

CLC Genomics Workbench CLC Genomics Server

様々なNGSアプリケーション解析に対応

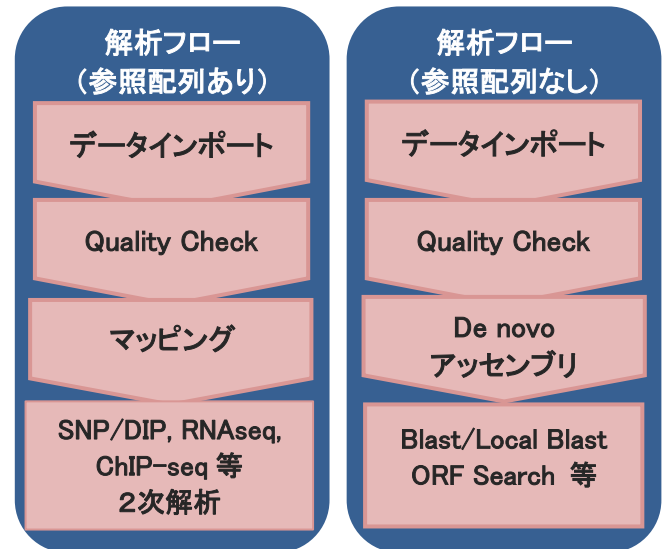
キアゲン社の CLC GenomicsWorkbench および提供されるプラグインは、次世代シーケンサーを利用した様々なアプリケーションの解析に対応しています。

解析は流れに沿った操作を画面から実施でき、各工程の結果の参照、サンプル間の比較、さらにはデータベースとの比較など、わかりやすいビューワーで確認することが可能です。従来法の複雑なコマンドによる解析や、複雑なパラメータの調整、さらには煩雑になる解析結果の比較など不要です。

毎年新規機能が追加され、時代に沿ったアプリケーションに対応したツールも開発されています。近年では特にクリニカルシーケンスに対応したツールの改良や、微生物解析に対応した新規ツールも提供されてきています。

ソフトウェアとして、デスクトップ型製品“CLC Genomics Workbench”、大量の解析処理に対応するサーバー/クライアント型製品“CLC Genomics Server”、アセンブリ専用コマンドラインツール“CLC Assembly Cell”は多くのユーザーに利用されています。さらに、メディカルリサーチ向けとして変異解析、RNAseq 解析に特化した姉妹品の Biomedical Genomics Workbench も準備されています。

ワールドフュージョンではインフォマティクスのスペシャリストとしてソフトウェアとともに、実際のプロジェクトに合った総合的なシステム構築のサポートおよびコンサルティングも実施しています。ハードウェアのインテグレーション相談など可能です。

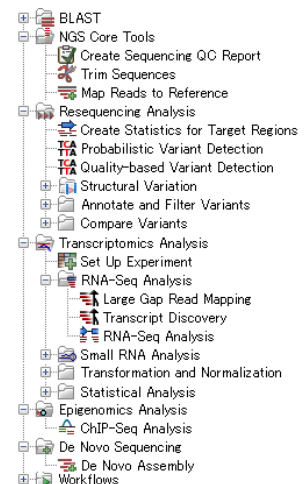


Gx CLC Genomics workbench

Genomics Workbench は初歩的な遺伝子解析ツールから次世代シーケンスデータ解析ツールまでを網羅し、多くの研究者に利用されている解析ソフトウェアです。DNA、RNA、およびタンパク質解析のための包括的なツールが多数搭載されています。

【次世代シーケンスデータ解析機能】

- *de-novo* アセンブルおよびマッピングツール搭載
- リファレンス配列・アノテーション情報・SRA 等の公共 NGS データはソフトウェア上でNCBIからそのままダウンロードしてインポート可能
- 独自の技術によりメモリの少ないPCでのアセンブル、マッピング作業可能
- イルミナ社、サーモフィッシャー社、PacBio 社等の高速シーケンスデータフォーマットに対応
- 従来型シーケンスデータ、FASTA、テキスト フォーマットに対応
- マッピング済みデータ(BAM, SAM)フォーマット取り込み可能
- 各種シーケンサーデータをまとめて解析可能 (ハイブリッドアセンブル対応)
- GFF/GTF 形式でユーザー独自のアノテーションが参照配列に追加可能、dbSNP 登録変異情報の参照やマッピング領域の指定(ex:エキソン領域・アンプリコン標的遺伝子等)が可能
- アライメントやマッピングビューワーのイメージは保存し利用可能
- ワークフローを作成することで複数の工程を必要とする解析を手軽に連続して実行可能
- リードのフィルタリング、トリミング、QC 機能有
- RNA-seq 解析結果は正規化・群間統計解析まで出力、GO annotationを加えることで Fisher 検定による GO 解析も可能
- ORF サーチや、Blast機能付属

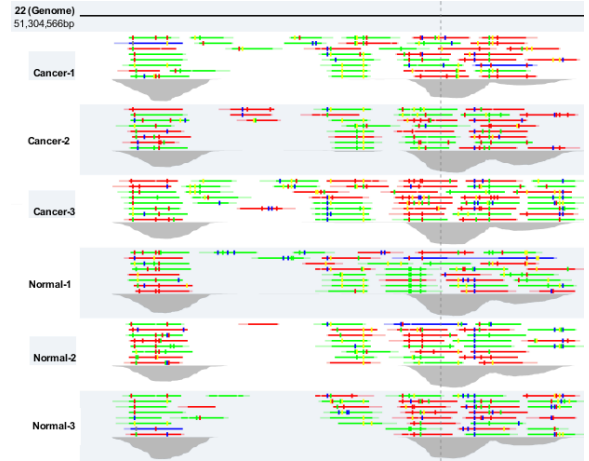


【対応解析アプリケーション】

- *de-novo* アセンブル
- リファレンスマッピング
- RNA-Seq 解析
- ChIP-seq/Histseq/Bisulfite seq
- SNP/DIP 検出
- ゲノム変異・構造変化

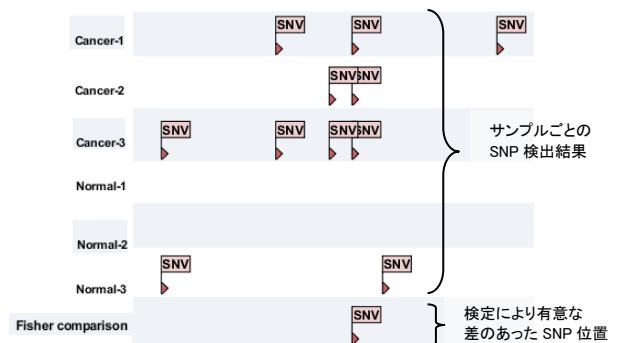
【変異解析機能】

- Probabilistic variant detection: クオリティベースの変異のツールに対し、確率モデルをベースとしたツール。ゲノムの倍数性は1倍体から4倍体まで対応
- Non-specific なマッチや、壊れたペアに対するフィルタリング機能
- Forward/Reverse リードのバランスに対するポストフィルタリング機能
- SNV/MNV/Deletion/Insertion 検出
- Trackでの解析: 興味の対象となる領域は、別途 Track として定義し対象ごとに別々のTrackを作成することが可能
- 変異のためのアノテーションとフィルタリングのツール
- サンプル間での変異の比較
- Compare variants within group: サンプル間で共通の変異の発見
- Fisher exact test: ケースとコントロールのグループ間で統計的に異なる変異の検出
- Trio analysis: 親子間での変異を比較し、遺伝によるものか、新規の変異によるものかを比較
- Filter against control reads: コントロール変異を取り除く
- GO enrichment analysis: 変異の存在した複数の遺伝子について、共通する機能面での役割を検証
- Amino acid changes: アミノ酸置換を引き起こさない変異 (synonymous) と アミノ酸置換を引き起こす変異 (non-synonymous) の検出
- Annotate with conservation scores: インポートされた Conservation スコアの情報から、変異へのアノテーション付け
- Predict splice site effect: 変異が intron-exon の境界に位置しているかどうかの探索
- Link Variants to 3D Protein: アミノ酸変異をタンパク質の3D構造上で可視化し、PDB database から取得した Drug interaction 情報とともに表示



複数サンプルのマッピング結果や変異検出結果を並べて表示可能。

SNP については Fisher's test に基づく 2 群間の出現頻度の差の検定や、親子関係情報を使用した Trio 解析が可能で GWB の検定機能により群間で有意な差の認められた SNP のフィルタもTrack上で可能です。



【トラック機能】

トラック (Track) は個々のサンプル解析結果や、アノテーションデータを視覚的に比較しやすくする為の機能で、リシーケンスに関するすべてのツール(例: read mapping, variant detection など)にはデータを Track として保存するオプションがあります。

【利用可能なデータベース】

ソフトウェアからは下記のデータベースをダウンロードして参照配列や既知変異情報として使用することができます。

Ensembl Reference sequence/ Genome annotations、Ensembl Variant (dbSNP)、HapMap、1000Genomes、ClinVar、NCBI SRA 上に登録されたシーケンスデータファイルも GWB からアクセスして Fastq ファイルとメタデータを取得できます。

【トランスクリプトーム解析】

- RNA-Seq 解析: 遺伝子発現値 (total read count/RPKM) の計算と遺伝子発現プロファイルの作成
- Small RNA 解析: アダプター除去、タグのカウント、miRBase やその他リソースを使用したアノテーション付け
- 単一サンプルから単一サンプル比較の割合差を調べる Kal らの方法を用いた統計検定
- レプリケートを含む 2 グループの割合差を調べる Baggerley らの方法を用いた統計検定
- サンプルの属性情報をエクセル形式で取り込み、Edge-R による群間比較に利用することが可能、さらにさまざまなクラスタリング解析上に属性情報を表示
- 既知のエクソンとのマッチングを比較、mRNA リードから染色体やゲノム全体のマッピングを通し新規の転写産物 / エクソンの発見
- リードがユニークに一致した遺伝子やエクソン上、またはエクソン間にまたがるリードの統計

【対応データフォーマット例】

- Illumina: fastq, qseq
 - Ion Torrent: fastq, .sff
 - Pacbio: fastq, bash5
 - Fasta, abi, ab, ab1
 - SAM/BAM (Sequence Alignment/Mapping)
- NGS データは、Paired-end / Mate-pair 対応、gzip 圧縮フォーマット一部対応、MultiPlex 対応

ハードウェア仕様(例)

最小解析モデル

パーソナル解析向き、RNA-seq の解析可能

OS: Windows10 Pro 64bit
CPU: Core i7 以上
メモリ: 16GB
HDD: 2TBx2(RAID1)

パーソナル解析モデル

微生物、mammal のマッピング可能
但し、WGS/WES データ解析には不向き

OS: Windows10 Pro 64bit
CPU: Core i7 以上
メモリ: 32GB
HDD: 4TBx2(RAID1)

推奨解析モデル

Mammal Exome の解析可能

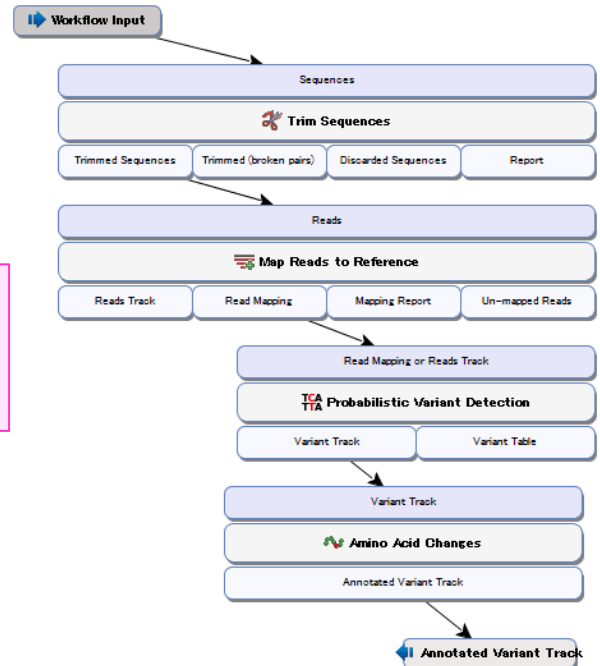
OS: Windows10 Pro 64bit
CPU: intel Xeon E5 以上
メモリ: 64GB
HDD: 10TB

【ワークフローツール】

Workflows ツールは、Genomics Workbench 内の個々の処理作業を連結し、一連の流れの解析として実行できるツールです。

- 独自の解析処理を作成可能
- 個々の処理ステップのパラメータ設定も自由に設定
- 結果、レポート保存有無の設定可能
- 作成したワークフローを配布することも可能
- 多数のサンプルをバッチでワークフローに投入することが可能なため、多量の検体も自動処理することができる

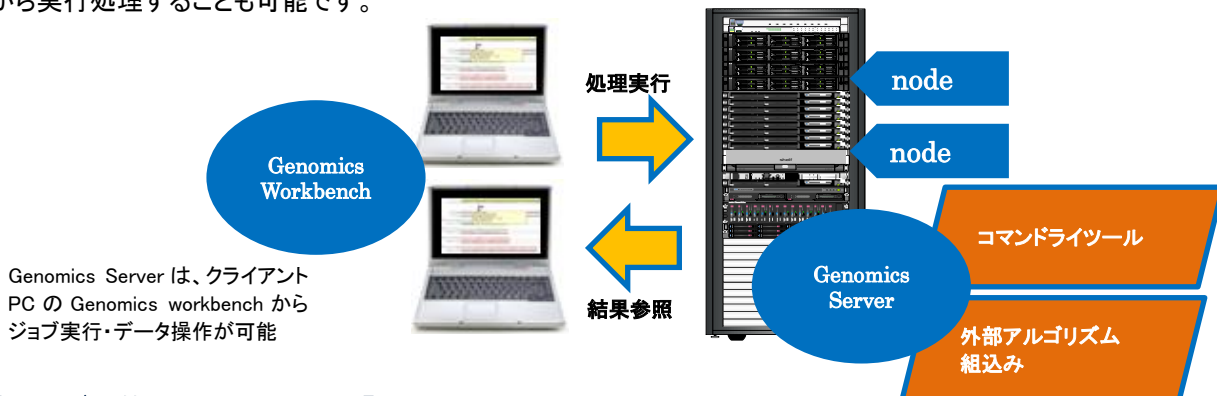
リードデータのクオリティトリミング→マッピング→変異検出→検出された変異がアミノ酸変化を伴うかの確認まで、一連の処理を連結したワークフローの例。
ワークフローを一度作製すると、処理に投入するリードデータを選択するだけで最終的な解析結果が得られます



GxS CLC Genomics Server

CLC Genomics Server は大量データ処理、データの共有化、アクセスセキュリティを考慮したシステム構築を検討されている方々に最適なソフトです。計算処理ツールと解析ツールを分離することでより効率よくデータの解析が行えます。

Genomics Server には**コマンドラインツール**が付属しており、処理のスキプトの組み込みが可能です。さらに、Bowtie や Velvet などの**外部アルゴリズムの組み込み**が可能で、これらアルゴリズムを Genomics Server/Workbench から実行処理することも可能です。



Genomics Server は、クライアント PC の Genomics workbench からジョブ実行・データ操作が可能

【ユーザー管理とアクセス制限】

- Genomics Server へのアクセスはユーザー名とパスワードで管理
- ユーザー毎のワーク領域設定可能
- データフォルダへのアクセス制限およびジョブ実行権限はユーザーおよびグループで割り当て可能
- フォルダごとに異なるアクセス権の設定可能

【ジョブ管理】

- Genomics Server の管理者はジョブの管理が可能
- ジョブの中止、削除が可能
- ノードを追加することでジョブの並列処理が可能
- ノードへの処理設定は管理者が設定
- グリッド利用による処理分散も可能
- 対応ジョブスケジューラーについては別途お問い合わせください。

ユーザー・グループ管理機能：データフォルダへのアクセス権限の割り当てやジョブ実行権限の設定可能

ゲノム解析モデル

Mammal Exome までの解析であれば、HiSeq 等のデータで対応可能。
大量サンプル処理の場合に推奨

OS: Windows10 Pro 64bit
CPU: intel Xeon E5 x 2 個
メモリ: 128GB
HDD: 20TB

大規模解析モデル

HiSeq 等のデータ解析をスムーズに実行可能。Whole ゲノム解析する方推奨。

OS: LINUX または Windows Server 12 (64bit)
CPU: intel Xeon E5 x 2 個
メモリ: 256GB
HDD: 4TBx8 (RAID5)

ワールドフュージョンでは Genomics Workbench および様々なモジュールを組み合わせ、オリジナルの「次世代シーケンスデータ解析システム」パッケージを構成いたします。詳細はお問い合わせください。

ライセンスラインナップ (一部)

CLC Genomics Workbench

様々なシーケンサーからのデータのアセンブル、マッピングに対応。de novo、reference mapping、miRNA 解析、変異同定、RNA-seq、ChIP-seq 解析などに対応。ワークフローツールが付属し、煩雑であった解析処理の工程をスムーズに行えまる。(一部ワークフロー不可能な工程有り)。

Genomics Gateway ビューワーは複数サンプルのマッピング結果とリファレンス結果を並べて表示、解析することも可能。また、多種多様なプラグイン(一部有料)は解析をバックアップ。

Genomics Workbench ネットワーク版

Genomics Workbench を複数で利用したい、でも実際に起動するのは数台でよい。といったニーズにお応えするライセンス形態。GWB のインストールは無制限。ライセンス数により同時アクセス数が制限される。リモートログインおよびアクセスでの利用可能。

Biomedical Genomics Workbench

変異解析、RNAseq 解析に特化した Genomics Workbench の姉妹品。1000genome、dbSNP、HapMap、Clinvar を搭載し、Trio 解析や家族性変異、体細胞変異の検出を行う専用のワークフローを搭載。

QIAseq DNA Panel 専用のワークフローを搭載しているため、スーパーリードの作成と分子バーコーディングを用いた低頻度変異の検出が可能。NGS データからのコピー数解析ツールが含まれる。

CLC Genomics Server

大量のデータ処理、複数のプロジェクト、データの共有化を行う場合に最適化されているサーバー型ソフトウェア。ユーザー管理、グループ管理、データの保存領域の設定、アクセスログ管理などの管理機能が充実 Genomics Workbench から操作可能。さらに、コマンドラインツールや External tool によるフリーツールの組み込みなども可能。Node 追加で並列処理可能。

CLC Assembly Cell

De novo アセンブルやマッピングをコマンドで処理したい方に最適のパッケージ。ルーチンで行うデータ処理や、様々なパラメータで解析および分析を行いたい場合に有効。

Blast2GO (GX プラグイン)

塩基配列からBLASTによってGOアノテーションを行うことが可能。GO グループの集計パイチャートの表示や、GO ツリー表示が可能。RNA-Seqの結果にBLAST2GOでまとめたアノテーション追加可能。

Microbial Genome Finishing Module (GX プラグイン)

微生物ゲノムまたはそれと同程度のサイズの全ゲノム決定のための支援ツール。コンティングの整理や、アセンブリに問題がある部分の検証を実施することができ、ゲノムの Finishing 工程を自動的に進めることが可能。

MetaGene Mark (GX プラグイン)

イントロンを含まないタンパク質コード領域を予測する Gene Finding tool。微生物や DNA ウィルスに対応しており、GeneMark.hmm prokaryotic algorithm により高精度な ORF 検出を実行。

Microbial Genomics Module (GX プラグイン)

16s rRNA のアンプリコンシーケンスデータを用いたメタゲノム解析を実行する Genomics Workbench プラグイン。GreenGene や Silva を参照配列として階層ごとの菌叢解析を実施。プラグインには NGS-MLST typing 機能と抗生物質抵抗性遺伝子を特定する機能も含まれており、採取した菌の種名が不明でも、近縁種の参照ゲノム配列を NCBI refseq で探索し、MLST.net や PubMLST.org から対応する MLST scheme を検出。

MLST (GX プラグイン)

バクテリア、酵母などについて DNA 配列に基づいた細菌の株タイプングツール。

QIAseq DNA V3 Panel Analysis[beta] (BX プラグイン)

QIAseq Targeted DNA Panel 専用の解析ツール。(イルミナ社シーケンサーからのリードのみ対応。)リードに付与された分子バーコードを基にスーパーリードを作成し、頻度の低い変異でも高い信頼性で検出。体細胞変異を検出するための一連の処理過程は付属のワークフローになっており、入力するシーケンスデータを指定することで複雑な操作を経ずに変異検出まで可能。

NOTE:

GX プラグインは Genomics Workbench のプラグインです。

BX プラグインは Biomedical Genomics Workbench のプラグインです。

上記の他、多数の有料または無料のプラグインが用意されています。

製品の詳細、およびデモ依頼は弊社までお問い合わせください。

記載されている製品の内容および仕様は変更になる場合がございます。最新の情報につきましてはお問い合わせください。



株式会社ワールドフュージョン

TEL: 03-3662-0521

E-mail: oshirase@w-fusion.co.jp

URL: <http://www.w-fusion.co.jp>