

# Chromatin Compartments and Selection on X

How Chromatin Compartments Align with Regions Under Selection in  
Primates

Author:

**Søren Jørgensen**

Student ID: 201906763

Supervisor:

**Kasper Munch**

Bioinformatics Research Center (BiRC)

Molecular Biology and Genetics

Aarhus University, Denmark



***MSc. Bioinformatics***

2025-01-15

Submitted in fulfillment of the requirements of the degree of MSc. Bioinformatics

---

## ABSTRACT

The X chromosome is uniquely exposed to selective pressures due to the lack of a second copy in the hemizygous sex, leaving no buffer against deleterious mutations, and giving unique inheritance patterns. Combined with its high density of essential genes related to reproduction and brain function, this suggests the presence of biological mechanisms that safeguard the integrity of the X chromosome.

In this study, the 3D chromatin architecture of the X chromosome in rhesus macaque (*Macaca mulata*) is investigated in the context of evolutionary pressures and genetic drivers. To ensure transparency and reproducibility, we adopt a comprehensive computational framework for publishing a version-controlled, fully reproducible analysis.

We compare two Hi-C analysis frameworks, *HiCExplorer* and *cooler/cooltools* (Open2C), on a subset, finding Open2C to be most flexible. The ICE method (Iterative Correction and Eigendecomposition) was used to infer conventional and refined A/B compartments for fibroblast and four stages of spermatogenesis. We find 200 kbp transition-zones between A/B-compartments on the X chromosomes in both fibroblasts and round spermatids that align well with strong selective sweeps in humans (ECH-regions), but not with strong negative selection in baboons (*Papio* spp.). We find that most edges either overlap or are in significant proximity of each other when comparing regions under selection in human and baboons with A/B-compartments inferred Hi-C matrices at 100kb resolution and restricting eigendecomposition along the X chromosome to 10Mb windows. We discuss the biological meaning of these findings, where conserved chromatin features may help to retain non-advantageous alleles, hinting to the role of structural features aiding in genome evolution.

---

# Table of contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Evolution on X . . . . .	1
1.1.1	Extended Common Haplotypes (ECH) on human X . . . . .	2
1.1.2	Disproportionate minor parent ancestry on baboon X . . . . .	3
1.2	Chromatin Architecture . . . . .	4
1.3	3C: Chromatin Conformation Capture . . . . .	5
1.4	Hi-C: High-Throughput 3C . . . . .	6
1.4.1	Hi-C Library preparation . . . . .	6
1.4.2	Hi-C Data Analysis . . . . .	7
1.5	Reproducibility Infrastructure . . . . .	12
1.5.1	GWF: workflow management for High-Performance Computing (HPC) . . .	13
1.5.2	Project Initialization . . . . .	13
1.5.3	git and GitHub . . . . .	14
<b>2</b>	<b>Methods</b>	<b>15</b>
2.1	Fetching raw data . . . . .	15
2.1.1	Fetching and indexing the reference . . . . .	16
2.2	HiCExplorer trials . . . . .	16
2.3	Open2C pipeline . . . . .	18
2.4	Genomic Intervals . . . . .	23
2.4.1	Full Regions, Edges, and Limits . . . . .	23
<b>3</b>	<b>Results</b>	<b>25</b>
3.1	HicExplorer Trials . . . . .	25
3.1.1	Quality Control . . . . .	25
3.1.2	Correction . . . . .	27
3.1.3	Eigenvectors . . . . .	27
3.2	Open2c ecosystem . . . . .	28
3.2.1	Quality Control . . . . .	28
3.2.2	Correction . . . . .	30
3.2.3	Compartments (Eigenvectors) . . . . .	33
3.3	Comparing Genomic Intervals . . . . .	35
3.3.1	Compartment Edges (transition zones) . . . . .	36
3.3.2	Testing against baboon regions under strong selection . . . . .	37
3.3.3	Testing limits . . . . .	38
<b>4</b>	<b>Discussion</b>	<b>41</b>
4.1	On Reproducibility . . . . .	41
4.2	On Methodology . . . . .	42
4.3	On Results . . . . .	43
4.3.1	Narrowing the parameter space for testing . . . . .	44
4.3.2	Correlating genomic regions . . . . .	45
4.3.3	Gene Drive, Selfishness and its Effect on Selection . . . . .	47
<b>5</b>	<b>Conclusion and future work</b>	<b>49</b>

<b>Bibliography</b>	<b>51</b>
<b>A Appendix</b>	<b>55</b>
Data Availability . . . . .	55
Acknowledgements . . . . .	55
People . . . . .	55
Compute . . . . .	55
Use of Generative Artificial Intelligence . . . . .	55