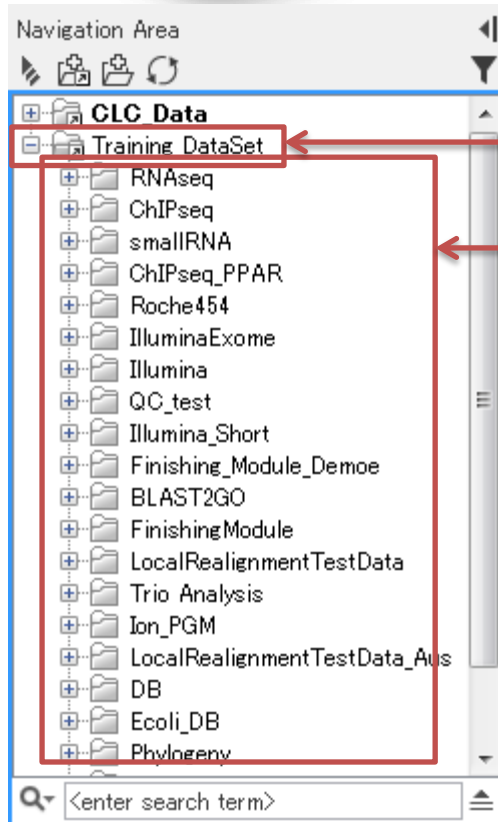
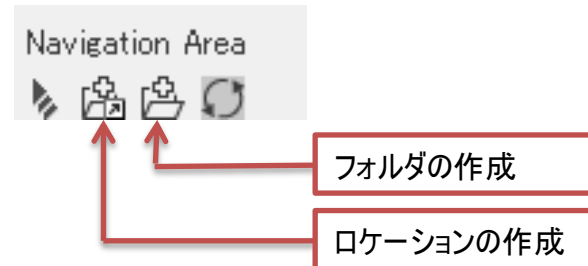


リード・ゲノム・アノテーションインポート

LocationとFolder



- Genomics Workbenchではデータを以下のような階層構造で保存可能です。フォルダの一番上位の階層を「Location」と呼び、その下の階層を「Folder」と呼びます。
- データの保存場所はロケーション毎に設定可能です。たとえばあるデータはCドライブに保存し、あるデータはDドライブに保存するといった事が可能です。
- ロケーション、フォルダの作成は以下のアイコンから作成できます。



データインポート

The screenshot shows a software menu bar with 'File', 'Edit', 'View', 'Download', 'Toolbox', 'Workspace', and 'Help'. Below the menu bar is a toolbar with icons for 'Show', 'New', 'Save', 'Import', 'Export', 'Graphics', 'Print', 'Undo', 'Redo', 'Cut', 'Copy', 'Paste', and 'Delete'. The 'Import' menu is open, showing the following options:

- Standard Import... Ctrl+I
- Tracks...
- Roche 454...
- Illumina...
- SOLiD...
- Fasta Read Files...
- Sanger...
- Ion Torrent...
- SAM/BAM Mapping Files...

Red boxes and arrows point from the menu items to descriptive text boxes:

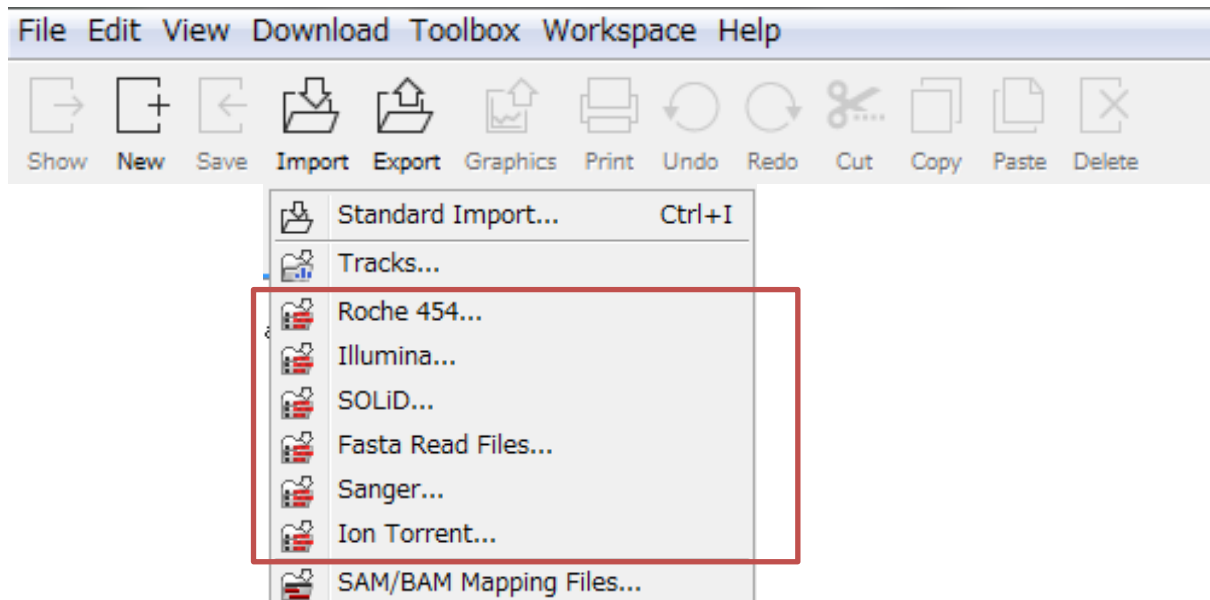
- Standard Import... points to "NGSデータ以外のインポーター"
- Tracks... points to "アノテーションファイルやBEDフォーマットのファイルなど"
- Sanger... points to "Sanger, NGSデータインポーター"
- SAM/BAM Mapping Files... points to "外部マッピングデータインポーター"

- Standard Importは、サンガーシーケンサー、次世代シーケンサー以外のファイルのインポートに利用します。



リードデータインポート

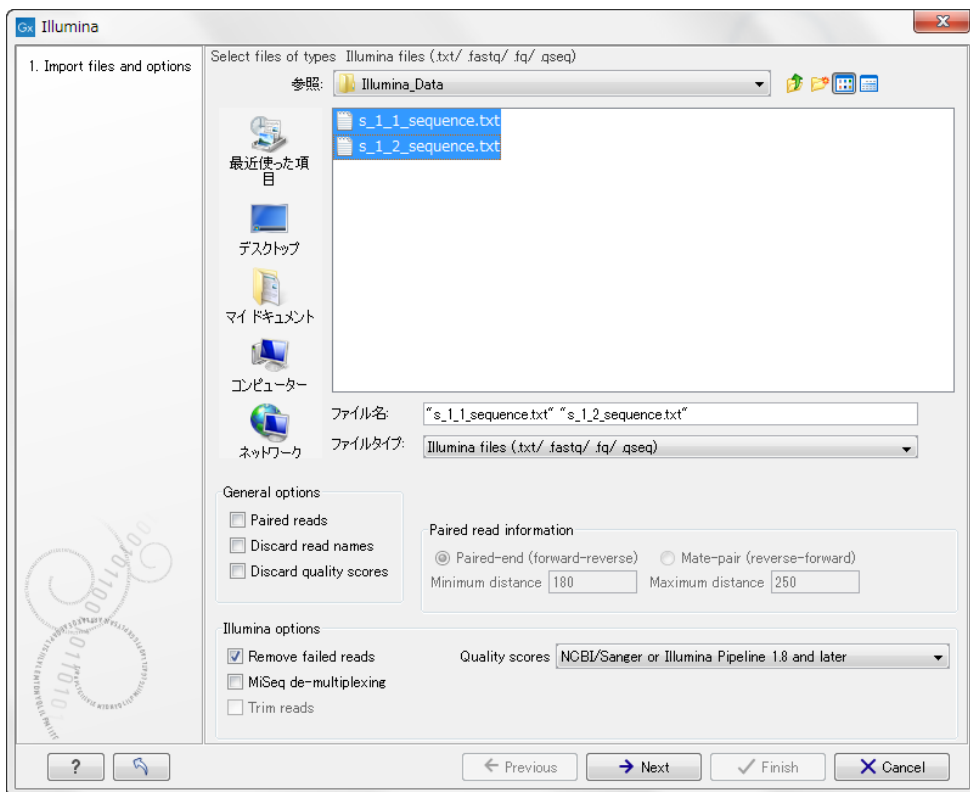
- Importからインポートしたいリードのシーケンサータイプを選択。





リードデータインポート

• Illuminaデータのインポート



General options

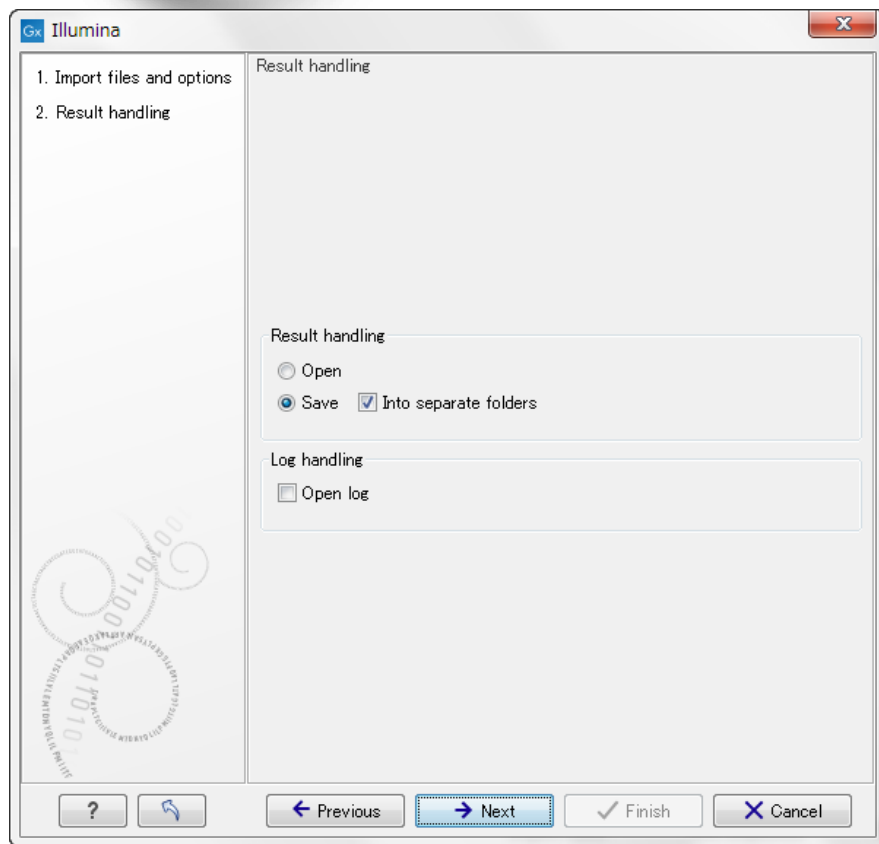
- Paired reads:ペアかどうか。
- Discard reads names:リードについている名前を捨てるかどうか。デフォルトでは捨てるとなっていますが、マッピング後、SAMにてExportした際など、リード名で確認したい場合があるため、最初は保存しておきましょう。
- Discard quality scores:Quality Scoreが必要ない場合はこのオプションにチェック。通常、インポート後にクオリティスコアが必要な事が多いです。
- Paired read orientation:ペアの距離と向きを指定。

Illumina options

- Remove failed reads:シーケンサーでfailとマークされたリードを除去するかどうか。
- MiSeq de-multiplexing:MultiplexingされたデータをDe-multiplexingするかどうか。
- Quality Score:使用するQuality Scoreのバージョンの選択。

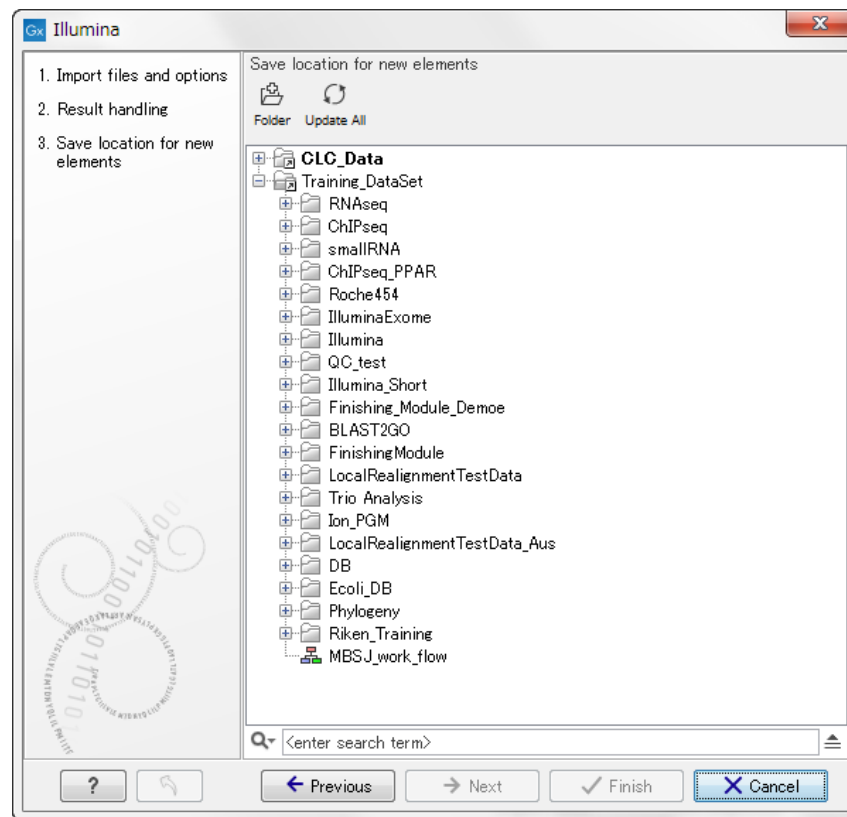


リードデータインポート



Result handling

- データを開くか、保存の選択
- Into separate folders では、別々のフォルダへ保存するかどうかを選択できます。バッチ処理を行う際に便利です。

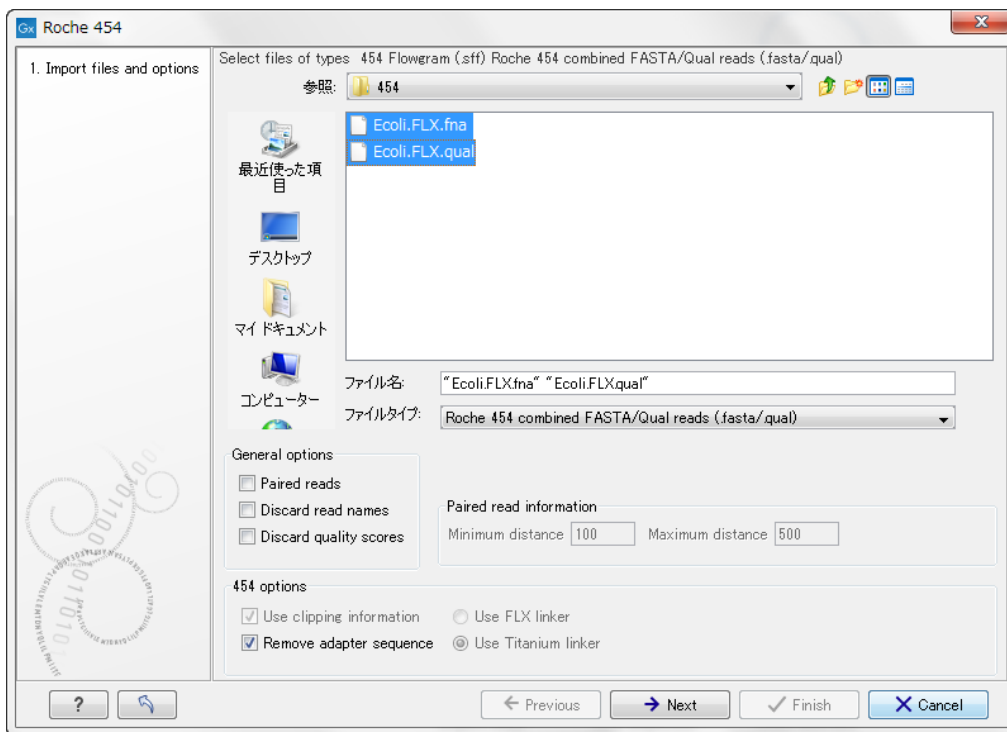


保存先の指定



リードデータインポート

- Roche 454/Juniorのインポート (.fna/.qual または .sff)



General options

- Paired reads:ペアかどうか。
- Discard reads names:リードについている名前を捨てるかどうか。デフォルトでは捨てるとなっていますが、マッピング後、SAMにてExportした際など、リード名で確認したい場合があるため、最初は保存しておきましょう。
- Discard quality scores:Quality Scoreが必要ない場合はこのオプションにチェック。通常、インポート後にクオリティスコアが必要な事が多いです。
- Paired read orientation:ペアの距離と向きを指定。

454 options

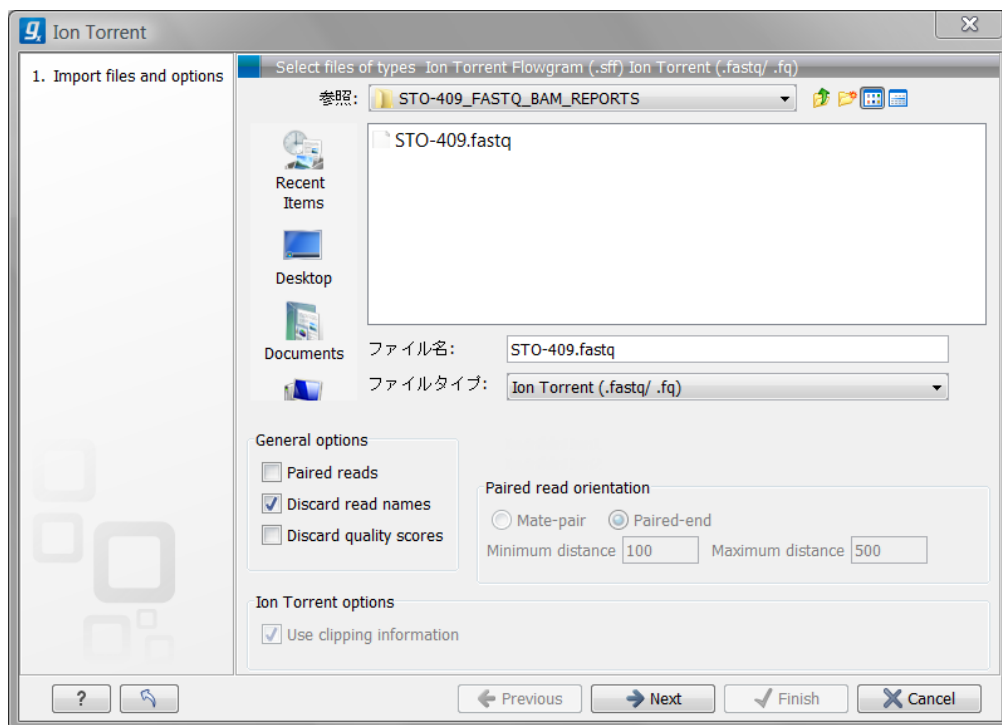
- Use clipping information:.sffファイルに含まれるクリッピングの情報を利用するとき。
 - Use FLX linker:FLXのリンカーを使用するとき
 - Use Titanium linker:Titaniumのリンカーを使用するとき
- Remove adapter sequence:アダプター配列を取り除きたいとき

この次のウィザードの画面はイルミナと共通です。



リードデータインポート

- Ion Torrentのインポート (.fastq または .sff)



General options

- Paired reads:ペアかどうか。
- Discard reads names:リードについている名前を捨てるかどうか。デフォルトでは捨てるとなっていますが、マッピング後、SAMにてExportした際など、リード名で確認したい場合があるため、最初は保存しておきましょう。
- Discard quality scores:Quality Scoreが必要ない場合はこのオプションにチェック。通常、インポート後にクオリティスコアが必要な事が多いです。
- Paired read orientation:ペアの距離と向きを指定。

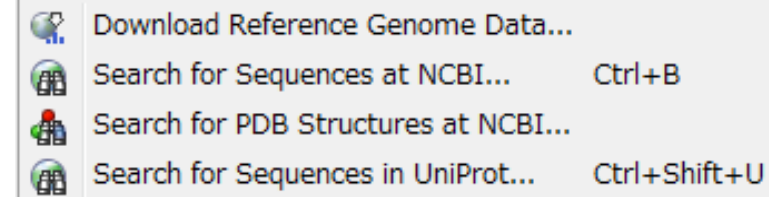
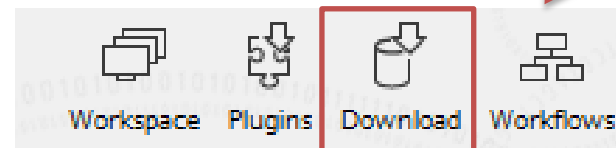
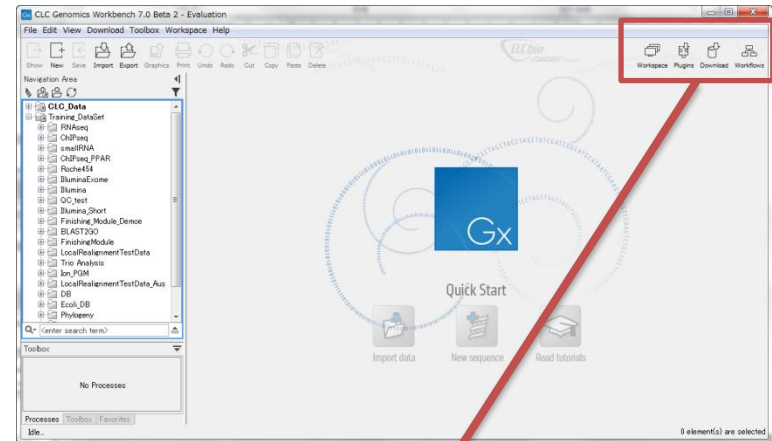
Ion Torrent options

- Use clipping information:.sffファイルに含まれるクリッピングの情報を利用するとき(.sffファイルを選択した場合)。

この次のウィザードの画面はイルミナと共通です。

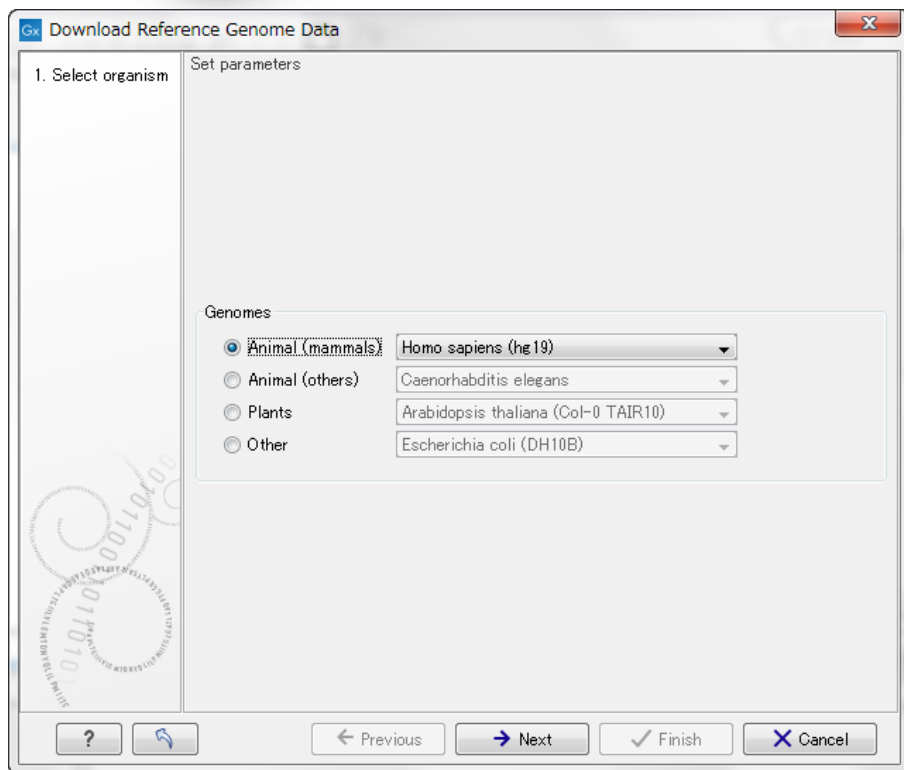
ゲノム・アノテーションインポート

- ゲノムはダウンロードアイコンより、生物種を指定してアノテーションと共にインポートすることが可能です。
- ゲノム配列とともに、アノテーションファイルをダウンロードすることも可能です。
- すでにGenomics Workbenchへ取り込んでいるゲノム配列について、アノテーションを付加することも可能です。

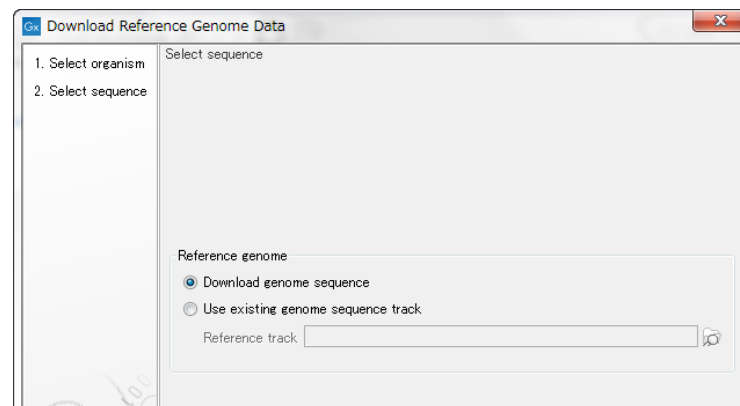




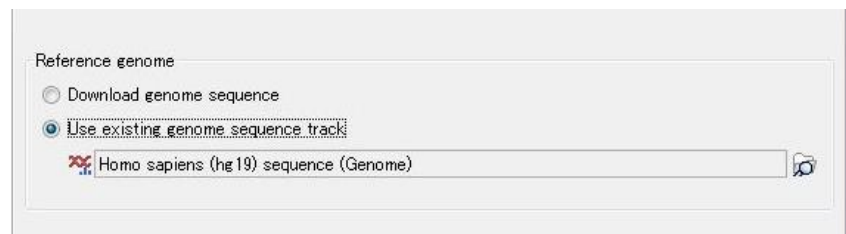
ゲノム・アノテーションインポート



- ドロップダウンリストから生物種を選択。



- Download genome sequence: 新規にゲノムをダウンロードする場合。
- Use existing genome sequence track: すでにダウンロードしたゲノムにアノテーションを追加する場合。以下のようにトラックのフォーマットになっているゲノムを選択。



ゲノム・アノテーションインポート

Download Reference Genome Data

1. Select organism
2. Select sequence
3. Select annotations

Select annotations

Download	Name	Version	Provider	Size (in Mb)
<input type="checkbox"/>	Sequence	74	Ensembl	827
<input checked="" type="checkbox"/>	Gene annotation	74	Ensembl	28
<input checked="" type="checkbox"/>	Dbsnp (common) variants	137	UCSC	528
<input checked="" type="checkbox"/>	Dbsnp variants	137	UCSC	1472
<input checked="" type="checkbox"/>	COSMIC	v67_241013	SANGER	50
<input checked="" type="checkbox"/>	Clinical variants in dbSNP		NCBI	1
<input checked="" type="checkbox"/>	HapMap Variants		Ensembl	441
<input checked="" type="checkbox"/>	1000genomes	phase1	Ensembl	1912

Total download size (in Mb): 5261

? ↶ ↷ ✓ ✕

Previous Next Finish Cancel

- 希望するアノテーションにチェックを入れる。ゲノム配列をダウンロードするときは、Sequencesにもチェックを入れる。
- 選択した生物種により、表示されるアノテーションの種類は異なります。



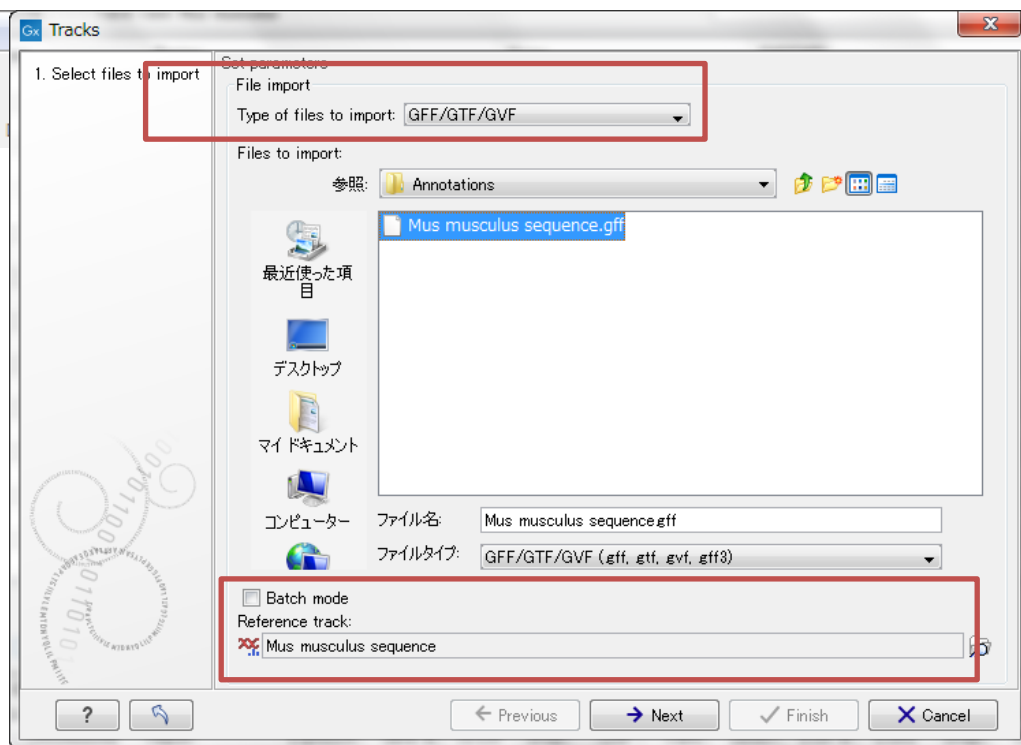
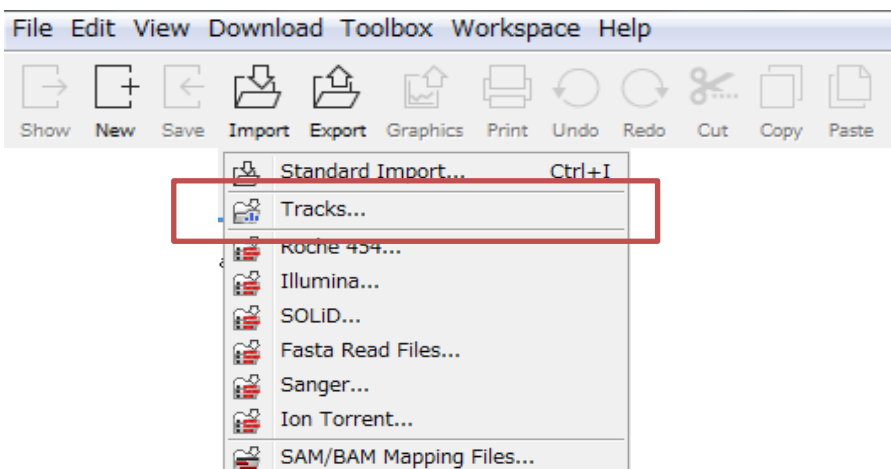
アノテーションインポート

- Download Genome 以外にも、アノテーションファイルをインポート可能です。
- アノテーションとして取り込めるファイルは以下のフォーマットです。
- アノテーションファイルをインポートする際には、対象となるゲノム配列がすでにインポートされ、Trackのフォーマットになっていることが前提です。
 - VCF
 - GFF/GTF/GVF
 - BED
 - Wiggle
 - Complete Genomics Var file
 - UCSC Variation table dump
 - COSMIC variation database

※変異のデータについても、アノテーションとして自分の変異へアノテーションとして情報の追加や比較ができるため、アノテーションのインポート可能フォーマットに含めています。

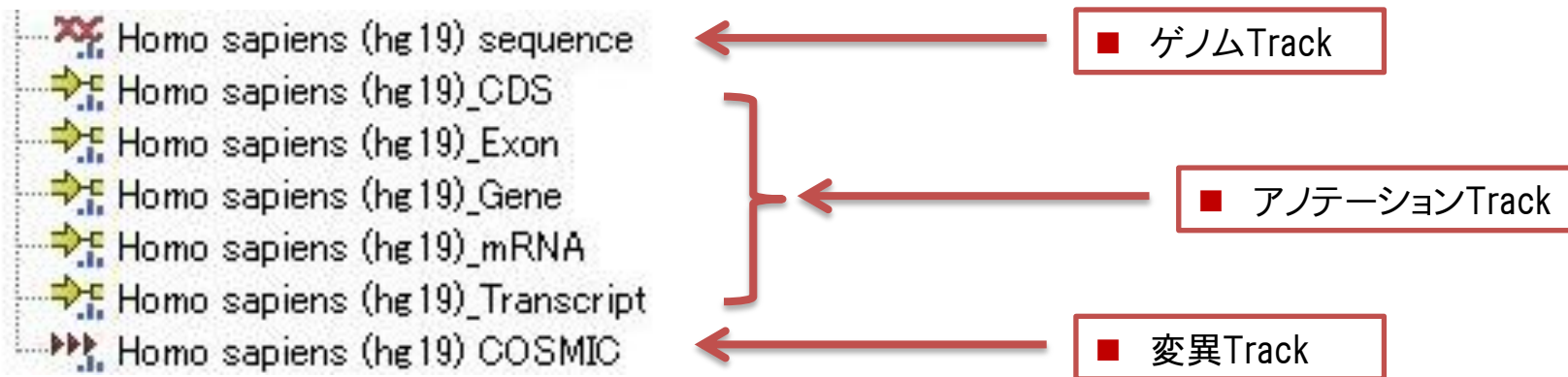


アノテーションインポート



- Type of files to importを選択
- インポートするファイルを選択
- Reference Track を選択

ゲノム・アノテーションインポート



- TrackにはGenomics Workbenchが認識しているタイプによりアイコンがそれぞれ異なります。解析によっては、Trackのタイプを認識して、解析に利用できるかどうか区別するため、データがどのようなタイプで認識されているか、アイコンで確認するようにしてください。