

Strategies for analyzing BS-seq data

Altuna Akalin

 @AltunaAkalin

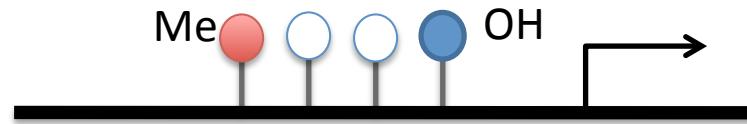
Outline

- Epigenetic dysregulation in acute myeloid leukemia (AML)
 - What is epigenetics
 - What is AML
 - Mutations causing epigenetic dysregulation and their effects
- Strategies for analyzing bisulfite-seq data
 - Align + Process raw data
 - Find interesting regions (changes in methylation)
 - Annotate

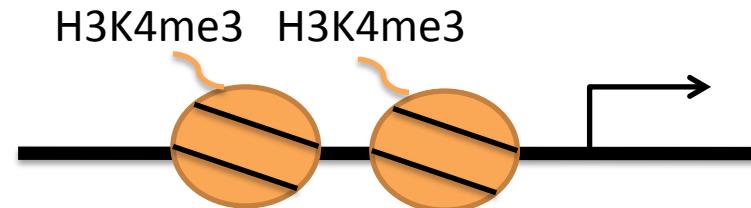
Epigenetics

Epigenetics is the study of heritable changes other than DNA sequence that affect gene activity.

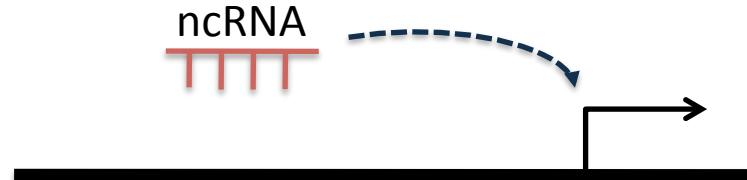
- DNA Methylation (and other base modifications)



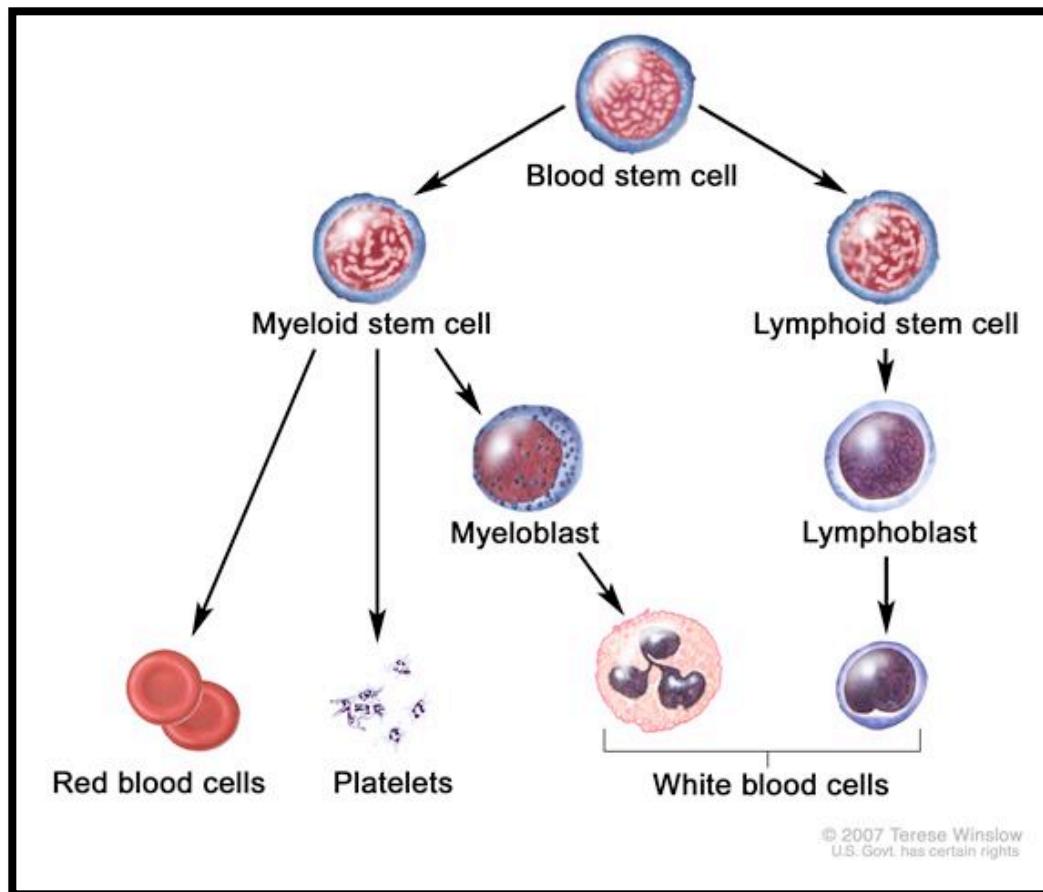
- Histone Modifications



- RNA-Associated control



Acute Myeloid Leukemia (AML)



From cancer.gov

- A malignant disease of the bone marrow where precursor cells arrested in early stage of development

How do normal cells transform to acute myeloid leukemia ?

Class I mutations

Growth advantage by
altering signaling
pathways

Class II mutations

Alterations in
differentiation control
genes

Class III ?

Alterations in
hydroxymethylation
IDH1, IDH2 & TET2

Class IV ?

Alterations in
epigenetic modifiers
MLL, DNMT3A

① Comparison of class III mutants
against class IV mutants

AML subtypes

IDH mutant

Known to induce DNA hypermethylation
via altering hydroxymethylation pathway

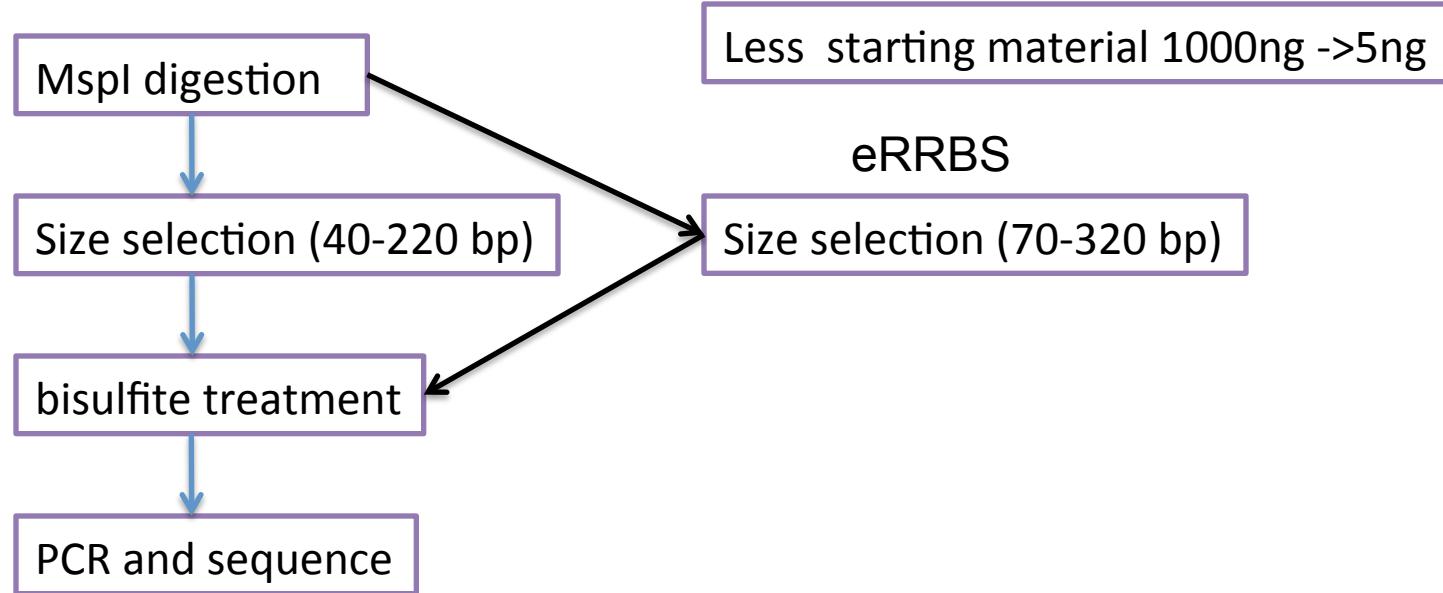
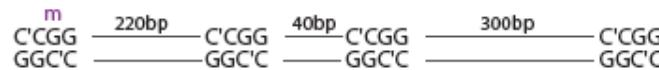
Class III

MLL rearrangement

fusion protein disrupting H3K4
methyltransferase domain

Class IV

RRBS (Reduced Representation Bisulfite Sequencing)



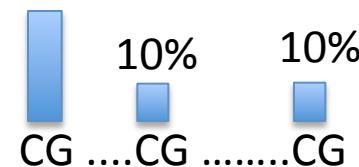
Bisulfite treatment and sequencing

A C_m G C_m G A C A C A

| | | | | | | | | | | |

A C G C G A T A T A

70%

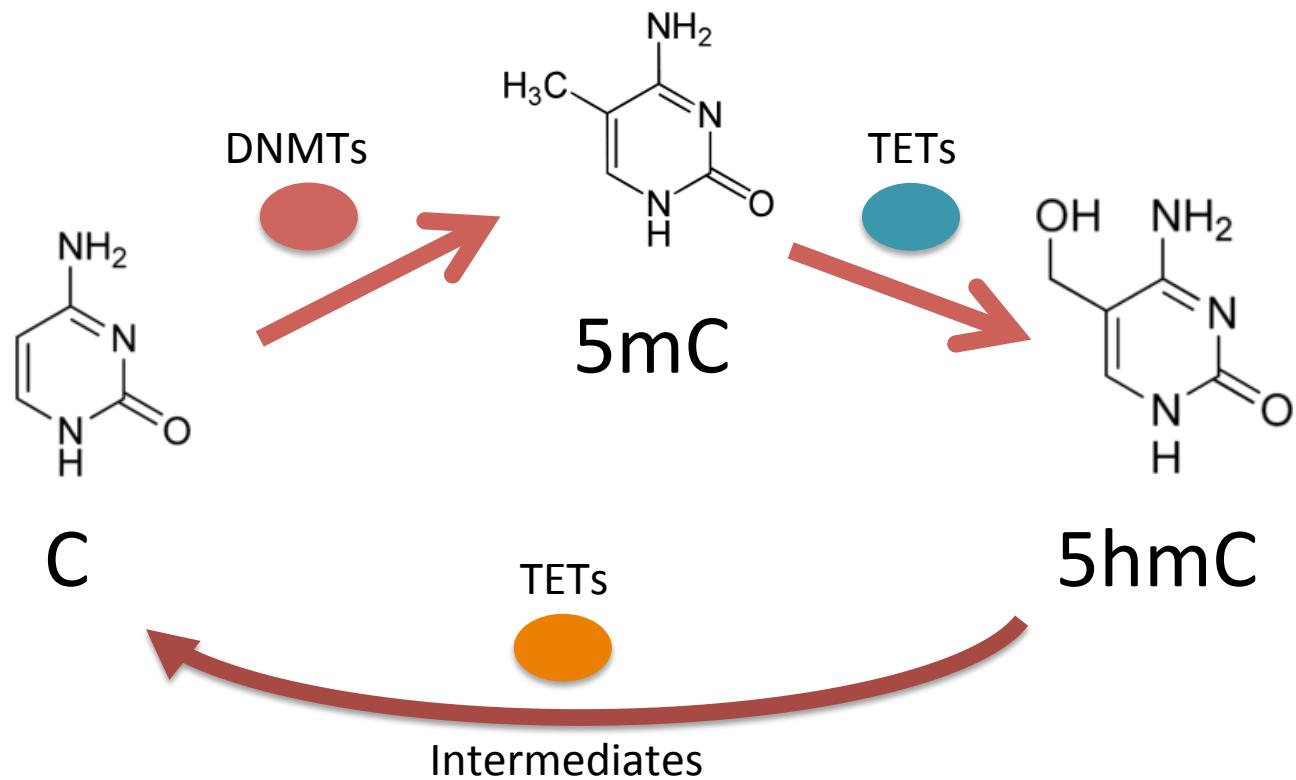


Differentially methylated CpGs (DMCs): IDH-mut vs normal

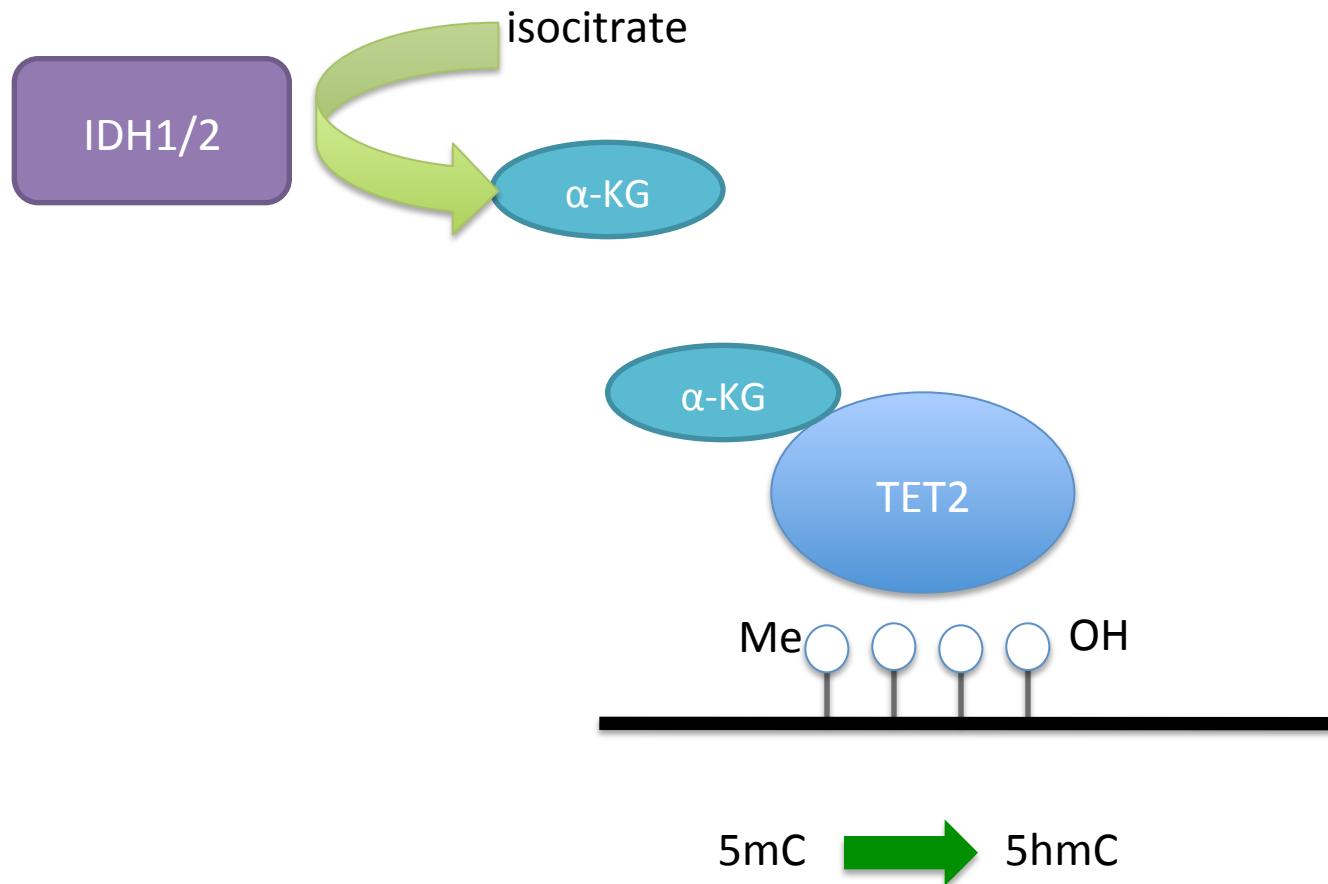
There is an exercise on this

② Focus on mutations that impair
hydroxymethylation pathway

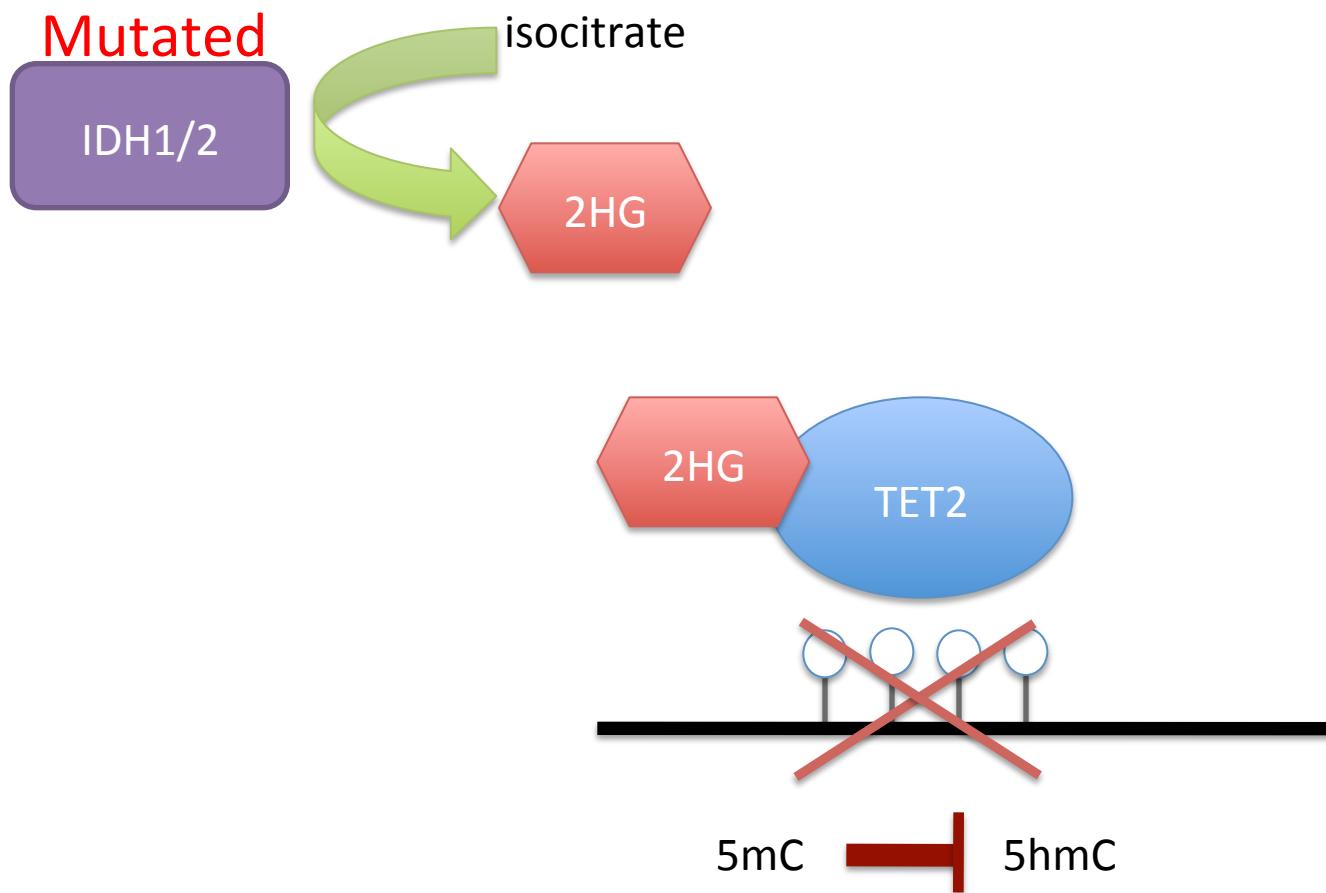
Demethylation



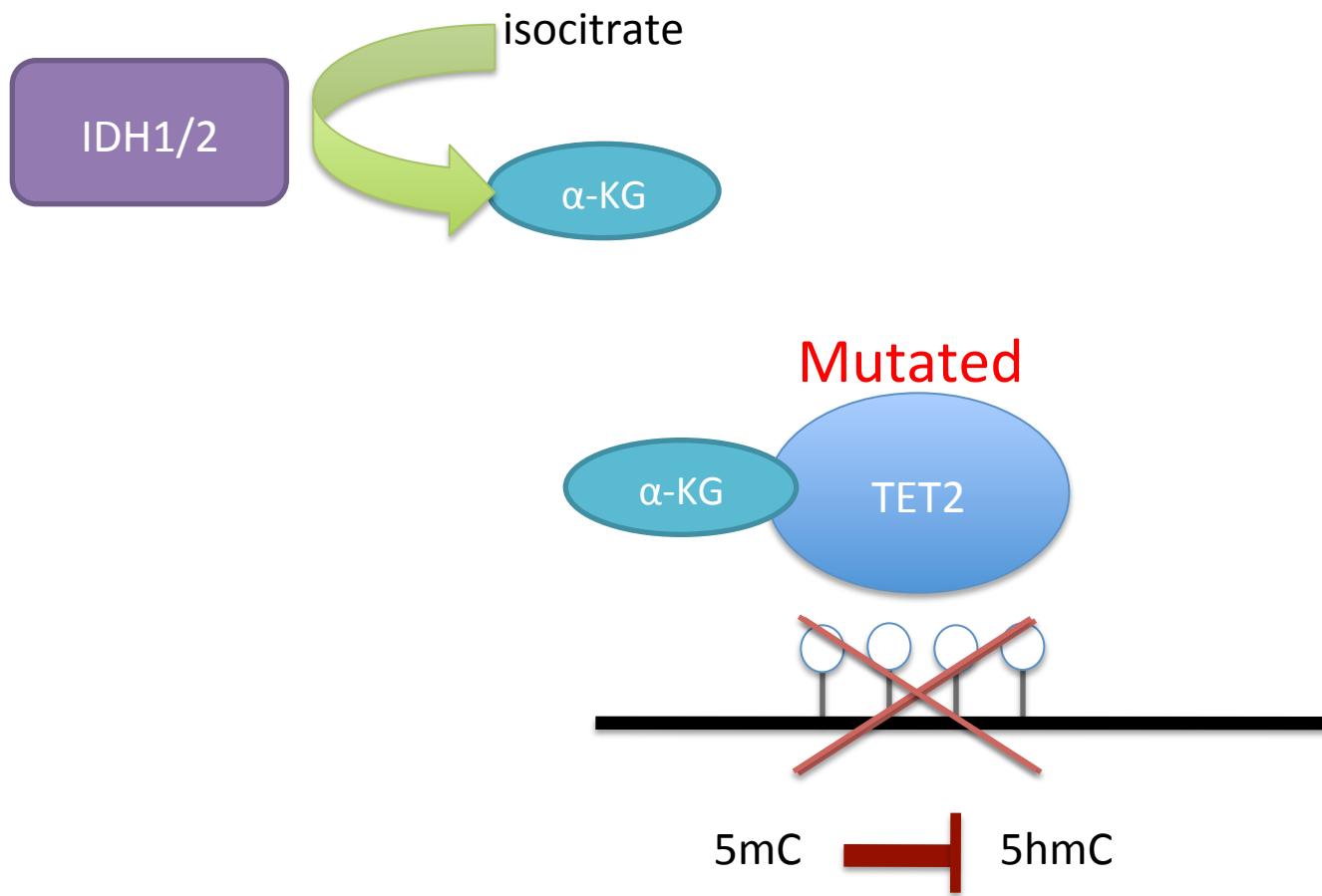
Conversion of 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC)



IDH1/2 mutations disrupt 5hmC production



TET2 mutations disrupt 5hmC production

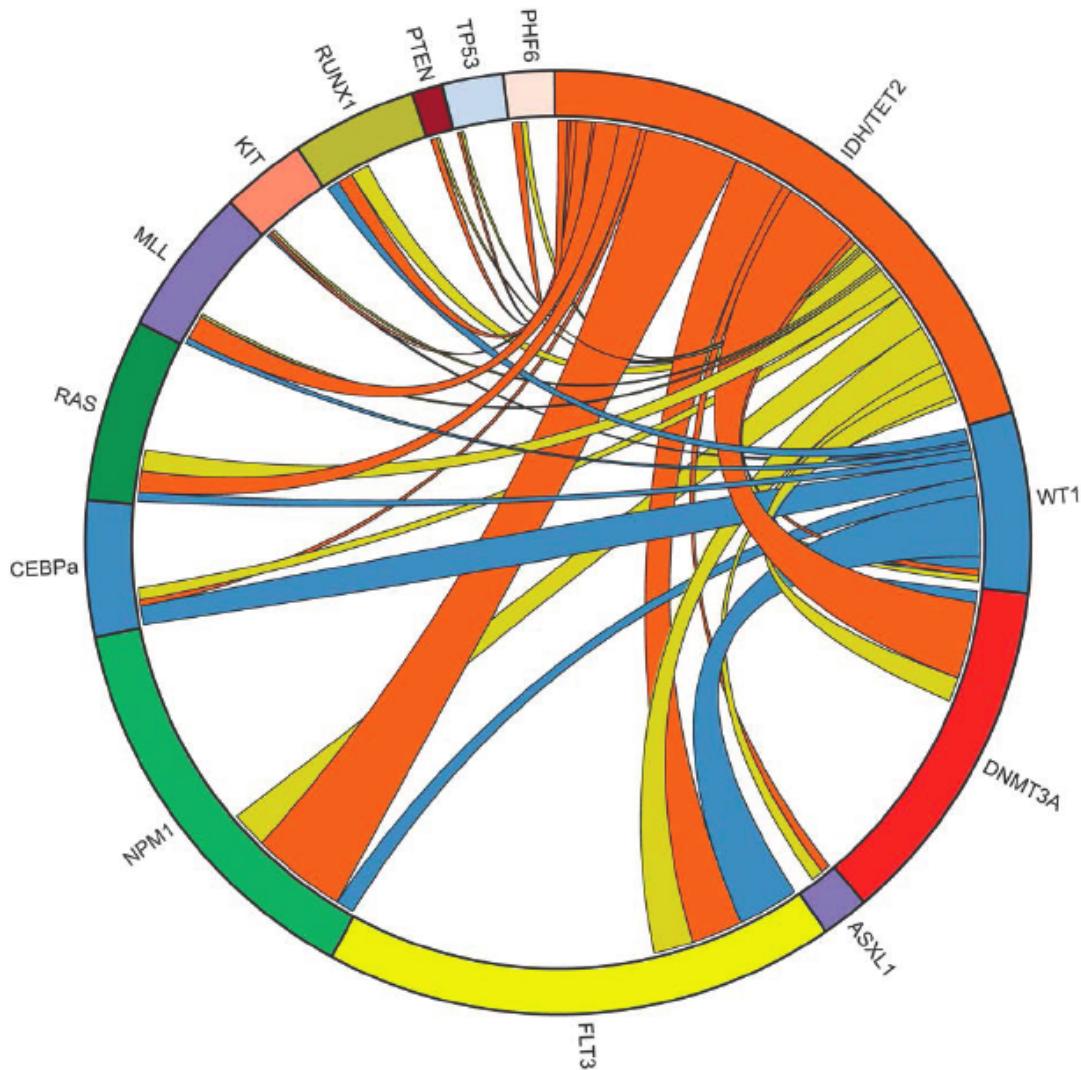


Mutations disrupting 5hmC production in AML

- IDH1/2,TET2 and WT1 mutations are mutually exclusive in AML => hints that they are on related pathway

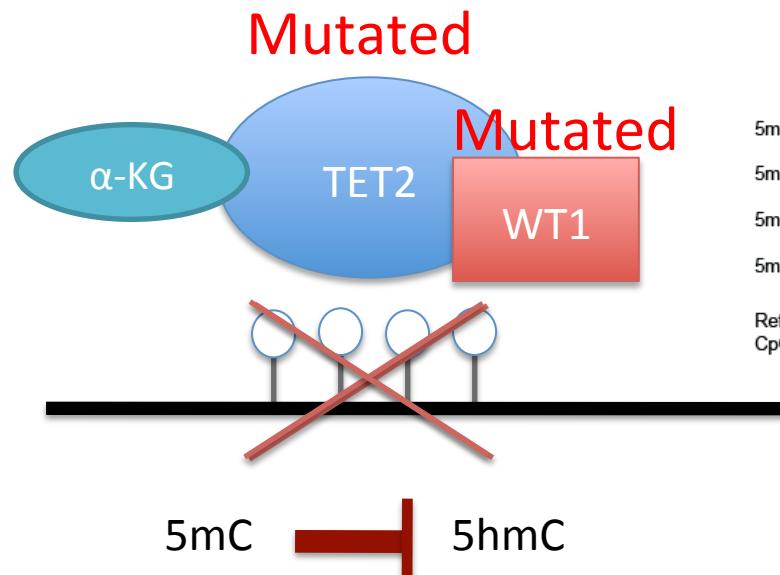
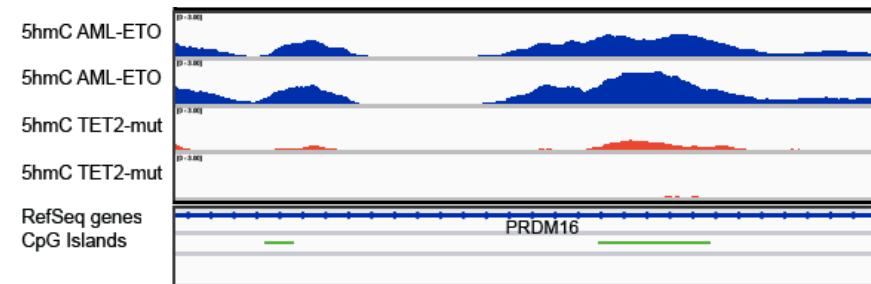
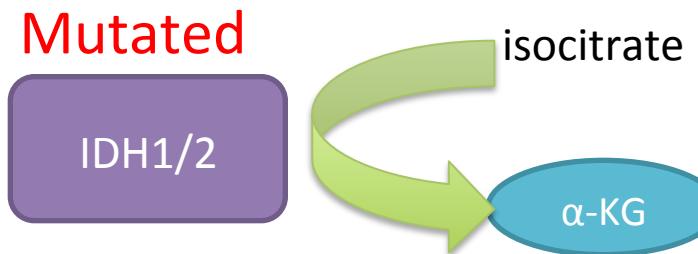
How these mutations affect genome-wide 5hmC profiles?

Does WT1 mutant has the same effect on 5hmC?

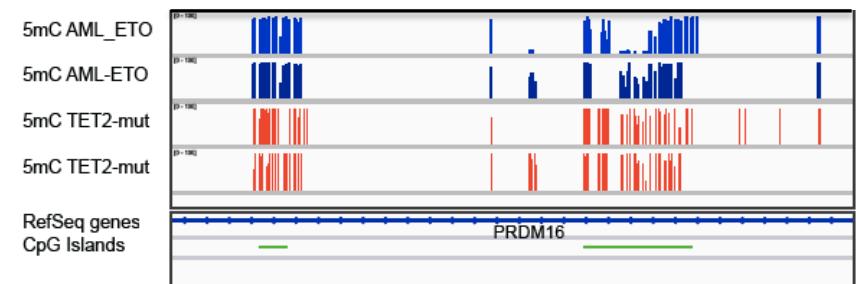


WT1 mutation could also cause methylation dysregulation

5hmC profiling

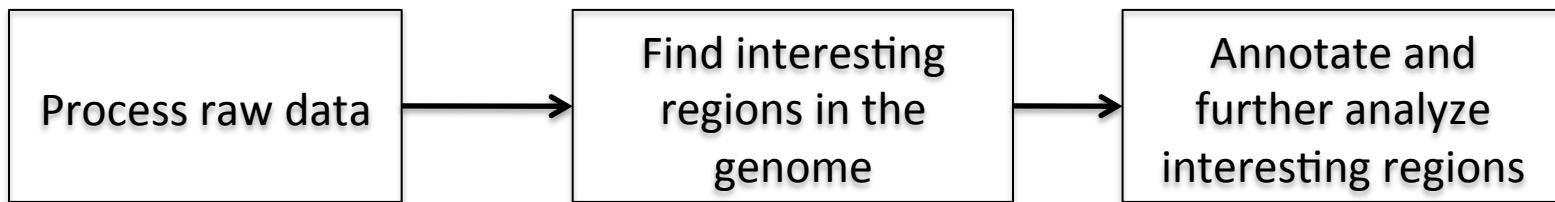


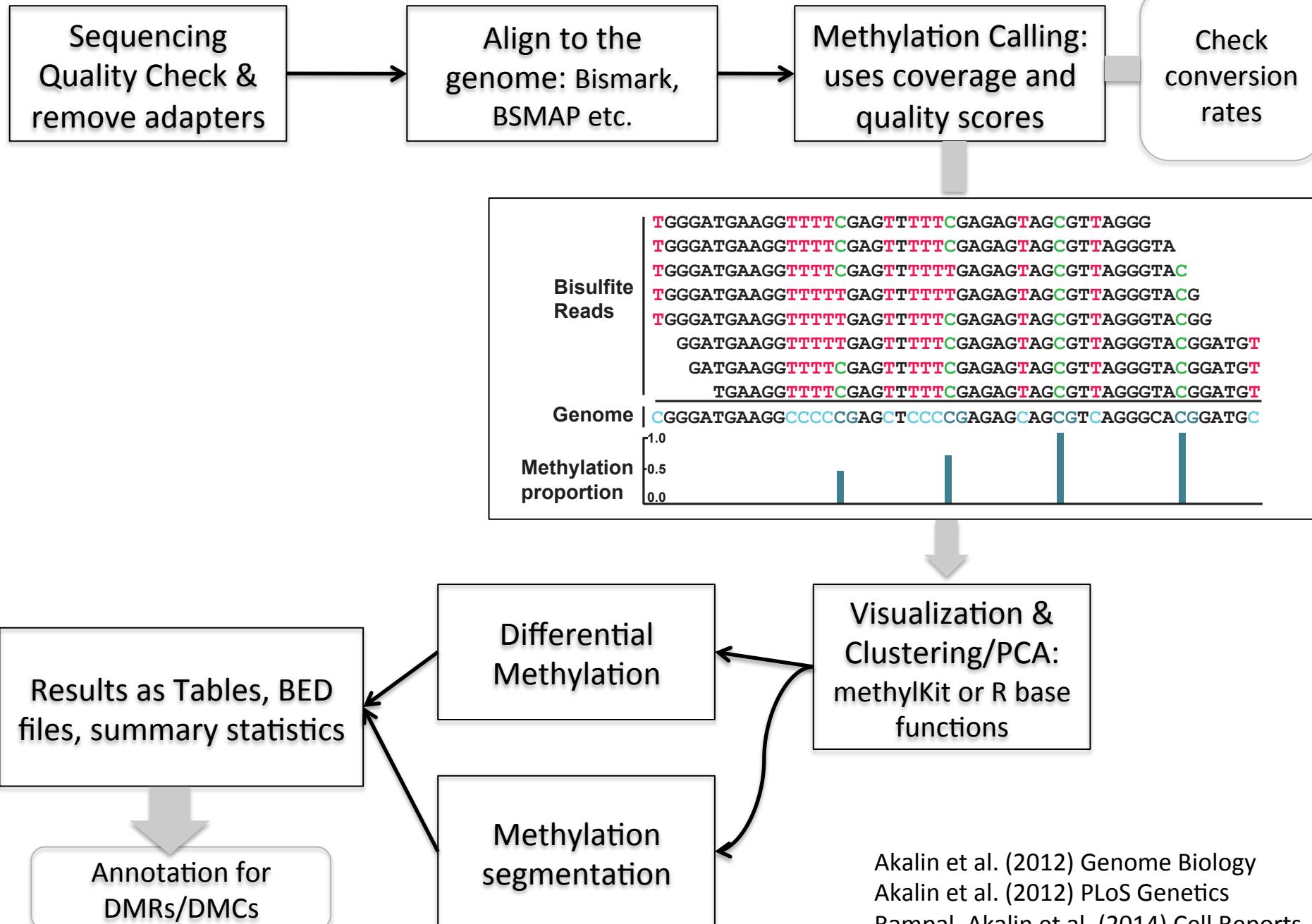
5mC profiling



Strategies for analyzing bisulfite-seq data

General workflow for HT-seq analysis





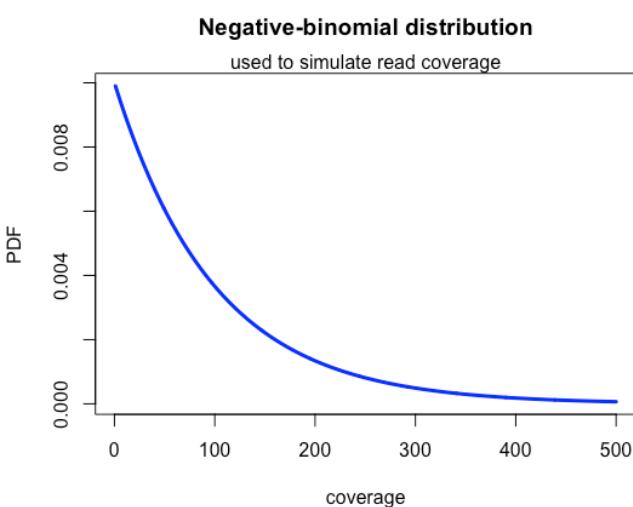
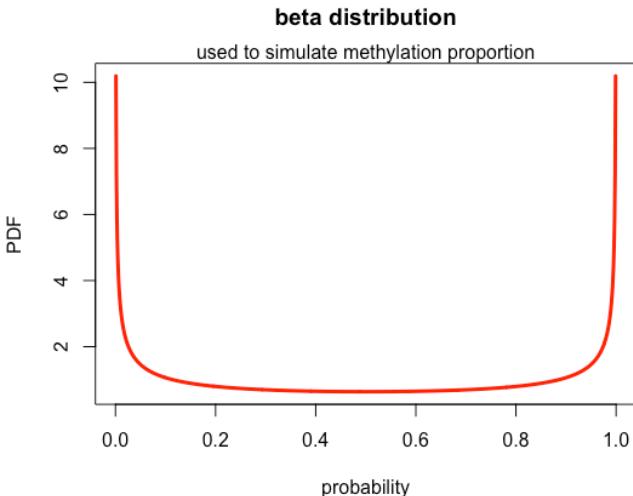
Differential methylation

- Differentially methylated regions could be associated with regulatory changes
- Abnormal methylation is hallmark of cancer

software	method	Language/Platform	reference
methylKit	Logistic regression with/without overdispersion correction / Fisher's Exact test	R package	(Akalin et al. 2012b)
BSseq	Smoothing +Linear regression + Empirical Bayes	R package	(Hansen et al. 2012)
BiSeq	Beta regression	R package	(Hebestreit et al. 2013)
DSS	Beta-binomial with Empirical Bayes	R package	(Feng et al. 2014)
MOABS	Beta-binomial with Empirical Bayes	C/C++/Command line	(Sun et al. 2014)
RADMeth	Beta-binomial regression	C/C++/ Command line	(Dolzhenko and Smith 2014)

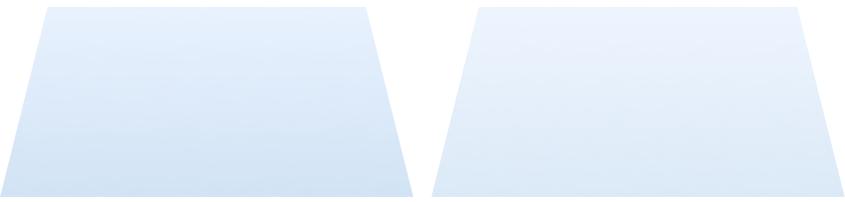
- Over-dispersion correction & effect-size cutoff

Comparison of Differential methylation methods



Methylation
Coverage

$$B(\mu = 0.1, \phi = 10) \quad B(\mu = 0.4, \phi = 10)$$

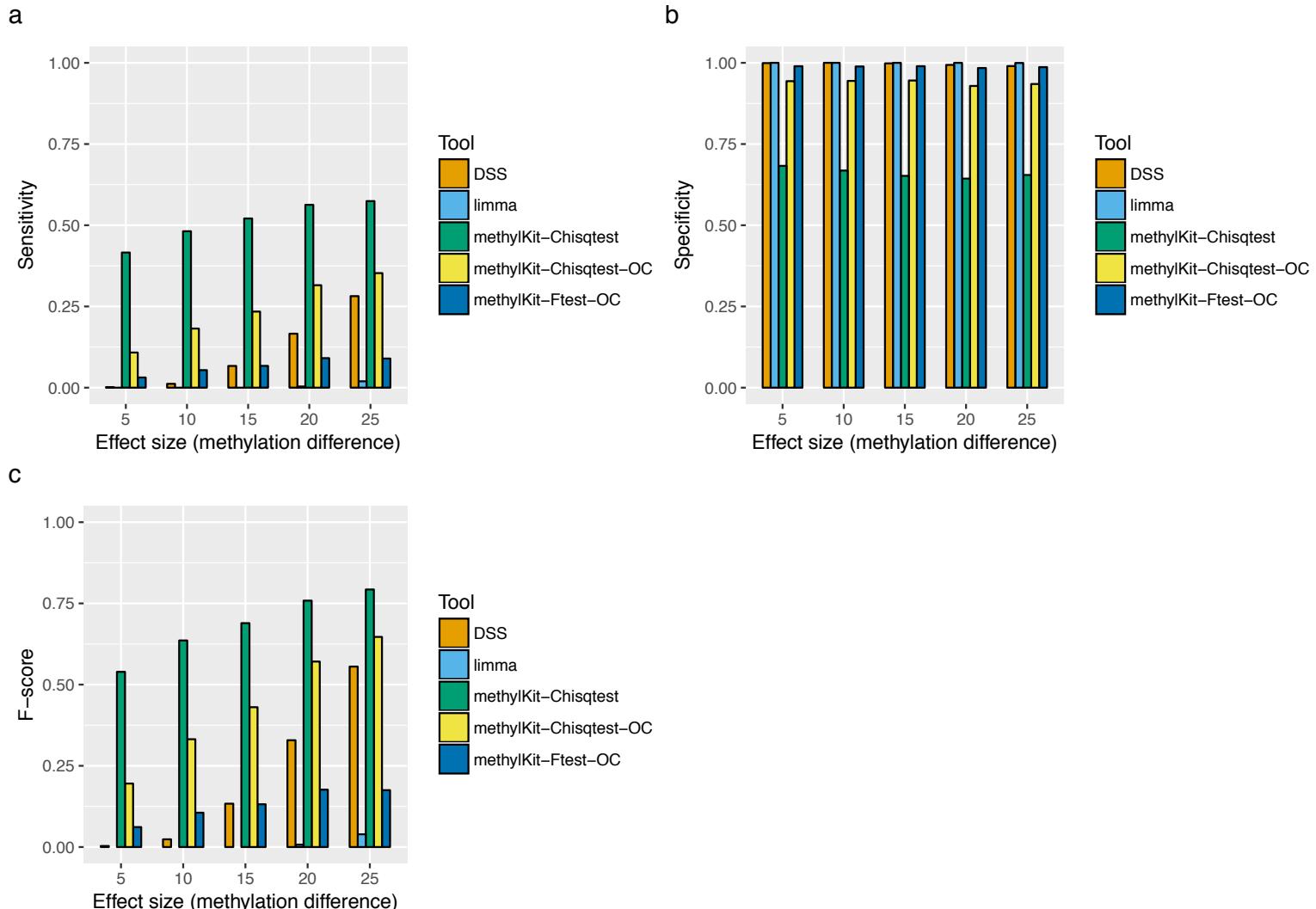


Test 1	Test 2	Ctrl 1	Ctrl 2
0.14	0.24	0.35	0.54
15	40	100	45

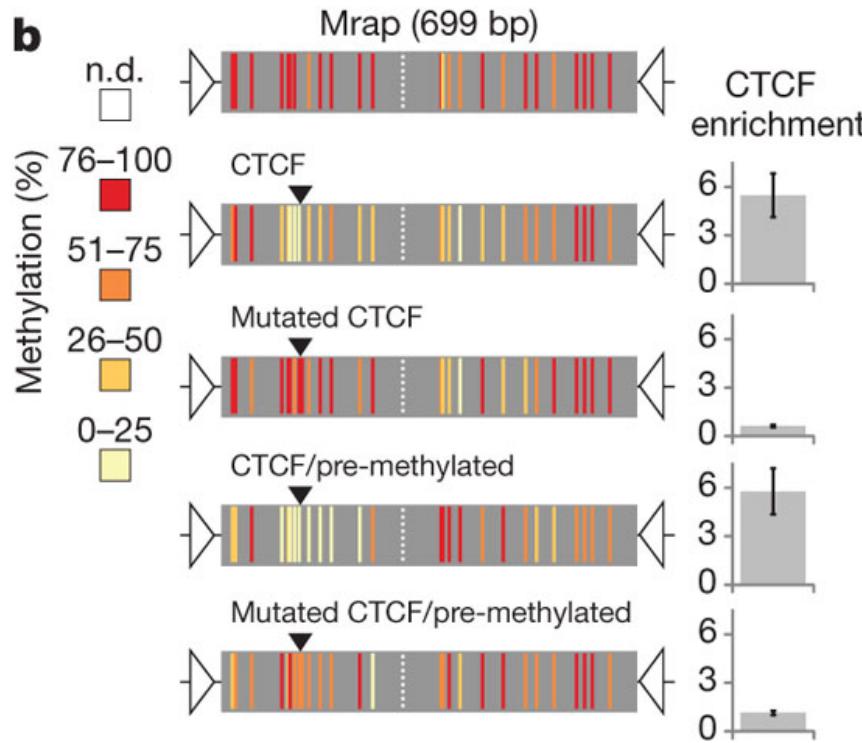
$$NB(r = 1, p = 0.01)$$

- Be cautious when simulated data is used to compare methods

Comparison of Differential methylation methods: Simulated data



Comparison of Differential methylation methods: CTCF binding dynamics



CTCF binding dynamics and methylation

	Peak has DMC	Peak has No DMC
CTCF Peak Lost/Gained	TP	FN
CTCF Peak No change	FP	TN

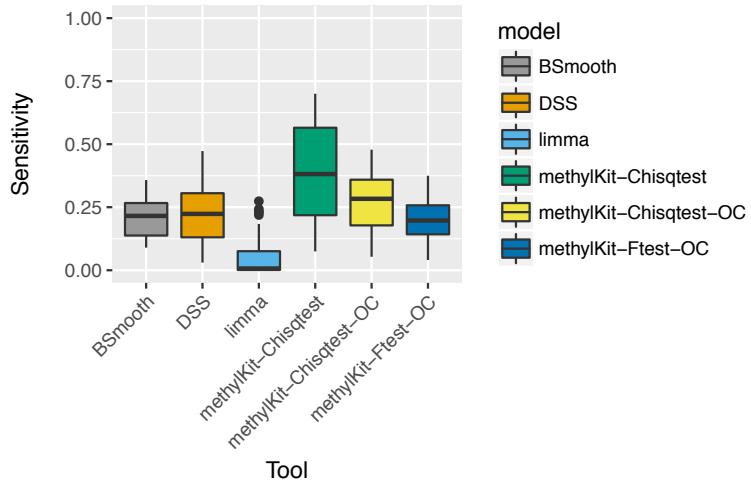
- Compared all possible pairs for 19 Human cell lines
- Calculated performance metrics

Caveats:

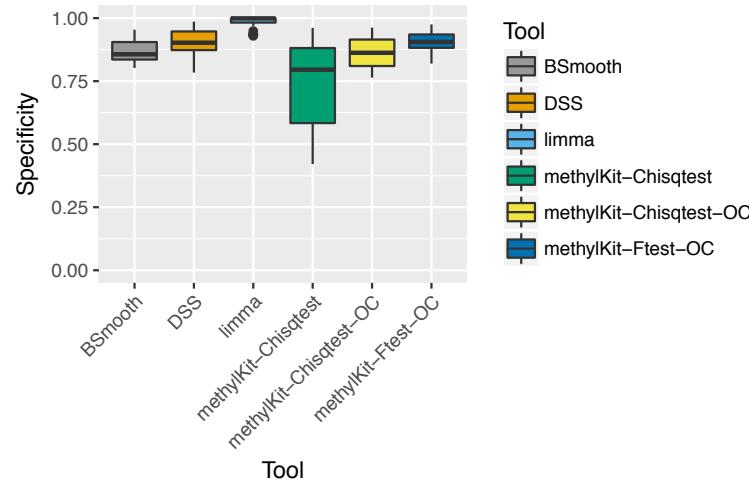
- We do not expect to cover all CTCF sites.
- Even if covered, CpGs might be far from the motif
- Due the sparse RRBS coverage, DMR approaches not well suited

Comparison of Differential methylation methods: CTCF binding dynamics

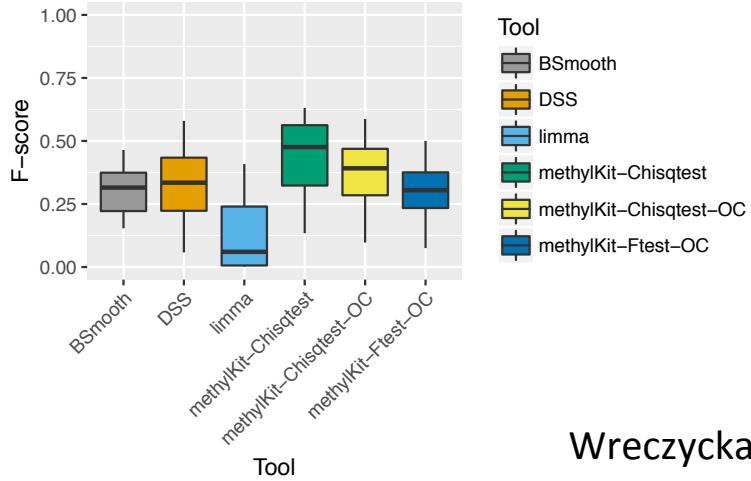
a



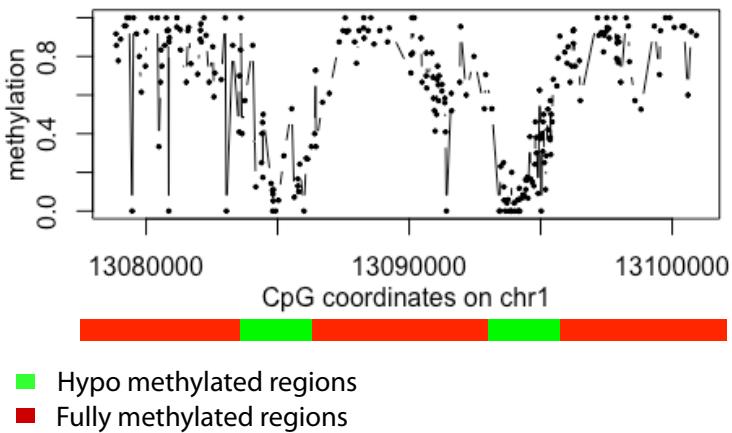
b



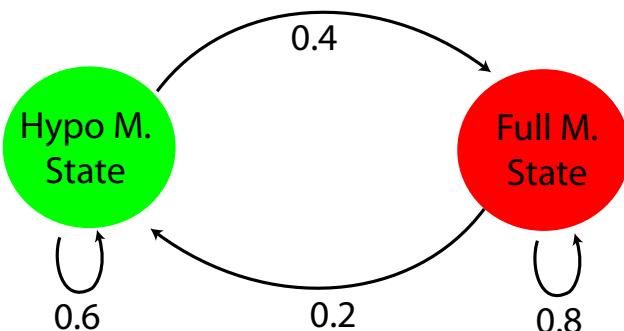
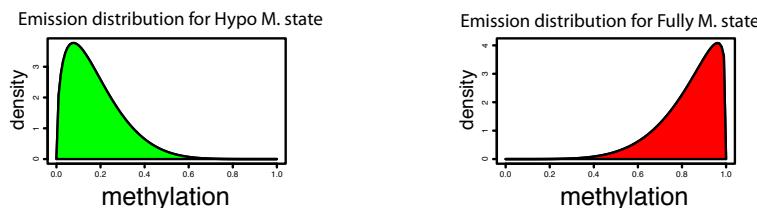
c



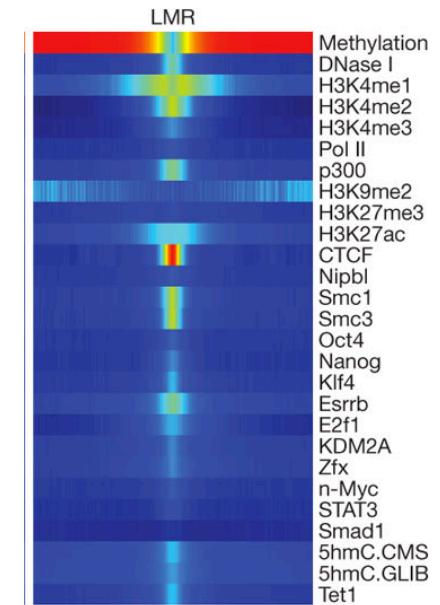
Segmentation of DNA methylation profiles



Dips in methylation profile marks regulatory regions

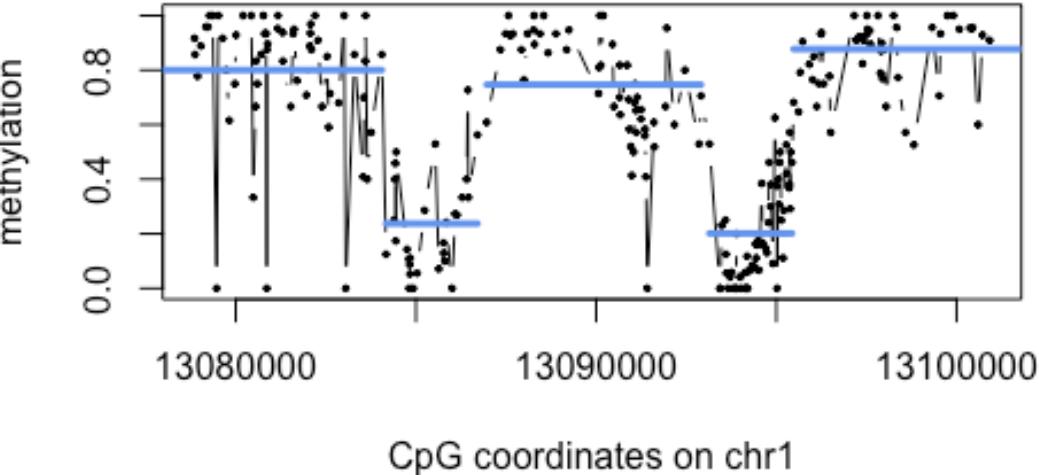


HMM produces the most probable state for each CpG

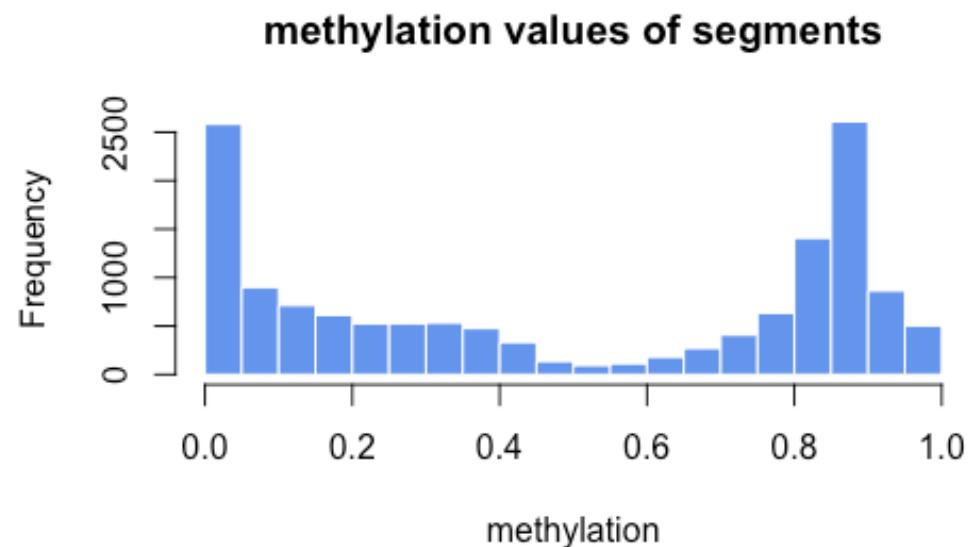


Stadler, 2011, Nature

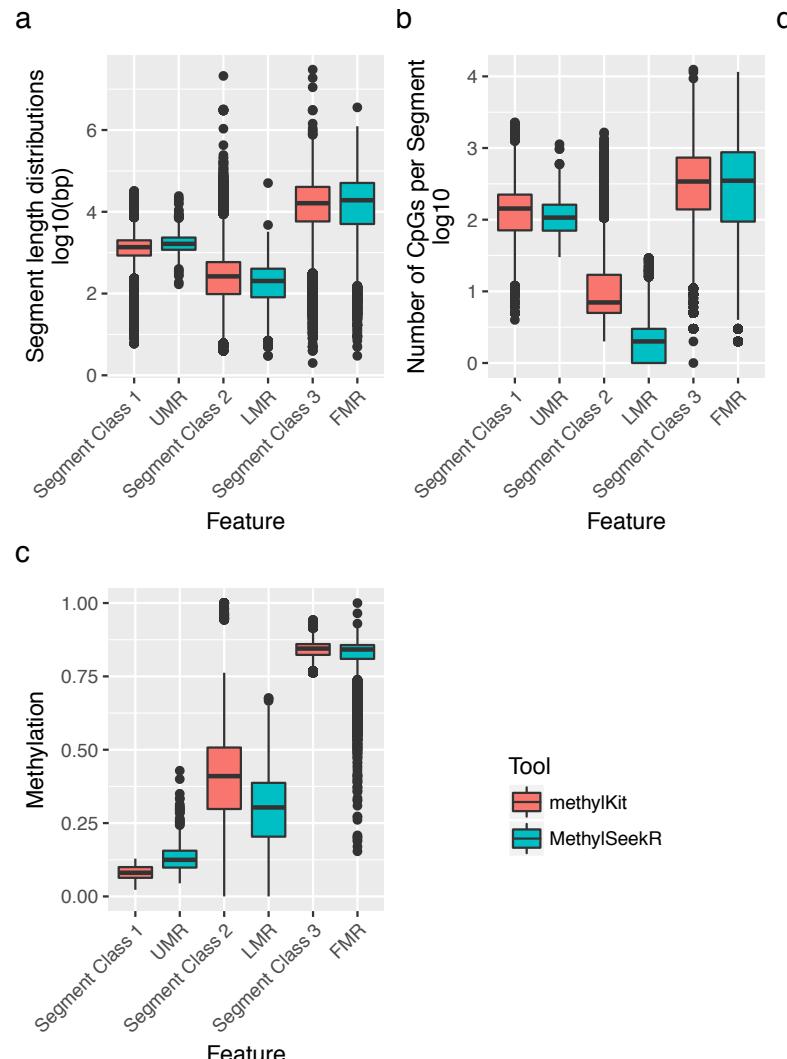
Change-point analysis for methylome segmentation

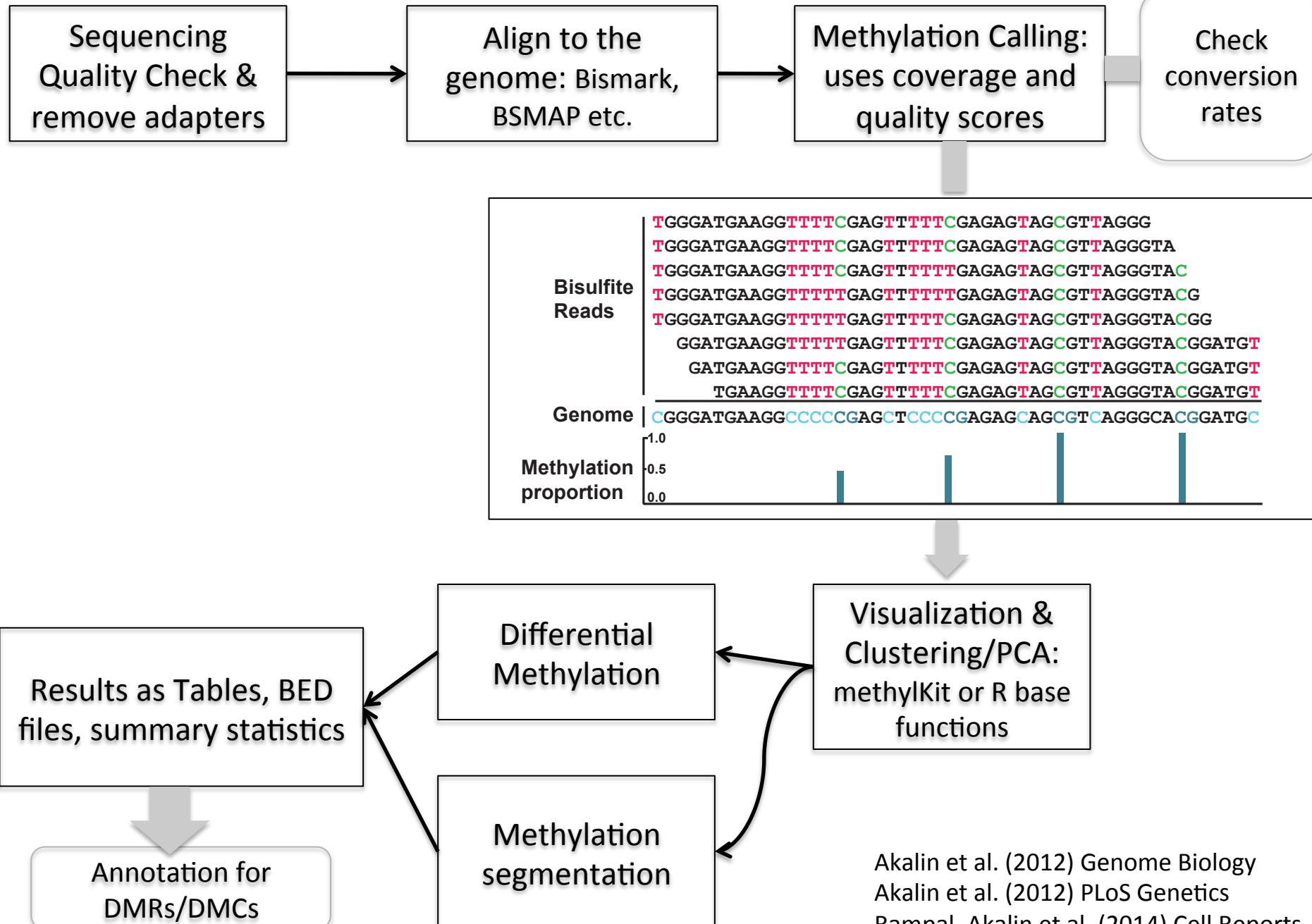


- Find change-points
- Cluster segments with similar methylation profiles using mixture modelling

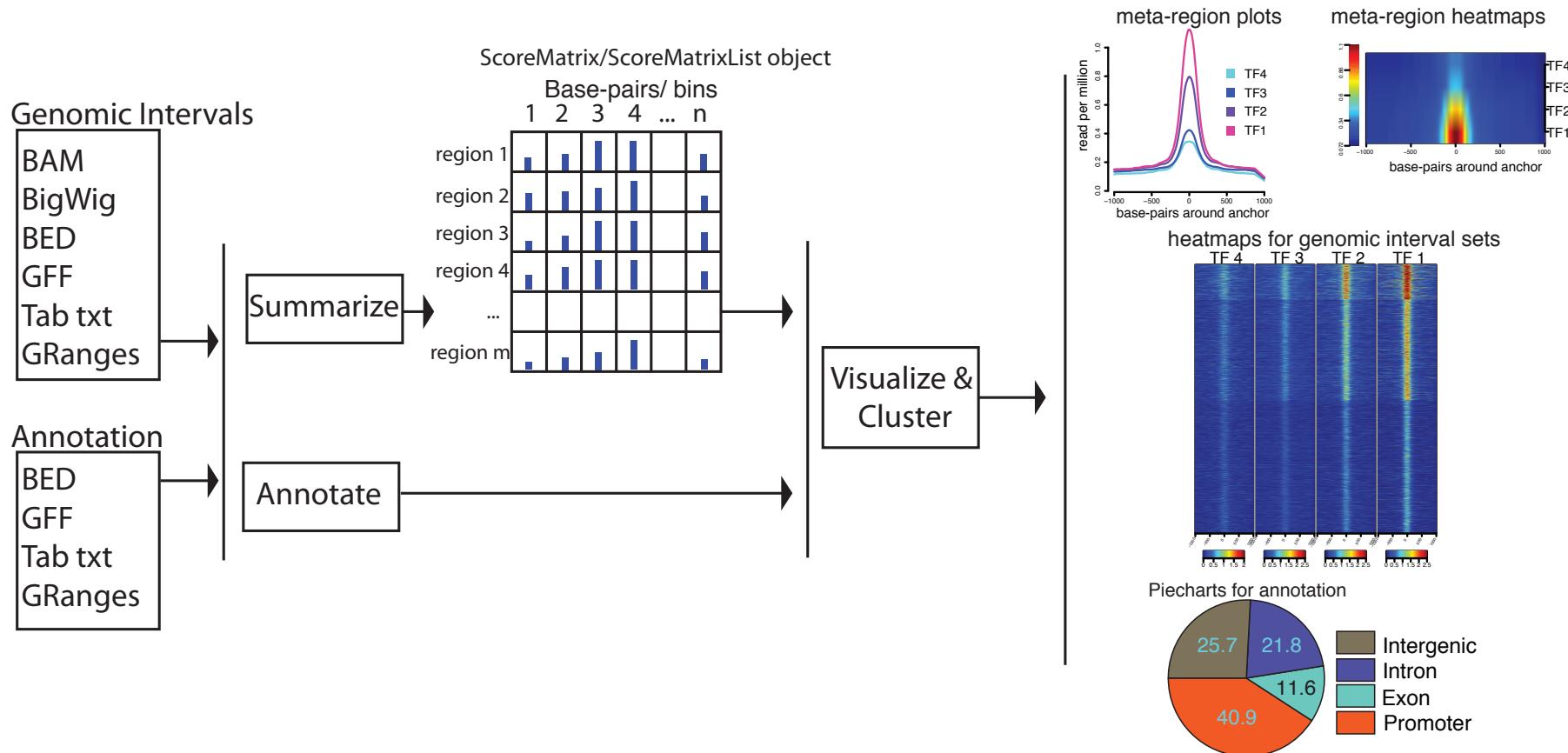


Comparison of segmentation methods





Annotation and integration with other genome-wide datasets



Acknowledgements



Weill Cornell Medical College

- Christopher Mason
- Ari Melnick



- Maria E. Figueroa



- Ross Levine

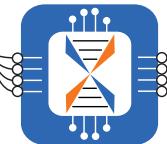
The slides are at: <http://www.slideshare.net/altunaakalin/>



Friedrich Miescher Institute
for Biomedical Research

- Dirk Schübeler

BIOINFORMATICS PLATFORM



- Vedran Franke
- Kasia Wreczycka
- Alex Godschan

Funding:



Federal Ministry
of Education
and Research