Design and calibration of stochastic models for DNA methylation patterns

Master seminar by Andrea Kupitz

Supervisors: Prof. Dr. Verena Wolf, Alexander Lück

Overview

- Problem:
  - Method of operation of DNMTs unclear
  - Measure distances between methylation pattern distributions
- Approach:
  - MCMC algorithm to simulate cell cycle
  - Pairwise distances for ABC
- Results:
  - Parameters of MCMC difficult to identify
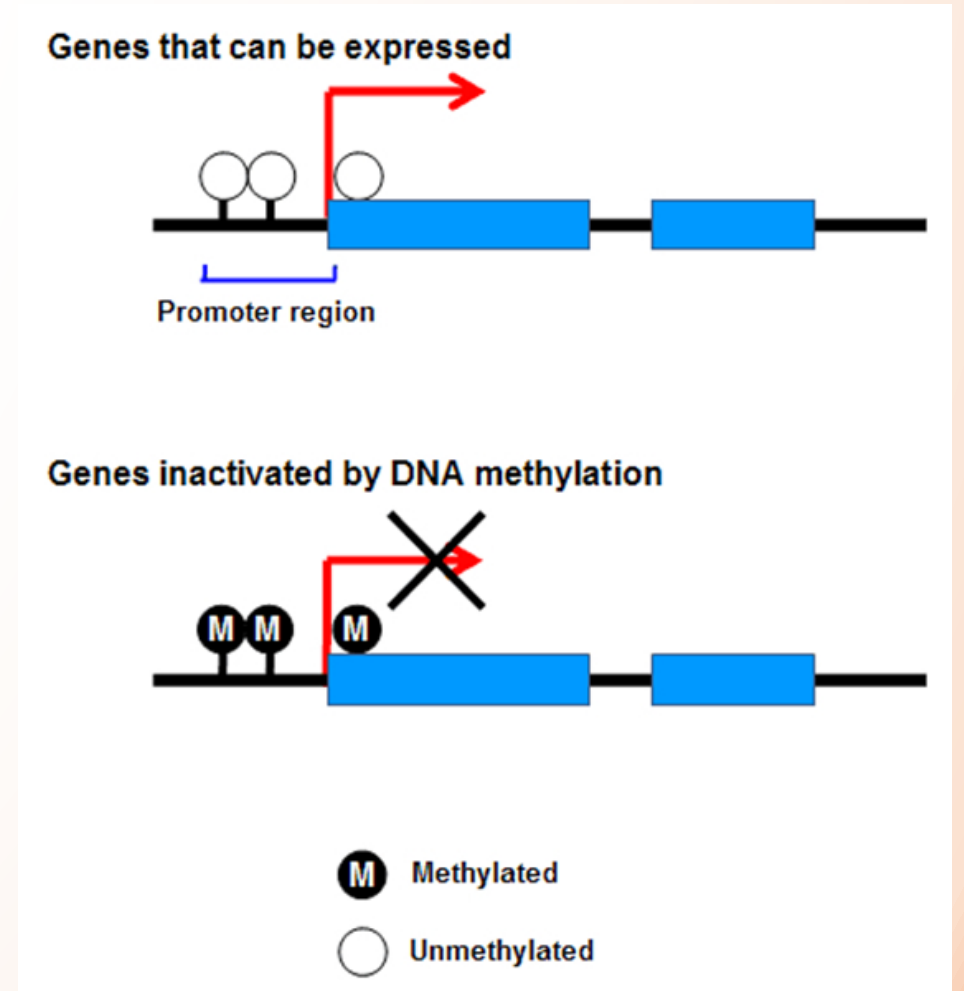- Outlook:
  - Validation of ABC method

Introduction – Epigenetics

- Histon modifications
  - Histon methylation
  - Histon acetylation
  - ...

- DNA modifications
  - **DNA methylation**

Parents

Environment

Epigenetik modifications

DNA

Cell differentiation differentiation
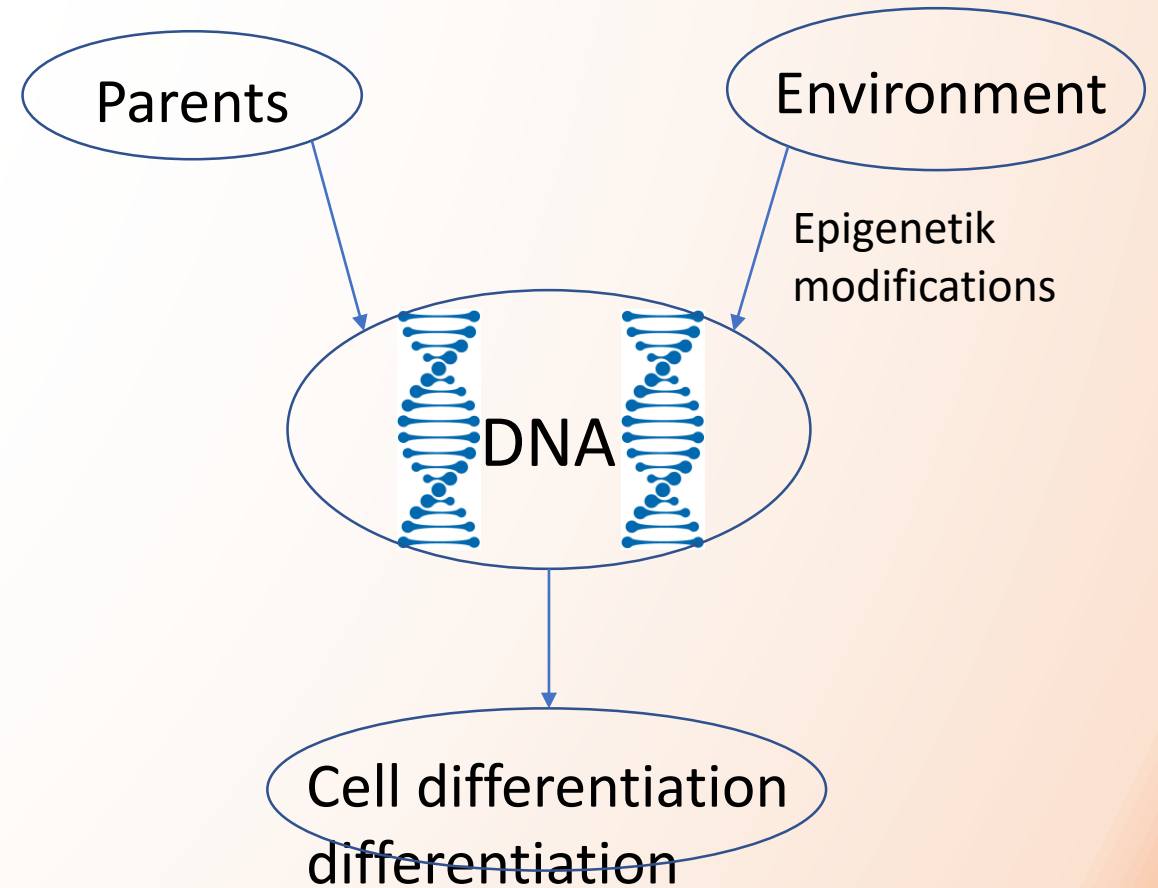
Introduction- Methylation

- Influences gene expression
- Indicator for diseases
- Heritable

Genes that can be expressed

Promoter region

Genes inactivated by DNA methylation

M Methylated

◯ Unmethylated

Introduction - Methylation

```
    M       M
    |       |
C-G-A-A-C-G-T-A
G-C-T-T-G-C-A-T
    |
    M
```
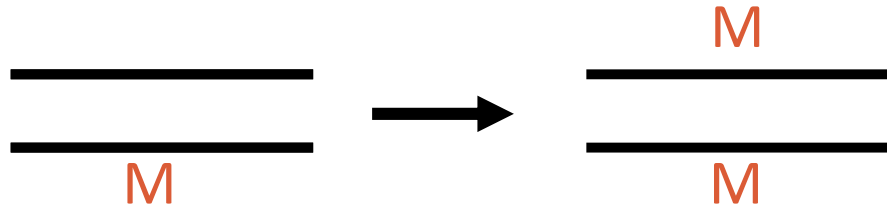
- DNA methylation at CpG
- Methylation transmitted by DNMTs
- Method of operation of DNMTs unclear
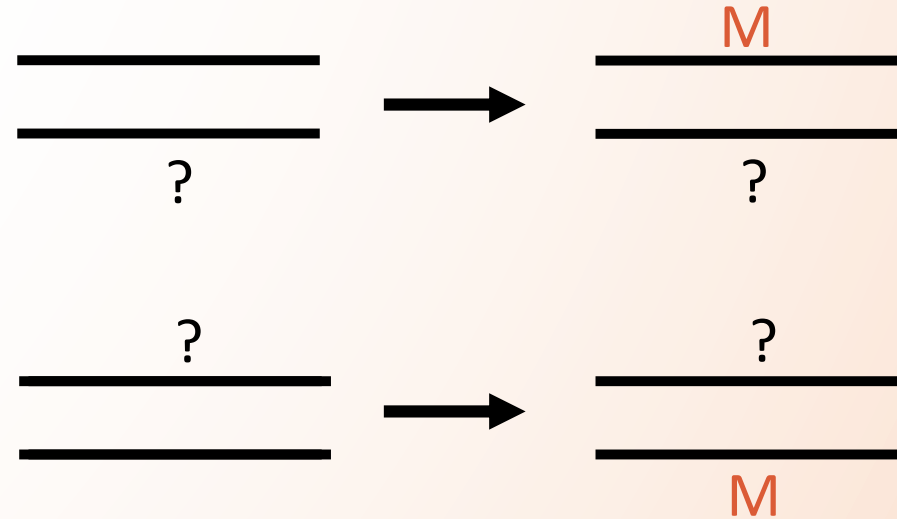
Introduction- DNMTs
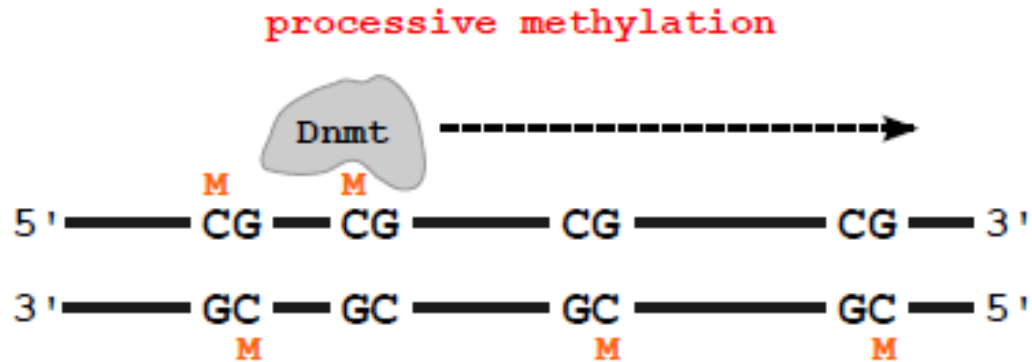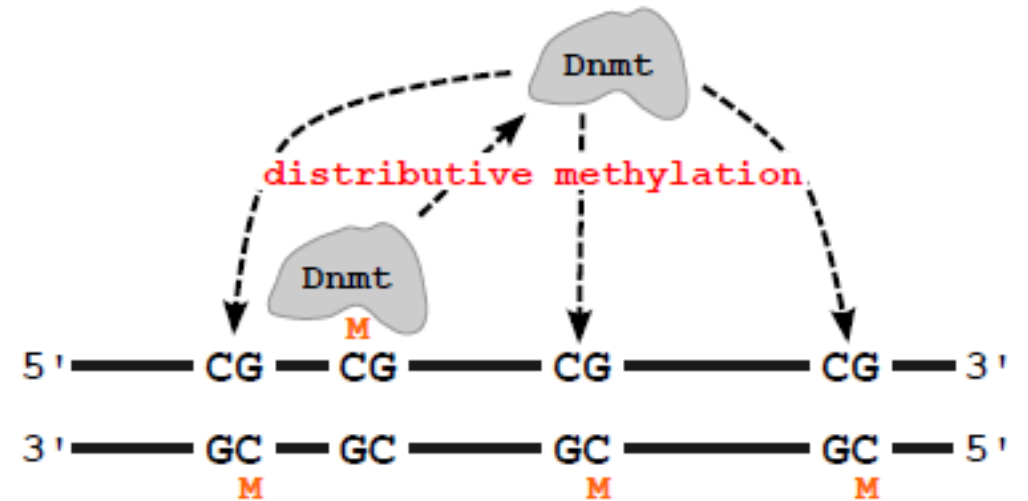
Methylation

- Maintainance:

- De novo:

Introduction- DNMTs

Processivity

- High:

- Low:

Methods – Markov Chain
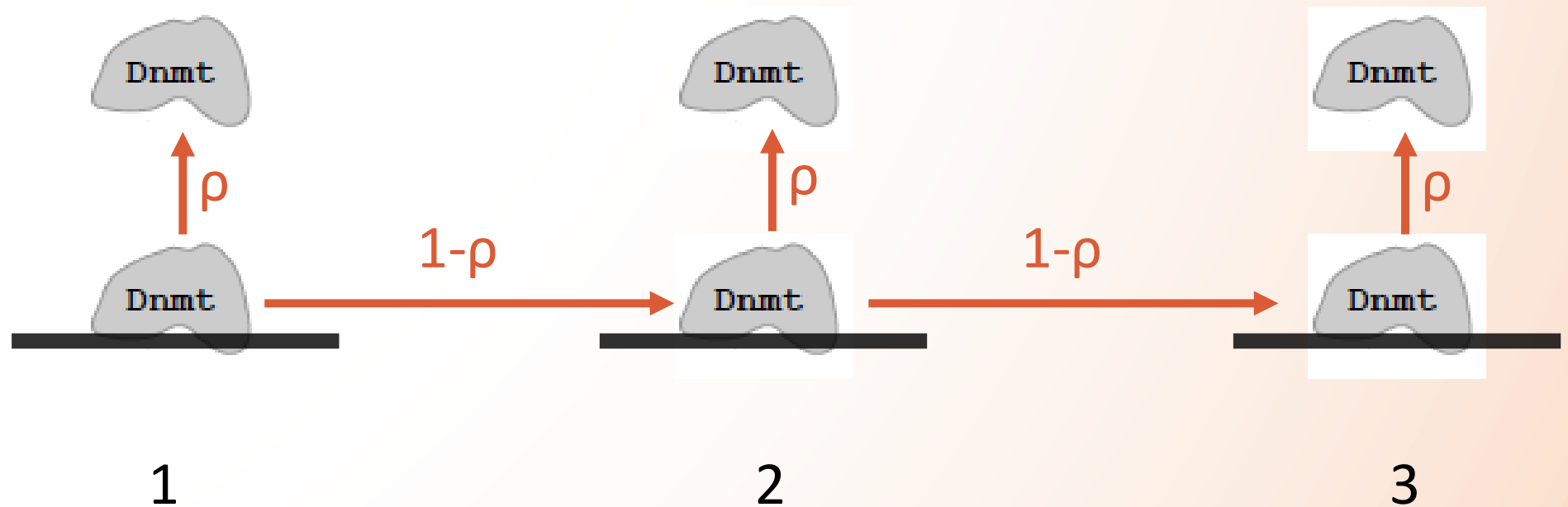
ρ – Dissociation probability

DNMT
Not bound

DNMT
Bound

Position   1   2   3

Methods – Markov Chain
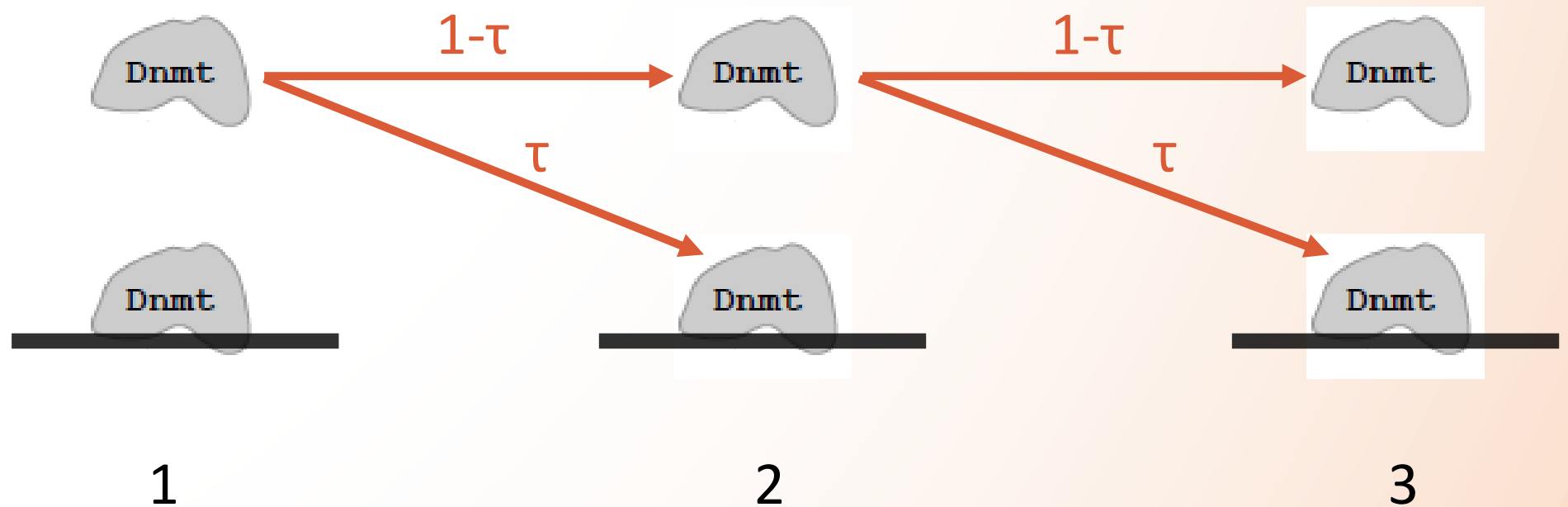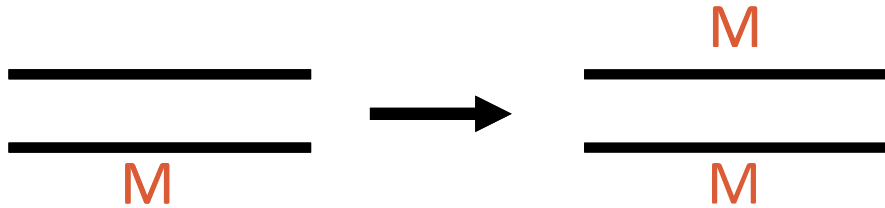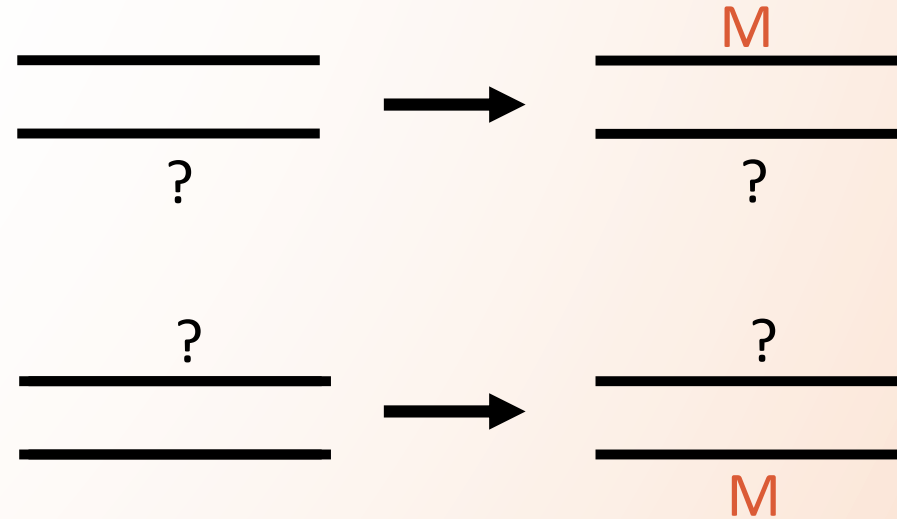
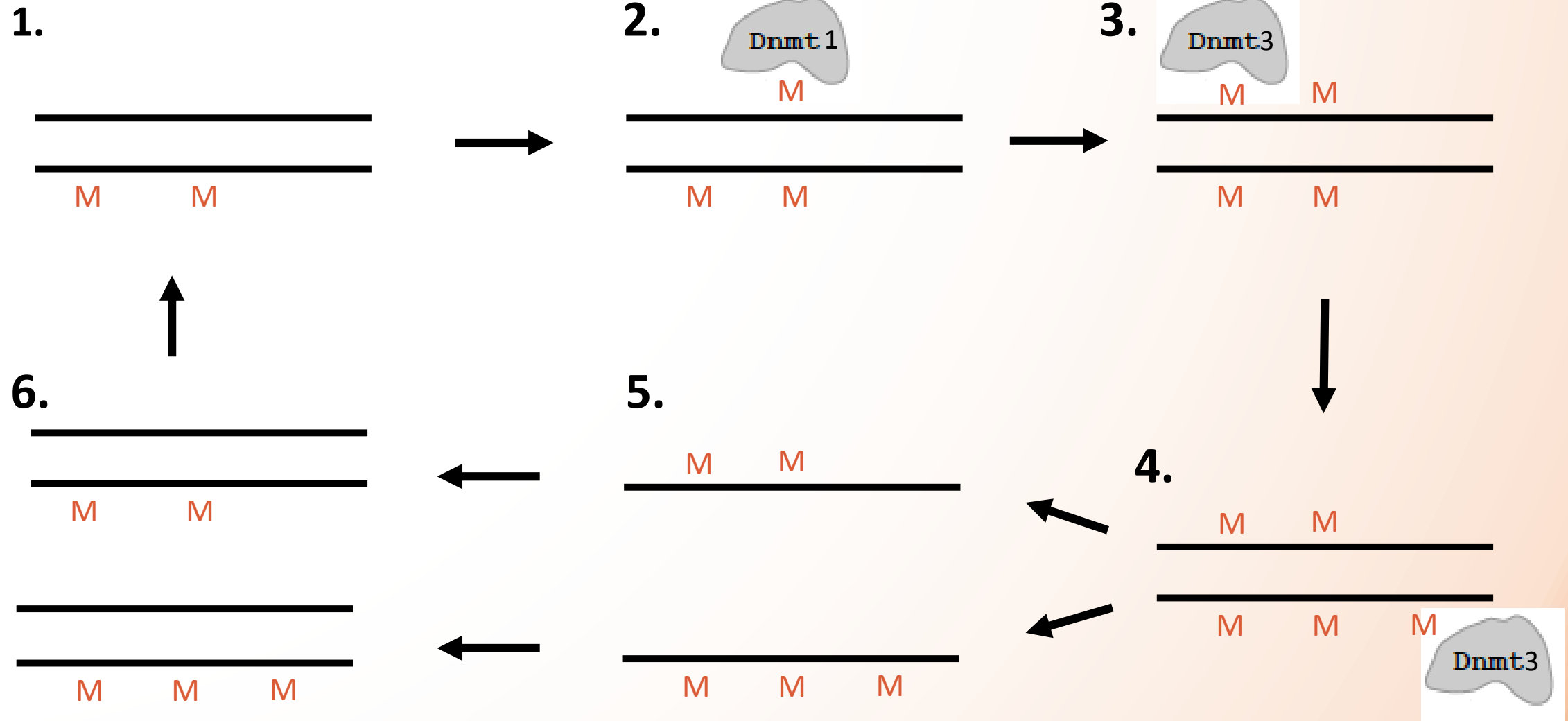**τ – Association probability**

Methods – Markov Chain

δ – De novo methylation probability

μ - Maintenance methylation probability

Methods – Cell Division

Methods - MCMC

Markov chain Monte Carlo

- Given:
  - Measured methylation pattern distribution before cell divisions
  - Simulation parameters ρ, τ, δ and μ
- Aim:
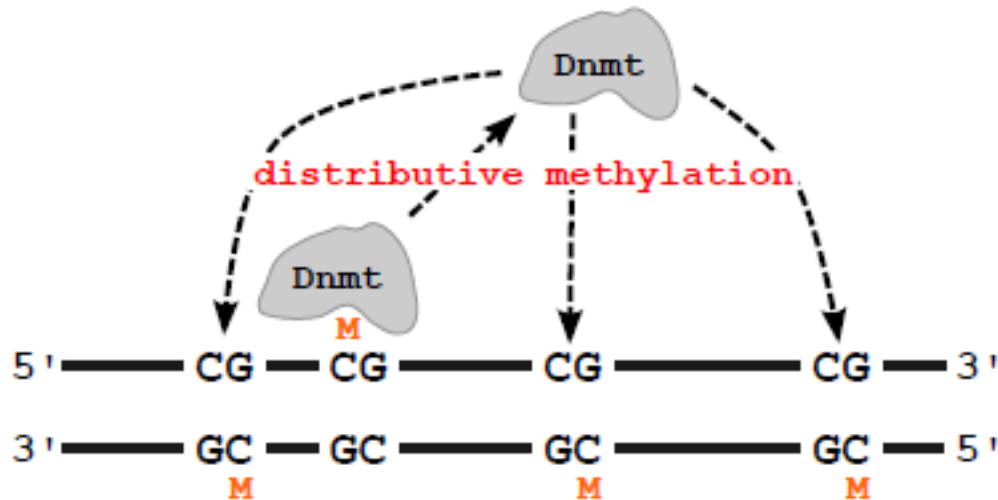 - Simulate a pattern distribution after cell division

Methods- MCMC

Workflow:

- 1. Sample from pattern distribution

- 2. Simulate t cell divisions
  - Draw RN and decide using ρ, τ, δ and μ if DNMT binds/methylates

- 3. Repeat 1 and 2 10000 times
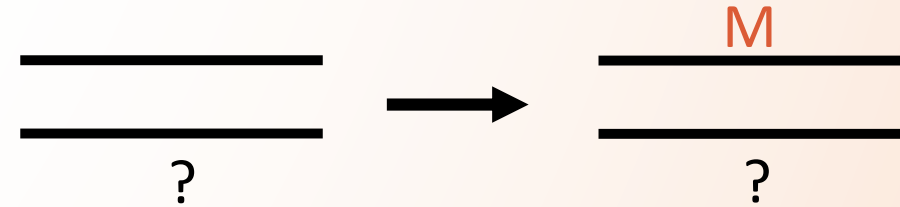
Results- MCMC

DMNT3 - Expectations

- Low processivity:

distributive methylation

- De novo methylation/few maintenance methylation activity:

Result:
- τ high (0.89)
- P low (0.28)

Result:
- μ high (0.78)
- δ high (1.00)

Results- MCMC

DMNT1 - Expectations

- High processivity:

processive methylation
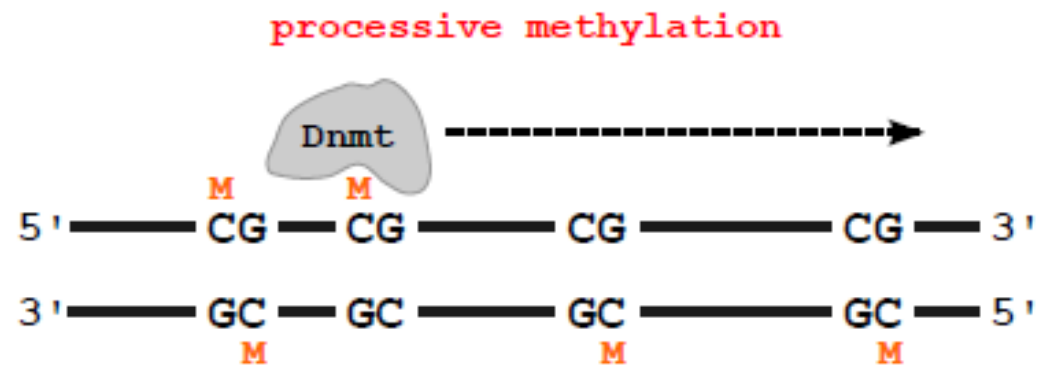
Result:
- τ low (0.23)
- P high (1.00)

- Maintenance methylation:
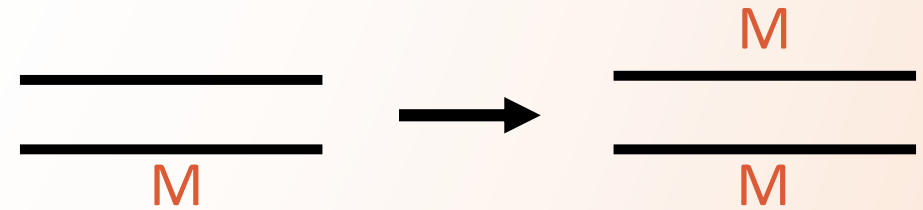
Result:
- μ high (0.75)
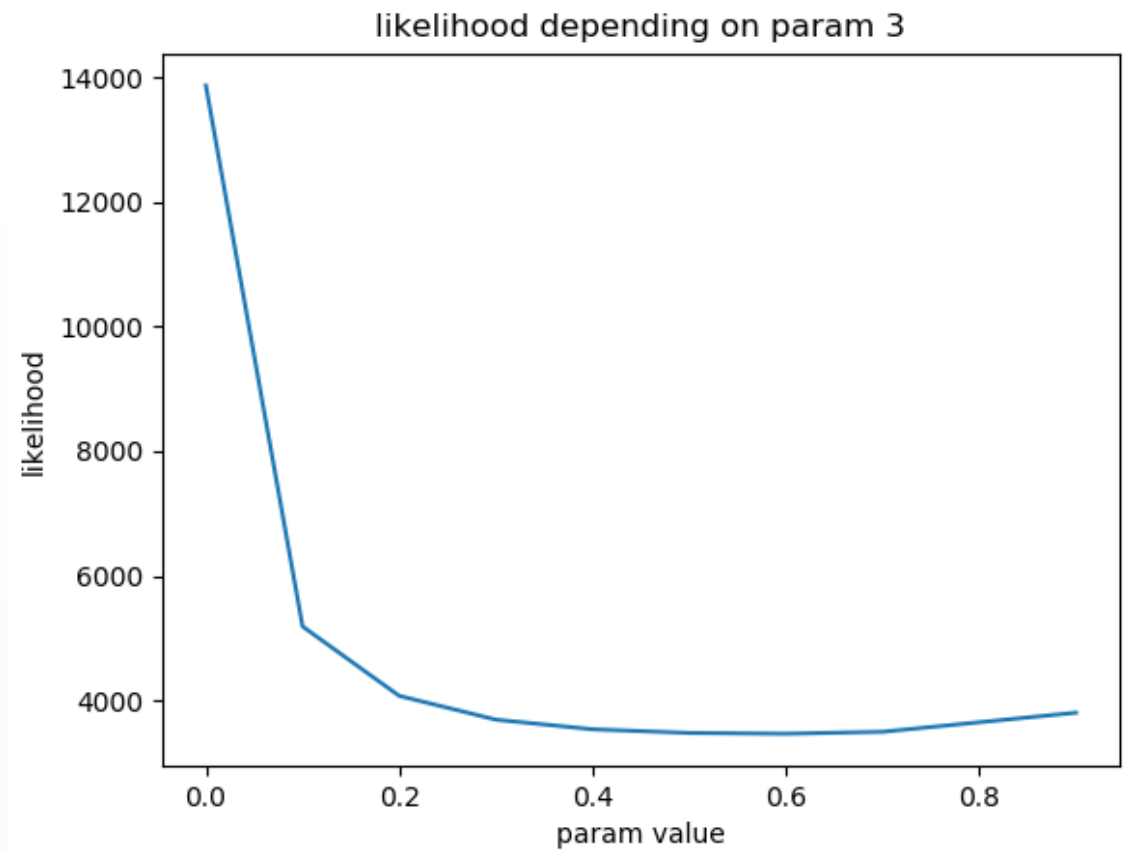- δ low (0.43)

Methods - MLE

Log-likelihood:

- L (θ) = $\sum_{i=1}^{4^l} log(\hat{\pi}_i(\theta)) * N_i$
- θ = (ρ, τ, δ, μ)
- l: number of CpGs
- $\hat{\pi}$: pattern distribution of simulation
- N_i : occurrences of pattern i in measured data

MLE: θ = arg max_θ L (θ)

Results- MLE

- Param 3: δ
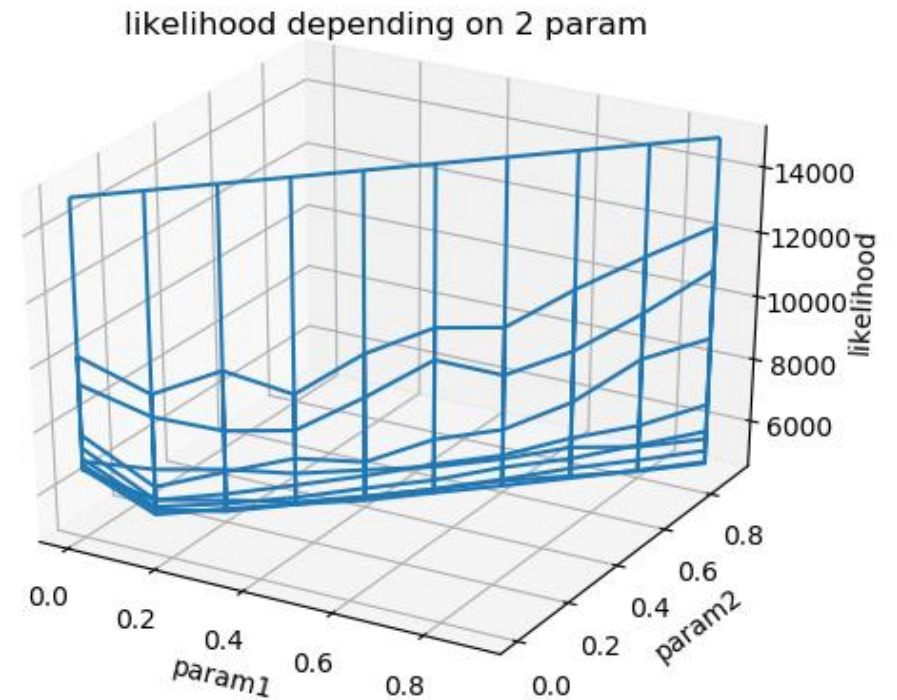- ρ, τ and μ fixed
- Neg. log-likelihood
- Large interval for δ

likelihood depending on param 3

Results- MLE

- Param 1: ρ
- Param 2: μ
- τ and δ fixed
- Neg. log-likelihood
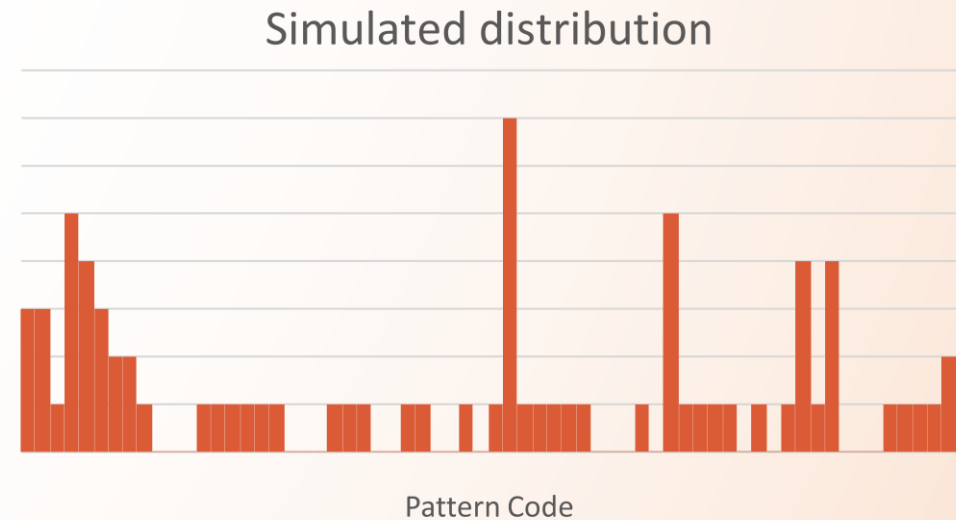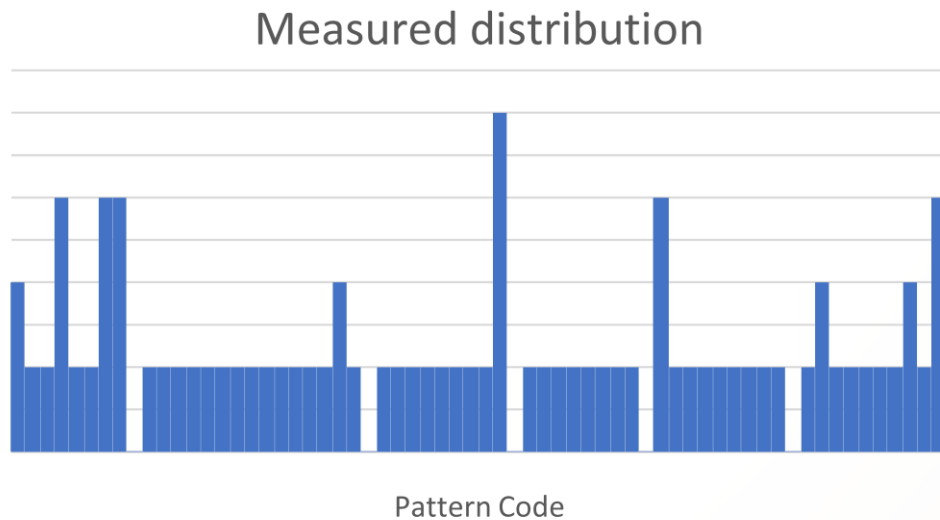- ρ and μ linear dependent

likelihood depending on 2 param

Methods- ABC

Approximate Bayesian Computation

- Given:
  - Measured methylation pattern distribution after cell division
  - Function simulating cell division

- Needed:
  - Distance between pattern distributions

- Output:
  - Best parameters

Methods - ABC

Pattern distributions – distance function

dist = ?

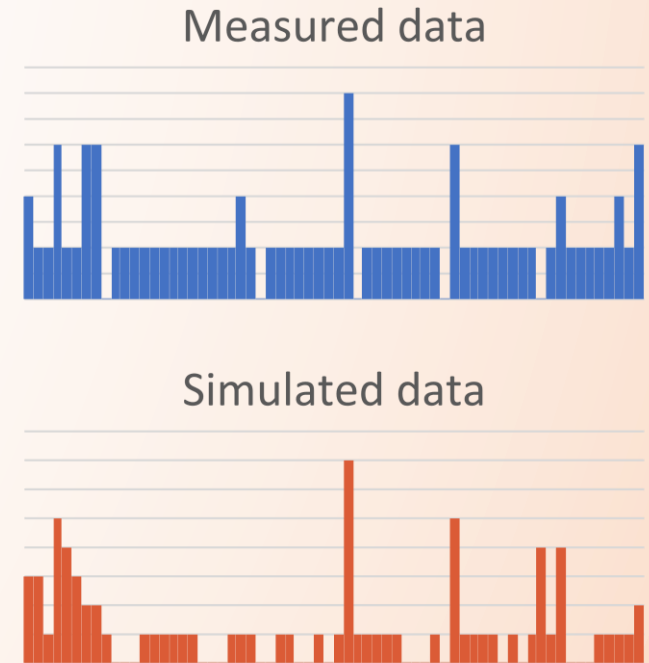# Methods - ABC

Pattern distributions – distance function

- $\text{dist}((X_1, \ldots, X_{l_1}), (Y_1, \ldots, Y_{l_2})) =$
  $$\sum_{i=1}^{l_1} \sum_{j=1}^{l_2} w_{ij}(X_i - Y_j)^2$$

- w_{ij}: distance between pattern i and j

Measured data

Simulated data

Methods - ABC

Methylation patterns – distance function

$$\text{dist}(\textcolor{blue}{i}, \textcolor{orange}{j}) = w_{ij}$$

Methods - ABC

Workflow:

- 1. Choose ρ, τ, δ and μ randomly

- 2. Simulate cell division

- 3. Compute dist$((X_1, \ldots, X_{l_1}), (Y_1, \ldots, Y_{l_2}))$

- 4. Repeat 1-3

- 5. Yield best ρ, τ, δ and μ

Outlook

- Implement distance function w_{ij} for ABC

- Use distance function to validate results from MCMC

- Test ABC for artificial data

Thank you!

Any Questions?

References

- A. Lück et al., A Stochastic Model for the Formation of Spatial Methylation Patterns

- A. Q. Fu et al., DNA Methyltransferases from Double-Stranded Methylation Patterns

- https://de.wikipedia.org/wiki/MCMC-Verfahren

- https://de.wikipedia.org/wiki/DNA-Methylierung

- https://en.wikipedia.org/wiki/DNA_methyltransferase

Pictorial sources

- Slide 2: https://www.biomol.de/die-top-10-der-dna-farbstoffe-und-sonden.html?id=821

- Slide 3: https://www.ncc.go.jp/en/ri/division/epigenomics/project/230/20170913152903.html

- Slide 7–9, 11, 14, 15: A. Lück et al., A Stochastic Model for the Formation of Spatial Methylation Patterns, p.3

- Slide 22: https://www.cartoonstock.com/cartoonview.asp?catref=cwln5040