)から調製し、8つのsc-GEXライブラリーを約5,000個のPMBCから調製しました(図2)。ライブラリー濃度はAgilent 2100 Bioanalyzer Instrument(カタログ番号G2939BA)で測定し(qPCRによる定量も代替法として実施可

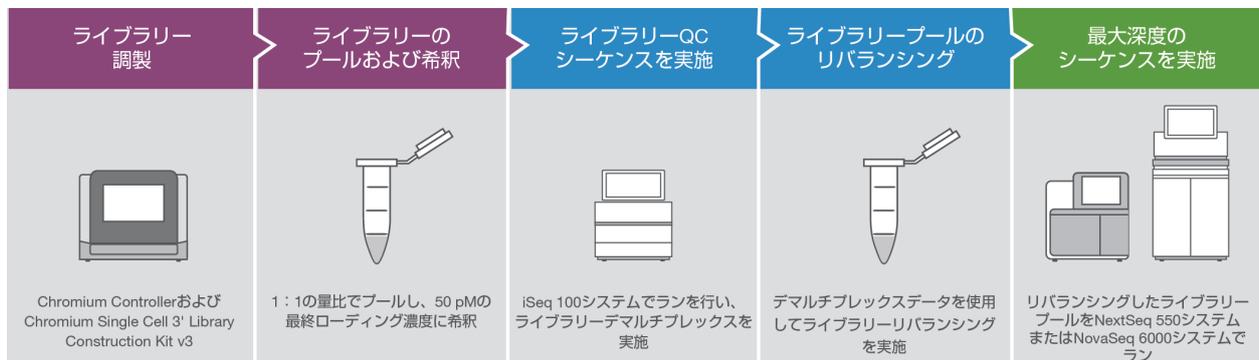


図1: iSeq 100システムによるシングルセルライブラリーのQC: iSeq 100システムは、高い深度のシーケンスを行う前にsc-GEXライブラリーのクオリティコントロールを行うための迅速かつシンプルなワークフローの一部です。

* NGSを使用してさまざまなシングルセルプロファイリングを行うアプリケーションが登場していますが(例、sc-ATAC、sc-CNV、sc-免疫プロファイリング)、本アプリケーションノートはsc-GEXに重点を置いています。その他のライブラリータイプに対して、ここで記述された技術を適用する際のベストプラクティスについては、イルミナのアプリケーションサイエンティストまたはイルミナテクニカルサポートにお問い合わせください。

能)、次にそれぞれ2 nMに希釈しました。約1,000個のPBMCからなる16個のsc-GEXライブラリーを、4つのサンプルごとに4つの異なるサブプールに1:1の容量比でプールし、50 pMの最終ローディング濃度に希釈しました。さらに、約1,000個のPBMCからなる16個のsc-GEXライブラリーを、1:1の容量比で1つのプールにプールし、50 pMの最終ローディング濃度に希釈しました。約5,000個のPBMCからなる8つのsc-GEXライブラリーを、1:1の容量比でプールし、50 pMの最終ローディング濃度に希釈しました。

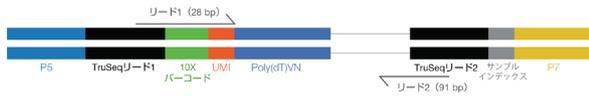


図2: Chromium Single Cell 3' Gene Expression Libraryの概要: リード1は28 bp長で10x Genomicsの細胞バーコードと固有識別子(UMI)をコードしています。インデックス1は8 bpであり、同じシーケンスランで複数の10xライブラリーをマルチプレックスすることができます。リード2は91 bp長で、発現したRNAに由来するインサートを含んでいます。本アプリケーションノートで示したデータはSingle Cell Gene Expression v3 Libraryに基づくものですが、10x GenomicsではSingle Cell Gene Expression v3.1 (Next GEM)と同等の性能を予測しています。

シーケンス

本アプリケーションノートで記述したすべてのシーケンスランは、BaseSpace™ Sequence Hubで公開されています(詳細についてはAppendixをご覧ください)。約1,000個のPBMCの各サブプールは、高深度QCとしてライブラリーあたり100万リードをターゲットにしてiSeq 100システム(ソフトウェアv1.4)で別々にランを行いました(ラン1~4)。約1,000個のPBMCからなる16個のライブラリーを統合したライブラリー、および約5,000からなる8つのライブラリーを統合したライブラリーはiSeq 100システムで別々にランを行いました(ラン5、7、および9)。リバランスしたライブラリープールは、NextSeq 500/550 High Output Kit v2.5(150 cycles, カタログ番号20024907, ラン6および8)を用いてNextSeq 550システム(RTAバージョン2.4.11)、またはNovaSeq S2 Reagent Kit(200 cycles, カタログ番号20012861, ラン10)を用いてNovaSeq 6000システム(RTAバージョン3.4.4)でランを行いました。

データ解析

Cellrangerソフトウェアパッケージ(v3.0.2, 10x Genomics Inc)でmkfastqコマンドを使用してデマルチプレックスを行い、BCLファイルをFASTQファイルに変換しました。FASTQファイルの生成後、iSeq 100システム、NextSeq 550システムおよびNovaSeq 6000システムのランからデマルチプレックスしたリードについてさまざまなリード深度でseqtkにあるソフトウェアツールを用いてサブサンプリングしました。リード深度は、iSeq 100システムについては細胞あたり約1,500リードから5リード、NextSeq 550システムについては細胞あたり約25,000リードから5リード、NovaSeq 6000システムについては細胞あたり約40,000リードから5リードまでの範囲としました。Cellrangerのカウントコマンドを使用し、refdata-cellranger-GRCh38-3.0.0の参照トランスクリプトームの指定とシングルセルフェューチャーカウントを生成するためにデフォルトの推定細胞数を3,000に指定しました。

結果

インデックス数に基づくライブラリーのリバランス

iSeq 100システム(ラン5)でランを行った約1,000個のPBMCからなる全16ライブラリーを統合したプールから得られたシーケンス結果は、95%以上の塩基がQ30以上のスコアを示しました(図3)。統合したプールを用いたiSeq 100システムのランをデマルチプレックスして得られたライブラリー数は、そのプール中の約1,000個のPBMCからなる16個のライブラリーをリバランスするために使用しました。初めに、各ライブラリーに対してノーマライズしたインデックス数の算出は、全16インデックス中のフィル

ターをパスした(PF)リードの合計で各インデックスのPFリードを割って求めました。次に、各ライブラリーに対するローディング係数は、すべてのサンプルでノーマライズされた最も高いインデックス数と、該当するインデックス数の割合から算出しました(表1)。

$$\text{Norm. \% Index}_i = \frac{\text{reads PF}_i}{\sum_{i=1}^n \text{reads PF}_n}$$

$$\text{Loading Factor}_i = \frac{\max(\text{Norm. \% Index}[n])}{\text{Norm. \% Index}_i}$$

各ライブラリーについて、元のインプット容量に計算したローディング係数をかけることで、NextSeq 550システムで高い深度のシーケンスを行う前にインデックスのリバランスを行ってプールするための新しい容量を得ることができます。インデックスをリバランスしたプールはBioanalyzerで定量を行い、初めは1.8 pMの最終ローディング濃度でシーケンスを行いましたが、クラスター密度を最適化するために、最終濃度を3.6 pMに増加しました。10x Genomicsによって推奨されているため、ユーザーは1.8 pMの最終ローディング濃度を遵守する必要がありますが、必要に応じて調整します。3.6 pMでのランのシーケンス結果は、90%以上の塩基がQ30以上のスコアを示しました(ラン6、図4)。約1,000個のPBMCからなる16個のライブラリーのインデックス数についてリバランスを行う前後の変動係数(CV)を比較したところ、10.41%から3.65%に減少することが示されました(図5)。

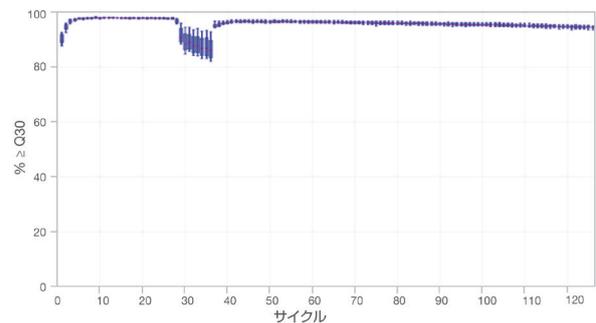


図3: iSeq 100システムのラン5のQ30プロット: iSeq 100システムでランを行った約1,000個のPBMCからなる16個のライブラリーすべてを統合したプールのシーケンス結果は、95%以上の塩基がQ30以上のスコアを示しました。

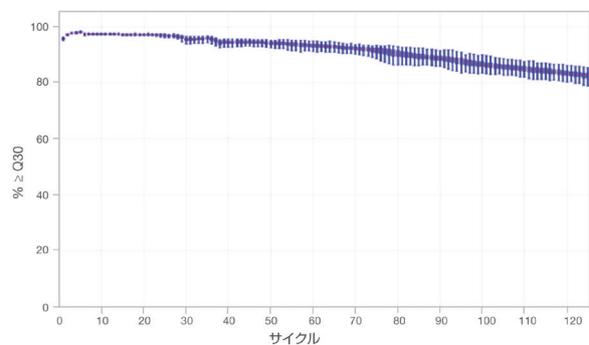


図4: NextSeq 550システムのラン6のQ30プロット: 約1,000個のPBMCからなる16個のライブラリーすべてについてインデックスをリバランスしたプールをNextSeq 550システムでランを行ったシーケンス結果は、90%以上の塩基がQ30以上のスコアを示しました。

表1: iSeq 100システムのインデックス数に基づく約1,000個のPBMC sc-GEXライブラリーのリバランシング

ライブラリー名	PFリード	ノーマライズしたインデックスの割合(%)	ローディング係数
S1_W1_R1	264,639	5.51%	1.43
S1_W1_R2	310,292	6.46%	1.22
S1_W1_R3	289,073	6.02%	1.31
S1_W1_R4	252,327	5.25%	1.50
S1_W2_R1	295,547	6.15%	1.28
S1_W2_R2	284,929	5.93%	1.33
S1_W2_R3	293,135	6.10%	1.29
S1_W2_R4	299,982	6.25%	1.26
S2_W1_R1	379,568	7.90%	1.00
S2_W1_R2	359,625	7.49%	1.05
S2_W1_R3	303,996	6.33%	1.25
S2_W1_R4	310,468	6.47%	1.22
S2_W2_R1	281,004	5.85%	1.35
S2_W2_R2	272,773	5.68%	1.39
S2_W2_R3	317,334	6.61%	1.19
S2_W2_R4	287,337	5.98%	1.32

ノーマライズした最も高いインデックス数は赤で示しています。

細胞あたりのリード数に基づくライブラリーのリバランシング

リバランシングを行う別の方法を用いて、インデックスあたりのリード数ではなく細胞あたりの平均リード数を求めることでより均一な分布を達成することができます。これを立証するために、約1,000個のPBMCからなる16個のライブラリーを統合した2つ目のライブラリーについてiSeq 100システムでラン(ラン7)を行い、Cellranger解析によって得られた細胞あたりの平均リード数の出力について、2つ目のライブラリープールをリバランシングしました。各ライブラリーについて、すべてのサンプル中の細胞あたりの最も高い平均リード数と該当するライブラリーの細胞あたりの平均リード数との割合としてローディング係数を算出しました(表2)。

$$\text{Loading Factor}_i = \frac{\max(\text{Mean Reads per Cell}[n])}{\text{Mean Reads per Cell}_i}$$

各ライブラリーに対して元のインプット容量に計算したローディング係数をかけることで、NextSeq 550システムでシーケンスを行う前に細胞あたりの平均リード数をリバランシングしてプールするための新しい容量を得ることができます。

細胞あたりの平均リードに対してリバランシングしたライブラリープールも3.6 pMの最終ローディング濃度に希釈し、NextSeq 550システムでランを行いました(ラン8)。約1,000個のPBMCからなる16個のライブラリーの細胞あたりの平均リード数についてリバランシングを行う前後のCVを比較したところ、13.87%から5.43%の減少が示されました(図6)。

より複雑なsc-GEXライブラリーのライブラリーリバランシング

より複雑なPBMCライブラリーについて同様の手法で検証しました。約5,000個のPBMCからなる8個のライブラリーをプールし、上述のようにiSeq 100システムでランを行いました(ラン9)。このランを用いて、インデックス数(表3)または細胞あたりの平均リード数(表4)のどちらに対してもライブラリーのリバランシングを行いました。リバランシングした約5,000個のPBMCからなるsc-GEXライブラリーの各プールを133 pMに希釈し、NovaSeq XP ローディングプロトコルを用いてS2フローセルの別々のレーンにロードし、NovaSeq 6000システムでランを行いました(ラン10)。シーケンス結果では90%以上の塩基がQ30以上のスコアを示しました(図7)。8つのライブラリー中のリバランシングを行う前後のCVを比較すると、インデックス数については15.01%から6.79%に減少し(図8A)、細胞あたりの平均リード数については15.56%から8.86%に減少したことが示されました(図8B)。

表2: iSeq 100システムのランから得られた細胞あたりの平均リード数に対する約1,000個のPBMC sc-GEXライブラリーのリバランシング

ライブラリー名	PFリード	推定細胞数	細胞あたりの平均リード数	ローディング係数
S1_W1_R1	218,691	841	260	1.64
S1_W1_R2	256,639	733	350	1.22
S1_W1_R3	248,325	707	351	1.21
S1_W1_R4	284,751	729	390	1.09
S1_W2_R1	265,240	857	309	1.38
S1_W2_R2	200,026	757	264	1.61
S1_W2_R3	266,009	800	332	1.28
S1_W2_R4	246,575	874	282	1.51
S2_W1_R1	242,390	846	286	1.49
S2_W1_R2	248,237	739	335	1.27
S2_W1_R3	204,978	715	286	1.49
S2_W1_R4	262,200	731	358	1.19
S2_W2_R1	275,822	854	322	1.32
S2_W2_R2	330,183	774	426	1.00
S2_W2_R3	262,099	807	324	1.31
S2_W2_R4	242,838	867	280	1.52

細胞あたりの最も高い平均リード数は赤で示しています。

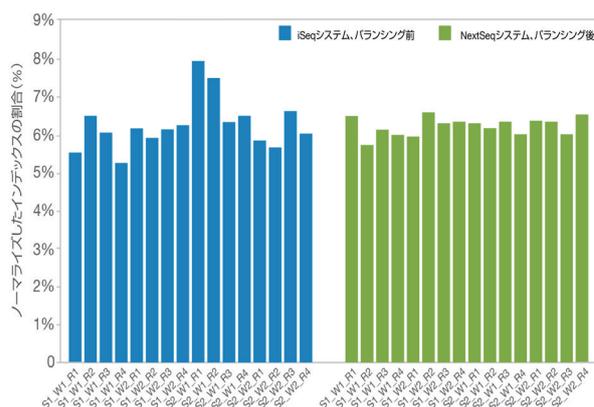


図5: 約1,000個のPBMC sc-GEXライブラリーのインデックスリバランシング: iSeq 100システムでランを行った16個のライブラリーのプール(ラン5)のインデックス数のCVは10.41%となりました(平均値=6.25%, SD=0.65%)。これらライブラリーのインデックスをリバランシングし、NextSeq 550システムでラン(ラン6)を行った後、CVは3.65%となりました(平均値=6.25%, SD=0.23%)。

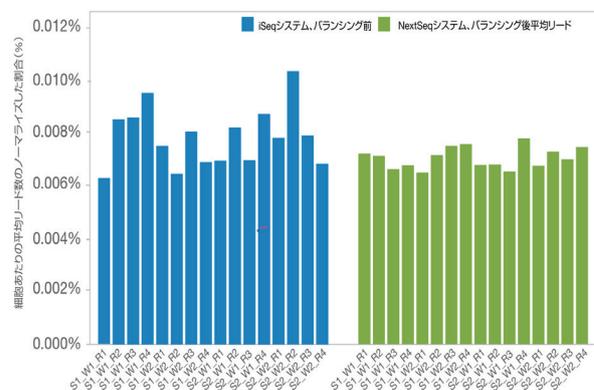


図6: 約1,000個のPBMC sc-GEXライブラリーの細胞あたりの平均リード数: iSeq 100システムでランを行った16個のライブラリーのプール(ラン7)は、ノーマライズ後の細胞あたりの平均リード数のCVが13.87%となりました(平均値=7.95E-3%, SD=1.1E-3%)。リバランシングを行ったライブラリーをNextSeq 550システムでラン(ラン8)を行った場合、CVは5.43%でした(平均値=7.19E-3%, SD=3.91E-4%)。

表3：iSeq 100システムのインデックス数に基づく約5,000個のPBMC sc-GEXライブラリーのリバランスング

ライブラリー名	PFリード	ノーマライズしたインデックスの割合(%)	ローディング係数
T1_W1_R1	554,168	10.69%	1.54
T1_W1_R2	683,013	13.18%	1.25
T1_W1_R3	853,552	16.47%	1.00
T1_W1_R4	625,127	12.06%	1.37
T1_W2_R1	622,075	12.00%	1.37
T1_W2_R2	528,790	10.20%	1.61
T1_W2_R3	596,223	11.50%	1.43
T1_W2_R4	720,780	13.90%	1.18

ノーマライズした最も高いインデックス数は赤で示しています。

表4：iSeq 100システムのランから得られた細胞あたりの平均リード数に対する約5,000個のPBMC sc-GEXライブラリーのリバランスング

ライブラリー名	PFリード	推定細胞数	細胞あたりの平均リード数	ローディング係数
T1_W1_R1	554,168	4545	121	1.64
T1_W1_R2	683,013	4681	145	1.37
T1_W1_R3	853,552	4286	199	1.00
T1_W1_R4	625,127	4194	149	1.34
T1_W2_R1	622,075	4132	150	1.33
T1_W2_R2	528,790	4206	125	1.59
T1_W2_R3	596,223	4148	143	1.39
T1_W2_R4	720,780	4203	171	1.16

細胞あたりの最も高い平均リード数は赤で示しています。

sc-GEXライブラリーの二次解析

推定細胞数

ライブラリーを構成する細胞のsc-GEXライブラリーメトリクスを予測するため、ライブラリー細胞の複雑性の制限や浅いシーケンスを行う能力を評価するための推定を指定せず、追加のCellranger解析を実施しました。約1,000個のPBMCからなる16個のsc-GEXライブラリー中では、iSeq 100システム(ラン5)によって推測される細胞数(約20リード以上/細胞)とNextSeq 550システム(ラン6)によって推測される細胞数(約25,000リード/細胞)の間に約7.8~8.9%の平均変動がありました(図9)。約5,000個のPBMCからなる8つのsc-GEXライブラリー中では、iSeq 100システム(ラン9)によって推測される細胞数(約20リード以上/細胞)と、NovaSeq 6000システム(ラン10)によって推測される細胞数(約40,000リード/細胞)の間に約21~23%の平均変動がありました(図10)。細胞数の絶対数の推定では浅い深度と高い深度で行ったシーケンスの間に、このような高い平均値の変動が示されましたが、相対的な推定細胞数は、1細胞あたり約20リードから約40,000リードにサブサンプリングされた結果において一定でした。

細胞あたりの平均リード数の範囲内(図9および10)で推定細胞数を比較することにより、リード深度に対して異なる反応を示す3つの領域が明らかになりました。細胞あたり20リードを下回る極端に低い深度では、認められたすべてのバーコードが細胞として同定されているため、推定細胞数は不安定です(Barcode Rankプロットは紺青色の細胞のみを示しており、バックグラウンドがありません)。細胞あたり20リードを超えるセルコーリングアルゴリズムでは細胞とバックグラウンドの識別が認められるようになります(Barcode Rankプロットは紺青色の細胞と灰色のバックグラウンドのいずれも示しています)。

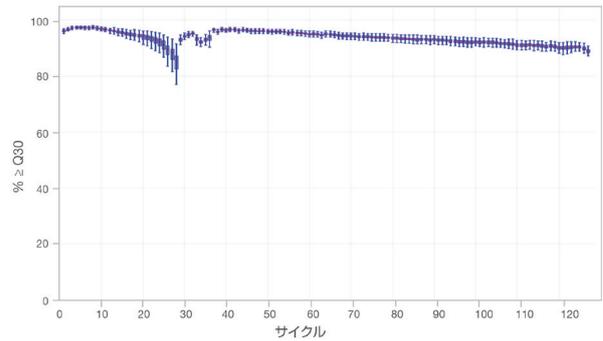


図7：NovaSeq 6000システムのラン10のQ30プロット：約5,000個のPBMCからなる8つのライブラリーすべてについてリバランスングしたプールをNovaSeq 6000システムでランを行ったシーケンス結果は、90%以上の塩基がQ30以上のスコアを示しました。

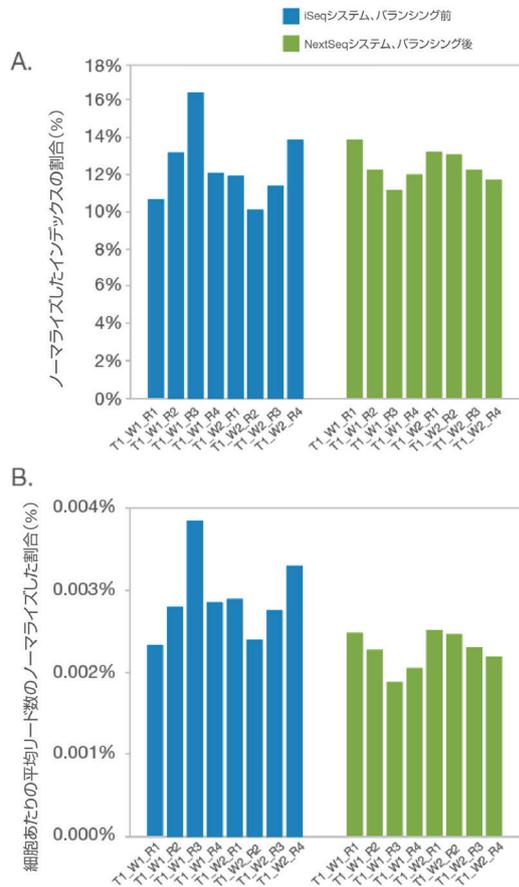


図8：約5,000個のPBMC sc-GEXライブラリーのリバランスング：iSeq 100システムでランを行った8つのライブラリーのプールは(ラン9)、(A)インデックス数のCVが15.01%であり(平均値=12.5%、SD=1.88%)、(B)細胞あたりの平均リード数をノーマライズした割合のCVが15.56%となりました(平均値=2.9E-3%、SD=4.51E-4%)。リバランスングを行ったライブラリーをNovaSeq 6000システムでラン(ラン10)を行った場合、(A)インデックス数のCVおよび(B)細胞あたりの平均リード数をノーマライズした割合のCVはそれぞれ、6.79%(平均値=12.5%、SD=0.85%)および8.86%(平均値=2.3E-3%、SD=2.04E-4%)でした。注記：より高い深度のシーケンスによってより多くの細胞が特定されるため、NovaSeq 6000システムでランを行った細胞あたりの平均リード数をノーマライズした割合は全体的に減少します。

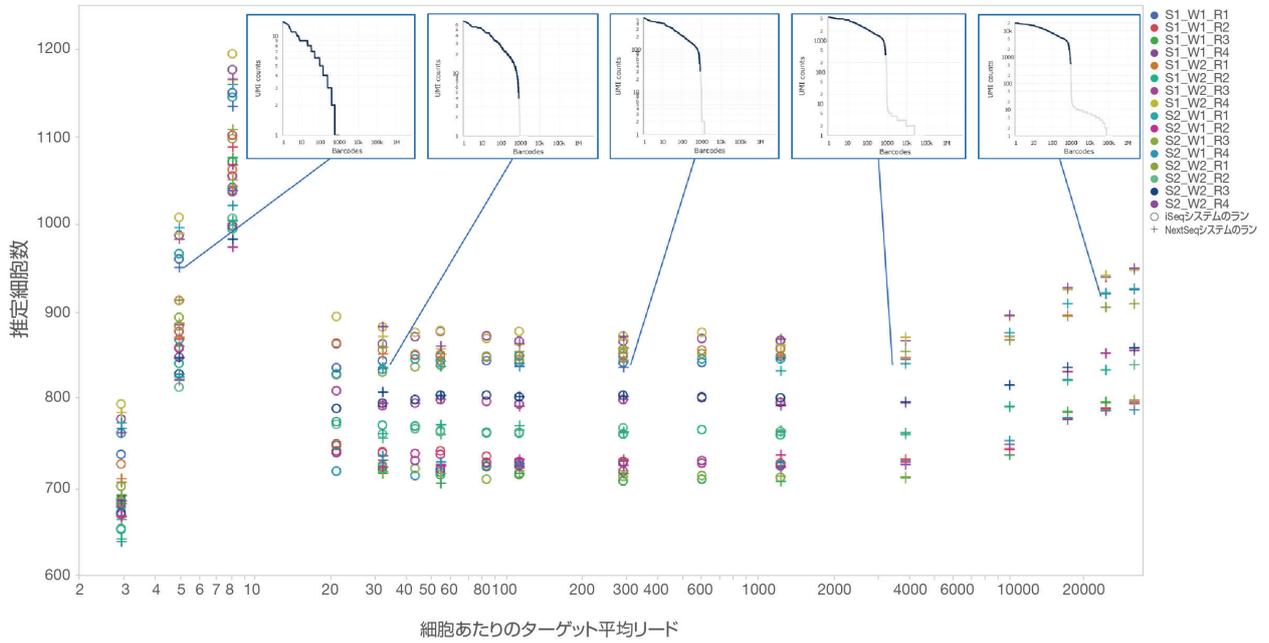


図9：約1,000個のPBMC sc-GEXライブラリーの推定細胞数：iSeq 100システム(ラン5)およびNextSeq 550システム(ラン6)で行ったすべてのランのデータおよびサブサンプリングしたデータから得られた各ライブラリーに対する推定細胞数のプロット。サブサンプリングしたiSeq 100システムとNextSeq 550システムのランから得られた細胞あたり20リード以上の推定細胞数とNextSeq 550システムの最大深度のランとの比較では平均変動は7.8~8.9%を示しています。プロットを図示するために示された選択したリード深度でのS1_W1_R1に対するBarcode Rankプロットはリード深度に反応しています。バーコードラベルされた細胞は紺青色線で示され、バックグラウンドは灰色で示されています。

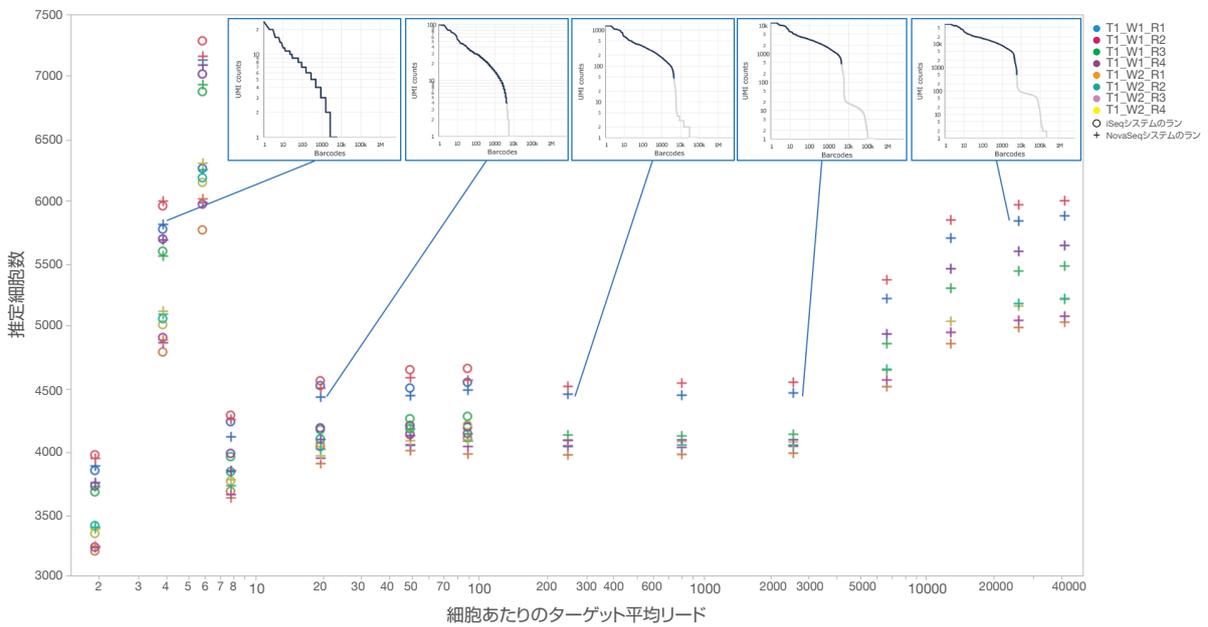


図10：約5,000個のPBMC sc-GEXライブラリーの推定細胞数：iSeq 100システム(ラン9)とNovaSeq 6000システム(ラン10)で行ったすべてのランのデータおよびサブサンプリングしたデータから得られた各ライブラリーに対する推定細胞数のプロット。サブサンプリングしたiSeq 100システムとNovaSeq 6000システムのランから得られた細胞あたり20リード以上の推定細胞数とNovaSeq 6000システムの最大深度のランとの比較では平均変動は21~23%を示しています。プロットを図示するために示された選択したリード深度でのT1_W1_R1に対するBarcode Rankプロットはリード深度に反応しています。バーコードラベルされた細胞は紺青色線で示され、バックグラウンドは灰色で示されています。

最終的に、リード深度が最大のラン深度に到達すると、推定細胞数が漸的に増加します。これは、**エンブティードロップ法**によってRNAプロファイルに基づき細胞としてUMIカットオフを下回るバーコードが特定されるためです（Barcode Rankプロットでは紺青色の細胞が低いUMI値まで伸長していることを示しています）。これらの3領域の推定細胞数の精度を考慮して、イルミナでは生成したBarcode Rankプロットを確認して、バックグラウンドが識別されていることを確認することを推奨しています。細胞あたり20リードの閾値では、複数のsc-GEXライブラリー中で理論上最大200,000個の細胞がiSeq 100システムで1回のランでアッセイできる可能性を示しています。sc-GEXライブラリー中の細胞あたりの平均リード数がiSeq 100システムでランを行う前にバランスがされていないと仮定すると、iSeq 100システムですべてのsc-GEXライブラリーのランを同時に行う場合、上限設定は合計50,000細胞にすることが推奨されます。

細胞あたりのリード分布がサンプルタイプ、サンプルクオリティおよびライブラリーの複雑性に基いて変動することがあることに留意してください。ここではヒトのPBMCに基づく推奨を概説していますが、他のサンプルタイプについては、追加のシーケンスや他のクオリティコントロールが必要になる場合があります。特に、Barcode Rankプロットの予測される「ニー」（曲がり部分）を検証することによりシーケンスQCが低い実験からの結果をサポートすることができます。

シーケンス深度およびプラットフォーム間の細胞バーコード数

同一の細胞バーコードが異なる深度および複数のプラットフォーム間で同一の相対的比率でシーケンスされていることを確認するために、iSeq 100システム（ラン5）とNextSeq 550システム（ラン6）（図11）間、およびiSeq 100システム（ラン9）とNovaSeq 6000システム（ラン10）（図12）間で最大

深度での細胞バーコードあたりのUMI数を比較して相関プロットを作成しました。

予測可能なライブラリーメトリクス

Cellrangerソフトウェアはsc-GEXライブラリーを解析する際に、推定細胞数だけでなくさまざまなメトリクスを出力します。浅いシーケンス深度にダウンサンプリングした場合、これらのメトリクスはさまざまな割合で反応します。ゲノム、遺伝子間、イントロン領域、およびエクソン領域にマップしたリードの割合など、一部のsc-GEXライブラリーメトリクスでは、検証した深度の範囲すべてにわたって均一性を保っています（図13）。このような安定性から、浅いシーケンス深度であっても、これらの特定のメトリクスは同一ライブラリーに対してより高い深度のランメトリクスも予測できることが示されています[†]。

シーケンス飽和、細胞あたりの遺伝子数中央値、および細胞あたりのUMI数中央値など、他のsc-GEXライブラリーメトリクスは、シーケンス深度に対してよりリニアな相関を示した後、高い深度で飽和することを示します（図14）。浅い深度でのこのような線形依存性は、これらのメトリクスが同一ライブラリーに対してより高い深度ではランメトリクスを予測できない可能性があることを示しており、注意して評価する必要があります。

細胞のリード割合などのメトリクスは推定細胞数と同様にシーケンス深度に反応し、細胞あたり20リード以上で相対的安定性および優れたパフォーマンスを示します（図15A）。最後に、検出された合計遺伝子数は浅い深度で細胞あたりのリード数に対して高い感度を示し、高い深度で平坦化します（図15B）。浅い深度で検出された合計遺伝子数は、同一ライブラリーに対してより高い深度ではランメトリクスを予測できない可能性があることを示しており、注意して評価する必要があります。

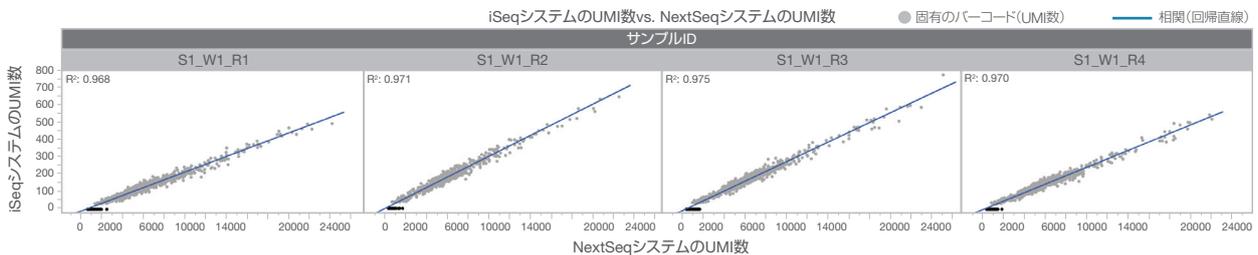


図11：iSeq 100ラン5およびNextSeq 550ラン6からの細胞あたりのUMI数：S1_W1_R1からS1_W1_R4までのプロットしたサンプルはシーケンスシステム間で高いR²値 (> 0.96)を示しています。iSeq 100システムでは認められず、NextSeq 550システムで認められる細胞（黒色）はNextSeq 550システムでのフルランで最も低い存在量にあるすべての細胞を示しています。

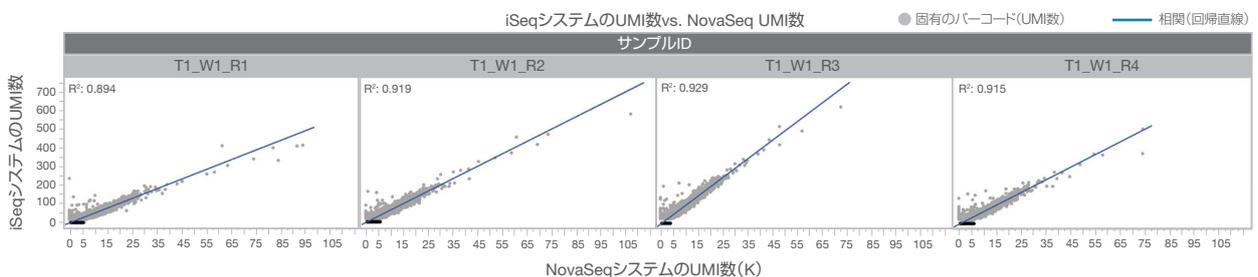


図12：iSeq 100ラン9およびNovaSeq 6000ラン10からの細胞あたりのUMI数：T1_W1_R1からT1_W1_R4までのプロットしたサンプルはシーケンスシステム間で高いR²値 (> 0.89)を示しています。iSeq 100システムでは認められず、NovaSeq 6000システムで認められる細胞（黒色）はNovaSeq 6000システムでのフルランで最も低い存在量にあるすべての細胞を示しています。

[†] 残りのすべてのグラフ（図13～15）では、iSeq 100システムおよびNextSeq 550システムでのサンプルS1_W1_R1のランを示していますが、認められる傾向はすべてのサンプルおよびプラットフォームの組み合わせでも同一性を保っています（データ示さず）。

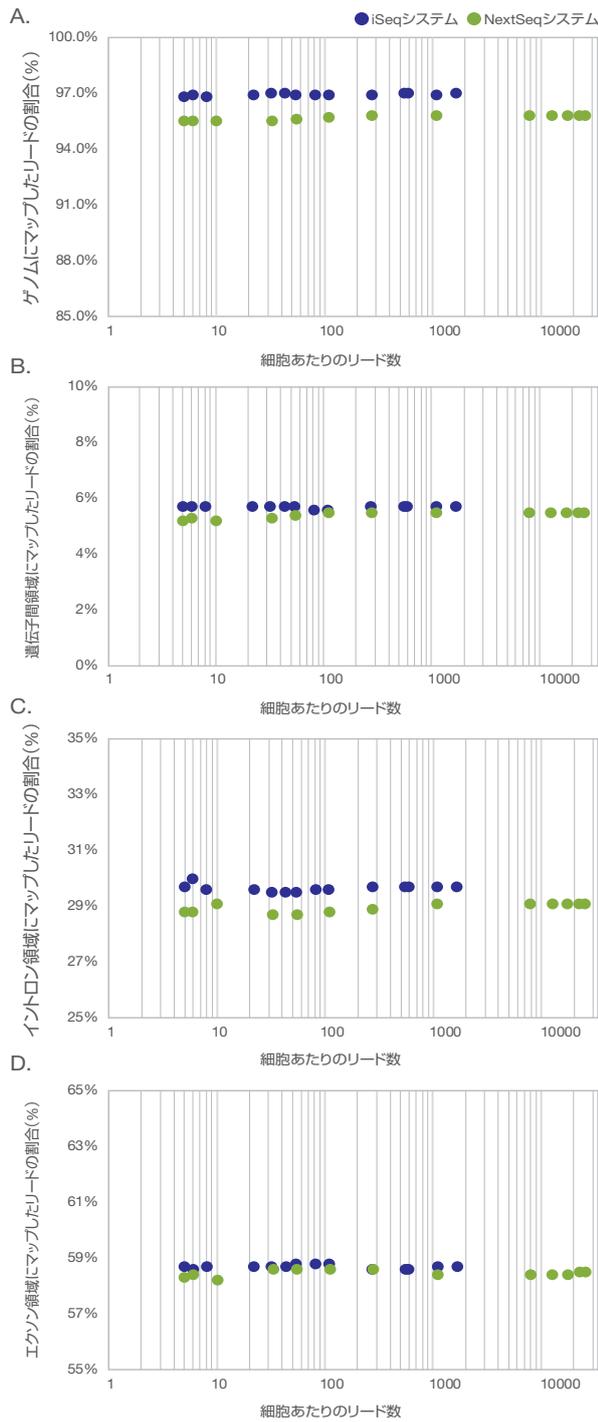


図13：すべてのシーケンス深度でのsc-GEXライブラリーマトリクスの安定性：(A)ゲノム、(B)遺伝子間領域、(C)イントロン領域、および(D)エクソン領域にマップしたリードの割合。

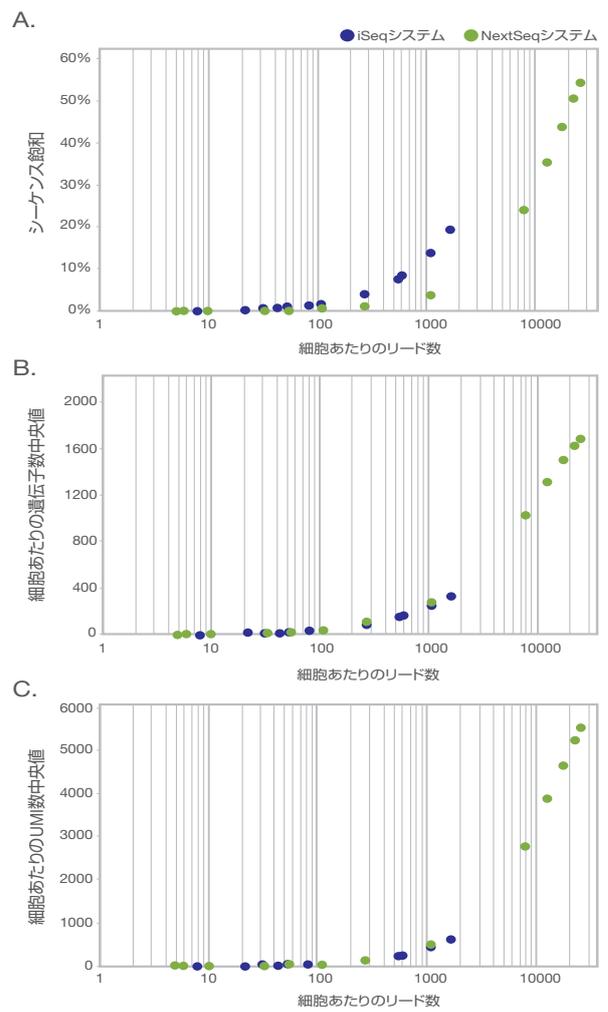


図14：シーケンス深度に依存して線形を示すsc-GEXライブラリーマトリクス：(A)シーケンス飽和、(B)細胞あたりの遺伝子数中央値、および(C)細胞あたりのUMI数中央値。

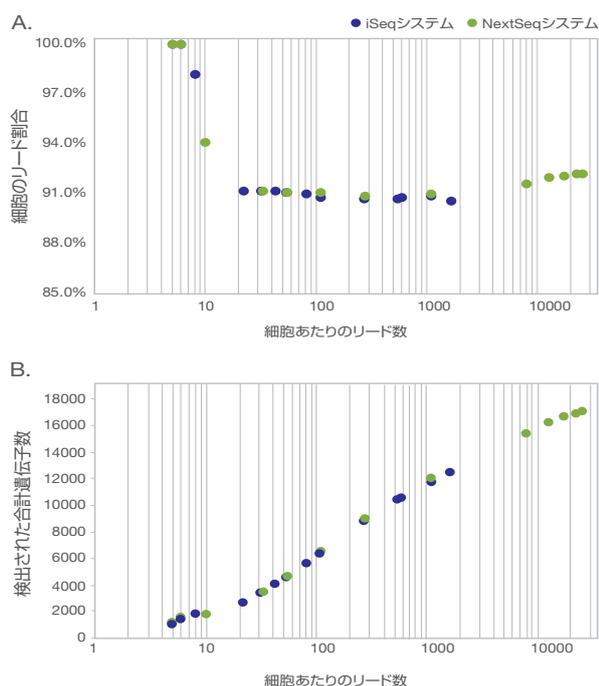


図15：細胞のリード割合および検出した合計遺伝子数：(A)細胞のリード割合は推定細胞数と同様にリード深度に反応しますが、(B)検出した合計遺伝子数は浅い深度で高い感度を示し、高い深度では平坦化します。

まとめ

iSeq 100システムでは、高い深度でNGSのマルチプレックスランを行う前にライブラリープールのリバランシングを行うことが可能です。リバランシングを行うことにより、それぞれのサンプルのリード数をインデックス数をターゲットにして均一化したり、サンプル内の細胞あたりリード数をターゲットにして均一化をすることができます。sc-GEXライブラリーでは、iSeq 100システムで低い深度でシーケンスを行うことにより、特定のライブラリーメトリクスを予測することができ、その後高い深度でのシーケンス解析に移ることができます。iSeq 100システムによる低い深度でのシーケンスがその他のシングルセルシーケンスアッセイに適用できるかどうかを検証し、特定のsc-ライブラリーの目的のメトリクスに対する予測を維持するために、既存の高い深度のシーケンスランからデータをサブサンプリングし、本アプリケーションノートで記述したようにCellrangerソフトウェアを用いてBarcode Rankプロット(図13~15)を作成して解析することが推奨されます。サンプルのマルチプレックスを行う量およびサンプルあたりの予測される細胞数を考慮すると、既存のデータから作成されたBarcode Rankプロットに関してiSeq 100システム上で予測されるリード深度を測定して評価することで、iSeq 100システムの能力を評価し、sc-ライブラリー中の特定のプールしたセットについてsc-ライブラリーメトリクスを予測することができます。

詳細について

イルミナのシーケンスシステムに関するより詳細な情報につきましては、jp.illumina.com/systems/sequencing-platforms.htmlにアクセスしてください。

参考文献

1. Satijalab.org. Accessed December 5, 2019.
2. github.com/lh3/seqtk. Accessed December 5, 2019.

Appendix

ラン番号	ラン名	リンク
ラン1	4-plex_iSeq_1k-PBMC-Unbal_subpool1	basespace.illumina.com/s/nyCmXl8LacEK
ラン2	4-plex_iSeq_1k-PBMC-Unbal_subpool2	basespace.illumina.com/s/Hn6TazAxiHKB
ラン3	4-plex_iSeq_1k-PBMC-Unbal_subpool3	basespace.illumina.com/s/7sW98OwhVcHk
ラン4	4-plex_iSeq_1k-PBMC-Unbal_subpool4	basespace.illumina.com/s/D5uIA5nGIPl
ラン5	16-plex_iSeq_1k-PBMC-Unbal_pool1	basespace.illumina.com/s/HPuoSEsrTv0k
ラン6	16-plex_NextSeq_1k-PBMC-Index-Bal	basespace.illumina.com/s/MUWFoQpDJip7
ラン7	16-plex_iSeq_1k-PBMC-Unbal_pool2	basespace.illumina.com/s/r1zw0mUqAbtL
ラン8	16-plex_NextSeq_1k-PBMC_MRPC-Bal	basespace.illumina.com/s/qkChF0pogWPF
ラン9	8-plex_iSeq_5k-PBMC-Unbal	basespace.illumina.com/s/tyxPesQs00g6
ラン10	8-plex_NovaSeq_5k-PBMC_Lane1-Index-Bal_Lane2-MRPC-Bal	basespace.illumina.com/s/fbRPbl3Y59Yf

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. 770-2019-029-A-JPN QB9175 19AUG2020

illumina®