

Correlation Engine対応の RNA-Seqデータ処理

Correlation Engineで使用するデータを準備するための
RNAシーケンスパイプラインの概要

自動化されたマルチステップの RNA-Seqデータ処理

配列の取得、品質管理、アライメント
および発現量を単一パイプラインに
統合

DRAGEN™ RNAパイプラインを 用いた高速アライメント

ハードウェアアクセラレーションで高速
化された処理により、ハイスループット
RNA-Seqデータセットの高速かつ
正確なマッピングを提供

Correlation Engine対応の インポート用バイオセット

生物学的背景調査のための発現差解析
からの倍数変化、FPKM、q値などを
含む

はじめに

Correlation Engine対応のRNAシーケンス (RNA-Seq) パイプラインは、mRNAからの次世代シーケンサー (NGS) データを処理して転写産物の存在量を推測し、サンプル内で発現差のある転写産物を同定します。Correlation Engineは、研究、条件および生物学的背景にわたる遺伝子発現の比較解析を可能にし、研究者は大規模な公開および非公開データセットから分子パターンを探索し、仮説を生み出すことができます。

公開されている情報源から取得されたRNA-Seqデータは、Illumina Connected Analyticsワークフローに取り込まれ、DRAGEN解析ツールを使って処理されます。このテクニカルノートでは、RNA-Seqパイプラインワークフローの個々のステップの詳細を説明します (図1)。

生配列の抽出

RNA-SeqパイプラインはFASTQフォーマットの生のNGSデータのインプットに対応しています。Gene Expression Omnibus (GEO)¹ で公開されている多くのNGS研究は、Sequence Read Archive (SRA)² に保存された生のシーケンスデータとの直接のリンクを提供します。SRAからのシーケンスデータは一般的に圧縮が必要であり、使用可能なFASTQファイルを生成するには分割しなければならない場合があります。インプットFASTQファイルを非公開の情報源から取得した場合、配列をトリミングし、アダプター配列と低品質な末端を除去する必要があります。

RNA-Seqパイプラインは、GEO/SRAから取得するサンプル、サンプルのグループ分け、および下流の比較解析の実験デザインを識別する、処理用アノテーションファイルを受け付けます。最初のステップでは、これらのファイルが取得できることを確認し、次にそれぞれのSRAフォーマットのファイルをダウンロードして、アメリカ国立生物工学情報センター (NCBI: National Center for Biotechnology Information) から提供されている標準のSRA Toolkitを使ってFASTQ形式に変換します。

配列アライメント

インプットNGS配列は、Connected AnalyticsのDRAGEN RNA-Seq MAP Align Analysisパイプラインを使って、リファレンスゲノムに対してアライメントされます。DRAGEN RNAパイプラインは、DNAアライメントと同様のメソッドを用いてRNA-Seqリードからの短いシード配列をマッピングする、スプライスアライナーを使用します。次に、スプライスジャンクション (RNAスプライシングでエクソンが結合した境界) を最終リードアライメントに取り込みます。³ ワークフローの分岐点において、分解、リボヌクレオチド構成、インサートサイズ、マッピングされたリードの割合などのサンプル品質パラメーターを評価します。

このワークフローは、複数種類のインプットに対して効率的にデータ処理を行う標準化されたインターフェースを使用し、RNA-Seqデータの高速アライメントと定量を実施することでCorrelation Engineとの統合を実現します (図2)。本パイプラインは、ハードウェアアクセラレーションで高速化したアルゴリズムを用いてRNA-Seqリードの高速かつ正確なマッピングとアライメントを実施し、多くの一般的に使用されるソフトウェアツールよりも優れた性能を発揮します。例

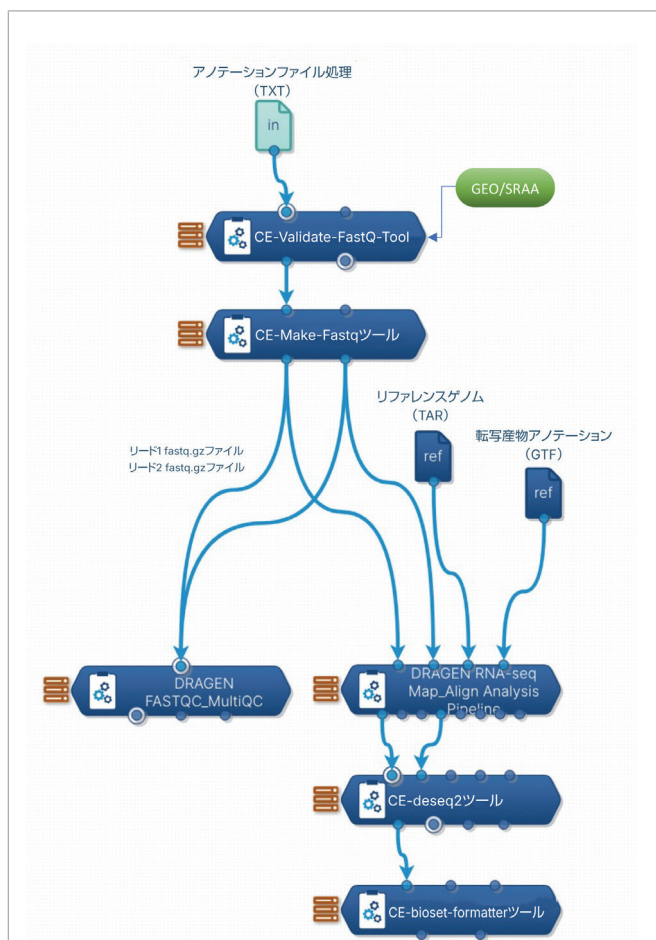


図1: Illumina Connected AnalyticsのRNA-Seqワークフロー

このワークフローは、生のFASTQファイルのインプットから、アライメント、定量および発現差を介し、Correlation Engine解析用のデータを準備するための拡張可能な標準化されたメソッドを提供します。

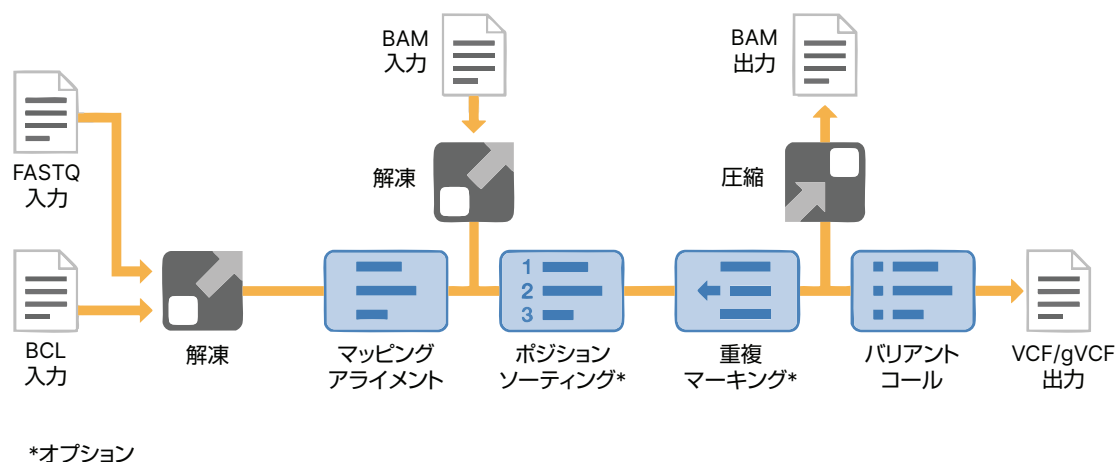


図2: DRAGEN RNAパイプラインを使ったデータ解析

このインターフェースは、RNA-Seqデータの迅速かつ正確なアライメントと定量が可能であり、Correlation Engineへの統合を実現します。

例えば、約3分で1億ペアエンドのRNA-Seqリードをアライメントできます。Correlation Engine対応の場合、カスタムパイプラインは、転写産物の定量に重点を置き、DESeq2を用いて下流の差異シグネチャーを生成します。具体的な内容については、[DRAGEN RNAパイプライン](#)ウェブサイトを参照してください。

転写産物存在量の推定

DRAGEN RNAパイプラインは遺伝子発現定量モジュールを搭載しており、RNA-Seqデータセットの各転写産物と遺伝子の発現を見積もります。初めに各リード（リードペア）のゲノムマッピングを、対応する転写産物マッピングにパイプライン内で変換します。次にExpectation-Maximization (EM) アルゴリズムを使って、観察されたすべてのリードに一番よく一致する転写産物発現値を割り出します。EMアルゴリズムはGCバイアスをモデル化し、レポートされた定量結果のGCバイアスを修正することもできます。

リードカウントをインプットとして使用し、RとDESeq2を使ってテスト群とコントロール群間の発現差解析を行います。塩基平均リードカウント、倍数変化、p値およびq値（Benjamini-Hochberg法で調整）はこの解析に由来します。群ごとにマッピングされた100万リードあたりの転写産物のキロベースあたりの断片平均値（FPKM）は、ノーマライズしたリードカウント、アライメントされたリード数および遺伝子全長に基づいて別々に計算します。

バイオセット生成

発現差解析の出力は単一のテキストファイルに統合され、以下の列に分類された処理用メタデータと発現データを含むバイオセットファイルに変換されます。

- 転写産物ID
- 倍数変化
- テスト発現（FPKM）
- コントロール発現（FPKM）
- p値
- q値

バイオセットファイルにフィルターを適用し、倍数変化値±1.2以上、q値0.05以下の遺伝子をパスさせます。最終のバイオセットファイルはCorrelation Engineに直接インポートされます。

まとめ

Correlation Engine用のRNA-Seqパイプラインは、DRAGEN RNAパイプラインを使用して高速アライメントと発現解析を実施し、下流の解析用に標準化した発現データを生成します。このワークフローでは公開および非公開のRNA-Seqデータセットの効率的な統合が可能です。

詳細はこちら →

[イルミナRNAシーケンス](#)

[DRAGEN RNAパイプライン](#)

[Correlation Engine](#)

参考文献

1. Edgar R, Domrachev M, Lash AE. [Gene Expression Omnibus: NCBI gene expression and hybridization array data repository](#). *Nuc Acids Res*. 2002;30(1):207-210. doi:10.1093/nar/30.1.207
2. Leinonen R, Sugawara H, Shumway M. [The Sequence Read Archive](#). *Nuc Acids Res*. 2011;39(Database issue):D19-D21. doi:10.1093/nar/gkq1019
3. Illumina. Illumina DRAGEN Bio-IT Platform Support. [support.illumina.com/sequencing/sequencing_software/dragen-bio-it-platform.html](#). Published 2025. Accessed May 7, 2025.

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/illumina

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

© 2025 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。

illumina[®]