

Saarland University  
Center for Bioinformatics  
Master's Program in Bioinformatics



Master Thesis in Bioinformatics

**Design and calibration of stochastic models for  
DNA methylation patterns**

submitted by

**Andrea Kupitz**

on March 2019

***Supervisor***

Professor Dr. Verena Wolf

***Advisor***

Alexander Lück

***Reviewers***

Professor Dr. Verena Wolf  
Professor Dr. Volkhard Helms

Zentrum für Bioinformatik



**Kupitz, Andrea**

*Design and calibration of stochastic models for DNA methylation patterns*

Master Thesis in Bioinformatics

Universität des Saarlandes

Saarbrücken, Germany

March 2019

## Declaration

*I hereby confirm that this thesis is my own work and that I have documented all sources used.*

*I hereby declare that the submitted digital and hardcopy versions of this thesis correspond to each other. I give permission to the Universität des Saarlandes to duplicate and publish this work.*

Saarbrücken, on March 2019

Andrea Kupitz

## Abstract

The expression of genes in the human genome is not only based on the DNA sequence; it relies on epigenetic modifications like DNA-methylations. Hereby, gene expression is inactivated by binding of methyl-groups to a cytosine phosphate guanine (CpG) dinucleotide at the promoter region of the concerned gene. The binding is performed by specific enzymes - the DNA Methyltransferases (DNMTs). The specific function of DNMTs is not fully determined.

In the following, the methylation of DNA is modelled using an Markov Chain Monte Carlo (MCMC) algorithm and parameters are estimated by Maximum Likelihood Estimation (MLE). Alternatively, parameter estimation is performed with an implementation of the Approximative Bayesian Computation (ABC) method. Moreover, a method to compare distributions of methylation patterns is provided.

We find differences in the function of the different DNMTs that are consistent with current biological findings. Thus, using this model, some parameters are difficult to identify because they seem to be conditionally dependant.



## Acknowledgments

I would like to thank deeply Prof. Dr. Verena Wolf and Alexander Lück, who introduced me to my master theme and always offered me their advice and support which I appreciated very much.

I owe thanks to both of them for the opportunity to develop this master thesis at the *Center for Bioinformatics*.

Furthermore I would like to thank all other people who put in touch with me during the development of this master thesis and who were very helpful in providing information.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Methods</b>	<b>3</b>
<b>3</b>	<b>Evaluation</b>	<b>5</b>
<b>4</b>	<b>Discussion</b>	<b>7</b>
<b>A</b>	<b>Regulation</b>	<b>9</b>





# List of Tables



# List of Figures



# List of Abbreviations

<b>ABC</b>	Approximative Bayesian Computation .....	ii
<b>CpG</b>	cytosine phosphate guanine .....	ii
<b>DNMTs</b>	DNA Methyltransferases .....	ii
<b>MCMC</b>	Markov Chain Monte Carlo .....	ii
<b>MLE</b>	Maximum Likelihood Estimation .....	ii



## Chapter 1

# Introduction





## Chapter 2

# Methods



# Chapter 3

## Evaluation



## Chapter 4

# Discussion

todo



# Appendix A

## Regulation

