

MiSeqシステムで行うシーケンスライブラリーのQC

大規模シーケンス解析前の、効率のよいライブラリー品質測定のための自動化されたQCワークフロー

はじめに

HiSeq® 2000システムの高いデータ産出性は、現時点で最大規模かつ最も複雑なゲノム研究を可能にしました。こうしたハイスループットシーケンスの効率を最大化するためには、ライブラリーの品質を知ることが重要です。低品質のライブラリーは大規模シーケンスの成功を揺るがし、高コストで時間がかかる複数回の再実験につながる可能性があります。HiSeqを使ったランに進む前にMiSeqシステムを使用してライブラリーの品質管理 (QC) を実施すれば、時間と費用を節約することができます。より優れたシーケンス結果が得られます。

MiSeqでは、わずか1日でクラスター密度、ライブラリーの多様性、重複率、GCバイアス、インデックスの表示などといった詳細なQCパラメーターを確認できます。MiSeqには、インサートサイズを正確に評価するためにペアエンド (PE) シーケンスを行えるという独自のすぐれた機能があります。ライブラリーのクラスター密度を求めるとも可能で、HiSeqにおけるクラスター密度の予測に利用できるように、データ産出性を最大化しながら再実験の回数を減少させることができます。このアプリケーションノートでは、MiSeqで行われるライブラリーのQCランについて説明しています。

QC基準

Illumina Experiment Managerというソフトウェアを使用して、ライブラリー内のサンプルをサンプルシートに定義します。次に、調製したライブラリーをアプライした試薬カートリッジとフローセルをMiSeqにセットし、クラスター形成およびシーケンスを自動で行います。2x26bpのランの後、MiSeq Reporterソフトウェアは装置上のコンピューターで自動的にライブラリーQCレポートを作成します。ライブラリーのQCワークフローでは変異検出を行いませんが、ライブラリーの品質とシーケンス性能に関係する重要なQCパラメーターを幅広く表示します (表1および図1A)。SUMMARYタブ (図1B) はランパラメータ、クラスターデータを示すグラフ、高値および低値、ミスマッチに関する情報を表示します。さらに、SAMPLEタブはリファレンス配列内の各位置をカバーするアライメントされたリード数と各サンプルのクオリティ値データを示すカバレッジの統計値を表示します (データ非掲載)。ユーザーはこれらの基準を定期的にモニターでき、将来的なハイスループットシーケンスランの指針となりうる標準的な基準のセットを確立することも可能です。

まとめ

フルスケールのシーケンスランに進む前に、MiSeqでシーケンスライブラリーの品質を評価することによって最大のシーケンス効率を確保し、正確なシーケンスデータとより均一なカバレッジをもたらします。MiSeqはHiSeqシステムと同じサンプル調製法と実績のあるSequencing by Synthesis (SBS) ケミストリーによるシーケンスを採用しており、ハイスループットシーケンスの実施前に調製済みライブラリーを1日で解析する

表1: シーケンスライブラリーのQC基準

基準	定義
Sample Name	サンプルシートに記載されたサンプル名
Clusters	サンプルあたりのクラスター数
Clusters %	インデックスが成功したクラスターの割合
% PF	パスフィルターのクラスターの割合
% Aligned R1	アライメントの割合 (リード1)
% Aligned R2	アライメントの割合 (リード2)
Length Median	ライブラリー断片長の中央値
Length Min	ライブラリー断片長の低パーセンタイル (標準偏差の3倍)
Length Max	ライブラリー断片長の高パーセンタイル (標準偏差の3倍)
Mismatch R1	ミスマッチ率 (リード1)
Mismatch R2	ミスマッチ率 (リード2)
Observed Diversity	リード1とリード2に関して異なるアライメント位置 (x, y) の数。アライメント位置が同じリードはPCR duplicationに相当すると考えられます。
Estimated Diversity	PCR duplicationを考慮したライブラリーの多様性の推定

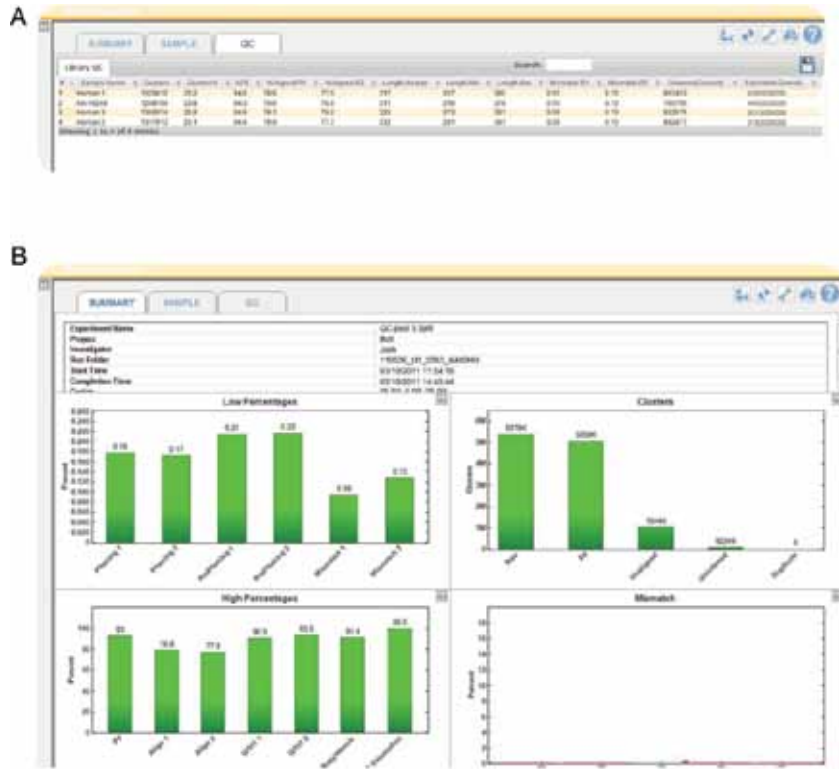
のに理想的です。すべての品質基準は、MiSeq Reporterが作成する自動レポートから簡単に入手およびエクスポート可能です。MiSeqから得られるデータはイルミナが提供するほかの次世代シーケンスプラットフォームに匹敵し、ある装置から別の装置へのスムーズな移行が保証されています。お客様ごとの実験要件に基づいて、シンプルなQCを実施して得られた基準をお客様のシーケンスプロジェクトを合理化および改善するために活用することも可能です。

詳細情報

革新的なデスクトップ型次世代シーケンサーMiSeqの詳細については、<http://www.illumina.com/product/system/miseq>をご参照ください。



図1：MiSeq Reporterが作成した自動ライブラリーQCレポート



A. Quality Control (QC) タブはサンプル名 (カラム1)、存在比 (カラム2~6)、断片長 (カラム7~9)、ミスマッチ率 (カラム10~12)、ライブラリーの多様性 (最後の2カラム) などといったシーケンスライブラリーの品質基準を表示しています。

B. SUMMARYタブはランの低い統計値 (左上のグラフ)、クラスタのパフォーマンス (右上)、ランの高い統計値 (左下)、リードとリファレンスシーケンス間のミスマッチ (右下) などといった高度なラン情報を表示しています。

イリミナ株式会社

〒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られます。

© 2013 Illumina, Inc. All rights reserved.

Illumina, illumina Dx, BaseSpace, BeadArray, BeadXpress, cBot, CSPPro, DASL, DesignStudio, Eco, GALiX, Genetic Energy, Genome Analyzer, GenomeStudio, GoldenGate, HiScan, HiSeq, Infinium, iSelect, MiSeq, Nextera, NuPCR, SeqMonitor, Solexa, TruSeq, TruSight, VeraCode, the pumpkin orange color, the Genetic Energy streaming bases design は Illumina, Inc. の商標または登録商標です。

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. 770-2011-J014 25JAN12

