

# Was kann durch eine C++ Implementierung der DNA Histonmodifikation erreicht werden?

Max Hild

*Abgegeben bei: Dr. Jörg Galle, Prof. Dr. Markus Scholz*

Universität Leipzig, Institut für medizinische Informatik, Statistik und Epidemiologie  
Neues Augusteum, Augustuspl. 10, 04109 Leipzig - Germany

**Abstract.** GWAS (Genome-Wide Association Studies) sind eine der am weitesten verbreiteten Methoden zur Identifizierung von genetischen Varianten, die mit komplexen Erkrankungen assoziiert sein könnten. Einer der treibenden Faktoren der Weiterentwicklung der Forschung zum menschlichen Genom war in den letzten Jahren jedoch eine Abkehr von der klassischen Sichtweise, dass die gesamte Heritabilität von Phänotypen lediglich durch die genetische Kodierung erklärt werden kann. In GWAS erreichten die relevanten Maßzahlen für die Heritabilität wiederholt lediglich kleine Werte. [1]

Hier zeigte sich, dass es eine Reihe weiterer Forschungszweige wie die Transkriptionsanalyse sowie die Epigenetik notwendig sind, um die fehlende Heritabilität der Phänotypen zu erklären. Epigenetische Mechanismen wie DNA-Methylierung und Histonmodifikationen spielen eine entscheidende Rolle bei der Regulation der Genexpression und könnten somit einen Teil der "verlorenen" Heritabilität aufklären. Diese Mechanismen sind dynamisch und können durch Umweltfaktoren beeinflusst werden, was sie zu einem spannenden Forschungsfeld macht.

## 1 Einleitung

Die Heritabilität von Phänotypen besser zu verstehen lässt Forscherinnen und Forscher leichter nachvollziehen, wie die genetische Information kodiert ist. Prohaska et al. argumentieren, dass die Bedeutung des regulatorischen Systems der Epigenetik für die Vererbung groß ist. [2] Diese Sichtweise deckt sich mit der Ansicht von McClellan et al, die in ihrer Arbeit die Ergebnisse der GWAS relativieren.

Eine der größten Hoffnungen an die GWAS war, dass man - genauso wie eine Vielzahl von mendelschen Erkrankungen auf DNA-Ebene eingegrenzt und das beteiligte Gen samt den Mutationen identifiziert werden konnte - einfach von Einzelgen-Erkrankungen auf komplexe multigenetische Erkrankungen schließen könnte. Das ist jedoch nicht eingetreten. Befürworter werden argumentieren, dass es funktioniert hat und dass allerlei faszinierende Gene identifiziert wurden, die beispielsweise eine Prädisposition zu oder einen Schutz vor Diabetes oder Brustkrebs verleihen, aber die Tatsache bleibt, dass der Großteil der Erbllichkeit in diesen Erkrankungen nicht den durch GWAS identifizierten Loci zugeschrieben werden kann, was eindeutig zeigt, dass dies nicht die universelle Lösung sein wird.

[1] Zu verstehen, wie die Modifikation der Histone und DNA-Methylierung funktioniert, ist daher ein wesentlicher Schritt, um die Lücke in unserem Verständnis der Heritabilität zu schließen.

## 2 Methoden

### 2.1 Modellierung der Histonmodifikation

Zur Modellierung der DNA-Histonmodifikation wurde eine C++ Implementierung entwickelt, die auf dem Modell von Prohaska et al. basiert. Das Modell simuliert die dynamischen Veränderungen der epigenetischen Zustände an CpG-Stellen über die Zeit. Jede CpG-Stelle kann einen von fünf möglichen Zuständen annehmen:

- U: Unmethyliert
- M: Methyliert
- H: Histon-modifiziert
- F: Vollständig modifiziert (methyliert und Histon-modifiziert)
- C: Komplex-gebunden

Die Übergänge zwischen diesen Zuständen werden durch ein stochastisches Modell gesteuert, das die biologischen Prozesse der Modifikation und Demodifikation nachbildet. Die Implementierung verwendet einen Markov-Prozess, bei dem die Übergangswahrscheinlichkeiten von den aktuellen Zuständen benachbarter CpG-Stellen abhängen.

### 2.2 Implementierungsdetails

Die C++ Implementierung besteht aus mehreren Klassen, die die verschiedenen Komponenten des Modells repräsentieren:

- Eine `CpGSite`-Klasse, die den Zustand einer einzelnen CpG-Stelle verwaltet
- Eine `Genome`-Klasse, die eine Sammlung von CpG-Stellen darstellt
- Eine `Simulator`-Klasse, die die zeitliche Entwicklung des Systems steuert

Die Simulation erzeugt bei jedem Zeitschritt eine neue Konfiguration aller CpG-Zustände basierend auf den Übergangsregeln. Die Ergebnisse werden in einer CSV-Datei gespeichert, die anschließend zur Visualisierung verwendet wird.

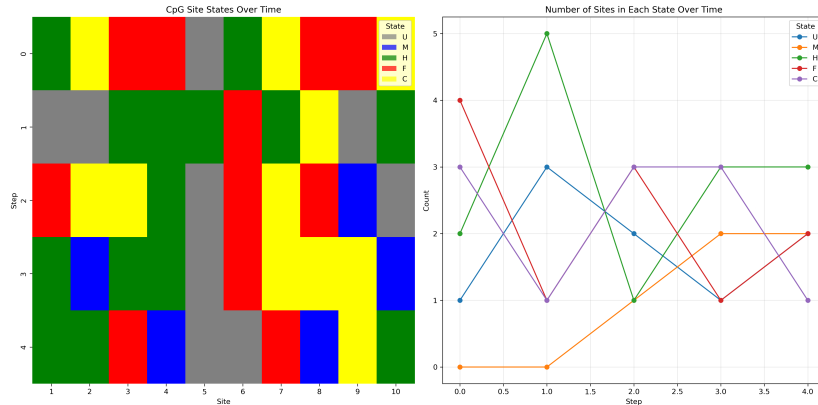


Fig. 1: Visualisierung der CpG-Zustände über die Zeit. Links: Heatmap der Zustände an verschiedenen CpG-Stellen über die Zeit (U: grau, M: blau, H: grün, F: rot, C: gelb). Rechts: Anzahl der CpG-Stellen in jedem Zustand über die Zeit.

### 3 Ergebnisse

Die Simulation ermöglicht es, die zeitliche Entwicklung der epigenetischen Zustände über mehrere Generationen zu beobachten. Abbildung 1 zeigt die Visualisierung der Simulation.

Die Visualisierung besteht aus zwei Teilen:

1. Eine Heatmap, die den Zustand jeder einzelnen CpG-Stelle über die Zeit darstellt
2. Ein Liniendiagramm, das die Häufigkeit jedes Zustands über die Zeit zeigt

Die Ergebnisse zeigen deutliche Muster in der räumlichen und zeitlichen Verteilung der epigenetischen Zustände. Insbesondere kann man beobachten, wie sich die Methylierungszustände in bestimmten Regionen zusammenhäufen und wie sich diese Cluster über die Zeit entwickeln.

### 4 Diskussion

Die C++ Implementierung des Modells bietet mehrere wichtige Vorteile:

#### 4.1 Recheneffizienz

Durch die Verwendung von C++ kann die Simulation effizienter durchgeführt werden als mit interpretierten Sprachen wie Python oder R. Dies ermöglicht die Simulation größerer Genome über längere Zeiträume, was für ein realistisches Modell der epigenetischen Prozesse unerlässlich ist.

## 4.2 Parametrisierbarkeit

Das Modell erlaubt die einfache Anpassung verschiedener Parameter, wie z.B. die Übergangsraten zwischen den Zuständen oder die Reichweite der Interaktionen zwischen benachbarten CpG-Stellen. Dies ermöglicht die systematische Untersuchung verschiedener biologischer Szenarien.

## 4.3 Erweiterbarkeit

Die objektorientierte Struktur der Implementierung macht es einfach, das Modell um zusätzliche Funktionalitäten zu erweitern, wie z.B. die Einbeziehung weiterer epigenetischer Mechanismen oder die Simulation von Umwelteinflüssen auf die epigenetischen Zustände.

## 5 Fazit

Die C++ Implementierung der DNA-Histonmodifikation ermöglicht ein tieferes Verständnis der dynamischen Prozesse, die der epigenetischen Regulation zugrunde liegen. Durch die Kombination von effizienter Simulation und anschaulicher Visualisierung können komplexe epigenetische Muster erkannt und analysiert werden.

Diese Art der Modellierung kann dazu beitragen, die "verlorene" Heritabilität besser zu verstehen, indem sie aufzeigt, wie epigenetische Mechanismen zur Vererbung phänotypischer Merkmale beitragen können. Die Ergebnisse unterstützen die These von Prohaska et al., dass das regulatorische System der Epigenetik eine wesentliche Rolle bei der Vererbung spielt, die über die klassische genetische Kodierung hinausgeht.

Zukünftige Erweiterungen der Implementierung könnten die Integration von experimentellen Daten zur Validierung des Modells sowie die Simulation spezifischer genetischer Erkrankungen umfassen, um potenzielle epigenetische Therapieansätze zu identifizieren.

## References

- [1] Jon McClellan and Mary-Claire King. Genetic heterogeneity in human disease. *Cell*, 141(2):210–217, 4 2010.
- [2] Sonja J. Prohaska, Peter F. Stadler, and David C. Krakauer. Innovation in gene regulation: The case of chromatin computation. *Journal of Theoretical Biology*, 265(1):27–44, 3 2010.