

CLC Genomics Workbenchを用いた リシークエンス解析

フィルジェン株式会社 バイオインフォマティクス部
(support@filgen.jp)

シーケンスデータのインポート

Import

クオリティチェック

QC for Sequencing Reads
Trim Reads

マッピング・再アライメント

Map Reads to Reference
Local Realignment

変異検出

Basic Variant Detection
Fixed Ploidy Variant Detection
Low Frequency Variant Detection

遺伝子名・アミノ酸変異情報の付加

Amino Acid Changes
Annotate with Overlap Information

シーケンスデータのインポート

Import

クオリティチェック

QC for Sequencing Reads
Trim Reads

マッピング・再アライメント

Map Reads to Reference
Local Realignment

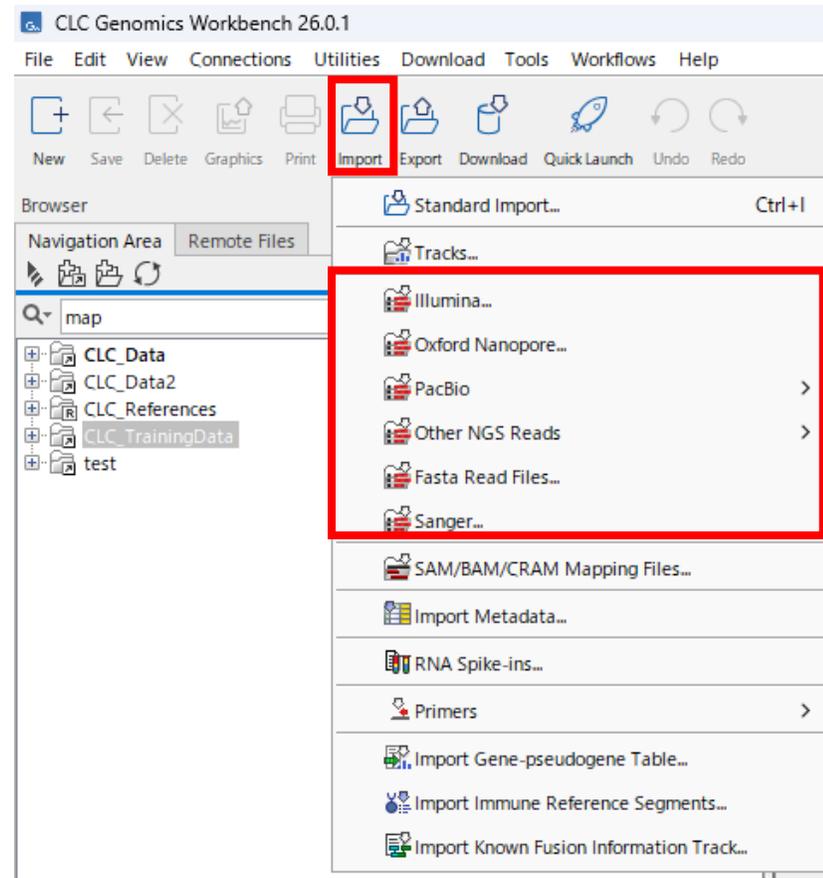
変異検出

Basic Variant Detection
Fixed Ploidy Variant Detection
Low Frequency Variant Detection

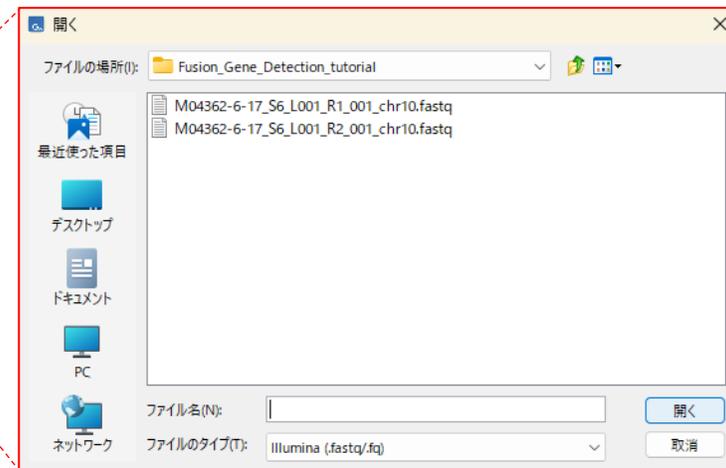
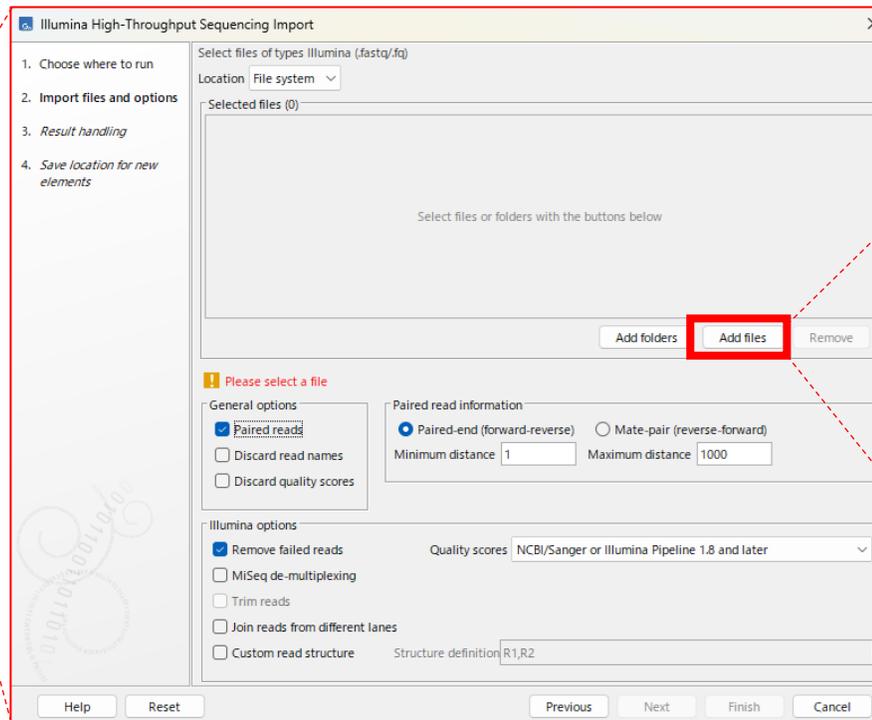
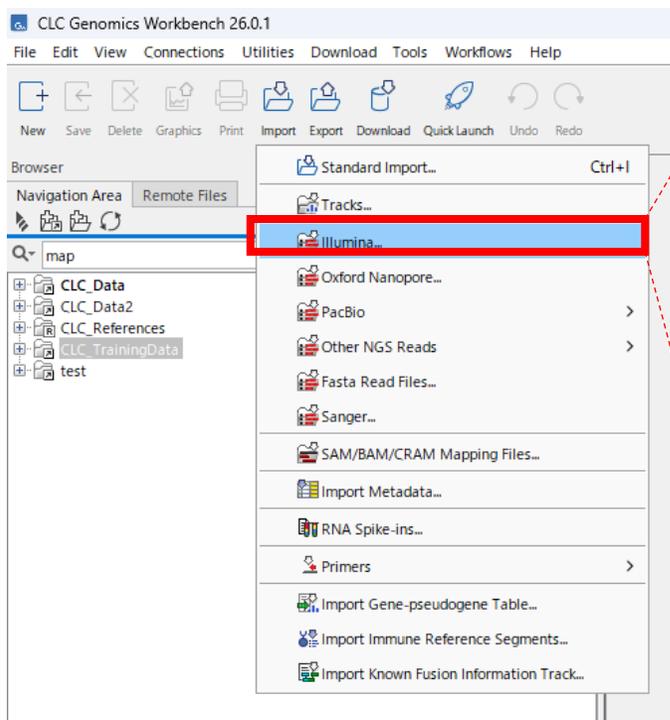
遺伝子名・アミノ酸変異情報の付加

Amino Acid Changes
Annotate with Overlap Information

- ✓ CLC Genomics Workbenchでは、シークエンサー機種やに合わせたインポートメニューを利用可能
- ✓ ToolbarのImportアイコンから表示されるインポーターから選択して、インポートを実行



- ✓ シーケンサー機種などに合わせてインポーターを選択し、シークエンスデータファイルを選択
- ✓ ペアエンドシークエンスデータのインポートにも対応



- ✓ シークエンスデータがインポートされ、各種解析に使用できるようになる
- ✓ 各リードの塩基配列やクオリティスコアを確認できる

The screenshot displays the CLC Genomics Workbench 26.0.1 interface. The main window shows a list of sequence reads with their corresponding quality scores. The reads are displayed in a grid format, with the sequence and quality score for each read. The quality scores are represented by a bar chart below the sequence. The reads are sorted by their position on the chromosome. The interface includes a navigation area on the left, a toolbox at the bottom left, and a settings panel on the right. The settings panel shows options for sequence layout, annotation layout, and position. The sequence layout settings include options for spacing, double stranded, numbers on sequences, relative to, numbers on plus strand, lock numbers, hide labels, and lock labels. The annotation layout settings include options for show annotations, position, offset, label, show arrows, and use gradients. The position settings include options for next to sequence, little offset, and stacked. The restriction sites, motifs, residue coloring, nucleotide info, positional stats, find, and text format settings are also visible.

シーケンスデータのインポート

Import

クオリティチェック

QC for Sequencing Reads
Trim Reads

マッピング・再アライメント

Map Reads to Reference
Local Realignment

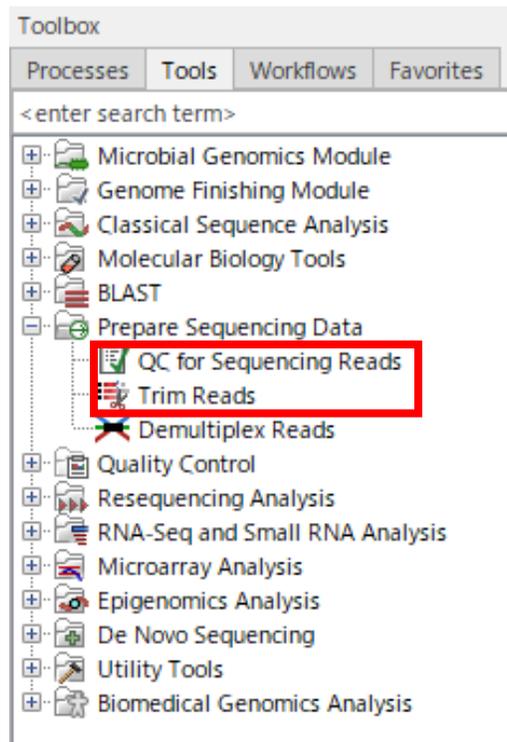
変異検出

Basic Variant Detection
Fixed Ploidy Variant Detection
Low Frequency Variant Detection

遺伝子名・アミノ酸変異情報の付加

Amino Acid Changes
Annotate with Overlap Information

- ✓ インポートしたシーケンスデータに対して、クオリティチェックレポートの作成や、低クオリティリードの除去などを行う
- ✓ その他、重複リードの除去や、マルチプレックスシーケンス時のサンプルバーコードのソートなどの、各種データ前処理用ツールなども利用が可能



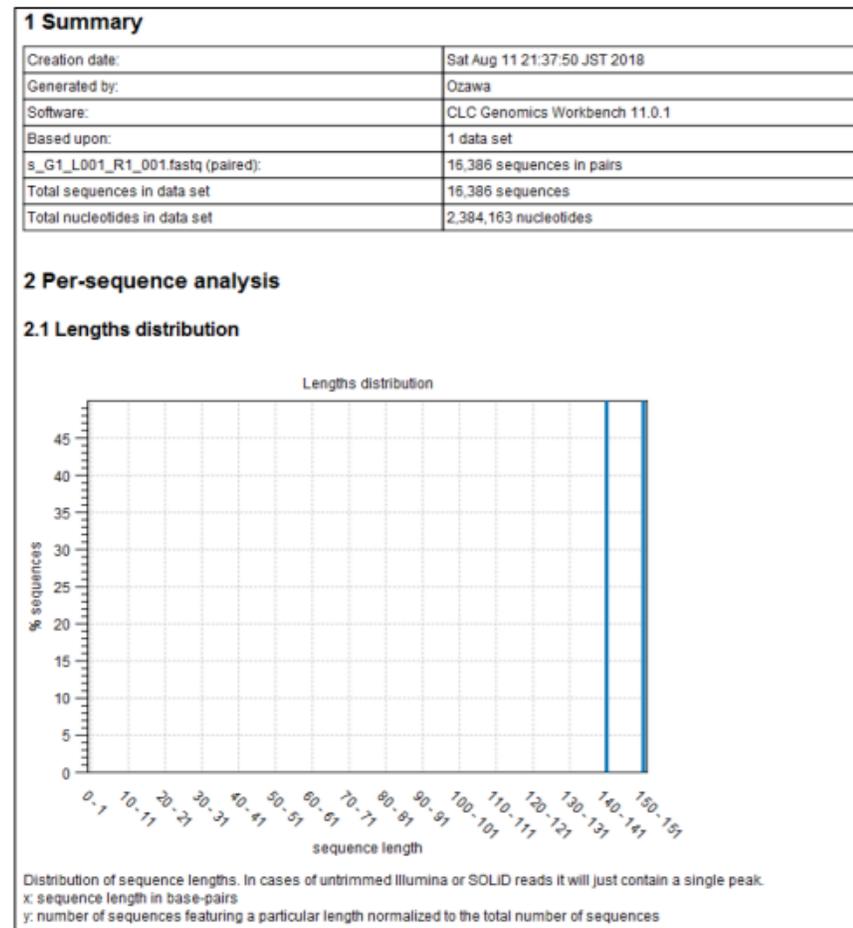
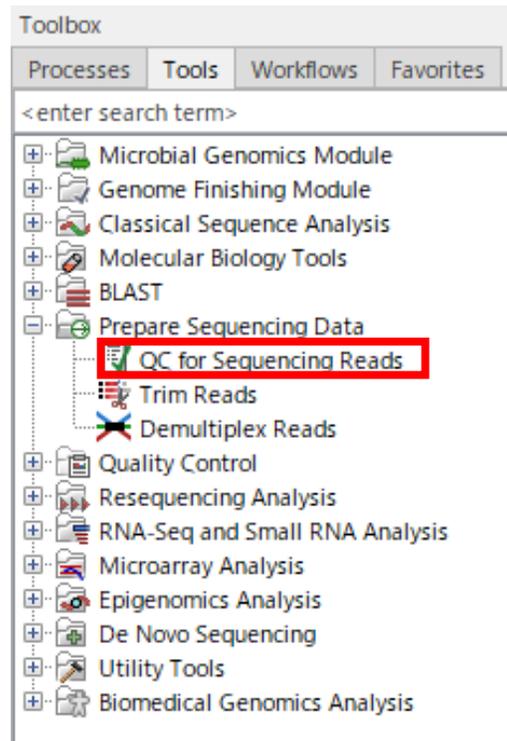
QC for Sequencing Reads

インポートしたシーケンスデータのクオリティやPCR Duplicateの状況を確認するためのレポートを作成

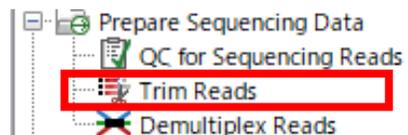
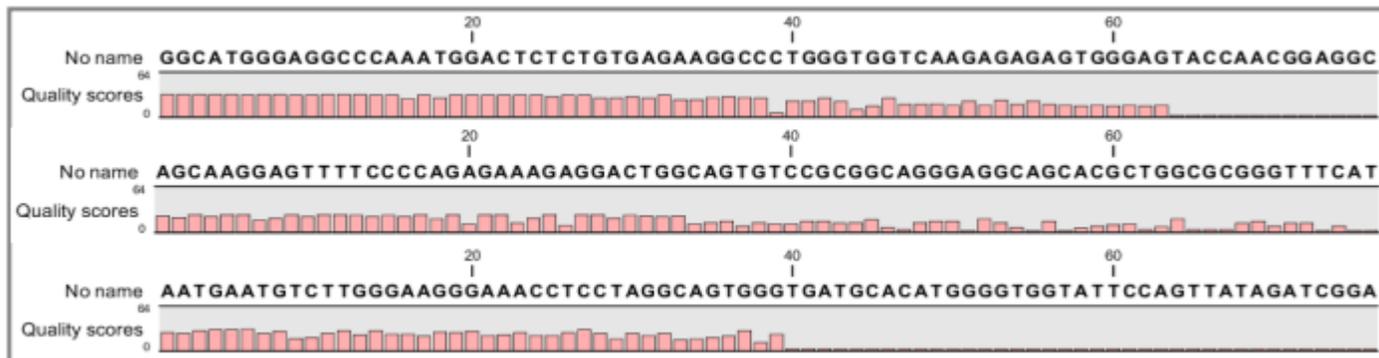
Trim Reads

アダプターの除去、クオリティスコアによる除去、長さを指定した除去などを選択・組み合わせて、リードのトリミングを実行

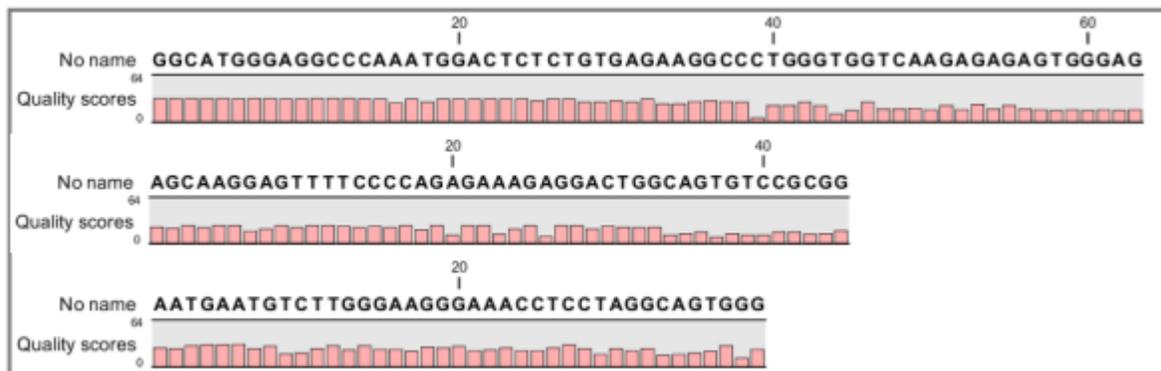
- ✓ QC for Sequencing Readsでは、シークエンスデータのクオリティ情報をまとめたレポートが作成される
- ✓ GC含量やクオリティスコア分布などのグラフデータや数値データを確認が可能



リードのQCCとトリミング



- ✓ Trim Readsの使用により、各リードの低クオリティ部分がカットされる
- ✓ アダプター配列の除去なども可能



シーケンスデータのインポート

Import

クオリティチェック

QC for Sequencing Reads
Trim Reads

マッピング・再アライメント

Map Reads to Reference
Local Realignment

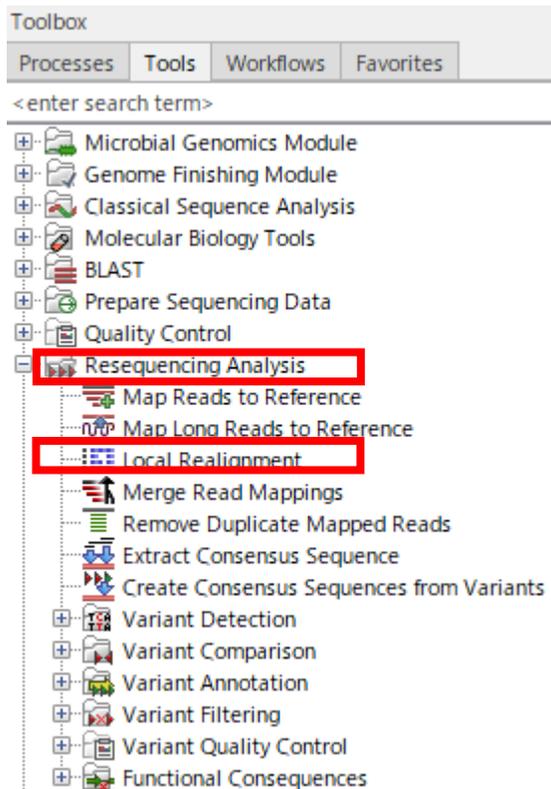
変異検出

Basic Variant Detection
Fixed Ploidy Variant Detection
Low Frequency Variant Detection

遺伝子名・アミノ酸変異情報の付加

Amino Acid Changes
Annotate with Overlap Information

- ✓ 変異の検出前に、シーケンスデータの各リードが、ゲノム上のどの部分を読んだものか、さらにどの部分で塩基が違っているかを調べるために、参照ゲノム配列へのリードのマッピングを行う
- ✓ サイズの大きいInsertion / Deletionを含んだリードは、マッピングが正確に行えない場合もあるので、必要に応じて再アライメントも行う



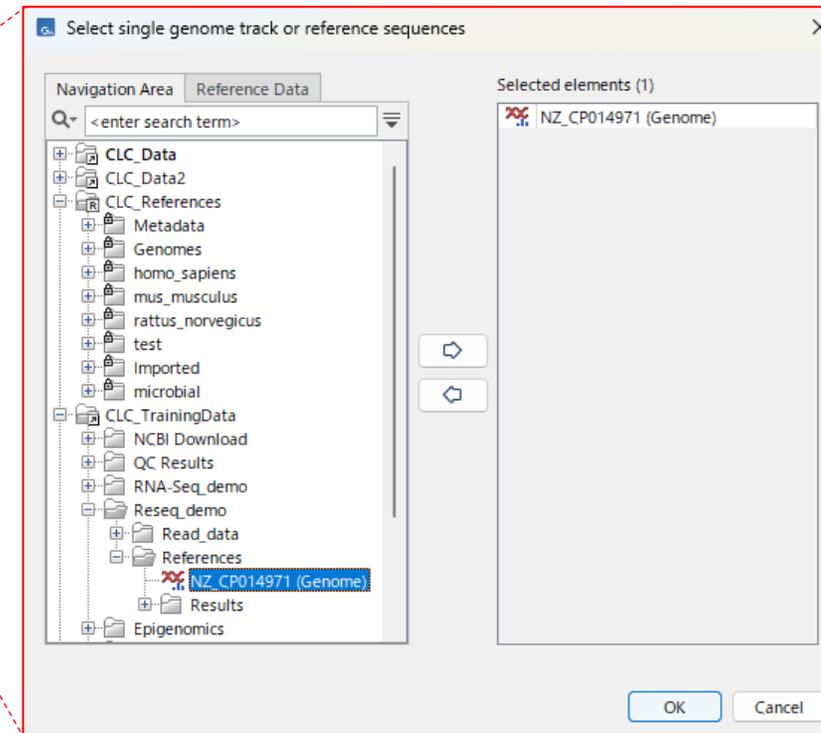
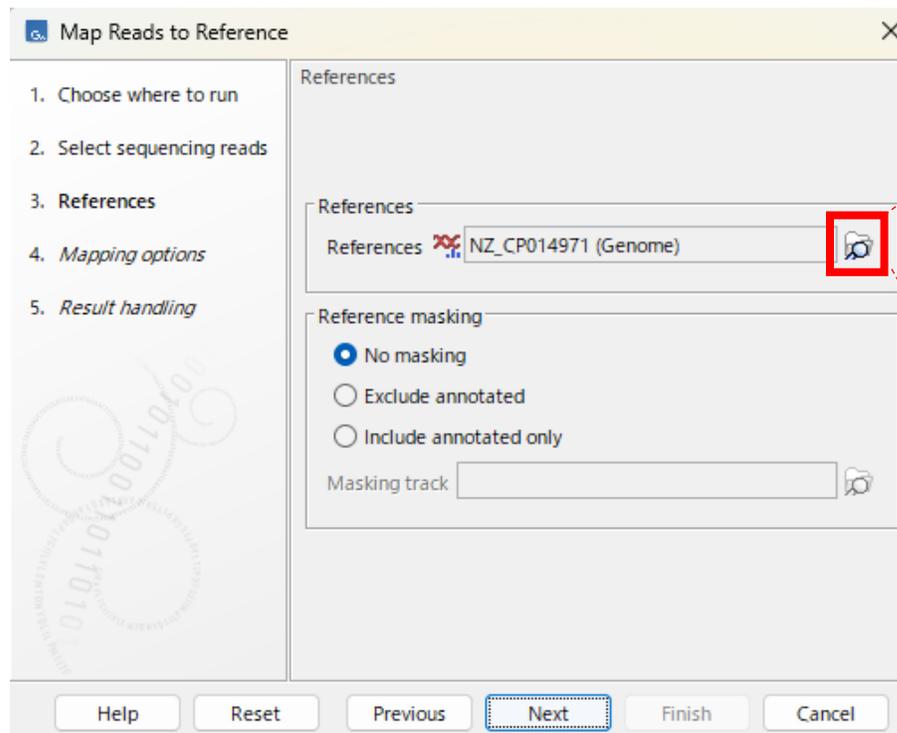
Map Reads to Reference

任意の参照ゲノム配列に対して、シーケンスデータのマッピングを行う

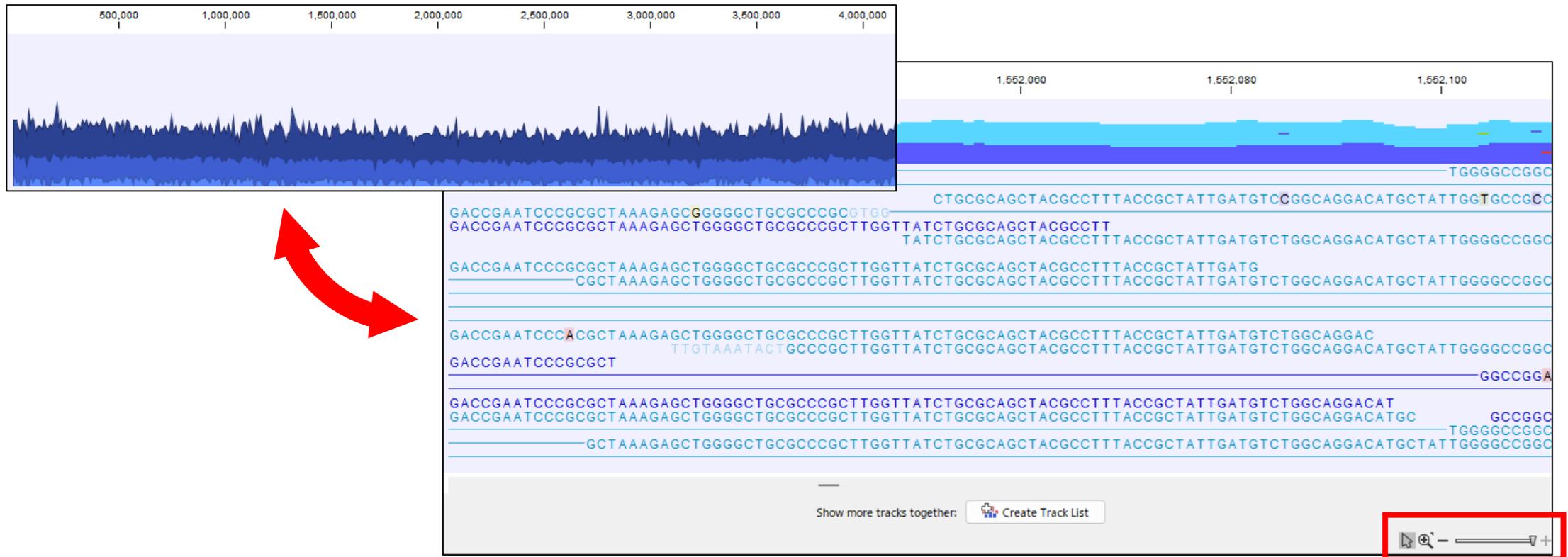
Local Realignment

すでにマッピングを実行したデータを使用し、リードの再アライメントを行う

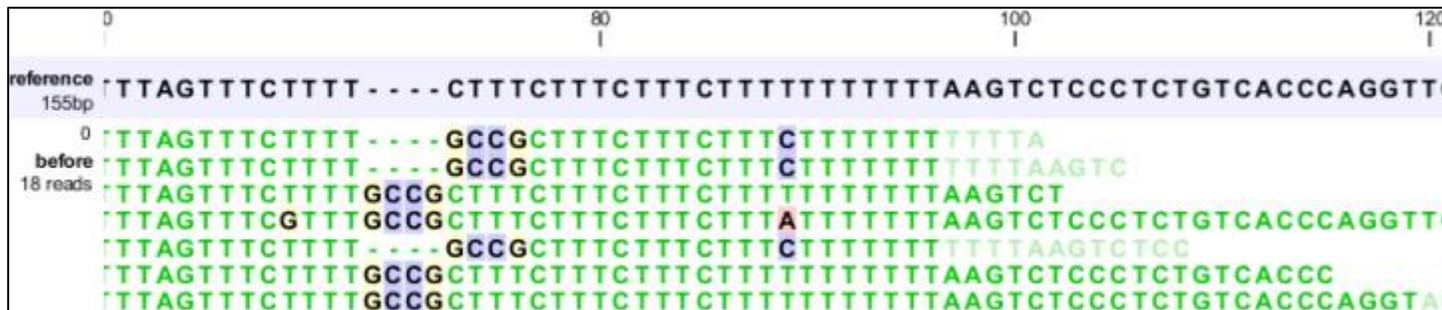
- ✓ Map Reads to Referenceでは、実行時のオプションパラメータで、任意の参照ゲノム配列データを選択可能
- ✓ ヒト、マウス、ラットなどのモデル生物の参照ゲノム配列データは、ソフトウェア標準搭載のダウンロードツールから取得でき、その他NCBIに登録されている参照ゲノム配列データや、ユーザーカスタム作成の配列データを使用することも可能



- ✓ Map Reads to Referenceを実行すると、ゲノム配列（染色体）ごとに、マッピングされたリード配列がどの領域に多いのかを示したカバレッジグラフが表示される
- ✓ カバレッジグラフを拡大していくと、各リードの塩基配列も表示され、変異部位なども確認できる



- ✓ Local Realignmentの使用により、すでにマッピングされた各リードに対して、同じ領域にマッピングされた周辺のリードに基づき、再アライメントが実行される
- ✓ サイズの大きいInsertion / Deletionの検出に対して有効



Local Realignment前



Local Realignment後

シーケンスデータのインポート

Import

クオリティチェック

QC for Sequencing Reads
Trim Reads

マッピング・再アライメント

Map Reads to Reference
Local Realignment

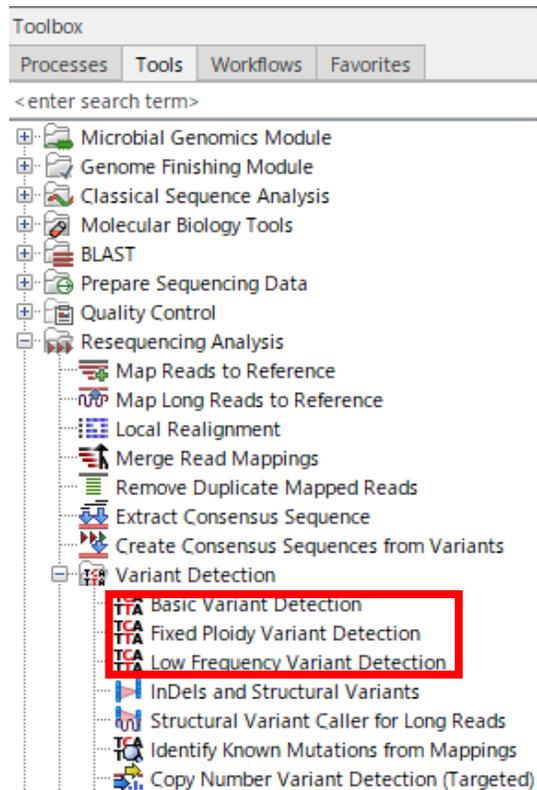
変異検出

Basic Variant Detection
Fixed Ploidy Variant Detection
Low Frequency Variant Detection

遺伝子名・アミノ酸変異情報の付加

Amino Acid Changes
Annotate with Overlap Information

- ✓ マッピング・再アライメント後のデータより変異の検出を行うためのツールは6種類あり、それぞれ検出可能な変異の種類やアルゴリズムが異なっているが、そのうち3種類がSNVまたは小サイズのInsertion / Deletionの検出に用いられる
- ✓ 変異検出時の感度や特異度、さらにサンプルの種類などに合わせて、どのツールを用いるのかを判断する必要がある



Basic Variant Detection

特殊な統計モデルを使用せずに、SNV, Small InDelを検出する
パラメータの設定を調整することで、検出可能な変異に制限を設けずに解析可能

Fixed Ploidy Variant Detection

確率モデルを用いてSNV, Small Indelを検出する
パラメータで指定したPloidy (倍数体) の値に基づいて変異の検出を行う
カバレッジ中に低頻度 (15%以下) で存在する変異は検出できない

Low Frequency Variant Detection

確率モデルを用いてSNV, Small Indelを検出する
カバレッジ中に低頻度で存在する変異の検出が可能

- ✓ 3ツールともオプションパラメータで、変異をもつ最低リード数や頻度、リードのクオリティなどを設定可能
- ✓ 遺伝子パネル解析の場合、変異検出の対象とする、ゲノム上のターゲット領域などを指定する

The screenshot shows the 'Basic Variant Detection' window with the 'General filters' tab selected. The left sidebar lists steps: 1. Choose where to run, 2. Select read mappings, 3. Basic Variant Parameters, 4. General filters (selected), 5. Noise filters, and 6. Result handling. The main panel is divided into three sections: 'Reference masking' with 'Ignore positions with coverage above' set to 100,000 and 'Restrict calling to target regions' with a file icon; 'Read filters' with 'Ignore broken pairs' checked, 'Ignore non-specific matches' set to 'Reads', and 'Minimum read length' set to 20; and 'Coverage and count filters' with 'Minimum coverage' at 10, 'Minimum count' at 2, and 'Minimum frequency (%)' at 35.0. The bottom bar contains buttons for Help, Reset, Previous (highlighted), Next, Finish, and Cancel.

The screenshot shows the 'Basic Variant Detection' window with the 'Noise filters' tab selected. The left sidebar lists steps: 1. Choose where to run, 2. Select read mappings, 3. Basic Variant Parameters, 4. General filters, 5. Noise filters (selected), and 6. Result handling. The main panel is divided into three sections: 'Quality filters' with 'Base quality filter' unchecked, 'Neighborhood radius' at 5, 'Minimum central quality' at 20, and 'Minimum neighborhood quality' at 15; 'Direction and position filters' with 'Read direction filter' unchecked, 'Direction frequency (%)' at 5.0, 'Relative read direction filter' checked, 'Significance (%)' at 1.0, 'Read position filter' unchecked, and 'Significance (%)' at 1.0; and 'Technology specific filters' with 'Remove pyro-error variants' unchecked, 'In homopolymer regions with minimum length' at 3, and 'With fraction below' at 0.8. The bottom bar contains buttons for Help, Reset, Previous, Next (highlighted), Finish, and Cancel.

- ✓ 検出された変異はテーブル形式で保存され、Excelなどにファイル出力が可能
- ✓ 変異テーブルには、変異の存在する位置や変異の種類、検出されたアリル、リード数などの情報が含まれる
- ✓ 変異テーブル上では、各テーブル項目に基づきフィルターをかけることができ、信頼性の高い変異のみを抽出することも可能

Chromosome: 変異の検出された染色体番号

Region: 変異のポジション

Type: 変異の種類(SNV, Insertion, Deletionなど)

Reference: リファレンスの塩基配列

Allele: 検出された塩基配列

Zygoty: 変異の接合性(HeteroかHomoか)

Count: マップされたリードのうち、変異を有するリードの数

Coverage: マップされたリード数

Frequency: 変異の頻度

Chromosome	Region	Type	Reference	Allele	Zygoty	Count	Coverage	Frequency
NZ_CP014971	10536^10537	Insertion	-	T	Homozygous	12	12	100.00
NZ_CP014971	13751	SNV	C	T	Homozygous	21	21	100.00
NZ_CP014971	16549	SNV	C	T	Homozygous	18	20	90.00
NZ_CP014971	20304	SNV	C	T	Homozygous	18	20	90.00
NZ_CP014971	43355	SNV	T	C	Homozygous	15	15	100.00
NZ_CP014971	92398	SNV	T	G	Homozygous	21	21	100.00

シーケンスデータのインポート

Import

クオリティチェック

QC for Sequencing Reads
Trim Reads

マッピング・再アライメント

Map Reads to Reference
Local Realignment

変異検出

Basic Variant Detection
Fixed Ploidy Variant Detection
Low Frequency Variant Detection

遺伝子名・アミノ酸変異情報の付加

Amino Acid Changes
Annotate with Overlap Information

- ✓ 変異テーブルには、変異部位に存在する遺伝子名や、変異によるアミノ酸置換などの情報を追加することが可能
- ✓ 変異データから、生物学的な解釈などを行う場合に、これらの情報が必要となる

変異データ

Chromosome	Region	Type	Reference	Allele	Count	Coverage	Frequency	Amino acid change	NZ_CP014971 (Gene)
NZ_CP014971	10536^10537	Insertion	-	T	12	12	100.00		
NZ_CP014971	13751	SNV	C	T	21	21	100.00	WP_000334066.1:p.Pro151Ser	AX01_RS00070
NZ_CP014971	16549	SNV	C	T	18	20	90.00	WP_000213163.1:p.Glu164Lys	AX01_RS00075
NZ_CP014971	20304	SNV	C	T	18	20	90.00		AX01_RS00090
NZ_CP014971	43355	SNV	T	C	15	15	100.00		gyrB
NZ_CP014971	92398	SNV	T	G	21	21	100.00	WP_063320106.1:p.Ile155Met	AX01_RS00435
NZ_CP014971	100693	SNV	C	T	11	11	100.00	WP_000997824.1:p.Pro39Leu	gatC
NZ_CP014971	131693	SNV	T	C	13	13	100.00		AX01_RS00610
NZ_CP014971	137541	SNV	C	T	23	24	95.83		AX01_RS00640
NZ_CP014971	147624	SNV	A	C	24	24	100.00		AX01_RS00670
NZ_CP014971	162996	SNV	A	C	17	18	94.44		AX01_RS00755
NZ_CP014971	164723	SNV	T	C	12	12	100.00	WP_063320107.1:p.Ser36Glv	AX01_RS00765

使用ツール

アミノ酸置換情報 遺伝子情報

- Amino Acid Changes
- Annotate with Overlap Information

- ✓ パラメータオプションで、遺伝子アノテーション情報データを指定して実行

Amino Acid Changes

The screenshot shows the 'Amino Acid Changes' window with the following settings:

- 1. Choose where to run**
- 2. Select variant track**
- 3. Set parameters**
 - References:**
 - CDS track: Homo_sapiens_ensembl_v74_CDS_chr17
 - mRNA track: Homo_sapiens_ensembl_v74_mRNA_chr17
 - Use transcript priorities
 - Sequence track: Homo_sapiens_sequence_hg19_chr17
 - Variant location:**
 - Move variants from VCF location to HGVS location
 - Flanking:**
 - Include upstream flanking positions: 5,000
 - Include downstream flanking positions: 2,000
 - Filtering and annotation:**
 - Filter away synonymous variants
 - Filter away CDS regions with no variants
 - Use one letter codon code
 - Sample type: DNA
 - Genetic code: 1 Standard
- 4. Result handling**

Buttons at the bottom: Help, Reset, Previous, Next, Finish, Cancel.



WP_063320106.1:p.Ile155Met	WP_063320106.1:p.Ile155Met	WP_063320106.1:c.465T>G
WP_000997824.1:p.Pro39Leu	WP_000997824.1:p.Pro39Leu	WP_000997824.1:c.116C>T
		WP_000774139.1:c.741A>G
		WP_061202184.1:c.270C>T
		WP_001068438.1:c.357T>G
		WP_001114508.1:c.771A>C
WP_063320107.1:p.Ser36Gly	WP_063320107.1:p.Ser36Gly	WP_063320107.1:c.106A>G
WP_063320108.1:p.Ile80Val	WP_063320108.1:p.Ile80Val	WP_063320108.1:c.238A>G

✓ パラメータオプションで、遺伝子アノテーション情報データを指定して実行

➡ Annotate with Overlap Information

1. Choose where to run

2. Select a track of annotations, variants, expressions or statistical comparisons

3. Overlap track

4. Result handling

Overlap track

Annotation options

Overlap track ➡ Homo_sapiens_ensembl_v74_Genes_chr17

Collapse duplicate attributes

Copy attributes

Help Reset Previous Next Finish Cancel

AX01_RS00070		AX01_RS00070	AX01_00070
AX01_RS00075		AX01_RS00075	AX01_00075
AX01_RS00090		AX01_RS00090	AX01_00090
gyrB	gyrB	AX01_RS00205	AX01_00205
AX01_RS00435		AX01_RS00435	AX01_00435
gatC	gatC	AX01_RS00480	AX01_00480
AX01_RS00610		AX01_RS00610	AX01_00610
AX01_RS00640		AX01_RS00640	AX01_00640
AX01_RS00670		AX01_RS00670	AX01_00670

- ✓ 変異テーブルのみではなく、マッピングされたリードを直接目視で確認し、実際にリードに変異が含まれているか、などの確認のために用いられる
- ✓ マッピングや変異テーブルデータのみではなく、データベースのアノテーション情報なども同時に表示させ、各変異の生物学的な解釈を行うことも可能

The screenshot illustrates the workflow for viewing genomic data in a browser. On the left, the 'Toolbox' menu is shown with 'Create Track List' highlighted in a red box. The central window, titled 'Select tracks from same genome', shows a tree view of genomic data for 'ERR277211.R1'. The 'Selected elements (6)' list includes: NZ_CP014971 (CDS), NZ_CP014971 (Gene), NZ_CP014971 (Genome), Mapped_reads-ERR277211.R1 (paired), Variants_passing_filters-ERR277211.R1 (Amino Acids), and Variants_passing_filters-ERR277211.R1 (paired). On the right, a multi-track visualization shows genomic tracks for coordinates 2,062,220 to 2,062,300. The tracks include reference sequences (e.g., NZ_CP014971 (CDS)), mapped reads (ERR277211.R1), and variant calls (Amino Acids). The amino acid sequence at the bottom is: K (H T G W G S T T P W F G T N Y W E Q F M A D L K D W K Q S V T W Y A I).

お問い合わせ先：フィルジエン株式会社

TEL: 052-624-4388 (9:00～17:00)

FAX: 052-624-4389

E-mail: support@filgen.jp