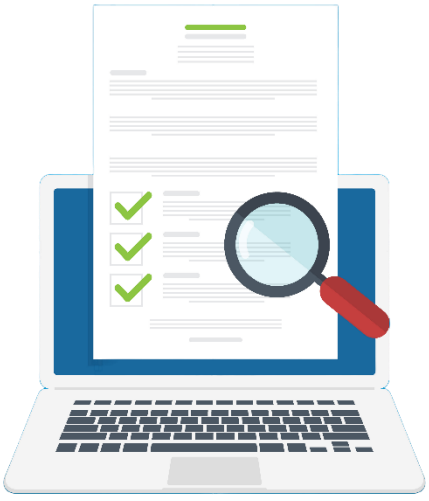
新世代のシーケンサーに対応した専用ツール

Third Generation Sequencing（TGS）は、NGSテクノロジーの制限をいくつか超えることで、関心が高まってきています。TGSの主な特徴は、ショートリードシーケンサーよりもはるかに長いリードを生成する能力です。このロングリードは、複雑なゲノムのアセンブリや選択的RNAスプライシングの研究を促進します。

しかし、TGSのロングリードは依然として高いシーケンスエラー率を示し、下流解析に悪影響を及ぼす可能性があります。この問題を克服するために、品質評価および制御ツール（LongQCなど）が、すべてのロングリードデータ解析に不可欠となっています。



▶次ページでロングリードデータのQCツールについてご紹介

biosupport@filgen.jp https://filgen.jp/ 052-624-4388 052-624-4389 フィルジェン株式会社

ロングリードデータの品質管理

ここ数年、ロングリードデータのクオリティチェックのためのツールが、いくつか提案されています。しかし、これらのツールは、異なるテクノロジー（主にPacBioとNanopore）によって生成されたデータ間の大きな違いや、配列された分子の性質（DNAまたはRNA）に互換があるか考慮する必要があります。

LongQCは、主要なTGSテクノロジーから得られるロングリードデータの品質管理を可能にするバイオインフォマティクスツールです。あらゆるデータセットの品質を効率的に評価するための有用な統計や図表を提供します。さらに、LongQC は解析に参照ゲノムを必要としないため、de novo ゲノムアセンブリを実施する前に特に有用なツールです。最後に、LongQC は PacBio と Nanopore の両方のデータに対応しており、ゲノム(DNA)とトランスクリプトーム(RNA)の両方のデータ解析が可能です。

LongQC を使用して ロングリードデータセットの品質を評価

その中で、著者らはcoverageモジュールをパイプラインの中核と位置づけています。このモジュールは、minimap2の修正版を用いてリード間のオーバーラップを計算することにより、カバレッジ統計とプロットを生成します。 coverageモジュールは、コンタミネーションやシーケンシングプロセスのアーチファクトとなりうるナンセンスリードを検出します。そのロジックでは、ナンセンスリードの割合から、シーケンスデータの品質を正確に推定することができます。

このパイプラインには、GC contentの分析、塩基あたりの品質値、リード配列長の分布など、より「古典的」な手法も含まれています。これらのほとんどは、NGSの品質管理ツールとして広く普及しているFastQCに含まれるいくつかのモジュールと非常に類似しています。

▶次ページで簡単にLongQCを実行できるツールを紹介

OmicBoxのLongQC

- LongQCは、FastQCやTrimomaticといった他の品質管理および前処理ユーティリティとともに、General Tools Moduleに含まれています。
- OmicsBoxでのLongQC分析の起動は簡単で、特定のデータ特性に合わせて分析の実行を迅速に調整することができます。
- OmicsBoxの実装により、複数のサンプルの同時解析が可能になりました。この機能により、同じ実験から得られた複数のファイルを解析した結果を比較することができます。

OmicsBox

実績は高いがコマンドライン型であったりOSに制限がある
オープンソフトウェアを多数搭載

それらの解析をマウス操作で簡単に解析できる

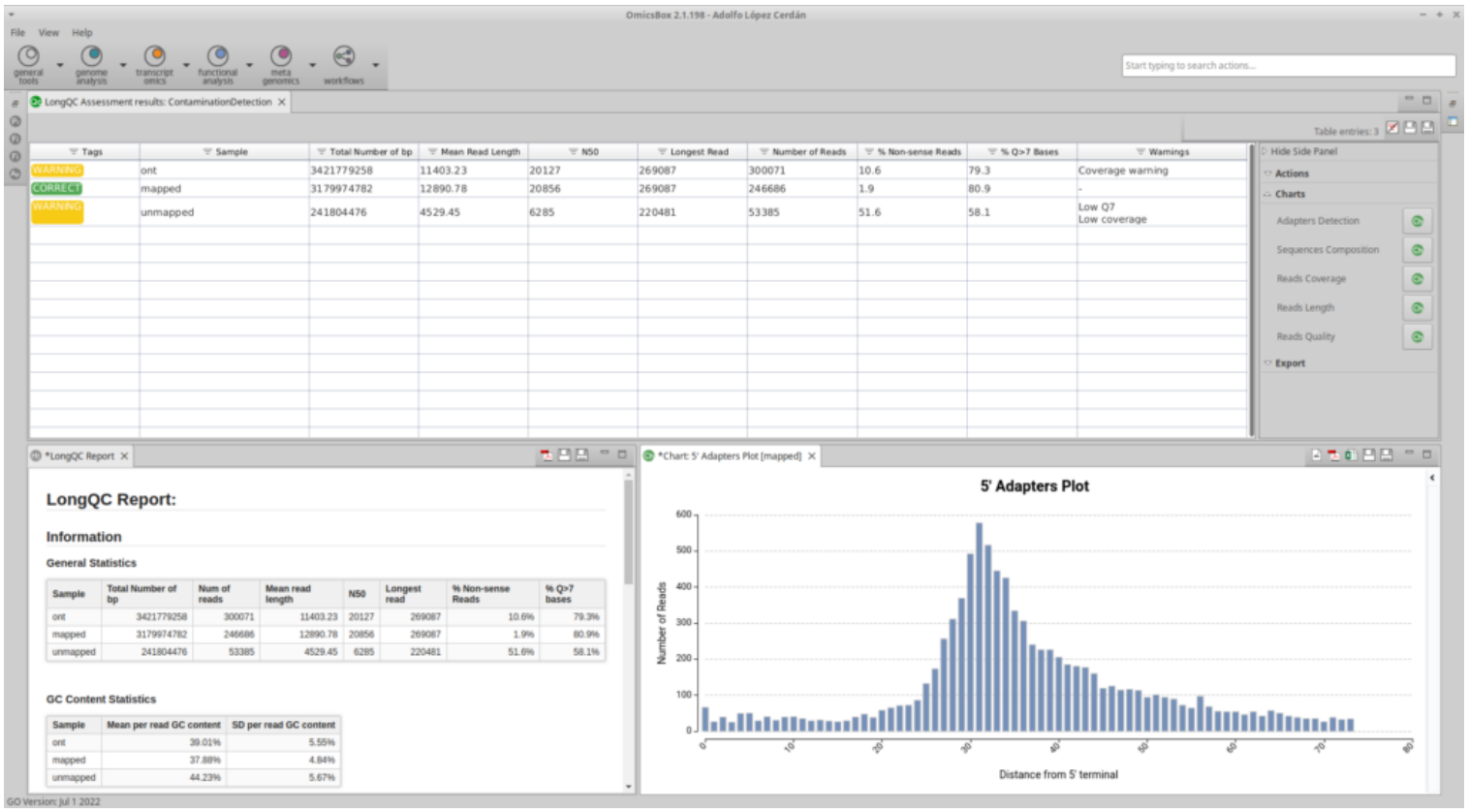


Figure 1. OmicsBox での LongQC 結果のオーバービュー

- LongQCを実行した結果、OmicsBoxはいくつかの有用な一般的統計情報を含む表と、より広範な品質指標のセットを含むHtmlレポートを表示します。

- 結果表から様々な記述統計量プロットを表示することができます。また、グラフに表示するサンプルを選択することができます。

- これらの情報をもとに、ユーザーはデータに対し前処理行うか、Flyeによるゲノムアセンブリや SQANTI3によるトランスクリプトーム前処理など、ロングリードデータ解析に進むことができます。

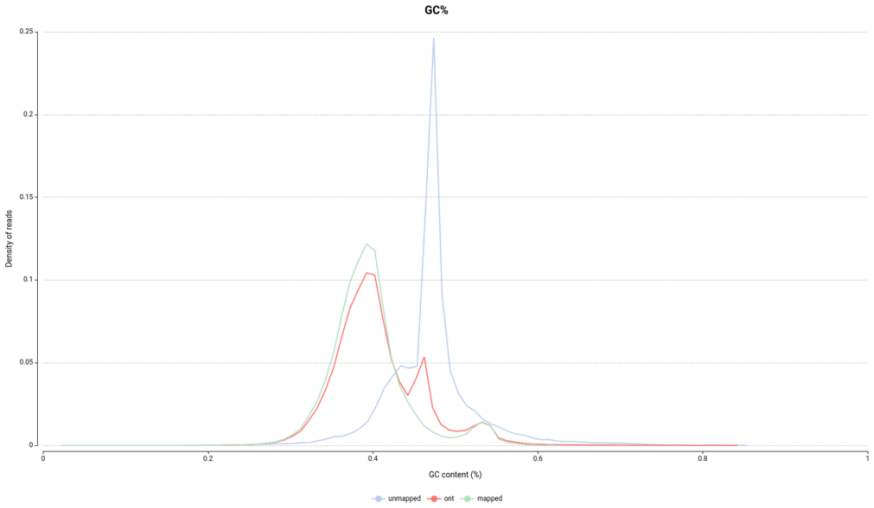


Figure 2.
複数のサンプルのGC含有量プロット。

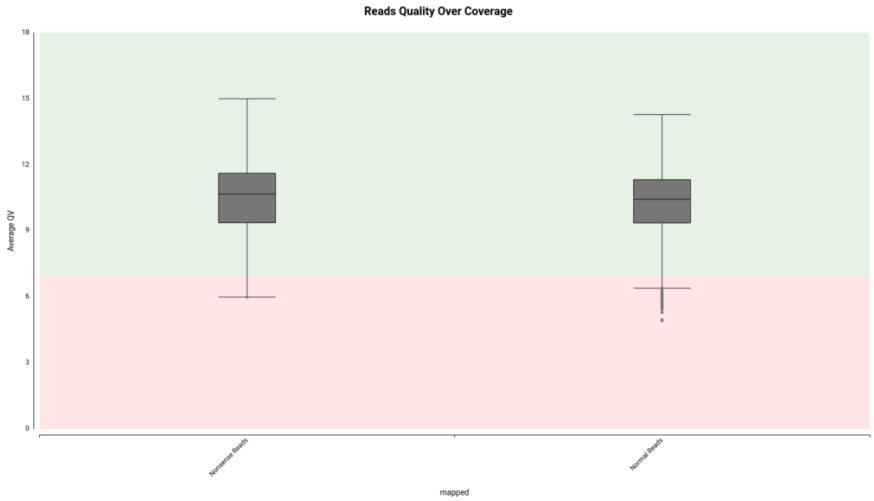


Figure 3.
カバレッジ別にグループ化された
Quality boxプロット

- 本解析の動画チュートリアル (YouTube)

- OmicsBoxの紹介ページはこちら