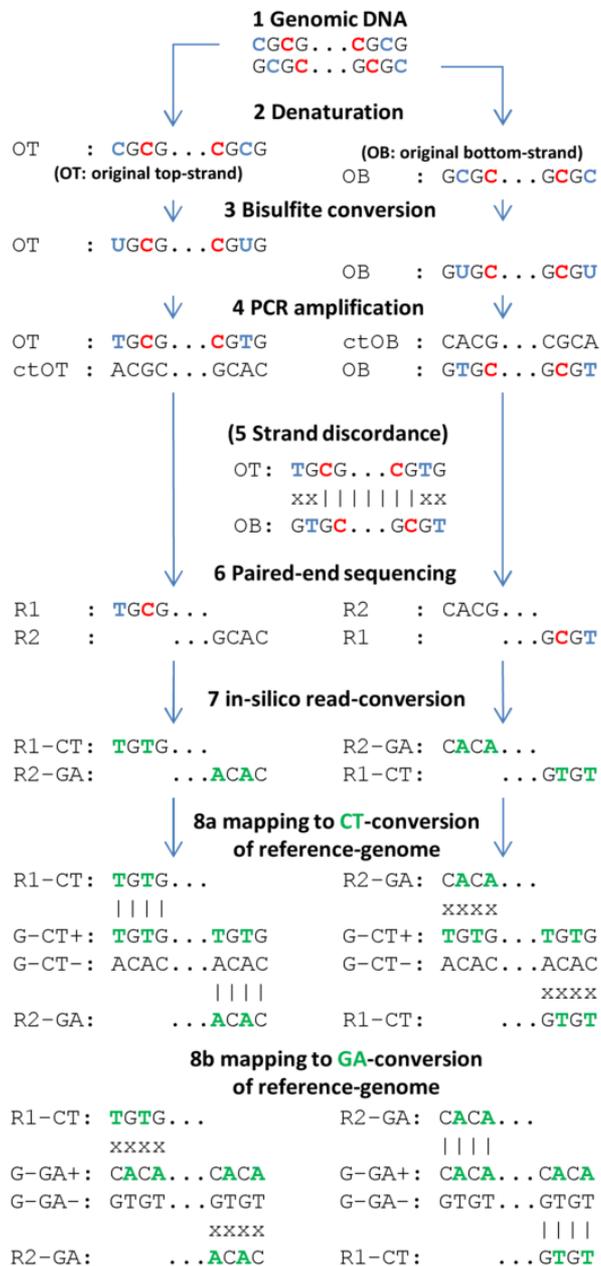


NGSデータ解析入門Webセミナー： バイサルファイトシーケンス編

バイサルファイトシーケンスの手順



- ゲノムDNA抽出、フラグメント化、エンドリペア、サイズ選択、アダプター付加
- 1本鎖に分離
- バイサルファイト処理
- PCR増幅
- 塩基配列の不一致の確認
- ペアエンドシーケンス
- ペアの片方のリードをC -> T変換し、もう片方をG -> A変換
- C -> TまたはG -> A変換された参照ゲノム配列に、リードをマッピング

✔	✘
✔	✘
✘	✔

赤 : メチル化シトシン
青 : 非メチル化シトシン、ウラシル、チミン
緑 : in-silico変換後の塩基

Directional protocol:

ペアエンドシーケンスを行ったリード1が original-top (OT)またはoriginal-bottom (OB)で、リード2が相補リード (ctOTまたはctOB)

キット例:

- QIAseq Methyl Library Kit
- Illumina TruSeq DNA Methylation Kit (formerly EpiGnome)
- Kits from the NuGen Ovation family of products
- Swift Accel-NGS Methyl-seq DNA Library Kit
- Libraries prepared by the 'Lister' method

Non-directional protocol:

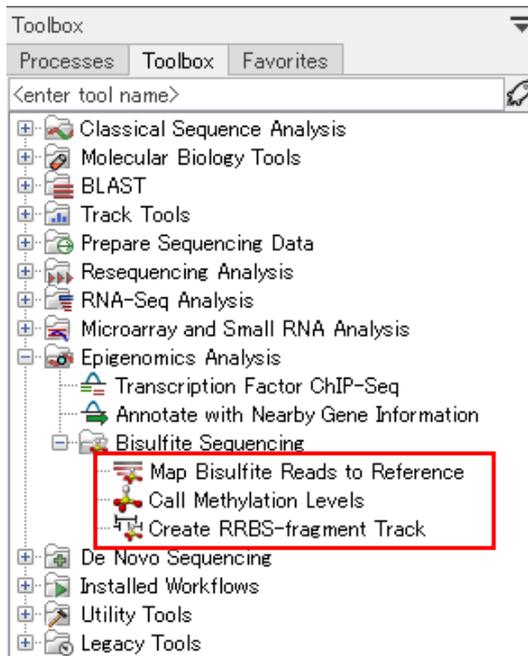
ペアエンドシーケンスを行ったリード1が OT、OB、ctOT、ctOBのいずれか

キット例:

- Zymo Pico Methyl-Seq Library Kit
- Bioo Scientific (Perkin Elmer) NEXTFlex Bisulfite-Seq Kit
- Libraries prepared by the 'Cokus' method

✓ 使用するプロトコールによって、鎖の方向を区別しているものとしていないものがある

- ✓ 塩基のin-silico変換は、ソフトウェアが自動で行うため、解析に用いるリード配列データと参照ゲノム配列データは、標準的な変換前のものを使用する
- ✓ 各ツールによる出力データのフォーマットは、その他のアプリケーションと同じトラックフォーマットのため、Track Toolsによるゲノムブラウザ表示など、他アプリケーションと共通の機能も使用できる



Map Bisulfite Reads to Reference

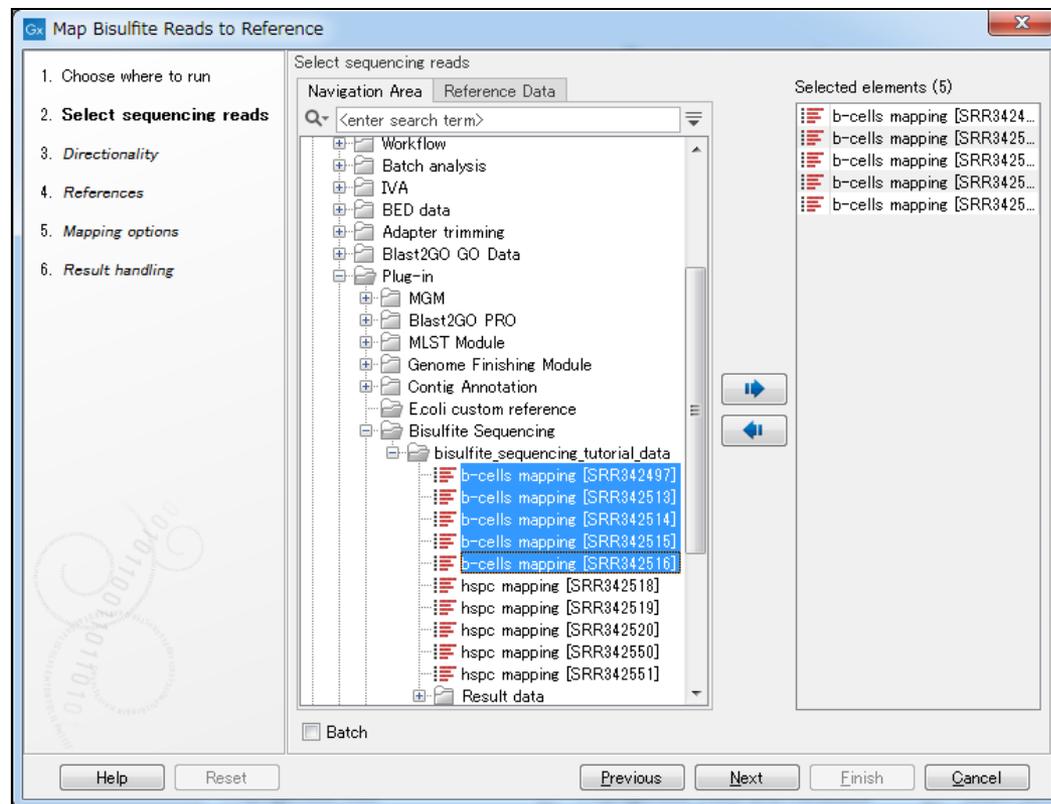
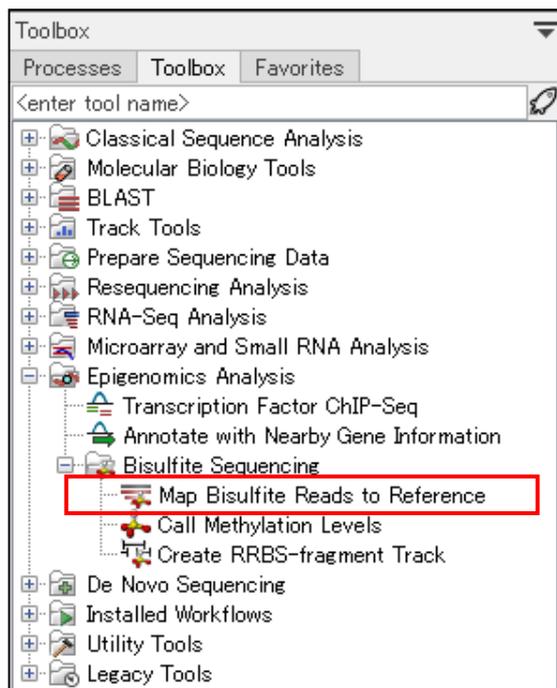
- 任意の参照ゲノム配列に対して、バイサルファイトシーケンスデータのマッピングを行う

Call Methylation Levels

- メチル化の検出、およびサンプル間比較を行う

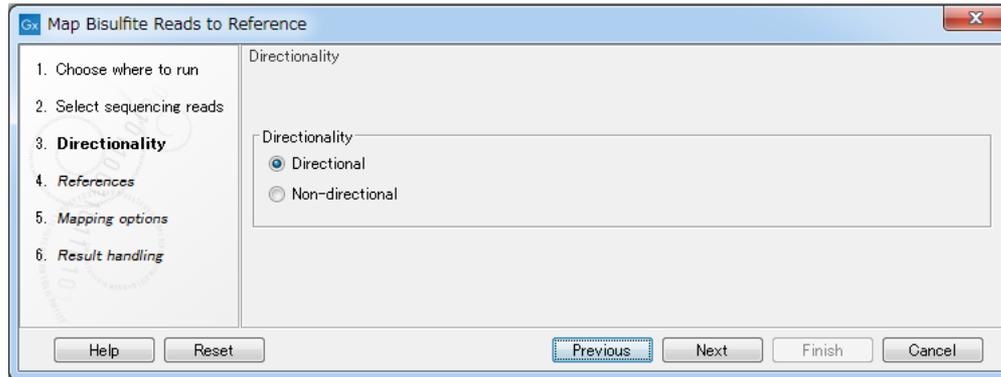
Create RRBS-fragment Track

- 参照ゲノム配列より、制限酵素処理によって生成する配列断片のトラックの作成を行う



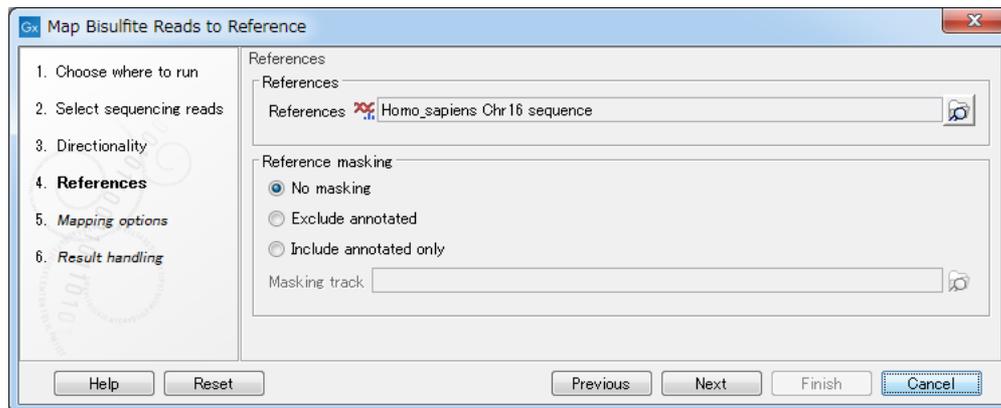
1. Map Bisulfite Reads to Referenceを選択し、ダブルクリック
2. リードデータを選択

Map Bisulfite Reads to Reference



Directionality:

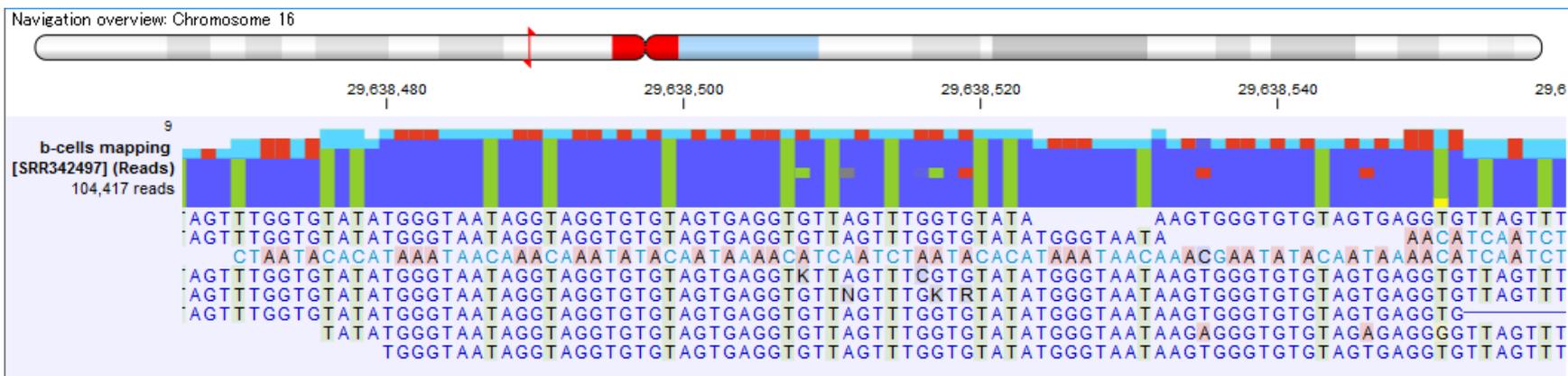
- シークエンスを、DirectionalとNon-directionalのどちらのプロトコールで実施したかを選択する



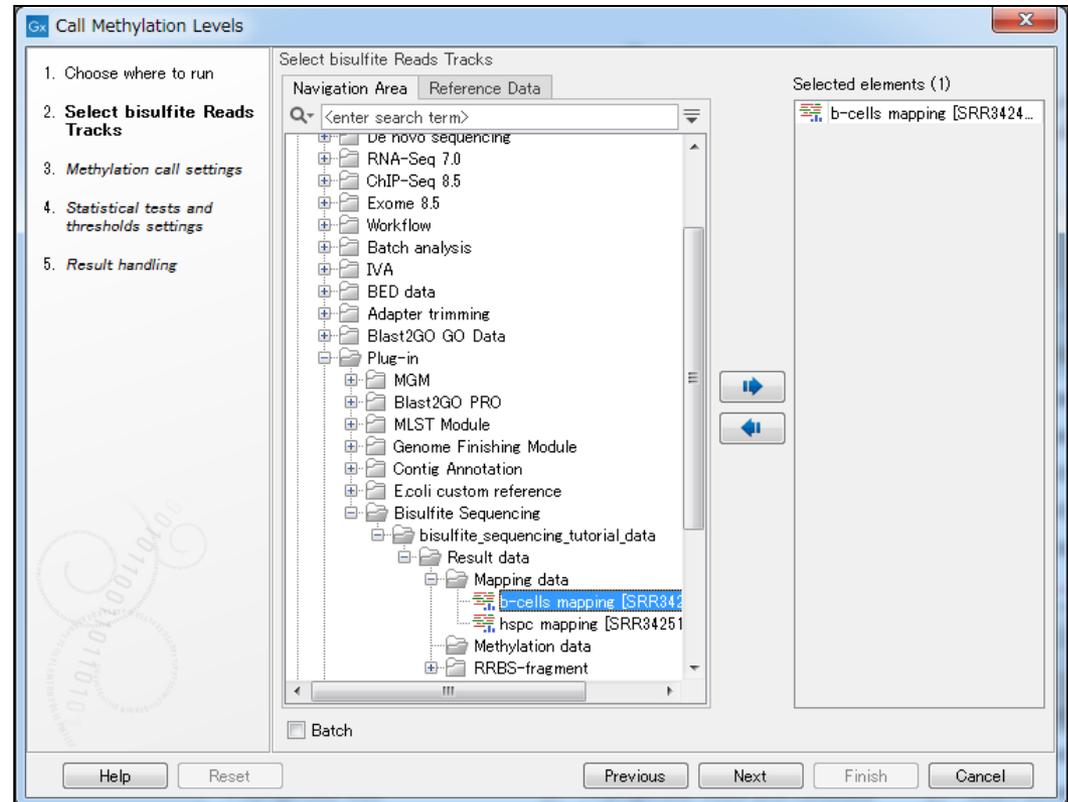
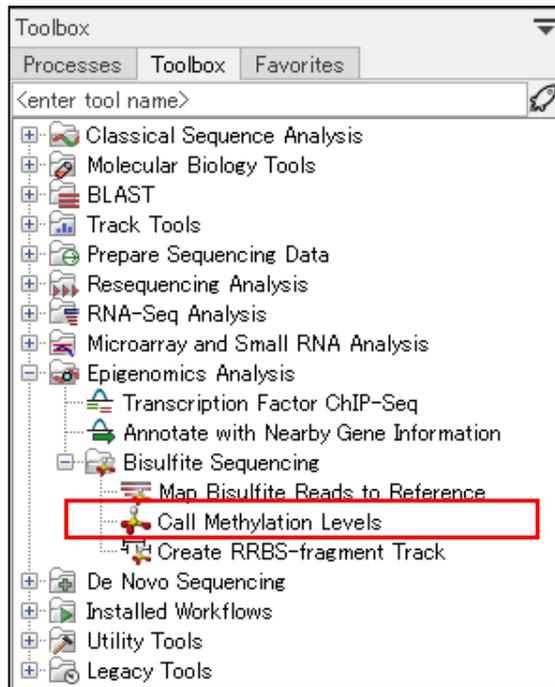
References:

- リファレンスとなるゲノム配列データを指定する

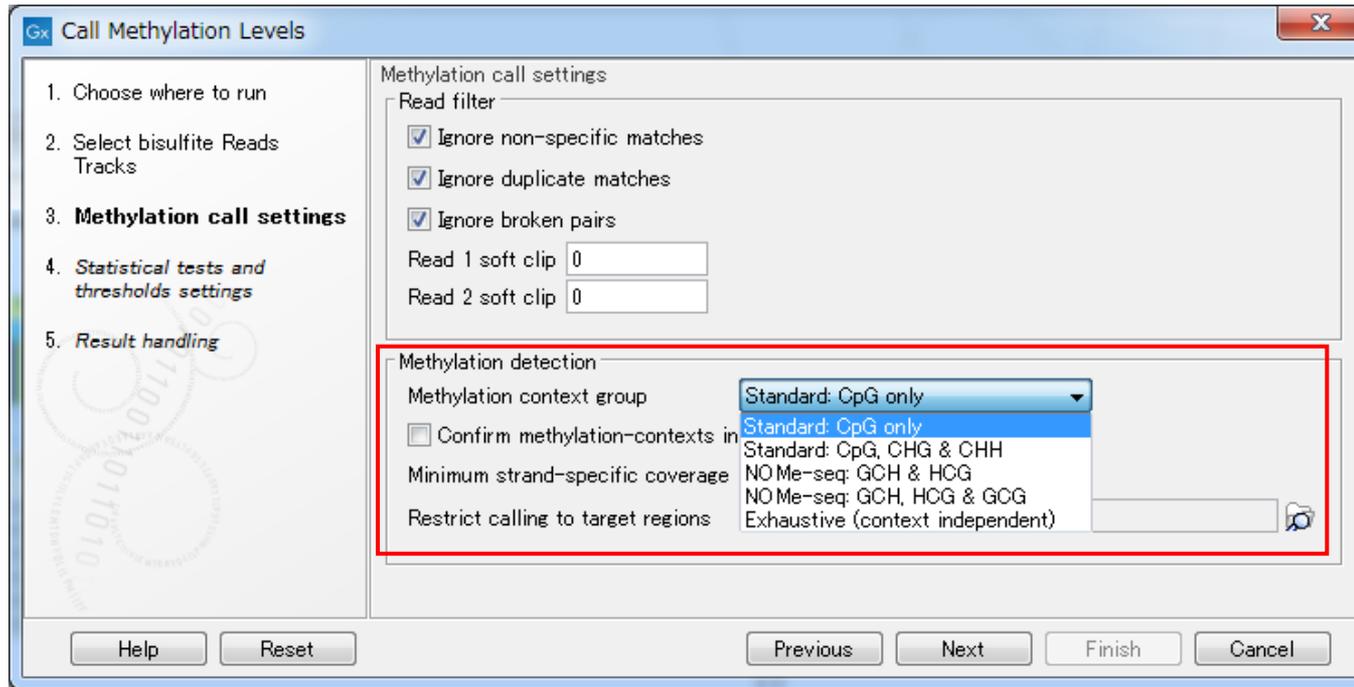
3. 使用した実験プロトコールに合うDirectionalityを選択し、加えて参照ゲノム配列データの指定を行う



4. バイサルファイトシーケンスリードのマッピングデータが得られる



1. Call Methylation Levelsを選択し、ダブルクリック
2. マッピングデータを選択



3. 参照ゲノム配列上のメチル化検出部位などを指定する

Standard:

- ✓ CpG Detects 5-methylated cytosines in CpG contexts
- ✓ CHG Detects 5-methylated cytosines in CHG contexts (H = A/C/T)
- ✓ CHH Detects 5-methylated cytosines in CHH contexts

NOMe-seq:

- ✓ GCH Detects enzymatic methylation in GCH contexts
- ✓ HCG Detects endogenous methylation in HCG contexts
- ✓ GCG Detects ambiguous methylation in GCG contexts

Call Methylation Levels

Statistical tests and thresholds settings

Statistical test

Statistic mode: Fisher exa...

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

4. 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

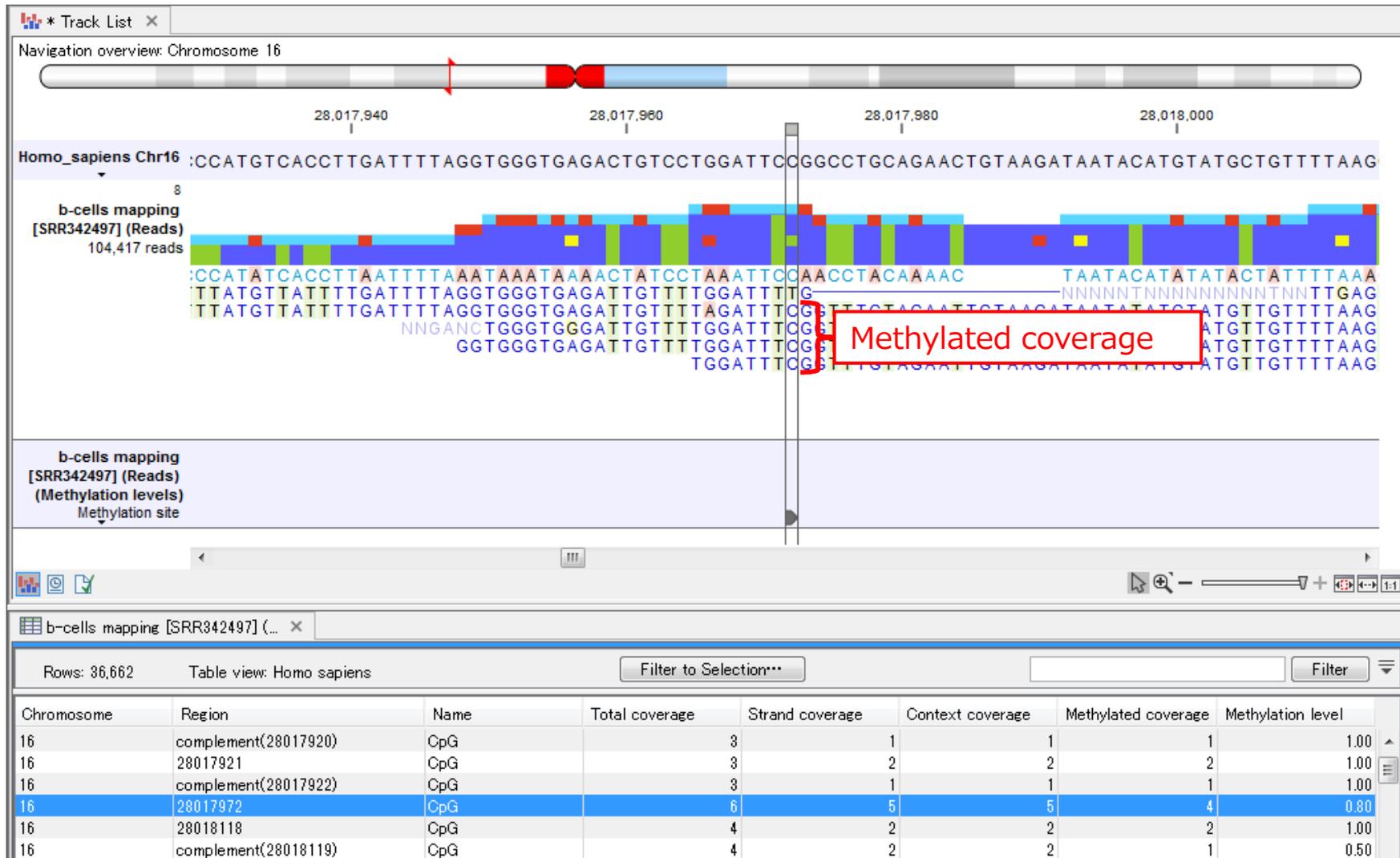
- Methylation data
 - Differential methylation (CG)
 - b-cells mapping [SRR342497] (Reads) (Methylation levels)
 - hspc mapping [SRR342518] (Reads) (Methylation levels)
 - b-cells mapping [SRR342497] (Reads) (Methylation-report)
 - hspc mapping [SRR342518] (Reads) (Methylation-report)

Differential methylationデータ

Chromosome	Region	Name	Cytosines	Case sampl...	Case covera...	Case covera...	Case methyl...	Case methyl...	Control sam...	Control cov...	Control cov...	Control met...	Control met...	p-value
16	28021001.28022000		34	1	80	2.35	61	0.76	1	34	1.00	19	0.56	0.03
16	28022001.28023000		44	1	121	2.75	82	0.68	1	41	0.93	17	0.41	2.73E-3
16	28059001.28060000		9	1	16	1.78	14	0.88	1	5	0.56	1	0.20	0.01
16	28072001.28073000		28	1	59	2.11	50	0.85	1	79	2.82	51	0.65	6.28E-3
16	28087001.28088000		31	1	69	2.23	63	0.91	1	25	0.81	16	0.64	3.07E-3
16	28093001.28094000		49	1	88	1.80	84	0.95	1	77	1.57	61	0.79	1.37E-3
16	28102001.28103000		41	1	91	2.22	85	0.93	1	35	0.85	28	0.80	0.03
16	28104001.28105000		47	1	114	2.43	108	0.95	1	83	1.77	68	0.82	4.20E-3
16	28110001.28111000		37	1	63	1.70	62	0.98	1	28	0.76	24	0.86	0.03
16	28120001.28121000		30	1	73	2.43	59	0.81	1	45	1.50	23	0.51	7.40E-4

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	complement(28010490)	CpG		2	2	2	2	1.00
16	complement(28010504)	CpG		3	3	3	2	0.67
16	28010534	CpG		4	1	1	1	1.00
16	complement(28010535)	CpG		4	3	3	3	1.00
16	complement(28010810)	CpG		7	7	7	6	0.86
16	28010928	CpG		6	2	2	2	1.00
16	complement(28010929)	CpG		6	4	4	4	1.00
16	complement(28011414)	CpG		1	1	1	1	1.00
16	complement(28011489)	CpG		1	1	1	1	1.00
16	28011568	CpG		1	1	1	1	1.00
16	28011617	CpG		3	2	2	2	1.00
16	complement(28011618)	CpG		3	1	1	1	1.00
16	28011707	CpG		4	3	3	2	0.67
16	complement(28011708)	CpG		4	1	1	1	1.00



- Regionが「complement」のものは、original-bottom (OB)由来となり、Aによるミスマッチが多数 (G -> Aのin silico変換のため) のリードとなる



- パラメータ設定したWindowサイズ（あるいはRRBSフラグメント）ごとに、サンプル間のメチル化の比較を行い、メチル化の有意差がある領域を決定する

- ✓ Output optionsの設定により、メチル化データに加えて、サンプルごとのレポート出力も行うことができる
- ✓ 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、リード配列上で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
- ✓ レポート内容を確認後、必要に応じてパラメータ設定を変更して、メチル化検出を再度行う

1 Summary

Software:	CLC Genomics Workbench 12.0.2
Creation date:	Thu Aug 15 11:29:44 JST 2019
Generated by:	Ozawa
Based upon:	b-cells mapping [SRR342497] (Reads)

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).

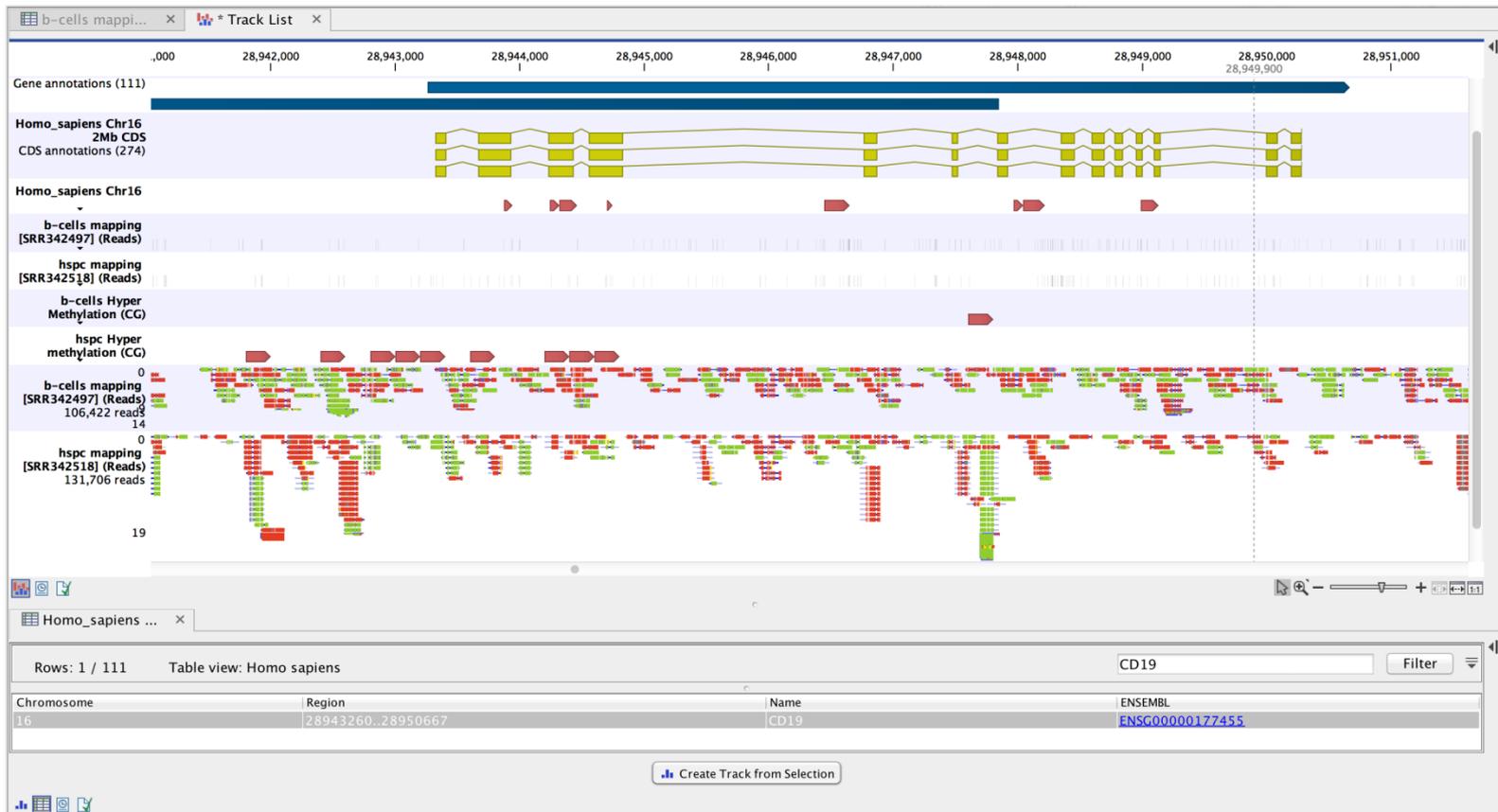
Mapped to CT-converted reference: The number of reads/pairs that were mapped to the CT-conversion of the reference.

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:	104,417	51,373
Duplicate:	5,989	2,988
Non-specific:	29,226	14,373
From broken pair:	1,178	
Included in analysis:	68,024	34,012
Mapped to CT-converted reference:	33,924	16,962
- as CT-converted read:	16,962	
- as GA-converted read:	16,962	
Mapped to GA-converted reference:	34,100	17,050
- as CT-converted read:	17,050	
- as GA-converted read:	17,050	

Track Listによる視覚表示

- ✓ マッピングデータやメチル化テーブルなどの各種トラックは、Track List機能を用いることにより、ゲノムブラウザ上に視覚表示が可能
- ✓ RRBSフラグメントトラックもTrack Listに含めることにより、参照ゲノム上のターゲット領域などを用意を確認できる



お問い合わせ先: フィルジェン株式会社

TEL 052-624-4388 (9:00~18:00)

FAX 052-624-4389

E-mail: biosupport@filgen.jp