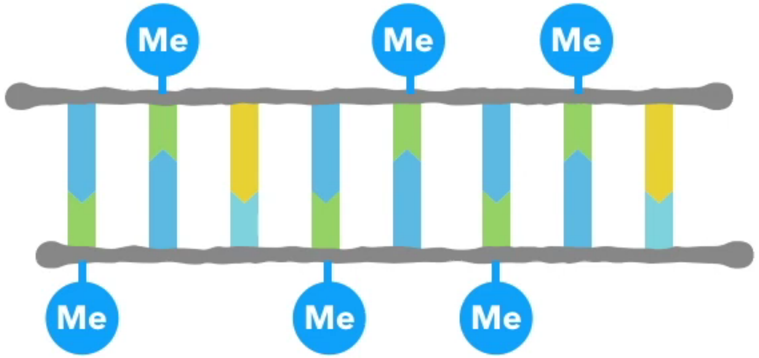
Methylation analysis

Architecture of large projects in bioinformatics

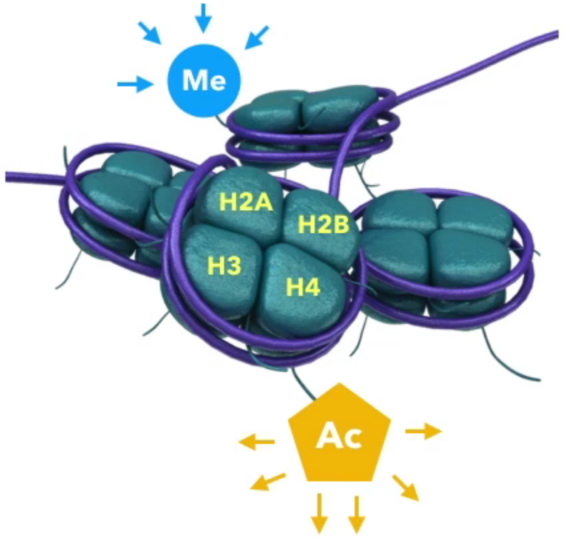
Krzysztof Łukasz

What is epigenetics

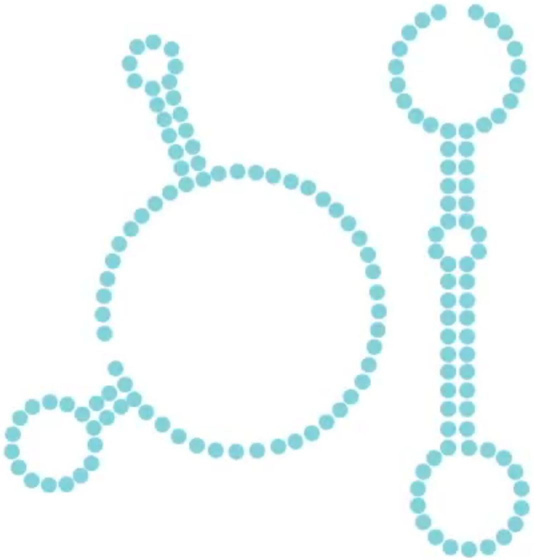
DNA Methylation



Histone Modification



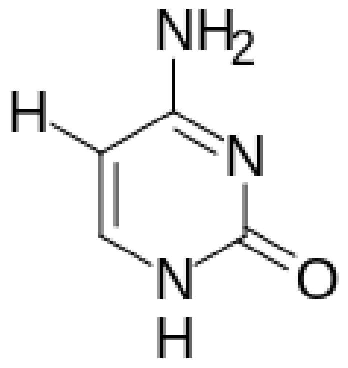
Non-coding RNA



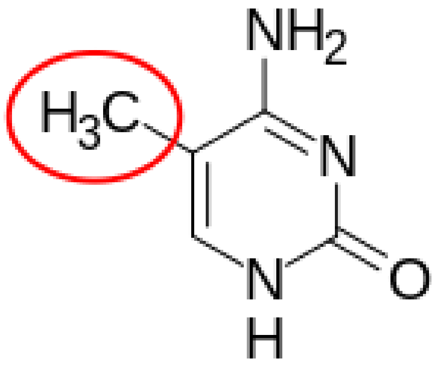
"the branch of biology which studies the causal interactions between genes and their products which bring the phenotype into being."

- Conrad Waddington 1940s

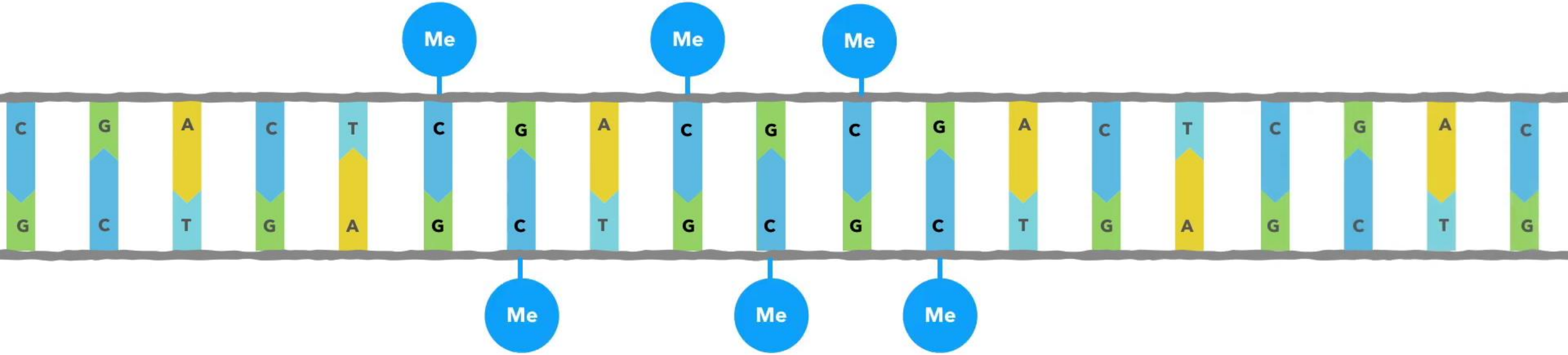
DNA methylation

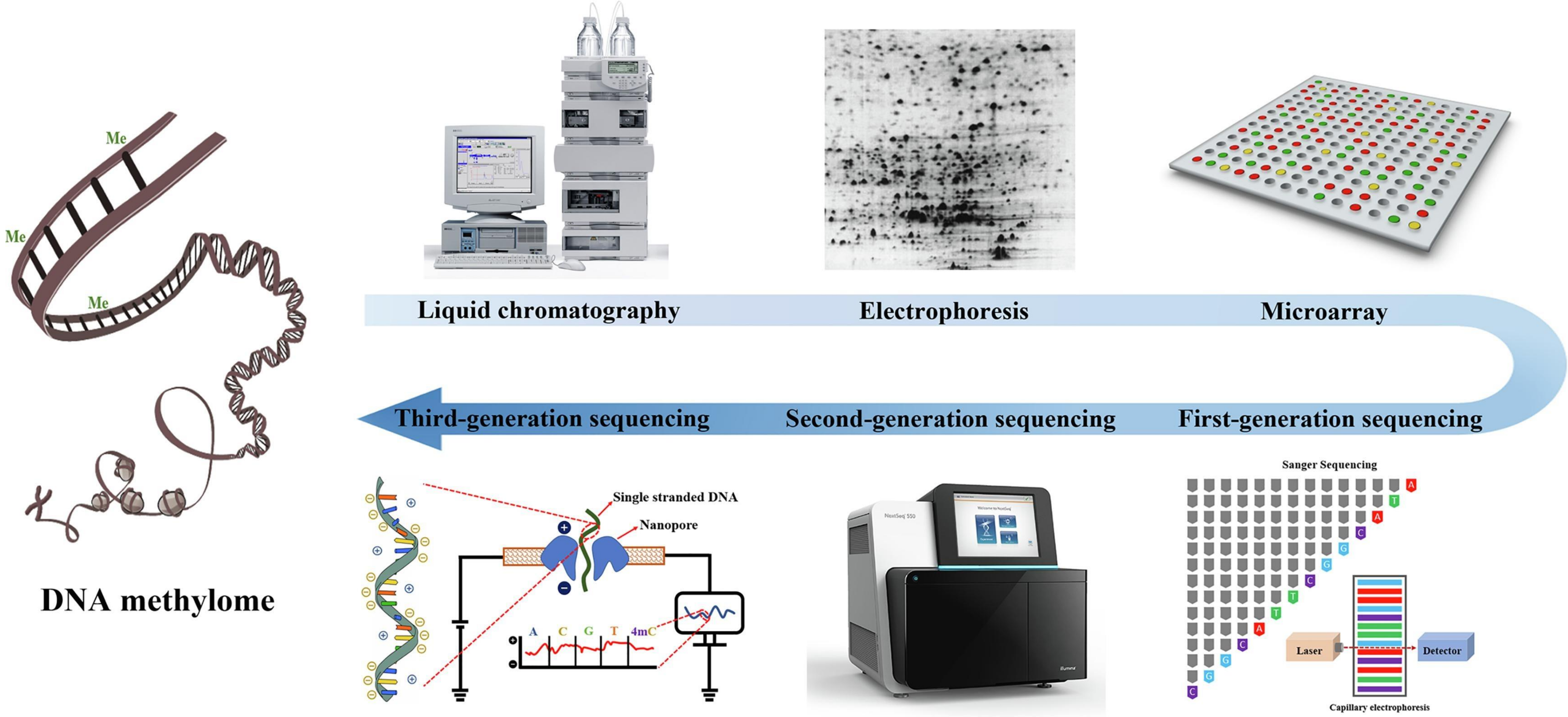


cytosine



methylated cytosine


Me
Me
Me
C G A C T C G A C G C G A C T C G A C
G C T G A G C T G C G C T G A G C T G
Me
Me
Me


Me
Me
Me
DNA methylome
Liquid chromatography
Electrophoresis
Microarray
First-generation sequencing
Second-generation sequencing
Third-generation sequencing
Single stranded DNA
Nanopore
A
C
G
T
4mC
Sanger Sequencing
Laser
Detector
Capillary electrophoresis

Epigenetic modification prediction tools based on the type of epigenetic mechanism (information until March 2020).

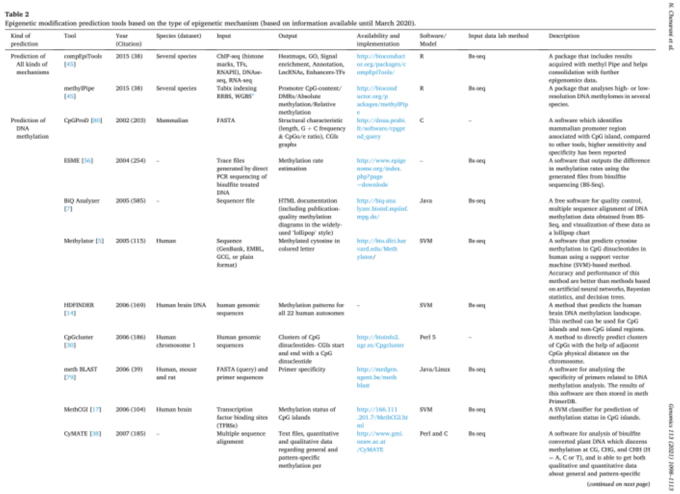


Table 2
Epigenetic modification prediction tools based on the type of epigenetic mechanism (based on information available until March 2020).

Kind of prediction	Tool	Year (Citation)	Species (dataset)	Input	Output	Availability and implementation	Software/ Model	Input data lab method	Description
Prediction of All kinds of mechanisms	compEpiTools [45]	2015 (38)	Several species	ChIP-seq (histone marks, TFs, RNAPII), DNAse-seq, RNA-seq	Heatmaps, GO, Signal enrichment, Annotation, LncRNAs, Enhancers-TFs	http://bioconductor.org/packages/compEpiTools/	R	Bs-seq	A package that includes results acquired with methyl Pipe and helps consolidation with further epigenomics data.
	methylPipe [45]	2015 (38)	Several species	Tabix indexing RRBS, WGBS[a]	Promoter CpG-content/ DMRs/Absolute methylation/Relative methylation	http://bioconductor.org/packages/methylPipe	R	Bs-seq	A package that analyses high- or low-resolution DNA methylomes in several species.
Prediction of DNA methylation	CpGProD [80]	2002 (203)	Mammalian	FASTA	Structural characteristic (length, G + C frequency & CpGo/e ratio), CGIs graphs	http://doua.prabi.fr/software/cpgprod_query	C	–	A software which identifies mammalian promoter region associated with CpG island, compared to other tools, higher sensitivity and specificity has been reported
	ESME [56]	2004 (254)	–	Trace files generated by direct PCR sequencing of bisulfite treated DNA	Methylation rate estimation	http://www.epigenome.org/index.php?page=downlode	–	Bs-seq	A software that outputs the difference in methylation rates using the generated files from bisulfite sequencing (BS-Seq).
	BiQ Analyzer [7]	2005 (585)	–	Sequencer file	HTML documentation (including publication-quality methylation diagrams in the widely-used 'lollipop' style)	http://biq-analyzer.bioinf.mpiinf.mpg.de/	Java	Bs-seq	A free software for quality control, multiple sequence alignment of DNA methylation data obtained from BS-Seq, and visualization of these data as a lollipop chart
	Methylator [5]	2005 (115)	Human	Sequence (GenBank, EMBL, GCG, or plain format)	Methylated cytosine in colored letter	http://bio.dfci.harvard.edu/Methylator/	SVM	Bs-seq	A software that predicts cytosine methylation in CpG dinucleotides in human using a support vector machine (SVM)-based method. Accuracy and performance of this method are better than methods based on artificial neural networks, Bayesian statistics, and decision trees.
	HDFINDER [14]	2006 (169)	Human brain DNA	human genomic sequences	Methylation patterns for all 22 human autosomes	–	SVM	Bs-seq	A method that predicts the human brain DNA methylation landscape. This method can be used for CpG islands and non-CpG island regions.
	CpGcluster [30]	2006 (186)	Human chromosome 1	Human genomic sequences	Clusters of CpG dinucleotides- CGIs start and end with a CpG dinucleotide	http://bioinfo2.ugr.es/Cpgcluster	Perl 5	–	A method to directly predict clusters of CpGs with the help of adjacent CpGs physical distance on the chromosome.
	meth BLAST [79]	2006 (39)	Human, mouse and rat	FASTA (query) and primer sequences	Primer specificity	http://medgen.ugent.be/methblast	Java/Linux	Bs-seq	A software for analyzing the specificity of primers related to DNA methylation analysis. The results of this software are then stored in meth PrimerDB.
	MethCGI [17]	2006 (104)	Human brain	Transcription factor binding sites (TFBSs)	Methylation status of CpG islands	http://166.111.201.7/MethCGI.html	SVM	Bs-seq	A SVM classifier for prediction of methylation status in CpG islands.
	CyMATE [38]	2007 (185)	–	Multiple sequence alignment	Text files, quantitative and qualitative data regarding general and pattern-specific methylation per	http://www.gmi.oeaw.ac.at/CyMATE	Perl and C	Bs-seq	A software for analysis of bisulfite converted plant DNA which discerns methylation at CG, CHG, and CHH (H = A, C or T), and is able to get both qualitative and quantitative data about general and pattern-specific

(continued on next page)

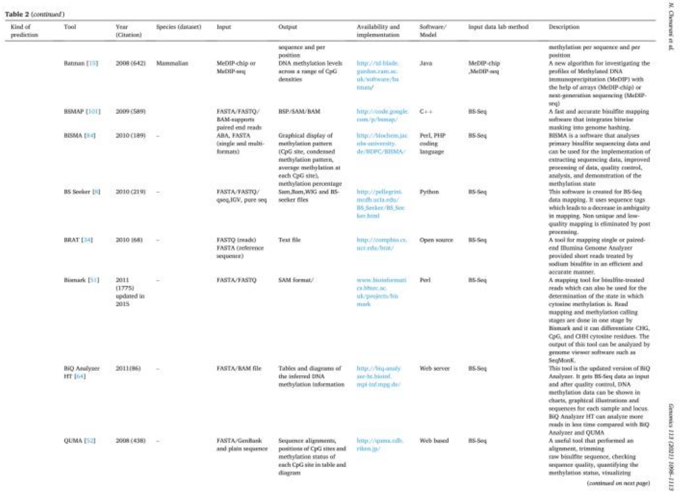


Table 2 (continued)

Kind of prediction	Tool	Year (Citation)	Species (dataset)	Input	Output	Availability and implementation	Software/ Model	Input data lab method	Description
					sequence and per position				methylation per sequence and per position
	Batman [15]	2008 (642)	Mammalian	MeDIP-chip or MeDIP-seq	DNA methylation levels across a range of CpG densities	http://td-blade.gurdon.cam.ac.uk/software/batman/	Java	MeDIP-chip ,MeDIP-seq	A new algorithm for investigating the profiles of Methylated DNA Immunoprecipitation (MeDIP) with the help of arrays (MeDIP-chip) or next-generation sequencing (MeDIP-seq)
	BSMAP [101]	2009 (589)		FASTA/FASTQ/ BAM-supports paired end reads	BSP/SAM/BAM	http://code.google.com/p/bsmap/	C++	BS-Seq	A fast and accurate bisulfite mapping software that integrates bitwise masking into genome hashing.
	BISMA [84]	2010 (189)	–	ABA, FASTA (single and multi-formats)	Graphical display of methylation pattern (CpG site, condensed methylation pattern, average methylation at each CpG site), methylation percentage	http://biochem.jacobs-university.de/BDPC/BISMA/	Perl, PHP coding language	BS-Seq	BISMA is a software that analyses primary bisulfite sequencing data and can be used for the implementation of extracting sequencing data, improved processing of data, quality control, analysis, and demonstration of the methylation state
	BS Seeker [8]	2010 (219)	–	FASTA/FASTQ/ qseq,IGV, pure seq	Sam,Bam,WIG and BS-seeker files	http://pellegrini.mcdb.ucla.edu/BS_Seeker/BS_Seeker.html	Python	BS-Seq	This software is created for BS-Seq data mapping. It uses sequence tags which leads to a decrease in ambiguity in mapping. Non unique and low-quality mapping is eliminated by post processing.
	BRAT [34]	2010 (68)	–	FASTQ (reads) FASTA (reference sequence)	Text file	http://compbio.cs.ucr.edu/brat/	Open source	BS-Seq	A tool for mapping single or paired-end Illumina Genome Analyzer provided short reads treated by sodium bisulfite in an efficient and accurate manner.
	Bismark [51]	2011 (1775) updated in 2015	–	FASTA/FASTQ	SAM format/	www.bioinformatics.bbsrc.ac.uk/projects/bismark	Perl	BS-Seq	A mapping tool for bisulfite-treated reads which can also be used for the determination of the state in which cytosine methylation is. Read mapping and methylation calling stages are done in one stage by Bismark and it can differentiate CHG, CpG, and CHH cytosine residues. The output of this tool can be analyzed by genome viewer software such as SeqMonK.
	BiQ Analyzer HT [64]	2011(86)	–	FASTA/BAM file	Tables and diagrams of the inferred DNA methylation information	http://biq-analyzer-ht.bioinf.mpi-inf.mpg.de/	Web server	BS-Seq	This tool is the updated version of BiQ Analyzer. It gets BS-Seq data as input and after quality control, DNA methylation data can be shown in charts, graphical illustrations and sequences for each sample and locus. BiQ Analyzer HT can analyze more reads in less time compared with BiQ Analyzer and QUMA
	QUMA [52]	2008 (438)	–	FASTA/GenBank and plain sequence	Sequence alignments, positions of CpG sites and methylation status of each CpG site in table and diagram	http://quma.cdb.riken.jp/	Web based	BS-Seq	A useful tool that performed an alignment, trimming raw bisulfite sequence, checking sequence quality, quantifying the methylation status, visualizing

(continued on next page)

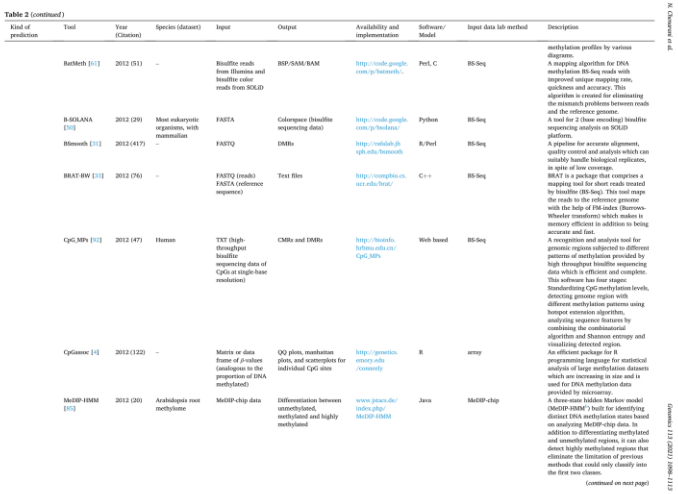


Table 2 (continued)

Kind of prediction	Tool	Year (Citation)	Species (dataset)	Input	Output	Availability and implementation	Software/ Model	Input data lab method	Description
									methylation profiles by various diagrams.
	BatMeth [61]	2012 (51)	–	Bisulfite reads from Illumina and bisulfite color reads from SOLiD	BSP/SAM/BAM	http://code.google.com/p/batmeth/.	Perl, C	BS-Seq	A mapping algorithm for DNA methylation BS-Seq reads with improved unique mapping rate, quickness and accuracy. This algorithm is created for eliminating the mismatch problems between reads and the reference genome.
	B-SOLANA [50]	2012 (29)	Most eukaryotic organisms, with mammalian	FASTA	Colorspace (bisulfite sequencing data)	http://code.google.com/p/bsolana/	Python	BS-Seq	A tool for 2 (base encoding) bisulfite sequencing analysis on SOLiD platform.
	BSmooth [31]	2012 (417)	–	FASTQ	DMRs	http://rafalab.jhsph.edu/bsmooth	R/Perl	BS-Seq	A pipeline for accurate alignment, quality control and analysis which can suitably handle biological replicates, in spite of low coverage.
	BRAT-BW [33]	2012 (76)	–	FASTQ (reads) FASTA (reference sequence)	Text files	http://compbio.cs.ucr.edu/brat/	C++	BS-Seq	BRAT is a package that comprises a mapping tool for short reads treated by bisulfite (BS-Seq). This tool maps the reads to the reference genome with the help of FM-index (Burrows-Wheeler transform) which makes is memory efficient in addition to being accurate and fast.
	CpG_MPs [92]	2012 (47)	Human	TXT (high-throughput bisulfite sequencing data of CpGs at single-base resolution)	CMRs and DMRs	http://bioinfo.hrbmu.edu.cn/CpG_MPs	Web based	BS-Seq	A recognition and analysis tool for genomic regions subjected to different patterns of methylation provided by high throughput bisulfite sequencing data which is efficient and complete. This software has four stages: Standardizing CpG methylation levels, detecting genome region with different methylation patterns using hotspot extension algorithm, analyzing sequence features by combining the combinatorial algorithm and Shannon entropy and visualizing detected region.
	CpGassoc [4]	2012 (122)	–	Matrix or data frame of β-values (analogous to the proportion of DNA methylated)	QQ plots, manhattan plots, and scatterplots for individual CpG sites	http://genetics.emory.edu/conneely	R	array	An efficient package for R programming language for statistical analysis of large methylation datasets which are increasing in size and is used for DNA methylation data provided by microarray.
	MeDIP-HMM [85]	2012 (20)	Arabidopsis root methylome	MeDIP-chip data	Differentiation between unmethylated, methylated and highly methylated	www.jstacs.de/index.php/MeDIP-HMM	Java	MeDIP-chip	A three-state hidden Markov model (MeDIP-HMM[b]) built for identifying distinct DNA methylation states based on analyzing MeDIP-chip data. In addition to differentiating methylated and unmethylated regions, it can also detect highly methylated regions that eliminate the limitation of previous methods that could only classify into the first two classes.

(continued on next page)

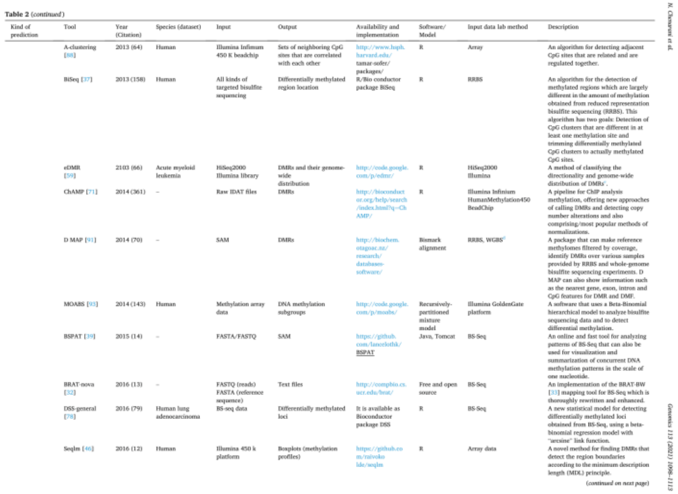


Table 2 (continued)

Kind of prediction	Tool	Year (Citation)	Species (dataset)	Input	Output	Availability and implementation	Software/ Model	Input data lab method	Description
	A-clustering [88]	2013 (64)	Human	Illumina Infimum 450 K beadchip	Sets of neighboring CpG sites that are correlated with each other	http://www.hsph.harvard.edu/tamar-sofer/packages/	R	Array	An algorithm for detecting adjacent CpG sites that are related and are regulated together.
	BiSeq [37]	2013 (158)	Human	All kinds of targeted bisulfite sequencing	Differentially methylated region location	R/Bio conductor package BiSeq	R	RRBS	An algorithm for the detection of methylated regions which are largely different in the amount of methylation obtained from reduced representation bisulfite sequencing (RRBS). This algorithm has two goals: Detection of CpG clusters that are different in at least one methylation site and trimming differentially methylated CpG clusters to actually methylated CpG sites.
	eDMR [59]	2103 (66)	Acute myeloid leukemia	HiSeq2000 Illumina library	DMRs and their genome-wide distribution	http://code.google.com/p/edmr/	R	HiSeq2000 Illumina	A method of classifying the directionality and genome-wide distribution of DMRs[c].
	ChAMP [71]	2014 (361)	–	Raw IDAT files	DMRs	http://bioconductor.org/help/search/index.html?q=ChAMP/	R	Illumina Infinium HumanMethylation450 BeadChip	A pipeline for ChIP analysis methylation, offering new approaches of calling DMRs and detecting copy number alterations and also comprising/most popular methods of normalizations.
	D MAP [91]	2014 (70)	–	SAM	DMRs	http://biochem.otagoac.nz/research/databases-software/	Bismark alignment	RRBS, WGBS[d]	A package that can make reference methylomes filtered by coverage, identify DMRs over various samples provided by RRBS and whole-genome bisulfite sequencing experiments. D MAP can also show information such as the nearest gene, exon, intron and CpG features for DMR and DMF.
	MOABS [93]	2014 (143)	Human	Methylation array data	DNA methylation subgroups	http://code.google.com/p/moabs/	Recursively-partitioned mixture model	Illumina GoldenGate platform	A software that uses a Beta-Binomial hierarchical model to analyze bisulfite sequencing data and to detect differential methylation.
	BSPAT [39]	2015 (14)	–	FASTA/FASTQ	SAM	https://github.com/lancelothk/BSPAT	Java, Tomcat	BS-Seq	An online and fast tool for analyzing patterns of BS-Seq that can also be used for visualization and summarization of concurrent DNA methylation patterns in the scale of one nucleotide.
	BRAT-nova [32]	2016 (13)	–	FASTQ (reads) FASTA (reference sequence)	Text files	http://compbio.cs.ucr.edu/brat/	Free and open source	BS-Seq	An implementation of the BRAT-BW [33] mapping tool for BS-Seq which is thoroughly rewritten and enhanced.
	DSS-general [78]	2016 (79)	Human lung adenocarcinoma	BS-seq data	Differentially methylated loci	It is available as Bioconductor package DSS	R	BS-Seq	A new statistical model for detecting differentially methylated loci obtained from BS-Seq, using a beta-binomial regression model with "arcsine" link function.
	Seqlm [46]	2016 (12)	Human	Illumina 450 k platform	Boxplots (methylation profiles)	https://github.com/raivokolde/seqlm	R	Array data	A novel method for finding DMRs that detect the region boundaries according to the minimum description length (MDL) principle.

(continued on next page)

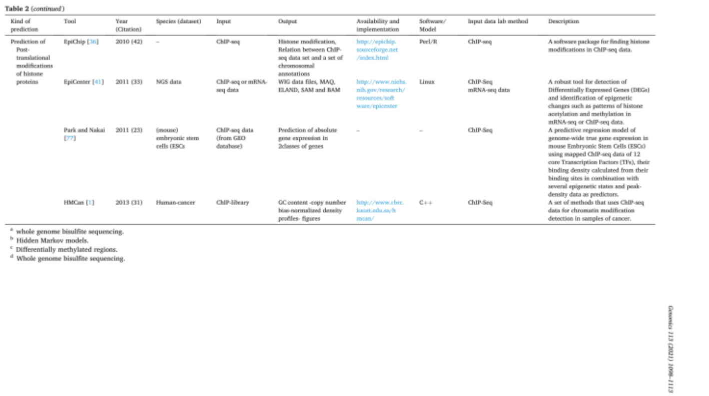


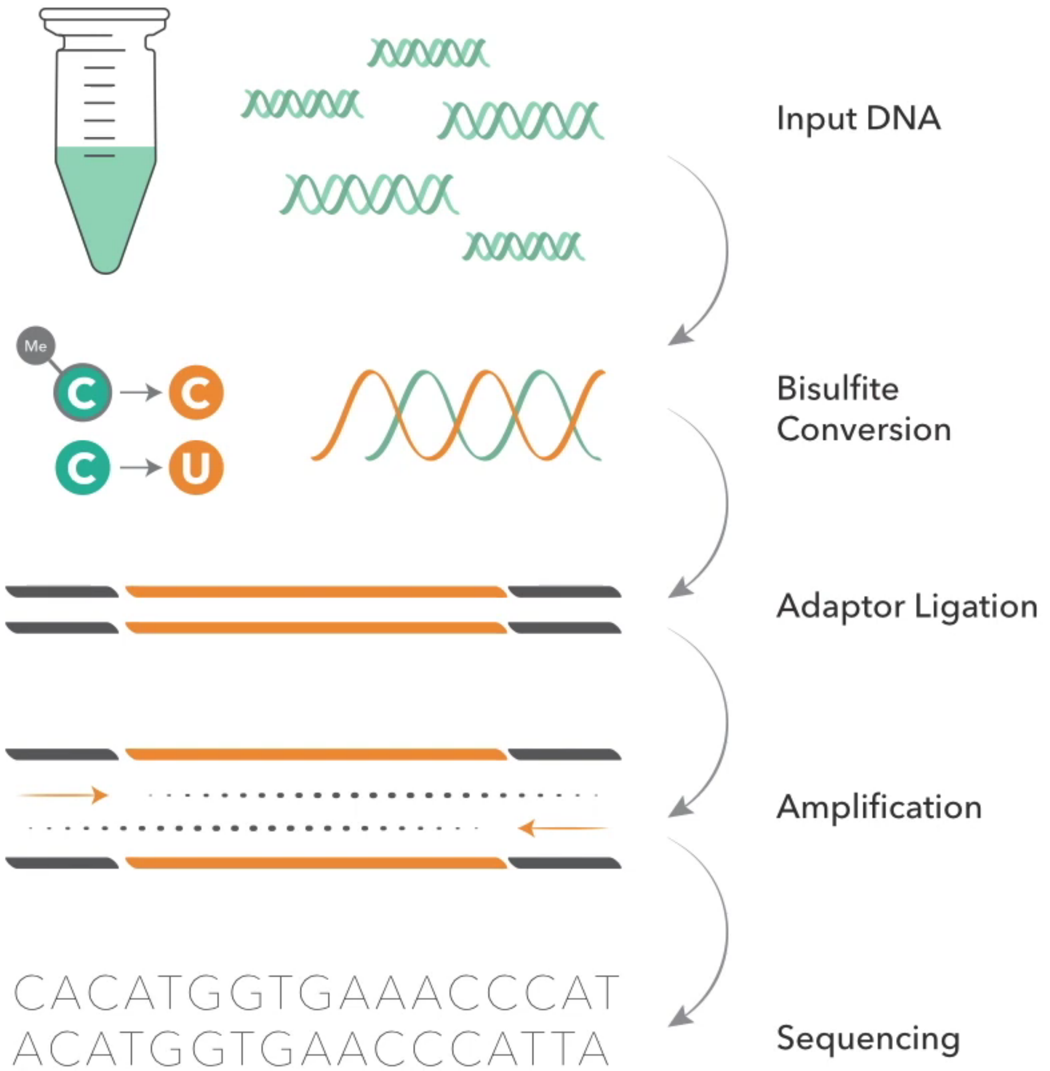
Table 2 (continued)

Kind of prediction	Tool	Year (Citation)	Species (dataset)	Input	Output	Availability and implementation	Software/ Model	Input data lab method	Description
Prediction of Post-translational modifications of histone proteins	EpiChip [36]	2010 (42)	–	ChIP-seq	Histone modification, Relation between ChIP-seq data set and a set of chromosomal annotations	http://epichip.sourceforge.net/index.html	Perl/R	ChIP-seq	A software package for finding histone modifications in ChIP-seq data.
	EpiCenter [41]	2011 (33)	NGS data	ChIP-seq or mRNA-seq data	WIG data files, MAQ, ELAND, SAM and BAM	http://www.niehs.nih.gov/research/resources/software/epicenter	Linux	ChIP-Seq mRNA-seq data	A robust tool for detection of Differentially Expressed Genes (DEGs) and identification of epigenetic changes such as patterns of histone acetylation and methylation in mRNA-seq or ChIP-seq data.
	Park and Nakai [77]	2011 (23)	(mouse) embryonic stem cells (ESCs	ChIP-seq data (from GEO database)	Prediction of absolute gene expression in 2classes of genes	–	–	ChIP-Seq	A predictive regression model of genome-wide true gene expression in mouse Embryonic Stem Cells (ESCs) using mapped ChIP-seq data of 12 core Transcription Factors (TFs), their binding density calculated from their binding sites in combination with several epigenetic states and peak-density data as predictors.
	HMCan [1]	2013 (31)	Human-cancer	ChIP-library	GC content -copy number bias-normalized density profiles- figures	http://www.cbrc.kaust.edu.sa/hmcan/	C++	ChIP-Seq	A set of methods that uses ChIP-seq data for chromatin modification detection in samples of cancer.

[a] whole genome bisulfite sequencing.
[b] Hidden Markov models.
[c] Differentially methylated regions.
[d] Whole genome bisulfite sequencing.

Bioinformatic tools for DNA methylation and histone modification: A survey, Chenarani et al., Genomics, 2011

BS experiment design



Input DNA

Bisulfite Conversion

Adaptor Ligation

Amplification

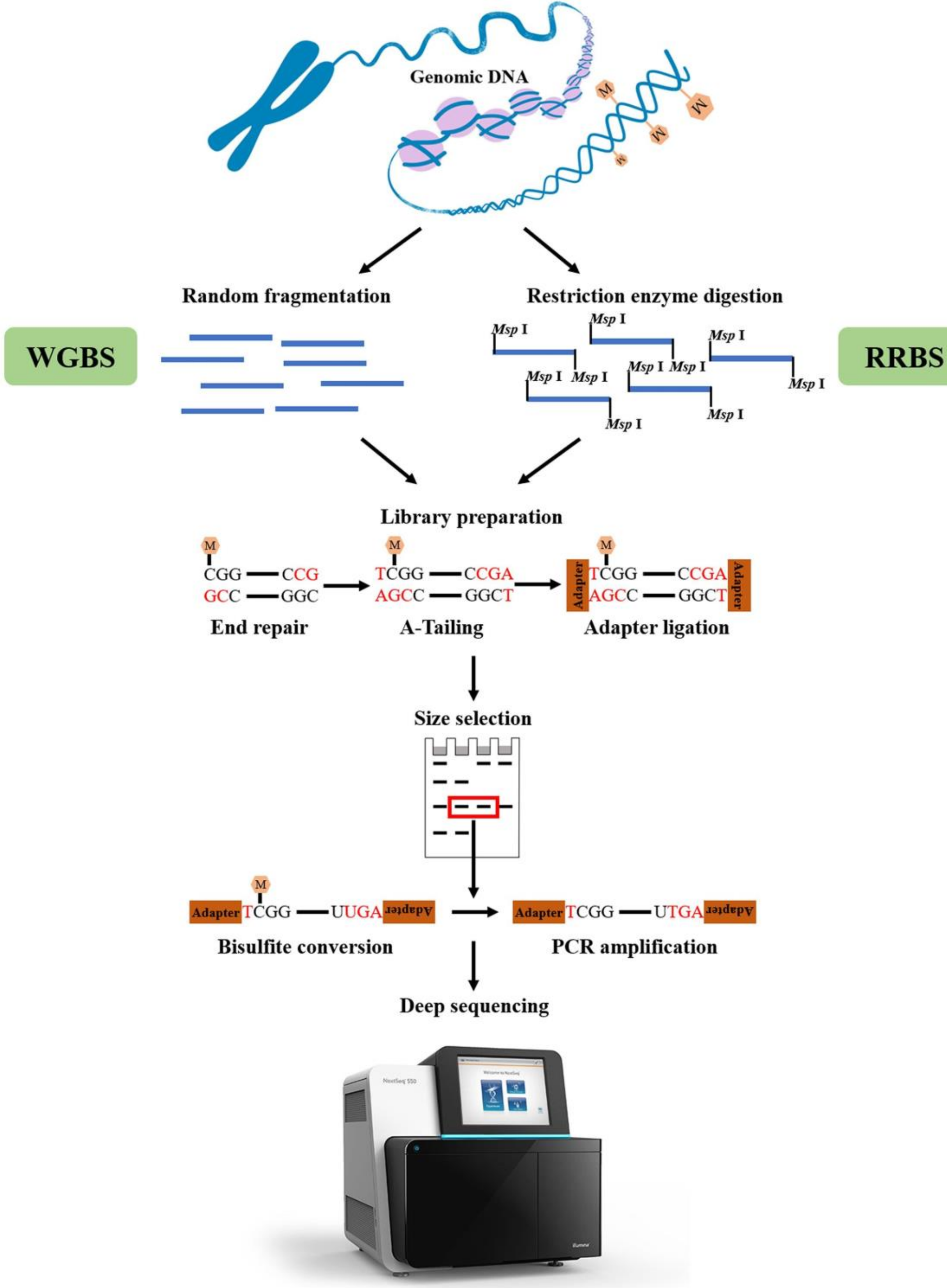
Sequencing

CACATGGTGAAACCCAT
ACATGGTGAACCCATTA

Next Generation Sequencing

Restrited BS

- Whole-genome methylation – expensive
- Restrictive enzymes
- Analyzing only CpG islands



Genomic DNA

Random fragmentation

Restriction enzyme digestion

WGBS

RRBS

Library preparation

End repair

A-Tailing

Adapter ligation

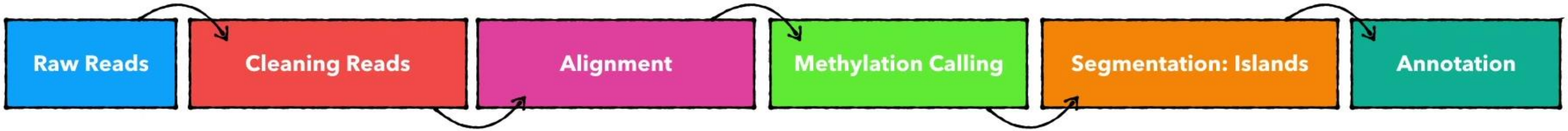
Size selection

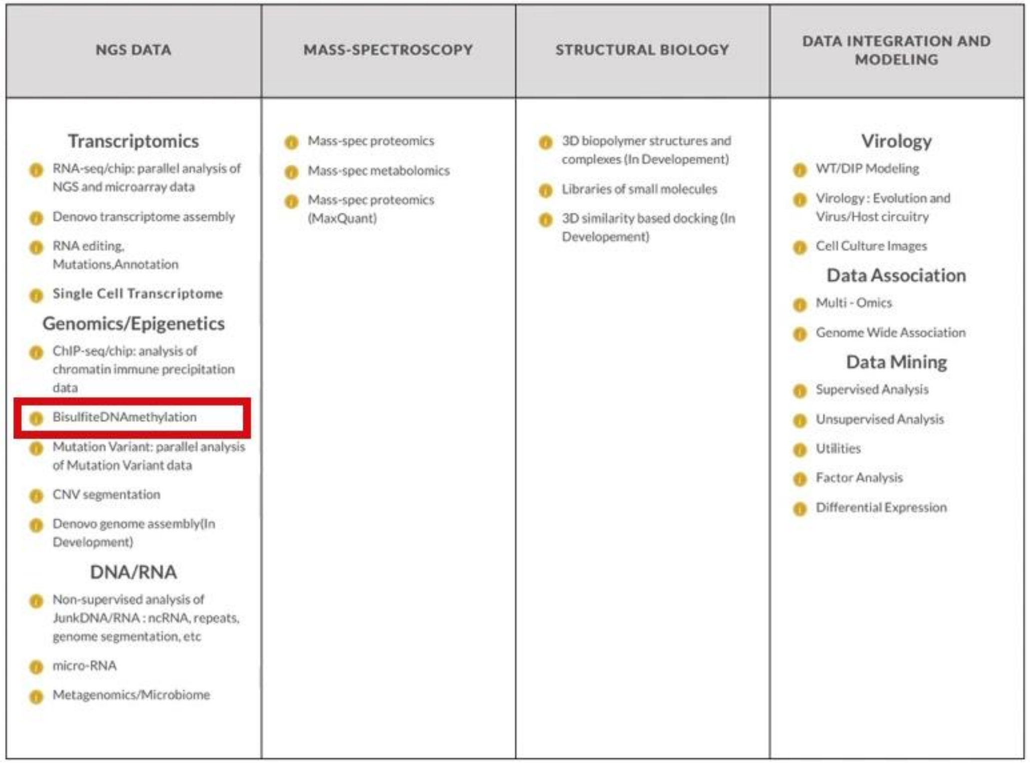
Bisulfite conversion

PCR amplification

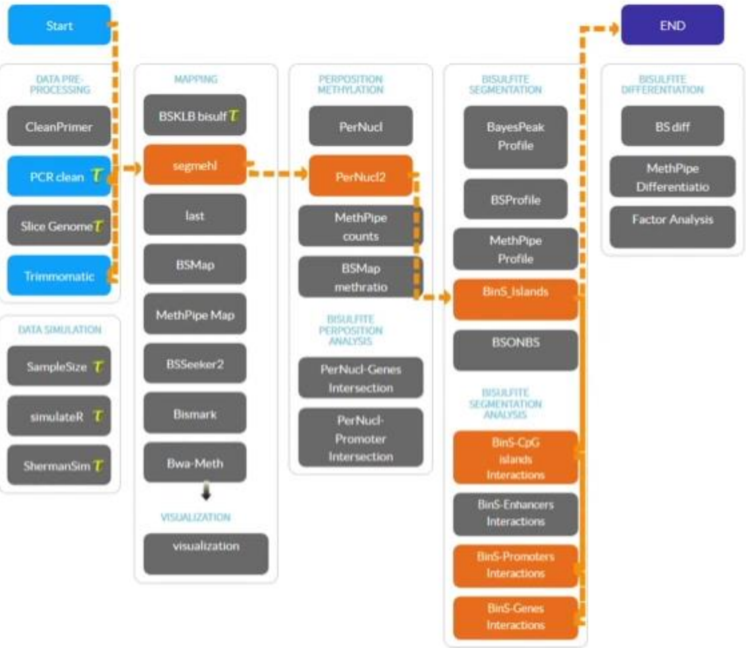
Deep sequencing

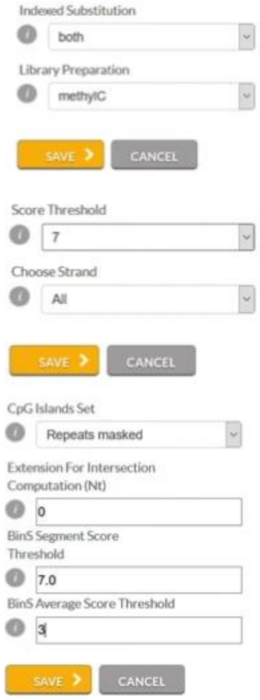
Bisulfite-seq workflow





NGS DATA	MASS-SPECTROSCOPY	STRUCTURAL BIOLOGY	DATA INTEGRATION AND MODELING
Transcriptomics RNA-seq/chip: parallel analysis of NGS and microarray data Denovo transcriptome assembly RNA editing, Mutations,Annotation **Single Cell Transcriptome** **Genomics/Epigenetics** ChIP-seq/chip: analysis of chromatin immune precipitation data BisulfiteDNAmethylation Mutation Variant: parallel analysis of Mutation Variant data CNV segmentation Denovo genome assembly(In Development) **DNA/RNA** Non-supervised analysis of JunkDNA/RNA : ncRNA, repeats, genome segmentation, etc micro-RNA Metagenomics/Microbiome	Mass-spec proteomics Mass-spec metabolomics Mass-spec proteomics (MaxQuant)	3D biopolymer structures and complexes (In Developement) Libraries of small molecules 3D similarity based docking (In Developement)	**Virology** WT/DIP Modeling Virology : Evolution and Virus/Host circuitry Cell Culture Images **Data Association** Multi - Omics Genome Wide Association **Data Mining** Supervised Analysis Unsupervised Analysis Utilities Factor Analysis Differential Expression





Indexed Substitution: both

Library Preparation: methylC

Score Threshold: 7

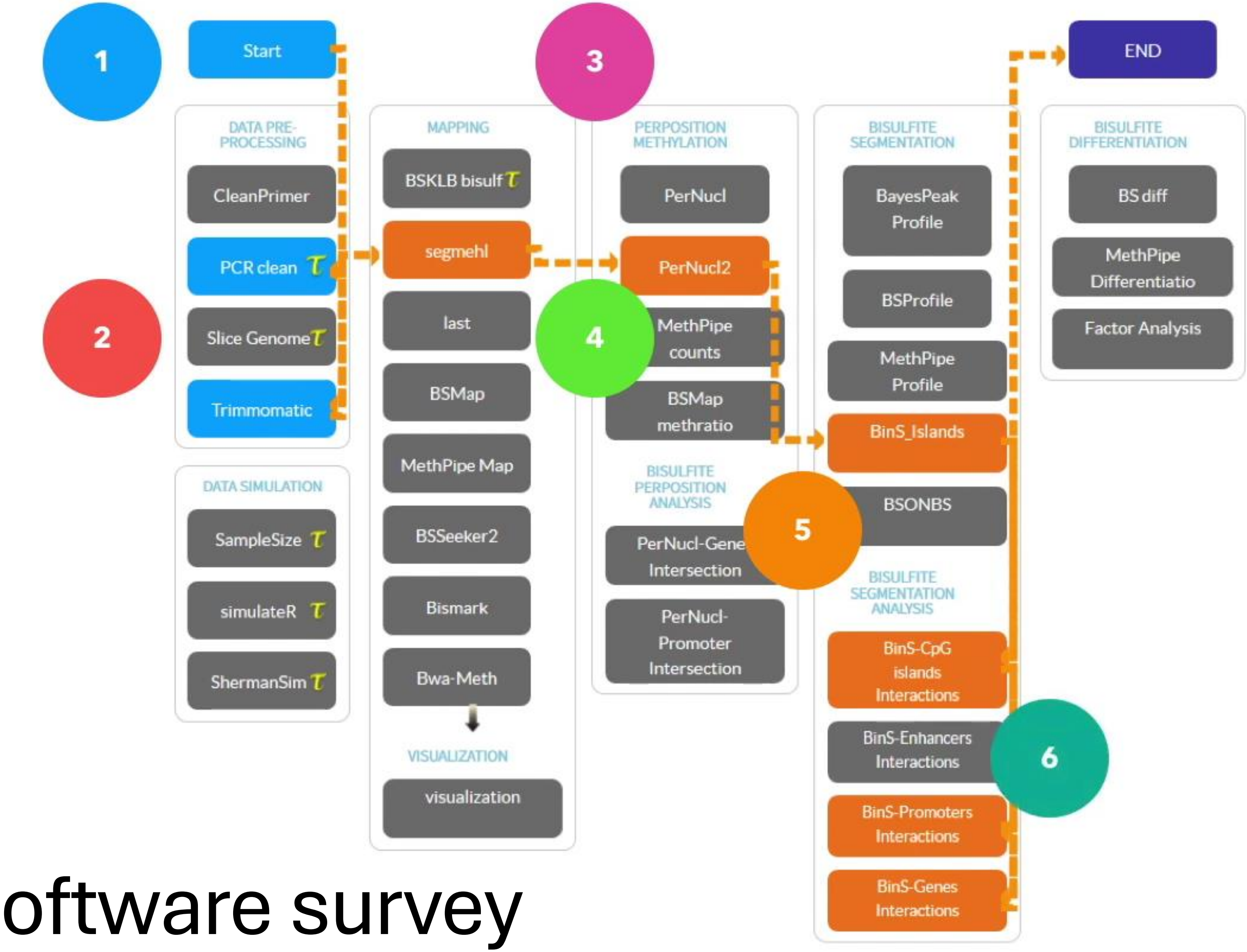
Choose Strand: All

CpG Islands Set: Repeats masked

Extension For Intersection Computation (Nt): 0

BinS Segment Score Threshold: 7.0

BinS Average Score Threshold: 3



1. Raw Reads

2. Cleaning Reads

3. Alignment

4. Methylation Calling

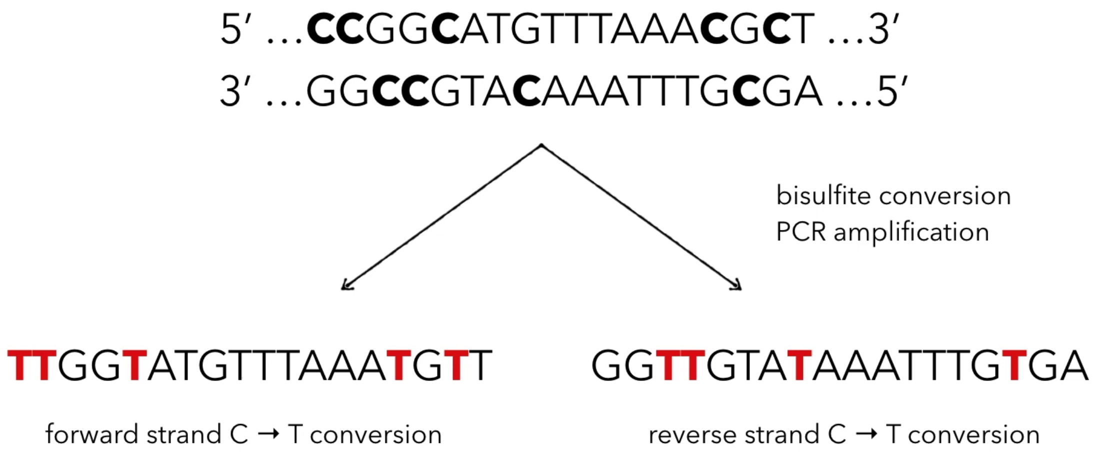
5. Segmentation: Islands

6. Annotation

Software survey

Library preparation

Directional Library Preparation



5′ …**CC**GG**C**ATGTTTAAA**C**G**C**T …3′
3′ …GG**CC**GTA**C**AAATTTG**C**GA …5′

bisulfite conversion
PCR amplification

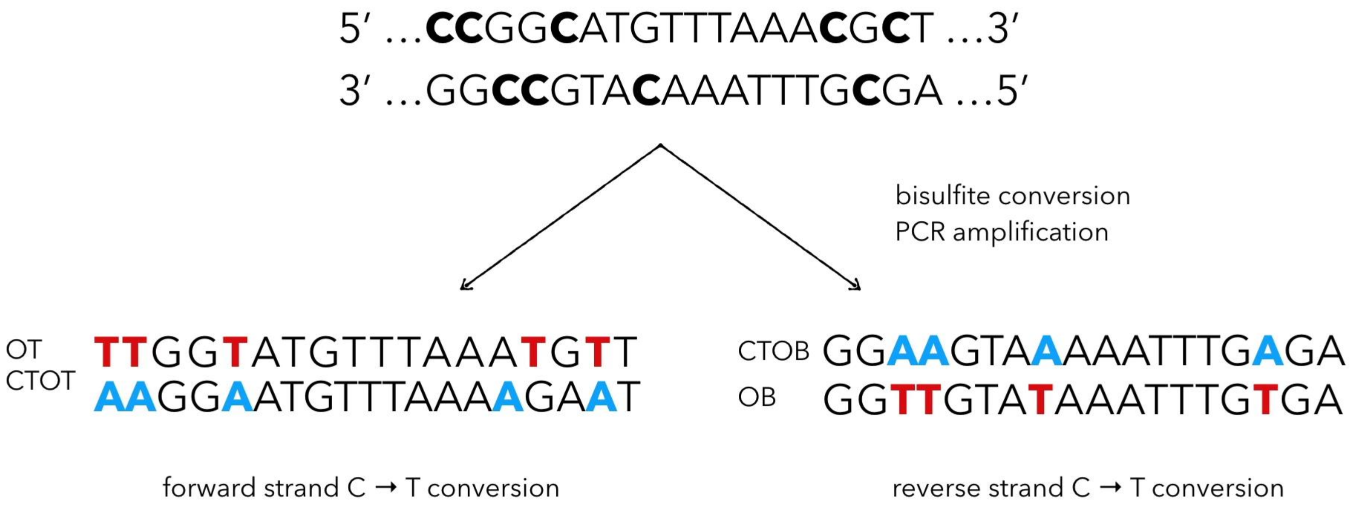
TTGG**T**ATGTTTAAA**T**G**T**T

forward strand C → T conversion

GG**TT**GTA**T**AAATTTG**T**GA

reverse strand C → T conversion

Non-directional Library Preparation



5′ …**CC**GG**C**ATGTTTAAA**C**G**C**T …3′
3′ …GG**CC**GTA**C**AAATTTG**C**GA …5′

bisulfite conversion
PCR amplification

OT **TT**GG**T**ATGTTTAAA**T**G**T**T
CTOT **AA**GG**A**ATGTTTAAA**A**GA**A**T

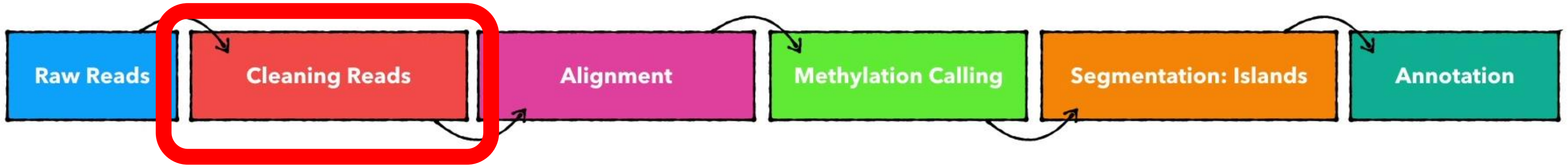
forward strand C → T conversion

CTOB GG**AA**GTA**A**AAATTTG**A**GA
OB GG**TT**GTA**T**AAATTTG**T**GA

reverse strand C → T conversion

Non-directional:

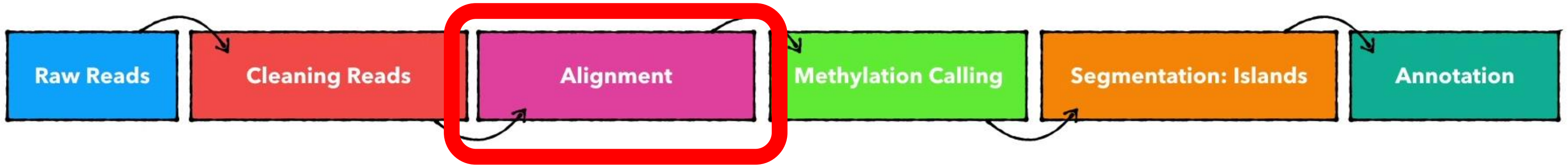
- Significantly cheaper
- By far more popular



Preparing reads

Standard approaches:

- FastQC
- Trimmomatic
- Cutadapt



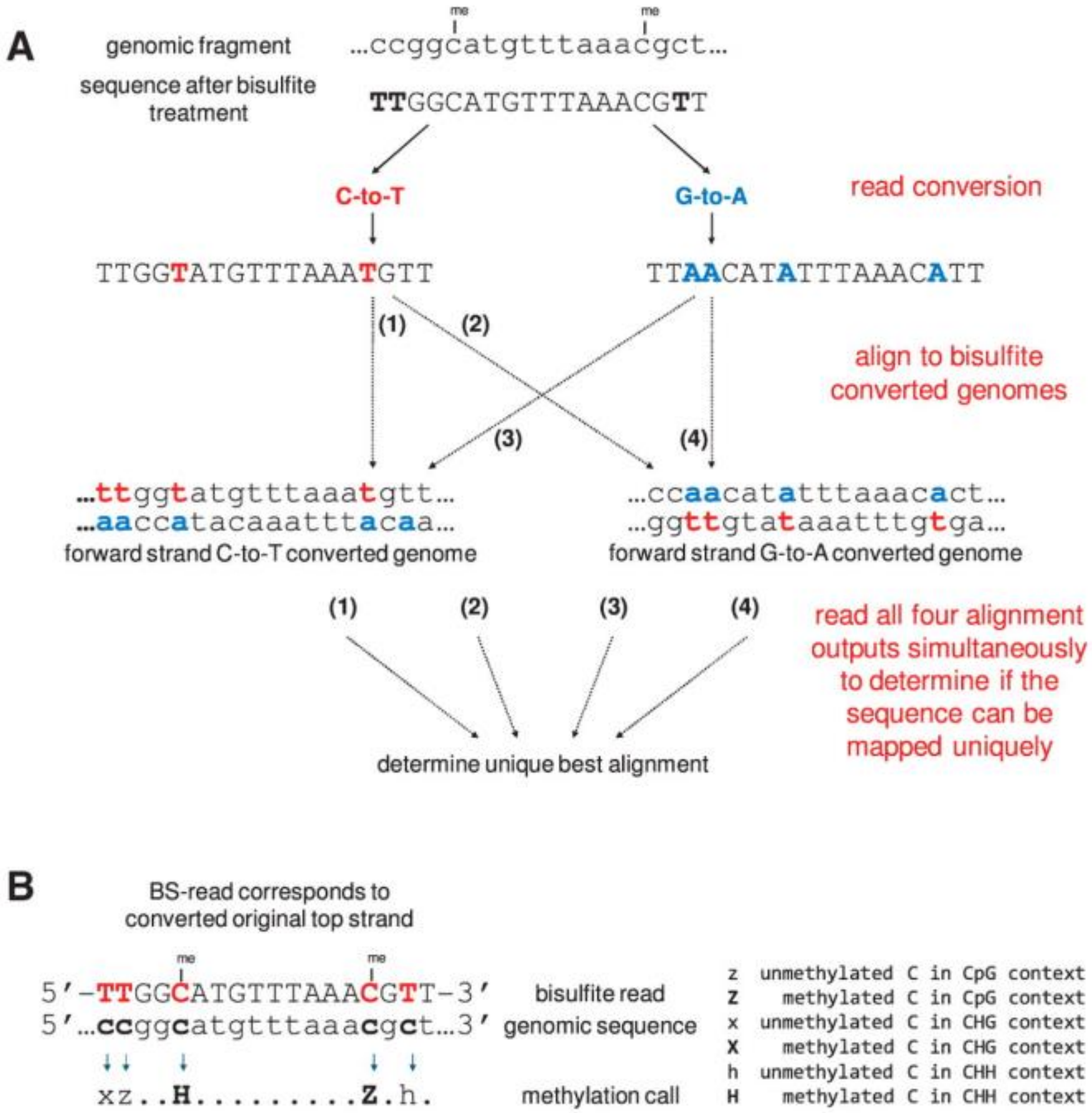
BS: Alignment

- Bisulfite treatment introduces mutations into genomic DNA in a methylation dependent manner
- Alignment of BS-seq reads is more challenging
  - Standard alignment methods cannot be used directly
- Standard tools: **Bismarck, BSMap, BWA-Metha**
- Bismark tool uses the following approach to map BS-seq reads
  - Reads from a BS-seq experiment are converted into a C-to-T version **and** a G-to-A version
  - The same conversion for the genome
  - External mapper (Bowtie) alignment to the genome
  - A unique best alignment is determined from four parallel alignment processes

Bismarck

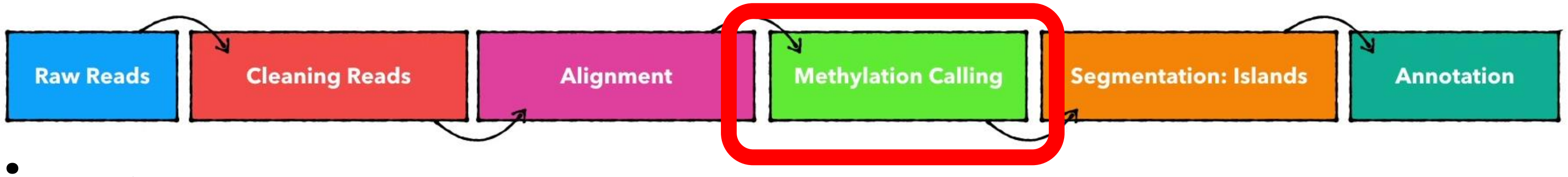
C-to-T **and** a G-to-A

- **Type**: Command-line tool.
- **Features**: Bismark is widely used for bisulfite sequencing data analysis. It aligns bisulfite-treated sequencing reads to a reference genome and can perform DNA methylation calling.


A
genomic fragment
...ccggcatgtttaaacgct...
me
me
sequence after bisulfite treatment
TTGGCATGTTTAAACGTT
C-to-T
G-to-A
read conversion
TTGGTATGTTTAAATGTT
TTAACATATTTAAACATT
(1)
(2)
(3)
(4)
align to bisulfite converted genomes
...ttggtatgtttaaatgtt...
...aaccatacaaatttacaa...
forward strand C-to-T converted genome
...ccaacatatttaaacact...
...ggttgtataaatttgtga...
forward strand G-to-A converted genome
(1)
(2)
(3)
(4)
read all four alignment outputs simultaneously to determine if the sequence can be mapped uniquely
determine unique best alignment
B
BS-read corresponds to converted original top strand
me
me
5′-TTGGCATGTTTAAACGTT-3′
bisulfite read
5′...ccggcatgtttaaacgct...3′
genomic sequence
xz..H.........Z.h.
methylation call

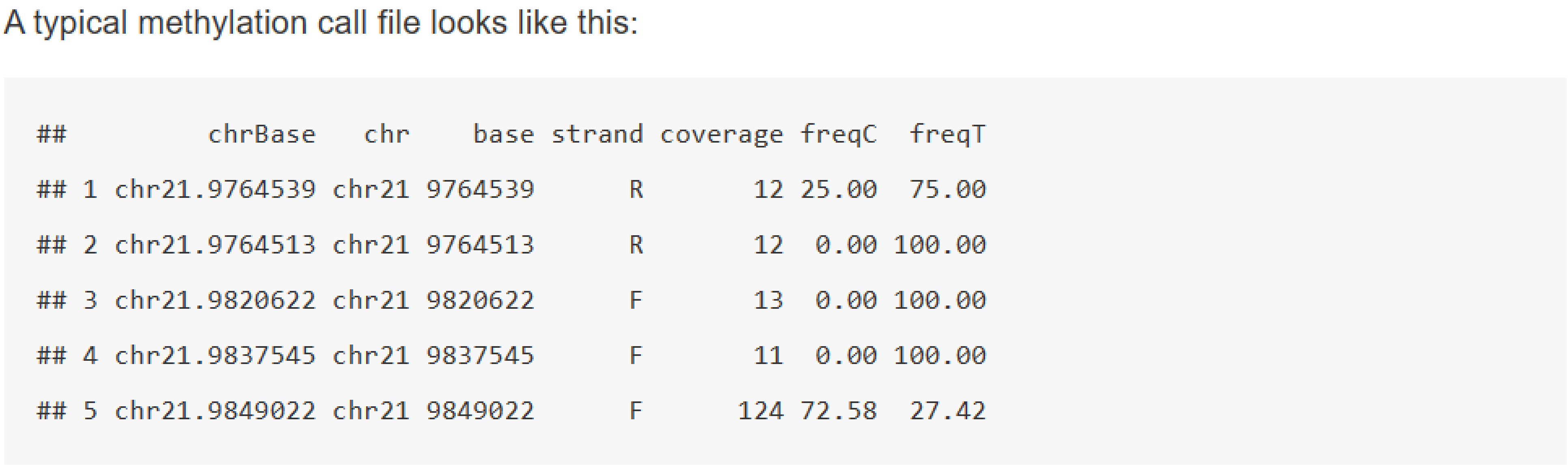
z	unmethylated C in CpG context
Z	methylated C in CpG context
x	unmethylated C in CHG context
X	methylated C in CHG context
h	unmethylated C in CHH context
H	methylated C in CHH context

Figure from (Krueger & Andrews, 2011)



Methylation calling

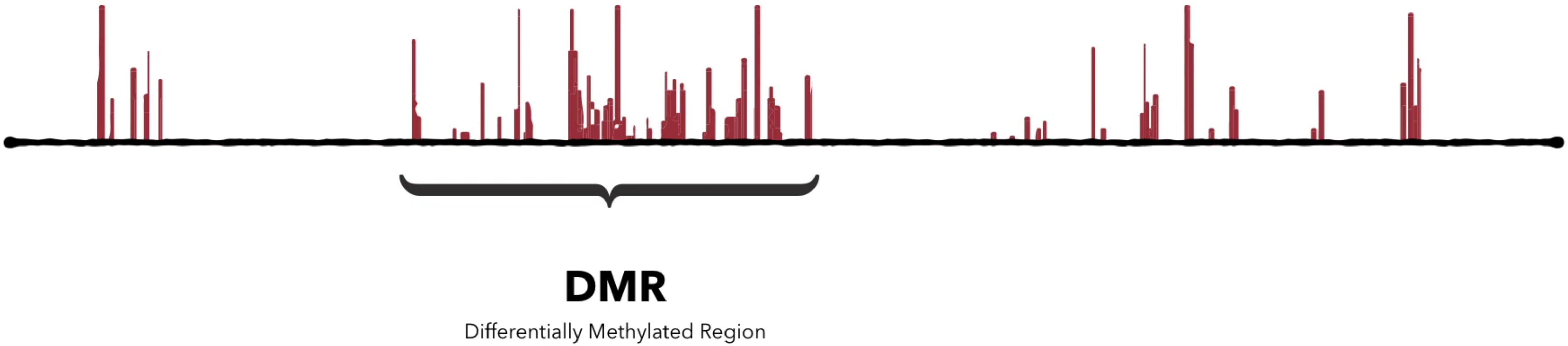
- Bismarck, BSMAP and other mappers also perform calling
- Separate software, such as Bis-SNP also available
- For each initial fastq file, we get a call that for each cytosine includes coverage, and methylation frequency
- Now, we need some way to do statistical analysis of such results



A typical methylation call file looks like this:

```
##           chrBase   chr     base strand coverage freqC  freqT
## 1 chr21.9764539 chr21 9764539      R       12 25.00  75.00
## 2 chr21.9764513 chr21 9764513      R       12  0.00 100.00
## 3 chr21.9820622 chr21 9820622      F       13  0.00 100.00
## 4 chr21.9837545 chr21 9837545      F       11  0.00 100.00
## 5 chr21.9849022 chr21 9849022      F      124 72.58  27.42
```

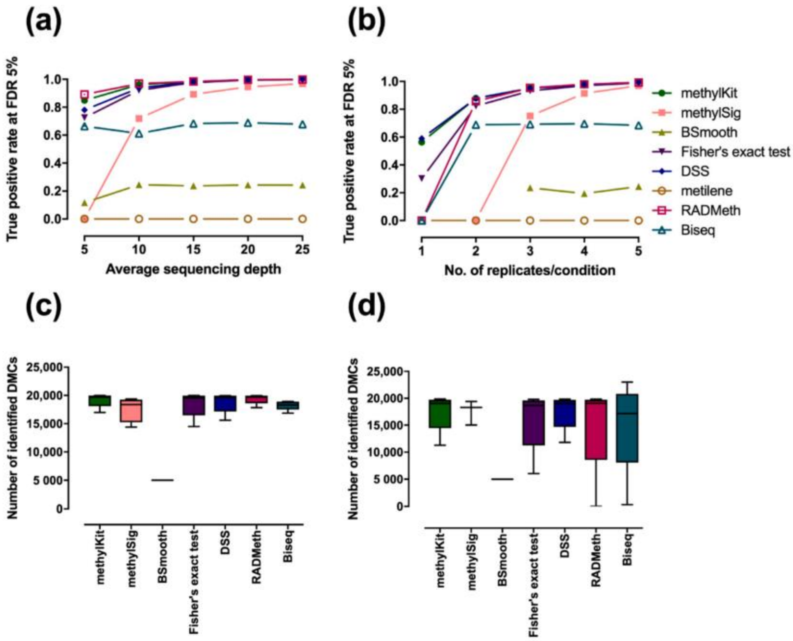
Segmentation: identification of DMR



- Different statistical models
- Fisher’s exact test, BSmooth, methylKit, methylSig, DSS, metilene, RADMeth, and Biseq

Different tools for DMR calling

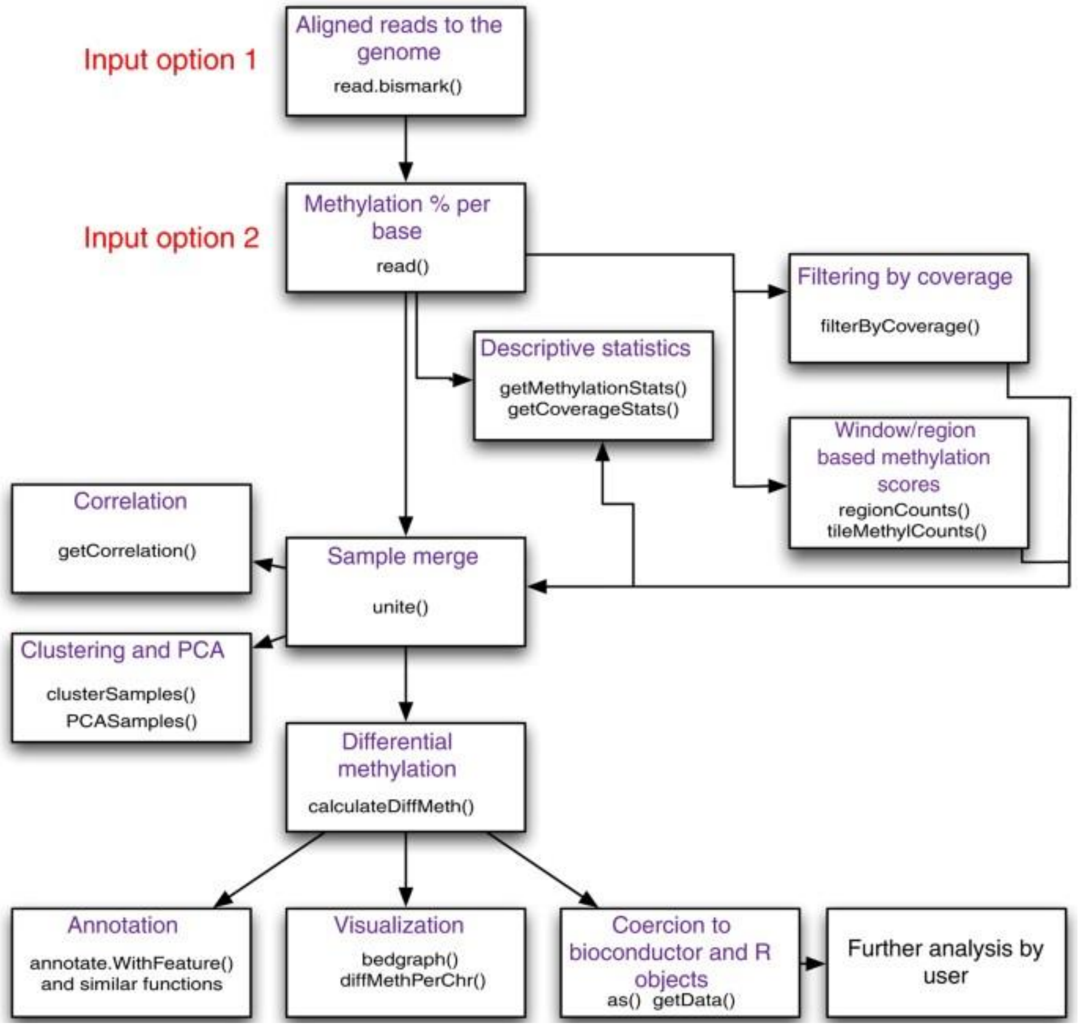
Tool	Version	Model Assumption	Differential Methylation Test	Segmentation	Language	Smoothing
Fisher's	1.8.2	-	Fisher's exact test	tilling window	R	No
BSmooth	1.8.2	binomial distribution	modified t-test	merging consecutive CpGs	R	Yes
methylKit	0.99.2	logistic regression	logistic regression test	tilling window or predefined regions	R	No
methylSig	0.4.4	beta-binomial model	likelihood ratio test	tilling window	R	No
DSS	2.12.0	Bayesian hierarchical model	Wald test	merging CpGs based on *p*-value	R	No
metilene	0.2–6	Nonparametric method	2D Kolmogorov–Smirnov	circular binary segmentation	C	No
RADMeth	-	beta-binomial regression	log-likelihood ratio test	correlation between *p*-value pairs within a bin	C++	No
Biseq	1.12.0	Beta regression model	Wald test	merging consecutive CpGs	R	Yes

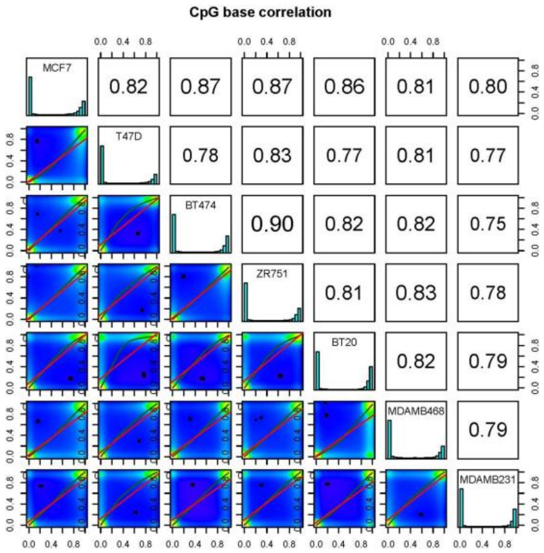


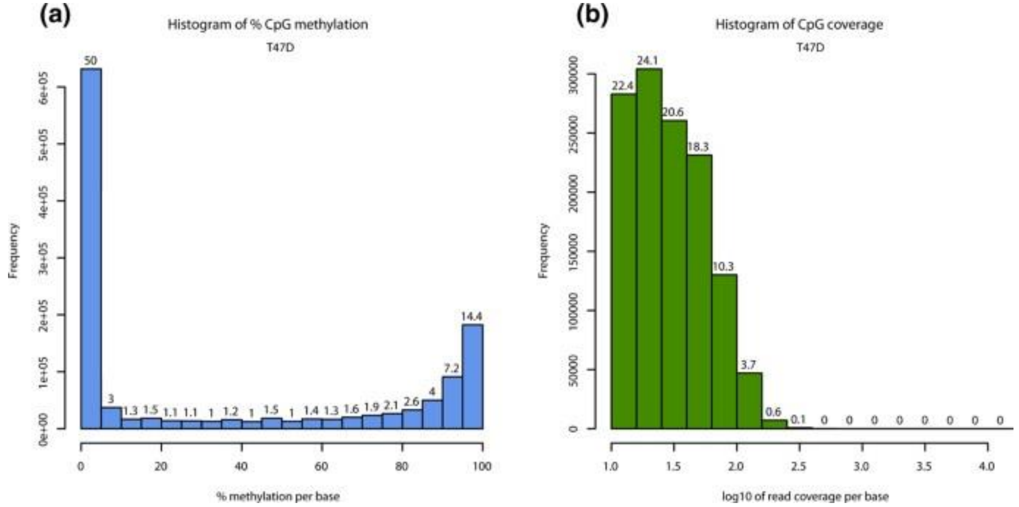
Comprehensive Evaluation of Differential Methylation Analysis Methods for Bisulfite Sequencing Data, Int. J. Environ. Res. Public Health 2021, 18(15), 7975;

- Notable variations among methods, and no single method consistently performed best in all benchmarking
- For DMR analysis, methylKit and Fisher's exact test covered more DMRs than other methods

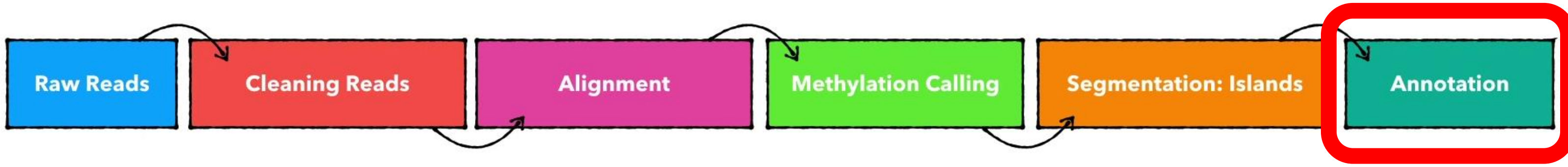
Methylkit for BS-seq data analysis







Annotation



Raw Reads

Cleaning Reads

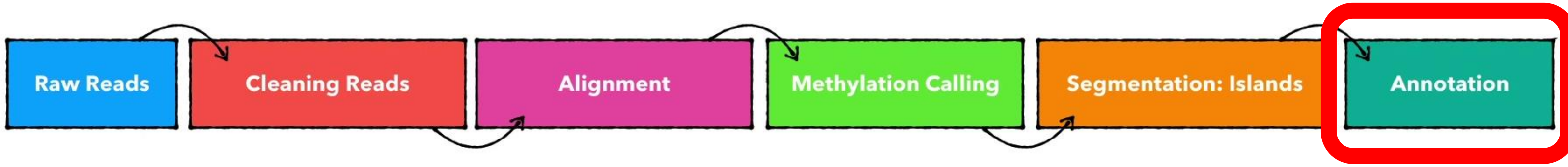
Alignment

Methylation Calling

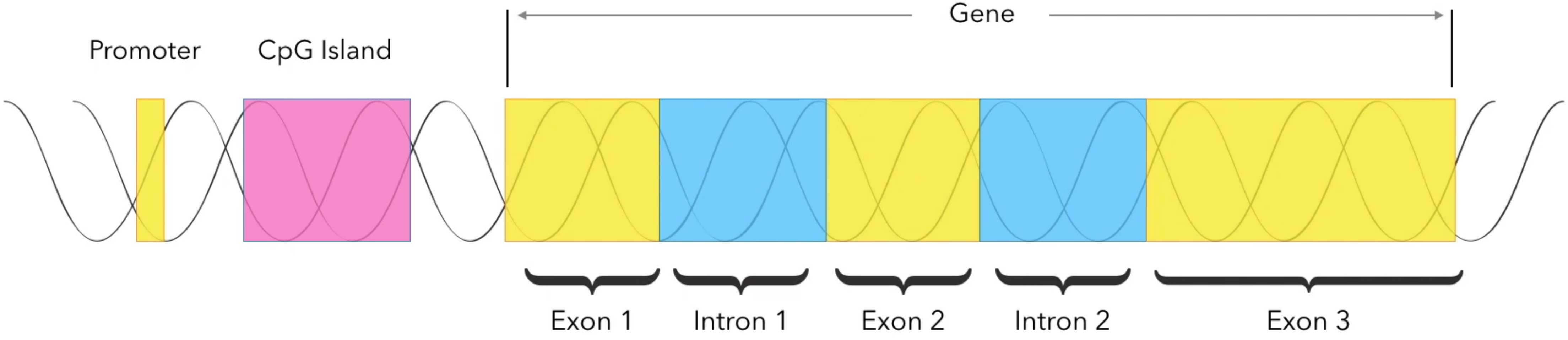
Segmentation: Islands

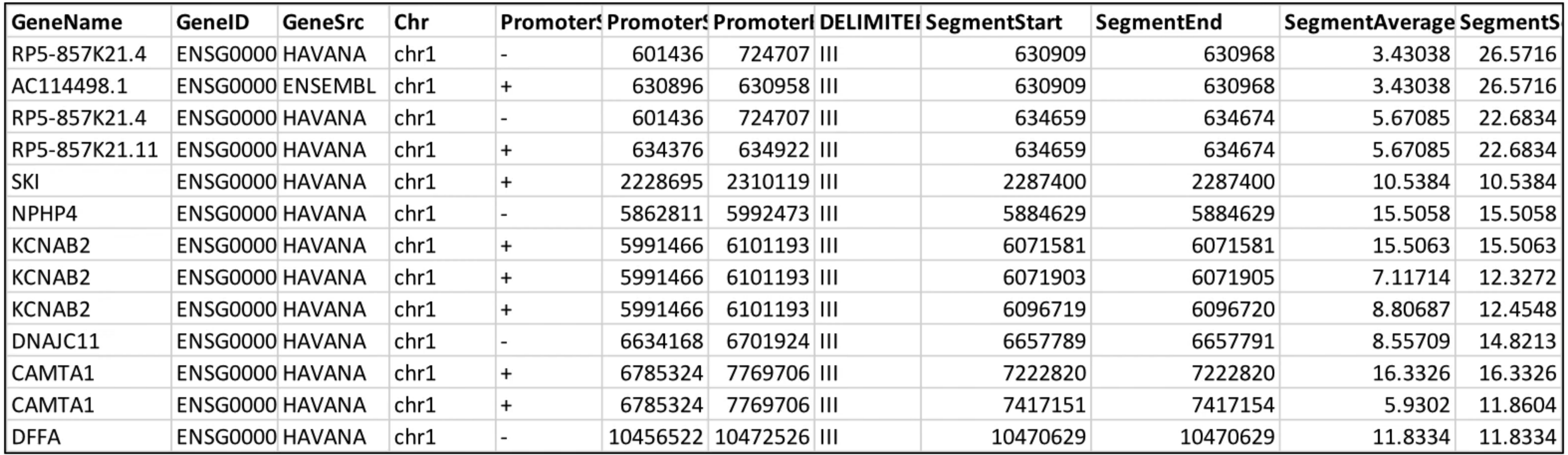
Annotation

- Function of differently methylated genes, comparing with known databases
- “clusterProfiler” is an example of R package to perform:
  - single-gene GO
  - KEGG enrichment.
  - GSEA enrichment analysis



Annotation





GeneName	GeneID	GeneSrc	Chr	PromoterS	PromoterS	PromoterI	DELIMITEI	SegmentStart	SegmentEnd	SegmentAverage	SegmentS
RP5-857K21.4	ENSG0000	HAVANA	chr1	-	601436	724707	III	630909	630968	3.43038	26.5716
AC114498.1	ENSG0000	ENSEMBL	chr1	+	630896	630958	III	630909	630968	3.43038	26.5716
RP5-857K21.4	ENSG0000	HAVANA	chr1	-	601436	724707	III	634659	634674	5.67085	22.6834
RP5-857K21.11	ENSG0000	HAVANA	chr1	+	634376	634922	III	634659	634674	5.67085	22.6834
SKI	ENSG0000	HAVANA	chr1	+	2228695	2310119	III	2287400	2287400	10.5384	10.5384
NPHP4	ENSG0000	HAVANA	chr1	-	5862811	5992473	III	5884629	5884629	15.5058	15.5058
KCNAB2	ENSG0000	HAVANA	chr1	+	5991466	6101193	III	6071581	6071581	15.5063	15.5063
KCNAB2	ENSG0000	HAVANA	chr1	+	5991466	6101193	III	6071903	6071905	7.11714	12.3272
KCNAB2	ENSG0000	HAVANA	chr1	+	5991466	6101193	III	6096719	6096720	8.80687	12.4548
DNAJC11	ENSG0000	HAVANA	chr1	-	6634168	6701924	III	6657789	6657791	8.55709	14.8213
CAMTA1	ENSG0000	HAVANA	chr1	+	6785324	7769706	III	7222820	7222820	16.3326	16.3326
CAMTA1	ENSG0000	HAVANA	chr1	+	6785324	7769706	III	7417151	7417154	5.9302	11.8604
DFFA	ENSG0000	HAVANA	chr1	-	10456522	10472526	III	10470629	10470629	11.8334	11.8334

- Akalin, A., Kormaksson, M., Li, S. *et al*. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol* **13**, R87 (2012). https://doi.org/10.1186/gb-2012-13-10-r87
- Nasibeh C. *et al.,* Bioinformatic tools for DNA methylation and histone modification: A survey, *Genomics*, 2021. https://doi.org/10.1016/j.ygeno.2021.03.004
- Piao, Y.; Xu, W.; Park, K.H.; Ryu, K.H.; Xiang, R. Comprehensive Evaluation of Differential Methylation Analysis Methods for Bisulfite Sequencing Data. *Int. J. Environ. Res. Public Health* **2021**, *18*, 7975. https://doi.org/10.3390/ijerph18157975
- Shizhao Li, Trygve O. Tollefsbol, DNA methylation methods: Global DNA methylation and methylomic analyses, Methods, 2021, https://doi.org/10.1016/j.ymeth.2020.10.002.