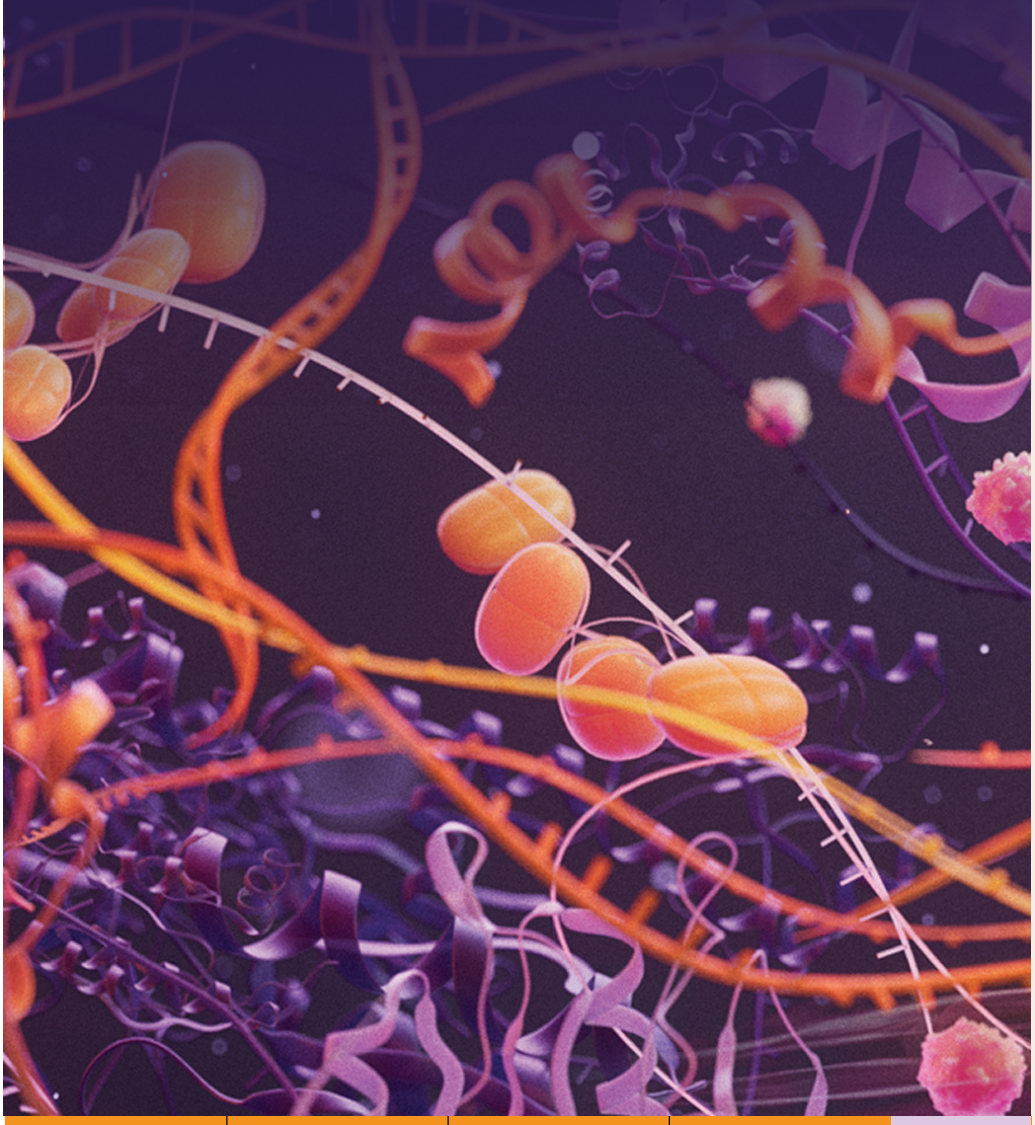
マルチオミクスのさらなる活用：マルチオミクスデータ解析入門

ゲノムおよびトランスクリプトームデータの統合事例を通じて、マルチオミクスデータ解析プロセスを学ぶ



illumina®

ゲノミクス | エピゲノミクス | トランスクリプトミクス | プロテオミクス

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。

M-AMR-01553 v2.0-JPN

目次

はじめに 3

シーケンスから生物学的洞察まで 4

マルチオミクスデータ解析ワークフロー 5

→ 品質評価 5
→ リードフィルタリング 6
→ ゲノムアライメント 6

ゲノムおよびトランスクリプトームデータの独立解析 7

→ ゲノムBAMファイルの解析：バリアントコール 7
→ 精度の最大化：すべての塩基から価値を引き出す 8
→ バリアントの精緻化とアノテーション 9
→ トランスクリプトームBAMファイル解析：定量 9
→ 遺伝子発現差解析 9

マルチオミクスデータの統合解析 10

まとめ 11

参考文献 12

はじめに

次世代シーケンス（NGS）テクノロジーは、ゲノム、エピゲノム、トランスクリプトーム、プロテオームなどのディープシーケンスを可能にし、生命の複雑さに対する私たちの理解に革新をもたらしました。[1] 生物学的システムをより包括的に捉えるため、研究者は、複数の分子層（オーム）にわたるデータを統合する、マルチオミクスアプローチの導入を進めています（図1）。[2-4] これらのアプローチは、遺伝子発現の測定値と空間情報を組み合わせることで、さらなる生物学的背景情報を提供する空間トランスクリプトミクスなどの新たな技術の登場に伴い、発展し続けています。公表文献や市場分析により、シーケンステクノロジーがより利用しやすく、かつコスト効率が高いものになるにつれて、マルチオミクスアプローチの導入が進んでいることが示されています。[5] イルミナのMiSeq™ i100システム、NextSeq™ 1000およびNextSeq 2000システム、NovaSeq™ 6000システムおよびNovaSeq Xシリーズは、幅広いアプリケーションに柔軟に対応し、コスト効率の高いマルチオミクスアプローチを可能にします。

複数のオームからシーケンスデータを生成した後は、正確かつ再現性の高い結果を提供する解析ワークフローが必要です。複数のオームにわたるデータを統合することにより、生命の最も基本的なプロセスに対する強力な洞察が生み出され、生物学的現象の包括的な理解が実現します。本eBookでは、マルチオミクスデータ解析ワークフローの一例として、ゲノムおよびトランスクリプトームシーケンスデータの統合について取り上げます。

プロテオミクス
タンパク質発現や特性評価を通じて、細胞機能および疾患生物学への理解を深めます。

トランスクリプトミクス
RNA転写産物の測定を通じて、遺伝子発現や組織内の空間的に分解されたパターンを含むトランスクリプトーム構成を評価します。

エピゲノミクス
遺伝子活性を制御するDNAメチル化およびクロマチン修飾を解析します。

ゲノミクス
次世代シーケンスおよびアレイベースの手法を用いてDNA配列バリエーションとゲノム構造を研究します。

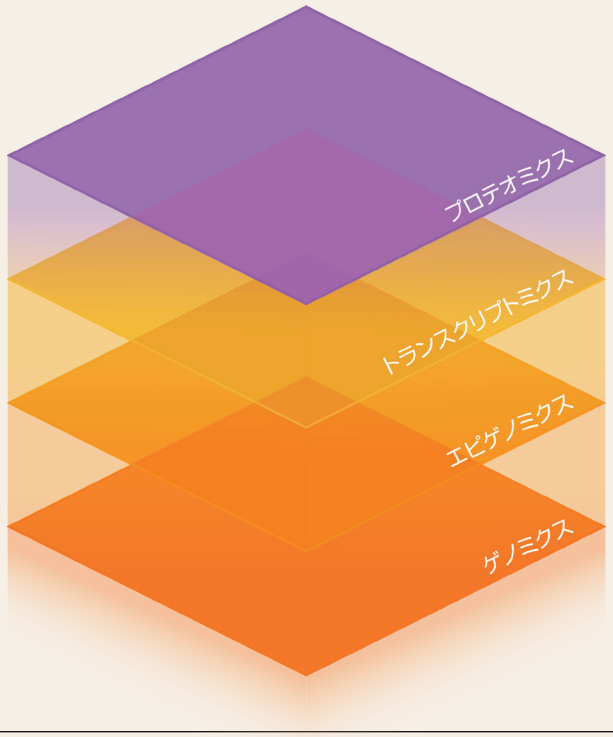


図1: マルチオミクスは、複数の層にわたる情報を繋げ、複雑なシステムの包括的な理解を実現します。

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。

シーケンスから生物学的洞察まで

マルチオミクスの研究デザインは多岐にわたりますが、ゲノミクスおよびトランスクリプトミクスの解析ワークフローの多くは、共通の解析フレームワークに基づいています。この一般的なフレームワークにおいても、データ解析は従来、複雑で手間がかかるものでした。技術の進歩により、これらのワークフローが効率化され、メンテナンスや日常的なパイプライン管理の負担が軽減されたことで、科学的な解釈やイノベーションに重点を置くことができるようになりました。[6-7] イルミナのクラウドベースおよび装置上の解析アプリケーションは、標準化されたワークフローを支援し、研究チーム間のコラボレーションを促進するように設計されています。

一般的なデータ解析ワークフローは、FASTQフォーマットのシーケンスデータから開始します（図2）。リードの評価、フィルタリング、リファレンスゲノムへのアライメントにより、Binary Alignment Map（BAM）ファイルが生成されます。その後の下流解析は、データの種類によって分岐します。ゲノムBAMファイルは、バリアントコールを通じて処理され、Variant Call Format（VCF）ファイルが生成されます。一方トランスクリプトームBAMファイルは、定量および正規化され、遺伝子発現量が測定されます。最終的に、ゲノムおよびトランスクリプトームの解析結果を統合することで、視覚化、解釈、生物学的発見に役立つようになります。本eBookでは、データ解析ワークフローの各段階についてさらに詳しく解説します。

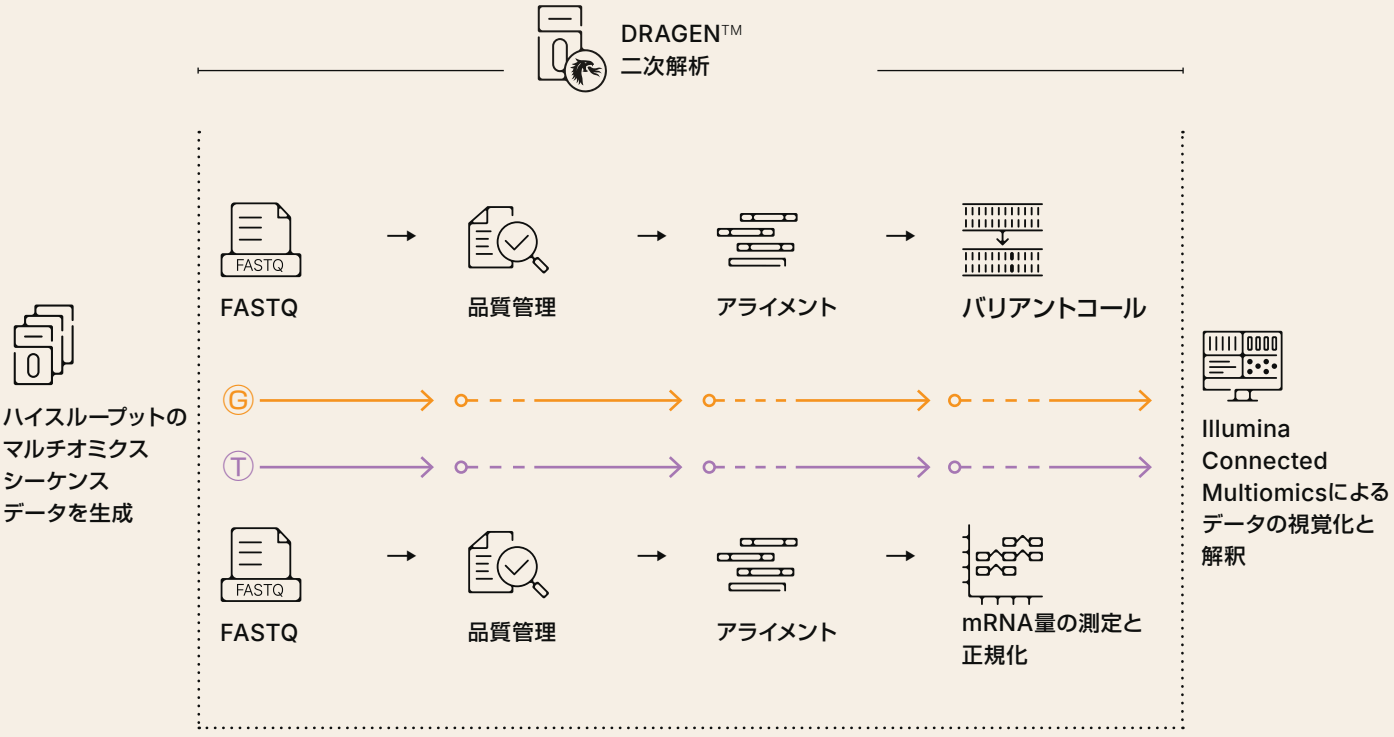


図2：ゲノムとトランスクリプトームのデータのためのマルチオミクス解析ワークフロー

ゲノムとトランスクリプトームのデータを統合するマルチオミクスワークフローには、生のシーケンスデータから生物学的洞察を得るまでのプロセスを支援する、複数の解析ステップが含まれています。DRAGEN二次解析は、シーケンスの品質評価、リードアライメント、高感度かつ高精度なバリアントコール、および正規化された発現量を出力するRNA定量を実施し、ゲノムとトランスクリプトームの統合解釈を可能にします。Illumina Connected Multiomicsにより、これらのマルチオミクスデータセットの探索と視覚化が実現します。

マルチオミクスデータ解析ワークフロー

DRAGENソフトウェアによるFASTQからBAMへの変換

品質評価

イルミナシーケンスシステムは、ラン終了後、バイナリーベースコール（BCL）ファイルを生成します。BCLファイルは通常、コアファシリティやサービスプロバイダーによってFASTQファイルに変換されます。FASTQファイルには、塩基配列のリストとクオリティスコアなどのメトリクスが含まれます。ほとんどの研究チームでは、このテキストベースのFASTQファイルを解析の標準的な出発点として使用しています。

あらゆるNGS実験において、FASTQファイル内のリードの品質評価は、データ解析の重要な第一段階です。品質評価には、平均リード長、平均塩基品質、グアニン/シトシン含有量の評価が含まれます。これらのメトリクスを評価することで、下流解析に影響を及ぼす可能性のある低品質リードを同定しやすくなります。イルミナのDRAGEN二次解析により、タッチパネル操作でオプションを選択し、わかりやすいグラフを生成することで、品質評価を効率化できます（図3および図4）。

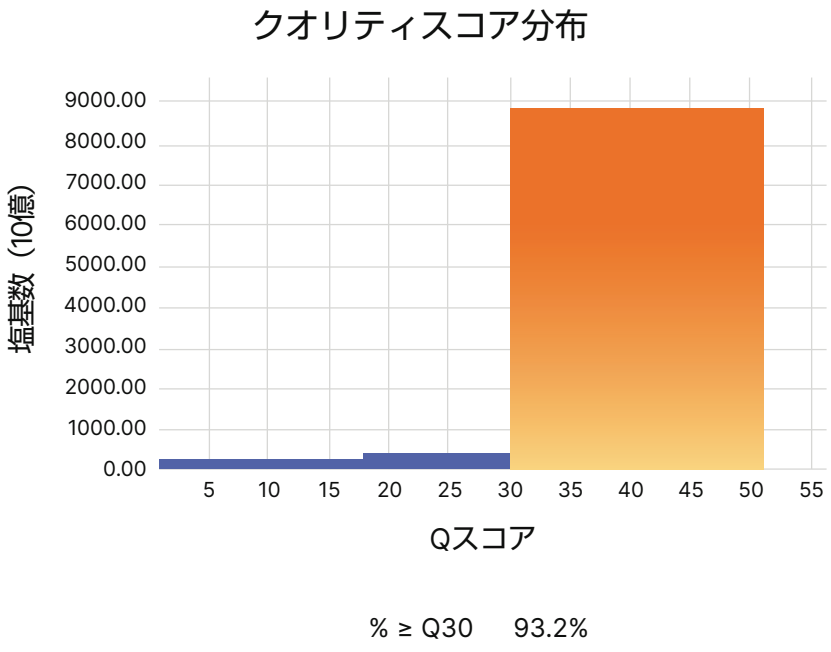


図3：シーケンスデータの品質を示したグラフの一例。

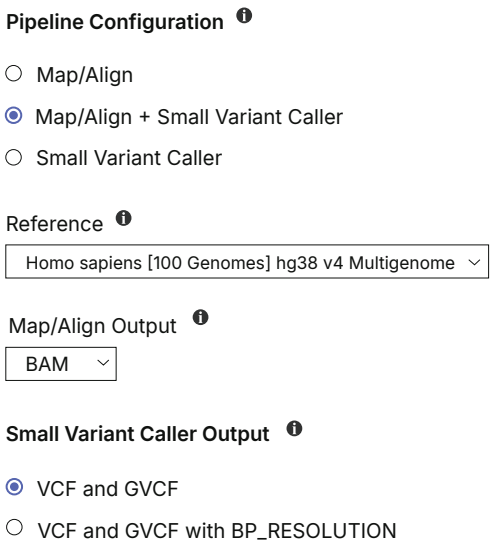
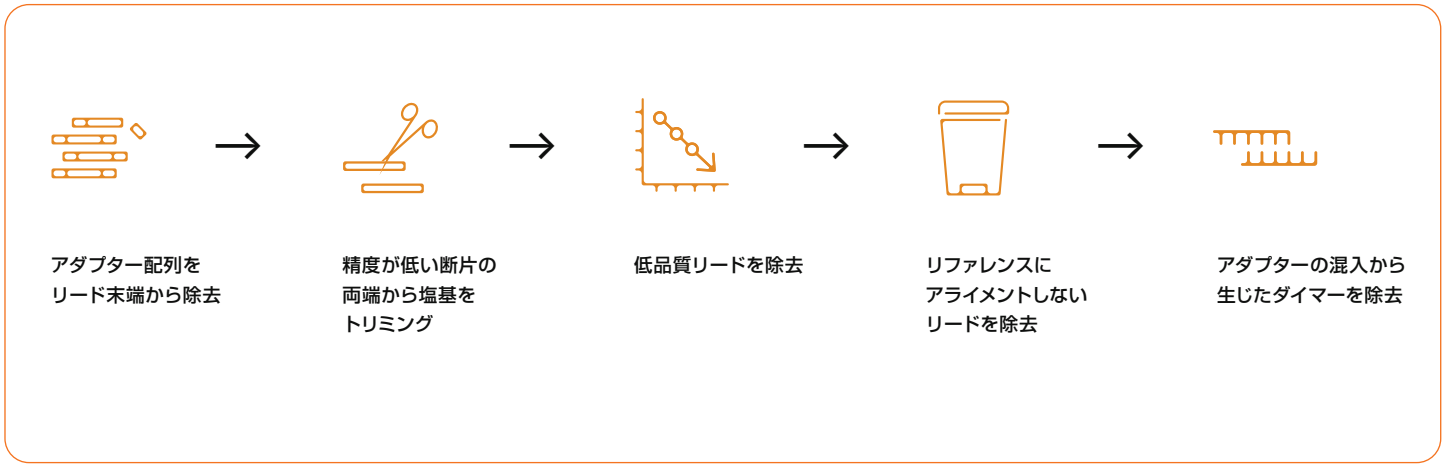


図4：DRAGEN二次解析を設定するための、使いやすいタッチパネルインターフェースの一例。利用可能なインターフェースと設定オプションは、ソフトウェアのバージョンによって異なる場合があります。

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。

リードフィルタリング

FASTQファイル内のシーケンスリードの品質を評価した後、解析に不要な配列や低品質な配列を除外します。これには以下の工程が含まれます。



ゲノムアライメント

低品質リードがフィルタリングされた後で、シーケンスされた断片がリファレンスゲノムにアライメントされます。アライメント精度は、シーケンス品質と使用するリファレンスゲノムの両方に依存し、特に遺伝的多様性が高い領域や複雑な構造変異を有する領域でその影響が顕著になります。アライメントされたリードはBAMファイルに保存され、ゲノムおよびトランスクリプトームの下流解析に使用されます。

多くの利用可能な公開リファレンスゲノムは、限られたサンプルグループのため、祖先集団や世界集団のバリエーションを十分に捉えたものではないことに注意が必要です。[8] さらに、多くのプログラムでは、ゲノムのマッピングおよびアライメントに単一のリファレンス（別名：線形リファレンス）を使用しますが、このアプローチでは、ゲノムの多型性が高い領域や、挿入、逆位、転座などの構造変異を解明することが難しい場合があります。[9]

マルチゲノムマッピングによるアライメントの向上

複雑で多様性が高い領域において、より高い解像度でより堅牢な解析を実現する1つの方法として、マルチゲノムマッピングの使用が挙げられます。[9] 1つのリファレンスゲノムを使用する従来のリードアライメント法とは異なり、マルチゲノムマッピングは、多くのゲノム（パンゲノム）をリファレンスソースとして使用できます。

以下の理由により、さまざまなゲノミクスアプリケーションでマルチゲノムマッピングの導入が進んでいます。

- 遺伝的多様性をより適切に表現し、祖先に関連するリファレンスバイアスを低減できる。[10]
- 線形リファレンスを用いたアプローチと比較して、特定の1塩基変異（SNV）の検出を改善できる。[9]

マルチゲノムマッピングは、科学文献での報告が増加しており、DRAGEN二次解析でもすぐに利用可能です。[11]

ゲノムおよびトランスクリプトームデータの独立解析

ゲノムBAMファイルの解析：バリアントコール

BAMファイルが生成された後で、DNAシーケンスデータは、バリアントコールと呼ばれるプロセスを経てVCFファイルに変換されます。[12] バリアントコールにより、得られた配列を1つまたは複数のリファレンスゲノムと比較することで、変異、欠失および挿入を同定できるようになります。バリアント検出は、シーケンス深度およびアライメント品質の影響を受けます（図5）。

バリアントコールにより、生殖細胞系列変異と体細胞変異の両方を同定できます。生殖細胞系列変異は、個体のすべての細胞に存在します。このため、生殖細胞系列のバリアントコールでは、アリル頻度は通常0、0.5または1付近を示します。これに対し体細胞変異は、生殖細胞系列以外の細胞で生じ、生物学的要因やサンプル不均一性のため、より低いアリル頻度で生じる場合があります。このため、低頻度バリアントの検出に対応できるように、バリアント検出閾値を調整する場合があります。DRAGEN二次解析には、バリアントコール精度の向上を目的として設計された、機械学習ベースのアプローチが取り入れられています。

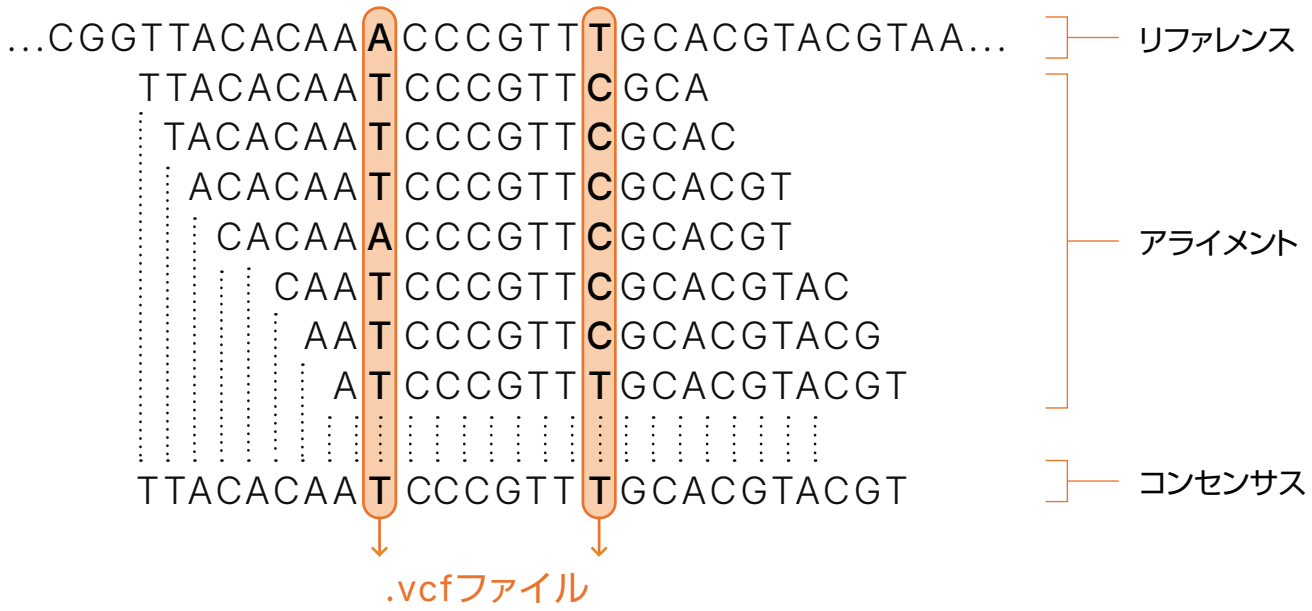


図5：バリアントコールからコンセンサス配列を取得するプロセスの一例。

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。

精度の最大化：すべての塩基から価値を引き出す

FASTQファイルの品質評価により、データ解析パイプラインへの高品質な入力が担保できます。これと同様に、その後の解析を通じて品質を維持することが重要です。シーケンスデータ処理に用いられるプログラムは、下流解析の精度に大きな影響を及ぼす場合があります。解析プログラムの信頼性が低い場合、特にシーケンスが困難なゲノム領域においては、偽陰性や偽陽性がバリアントコールや発現レベルに影響を及ぼす場合があります。

DRAGEN二次解析（v4.5）は、評価したベンチマークにおいて、検証した比較対象のワークフローと比べてバリアントコールエラーが少ないことを示しています（図6）。[13]

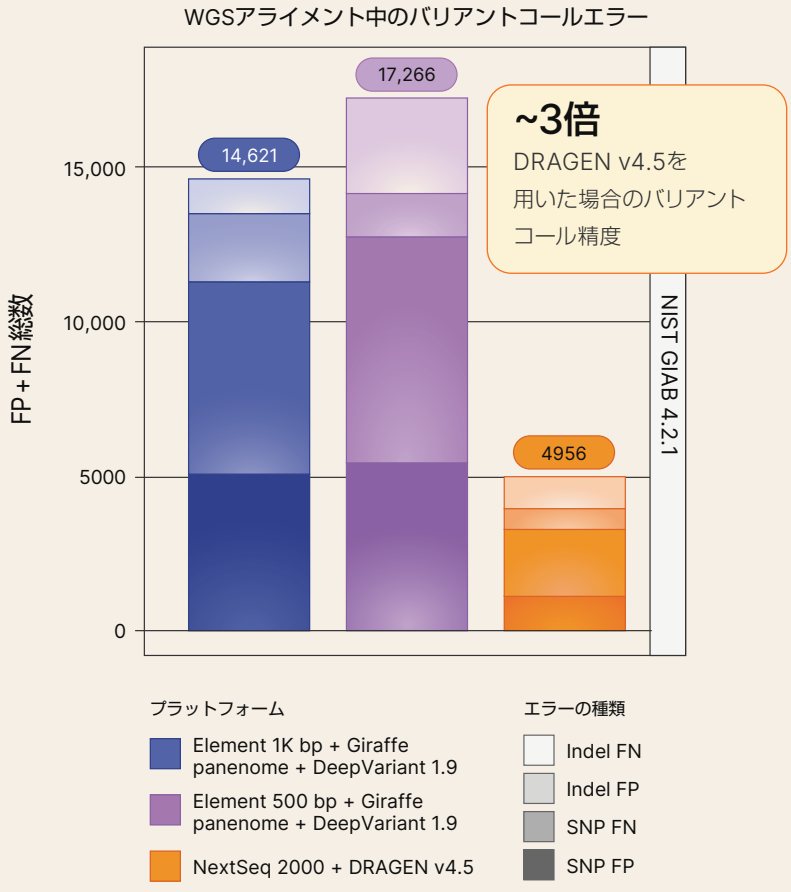


図6：DRAGEN二次解析（v4.5）は、米国国立標準技術研究所（NIST）のGenome in a Bottle（GIAB） v4. 2.1ベンチマークに基づきElement AVITI + DeepVariantと比較して、XLEAP-SBS™ ケミストリーを採用したNextSeq 2000システムで約3倍高い精度でバリアントコールを実現します。[13]

結果は、評価した条件下でのNIST GIAB v4.2.1ベンチマークにおける性能を反映したものであり、すべてのワークフロー、検体、アプリケーションでの性能を予測するものではありません。

FN：偽陰性、FP：偽陽性、Indel：挿入/欠失、SNP：1塩基多型、WGS：全ゲノムシーケンス。



バリアントの精緻化とアノテーション

バリアントコール後は、技術的アーティファクトやバックグランドノイズを低減するために、バリアントフィルタリングを実施します。残ったバリアントには、遺伝子と転写産物との関連性、予測される機能的影響、集団内頻度、およびキュレーションされたバリアントデータベースへのリンクなど、生物学的情報がアノテーションされます。これらのアノテーションは、大量のバリアントデータセットの整理や、研究対象の生物学的課題に関連する可能性のあるバリアントへの優先順位付けに役立ちます。

クオリティフィルタリングと生物学的アノテーションを組み合わせてバリアントを精緻化すると、特定の遺伝的変化と観察された生物学的特性との関連性が解釈しやすくなります。構造バリアントデータは、サンプル間で比較したり、下流のマルチオミクス解析で他の分子データセットと統合したりすることができます。

トランスクリプトームBAMファイルの解析：定量

RNAシーケンスデータをBAMファイルにアライメントした後は、リード数の定量が必要です。[14] トランスクリプトームをシーケンスする場合、一般的に特定の転写産物に対するリード数が多いほど、遺伝子発現レベルは高くなります。ただし、総リード数の定量後は、遺伝子長、RNA組成、サンプル間のリード深度の違いなどの要因を考慮するため、リード数を正規化することが重要です。

転写産物を定量し、発現レベルを正規化すると、複数の統計学的解析手法を用いて、遺伝子発現における重要なパターンを発見できるようになります。

遺伝子発現差解析

下流解析の最も一般的なアプローチの1つとして、統計学的に有意な差の同定を目的とした、条件間での遺伝子発現パターンの比較が挙げられます。サンプル間および実験条件間で遺伝子発現を正確に比較できるように、トランスクリプトームデータを処理します。発現測定値のフィルタリングおよび正規化により、技術的なばらつきが低減されるため、発現差には、シーケンスまたはサンプル調製による影響ではなく、基盤となる生物学的シグナルが確実に反映されます。これらの前処理により、一貫性のある下流解析の基盤を確立できます。

発現データの準備が完了すると、全体的な転写パターンの調査や、発現レベルが有意に変化した遺伝子の同定が可能になります。発現差解析では、生物学的応答や状態に関連する遺伝子を明らかにすることで、これらの転写変化をもたらすパスウェイやプロセスについての洞察が得られます。

これらの解析を組み合わせることで、転写活性を定量的に捉え、トランスクリプトームデータの生物学的解釈に役立てられるようになります。

マルチオミクスデータの統合解析

マルチオミクス解析による洞察は、ゲノムおよびトランスクリプトームのデータセットを同時に解析し、遺伝的変異が遺伝子発現に与える影響を検討することで得られます。ゲノム変異を対応する発現変化に関連付けることで、遺伝型と転写活性との関係性を探索し、それらの潜在的な生物学的意義を評価できます。これらの相補的なデータタイプを組み合わせることで、単独のデータセットを用いる場合よりも、分子機能をより包括的に捉えることができます。

データの探索と解釈

Illumina Connected Multiomicsは、モダリティ間でのマルチオミクスデータの解析、視覚化、統合のための共有環境を研究チームに提供します。[15] 厳密な統計手法とAI解析ツールが、複雑なマルチオミクスデータセットの解釈を支援します（図7）。研究者は遺伝子発現パターン、変異との関連性、パスウェイ活性、表現型との関連性、その他のオミクス間の関係性を評価できます。

イルミナのソフトウェアおよびインフォマティクス製品群では、幅広い研究アプリケーションにおける解析、解釈、データ集計に対応します。Correlation Engineは、Illumina Connected Multiomicsに統合されており、三次解析の一環として、解析結果に対する生物学的背景を提供します。研究者は、トランスクリプトーム解析から得られた遺伝子レベルの結果とキュレーションされた公開データセットとを比較し、一致するシグネチャー、パスウェイ、制御関係、薬剤や表現型との関連性を同定できます。Correlation Engineは、27,000件を超える科学研究をマイニングし、遺伝子、実験、薬剤、表現型に関するデータに基づく洞察を提供することで、より広範な科学的エビデンスにおける個々の実験結果の位置づけに役立ちます。

統合解析と知識に基づく解釈を組み合わせることで、ゲノムとトランスクリプトームのデータセットを、複雑な生物学的システムに対するより包括的な理解へと変換できるようになります。

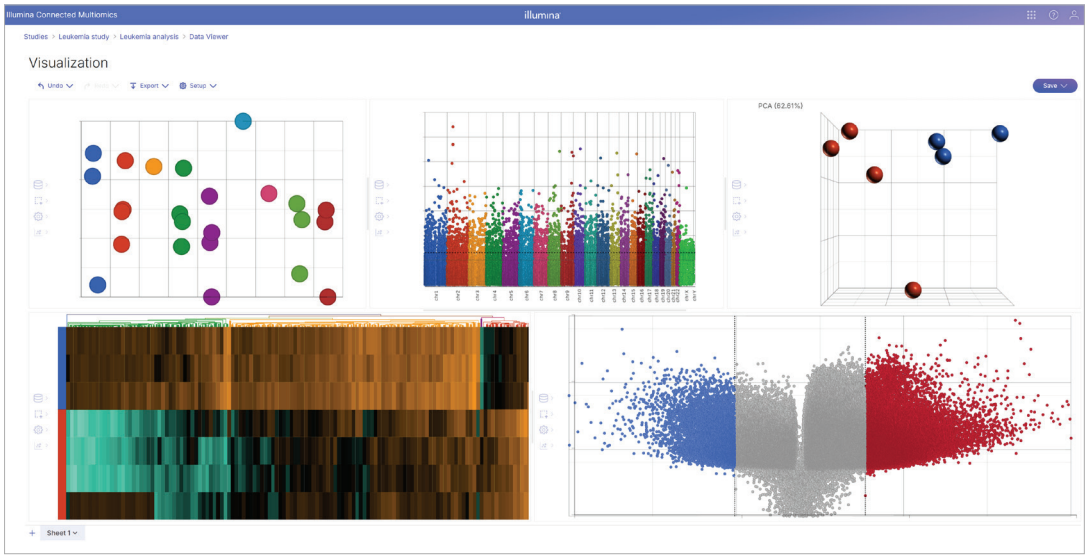


図7：Illumina Connected Multiomicsの視覚化出力例。ドットプロット、マンハッタンプロット、主成分分析（PCA）プロット、ヒートマップ、ボルケーノプロットなどがあり、マルチオミクスデータセット間の発現パターン、関連シグナル、サンプル構造、差次的特徴の評価を支援します。

「**現在では、マルチオミクスはさらに利用しやすくなっています。**要するに、誰もが活用できる時代になったということです。ワークフローは5年前でさえ不可能だった方法で効率化されています。ワークフローが効率化され、テクノロジーが統合されてシームレスに連携するようになれば、研究上の問いを追求できるようになります。実現可能だと分かれば、以前ほど難しく感じることもありません。」



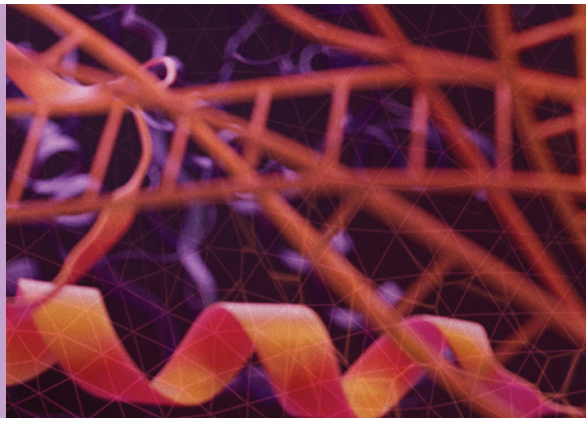
Bodour Salhia, PhD
Professor of Cancer Biology
Royce and Mary Trotter Chair in Cancer Research, Keck School of Medicine of USC

まとめ

マルチオミクスは、複数の分子層にわたる情報を統合することで、複雑な生物学的システムのより包括的な理解を実現します。マルチオミクスの研究デザインは多岐にわたりますが、いずれも複雑で多層的なデータセットを正確かつ一貫して解析できる、解析フレームワークの恩恵を受けられます。イルミナのアプリケーションベースのソリューションは、データ解析を効率化し、確信を持った解釈を支援することで、シーケンスデータから生物学的洞察へと効率的に進められるよう支援します。その結果、複雑なマルチオミクスデータから新たな洞察や発見を導き出すまでのプロセスが、これまでになくシンプルになります。

今日からマルチオミクス データ解析を始めましょう

イルミナマルチオミクスの詳細はこちら→



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。

参考文献

1. Ohashi H, Hasegawa M, Wakimoto K, et al. Next-generation technologies for multiomics approaches including interactome sequencing. *Biomed Res Int*. 2015;2015:104209. doi:10.1155/2015/104209
2. Dar MA, Arafah A, Bhat KA, et al. Multiomics technologies: role in disease biomarker discoveries and therapeutics. *Brief Funct Genomics*. 2023;22(2):76-96. doi:10.1093/bfgp/elac017
3. Flynn E, Almonte-Loya A, Fragiadakis GK. Single-cell multiomics. *Annu Rev Biomed Data Sci*. 2023;6:313-337. doi:10.1146/annurevbiodatasci-020422-050645
4. Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. *Genome Biol*. 2017;18(1):83. doi:10.1186/s13059-017-1215-1
5. Digital Science. Dimensions [database]. app.dimensions.ai. 2015–.2025 Accessed December 4, 2025, under license agreement.
6. Illumina. No limits in the cloud: Insights proliferate at Eurofins Viracor. illumina.com/company/news-center/feature-articles/no-limits-in-the-cloud--insights-proliferate-at-eurofins-viracor0.html. Published May 1, 2023. Accessed June 15, 2026.
7. Illumina. Efficient cloud data analysis for COPD multiomics project. illumina.com/science/customer-stories/icommunity-customer-interviews-case-studies/okayama-cloud-data-analysis-copd.html. Published 2023. Accessed June 15, 2026.
8. Fatumo S, Chikowore T, Choudhury A, Ayub M, Martin AR, Kuchenbaecker K. A roadmap to increase diversity in genomic studies. *Nat Med*. 2022;28(2):243-250. doi:10.1038/s41591-021-01672-4
9. Miga KH, Wang T. The need for a human pangenome reference sequence. *Annu Rev Genomics Hum Genet*. 2021;22:81-102. doi:10.1146/annurev-genom-120120-081921
10. Wong KHY, Ma W, Wei C-Y, et al. Towards a reference genome that captures global genetic diversity. *Nat Commun*. 2020;11(1):5482. doi:10.1038/s41467-020-19311-w
11. Illumina. The quest for accuracy gains in the dark regions of the genomes: Presenting the DRAGEN multigenome mapper and pangenome reference updates in version 4.3. illumina.com/science/genomics-research/articles/second-gen-multigenome-mapping.html. Published August 12, 2024. Accessed June 15, 2026.
12. Zverinova S, Guryev V. Variant calling: Considerations, practices, and developments. *Hum Mutat*. 2022;43(8):976-985. doi:10.1002/humu.24311
13. Illumina, Inc. Data on file, 2026.
14. Koch CM, Chiu SF, Akbarpour M, et al. A beginner's guide to analysis of RNA sequencing data. *Am J Respir Cell Mol Biol*. 2018;59(2):145-157. doi:10.1165/rcmb.2017-0430TR
15. Illumina. Illumina Connected Multiomics. illumina.com/products/by-type/informatics-products/connected-multiomics.html. Published 2026. Accessed June 15, 2026.

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

© 2026 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. M-AMR-01553 v2.0-JPN 27OCT2026