

CLC Genomics Workbenchを用いた バイサルファイトシーケンス解析

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)

M M M
ATGCGCGTAGCGATGACGACA

M : メチル基

非メチル化シトシンをチミンに変換

ATGCGTGTAGCGATGACGATA

ATGCGTGT
GCGTGTAGC
CGTAGCGATG
TAGCGATGA
CGATGACGAT
ATGACGATA
ATGCGCGTAGCGATGACGACA

リファレンスへのマッピング



断片化してシーケンス

CGATGACGAT
CGTAGCGATG
ATGCGTGT
TAGCGATGA
GCGTGTAGC
ATGACGATA

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

リファレンスへのマッピング

Map Bisulfite Reads to Reference

メチル化の検出

Call Methylation Levels

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

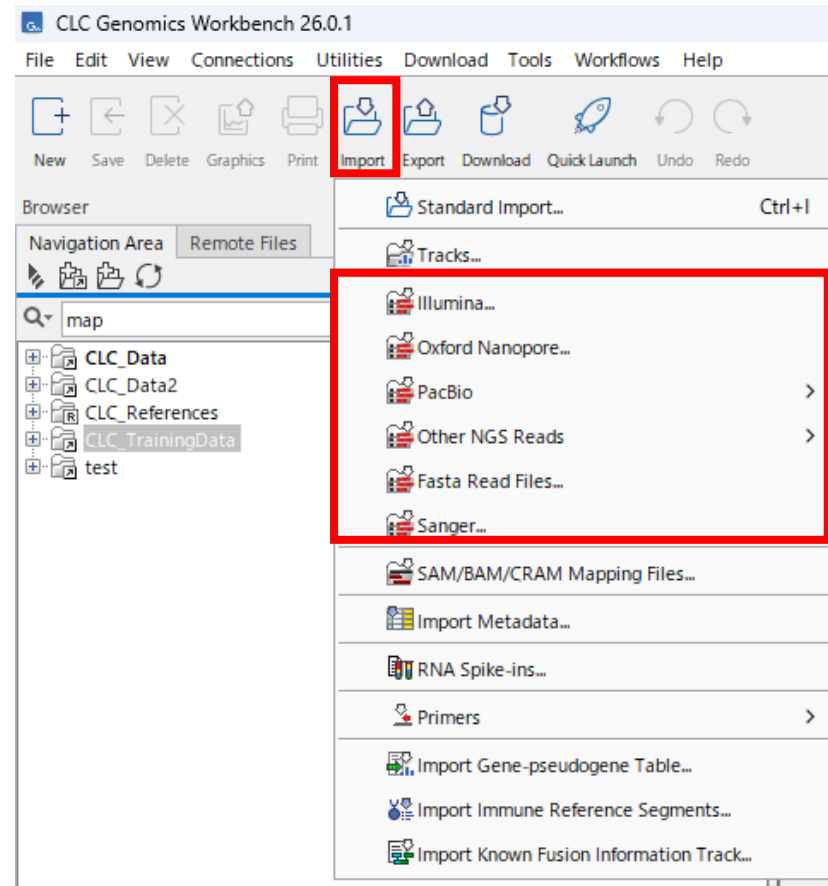
リファレンスへのマッピング

Map Bisulfite Reads to Reference

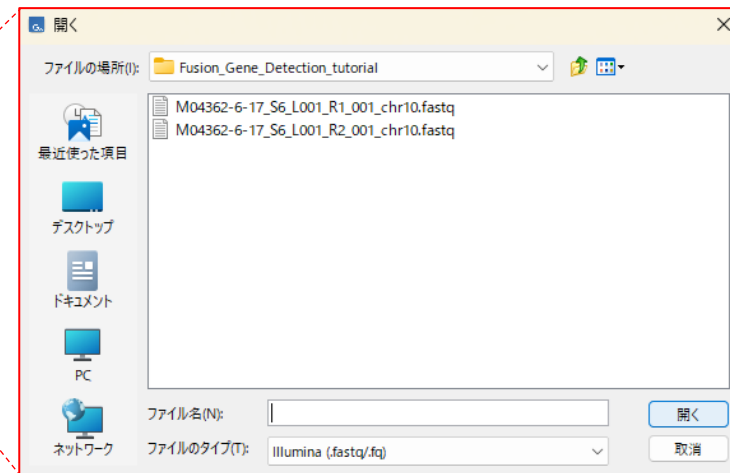
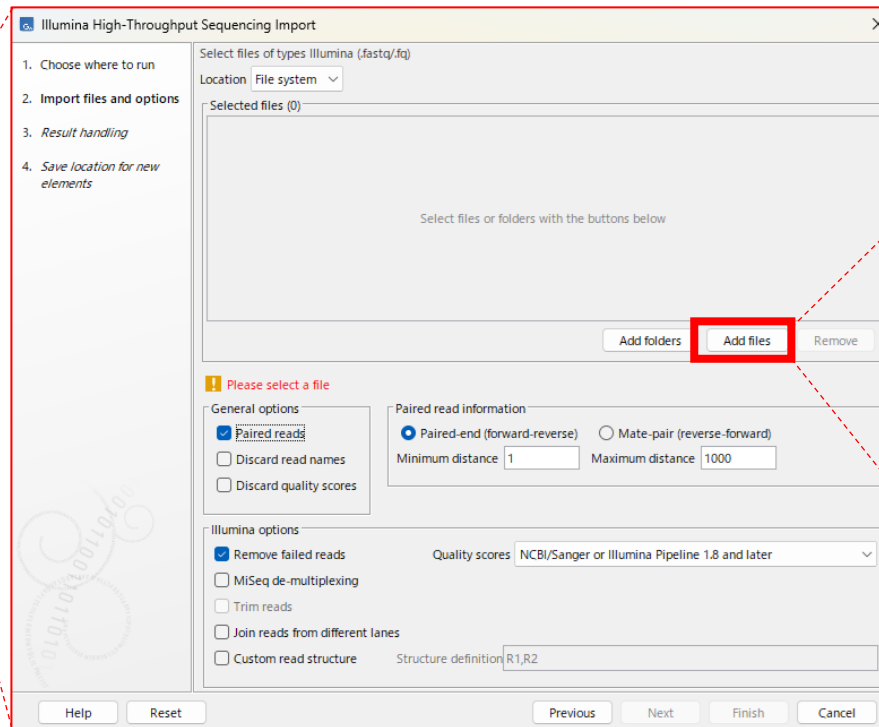
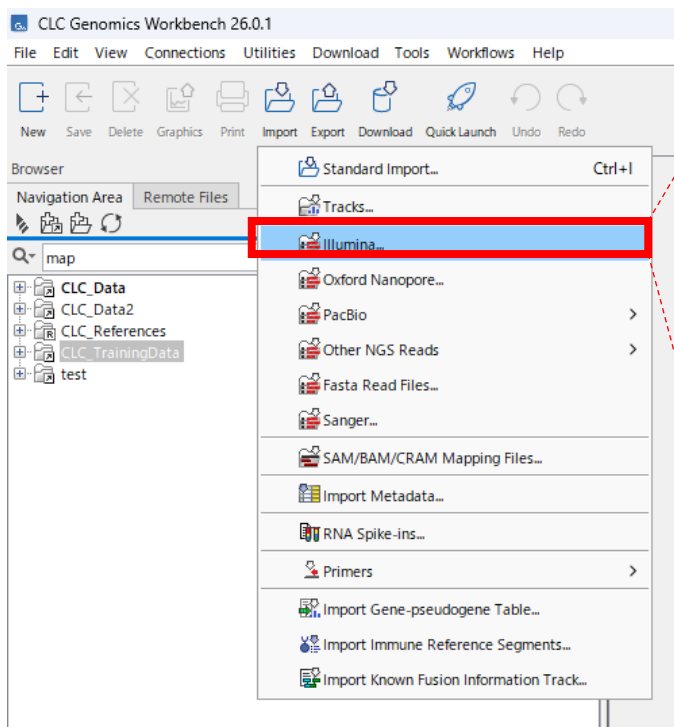
メチル化の検出

Call Methylation Levels

- ✓ 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 base sequences and quality scores. The reads are organized into a grid, with each row representing a different read. The quality scores are visualized as a bar chart below each sequence. The interface includes a navigation pane on the left, a toolbar at the top, and a settings panel on the right. The settings panel is currently open, showing options for sequence layout, alignment, and annotation. The sequence layout settings are set to 'No spacing', 'Numbers on sequences' is checked, and 'Numbers on plus strand' is also checked. The alignment settings are set to 'Next to sequence', 'Little offset', and 'Stacked'. The annotation settings are set to 'Show annotations' and 'Use gradients'.

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

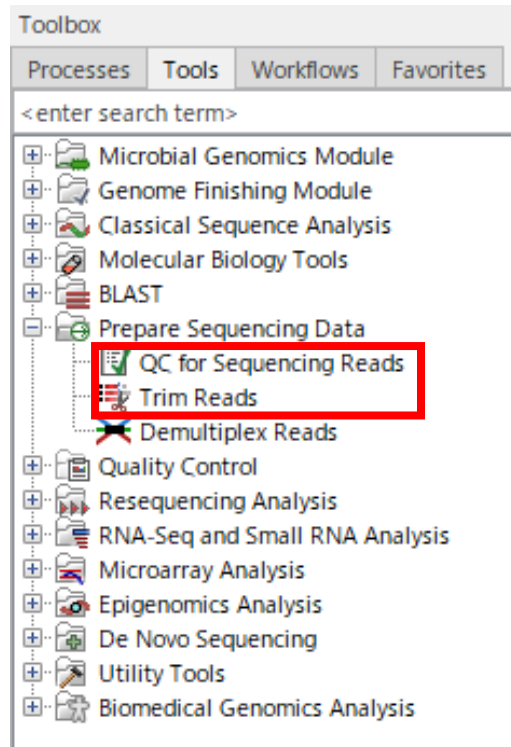
リファレンスへのマッピング

Map Bisulfite Reads to Reference

メチル化の検出

Call Methylation Levels

- ✓ 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



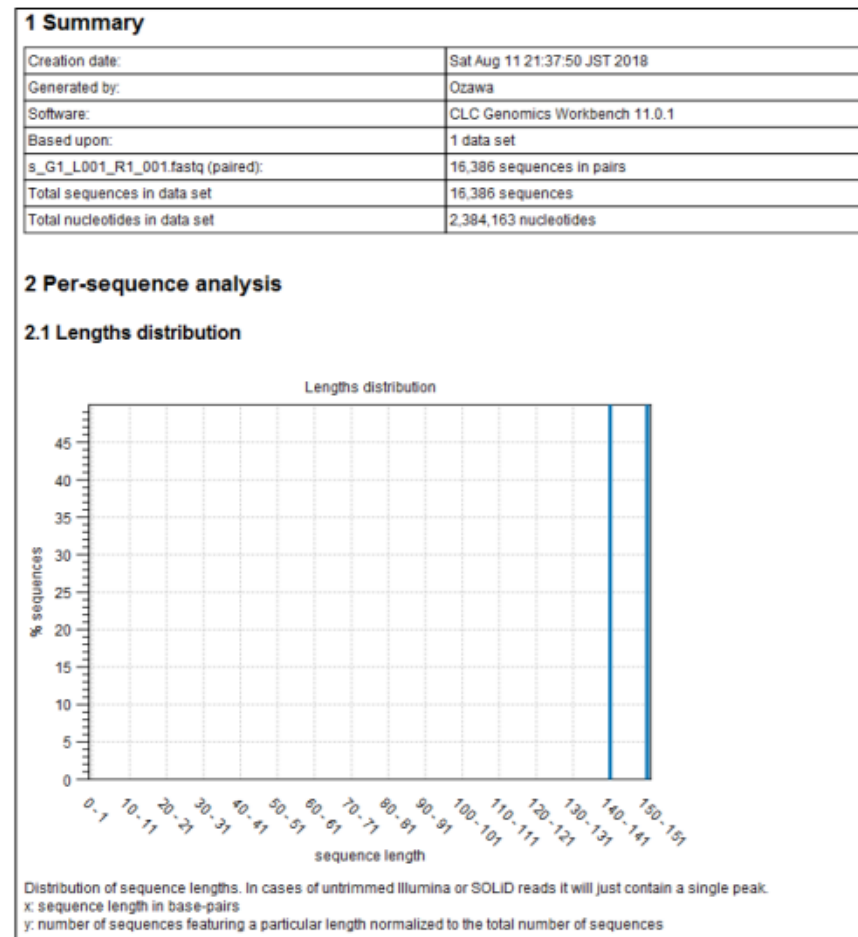
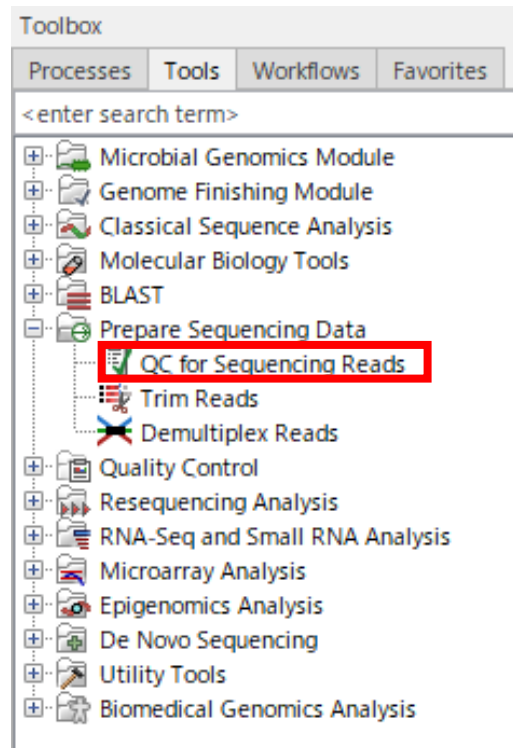
QC for Sequencing Reads

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

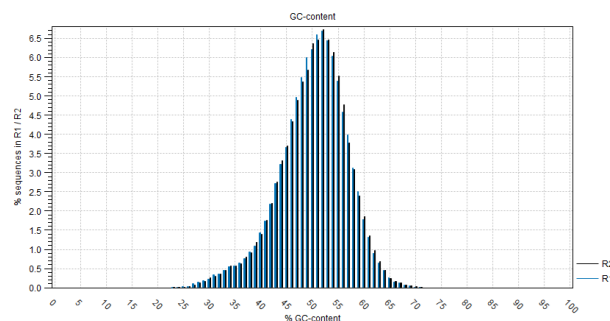
- ✓ QC for Sequencing Readsでは、シークエンスデータのクオリティ情報をまとめたレポートが作成される
- ✓ GC含量やクオリティスコア分布などのグラフデータや数値データを確認が可能



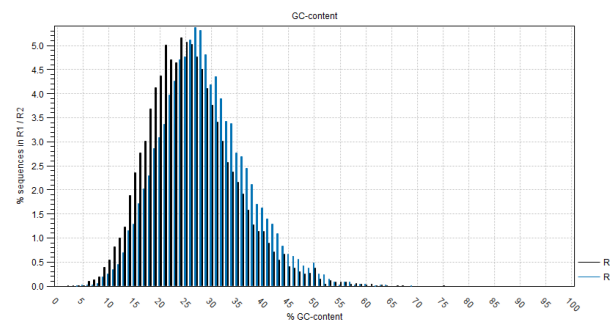
- ✓ バイサルファイト処理済み配列のQCチェックの解釈には注意が必要

GC含量

一般的なリード



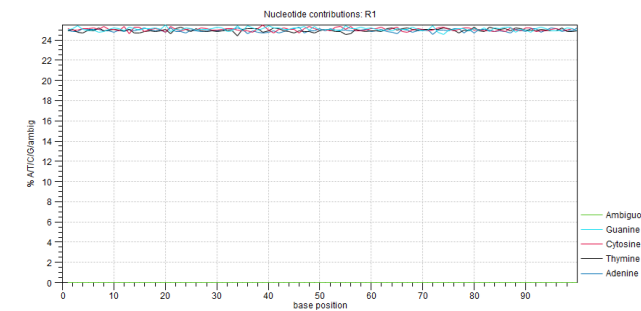
バイサルファイト処理済みリード



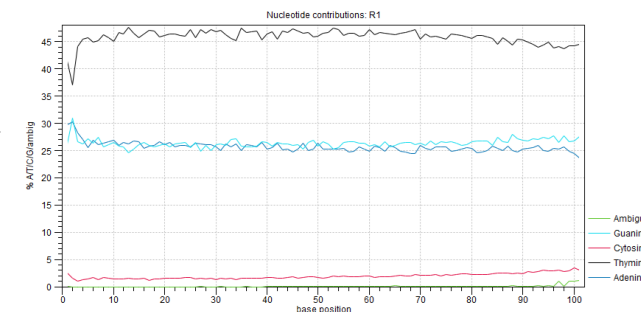
バイサルファイト処理済みリードでGC含量が低い

ポジション別ATGC割合

一般的なリード

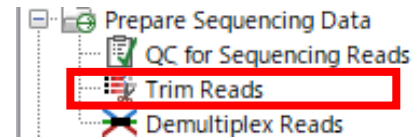
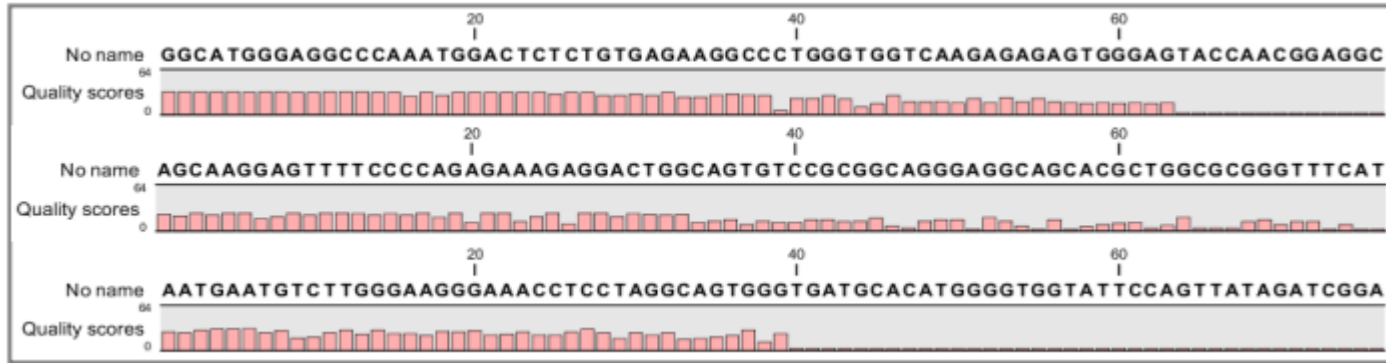


バイサルファイト処理済みリード

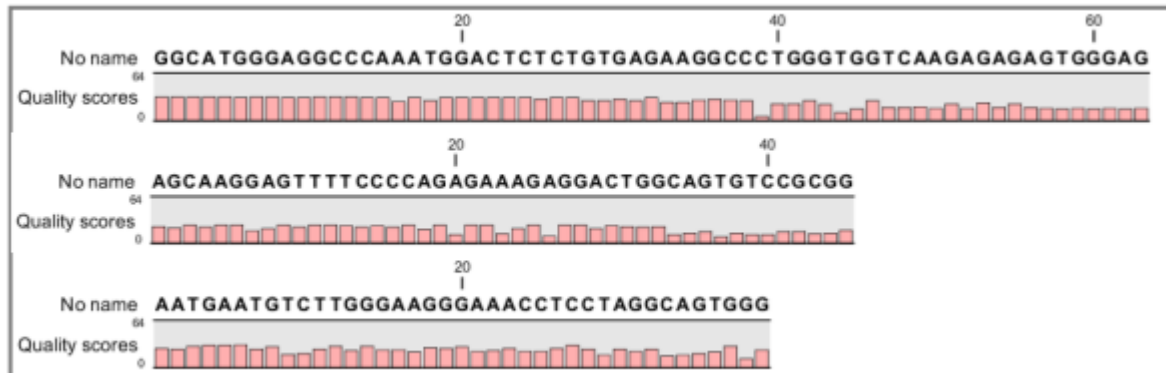


バイサルファイト処理済みリードでTの割合が著しく高く、Cの割合が著しく低い

リードのQC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

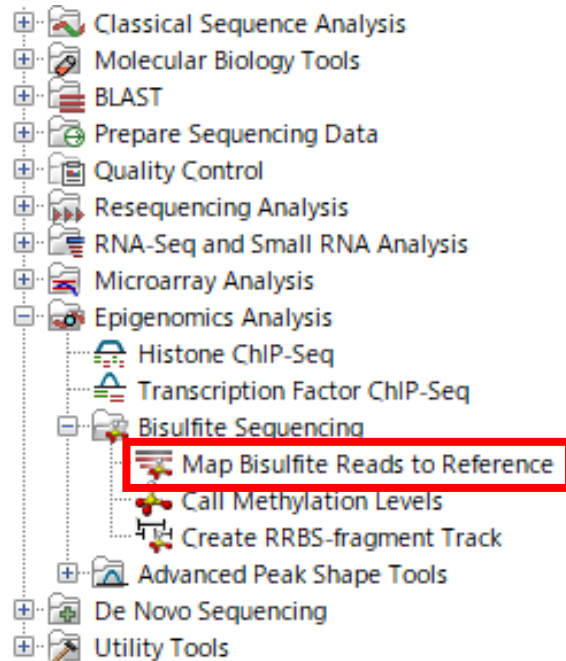
リファレンスへのマッピング

Map Bisulfite Reads to Reference

メチル化の検出

Call Methylation Levels

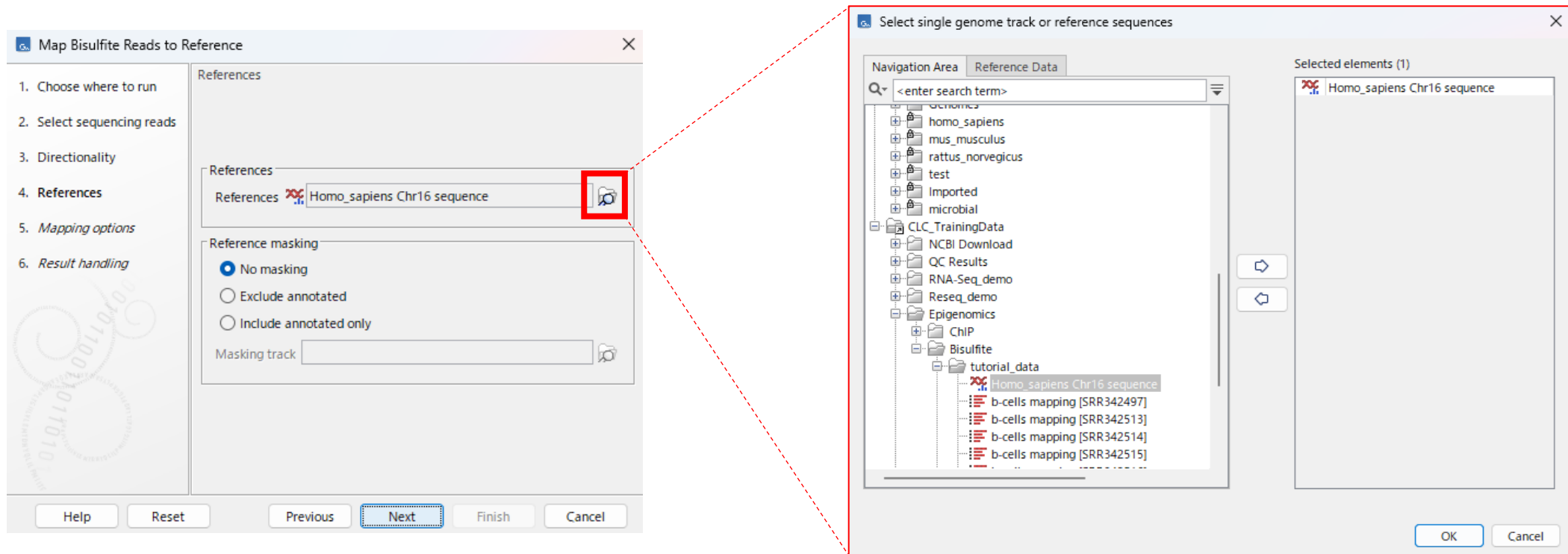
- ✓ 参照ゲノム上にバイサルファイト処理済みリードのマッピングを行う
- ✓ コントロール (Input) のデータがある場合は、それについてもマッピングを行う



Map Bisulfite Reads to Reference

参照ゲノム配列に対して、バイサルファイト処理済みリードのマッピングを行う

- ✓ 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



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

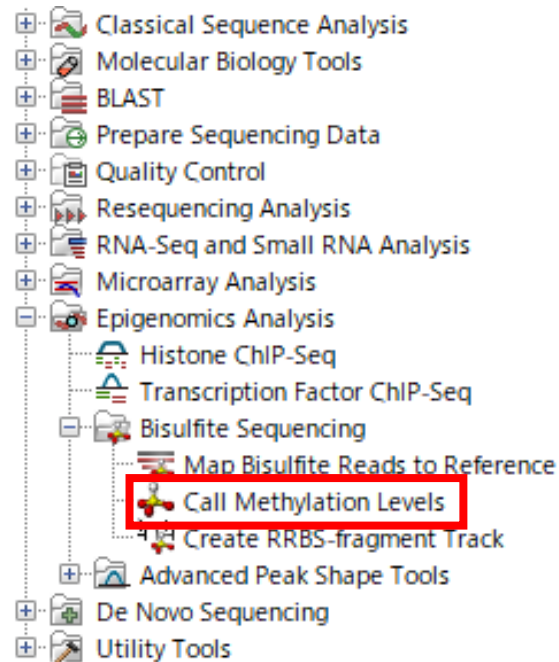
リファレンスへのマッピング

Map Bisulfite Reads to Reference

メチル化の検出

Call Methylation Levels

- ✓ マッピングデータ上のC→Tの変換が生じている箇所をたよりに、メチル化部位の検出を行う。



Call Methylation Levels

マッピングデータからメチル化部位を検出

- ✓ メチル化の検出に使用するリードのフィルターおよび検出したいメチル化のコンテキストを設定する

Call Methylation Levels

1. Choose where to run

2. Select bisulfite Reads Tracks

3. Methylation call settings

4. Statistical tests and thresholds settings

5. Result handling

Methylation call settings

Read filter

- Ignore non-specific matches
- Ignore duplicate matches
- Ignore broken pairs

Read 1 soft clip 0

Read 2 soft clip 0

Methylation detection

Methylation context group: Standard: CpG only

- Confirm methylation-contexts in
- Minimum strand-specific coverage
- Restrict calling to target regions

Methylation reporting

- Report unmethylated cytosines

Help Reset Previous Next Finish Cancel

検出するメチル化の種類（コンテキスト）を選択

Standard

- CpGのみ
 - CpG + CHG + CHH
- * H = A/C/T

NOMe-seq

- GCH + HCG
- GCH + HCG + GCG

Exhaustive（コンテキストに依らず検出）

- ✓ サンプル間比較を行う場合は、Statistic testにて計算手法とP-valueの閾値、さらにコントロールサンプルのマッピングデータの指定を行う

The screenshot shows the 'Call Methylation Levels' window with the following settings:

- Statistical tests and thresholds settings**
 - Statistical test**
 - Statistic mode: Fisher exact
 - Maximum p-value: 0.05
 - Control Reads Track: hspc mapping [SRR342518] (Reads)
 - Window thresholds**
 - Window length: 1,000
 - Minimum number of samples: 1
 - Sample thresholds**
 - Minimum high-confidence site-coverage: 1
 - Minimum high-confidence site-count: 1
 - Maximum mean site coverage: 0.0

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

- ✓ レポートには、参照ゲノムにマッピングされたリード配列の各種カウントデータや、メチル化のバイアスが発生している部位などの情報が含まれる

- b-cells mapping [SRR342497] (Methylation levels)
- hspc mapping [SRR342518] (Methylation levels)
- Differential methylation (CG)
- b-cells mapping [SRR342497] (Reads) (Methylation-report)
- hspc mapping [SRR342518] (Reads) (Methylation-report)

2 Read counts

The table below gives an overview of sequences analysed. The column 'Single Reads' counts individual reads, i.e. single-end reads once and paired-end reads twice. The column 'Read Pairs' counts each read pair once.

Total: The total number of reads/pairs in the input data set.

Duplicate: The number of duplicate reads/pairs. A single read is called a duplicate, when its mapping coordinates are identical to those of another single read. A read pair is called duplicate, when the mapping coordinates of the fragment (the outer mapping coordinates of both reads) are identical to another read pair.

Non-specific: Reads/pairs that had more than one optimal mapping.

From broken pair: Single reads that were mapped as single reads, but originated from a read pair.

Included in analysis: How many reads/pairs were included in the methylation call analysis after filtering of aforementioned reads/pairs (given user-defined filter settings).

Mapped to CT-converted reference: The number of reads/pairs that were mapped to the CT-conversion of the reference.

Mapped to GA-converted reference: The number of reads/pairs that were mapped to the GA-conversion of the reference.

	Single Reads	Read Pairs
Total:	103,804	50,588
Duplicate:	5,805	2,890
Non-specific:	29,180	14,210
From broken pair:	1,843	
Included in analysis:	66,976	33,488
Mapped to CT-converted reference:	33,354	16,677
- as CT-converted read:	16,677	
- as GA-converted read:	16,677	
Mapped to GA-converted reference:	33,622	16,811
- as CT-converted read:	16,811	
- as GA-converted read:	16,811	

2.1 Directional counts

This section provides details of the mapping of reads, both conversion and direction.

The following table shows read conversion and reference conversion for each mapped read. For paired reads, both R1 and R2 are tabulated separately, but note that the direction of each read refers only to the direction of R1 of the pair. The table is most useful for diagnosing whether reads that have been mapped as non-directional should actually have been mapped as directional. If this is the case, there will be few counts in the cells marked 'non-directional only' in the explanatory table below. Forward and reverse is the direction of the read, for paired reads the direction of R1. There should only be entries in the cells CT/CT Reverse, CT/GA Forward, GA/GA Forward, and GA/CT Reverse for non-directed sequencing.

Read/Reference	Forward (R1)	Reverse (R1)
CT/CT	16,660	17
CT/GA	26	16,785
GA/GA	26	16,785
GA/CT	16,660	17

- b-cells mapping [SRR342497] (Methylation levels)
- hspc mapping [SRR342518] (Methylation levels)
- Differential methylation (CG)
- b-cells mapping [SRR342497] (Reads) (Methylation-report)
- hspc mapping [SRR342518] (Reads) (Methylation-report)

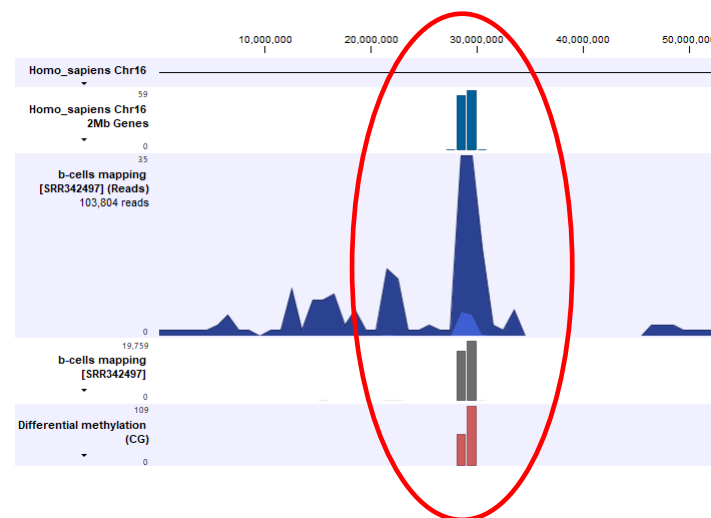
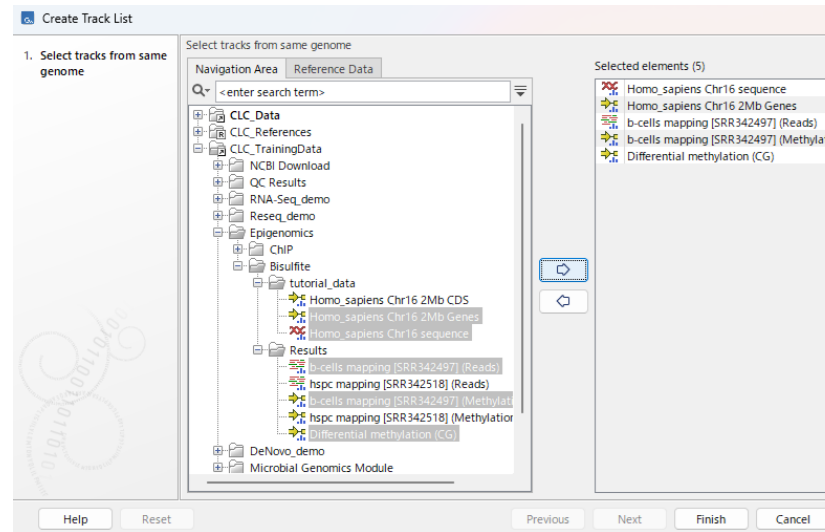
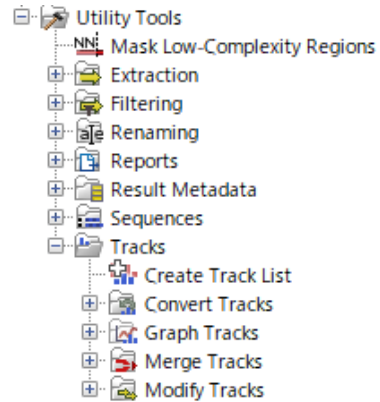
Differential methylationデータ

Chromosome	Region	Cytosines	Case sampl...	Case covera...	Case covera...	Case methyl...	Case methyl...	Control sam...	Control cov...	Control cov...	Control met...	Control met...	p-value
16	28007001..28008000	11	1	23	2.09	18	0.78	1	17	1.55	8	0.47	0.04
16	28021001..28022000	34	1	80	2.35	61	0.76	1	36	1.06	21	0.58	0.04
16	28022001..28023000	43	1	109	2.53	76	0.70	1	40	0.93	16	0.40	9.94E-4
16	28059001..28060000	8	1	16	2.00	14	0.88	1	4	0.50	1	0.25	0.03
16	28072001..28073000	28	1	57	2.04	49	0.86	1	79	2.82	51	0.65	4.03E-3
16	28087001..28088000	31	1	69	2.23	63	0.91	1	24	0.77	16	0.67	7.06E-3
16	28093001..28094000	49	1	88	1.80	84	0.95	1	75	1.53	59	0.79	1.10E-3
16	28102001..28103000	41	1	90	2.20	84	0.93	1	37	0.90	30	0.81	0.04

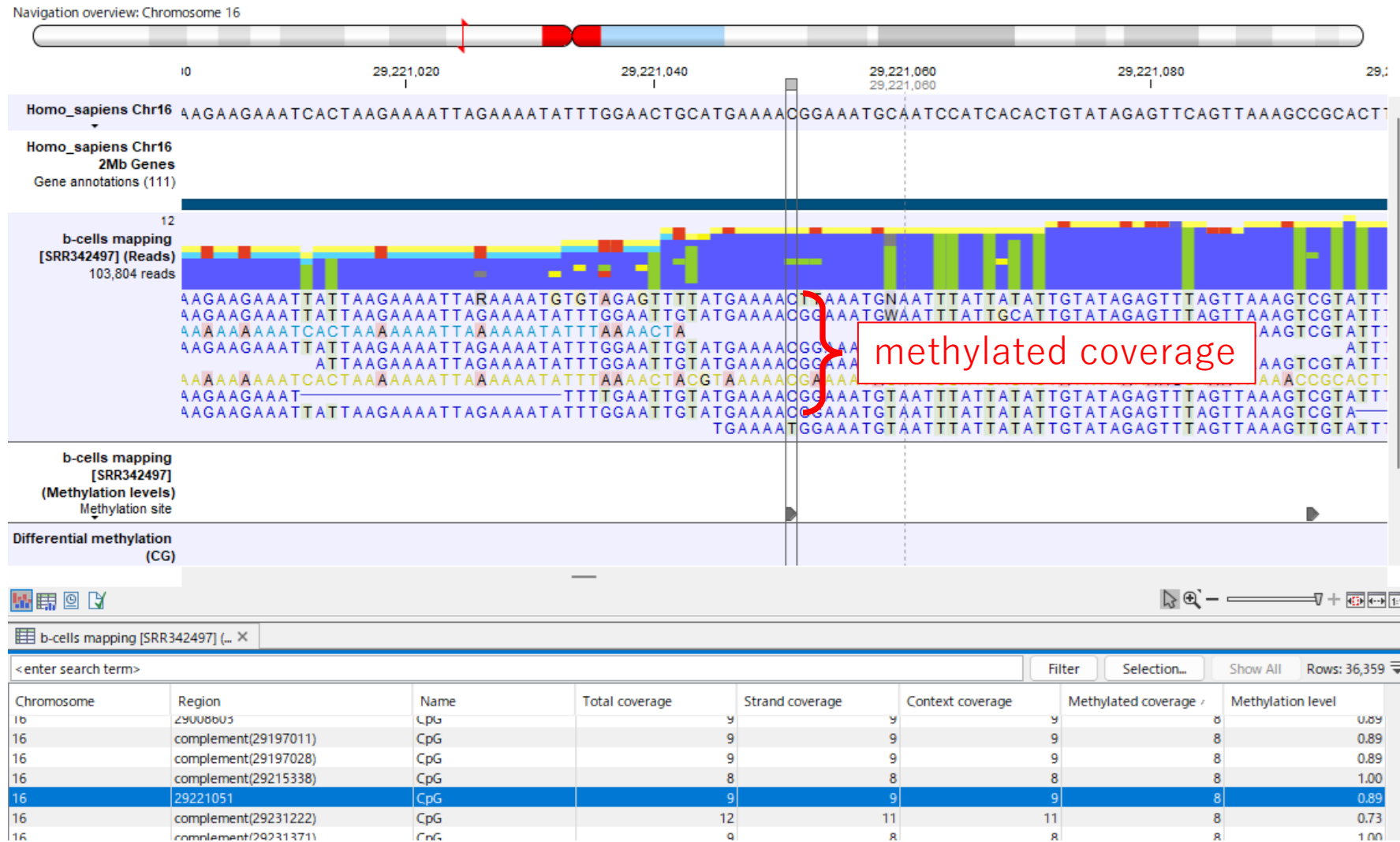
サンプルごとのMethylation levelsデータ

Chromosome	Region	Name	Total coverage	Strand coverage	Context coverage	Methylated coverage	Methylation level
16	28005482	CpG	4	2	2	2	1.00
16	complement(28005483)	CpG	3	1	1	1	1.00
16	28005504	CpG	2	1	1	1	1.00
16	complement(28005505)	CpG	2	1	1	1	1.00
16	complement(28005897)	CpG	6	6	6	5	0.83
16	complement(28005944)	CpG	3	3	3	3	1.00
16	28006150	CpG	5	1	1	1	1.00
16	complement(28006151)	CpG	5	4	4	3	0.75

✓ マッピング結果や、ピークの位置の遺伝子の位置関係を目視で確認できる



- ✓ Methylation levelsの表とリンクさせることも可能



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