Universität Rostock Traditio et Innovatio

Universitätsmedizin Rostock

Integrative Bioinformatics and Systems Biology

Dr. Mohamed Hamed

Lecture 10
Introduction to Epigenetic
DNA Methyaltion

Definition

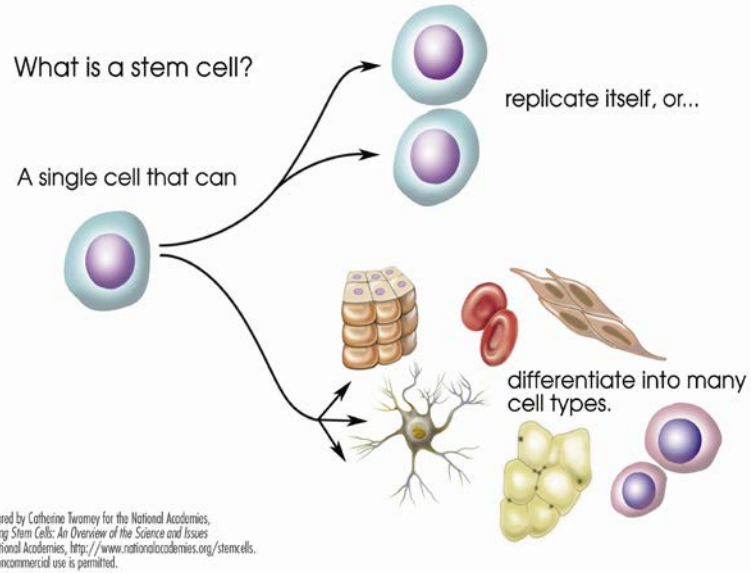
- Epigenetics

Is the study of heritable changes in gene function that cannot be explained by changes in DNA sequence (Riggs, 1996).

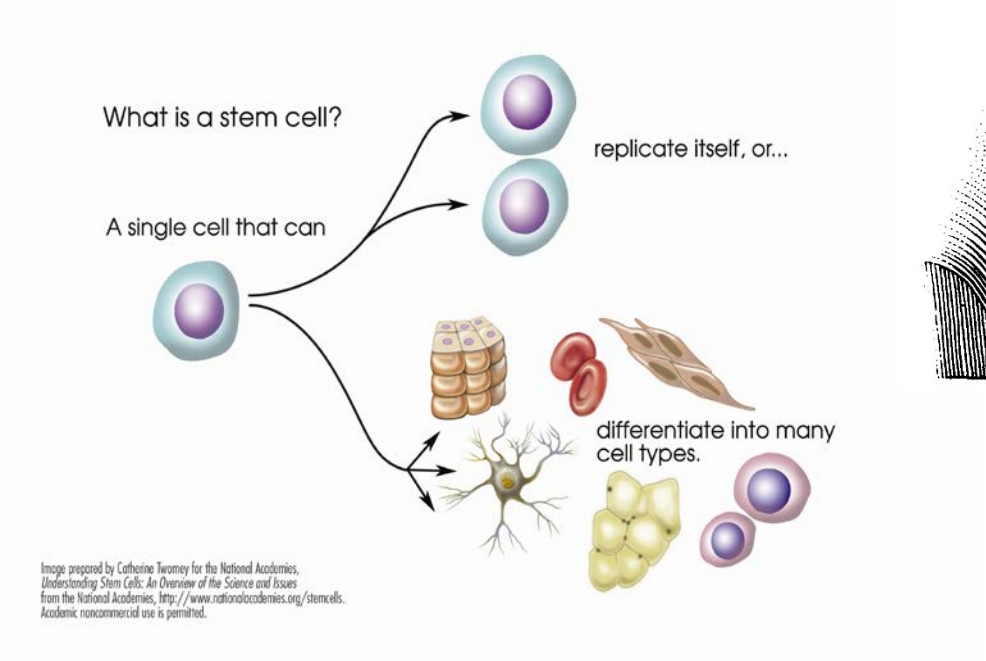
Definition II

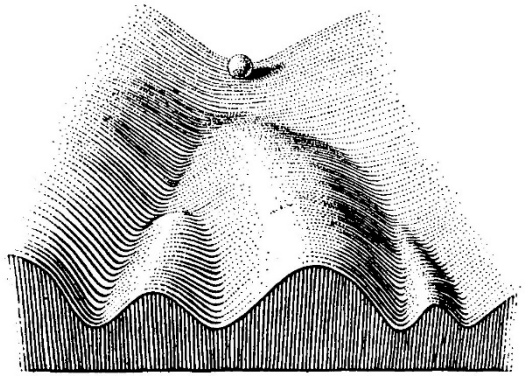
- The majority of cells of an individual have an identical or almost identical genome (exception: e.g. cancer cells)
- However, the cells are very heterogeneous, they have different functions, morphology, ...
- The part of the heritable information that encodes this difference is called epigenetic.

Stages of Cell development



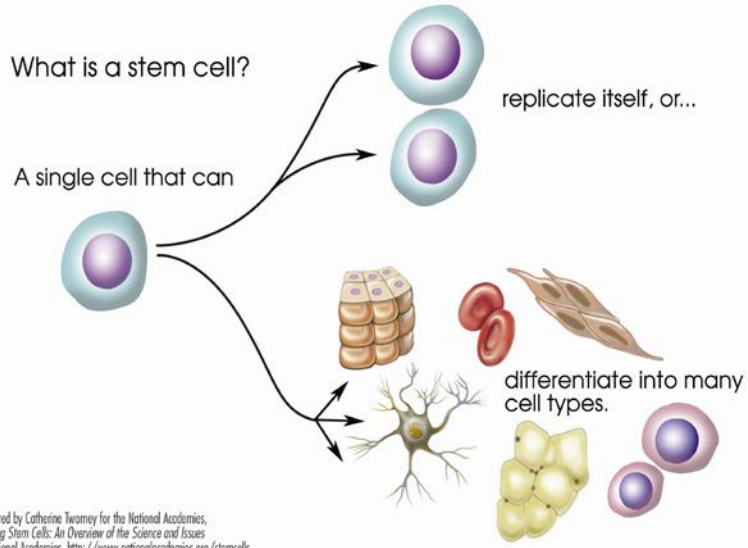
Stages of Cell development

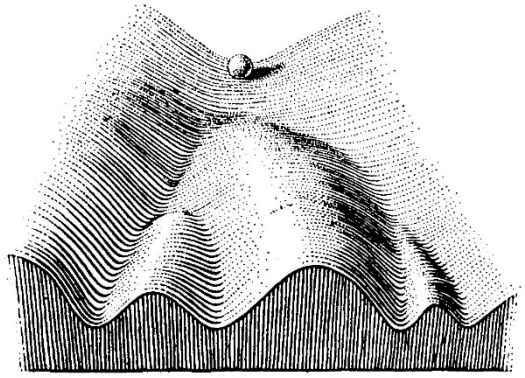




Conrad H. Waddington 1956

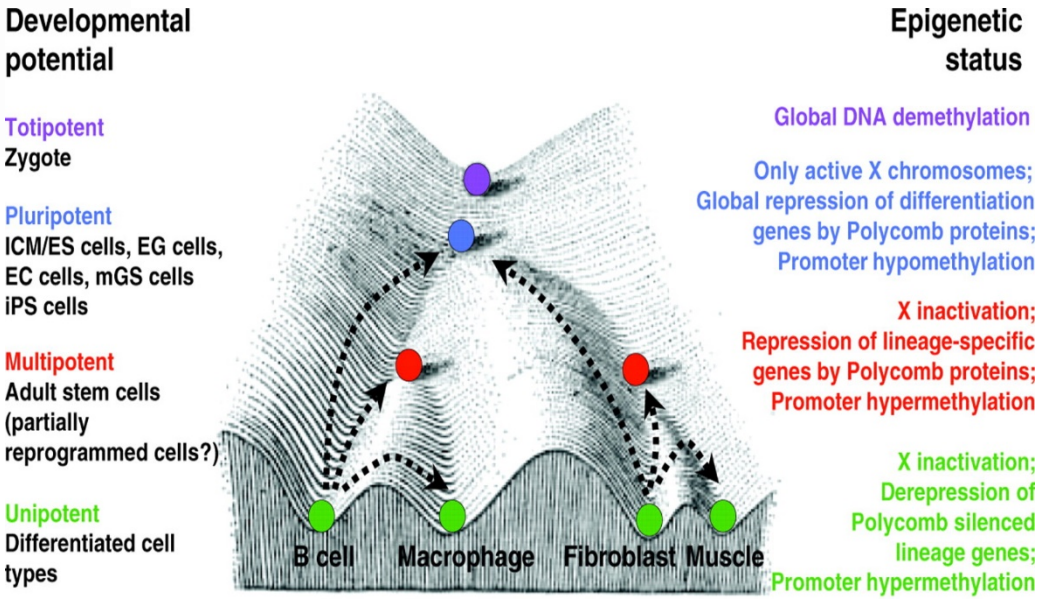
Stages of Cell development



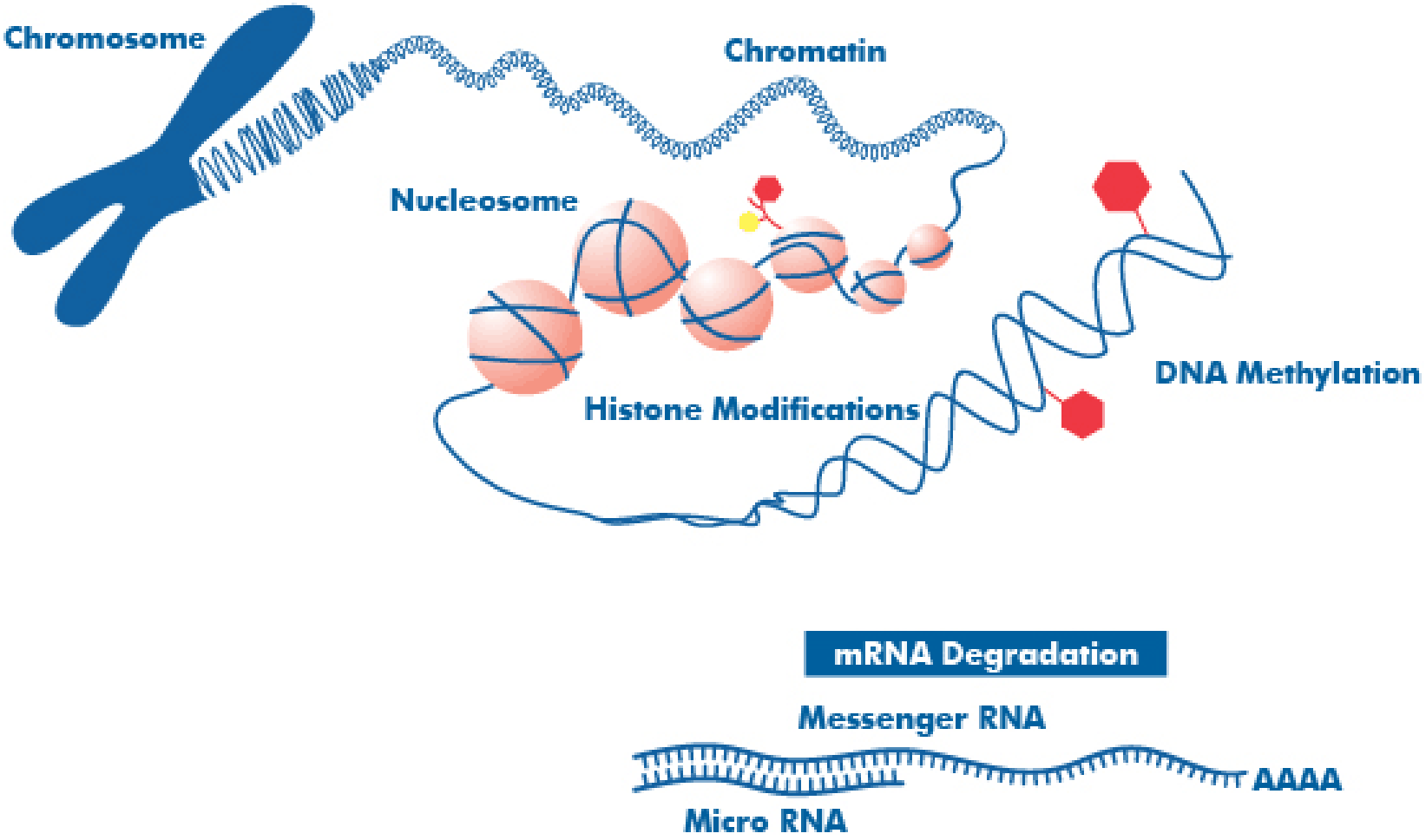


Conrad H. Waddington 1956

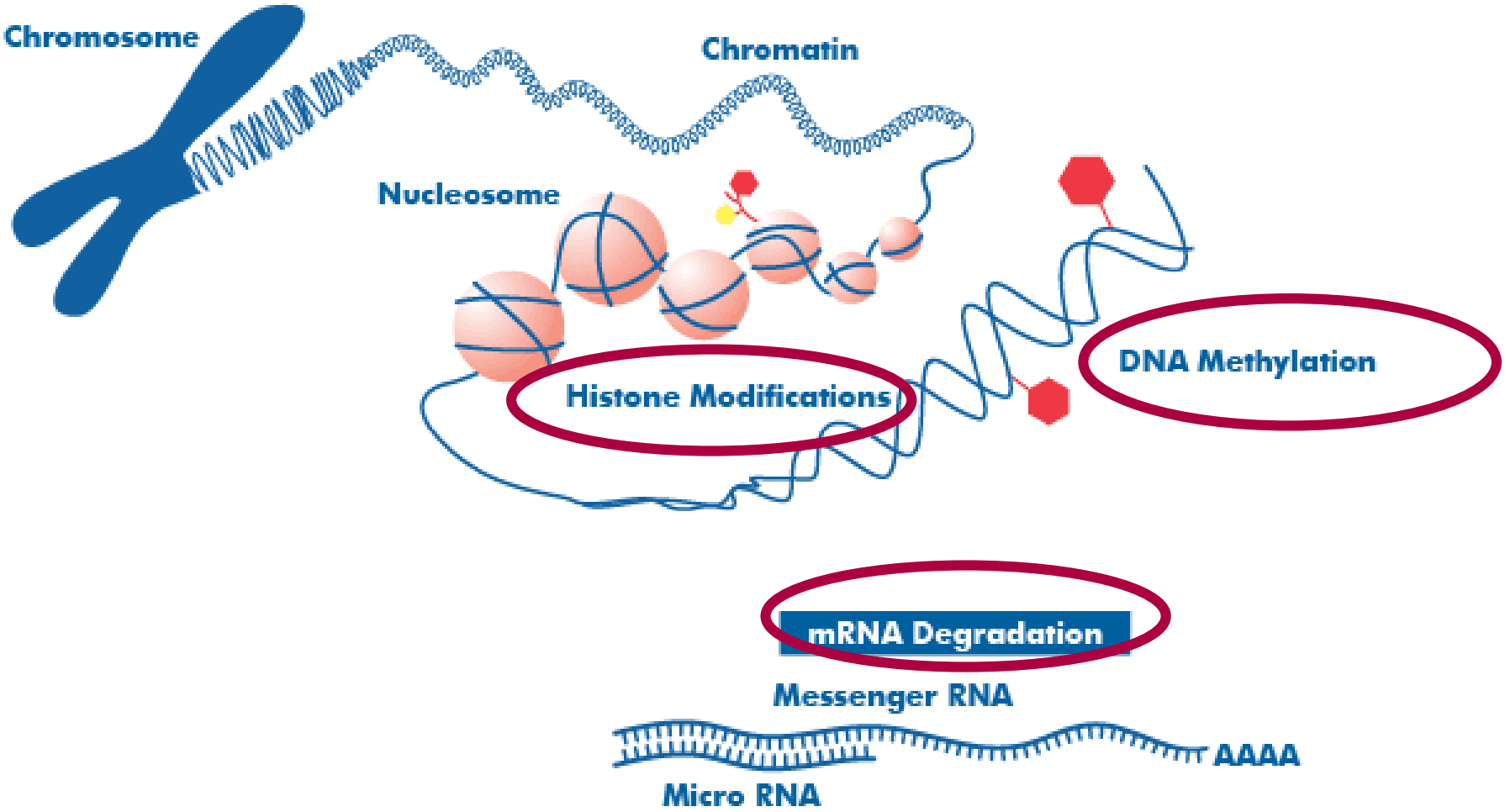
Konrad Hochedlinger and K. Plath,
Development 136, 509-523 (2009)



Epigenetics changes



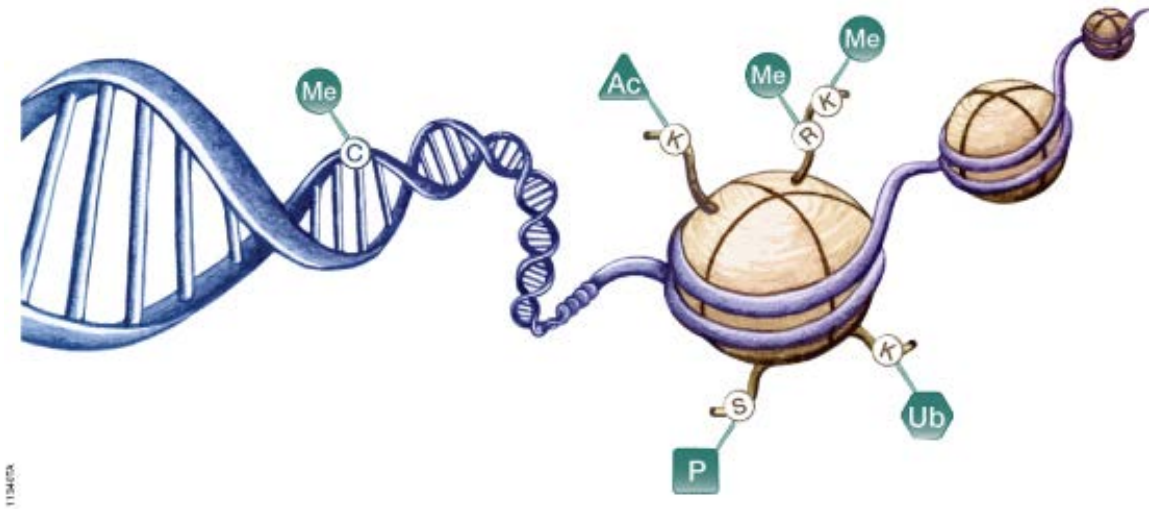
Epigenetics changes



Histone modifications

- Histones are chief protein components of chromatin that order the DNA into structural units called nucleosomes.

- Histone modifications control the gene expression of cells via the following modifications:
  - methylation
  - acetylation
  - ubiquitylation
  - phosphorylation
  - sumoylation



https://worldwide.promega.com/resources/product-guides-and-selectors/protocols-and-applications-guide/epigenetics/

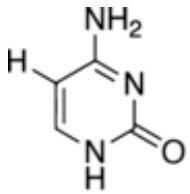
DNA Methylation

- It is the addition of methyl groups to the DNA
- happens mostly at CpG sites, such that cytosine is converted to 5- methylcytosine.

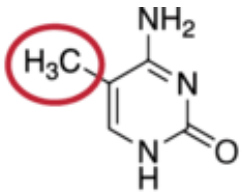
- Highly methylated areas in a genome are frequently less transcriped or silenced (switched off).

  High Methylation: Hyper methylation

  Low Methylation: Hypo methylation



Cytosine



methylated Cytosine

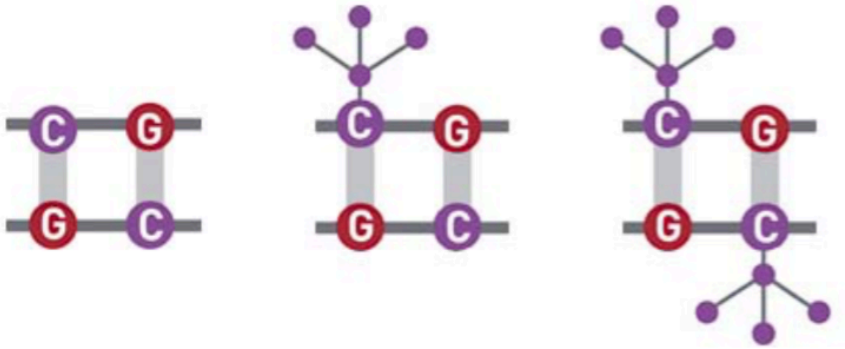
Methylation of cytosines can also persist from the germ line of one of the parents into the zygote, leading to **genetic imprinting.**

CpG islands

- Generally, CpG (5'->Cytosine -> Phosphate -> Guanine -> 3')
- CpG islands are defined as regions
  - >500 bp,
  - >55% GC content
- Length: 0.5-2KB
- 50% of gene promoters contain islands.

CpG methylation

- Methylation is mainly targeted at CpG dinucleotides
- CpGs are either unmethylated or methylated on both strands
- Hemi-methylated CpGs are rare



- DNA methyltransferases (DNMTs) bind hemi-methylated sites and modify the remaining position
- Thus the epigenetic information is inherited to daughter cells

miRNAs and CpG methylation

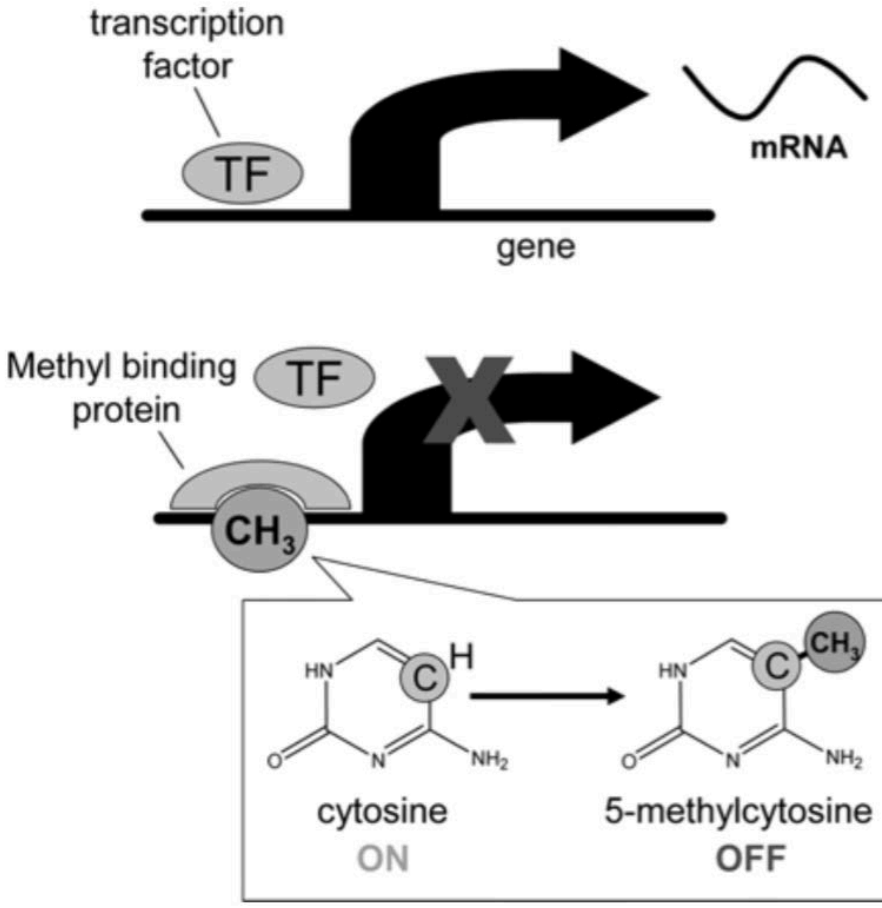
- About 50% of miRNAs are associated with CpG islands, that may be repressed by epigenetic methylation.
- Transcription from methylated CpG islands is strongly repressed.
- Other miRNAs are epigenetically regulated by either histone modifications or by combined DNA methylation and histone modification.
- **Interesting research topic for PhD or MSc : EpiMiR????**

Building a database for the association between epigenetic modifications and miRNA regualtions in normal and tumor tissues

Promotor methylation and expression regulation

50-60% of promotor regions are CpG islands /regions

Methylation level anti-correlated to expression in enriched CPG promoter regions.



http://www.39kf.com/uploadfiles/image/15902/TXT-20081228163836878.gif

DNA methylation

- DNA methylation is involved in
  - Gene expression
  - **Imprinting**
  - X-chromosome inactivation
  - Chromosome stability
  - **Development**
  - **Changes in cancer**
  - Changes with age
  - ……..

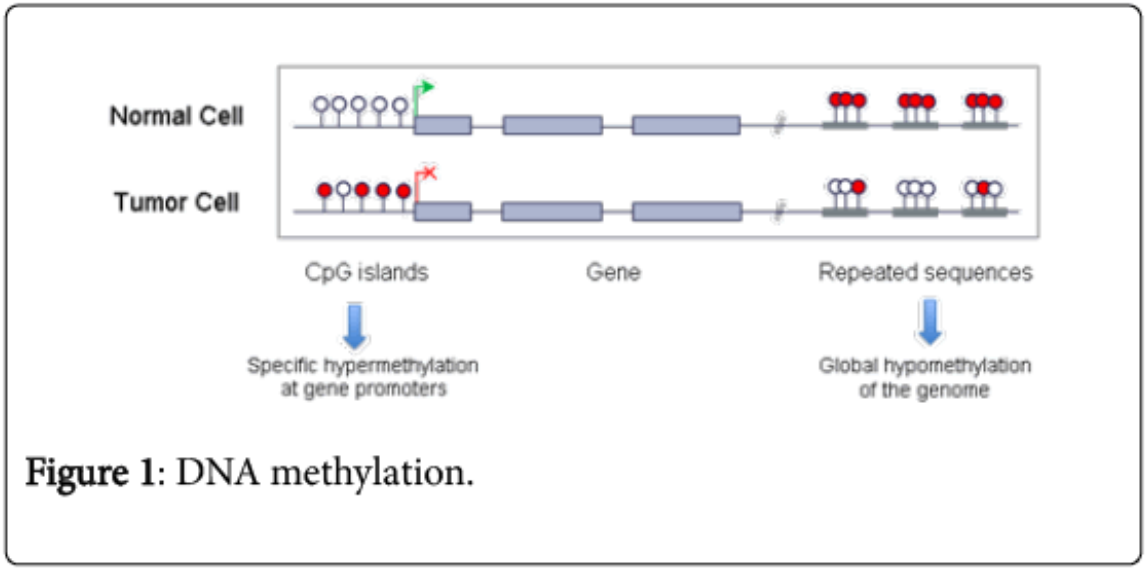
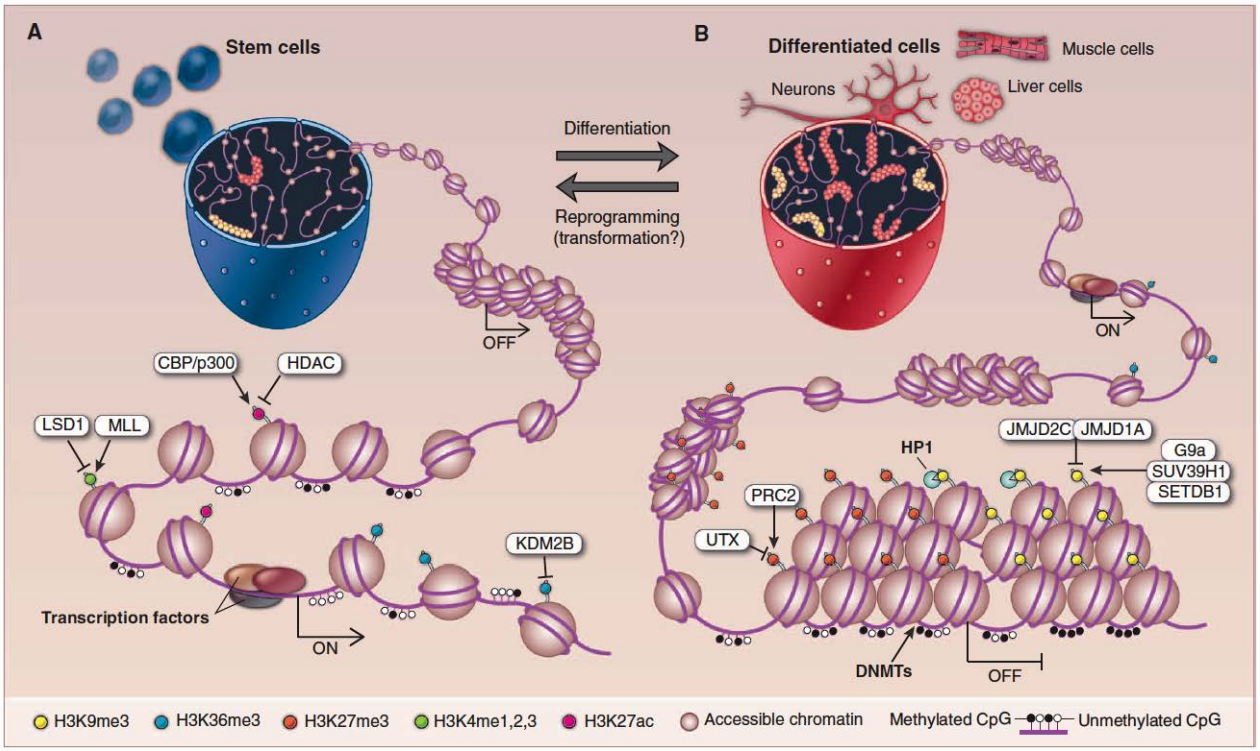


Figure 1: DNA methylation.

https://www.omicsonline.org/open-access/dna-methylation-in-cancer-tissues-2157-7013-1000272.php?aid=89525

Differentiation linked to alterations of chromatin structure



(A) In pluripotent cells, chromatin is hyperdynamic and globally accessible.

(B) Upon differentiation, inactive genomic regions may be sequestered by repressive chromatin enriched for characteristic histone modifications.

These global structures are regulated by DNA methylation, histone modifications, and numerous CRs whose expression levels are dynamically regulated through development.

ML Suva et al. Science 2013; 339:1567-1570

Genomic imprinting

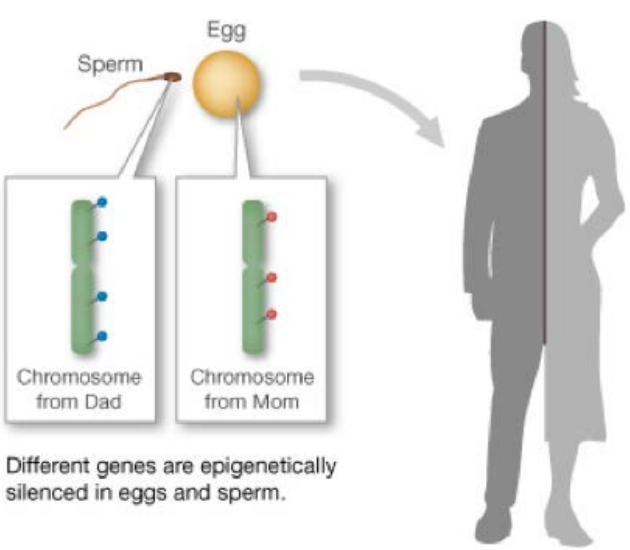
- Violate the usual rule of inheritance (only one allele is expressed)
- Bi-allelic Genes :

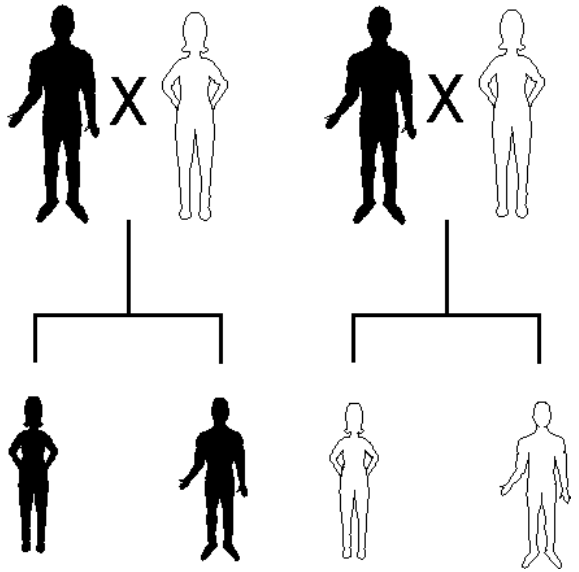
  gene encoding hemoglobin A from Dad

  gene encoding hemoglobin S from Mom

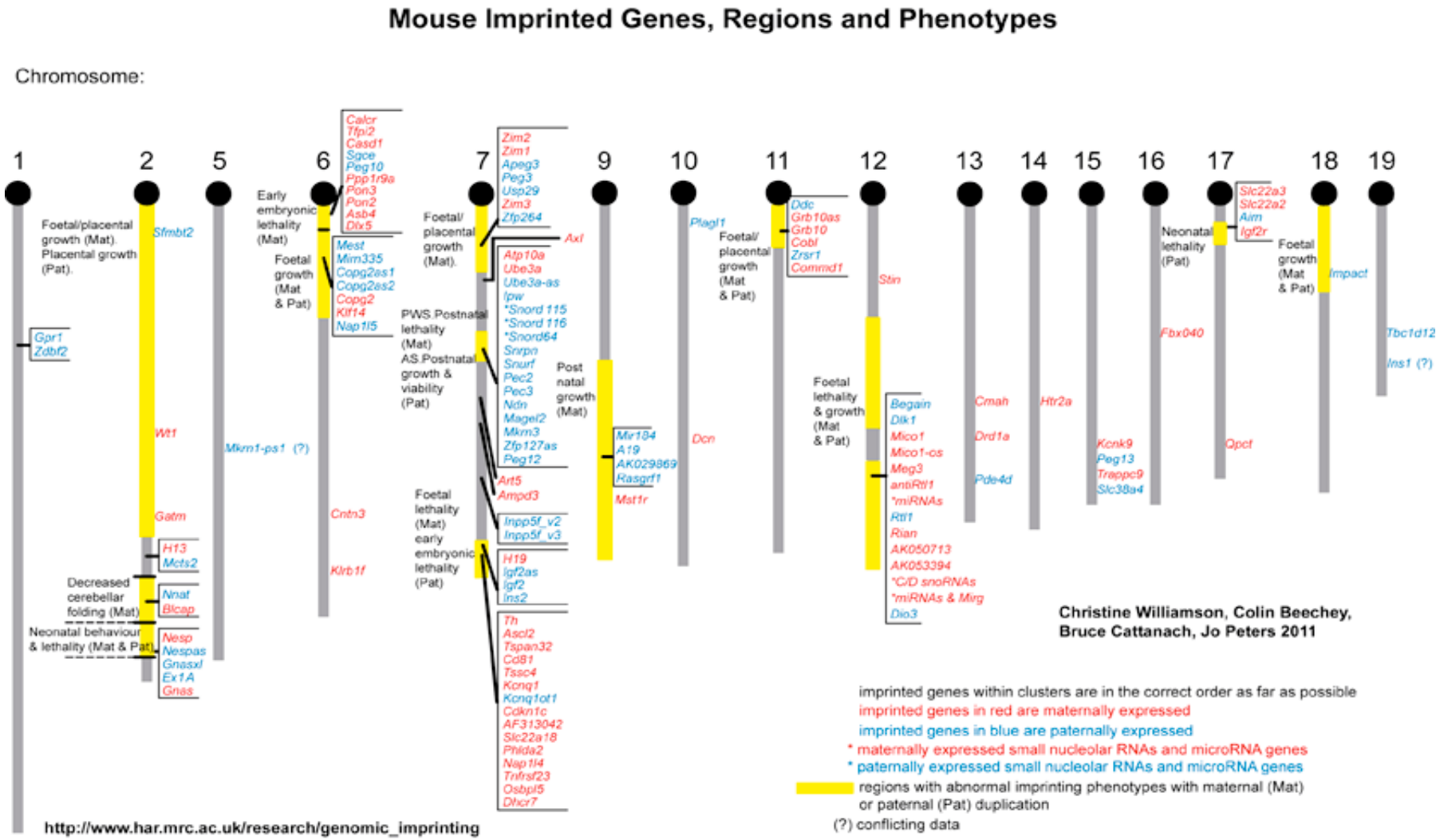
  Child: equal amount of the 2 types of hemoglobin

- Mono- allelic (Imprinted) Genes :



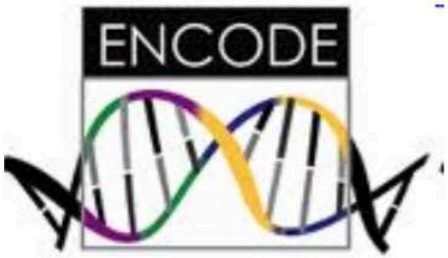


Imprinted genes distribution

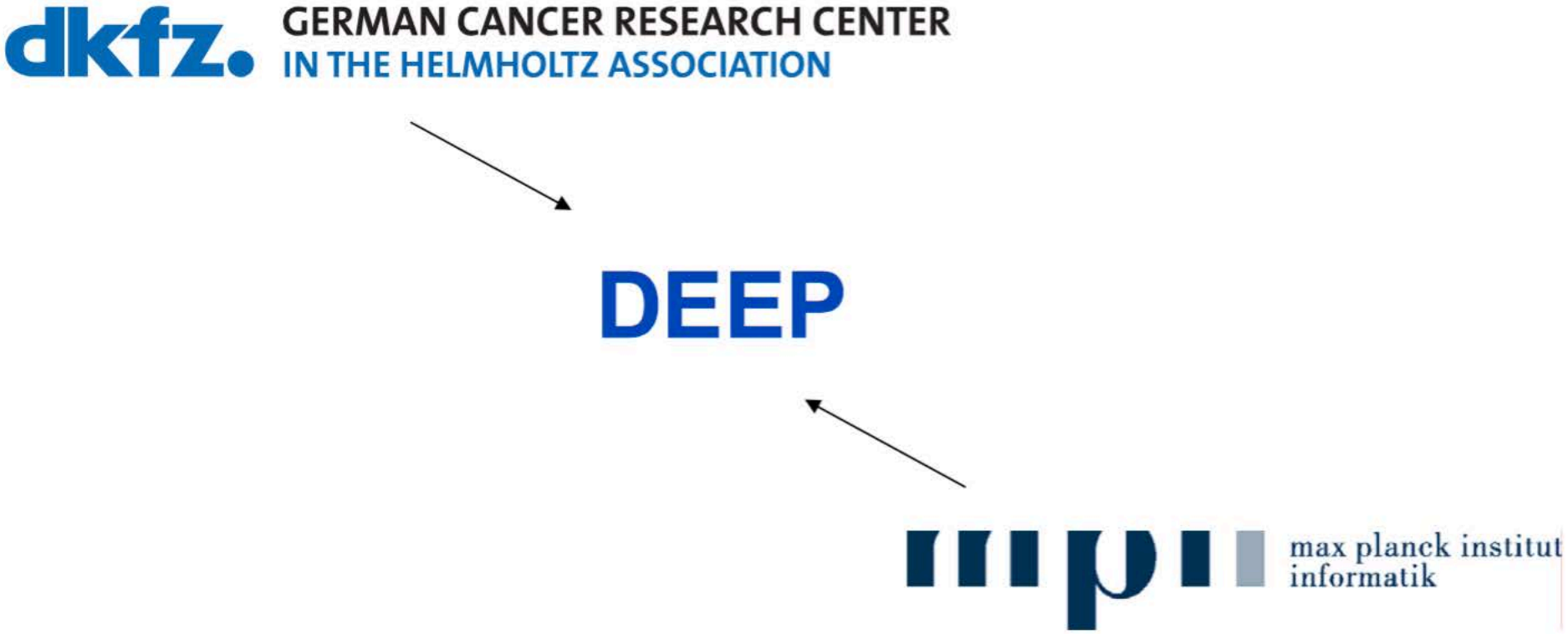


Epigenome Projects

- One Genome many Epigenomes
- Cataloguing epigenetic modifications in different tissues



German Epigenome project



Useful resources

- NIH
  - http://www.roadmapepigenomics.org/
- Genome Browsers
  - http://www.epigenomebrowser.org
  - http://genomebrowser.wustl.edu/
  - http://epigenomegateway.wustl.edu/
- Data Repositories
    - NCBI Epigenomics Gateway
    - Epigenome Atlas

Translation into NGS signal

NGS applications

DNA

Whole human genome sequencing
Exome sequencing
Gene panel sequencing
Pathogene sequencing

SNP calling
INDEL calling
Copy Number Variations
Genomic Rearrangements

RNA

Transcriptome sequencing
miRNOme sequencing

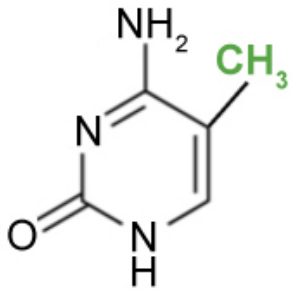
Expression level
Alternative splicing events
Fusion genes

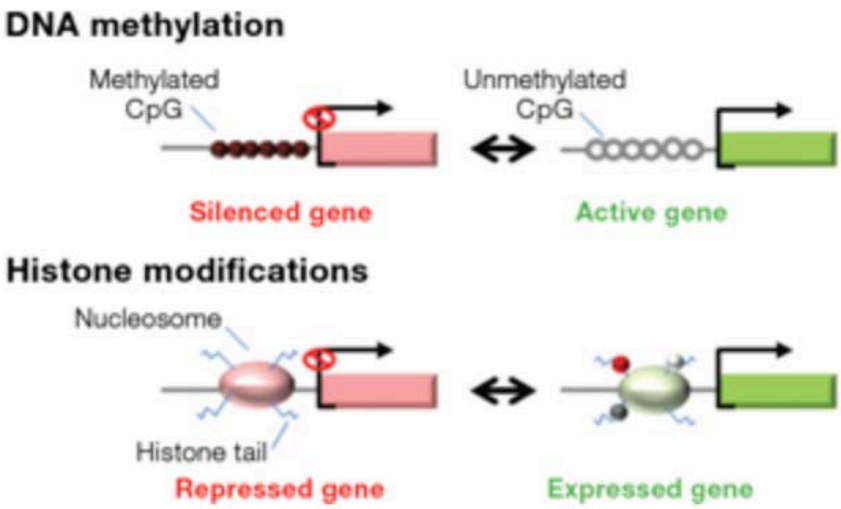
Others

interactions of proteins with DNA (ChIP seq)
interactions of proteins with RNA (PAR-CLIP)
Bisulfite sequencing (methylation)

NGS applications: Epigenomics

- Histone modifications
  - ChIP-Seq
- Methylome
  - 5-methylcytosine
  - whole genome or reduced representation bisulphite sequencing (WG-BS, RRBS)
  - meDIP-Seq
- Global methylation changes
- Differentially methylated regions
- Integration with gene expression





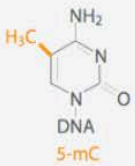
Experiments for quantifying DNA Methylation

- **Capture based sequencing**
  - Chromatin Immunoprecipitation (ChIP)
  - Metylated DNA Immunoprecipitation(MeDIP)
  - MBD chromatography

- **Conversion based sequencing**
  - Bisulfitesequencing(methyl-seq)
  - Reduced representation bisulfite sequencing (RRBS)
  - Ultra-deepamplicon sequencing

Experimental approaches for measuring DNA Methylation

Method	DNA Preparation	Considerations
WGBS-Seq Whole-Genome Bisulfite Sequencing	Convert non-methylated Cs to U with sodium bisulfite, which are read as T. 5-mC and 5-hmC are protected from conversion and read as C. Sequence with random-primer extension, 2 × 75–100 bp reads.	Queries >38 million CpGs for humans Requires > 90 Gb/sample Cost: US $5,000–6,000/ sample*
RRBS-Seq Reduced Representation Bisulfite Sequencing	Digest DNA with *MspI* and isolate 100–150 bp fragments, representing >85% CpG islands. Bisulfite treat fragments and sequence using 1 × 75 bp reads.	Queries ~85% CpG islands (~2 M CpGs) and 60% of RefSeq promoters Requires ~40–50 M reads, 3-5 Gb/sample Cost: US $300–360/sample*
MeDIP Methylated DNA Immunoprecipitation	Sonicate DNA to 100–300 bp, end-repair, ligawte to adapters, and denature. Capture 5-mC containing fragments with anti-5-mC magnetic beads, and sequence using 1 × 75–100 bp reads. May be combined with bisulfite conversion.	5-mC detected at ~150 bp resolution. Bias towards hypermethylated regions. Requires ~60 M reads, ~5 Gb/sample Cost: US $300–360/sample*
MIRA Methylated CpG Island Recovery Assay	Sonicate DNA to 100–300 bp, end-repair, ligate to adapters, and capture methylated dsDNA with GST-labeled MBD2b and His-tagged MBD3L1 proteins. Purify with glutathione beads and sequence with 1 × 75 bp reads.	Detects mCpGs at ~150 bp resolution. Bias towards hypermethylated CpGs; non-CpG methylation not detected. Requires ~60M reads, ~5 Gb/sample Cost: US $300–360/sample*
HM450 Array Human Methylation 450k	Convert DNA with sodium bisulfite to find C-to-T changes at defined genomic positions detected by the Infinium® assay.	Samples 96% CpG islands, 485,000 CpGs, 99% RefSeq promoters, 3' & 5' UTR, 1st exon, gene body, 3' UTR, shores and shelves Cost: US $300–360/sample*



http://res.illumina.com/documents/products/other/field_guide_methylation.pdf

1-Bisulfite Conversion

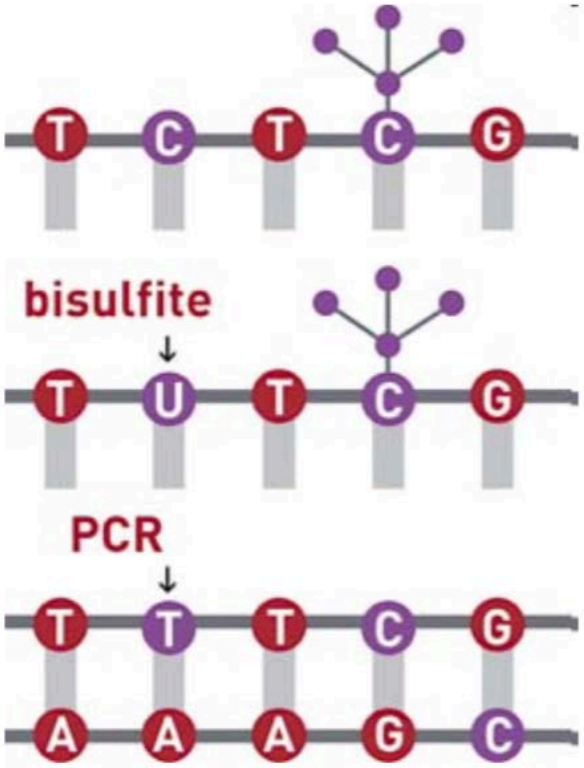
- Originally introduced by Frommer in 1992
- (http://www.ncbi.nlm.nih.gov/pmc/articles/PMC48546/)

Idea

bisulfite treatment specifically changes the DNA sequence dependent on the methylation status

- Treatment of DNA with bisulfite converts :
  - non-methylated cytosine to uracil
  - And keeps 5-methylcytosine unaffected.

On a single nucleotide resolution, bisulfite sequencing thus reveals information about the methylation status of a segment of DNA

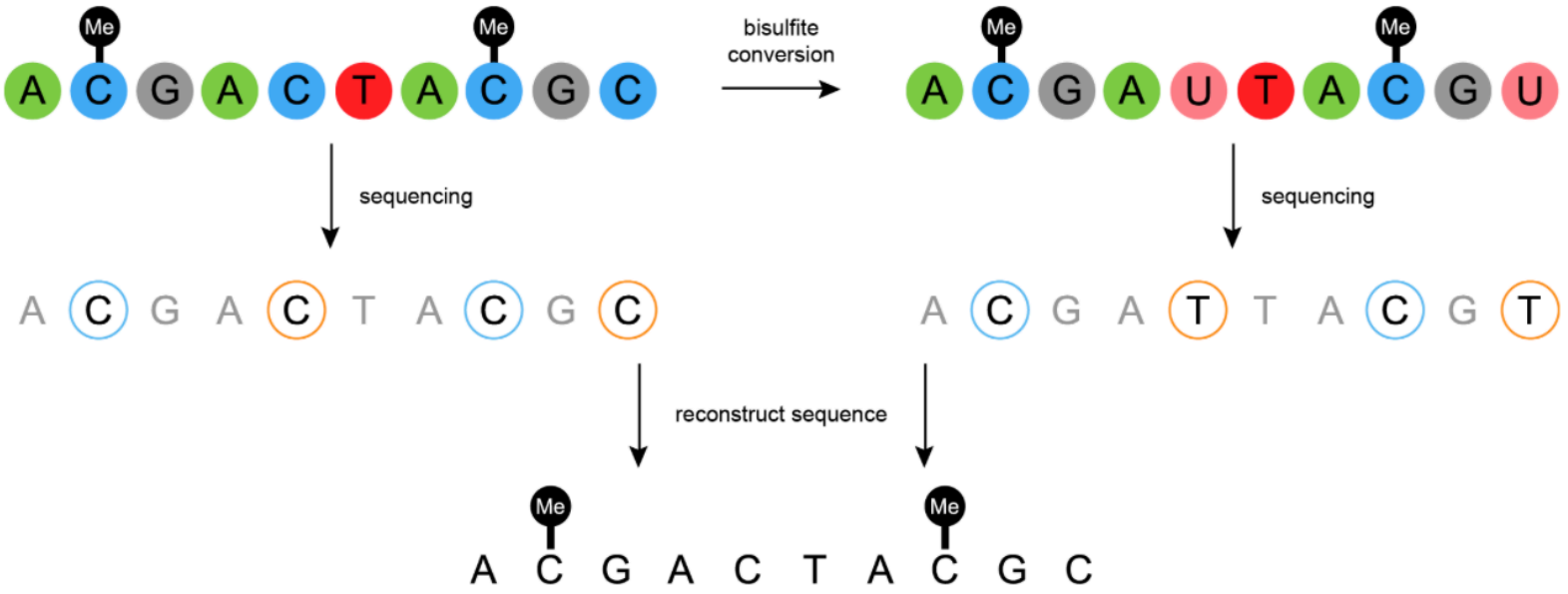


1-Bisulfite Conversion

Idea

bisulfite treatment specifically changes the DNA sequence dependent on the methylation status: non-methylated cytosine to uracil and

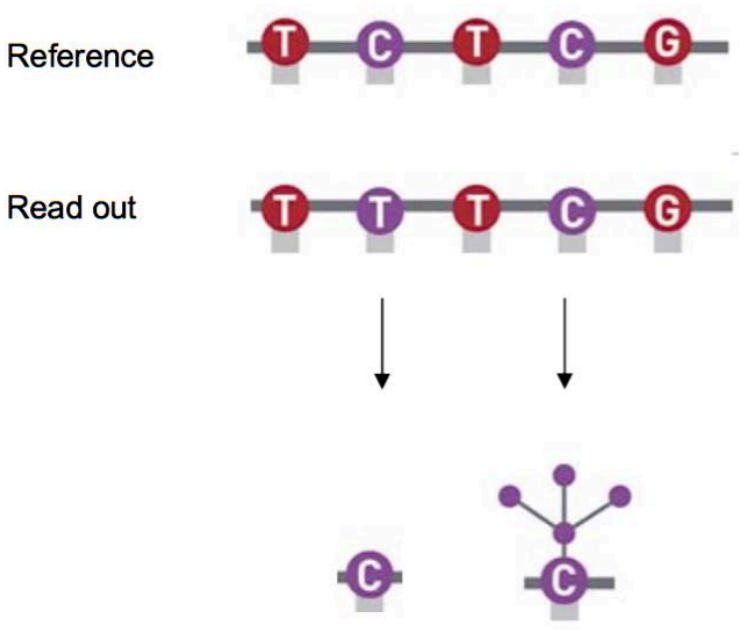
5-methylcytosine kept unaffected.



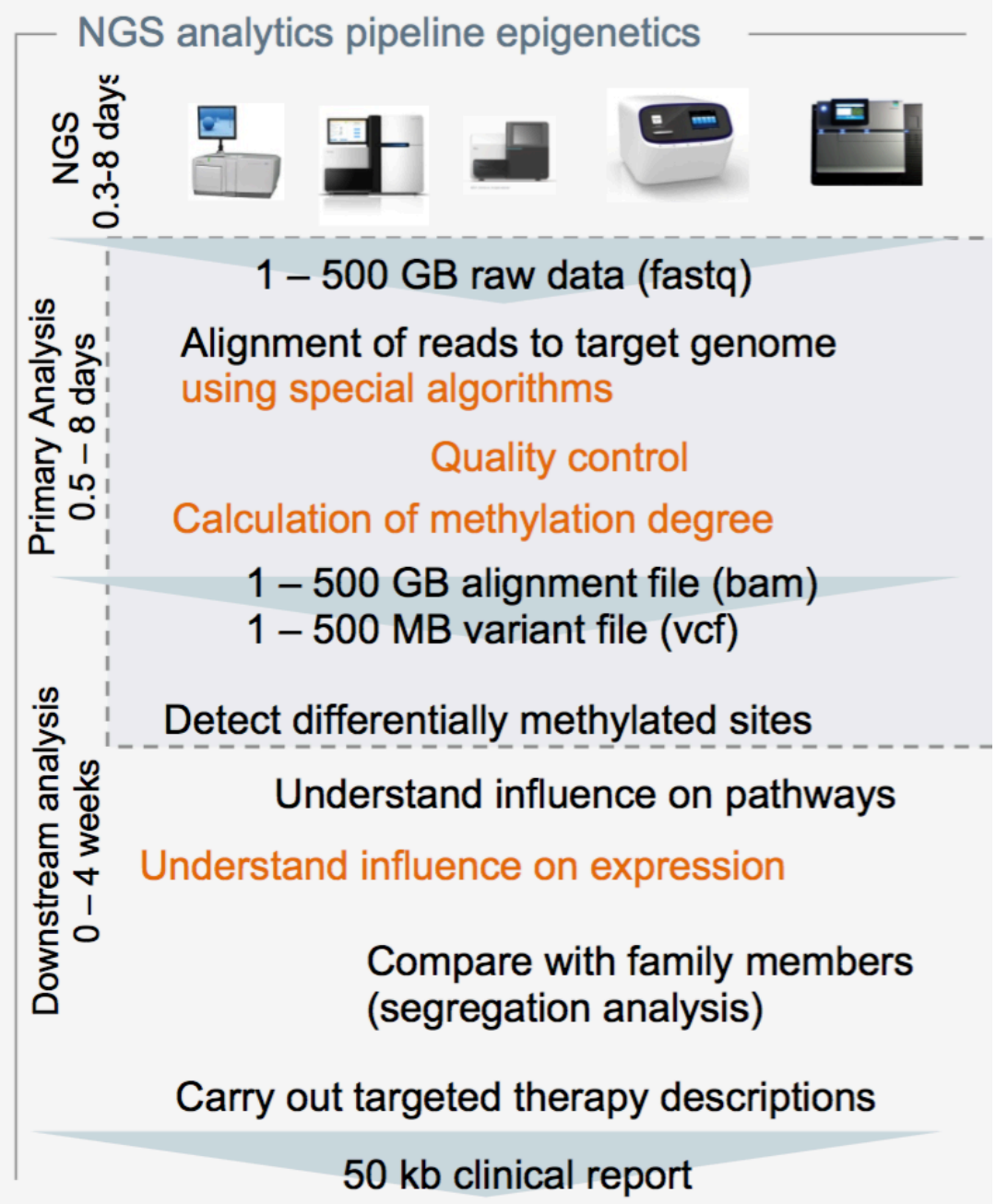
https://galaxyproject.github.io/training-material/topics/epigenetics/slides/introduction.html#11

Conversion-seq – Alignment

- Bisulfite treated DNA is mainly characterized by A, G and T nucleotides.
- Normal alignment to the reference is not possible
- Specialized alignment procedures have lower accuracy



NGS pipeline for Epigenetic Data

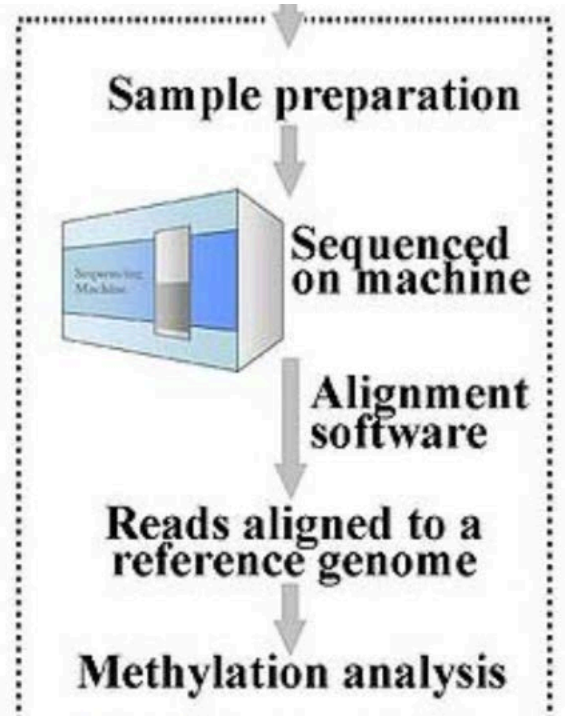


Credit: Dr. Keller, CBI, Saarland

Alignment tools for bisulfite sequences

A number of tools support the alignment of bisulfite treated DNA

- Bismark
- BSMAP
- BS Seeker
- Cokus
- Merman (also Color Space)
- BSmooth



Estimating of Methylation

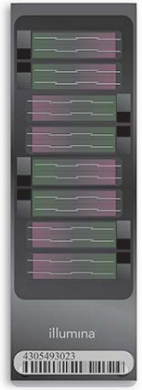
- Measure unmethylated Cs, denoted as #C
- Measure methylated Cs, denoted as #5mC
- Compute fraction of methylated Cs on all Cs.

Methylation ration = (Beta value)

$$\frac{\#5mC}{\#5mC+\#C}$$

2- Methylation Arrays

- Infinium HD Methylation (Illumina, Affymatrix,..)
- Robust methylation profiling microarray with extensive coverage of CpG islands, genes, and enhancers. Use for epigenome-wide association studies (EWAS).
- Different sizes (27k, 45k,850k)
- The HD EPIC: Targets 850,000 CpGs. Quantitatively interrogates selected methylation sites across the genome.

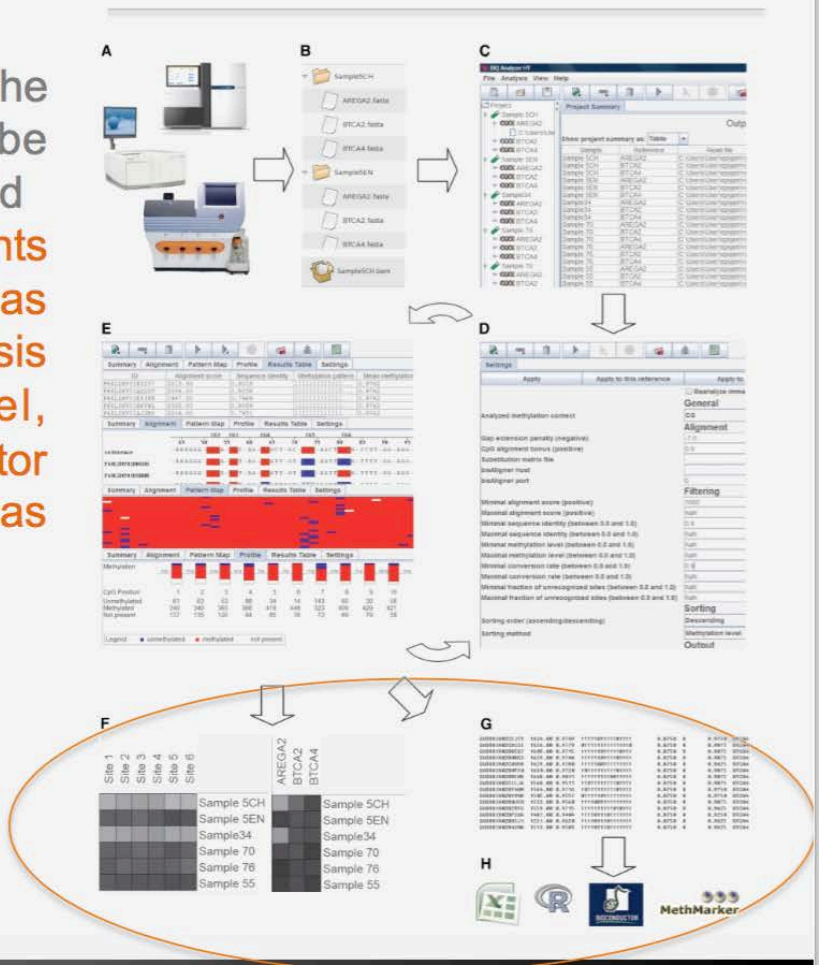


- Over 850,000 methylation sites per sample at single-nucleotide resolution

Integrative analysis developed at MPI Germany

Integrative pipeline developed at MPII

- The the analysis is carried out and the inferred DNA methylation data can be inspected and parameters can be adjusted
- Finally, the DNA methylation measurements can be visualized graphically, exported as tab-separated tables for in-depth analysis using spreadsheets such as Excel, statistical software such as R/Bioconductor and biomarker development tools such as MethMarker.



Relevant Bioconductor packages

- **methylKit** : DNA methylation analysis from high-throughput bisulfite sequencing results.
- **methylationArrayAnalysis**: A cross-package Bioconductor workflow for analysing methylation array data.
- **IMA**: An R package for high-throughput analysis of Illumina's 450K Infinium methylation data.
- **RnBeads**: Comprehensive analysis of DNA methylation data

RnBeads

MENU nature methods

Brief Communication | Published: 28 September 2014

Comprehensive analysis of DNA methylation data with RnBeads

Yassen Assenov, Fabian Müller, Pavlo Lutsik, Jörn Walter, Thomas Lengauer & Christoph Bock

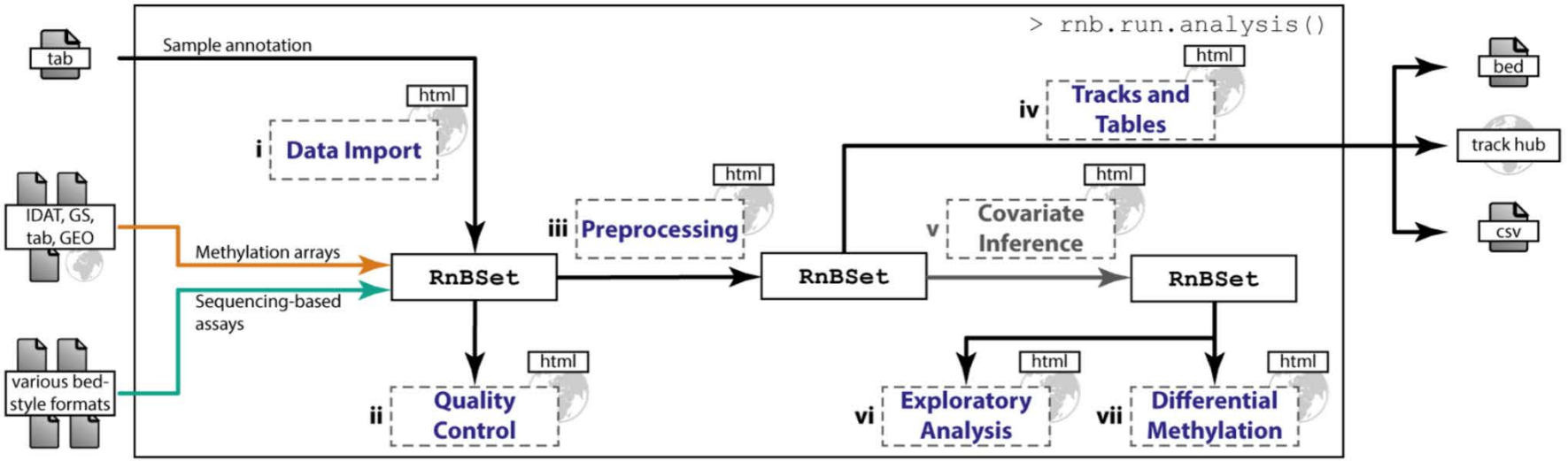
Nature Methods **11**, 1138–1140 (2014) | Download Citation

Abstract

RnBeads is a software tool for large-scale analysis and interpretation of DNA methylation data, providing a user-friendly analysis workflow that yields detailed hypertext reports (http://rnbeads.mpi-inf.mpg.de/). Supported assays include whole-genome bisulfite sequencing, reduced representation bisulfite sequencing, Infinium microarrays and any other protocol that produces high-resolution DNA methylation data. Notable

RnBeads

RnBeads About Installation Tutorials Examples Resources References FAQ Contact



Comprehensive Analysis of DNA Methylation Data with RnBeads

To Do

- 1- Download the epigenetics data of prostate cancer (only few samples)
- Load them into Rstudio
- Process them and perform differential methylation analysis
- Identify the DMR’s
- Identify their Genes
- Perform Gene Enrichment analysis

YOUR TURN
START WITH THE TUTORIALS