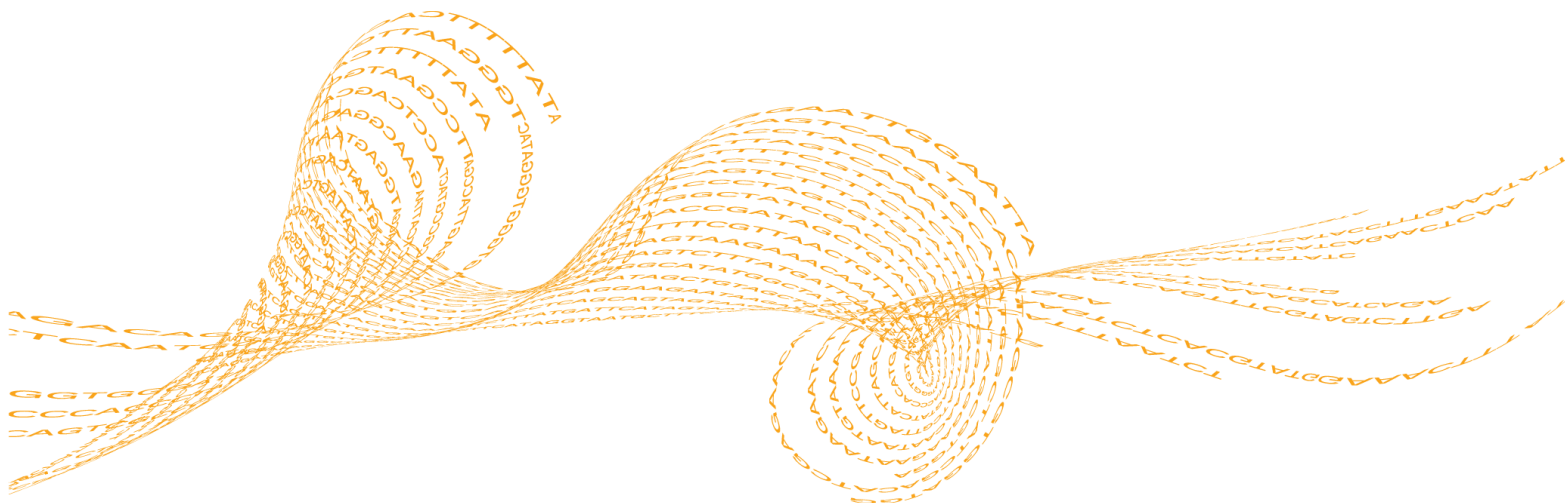


MethylSeq v1.0

App Guide

For Research Use Only. Not for use in diagnostic procedures.

Introduction	3
Run MethylSeq v1.0	5
MethylSeq v1.0 Output	6
MethylSeq Methods	6
Technical Assistance	18



This document and its contents are proprietary to Illumina, Inc. and its affiliates ("Illumina"), and are intended solely for the contractual use of its customer in connection with the use of the product(s) described herein and for no other purpose. This document and its contents shall not be used or distributed for any other purpose and/or otherwise communicated, disclosed, or reproduced in any way whatsoever without the prior written consent of Illumina. Illumina does not convey any license under its patent, trademark, copyright, or common-law rights nor similar rights of any third parties by this document.

The instructions in this document must be strictly and explicitly followed by qualified and properly trained personnel in order to ensure the proper and safe use of the product(s) described herein. All of the contents of this document must be fully read and understood prior to using such product(s).

FAILURE TO COMPLETELY READ AND EXPLICITLY FOLLOW ALL OF THE INSTRUCTIONS CONTAINED HEREIN MAY RESULT IN DAMAGE TO THE PRODUCT(S), INJURY TO PERSONS, INCLUDING TO USERS OR OTHERS, AND DAMAGE TO OTHER PROPERTY.

ILLUMINA DOES NOT ASSUME ANY LIABILITY ARISING OUT OF THE IMPROPER USE OF THE PRODUCT(S) DESCRIBED HEREIN (INCLUDING PARTS THEREOF OR SOFTWARE).

© 2015 Illumina, Inc. All rights reserved.

Illumina, 24sure, BaseSpace, BeadArray, BlueFish, BlueFuse, BlueGnome, cBot, CSPro, CytoChip, DesignStudio, Epicentre, GAllx, Genetic Energy, Genome Analyzer, GenomeStudio, GoldenGate, HiScan, HiSeq, HiSeq X, Infinium, iScan, iSelect, MiSeq, NeoPrep, Nextera, NextBio, NextSeq, Powered by Illumina, SeqMonitor, SureMDA, TruGenome, TruSeq, TruSight, Understand Your Genome, UYG, VeraCode, verifi, VeriSeq, the pumpkin orange color, and the streaming bases design are trademarks of Illumina, Inc. and/or its affiliate(s) in the U.S. and/or other countries. All other names, logos, and other trademarks are the property of their respective owners.

Introduction

The BaseSpace® app MethylSeq v1.0 analyzes DNA that has been sequenced using the TruSeq DNA Methylation Kit. Sequencing-based DNA methylation analysis applies the coverage density and flexibility enabled by next-generation sequencing to enhance epigenetic studies.

The process of bisulfite treatment denatures genomic DNA into single-stranded DNA (ssDNA). The TruSeq DNA Methylation Kit converts bisulfite-treated, ssDNA into an Illumina sequencing library. All ssDNA fragments are captured during the library prep procedure, eliminating sample loss associated with other methods.

The core algorithm used in MethylSeq v1.0 is Bismark, which maps bisulfite-treated sequencing reads to the genome of interest and performs methylation calls.

The alignment method used in MethylSeq v1.0 is Bowtie2. Bowtie2 is an ultrafast and memory-efficient tool for aligning sequencing reads to long reference sequences.

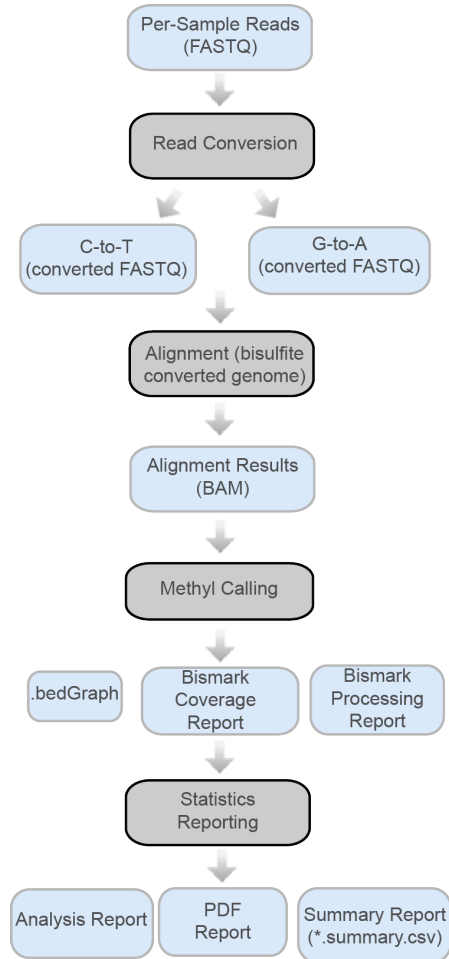
The MethylSeq v1.0 app generates the following output files:

- ▶ BAM files, which contain the reads after alignment.
- ▶ Bismark Processing Report, which contains a processing report generated by Bismark.
- ▶ Cytosine Report (optional), which contains methylation status for every cytosine in the genome, including both strands.
- ▶ bedGraph, which contains a cytosine methylation status report for only the cytosines that have sequencing coverage.

A summary page and other reports are available in multiple formats.

See *MethylSeq v1.0 Output* on page 6 and *MethylSeq Methods* on page 18.

Figure 1 MethylSeq v1.0 Analysis with Bismark



Versions

The following module versions are used in the MethylSeq v1.0 app:

- ▶ Bismark—v0.12.2
- ▶ Bowtie2—v2.2.2
- ▶ Isis (analysis software)—v2.5.61.0
- ▶ SAMtools—v0.1.19-isis-1.0.3
- ▶ bgzip
- ▶ tabix

Current Limitations

Before you run the MethylSeq v1.0 app, note the following limitations:

- ▶ hg19 reference only
- ▶ Methylation kits other than TruSeq DNA Methylation Kit are not supported
- ▶ Read length of 50–500 bp
- ▶ Maximum sample size of 200 gigabases
- ▶ No minimum number of reads; use a reasonable input size to get required coverage
- ▶ Only 1 sample per analysis

Run MethylSeq v1.0

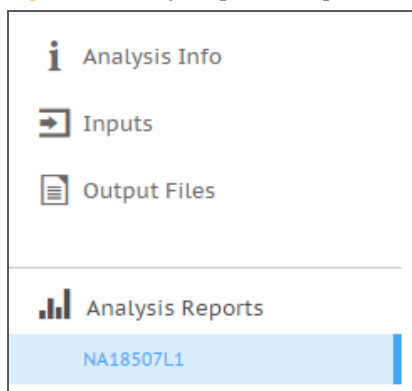
- 1 Navigate to the project or sample that you want to analyze.
- 2 Click the **Apps** tab and select **MethylSeq v1.0**.
- 3 Click **Launch** to open the app.
- 4 Enter the required fields in the MethylSeq v1.0 input form:
 - **Analysis Name**—Provide the analysis name. Default name is the app name with the date and time analysis was started.
 - **Save Results To**—Select the project that stores the app results.
 - **Sample**—Browse to the sample you want to analyze, and select the checkbox. You can analyze 1 sample per analysis.
 - **Library prep kit (human only) is directional (2 strands) or non-directional (4 strands)**—Select the directional or non-directional option. The TruSeq DNA Methylation Kit is directional.
- 5 [Optional] Click the arrow to expand the Advanced options, and provide parameters for Trim Options during the FASTQ trimming stage.
 - **Number of 5' bases to crop (after adapter trimming)**—Select the value at which the app crops a fixed number of 5' bases of each read.
 - **Minimum 5' base quality**—Select the minimum 5' base quality value. The app removes all 5' bases with quality scores less than this value.
 - **Minimum 3' base quality**—Select the minimum 3' base quality value. The app removes all 3' bases with quality scores less than this value.
- 6 [Optional] Enter the advanced fields for Methyl-Alignment Options.
 - **Flag and Remove PCR Duplicates**—When selected, PCR duplicates are flagged in the BAM files and not used for methyl calling. PCR duplicates are defined as 2 clusters from a paired-end run where both clusters have the exact same alignment positions for each read. Not applicable for single-end samples.
- 7 [Optional] Enter the advanced fields for Methyl-Call Options during the methylation calling stage.
 - **Remove duplicate methylation calls from overlapping regions of one pair of reads (two ends)**—Select to avoid scoring overlapping methylation calls on paired-end reads twice.
This option removes a bias towards more methylation calls towards the center of sequenced fragments.
 - **Number of 5' bases (post-trimming) whose methylation calls will be ignored**—The bismark_methylation_extractor ignores the number of 5' bases selected.
 - **Enable genome-wide cytosine methylation report for all cytosines in the genome**—Select to generate a Cytosine Report. The Cytosine Report contains methylation status for every cytosine in the genome, including both strands. For more information, see *Cytosine Report (*.CX_report.txt.gz)* on page 9.
- 8 Click **Continue**.

The MethylSeq v1.0 app begins analysis of your sample. When completed, the status of the app session is automatically updated, and you receive an email.

MethylSeq v1.0 Output

To view the results, click the **Projects** tab, then the project name, and then the analysis.

Figure 2 MethylSeq v1.0 Output Navigation Bar



When the analysis is completed, you can access your output through the left navigation bar.

- ▶ **Analysis Info**—Overview of the app session settings. For more information, see *Analysis Info* on page 6.
- ▶ **Inputs**—Overview of input settings. For more information, see *Inputs* on page 7.
- ▶ **Output Files**—Access to the output files for the sample. For more information, see *Output Files* on page 7.
- ▶ **Analysis Reports**—Access to an analysis report for the sample. For more information, see *Analysis Reports* on page 15.

Analysis Info

This app provides an overview of the analysis on the Analysis Info page.

A brief description of the metrics is below.

Row	Definition
Name	Name of the app session.
Application	App that generated this analysis.
Date Started	Date and time the app session started.
Date Completed	Date and time the app session completed.
Duration	Duration of analysis.
Session Type	The number of nodes used.
Size	Total size of all output files.
Status	Status of the app session.

Log Files

Click the **Log Files** link on the Analysis Info page to access the app log files. Log files may also be located in a folder in the Output Files section.

The key log files to help follow data processing and debugging are the following:

- ▶ **bsfs-{timestamp}.log**—File logging information pertaining to the BaseSpace file system.
- ▶ **CompletedJobInfo.xml**—Contains information about the completed job.
- ▶ **Logging.zip**—Contains all detailed workflow log files for each step of the workflow.
- ▶ **MethylSeqRunStatistics.xml**—Provides methylation-related information.
- ▶ **metrics-{timestamp}.log**—Internal application log file not for general customer usage.
- ▶ **output-{timestamp}.log**—Shows the raw console output from the app.
- ▶ **SampleSheet.csv**—Sample sheet.
- ▶ **SampleSheetUsed.csv**—A copy of the sample sheet, generated at the end of a run.
- ▶ **spacedock-{timestamp}.log**—Internal application log file not for general customer usage.
- ▶ **spacedock-infrastructure-{timestamp}.log**—Internal application log file not for general customer usage.
- ▶ **uploader-{timestamp}.log**—Internal application log file not for general customer usage.
- ▶ **WorkflowError.txt**—Workflow standard error output (contains error messages created while running the workflow).
- ▶ **WorkflowLog.txt**—Workflow standard output (contains details about workflow steps, command line calls with parameters, timing, and progress).

MethylSeq v1.0 Status

The status of the MethylSeq v1.0 app session can have the following values:

- Launching Isis
- TrimReads
- MethylAlignment
- SortAndMergeAlignment
- MethylCalling
- Statistics evaluation
- Report generation

Inputs

The Inputs page provides an overview of the input samples and settings that were specified when the MethylSeq v1.0 analysis was set up.

Output Files

The Output Files page provides access to the sample output files.

BAM Files

The Sequence Alignment/Map (SAM) format is a generic alignment format for storing read alignments against reference sequences, supporting short and long reads (up to 128 Mb) produced by different sequencing platforms. SAM is a text format file that is human-readable. The Binary Alignment/Map (BAM) keeps the same information as SAM, but in a compressed, binary format that is only machine readable.

If you use an app in BaseSpace that uses BAM files as input, the app locates the file when launched. If using BAM files in other tools, download the file to use it in the external tool.

Go to samtools.sourceforge.net/SAM1.pdf to see the exact SAM specification.

BAM Splitting Report (*.bam_splitting_report.txt)

The MethylSeq v1.0 app provides a BAM Splitting Report in a text (TXT) file format (*.bam_splitting_report.txt). The BAM Splitting Report contains a list of chromosomes analyzed, the parameters used to extract methylation information, and the total number of methylation call strings processed. The report also contains methylation and conversion information.

bedGraph (*.bedGraph.gz)

The MethylSeq v1.0 app provides a cytosine methylation status report in a compressed bedGraph format (*.bedGraph.GZ). This report contains only the cytosines that have sequencing coverage.

For more information on the bedGraph Track Format, visit genome.ucsc.edu/goldenpath/help/bedgraph.html.

Statistic	Definition
Chromosome	The chromosome name.
Start position	The genomic start position.
End position	The genomic end position.
Methylation Percentage	The percentage of methylation at that position.
Coverage	The total number of C bases at that position.

Bismark Processing Report

The MethylSeq v1.0 app provides a Bismark Processing Report. The report is available in the following formats:

- ▶ HTML (*.bismark.processing.report.html)
- ▶ TXT (*.bismark.processing.report.txt)

Bismark generates this analysis report. For more information, visit the Babraham Bioinformatics website at www.bioinformatics.babraham.ac.uk/projects/bismark/.

Bismark Coverage (*.bismark.cov.gz)

The MethylSeq v1.0 app provides a Bismark Coverage report in a GZIP compressed format (*.bismark.cov.gz). See *Bismark* on page 18.

Statistic	Definition
Chromosome	The chromosome name.
Start position	The genomic start position.
End position	The genomic end position.
Methylation Percentage	The percentage of methylation at that position.

Statistic	Definition
Count Methylated	The number of C bases that are methylated.
Count Unmethylated	The number of C bases that are unmethylated.

Coverage Histogram (*.CoverageHistogram.txt)

The MethylSeq v1.0 app provides a Coverage Histogram Report in a text (TXT) file format (*.CoverageHistogram.txt). This report includes the number of bases for each chromosome with a particular depth of coverage.

Cytosine Report (*.CX_report.txt.gz)

The MethylSeq v1.0 app provides a Cytosine Report in a GZIP archive file format (*.CX_report.txt.gz). The unzipped report is generated in a tab-delimited text (TXT) format.

The cytosine report contains methylation status for every cytosine in the genome, including both strands. This report is only generated when you select **Enable genome-wide cytosine methylation report for all cytosines in the genome** on the input form. See *Bismark* on page 18.

Statistic	Definition
Chromosome	The chromosome.
Position	The genomic position.
Strand	The strand.
Count methylated	The number of C bases that are methylated.
Count unmethylated	The number of C bases that are unmethylated.
C-context	CpG or alternative context.
Trinucleotide context	Trinucleotide context.

M Bias (*.M-bias.txt)

The MethylSeq v1.0 app provides an M-Bias Report in a tab-delimited text (TXT) file format (*.M-bias.txt). The M-Bias Report describes the methylation proportion across each possible position in the read. The report contains a numerical value for M-Bias plot that can indicate the presence of fundamental technical biases in the methylation calling of reads. See *Bismark* on page 18.

Statistic	Definition
Position	The position of the C base.

Statistic	Definition
Count methylated	The number of C bases that are methylated.
Count unmethylated	The number of C bases that are unmethylated.
% methylation	The percentage of methylation at that position.
Coverage	The total number of C bases at that position.

Sample Report

The MethylSeq v1.0 app provides a MethylSeq Report, containing an overview of statistics for the sample, for download. The report is available in the following formats:

- ▶ HTML (*.report.html)
- ▶ JS.HTML (*.report.js.html)
- ▶ PDF (*.report.pdf)

The report is also available in the Analysis Reports section. See *Analysis Reports* on page 15.

Sample Information

Statistic	Definition
Total PF Reads	The number of reads (2x the number of pairs for paired-end data) in the trimmed FASTQ files.
Percent Q30	The percentage of bases with a quality score of 30 or higher.

Cytosine Methylation

Provides the total number of C bases analyzed and the following metrics.

Statistic	Definition
Category	Describes whether the C base is methylated or unmethylated.
Cs in CpG	The number of C bases in aligned reads that are in the context of CpG.
Cs in CHG	The number of C bases in aligned reads that are in the context of CHG.
Cs in CHH	The number of C bases in aligned reads that are in the context of CHH.

Read and Base Alignment Statistics

Provides alignment statistics for read and base and the following metrics.

- ▶ Read Level Statistics

Statistic	Definition
Total Aligned Reads	The total number of reads present in the data set that are aligned to the reference genome.

Statistic	Definition
Percent Aligned Reads	The percentage of reads present in the data set that are aligned to the reference genome.

► Bisulfite-Treated Strand Alignment Statistics

Statistic	Definition
Top Strand	Number of aligned reads that are the original bisulfite-treated top strand.
Complementary Top Strand	Number of aligned reads that are complementary to the original bisulfite-treated top strand. When the library prep kit is directional, the value is 0.
Bottom Strand	Number of aligned reads that are the original bisulfite-treated bottom strand.
Complementary Bottom Strand	Number of aligned reads that are complementary to the original bisulfite-treated bottom strand. When the library prep kit is directional, the value is 0.

► Base Level Statistics

Statistic	Definition
Total Aligned Bases	The total number of bases present in the data set that are aligned to the reference genome.
Percent Aligned Bases	The percentage of bases present in the data set that are aligned to the reference genome.

Coverage Histogram (Mean Coverage)

Provides the mean coverage and the following metrics.

Statistic	Definition
Number of Bases at Coverage X	Number of bases that have at least the indicated depth of coverage.
Depth of Sequencing Coverage	The coverage depth of a position in the genome refers to the number of sequenced bases that align to that position.

Fragment Length Summary

Statistic	Definition
Fragment Length Median	Median length of the sequenced fragment. The fragment length is calculated based on the locations at which a read pair aligns to the reference. The read mapping information is parsed from the BAM files.
Minimum	Minimum length of the sequenced fragment.

Statistic	Definition
Maximum	Maximum length of the sequenced fragment.
Standard Deviation	Standard deviation of the sequenced fragment length.

Duplicate Information

Statistic	Definition
Percent Duplicate Paired Reads	Percentage of paired reads that have duplicates.

Summary Report (*.summary.csv)

The MethylSeq v1.0 app provides a MethylSeq Summary Report containing sample results in a comma-separated values (CSV) format (*.summary.csv). This report is an overview of statistics for the sample.

Statistic	Definition
Sample ID	IDs of samples reported on in the file.
Sample Name	Names of samples reported on in the file.
Run Folder	Run folders for samples reported on in the file.
Reference Genome	Reference genome selected.
Paired End	When true, the input sample is paired-end. When false, the input sample is single-end.
Number of 5' bases to crop during trimming	The number of 5' bases cropped from each read during the FASTQ trimming stage.
Minimum 5' base quality during trimming	All 5' bases with a quality score less than this value are removed during the FASTQ trimming stage.
Minimum 3' base quality during trimming	All 3' bases with a quality score less than this value are removed during the FASTQ trimming stage.
Input methylation library directional or not	When true, the methylation preparation library only contains reads from bisulfite-treated top and bottom strands. When false, the library contains additional reads complementary to bisulfite-treated top and bottom strand.
PCR duplicates flagged or not	When true, PCR duplicates are flagged in the final .bam files. If the input data are single-end, this setting is ignored and no duplicates are marked.
Remove duplicate methylation calls in overlapping regions of two ends	Applies to paired-end reads with an overlap for Read 1 and Read 2. When true, this option tells the app to score methylation calls in the overlap area only one time, which removes a bias towards more methylation calls towards the center of sequenced fragments. This option does not apply to single-end data.

Statistic	Definition
Number of 5' bases to ignore in methylation-calling	Number of 5' bases to ignore during the methylation calling stage.
Enable cytosine report output	When true, a Cytosine Report containing methylation status for every cytosine in the genome (including both strands) is generated, with the file name <code>SampleName_SampleNumber.CX_report.txt.gz</code> . See <i>Cytosine Report (*.CX_report.txt.gz)</i> on page 9.
Diversity	Measuring diversity of input library. It is the number of unique input DNA fragments. Because of the low complexity associated with MethylSeq data, this value might not apply.
Total PF reads	The number of reads (2x the number of pairs for paired-end data) in the trimmed FASTQ files.
Total aligned Read 1	The total number in Read 1 that are aligned to the reference genome.
Total aligned Read 2	The total number in Read 2 that are aligned to the reference genome.
Percent aligned Read 1	The percentage of Read 1 that are aligned to the reference genome.
Percent aligned Read 2	The percentage of Read 2 that are aligned to the reference genome.
Percent duplicate read pairs	Percentage of paired reads that have duplicates.
Fragment length median	Median length of the sequenced fragment. The fragment length is calculated based on the locations at which a read pair aligns to the reference. The read mapping information is parsed from the BAM files.
Fragment length min	Minimum length of the sequenced fragment.
Fragment length max	Maximum length of the sequenced fragment.
Fragment length SD	Standard deviation of the sequenced fragment length.
Total PF bases	The number of bases passing filter for the sample.
Total PF bases Read 1	The number of bases passing filter for Read 1.
Total PF bases Read 2	The number of bases passing filter for Read 2.
Percent Q30 bases Read 1	Percentage of bases from Read 1 with a quality score of 30 or higher.
Percent Q30 bases Read 2	Percentage of bases from Read 2 with a quality score of 30 or higher.
Total aligned bases	The total number of bases present in the data set that aligned to the reference genome.
Total aligned bases Read 1	The total number of bases that are aligned to the reference genome for Read 1.

Statistic	Definition
Total aligned bases Read 2	The total number of bases that are aligned to the reference genome for Read 2.
Percent aligned bases	The percentage of aligned bases for the reads.
Percent aligned bases Read 1	The percentage of aligned bases for Read 1.
Percent aligned bases Read 2	The percentage of aligned bases for Read 2.
Mean coverage	The total number of aligned bases divided by the genome size.
Top Strand	The number of read pairs (or reads when single-end) aligned to the bisulfite-converted top strand of reference genome.
Complementary Top Strand	The number of aligned reads that are complementary to the original bisulfite-treated top strand. When the library prep kit is directional, the value is 0.
Bottom Strand	The number of aligned reads that are the original bisulfite-treated bottom strand.
Complementary Bottom Strand	The number of aligned reads that are complementary to the original bisulfite-treated bottom strand. When the library prep kit is directional, the value is 0.
Number of Cs analyzed	The number of C bases in aligned reads.
Number of Methylated Cs in CpG	The number of methylated C bases in aligned reads that are in the context of CpG.
Number of Methylated Cs in CHG	The number of methylated C bases in aligned reads that are in the context of CHG.
Number of Methylated Cs in CHH	The number of methylated C bases in aligned reads that are in the context of CHH.
Number of Methylated Cs in others	The number of methylated C bases in aligned reads that are in other contexts.
Number of Unmethylated Cs in CpG	The number of unmethylated C bases in aligned reads that are in the context of CpG.
Number of Unmethylated Cs in CHG	The number of unmethylated C bases in aligned reads that are in the context of CHG.
Number of Unmethylated Cs in CHH	The number of unmethylated C bases in aligned reads that are in the context of CHH.
Number of Unmethylated Cs in others	The number of unmethylated C bases in aligned reads that are in other contexts.
Percent Methylated Cs in CpG	The percent of methylated C bases in aligned reads that are in the context of CpG.
Percent Methylated Cs in CHG	The percent of methylated C bases in aligned reads that are in the context of CHG.
Percent Methylated Cs in CHH	The percent of methylated C bases in aligned reads that are in the context of CHH.

Statistic	Definition
Percent Methylated Cs in others	The percent of methylated C bases in aligned reads that are in other contexts (besides contexts listed).

Workflow Status Report (WorkflowStatus.txt)

The MethylSeq v1.0 app provides a Workflow Status Report (WorkflowStatus.txt).

Statistic	Definition
JobEvent	Status of the run.
Duration	Duration of the run.
StartTime	Date and time the run started.
EndTime	Date and time the run ended.
JobID	Job number.
JobCategory	Category of job.
JobName	Name of job.
CoresRequested	Number of cores requested for the job.
MemoryRequested	Amount of memory requested for the job.
IORequested	The amount (a float between 0 and 1) of I/O (reading and writing) requested for this job. The amount is 0 if a job does not need to read or write the disks. The amount is 1 if a job reads or writes the disks so intensively that it blocks out any job that requires disk reading/writing.
MemoryRequestedPerMBInput	Amount of memory requested per MB of input for the job.
MemoryUsed	Amount of memory used for the job.
TotalInputSize	Total input size.
TotalOutputSize	Total output size.

Analysis Reports

The MethylSeq v1.0 app generates an overview of statistics for the sample on the Analysis Reports page. You can also download the MethylSeq Report as PDF. See *Sample Report* on page 10.

This page includes a link to the Bismark Processing Report. See *Bismark Processing Report* on page 8.

Sample Information

Statistic	Definition
Total PF Reads	The number of reads (2x the number of pairs for paired-end data) in the trimmed FASTQ files.
Percent Q30	The percentage of bases with a quality score of 30 or higher.

Cytosine Methylation

Provides the total number of C bases analyzed and the following metrics.

Statistic	Definition
Category	Describes whether the C base is methylated or unmethylated.
Cs in CpG	The number of C bases in aligned reads that are in the context of CpG.
Cs in CHG	The number of C bases in aligned reads that are in the context of CHG.
Cs in CHH	The number of C bases in aligned reads that are in the context of CHH.

Read and Base Alignment Statistics

Provides alignment statistics for read and base and the following metrics.

► Read Level Statistics

Statistic	Definition
Total Aligned Reads	The total number of reads present in the data set that are aligned to the reference genome.
Percent Aligned Reads	The percentage of reads present in the data set that are aligned to the reference genome.

► Bisulfite-Treated Strand Alignment Statistics

Statistic	Definition
Top Strand	Number of aligned reads that are the original bisulfite-treated top strand.
Complementary Top Strand	Number of aligned reads that are complementary to the original bisulfite-treated top strand. When the library prep kit is directional, the value is 0.
Bottom Strand	Number of aligned reads that are the original bisulfite-treated bottom strand.
Complementary Bottom Strand	Number of aligned reads that are complementary to the original bisulfite-treated bottom strand. When the library prep kit is directional, the value is 0.

► Base Level Statistics

Statistic	Definition
Total Aligned Bases	The total number of bases present in the data set that are aligned to the reference genome.
Percent Aligned Bases	The percentage of bases present in the data set that are aligned to the reference genome.

Coverage Histogram (Mean Coverage)

Provides the mean coverage and the following metrics.

Statistic	Definition
Number of Bases at Coverage X	Number of bases that have at least the indicated depth of coverage.
Depth of Sequencing Coverage	The coverage depth of a position in the genome refers to the number of sequenced bases that align to that position.

Fragment Length Summary

Statistic	Definition
Fragment Length Median	Median length of the sequenced fragment. The fragment length is calculated based on the locations at which a read pair aligns to the reference. The read mapping information is parsed from the BAM files.
Minimum	Minimum length of the sequenced fragment.
Maximum	Maximum length of the sequenced fragment.
Standard Deviation	Standard deviation of the sequenced fragment length.

Duplicate Information

Statistic	Definition
Percent Duplicate Paired Reads	Percentage of paired reads that have duplicates.

Analysis Details

Provides settings information and software versions.

MethylSeq Methods

The following methods are used in the MethylSeq v1.0 app.

Bismark

The MethylSeq workflow uses Bismark for methylation calling. Bismark is a tool that maps bisulfite-converted sequence reads and determines cytosine methylation states. The Bismark tool is a product of the Babraham Bioinformatics Group at the Babraham Institute.

For more information and to review the Bismark User Guide, see www.bioinformatics.bbsrc.ac.uk/projects/bismark/.

- 1 Krueger F. and Andrews S.R. (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*, 27, 1571–1572.

Bowtie2

The MethylSeq workflow uses the Bowtie2 aligner. Bowtie2 aligns sequencing reads to long reference sequences, and supports gapped, local, and paired-end alignment modes.

For more information, see bowtie-bio.sourceforge.net/bowtie2/index.shtml.

- 1 Langmead B. and Salzberg S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9, 357–359.

Technical Assistance

For technical assistance, contact Illumina Technical Support.

Table 1 Illumina General Contact Information

Website	www.illumina.com
Email	techsupport@illumina.com

Table 2 Illumina Customer Support Telephone Numbers

Region	Contact Number	Region	Contact Number
North America	1.800.809.4566	Italy	800.874909
Australia	1.800.775.688	Netherlands	0800.0223859
Austria	0800.296575	New Zealand	0800.451.650
Belgium	0800.81102	Norway	800.16836
Denmark	80882346	Spain	900.812168
Finland	0800.918363	Sweden	020790181
France	0800.911850	Switzerland	0800.563118
Germany	0800.180.8994	United Kingdom	0800.917.0041
Ireland	1.800.812949	Other countries	+44.1799.534000

Safety Data Sheets

Safety data sheets (SDSs) are available on the Illumina website at support.illumina.com/sds.html.

Product Documentation

Product documentation in PDF is available for download from the Illumina website. Go to support.illumina.com, select a product, then click **Documentation & Literature**.



Part # 15069929 Rev. A



Illumina
San Diego, California 92122 U.S.A.
+1.800.809.ILMN (4566)
+1.858.202.4566 (outside North America)
techsupport@illumina.com
www.illumina.com