

Collective Processes In Cellular Reprogramming

DISSERTATION

zur Erlangung des akademischen Grades

Doctor rerum naturalium

(Dr. rer. nat)

vorgelegt

der Fakultät Physik

dem Bereich Mathematik und Naturwissenschaften

der Technischen Universität Dresden

von

Aida Mohammadzadehhashtroud

geboren am 04.08.1994 in Tehran

Max-Planck-Institut für Physik komplexer Systeme

December 2023

Eingereicht am 14.12.2023

Verteidigt am 08.05.2024

Gutachter:

Prof. Dr. Frank Jülicher

Prof. Dr. Steffen Rulands

Prof. Dr. Benedikt Sabass

Acknowledgements

To begin with, I would like to express my gratitude to my Ph.D. supervisor, Steffen Rulands. Steffen, your scientific precision and inquiring mindset constantly motivated and challenged me. Thank you for introducing me to the fascinating world of epigenetics and giving me the opportunity to work with you. I am also grateful to my IMPRS thesis advisory committee members, Frank Jülicher, Meritxell Huch, and Carl Modes, for their valuable feedback and expert evaluation at the annual TAC meetings.

I thank all the former and current members of Rulands lab, the Biological Physics department at the Max Planck Institute for the Physics of Complex Systems, and the Department for Statistical and Biological Physics at Arnold Sommerfeld Center. The fruitful scientific discussions and friendship fostered during lunch and coffee gatherings, hikes, bouldering, boogie-woogie dance, and our joyful summer and Christmas parties have been invaluable. Going through my PhD journey, which started shortly before the Corona crisis, would have been more complicated without you. I must also extend my thanks to Hubert and Thomas from the IT department at the PKS for their tireless efforts in maintaining the institute's computational power.

My heartfelt appreciation goes to my cherished friends, outstanding scientists, and remarkable human beings. To Leonie, thanks for your kindness on the bad days and your joy on the good ones. To Sudarshana, thanks for being a good listener and consistently insightful suggestions, regardless of the situation. Our birthday trips with many museum visits are always among my best memories. To Felix, thanks for your constant support and empathetic understanding. I'm incredibly thankful for our regular discussions in the last months, as they enabled me to untangle my thoughts and write them into this thesis.

Finally, I am deeply grateful to my parents for their unwavering patience and trust. To my Mom, thank you for waking me up every morning with your love and care despite the poor internet connection. To my little sister Athena, thanks for your unconditional love. No matter the physical distance that separates us, we are always there for each other.

Lastly, I dedicate this thesis to all great, unknown women who bravely champion their rights and tirelessly pursue their dreams. May you embrace the life and freedom you rightfully deserve.

Abstract

Epigenetics comprises chemical modifications of the DNA and the proteins that the DNA is wrapped around them. These modifications play key roles in establishing and maintaining cellular identity throughout development and adulthood. In recent years, it has become increasingly clear that these actions are more dynamic than initially believed. The alteration of cellular identities during regeneration, ageing, and the formation of tumors is closely linked to systematic changes in epigenetic modifiers. The emergence of cutting-edge single-cell sequencing technologies has enabled thorough explorations of biological processes with high molecular precision. Nevertheless, the regulation of cellular behavior is intricately tied to collective processes occurring in both spatial and temporal dimensions, operating on the mesoscopic and macroscopic scales. However, these larger scales cannot be straightforwardly deduced from microscopic measurements along the DNA sequence [1]. Consequently, the findings obtained from sequencing experiments stay at the descriptive level until they are coupled with methodologies capable of discerning collective degrees of freedom. Here, using statistical physics tools and sequencing technologies, we study the collective processes underlying epigenetic dynamics in cells that change their identity over time. Specifically, we investigate collective epigenetic processes during ageing and the reprogramming of cells after injury.

In the first part of this thesis, we study the mechanistic basis of epigenetic modifications during ageing. Despite the accuracy of machine learning models in predicting the biological age based on epigenetic DNA methylation marks, these tools do not inform about the mechanistic basis of epigenetic ageing. We show that epigenetic ageing is reflected in systematic and collective changes in DNA methylation marks during ageing, which manifests in the stereotypical behavior of two-point correlation functions. We devise a stochastic theory that comprises competition of antagonistic enzymes at the boundaries of genomic regions with atypically high content of cytosine-guanine pairs. We systematically coarse-grain this theory to derive a macroscopic description in terms of a phase-field theory. This model predicts the changes in two-point correlation functions during ageing and explains diverse observations in the field of epigenetic ageing.

In the second part of this thesis, we study the collective epigenetic processes during the regeneration of the liver after injury. In particular, we study the interplay between DNA methylation and the accessibility of chromatin and show the necessity for emergent memory of past injuries in the system. This memory is achieved by considering an effective projection between different scales of epigenetic modifications.

In total, in this thesis, we derived theoretical descriptions of epigenetic processes that have so far only been studied descriptively. We showed that both epigenetic alterations during ageing and during reprogramming rely on an interplay between collective biochemical processes and the geometry of the DNA. With this work, we show how linear DNA sequencing can inform

about collective epigenetic processes in space and time.

Zusammenfassung

Die Epigenetik umfasst chemische Veränderungen der DNA und der Proteine, um die die DNA gewickelt ist. Diese Veränderungen spielen eine Schlüsselrolle bei der Schaffung und Erhaltung der zellulären Identität während der Entwicklung und im Erwachsenenalter. In den letzten Jahren wurde immer deutlicher, dass diese Vorgänge dynamischer sind als ursprünglich angenommen. Die Veränderung der zellulären Identität während der Regeneration, der Alterung und der Entstehung von Tumoren ist eng mit systematischen Veränderungen der epigenetischen Modifikatoren verbunden. Das Aufkommen modernster Einzelzell-Sequenzierungstechnologien hat eine gründliche Erforschung biologischer Prozesse mit hoher molekularer Präzision ermöglicht. Dennoch ist die Regulierung des zellulären Verhaltens eng mit kollektiven Prozessen verbunden, die sowohl in räumlicher als auch in zeitlicher Hinsicht auf mesoskopischer und makroskopischer Ebene ablaufen. Diese größeren Skalen lassen sich jedoch nicht ohne Weiteres aus mikroskopischen Messungen entlang der DNA-Sequenz ableiten [1]. Daher bleiben die Erkenntnisse aus Sequenzierungsexperimenten so lange deskriptiv, bis sie mit Methoden zur Identifizierung kollektiver Freiheitsgrade kombiniert werden. In dieser Arbeit kombinieren wir Werkzeuge aus der statistischen Physik und Sequenzierungstechnologien, um die kollektiven Prozesse zu untersuchen, die der epigenetischen Dynamik in Zellen zugrunde liegen, die ihre Identität im Laufe der Zeit ändern. Konkret untersuchen wir kollektive epigenetische Prozesse während der Alterung und der Reprogrammierung von Zellen nach Verletzungen.

Im ersten Teil dieser Arbeit untersuchen wir die mechanistischen Grundlagen der epigenetischen Veränderungen während des Alterns. Obwohl Modelle des maschinellen Lernens das biologische Alter auf der Grundlage von epigenetischen DNA-Methylierungsmarkierungen genau vorhersagen können, geben diese Werkzeuge keine Auskunft über die mechanistische Grundlage der epigenetischen Alterung. Wir zeigen, dass sich die epigenetische Alterung in systematischen und kollektiven Veränderungen der DNA-Methylierungsmarkierungen während des Alterns widerspiegelt, was sich in einem stereotypen Verhalten der Zweipunkt Korrelationsfunktionen zeigt. Wir entwickeln eine stochastische Theorie, die den Wettbewerb antagonistischer Enzyme an den Grenzen von Genomregionen mit atypisch hohem Gehalt an Cytosin-Guanin-Paaren umfasst. Wir vergrößern diese Theorie systematisch und leiten daraus eine makroskopische Beschreibung in Form einer Phasenfeldtheorie abzuleiten. Dieses Modell sagt die Veränderungen der Zweipunkt Korrelationsfunktionen während der Alterung voraus und erklärt verschiedene Beobachtungen auf dem Gebiet der epigenetischen Alterung. Im zweiten Teil dieser Arbeit untersuchen wir die kollektiven epigenetischen Prozesse während der Regeneration der Leber nach einer Verletzung. Insbesondere untersuchen wir das Zusammenspiel zwischen der DNA-Methylierung und der Zugänglichkeit des Chromatins und zeigen die Notwendigkeit eines emergenten Gedächtnisses für vergangene Verletzungen im System. Dieses Gedächtnis wird durch die Berücksichtigung einer effektiven Projektion zwischen verschiedenen Skalen epigenetischer Modifikationen erreicht.

Insgesamt haben wir in dieser Arbeit theoretische Beschreibungen von epigenetischen Prozessen

abgeleitet, die bisher nur deskriptiv untersucht worden sind. Wir haben gezeigt, dass sowohl epigenetische Veränderungen während der Alterung als auch während der Reprogrammierung auf einem Zusammenspiel zwischen kollektiven biochemischen Prozessen und der Geometrie der DNA beruhen. Mit dieser Arbeit zeigen wir, wie lineare DNA-Sequenzierung über kollektive epigenetische Prozesse in Raum und Zeit informieren kann.

Contents

1	Introduction	1
1.1	Epigenetics	3
1.1.1	DNA methylation	3
1.1.2	Histone modification	5
1.2	DNA sequencing experiments	6
1.2.1	Bisulfite sequencing experiment	6
1.2.2	Chromatin-immunoprecipitation sequencing	7
1.3	Theoretical background	8
1.3.1	The phase field approach	9
1.3.2	On the memory effect and generalized Langevin equation	9
1.4	Research goal and overview of the thesis	11
2	Collective epigenetic processes during ageing	13
2.1	Introduction	13
2.1.1	Epigenetic processes in ageing	13
2.1.2	DNA methylation ageing clocks	18
2.1.3	Research questions and overview of this chapter	20
2.2	Results	21
2.2.1	Sequencing experiment	21
2.2.2	DNA methylation ageing as a collective epigenetic process	22
2.2.3	Towards a biophysical mechanism of DNA methylation ageing	28
2.2.4	Application to single-cell data	38
2.3	Summary and discussion	44
3	Epigenetic reprogramming during regeneration	47
3.1	Introduction	47
3.1.1	Epigenetic regulation in liver regeneration	47
3.2	Research questions and overview of this chapter	49
3.3	Results	50
3.3.1	Sequencing experiment	50
3.3.2	Collective epigenetic processes during regeneration	52
3.3.3	Towards a biophysical mechanism of regeneration	55
3.4	Summary and discussion	60

4 Conclusion	61
Appendices	
A Multiplicative changes in correlation functions	65
B From the Master equation to the Fokker-Planck equation	67
C Preliminary analysis of sequencing experiments	68
List of Figures	70
Bibliography	76

Chapter 1

Introduction

“What is your secret?” the traveler asked.

“I am the Tree of Life.” it replied, “and my fruit grants immortality. Eat of it, and you shall never age, never die.”

– Rumi, *Masnavi*

The desire to prolong life has been a recurring topic in human history, illustrated by the age-old story “The Epic of Gilgamesh.” Although the quest for understanding life and conquering death has not always been based on scientific principles, scientists recognize it as a complex issue encompassing various fields, such as mathematics, physics, and biology [2–5]. Some of the questions to be tackled include how cells organize themselves into tissues and organs, how they make decisions in noisy environments, how aging occurs, and whether it can be reversed and at what cost. We need an in-depth understanding of cell fate regulation to interpret development, post-injury regeneration, aging, and diseases like cancer. However, due to the vast number of factors that affect cell behavior, we should narrow our focus to the molecular level, where existing technology can observe at the smallest scale [6]. Gene expression has traditionally been the main area of focus when studying molecular regulation [7], but recent advances in single-cell biology have allowed scientists to investigate further. These new studies investigate other regulations, including the dynamic changes in DNA folding, modifications in protein complexes encircling DNA, and chemical alterations to DNA. Collectively, these additional layers of regulation are known as epigenetics [8].

Recent advancements in single-cell biology have provided researchers with unprecedented molecular information regarding the mechanisms underlying cell fate regulation [9]. Due to advancements in technology, we can now acquire data on gene expression, DNA epigenetic modifications, and chromatin organization at the level of individual loci and from single cells within living organisms. Moreover, single-cell multi-omics sequencing allows researchers to examine various regulatory layers within the same cell [10, 11]. These breakthroughs have allowed us to gain valuable insights into molecular behavior and identity of the cells, such that single-cell genomics was named the scientific breakthrough of the year 2018 by Science.

However, it is important to recognize that biological function cannot be solely explained by the molecular processes taking place within individual cells. The emergent (collective) states, which result from interactions between molecular processes occurring at multiple layers, ultimately determine cellular function. Furthermore, identifying and interpreting specific regulatory processes on the molecular level requires a thorough understanding of how macroscopic processes emerge on the cellular and tissue scale. The current challenge lies in integrating precise quantitative data at the microscopic scale with the macroscopic processes that rule cellular and tissue function. To draw a parallel from physics, this concept is similar to comprehending collective phenomena, like high-temperature superconductivity, through an understanding of the governing processes at the atomic level [11]. The emergence of new collective properties in the *macroscopic* realm from the dynamics of the *microscopic* variables is the topic of statistical physics that provides a probabilistic connection between the two regimes. Another essential concept is the energy requirement for biological systems to construct and uphold intricate spatiotemporal structures [12], known as the systems outside of equilibrium. To describe systems out of equilibrium, physicists primarily employ either Langevin or master equations based on the specific nature of the problem. Langevin equations elucidate the stochastic dynamics of a field, providing a simplified representation of microscopic variables. In contrast, master equations describe the temporal and spatial alterations in a probability distribution. The description of Langevin equations can be expanded using Fokker-Planck equations, which mirror the probability distribution of the field like the master equation. The crucial distinction between these approaches is that master equations accommodate discrete changes in variables, whereas Fokker-Planck equations operate continuously [1, 11].

Nevertheless, advances in artificial intelligence (AI), may have a better performance in the analysis of a big data set, such as the the output of the sequencing experiments [13]. AI, however, is descriptive and lacks a general conceptual framework for bridging the gap between microscopic and macroscopic scales. To gain insight into cellular behavior mechanisms and find effective cures, we need a mechanistic approach to decipher biological processes.

In this thesis, we take an interdisciplinary approach and combine novel analysis of sequencing experiments and statistical physics to interpret collective processes underlying cellular behavior during biological ageing and regeneration after injury. In the following part of this chapter, we first describe the biology of epigenetics and their dynamic changes throughout life, which serves as a fundamental mechanism underlying cell behavior. Then, we will give a detailed description of current technologies in genomics. We will also discuss some theoretical background we applied in the next chapters. Finally, we will introduce the main research goals in this thesis.

1.1 Epigenetics

While every cell in our body possesses the same DNA sequence, they have different functionality. The process by which a stem cell, the precursor cell responsible for generating all differentiated cell types in the body, decides to commit to a specific cellular lineage over another is dictated by the intricate interplay of multiple regulatory layers. Epigenetics is related to phenotypic states not written into the genes but inherited as gene expression patterns influenced by environmental factors. [14]. Epigenetic processes involve a range of mechanisms, including alterations in chromatin structure, chemical modifications of histones, and DNA methylation [1, 15]. In this thesis, our primary focus lies in examining DNA methylation and histone methylation, which represent a key facet within the realm of epigenetic modifications. Epigenetics is a dynamic process that occurs throughout one's lifespan and can be altered at various time scales. In a short time scale (a few days), epigenetic changes occur immediately after fertilization during early development, in primordial germ cells during later development, and during erythropoiesis, which is the formation of red blood cells [16, 17]. These modifications also take place during the reprogramming process after injury. In addition to these rapid alterations, epigenetic modifications can also occur over a long time scale and play a crucial role in the ageing process. In the second chapter of this thesis, we will delve into the topic of epigenetic changes that occur during biological ageing. Following that, in the third chapter, our focus will shift to exploring the epigenetic modifications that the liver undergoes during the regeneration process after injury. Before diving into these topics, we first introduce DNA methylation and histone modification, which we will be studying and analyzing in the upcoming chapters.

1.1.1 DNA methylation

DNA methylation (DNAm) is a process that alters the structure of nucleic acids in DNA. In mammals, it primarily affects the cytosine (C) base when it is next to a guanine (G). This specific base pair is known as CpG, with the "p" representing the phosphorous between cytosine and guanine. When a methyl group binds to this cytosine, it will be methylated. This cytosine will be called 5-methylcytosine (5mC), and the process is associated with gene repression¹. The establishment of new DNA methylation marks is driven by a class of methyl transferase enzymes, DNMT3a, DNMT3b, and DNMT3l, and their maintenance after replication during cell division relies on DNMT1. In particular, DNMT3a/b plays an active role in modifying the epigenetic status of CpGs, converting them into methylated cytosines, and DNMT3l serves to recruit DNMT3a/b [18–21].

Additionally, DNA methylation is a reversible process through mechanisms known as demethylation. DNA demethylation might occur passively due to the lack of DNMT1 and the corresponding lack of DNA methylation maintenance during replication or actively via the

¹Switching off the genes, the DNA units containing information needed to specify physical and biological traits.

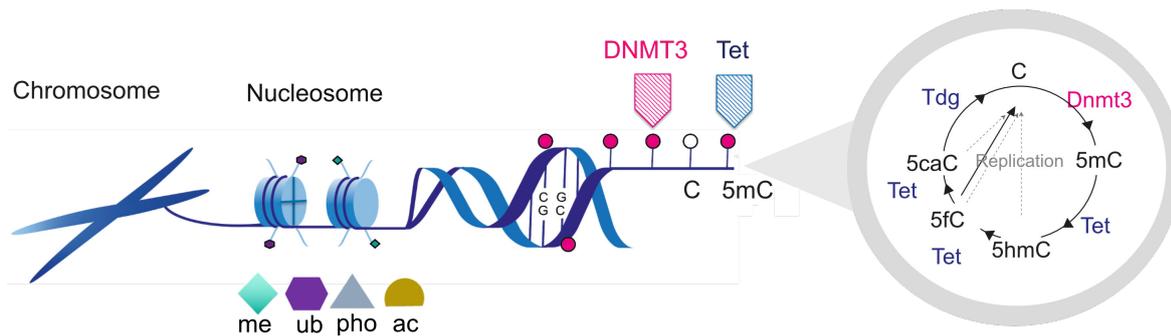


Figure 1.1: Illustration of two major epigenetic mechanisms (i) DNA methylation occurs when methyl groups tag DNA and repress (through the effect of DNMT3) or activate (through the effect of TET) genes. Removing the effect of methylation (demethylation) happens in different steps, which have been shown in the cycle at the right. (ii) Nucleosome or histone modification occurs when the binding of epigenetic factors (methylation, ubiquitination, phosphorylation, acetylation) to histone tails alters the extent to which DNA is wrapped around histones and the availability of genes in the DNA to be activated.

conversion of 5mC to 5hmC through the effect of the Ten-eleven translocation (TET) family of enzymes through oxidizing the methyl group. 5hmC can undergo additional oxidation and eventually turn an unmodified cytosine. [18–21], see Fig.1.1. Most CpG methylation occurs mainly in intergenic regions, repetitive elements, and gene bodies. In contrast, regions of the genome with high CpG density (known as CpG islands or CGIs), which colocalize with gene promoters, are unmethylated [22]; and CpG sites located in regions just outside CGIs (termed shores and shelves) or in distal regulatory elements, notably enhancers, exhibit the highest variability in DNA methylation [23].

DNA methylation is initially imposed during early development to ensure that differentiated cells remain lineage-restricted. During the development process, significant changes occur in the levels of DNA methylation both globally and locally. Specifically, embryonic cells acquire DNA methylation marks from both the father and mother after fertilization. However, in the initial days of mouse embryo development, these marks are erased and followed by a new round of DNA methylation. This results in a substantial increase in the overall level of DNA methylation, going from around 20% to approximately 80% [1]. This alteration in DNA methylation occurs before embryonic cells make their initial developmental decisions and play a crucial role in shaping their destiny. Furthermore, any disturbances in methylation patterns, such as DNMT3 mutations, can prove fatal for the embryo, and in adulthood, they serve as significant indicators for the onset of cancer. The existing DNA methylation pattern that is established during development can undergo modifications during different occasions in response to environmental changes, such as ageing and tissue regeneration. We will discuss ageing-associated changes in DNA methylation in chapter 2 with more detail.

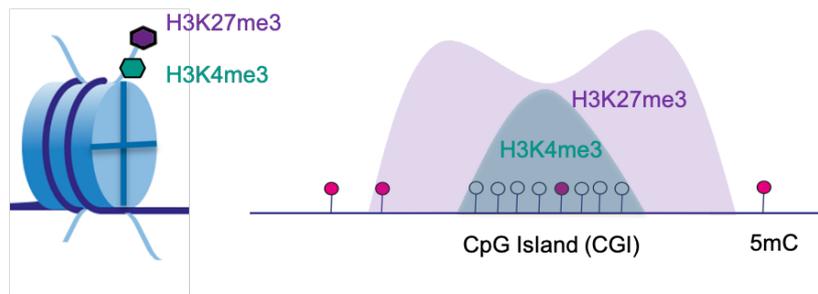


Figure 1.2: Illustration of a bivalent CpG island. H3K4me3 usually affects the center of the CpG islands, while H3K27me3 usually affects the shores in the case of bivalent CGI.

1.1.2 Histone modification

The DNA molecule is organized into chromatin. The basic building block of chromatin is the nucleosome, consisting of 147 base pairs of DNA wrapped around an octamer composed of two of each core histone protein (H2A, H2B, H3, and H4). [24]. In addition, the epigenetic modifications also occur at the level of individual histones. These modifications include lysine acetylation, methylation, ubiquitination, sumoylation, arginine methylation, as well as phosphorylation of serine and threonine residues. [25, 26]. While histone acetylation typically correlates with gene activation, the effects of histone methylation vary depending on the specific residue modified, being associated with either activation or repression. [27]. Histone methylation on lysine or arginine residues has been described to be associated with gene activation; for example, histone H3 lysine 4 (H3K4), histone H3 lysine 36 (H3K36), histone H3 lysine 79 (H3K79), histone 4 arginine 3 (H4R3), histone 4 arginine 17 (H4R17); or with gene repression, including histone H3 lysine 9 (H3K9), histone H3 lysine 27 (H3K27), histone H4 lysine 20 (H4K20), and histone 3 arginine 8 (H3R8) [24, 26, 28]. H3K4me3 and H3K27me3 are two of the most extensively studied histone modifications. Whereas H3K4me3 is characteristic of active promoters associated with the Trithorax group (TrxG), H3K27me3 is associated with transcriptional repression via the polycomb complex group (PcG). Because TrxG and PcG proteins act antagonistically to regulate the activated and repressed states of gene expression, H3K4me3 and H3K27me3 were thought to be mutually exclusive. But this assumption was challenged by the discovery of bivalent genomic regions characterized by the simultaneous presence of H3K4me3 and H3K27me3 (Fig.1.2), which can be found predominantly at CpG islands in embryonic stem cells (ESCs) [27, 29]. Almost all native CGIs in mouse ESCs are non-methylated at the DNA level and coincide with high levels of H3K4me3, which is often seen as the signature histone mark of CGIs [30]. About one-third of the H3K4me3-marked CGIs in mouse ES cells also carry H3K27me3 and are therefore defined as bivalent [31].

During ageing process, upon regeneration, or due to cancers, chromatin status, including variation in histone levels, exchange of histone variants, histone modifications, and nucleosome occupancy, is dynamically modified. Chapter 2 will discuss histone modification changes

during ageing, focusing on histone methylation.

1.2 DNA sequencing experiments

We introduced methylation as a primary layer of epigenetic modifications that dynamically regulate cellular behavior. Here, we will discuss how one can detect these marks through Bisulfite sequencing experiments (BS-Seq). Furthermore, we introduce the Chromatin-immunoprecipitation sequencing experiment (ChIP-seq) as a tool one can use to study the binding sites associated with chromatin modifications as another crucial layer of epigenetic modifications.

1.2.1 Bisulfite sequencing experiment

Bisulfite sequencing experiment (BS-Seq) is known as a gold-standard method for the detection of DNA methylation as it provides a quantitative and efficient approach to identify 5-methylcytosine at single base-pair resolution [32]. BS-Seq protocol can also be shared and parallelized as a powerful tool between many sequencing experiments to investigate regulatory relationships on top of chemical modifications such as profiling the chromatin accessibility [33]. In this method, sodium bisulfite is used as a chemical treatment for DNA after its extraction from the nucleus and then sequenced and amplified via polymerase chain reaction (PCR). Following bisulfite treatment, cytosines that are unmethylated undergo deamination, converting into uracils, which are subsequently transformed into thymidines. However, methylated cytosines resist deamination and are recognized as cytosines during sequencing [34]. The position of methylated cytosines can be discerned by comparing sequences before and after treatment. This involves aligning the treated and untreated sequences and mapping each fragment back to its location along the DNA sequence. [35]. The last step contains a significant amount of bioinformatics. Starting from raw FastQ files (the outcome of a sequencing facility), the quality of the sequenced reads must be determined, and low-quality reads should be removed using TrimGalore software. Then, the trimmed reads are aligned to the reference genome using Bismark. In the end, biological duplicated reads are merged, and the methylation marks will be extracted for unique mapped reads and will be ready for further analysis [36]. Each introduced step contains biological variability and technical noises; hence, advanced pipelines are necessary to derive accurate quantitative predictions from sequencing experiments. Furthermore, modern sequencing technologies generate substantial data volumes demanding that raw analysis be time-consuming. The time and computation cost depend on sequencing parameters like read length, genome size, duplicated content, and alignment parameters [37]. In Chapter 3, we will briefly provide some results of the preliminary analysis of raw data while working on the liver regeneration data set.

In general, bisulfite sequencing experiments can be conducted in two types: bulk and single-cell sequencing. In bulk experiments, methylation information comes from the average

1.2 DNA sequencing experiments

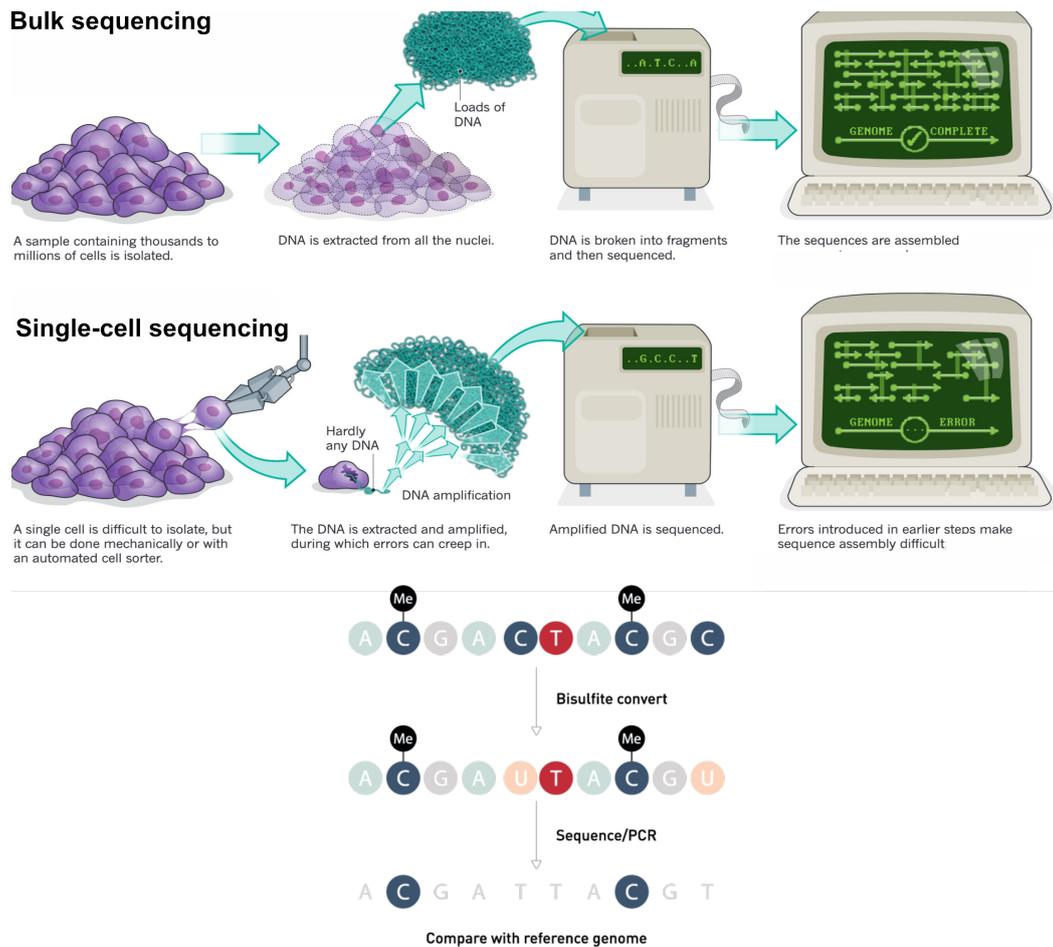


Figure 1.3: Sketch of the bisulfite sequencing experiments, adapted by permission from [38], (License Number 5659270261727).

methylation status of many copies of a particular fragment driven from many samples. Yet, single-cell sequencing has the capability to detect the methylated cytosines on individual DNA strands, see Fig 1.3.

Note that these sequencing experiments are static. To study the dynamic and track the changes in methylation information during aging, regeneration, or development, one needs to do a time course sequencing experiment by simply repeating the sequencing experiment at different time points.

1.2.2 Chromatin-immunoprecipitation sequencing

Chromatin-immunoprecipitation sequencing (ChIP-seq) is the most widely used technique for analyzing protein interactions with DNA. In this sequencing technique, cells are cross-linked, fragmented, and immunoprecipitated with an antibody specific to the target protein and then will be aligned to a reference genome. The corresponding computational analysis

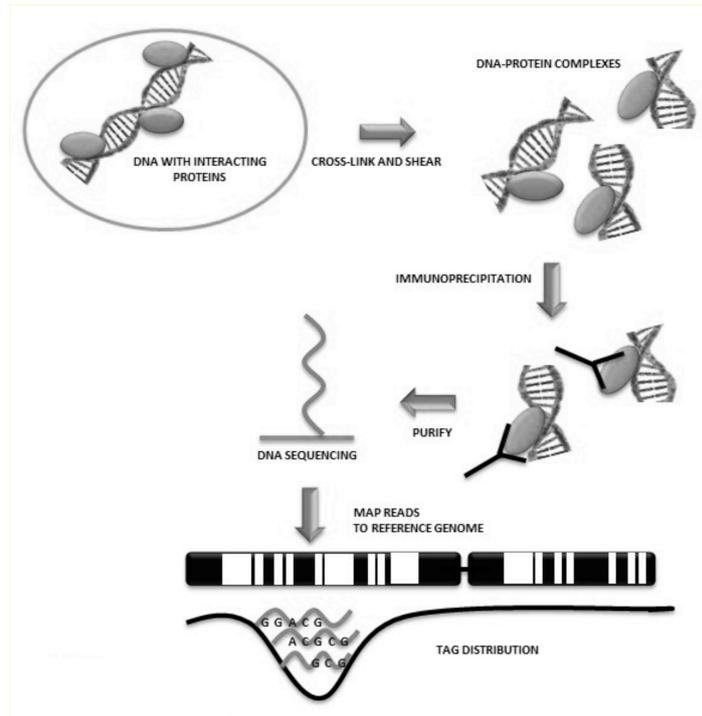


Figure 1.4: Sketch of ChIP-seencing experiments, adapted from [39]

depends on detecting "peaks", regions of the genome where multiple reads align that indicate protein binding enrichment. However, difficult computational challenges will be posed in identifying peaks due to irregular noise and bias on various levels, which are caused by biological variability, a variety of domains, and the experiment environment. Although many peak-calling methods have been developed, the current computational tools still require, in some cases, human manual inspection, such as determining if fewer high-quality peaks are preferred over lower-quality peaks [40].

1.3 Theoretical background

In this section, we discuss the key theoretical concepts from statistical physics that we have applied to genomics in this thesis. Statistical physics offers efficient conceptual and computational approaches to describe complex mechanisms by focusing on essential features and disregarding irrelevant details, a process known as coarse-graining. We begin by introducing the concept of phase field, which we apply in Chapter 2, and is categorized within the realm of coarse-grained approaches. Later, we briefly explore the concepts of Markovian and non-Markovian dynamics, highlighting how memory becomes necessary to describe the relationship between fast and slow degrees of freedom in a coarse-grained system which we use in Chapter 3.

1.3.1 The phase field approach

The phase field theories are continuum theories that describe a system by using a set of order parameters, denoted as $\varphi(\vec{r}, t)$. These parameters represent the local state/phase of the system and are normalized between 0 and 1. While the term "phase" typically refers to atomic order in physical systems, in this thesis, we define a phase as a genomic compartment with a distinct functionality pertaining to transcriptional activation or silencing.

Once we have characterized the system's state, our interest lies in understanding how these states change over time. The evolution equations of the order parameter can be derived through either explicit coarse-graining of microscopic models (the bottom-up approach, discussed in Chapter 2) or through top-down considerations (applied in Chapter 3), where an expansion in powers of φ and its spatial and temporal derivatives is included in the evolution equation. In both cases, the interaction between different system compartments may lead the system to demix and build phases of different compositions to lower the system's free energy, $F \equiv F[\varphi, \nabla\varphi]$. In this regard, it has been proven that phase field theories are powerful tools for visualizing the development of microstructures without the need to track individual interfaces [41, 42]. Specifically, in Chapter 2, we will see how the interaction between two antagonistic epigenetic modifiers regulates the interface between different genomic regions, and using phase field principles, we relate these regulations to the age biological age differences.

1.3.2 On the memory effect and generalized Langevin equation

Biological systems are complex structures that contain interactions between several components influenced by various factors, many of which are not known. When studying such a complex many-body system, it is often a pragmatic choice to describe a process of interest in terms of the evolution of a small set of relevant observable (usually called reaction coordinates), which capture the main features of the process [43]. However, the relation between the reaction coordinates and the microscopic variables is rarely trivial. The investigation of such relations has been extensively discussed for many decades. It started with Langevin's work [44] and was followed by resourceful theoretical studies. Specifically, the development of projection operator techniques and the Mori-Zwanzig formalism as an extension of Langevin's work has revealed that the equations governing the dynamics of coarse-grained variables and reaction coordinates must inherently display non-locality in time [45, 46]. This means that one typically deals with the coupling of non-completely time-separated variables, namely the "fast" microscopic degrees of freedom and the "slow" reaction coordinate variables. The consequence of this is known as non-Markovian or memory effects. Briefly, in a Markovian process, the future behavior of the system is determined solely by its present state, and any information from the past is irrelevant. In contrast, in a non-Markovian process, the future behavior of the system depends not only on its present state but also on its past history.

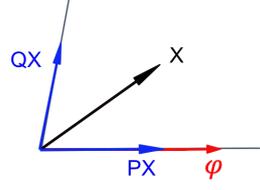


Figure 1.5: Schematic visualization of the projectors (blue arrows) in the space of the observables.

Here, following [47], we introduce the generalized Langevin equation and how memory emerges as the consequence of the projection technique. To begin with, consider the time evolution of a general phase-space observable φ_t as,

$$\frac{d\varphi_t}{dt} = i\mathcal{L}\varphi_t. \quad (1.1)$$

where $i\mathcal{L}$ is the Liouvillian operator. This equation has a formal solution as $\varphi_t = e^{i\mathcal{L}t}\varphi_0$. In the Mori-Zwanzig formalism, a common practice involves defining a stationary projection operator, which is utilized to partition the dynamics into a parallel (slow) component and an orthogonal (fast) component [48]. Assuming we have a projection operator $P^2 = P$, and $Q = 1 - P$, as it is demonstrated in Fig. 1.5, and using the inner product property, we define the operation of the projection operator on any function of phase space X as $PX = \langle \varphi^* X \rangle \langle |\varphi|^2 \rangle^{-1} \varphi$.

Following splitting the dynamic into the projected part P and orthogonal part Q , one can rewrite Eq.1.1 as,

$$\begin{aligned} \frac{d\varphi_t}{dt} &= e^{i\mathcal{L}t}[P + Q]i\mathcal{L}\varphi_0, \\ \frac{d\varphi_t}{dt} &= e^{i\mathcal{L}t}Pi\varphi_0 + \int_0^t d\tau e^{i\mathcal{L}\tau}Pi\mathcal{L}Qe^{i\mathcal{L}Q(t-\tau)}i\mathcal{L}\varphi_0 + Qe^{i\mathcal{L}Qt}i\mathcal{L}\varphi_0. \end{aligned} \quad (1.2)$$

By specifying the definition of projection operator and renaming variables here, one can simplify the Eq.1.2 as the famous generalized Langevin equation (GLE),

$$\frac{d\varphi_t}{dt} = \omega\varphi_t + \int_0^t d\tau K(t-\tau)\varphi_\tau + F_t. \quad (1.3)$$

where the first term accounts for the parallel contribution that is local in time. The second term is, however, non-local in time and can be called a friction term with the memory kernel K . The last term F_t refers to the fast variable is the orthogonal contribution to φ .

In a more complex version of the Eq.1.3, the memory kernel varies over time, denoted as $K(t, \tau)$, and the corresponding dynamic named non-stationary generalized Langevin. The nonstationarity, often referred to as stretched relaxation, is observed in materials such as quenching spin glasses and polymer melts with viscoelastic properties [49, 50]. However, the

origin of this peculiar memory behavior remains elusive at the microscopic level [51]. In order to apply the Mori-Zwanzig formalism to the nonstationary GLE, it is necessary to take into account a time-dependent projection operator P_t . This operator should satisfy the condition $P_t P_t' = P_t$, and all the preceding steps need to be rewritten accordingly. P_t can also be understood as an operator that projects onto the fixed vector φ , and the orientation of the projection changes over time [48].

In Chapter 3, where we explore the relationship between DNA structure and changes in epigenetic modifications during liver regeneration, despite the numerous variables involved, we describe our system effectively by projecting the fast epigenetic degrees of freedom onto the reaction coordinate variable, DNA structure. This is achieved using the Mori-Zwanzig formalism. To further understand this system, we study the perturbation caused by adding poison to the liver, which activates the regeneration process. We also consider changes in the amount of poison (injury) in the liver during the regeneration process equivalent to the changes in the projection operator with time, indicating a non-stationary memory kernel. Afterward, we explain in detail how to use perturbation techniques to solve such a dynamic and measure the corresponding statistical characteristics like correlation functions.

1.4 Research goal and overview of the thesis

Understanding the mechanistic basis of temporal changes in DNA methylation and their interplay with other epigenetic modifications, such as alteration in chromatin structure, is essential for comprehending the processes responsible for development, regeneration, and aging, as well as diseases occurring due to the irregular regulation of these processes. Sequencing technologies provide detailed and complex descriptions of molecular processes governing cellular fate across various stages of life. However, the descriptive nature of these insights persists until complemented by methods capable of identifying collective degrees of freedom that arise from microscopic interactions. In this thesis, using a combination of novel sequencing technologies and statistical physics, we studied collective behavior underlying cellular regulation during aging and in the regeneration of the liver after injuries.

Specifically, in Chapter 2, we investigate the relationship between changes in DNA methylation over time and biological age, showing that biological age can be predicted based on collective processes at the boundaries between genomic regions with different densities of cytosine-guanine pairs. We also develop a biophysical model to predict the evolution of DNA methylation patterns during aging using an interactive two-component phase field model. While our study is focused on bulk sequencing experiments, we explore its potential application to single-cell datasets to determine if biological aging can be inferred from the behavior of individual cells.

In Chapter 3, we shift our focus to rapid epigenetic alterations that occur during liver regeneration. We will expound on how liver cells retain a memory of injury repair occurrences

through a significant modification in the interplay between epigenetic marks and DNA topology.

In Chapter 4, we summarize our findings, and we propose potential future research inspired by the insights gained.

Chapter 2

Collective epigenetic processes during ageing

2.1 Introduction

Ageing is a complex biological process characterized by a buildup of deleterious molecular changes that result in a gradual decline of the function of various organs and systems and ultimately lead to death. There is no commonly accepted singular mechanism underlying the ageing process [53], and different genetic and non-genetic factors can affect ageing. Furthermore, the ageing process is not solely determined by the amount of time that has passed. While chronological age measures the time that has elapsed, it is not the most accurate indicator of the ageing process. In fact, biological age may provide a better representation of true ageing as it takes into account individual differences in the ageing process. So far, nine tentative cellular and molecular hallmarks of ageing have been found that act together as determinants of age-related phenotypes; these include genomic instability, telomere attrition, epigenetic alterations, loss of proteostasis, deregulated nutrient-sensing, mitochondrial dysfunction, cellular senescence, stem cell exhaustion, and altered intercellular communication (see [52] for a detailed description of each factor). However, a notable factor that stands out among others is the presence of epigenetic changes, as these changes are strongly linked to one's biological age [54–56], see Fig.2.1. In this chapter, we will pursue understanding biological ageing by focusing on epigenetic modifications during ageing.

2.1.1 Epigenetic processes in ageing

Despite longstanding interest in investigating the aging process and epigenetics separately, the field of aging epigenetics has only recently experienced a notable upswing because of advancements in sequencing technologies [24]. During ageing, there is a continuous change in epigenetic modifications of the DNA and chromatin structure, which gives rise to multiple age-related diseases [57]. This ranges from changes in DNA methylation and histone modi-

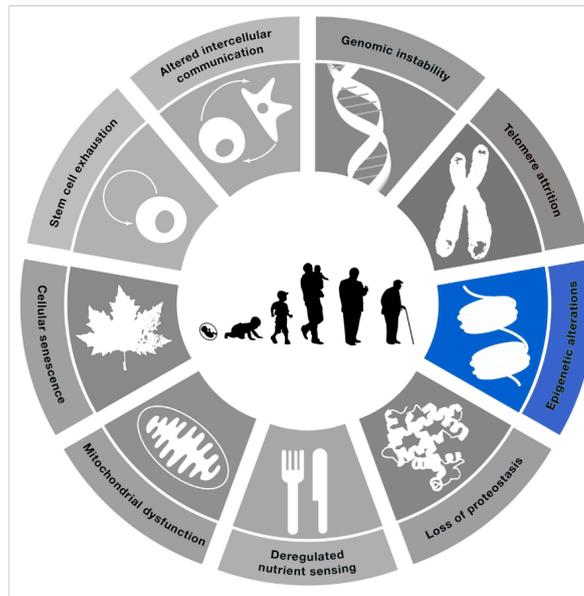


Figure 2.1: The scheme enumerates the hallmarks of ageing, adapted from [52](with licence CC-BY-ND).

fications to modes of modulating environmental factors ¹ like polycomb complex group that affect epigenetic processes, as well as cross-talk between different epigenetic modifiers, which we are going to discuss in detail below:

DNA methylation in ageing

In 1973, Vanyushin et al. [58] were the pioneers in documenting age-related global variations in 5-methylcytosine (5mC) in rats. Since then, an extensive body of literature has unveiled genome-wide and systematic DNA methylation alterations associated with aging across various species. These changes can either manifest independently of tissue or be confined to specific tissues or cell types. [54–56, 59]. Moreover, several studies revealed that identical twins, which arise from the same single cell and share almost all of their genetic variants, are not entirely similar and experience phenotypic discordances for many traits from birth weight to age-related diseases [23, 60]. These studies did not find any clear pattern of DNAm divergence, likely due to the study design, small sample size, and limited genomic coverage. The loci that showed divergence within one twin pair did not overlap with the loci that showed divergence in another twin pair. Therefore, the authors coined the term "epigenetic drift" to describe the random nature of age-related DNAm changes within twin pairs. This may be due to a global decrease in methylation comes with successive cell divisions [23]. In general, changes in DNA methylation contain either loss of DNAm (hypomethylation) at sites that are normally methylated or DNAm gains (hypermethylation) at sites that are usually unmethylated [24]. Global loss of methylation or hypomethylation of ageing cells and

¹These are the modifications that do not change the genetic information.

tissues has been documented in numerous studies and plays a major role in ageing [61]. Different mechanisms, such as genomic instability and a decline in DNMT1, may be related to a decrease in methylation status during ageing; however, the main cause of this phenomena is not known. DNA hypomethylation is also observed in many chronic age-related diseases, such as cancer, atherosclerosis, alzheimer, autoimmunity, and macular degeneration [62, 63]. Besides genomic hypomethylation, regional DNA hypermethylation is also important for biological ageing. Age-related hypermethylation has been estimated to be involved in colon cancer and other tissues such as the breast, kidney, and lung. [24, 64]. One of the important drivers of changes in DNAm during ageing is mutation in DNA methyltransferase 3a (DNMT3a), which has been associated with clonal hematopoiesis of indeterminate potential (CHIP) in ageing individuals. Somatic DNMT3a mutations arise in hematopoietic stem cells (HSCs) many years before the potential cancer development (malignancies). While common in hematologic malignances and CH, DNMT3a mutations are essentially absent from ageing skin and esophagus as well as associated malignancies; instead, in TP53 (a tumor suppressor protein) and NOTCH (a signaling pathway important for cell-cell communication), mutations dominate, indicating context specificity of the advantage conferred by mutations [65]. In principle, DNMT3a can harbor a wide range of mutations at varying frequencies with distinct predicted consequences to enzymatic function, such as R635G and R736H at the position of the DNA substrate and R771Q at the tetramer interface on DNMT3a, to name a few. Considering different CpG densities along the DNA, some of these mutations enhance, and some diminish, the activity of DNMT3a compared to the wild type (WT), which then leads to diverse levels of methylations [66]. Although DNMT3a mutations are one of the common drivers of age-related clonal hematopoiesis, in which specific HSC clones dominate blood, because these mutations are tissue-biased, it's impossible to consider them as a universal age-driven factor in multiple tissues. Therefore, One needs to be extra cautious and consider various other factors while studying ageing models.

Histone modifications in ageing

Chromatin status undergoes dynamic changes during ageing process, including variations in histone levels, exchange of histone variants, modifications of histones, and shifts in nucleosome occupancy. Comparative chromatin analysis of young and old cells through ChIP-seq analysis ¹ revealed the ageing-associated changes in histone marks and the alteration in their interplay with DNA methylation [67, 68]. Being aware of the different possible histone modifications we introduced in Chapter 1, we will narrow our focus on changes in H3K4me3 and H3K27me3 histone modifications as two important players in ageing process. Studies on old hematopoietic stem cells exhibited a significant increase in the number of H3k4me3 peaks from ChIP-seq analysis, many of which were considerably broader with age; more than half of all H3K4me3 peaks expanded, and only a few of them shrank, which suggests

¹To comprehend the functioning of this analysis, refer to chapter 1.

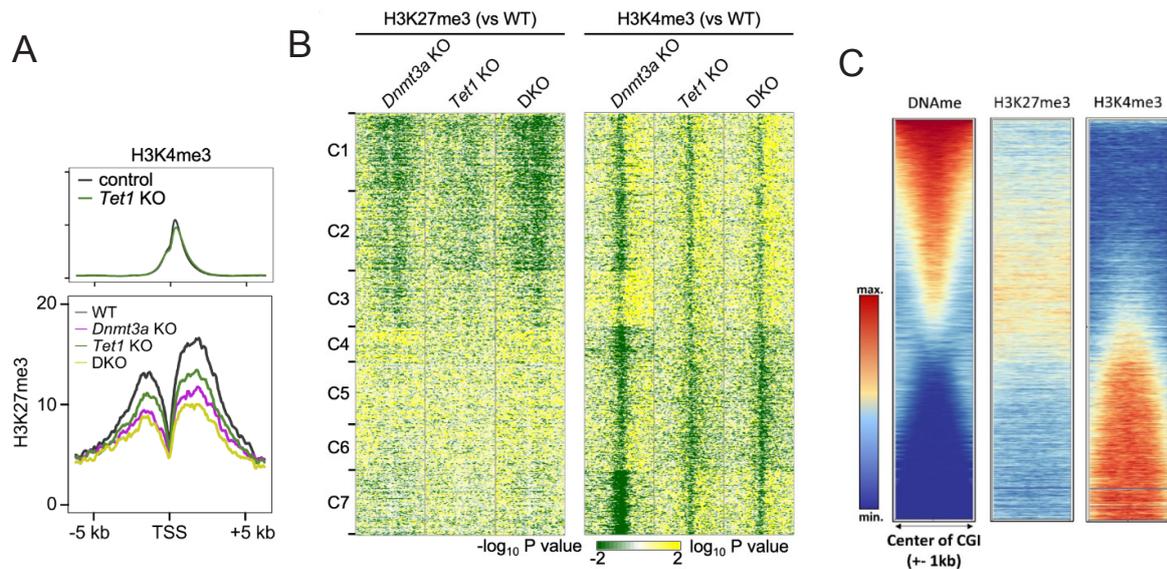


Figure 2.2: Relation of DNA Methylation and Chromatin Marks. (A) Average density profiles for H3K4me3 and H3K27me3 around TSSs. (B) Heatmaps showing the signal difference of H3K27me3 and H3K4me3 between KO and control cells surrounding TSSs analyzed in seven clusters based on their different binding properties. Green, decreased; yellow, increased signal intensity, compared to WT control. (C) Aggregating the density of DNA methylation along with the density of other chromatin marks within annotated CpG islands (CGIs). Figures A, B are adapted from [71], and C is adapted from [72].

that some previously unexpressed transcripts are activated [67]. It has also been found that there is a strong inverse correlation between DNA methylation and H3K4me3 levels during ageing [69, 70].

H3k27me3 is also likely to be actively modified during ageing. H3K27me3 designate transcriptional silencing, which is produced by polycomb repressive complex-2 (PRC2) and is maintained by PRC1. However, the results of studies on several models, including worms, flies, and fish, showed that the link between H3K27me3 levels and ageing is complex, and it is possible that depending on the specific loci and cells, different H3K27me3 regulators through different catalytic activities of PRCs subunits influence lifespan. Thus, ageing may be associated with both an increase and decrease in H3K27me3 [68]. In the same study by Sun et al. [67], it has been shown that although H3K27me3 shows similar peak counts with age, at the level of transcription, the genes that are enriched for the category of membrane protein lose H3K27me3, while the genes that are enriched for glycoprotein and cell adhesion gain H3K27me3.

Cross-talk between DNA methylation and histone modifications

It has been found that the complementary and competitive behavior in the binding of DNA methylation enzymes and methylation modifiers on histone modifications have a mutual ef-

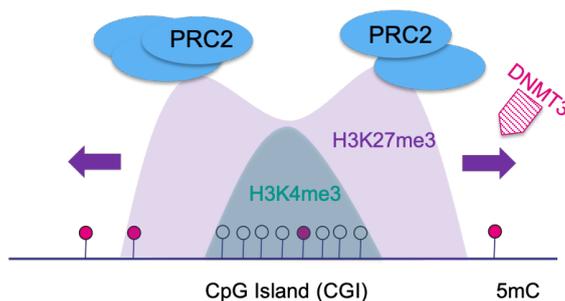


Figure 2.3: Schematic describes the interplay between different epigenetic modifiers at bivalent CpG islands (CGI). H3K4me3 affects the island's center, while H3k27me3 affects the shores, and both repress the methyltransferase (DNMT3). PRC2, in interaction with H3K27me3, regulates the shores.

fect on each other. Gu et al. studied the cross-talk between the binding of DNMT3a/TET1 enzymes and histone methylation modifiers in and outside of the bivalent regions¹ in mouse embryonic stem cell (ESCs) [71]. They found that the deletion of Tet1 increases DNMT3a1 binding capacity at and around regions with wild-type TET1 binding. However, deletion of Dnmt3a has a minor effect on TET1 binding on chromatin, indicating that TET1 could serve as an anchor protein to establish boundaries for DNMT3a binding in order to limit DNA methylation and to establish boundaries. Nevertheless, genes encoding DNA methyltransferases (Dnmts) as a group decrease during ageing. Concomitantly, genes encoding Tet1 and Tet3 DNA demethylation proteins are also decreased. Consistent with the reduction of Tet gene expression, 5-hydroxymethyl cytosine (5hmC) decreases with age.

It has also been shown that DNMT3a binding negatively correlates with the extent of H3K4me3 levels. However, loss of TET1 has little effect on the H3K4me3 levels. Loss of DNMT3a and/or TET1 also leads to diminished H3K27me3 levels, especially at the regions with intermediary levels of methylation and to a greater extent in Dnmt3a KO, see Fig.2.2. The interplay between DNMT3a and TET1 may also poise the expression of bivalent genes largely via PRC2-mediated H3K27me3 modification, which is critical for the repression of developmental genes. In this concept and as explained above, loss of DNMT3a diminishes H3K27me3 but extends H3K4me3 [67, 71], and that TET1 is also required for chromatin binding of EZH2 in mouse ESCs which is a core component of PCR2 and produce H3K27me3 [73], Fig. 2.3. Following the reduction of TET, PCR2 will also be depleted, which may ease the local increase of DNMT3A binding inside CpG islands (CGIs) [74]. However, the PRC2 modification is very slow in comparison to the histone and DNA modifiers [75]. We hypothesize that the shores at bivalent CpG islands are regulated by the slow depletion of the polycomb complex, which is a potential cue for methylation ageing. We will explain this concept more comprehensively later in Sec.2.2.3.

¹The regions that contain both active and repressive modifiers, see Sec. 1.1.2 for more detail.

2.1.2 DNA methylation ageing clocks

In the previous section, we introduced DNA methylation as a promising biomarker of ageing. In accord with DNA methylation being a dynamic modification, here we want to discuss whether it is possible to predict the progression of ageing based on a sequencing experiment. Horvath [76] was one of the pioneers that tackled this question using multivariate machine learning models to assay genome-scale temporal changes of DNA methylation marks and predict biological age. Using machine learning techniques, he identified collections of individual CpGs whose aggregate methylation status is highly correlated to chronological age in humans, and thus, they can be used as age estimators. Shortly after that, similar DNA methylation age estimators have been reported for mice [77,78] and extended to many other mammalian species [79–81]. Because of the high accuracy of these models, DNA methylation age estimators are often called as *methylation ageing clocks* [56]. We do not enter into the technical details of different machine learning age-estimators as it is beyond the scope of this thesis, but we discuss them here only briefly. Methylation ageing clocks are large linear machine learning models fitted to sequencing data,

$$Age \sim \beta_1 m_1 + \beta_2 m_2 + \dots + \beta_n m_n. \quad (2.1)$$

where m_i is the DNA methylation state at position i along the genome, and each site has its own associated weight β_i , which indicates the amount that age changes in response to the changes in methylation values. Using regression analysis, such as ElasticNet, the associated weights can be learned, and it can be understood which sites are more relevant for age prediction. To this end, one needs to optimize the following cost function through the preferred regression algorithms,

$$\hat{\beta} = ||Age - \vec{m} \cdot \vec{\beta}||^2 + \lambda_1 ||\vec{\beta}|| + \lambda_2 ||\vec{\beta}||^2. \quad (2.2)$$

If $\lambda_2 = 0$ and $\lambda_1 = \lambda$, it is Ridge regression, and if $\lambda_1 = 0$ and $\lambda_2 = \lambda$, it is LASSO [82]. Regularisation selects features by decreasing the weights of least informative CpGs with age to 0. That is in spite of the fact that this model, in principle, contains all CpG sites, which can be approximately 28 million sites throughout the mammalian genome; in the end, a few hundred highly informative CpGs are sufficient for age prediction. These sites are named clock sites. A graphical explanation of how these methylation clocks work can also be found in Fig. 2.4.

It is also possible to make universal mammalian clocks using a single model similar to Eq. 2.2, that demonstrate accurate predictions of chronological age across different species [83]. To this end, various tissues have been examined, encompassing blood, liver, kidney, muscle, skin, and the brain, including multiple regions such as cortex, hippocampus, and hypothalamus; see Fig. 2.5.

Note that for most of the reported methylation ageing clocks so far, the utilized methylation

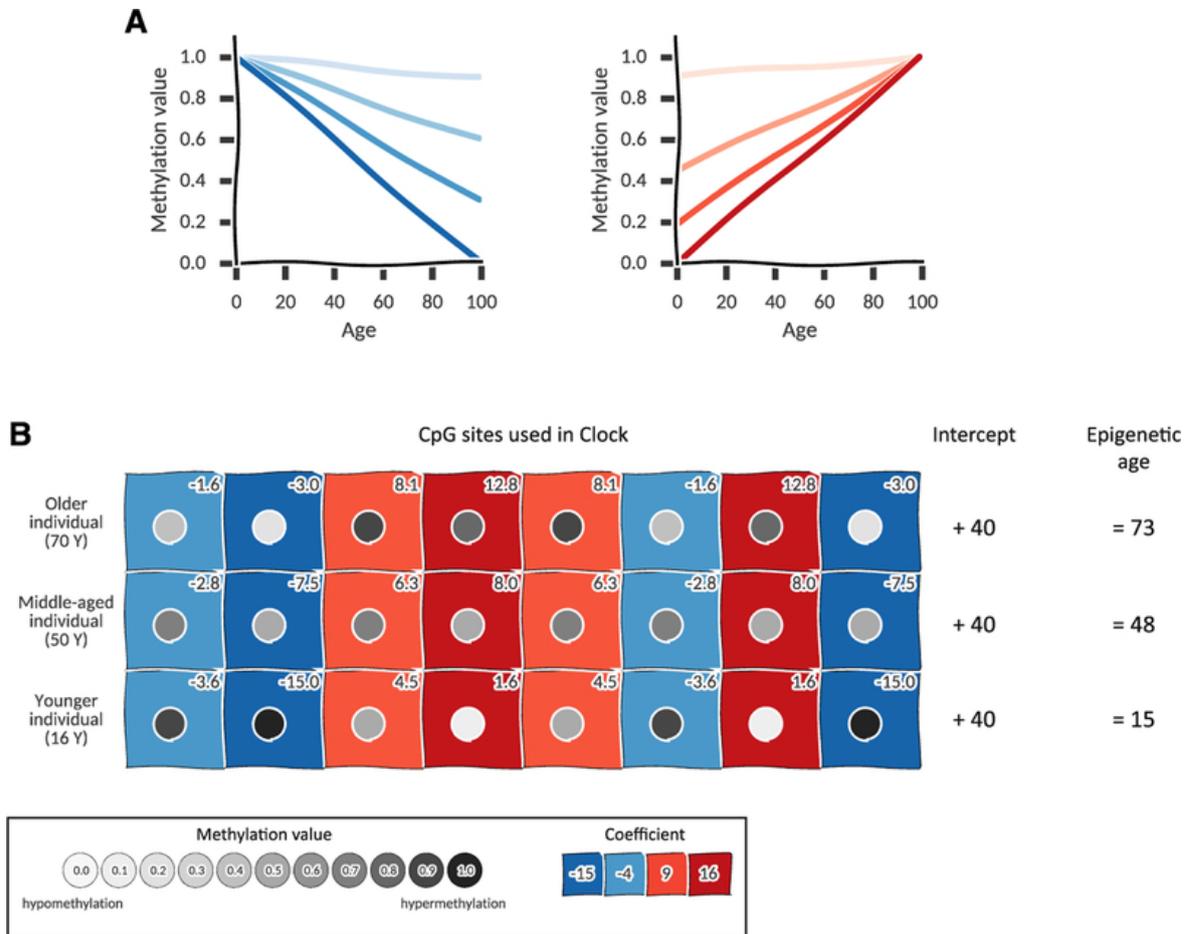


Figure 2.4: Illustration of DNA methylation ageing clocks: (A) Depicts the age-related methylation trajectory of eight individual CpGs, divided into four with negative coefficients (blue) and four with positive coefficients (red). The colors reflect the rate of change with age, with darker shades indicating faster rates and stronger weights in the epigenetic clock. (B) Displays the methylation values of these eight clock CpGs (columns) in three individuals of different chronological ages (rows). The shading of the filled circles represents the fractional methylation (ranging from 0.0 to 1.0). The box color signifies the coefficient of change over age (slope of the line, top). In each colored box, the numerical product of the methylation value and its corresponding coefficient is presented. The summation of these products, combined with the intercept learned during the construction of the clock, results in an epigenetic age approximation that corresponds closely with chronological age. Figures are adapted from [82].

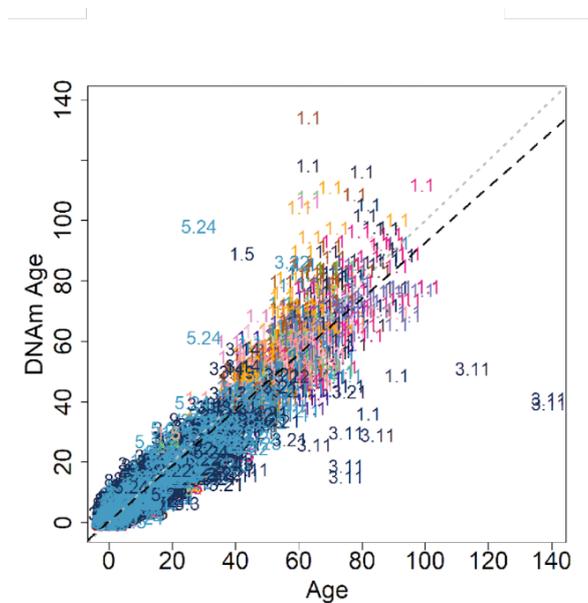


Figure 2.5: Universal clock: Every point on the graph corresponds to a tissue sample identified by the mammalian species index. For example, 1 denoting primates, 2 indicating elephants, 3 representing cetaceans, and so on. Figure is adapted from [83].

values at each site in Eq. 2.1 are usually acquired from bulk sequencing experiments, as explained earlier in the first chapter. In fact, having a higher coverage in an experiment would increase the accuracy of the predictions. However, having model’s inputs from single-cell experiments would be challenging, as the intrinsically sparse and binarized methylation profiles of individual cells precluded the assessment of ageing in single-cell data [84]. In section 2.2.4, We will discuss the analysis of single-cell ageing in more detail. In particular, we will discuss the results of our collaboration with Ferdinand von Meyenn’s group at ETH, Wolf Reik’s group (Babraham Institute and AltosLabs), and Oliver Stegles division (DKFZ and EMBL Heidelberg), where we investigated the long-standing question as to whether DNA methylation age is also measurable in single cells and whether these measurements are biologically relevant.

2.1.3 Research questions and overview of this chapter

So far, we have discussed how epigenetic modifications can be a valuable indicator of the ageing process. We also learned that one needs to consider the interplay between different modifiers to clearly understand the different alterations these modifications may face during ageing. Additionally, the alterations of these modifications are tightly correlated to biological age, and machine learning models can predict biological age by tracking these modifications. However, machine learning ageing models are, in general, purely descriptive and have lots of parameters. The mechanistic basis underlying these methylation clocks is indeed poorly understood despite the fact they are quite accurate and are able to predict biological age

and time of death with an uncertainty of three to four years [76,85]. Besides, the field has advanced beyond questioning the feasibility of developing DNA methylation aging clocks to exploring the reasons behind the potential of DNA methylation contribution to biological ageing. [56]. Probably one of the most interesting prospects of understanding the ageing mechanism is that it can be useful for identifying or validating anti-ageing interventions. To this end, we designate this chapter to answer the question: *What is the mechanistic basis underlying DNA methylation during ageing?* Using a combination of tools from statistical physics, we show that biological age can be predicted as a result of collective processes in the boundaries between genomic regions with different densities of CpGs. We first explicitly study epigenetic ageing using bulk sequencing experiments, and later in this chapter, we will study single-cell ageing.

2.2 Results

2.2.1 Sequencing experiment

We started our analysis with bulk sequencing experiments, which have higher coverage and lower technical noise compared to single-cell sequencing experiments. We utilize a DNA methylation sequencing data set of mouse blood cells taken at various ages, as previously published by Petkovich et al. (2017). In this experiment, in order to examine the overall changes in the DNA methylome on a global scale, the authors used a modified form of reduced representation bisulfite sequencing (RRBS)¹ to analyze DNA methylation from the whole blood of 141 male C57BL/6 mice. These mice ranged in age from 3 to 35 months and were divided into 16 age groups. Within each individual sample, they detected between 3.9 and 13.0 million CpG sites, with over 1.9 million sites being present in all samples. As a result, their RRBS method effectively covered a significant portion of the DNA methylome in the blood of mice, allowing us to examine how it changes with age. In Fig. 2.6, we present the results obtained from conventional bulk bisulfite sequencing data. For each sample/replicate, we have the start and end position on the chromosome, percentage of methylation, age and other corresponding information. Usually, instead of methylation, we get information about the number of number of methylated and non-methylated CpGs. Note that due to technical errors in the sequencing experiment, we may get misinformation about certain cytosines from the corresponding table. Thus, it is typically important to conduct thorough quality control before proceeding with further analysis. Regarding quality control, it is advisable to retain samples that exhibit substantial coverage, efficient mapping, and a low number of

¹A bisulfite-based method that focuses on CG-rich regions of the genome, in order to decrease the necessary sequencing while encompassing most promoters and other important genomic areas.

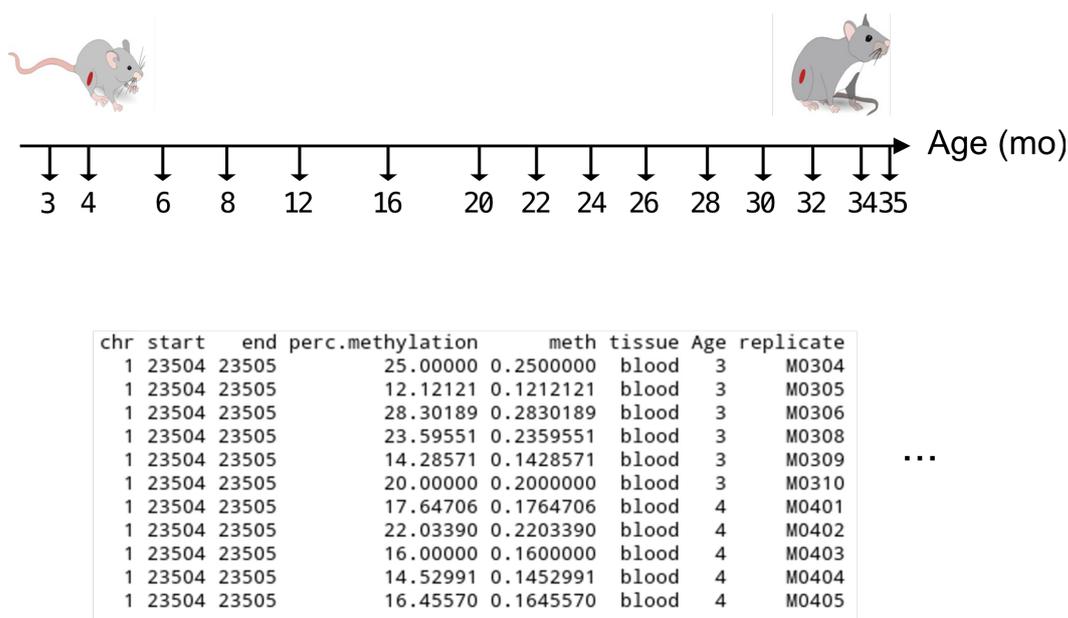


Figure 2.6: Illustration of the data from Petkovich et al., we used for this study. In the table, we show typical information one gets from sequencing experiments.

CHH methylated sites (where H is any base except G).¹ In Fig.2.7, we present a subset of data that displays the overall measurement of DNA methylation across a wide genomic area and at various ages. Methylation marks, however, are affected by different factors in specific regions, like CpG density, transcription, and histone modifications. Consequently, the representation of methylation densities may not offer meaningful insights because it averages across distinctly functional genomic regions. Although this map (Fig. 2.7) would be sufficient for machine learning models, we here do not want to treat each CpG individually and track their changes over time like in machine learning models. Instead, we here hypothesize that *DNA methylation ageing is a result of a collective process*, and to understand the DNA methylation ageing process, one needs to study changes in the interaction between different CpGs. To test this hypothesis, we need to quantify the collective dynamics of DNA methylation states through statistical measurements like two-point correlation functions, which we elaborate on in detail in the next section.

2.2.2 DNA methylation ageing as a collective epigenetic process

As we discussed earlier, and in order to examine our hypothesis that DNA methylation ageing is a result of a collective epigenetic process, we calculated connected two-point correlation functions. These functions help us observe variations in the relative arrangement of DNA

¹We didn't apply any explicit quality control for the data set from Petkovich et al., as it was trimmed before publication. However, for the analysis of ageing single-cell data set, which we discuss later in this chapter, and for the liver regeneration data set in the chapter3, we developed computational pipelines, which entail quality control, normalization, overlapping of reads with genomic annotations, and propagation of uncertainties as well as the computation of statistical summaries.

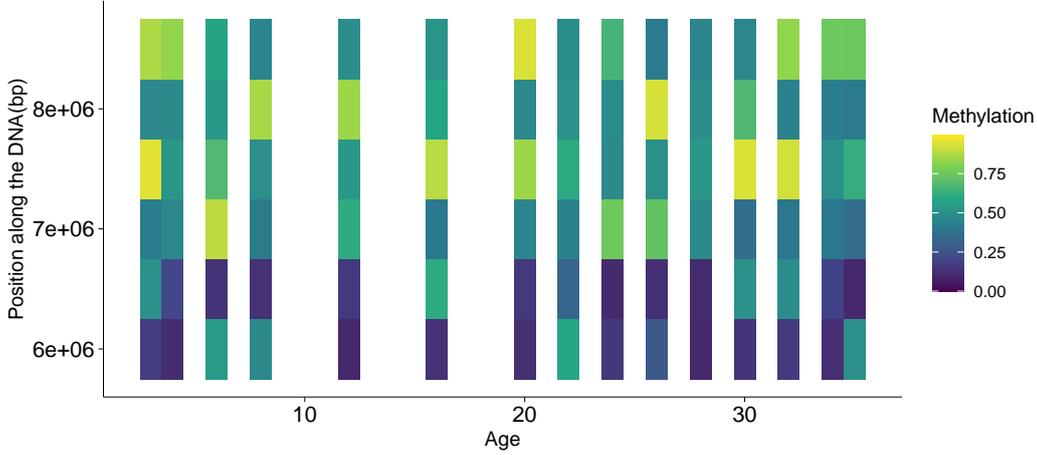


Figure 2.7: Methylation density during a 32-month time-course sequencing experiment for chromosome 1 and on a broad genomic region. For visualization, the genome is re-scaled into bins with 100,000 base pairs.

methylation marks during ageing. Connected correlation functions are defined as,

$$C(j) = \langle m_i m_{i+j} \rangle - \langle m_i \rangle \langle m_{i+j} \rangle, \quad (2.3)$$

where i and $i + j$ are the genomic positions of CpG sites, and m_i and m_{i+j} are the corresponding DNA methylation values. The averages are performed over all pairs of CpGs that are a distance of j base pairs apart and over all samples. We find a clear, genome-wide ageing signal by calculating this two-point correlation function over the Petkovich data set we introduced above; see Fig. 2.8. However, we need to quantify these systematic changes in two-point correlations over time. In order to achieve this, we focus on a measure that indicates only uncorrelated changes during ageing. We consider multiplicative changes as the primary indicators of ageing; see Appendix A for a detailed consideration.

Therefore, the ratios of correlations at a given distance, d , and consecutive time points as a measure to quantify the changes in correlation function becomes,

$$\frac{C(d, t')}{C(d, t_0)} = A(t', t_0). \quad (2.4)$$

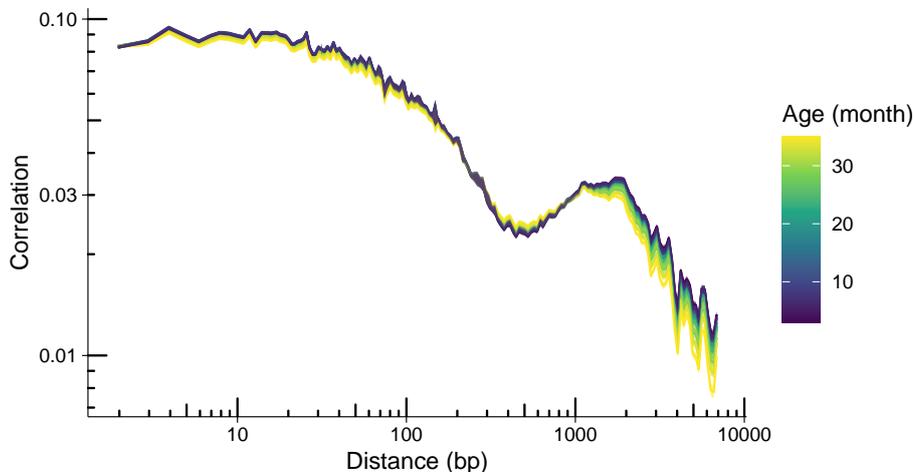


Figure 2.8: Connected two-point correlation functions across the genome between DNA methylation marks at different ages for mouse blood cells, which shows a clear ageing signal (Data from Petkovich et.al).

where $A(t', t_0)$ is a function of the new age and the initial one. By computing these correlation ratios with respect to the first age sampled in the experiment, Petkovich et. al, the systematic genome-wide ageing signal we observed in Fig.2.8 becomes more detectable in Fig. 2.9. Specifically, we observe a loss of correlations for distances smaller than roughly 400 bp and larger than 1000 bp while correlations increase over time around a characteristic distance of 500 bp.

An important inquiry is whether this behavior is prevalent across the genome or if it is limited to specific genomic regions. This query may seem peculiar from a physics standpoint. This is because when calculating correlation functions or other statistical measurements in a spin system, the direction along the spin domain is not a concern, and it is typically assumed that the system exhibits translational symmetry. However, this assumption no longer holds for genomic regions. This is due to the heterogeneous nature of the genome and the presence of distinct functionally relevant regions, which are mapped in DNA annotations. DNA annotation serve to identify the location and function of genes and other regions of the genome that are relevant for understanding the biological context, such as gene bodies, promoters (the regions of the genome where transcription of the gene is initiated), TSSs (transcription start sites), and CPG islands (CGIs, regions that are more dense in terms of having CpGs). Hence, to reveal the genomic origin of these ageing patterns from multiplicative changes in correlations, we analyze the correlation ratio separately for different genomic annotations classified by the density of CpGs. As mentioned earlier, the genome is not uniform, and certain areas called CpG islands have a higher concentration of CpGs while others are less

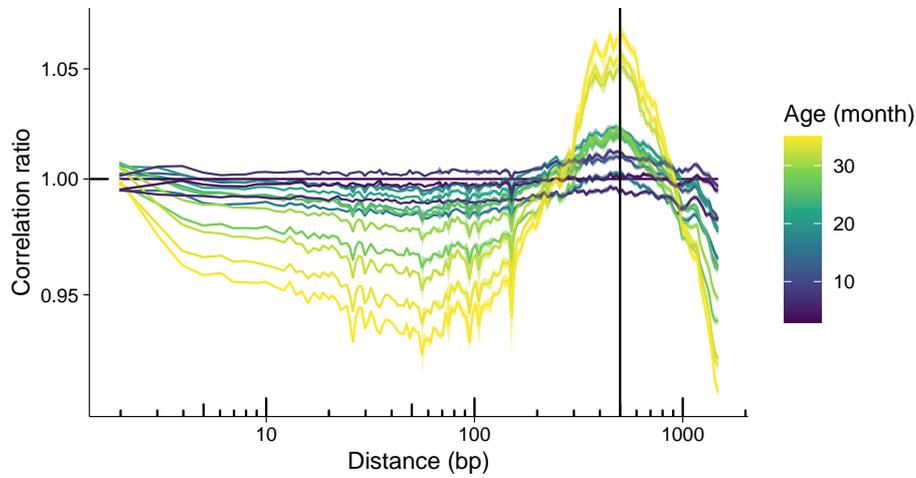


Figure 2.9: Characteristic behavior of genome-wide multiplicative changes in connected two-point correlation functions as quantified in correlation ratios, Eq.2.4 (Data from Petkovich, et.al). There is a striking increase in the correlations around 500 bps.

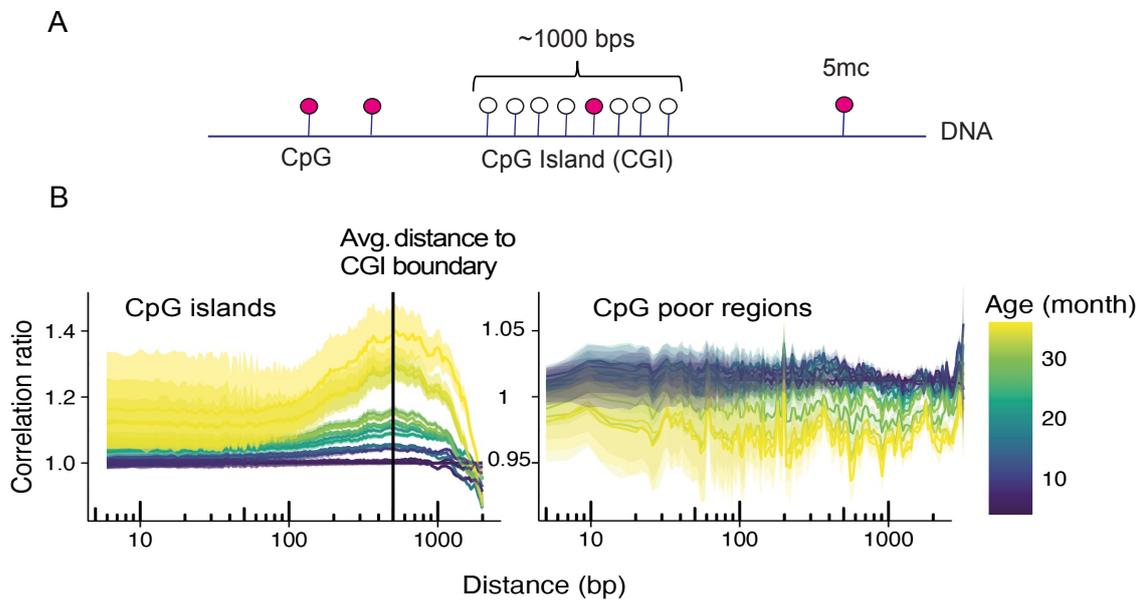


Figure 2.10: (A) CpG islands have a higher concentration of CpGs along the genome, and the average length of this region is 1000 bps. (B) The increase in correlations observed at a 500bp distance is unique to regions characterized by high CpG density, known as CpG islands (CGIs). The distance of 500bp can be understood as the average distance between sites within a CpG island and its boundary. Correlations decrease over time in regions with low CpG density (CpG-poor regions), but genomic proximity does not influence this decay.

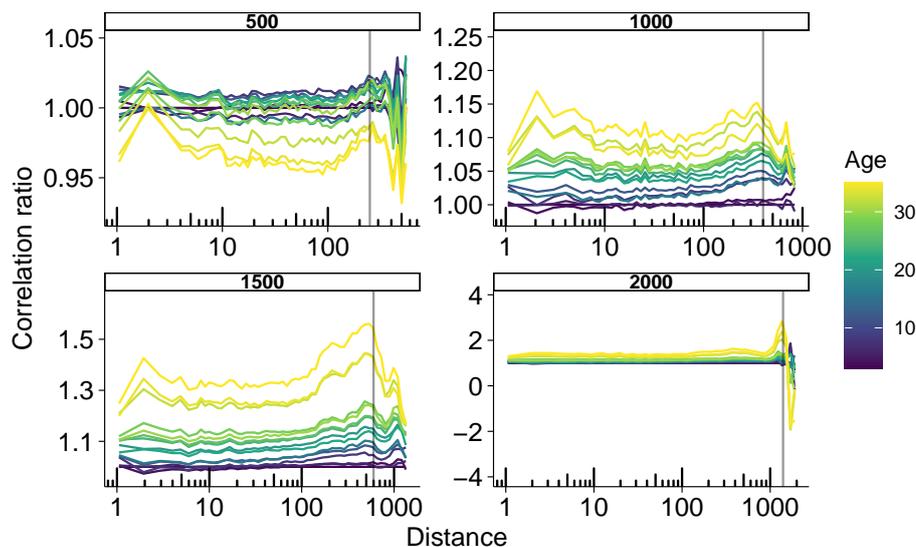


Figure 2.11: Position of the peak in multiplicative changes of correlations scales with the size of CpG islands, which indicates that an increase in correlations at a distance of 500bp is associated with DNA methylation dynamics in the boundaries of CpG islands. The position of this peak aligns with the half-length scale of each group of CpG islands with different lengths (500, 1000, 1500, 2000), as indicated by the vertical line for better observation.

dense. In the following section, we explore the connection between CpG islands and the ageing process and how it affects the striking patterns in the correlation ratios.

Role of CpG islands in ageing

CpG islands (CGIs) or CpG-rich regions have an average length of 1000 bps. These regions are primarily found in the vicinity of gene promoters, which are the regions responsible for initiating gene expression. Therefore, unlike the rest of the genome where CpGs are highly methylated, CGIs tend to be lowly methylated. By filtering the genome based on the density of CpGs, that is, splitting the genome into CpG islands and the rest of the genome (open sea), and calculating the correlation ratios for each region separately, we find that the gain of correlations at a distance of 500bp is specific to CpG rich regions where correlation increases, to a lesser degree, also for distances smaller than 500bp. In contrast, correlations in CpG poor regions decay over time unrelated to the genomic distance, Fig. 2.10. 500bp can also be interpreted as the average distance of sites inside a CpG island to the boundary is 500bp. We, therefore, hypothesized that the increase in correlations at a distance of 500bp is associated with DNA methylation dynamics in the boundaries of CpG islands (shores). To prove this hypothesis, we conducted a numerical experiment in which we calculated the correlation ratios for different groups of CGIs distinguished by their average length. As depicted in Fig.2.11, we observe that the position of the peak in multiplicative changes in correlations

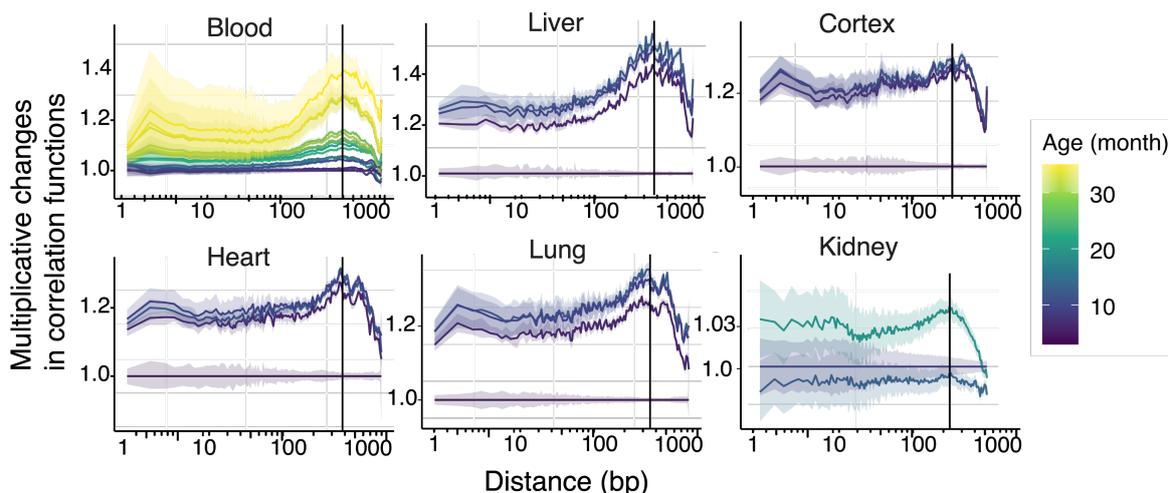


Figure 2.12: Generic changes in DNA methylation correlations across tissue. A clear trend of ageing and a characteristic gain of correlation (bump) around 500 bps is evident (indicated by the vertical lines). Data is collected from Stubbs et al. [86], except for blood from Petkovich et al.

scales linearly with the size of CpG islands, which supports our hypothesis. However, one may be cautious about these results, thinking they embedded some cell type or tissue-specific property. To investigate this issue, we apply our analysis to DNA methylation sequencing experiments from a broad range of mouse tissues published by Stubbs et al. [86]. We find that the stereotypical behavior in changes of the two-point correlation function is not specific only to blood cells with which we stated our analysis. Rather, it is a generic behavior that characterizes ageing in a broad range of tissues, see Fig.2.12. This shows that the dynamic of DNA methylation at the boundaries of CpG islands plays a role in DNA methylation ageing. We now investigate the mechanism governing these changes more precisely. As discussed earlier in this chapter, the interplay between DNA and histone modification is an essential mutual effect that one must consider when studying epigenetic alterations. To check the impact of histone modifiers on CGI methylation, we filter CGIs based on the presence of H3K4me3, which mostly can be found at the center of CGIs and is associated with gene activation. Detection of the histone modifiers like H3K4me3 is possible through a Chip seq experiment as introduced in Sec. 1.4. We also tested for the presence of H3K27me3, which is associated with gene repression but can be found at the bivalent CpG islands, as introduced before. Fig. 2.13 shows the result of our correlation ratio analysis for these different groups of CGIs. It is evident that the ageing signal which appears in the characteristic peak of correlation ratios is specific to the bivalent CGI. Combining what we learned from the general ChIP-seq analysis of histone modifiers and our genuine observations about CGI alteration during ageing, we take quantitative steps toward making a biophysical ageing model in the next section.

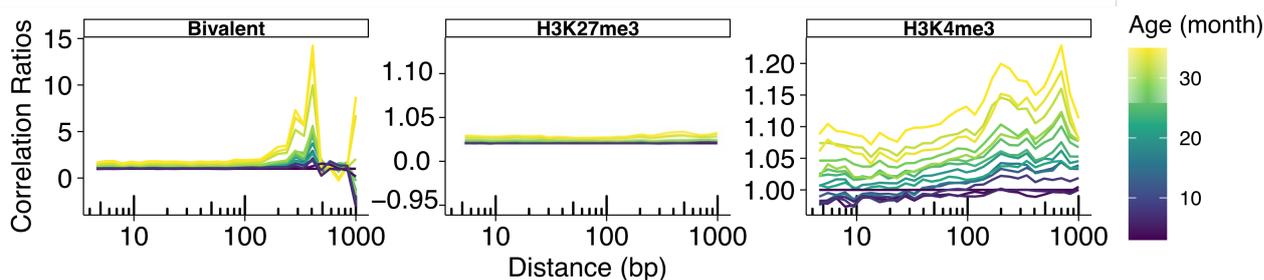


Figure 2.13: Correlation ratios for three groups of CpG islands based on the enrichment of H3k4me3 and H3k27me3. Ageing signals are specific to bivalent CGIs, where one can find both of these modifiers together.

2.2.3 Towards a biophysical mechanism of DNA methylation ageing

As we learned in the previous section, collective dynamics occurring at the boundaries of CpG islands, which encompass both active histone modifiers (H3K4me3) and repressive modifiers (H3K27me3), are linked to the characteristic alterations in two-point correlation functions of methylation marks. In section 2.1.1, we also found that active modifiers predominantly occupy CpG islands, while repressive modifiers are predominant outside of these regions. However, in the transitional border areas between CpG-rich and CpG-poor regions, there is an enrichment of both active and repressive enzymes, indicating competition for binding between the proteins depositing them in these regions. Now, we want to investigate if this potential competition would be ever tilted during ageing. Therefore, we measure changes in the length of lowly methylated regions from the Petkovich et al. data set. The reason is that lowly methylated regions are linked to CpG islands and the areas bordering these islands. To determine the size of the methylated region linked to each CpG island, we employ an algorithm in R that calculates the methylation values and the total length of the region that exhibits lower methylation levels compared to the average within that region. This algorithm encompasses the CpG islands as well as the shores defined to be the regions of 5 kbps¹ size adjacent to both sides of the CpG islands. From this, we find that the size of the lowly methylated region increases with time; see Fig.2.14, blue line. Furthermore, we measure the changes in DNA methylation status inside the CpG islands; see Fig.2.14, red line. From this figure, determining the competition's winner may not be simple, as the growth of the lowly methylated region is associated with an increase in DNA methylation levels in that region. This is an interesting observation that is, at first sight, paradoxical. However, the potential depletion of the PRC2 group, as we discussed earlier in section 2.1.1, gives a potential molecular mechanism for this outcome. It suggests that the decrease in PRC2 would lead to a reduction in H3K27me3 levels [73], possibly due to a decrease in binding affinity and ineffective recruitment of the enzyme to the DNA, that consequently modifies the shores, see Fig. 2.3. Additionally, because of the decrease in H3K27me3, DNMT3A would

¹5 kilo base pairs = 5000 base pairs

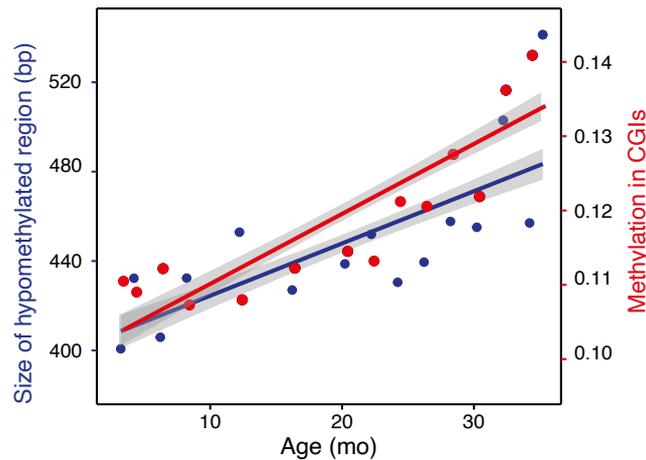


Figure 2.14: During ageing, expansion of the lowly methylated region goes along with an increase of methylation in this region.

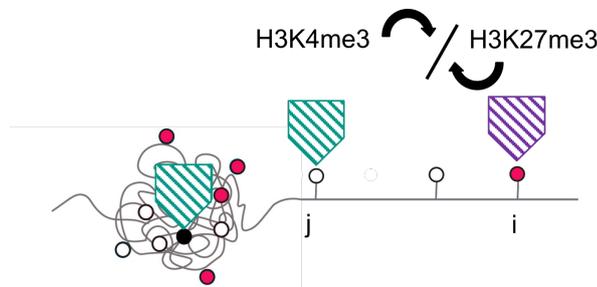


Figure 2.15: Microscopic model that accounts for the exchange dynamics and interactions between different histone modifiers. These interactions are long-ranged due to the chromatin structure in three dimensional space.

have an easier time binding to the CGI [74]. Overall, the competition between H3K4me3 and H3K27me3 during ageing may lean in favor of H3K4me3, which we are going to explore later in this section.

Before we continue with the definition of the model, it is important to recapitulate the main points discussed in this chapter. A basic biophysical model of DNA methylation ageing should fulfill three essential criteria:

- It should explain the consistent patterns observed in the relative changes of methylation marks that are found in correlation ratios.
- The model should apply to a wide range of tissues and cells without bias to any specific type.
- Additionally, the model should account for the expansion of regions with low levels of methylation, as well as the acquisition of methylation within these regions.

To continue with the definition of the model, we begin by considering a one-dimensional lattice model comprising binary state variables. Each site in the lattice corresponds to a nu-

cleosome along the DNA strand, which either carries an H3k27me3 or H3K4me3 mark. We assign indices to the lattice sites using the notation $i = 1, \dots, N$, and the overall configuration of the lattice is represented by the set of all variables, $\sigma = \{s_1, \dots, s_i, \dots, s_N\}$. In this notation, the variable $s_i = 1$ signifies that a repressive modifier, denoted as "r" or H3K27me3, occupies lattice site i , while $s_i = 0$ indicates that an active modifier, represented as "a" or H3K4me3, occupies the site. Moreover, the interplay between two modifiers at sites i and j is represented by an interaction energy J_{ij} . We use J_{ij}^{aa} as the interaction energy between two "a" modifiers separated by a distance $|i - j|$, where we specified it for a long-range interaction as $J_{ij}^{aa} = \frac{J_0^{aa}}{|i-j|^{d+\lambda}}$. Here, J_0^{aa} is a constant coefficient that sets the strength of the interaction. The value of d is 1, which indicates the dimension of the model, and considering previous studies [87, 88], λ falls within the range of $0 < \lambda < 1$, that will allow using a mean-field approximation in this 1d model, as we applied later in this section ¹. Similarly, the interactions between two "r" modifiers are denoted as J_{ij}^{rr} , and those between an "ar" pair are written as J_{ij}^{ar} . These interactions are all attractive, but the attractive interaction between the modifiers of the same kind is stronger than the attraction between the antagonistic ones, $J_0^{rr}, J_0^{aa} \gg J_0^{ar} > 0$, that satisfy the existence of the bistable islands, Fig.2.3 and [71, 89].

As we argued above, and due to the polymeric characteristics of DNA with a three-dimensional structure (Fig.2.15), we account for long-range interaction in 1d. Specifically, and in our case study, all the modifiers we are interested in are associated with large genomic regions rather than point objects, and they affect chromatin regulation in three dimensions [90]. For example, H3k27me3 is linked to regions of approximately 150 kbps in size [90, 91], while the average effect length of H3K4me3 spans from 1.5 to 12 kbps [92]. A representative microscopic model, however, can be intuitive to understand the mutual interactions between the modifiers and DNA. The gap bridging between these two pictures of the same phenomena requires a multiscale approach. A promising route toward this goal is using coarse-grained models in the spirit of the phase field method, in which one can compute quantitative aspects of the evolution of microstructure without explicit intervention. A significant advantage of this method is its ability to handle dynamics at various time scales, as the rapid fluctuations will be incorporated in a "mesoscopic" ² free energy of the system [93].

Back to the microscopic picture as the starting point, and taking into account the alterations in the interaction among these modifiers influenced by PRC2 (Section 2.1.1), the aging system experiences the following changes,

- Exchange dynamic between histone modifiers.
- The slow effect of the depletion of PRC2 on H3K27me3 causes a reduction in H3K27me3 concentration; however, the concentration of H3K4me3 remains constant.

¹The long-range interaction linked to the spatial structure of chromatin is consistent across all histone modifiers, such that they all have the same arbitrary λ that falls between 0 and 1. The exact value of λ is unimportant in this work; we refer to the previous study of 1d long range interactions in genomics [1], in which they discovered $\lambda = \frac{1}{3}$.

²Mesosopic scales are larger than the lattice spacing but much smaller than the system size.

The evolution of a given initial configuration of this system is described by the master equation as,

$$\frac{\partial P(\sigma, t)}{\partial t} = \sum_{ij} (\omega(\sigma', \sigma) P(\sigma', t) - \omega(\sigma, \sigma') P(\sigma, t)), \quad (2.5)$$

where $\omega(\sigma, \sigma')$ is the transition rate from configuration σ to σ' . To determine this rate, we assume the local thermalization following a successful exchange as a way of restricting the parameter space. Note that biological systems like the one we introduced here are usually out of equilibrium. Nevertheless, due to the distinct time scales involved in the exchange dynamics and the changes in H3K27me3 concentrations, we suggest achieving a quasi-equilibrium state at each age, that is $\omega(\sigma, \sigma') \propto e^{-\beta \delta H}$; where β is the Boltzmann factor and H is the Hamiltonian. To explicitly express the transition rate, we begin by defining the system's Hamiltonian, taking into account the aforementioned characteristics of the system as follows,

$$H(\sigma) = -\frac{1}{2} \sum_{ij} [J_{ij}^{aa}(1-s_i)(1-s_j) + J_{ij}^{rr}s_i s_j + J_{ij}^{ar}[s_i(1-s_j) + s_j(1-s_i)]] - h(\tau) \sum_i s_i. \quad (2.6)$$

Note that the first summation in the Hamiltonian includes all sites in the lattice to account for long-range interaction in 1d, as we discussed above. Later, we will show how we simplify these long-range interactions using a mean-field approach. The negative sign in the Hamiltonian indicates that interactions are attractive. The first summation can also be understood as the exchange dynamic, while the second describes the interaction between modifiers and PRC2 in the content of interacting with an external field, h , as a function of τ , the time scale of ageing in the system and is modifying adiabatically/ very slowly compared to the exchange dynamic; that is, it accounts for the separation of the time scales as $\tau = \epsilon t$. Considering the Hamiltonian in Eq. 2.6, the transition rate $\omega(\sigma, \sigma')$ is,

$$\omega(\sigma, \sigma') = \delta(s_i) \delta(s_j - 1) \theta e^{-\beta(2h - u_i^a(\sigma) - u_j^r(\sigma))}. \quad (2.7)$$

Where θ incorporates the entropy changes in the system due to the exchange. $u_i^a(\sigma)$ is the interaction energy between site i and the rest of the system in configuration σ , when site i is occupied by the modifier a , that is $u_i^a(\sigma) = \sum_j J_{ij}^{aa}(1-s_j) + J_{ij}^{ar}(s_j)$. A similar definition works for $u_j^r(\sigma)$. Since the reduction of PRC2 is a very slow process compared to the exchange dynamic between H3K27me3 and H3k4me3 [75], in order to average out microscopic fluctuations, we now divide the system into compartments of linear size d , and define a mesoscopic configuration $\tilde{\sigma} = \{\dots c_n \dots\}$, where c_n is the concentration of repressive (r) modifier, H3K27me3, in compartment n ,

$$c_n = \frac{1}{N_d} \sum_{i \in n} s_i. \quad (2.8)$$

where N_d is the number of lattice sites in each compartment. With this, we follow the steps in [94] and rewrite the master equation 2.5 as,

$$\begin{aligned} \frac{\partial P(\tilde{\sigma}, t)}{\partial t} = & -e^{(-2\beta h)} \sum_{nm} \sum_{i \in n, j \in m} Tr_{\sigma/\tilde{\sigma}} \delta(s_i) \delta(s_j - 1) \theta \\ & e^{(\beta u_i^a(\sigma))} e^{(\beta u_j^r(\sigma))} P(\sigma, t) + \text{gain term.} \end{aligned} \quad (2.9)$$

where $Tr_{\sigma/\tilde{\sigma}}$ indicates that the sum runs over all the microscopic configurations σ , which are compatible with the mesoscopic configuration $\tilde{\sigma}$. The gain term also represents the second contribution from the master equation, that is, the transition from state $\tilde{\sigma}$ to σ . Using mean-field approximation ¹, the transition probability is extracted from the Eq. 2.9 as,

$$\langle \delta(s_i) \delta(s_j - 1) \theta e^{-2\beta h} e^{(\beta u_i^a)} e^{(\beta u_j^r)} \rangle_{\sigma/\tilde{\sigma}} \sim \theta e^{-2\beta h} e^{(\beta g_i^a(\tilde{\sigma}))} e^{(\beta g_j^r(\tilde{\sigma}))}. \quad (2.10)$$

where $\langle \dots \rangle_{\sigma/\tilde{\sigma}}$ means the average transition rate that governs the exchange between two modifiers at sites i and j when they belong to the adjacent cells n and m , respectively. Besides, $g_i^a(\tilde{\sigma})$ and $g_j^r(\tilde{\sigma})$ are expressions for the chemical potentials of histone modifiers a and r on nucleosomes i and j ,

$$\begin{aligned} g_i^a(\tilde{\sigma}) &= KT \ln \langle \delta(s_i) e^{(\beta u_i^a)} \rangle_{\sigma/\tilde{\sigma}}, \\ g_j^r(\tilde{\sigma}) &= KT \ln \langle \delta(s_j - 1) e^{(\beta u_j^r)} \rangle_{\sigma/\tilde{\sigma}}. \end{aligned} \quad (2.11)$$

which can be extrapolated to the average chemical potentials within the defined coarse-grained compartments, $g_n^a = \frac{1}{N_d} \sum_{i \in n} g_i^a(\tilde{\sigma})$ and $g_m^r = \frac{1}{N_d} \sum_{j \in m} g_j^r(\tilde{\sigma})$. By substituting these into master equation 2.9, and defining $\mu_n(\tilde{\sigma}) = g_n^a(\tilde{\sigma}) - g_n^r(\tilde{\sigma})$, the average chemical potential within each compartment, the master equation becomes,

$$\begin{aligned} \frac{\partial P(\tilde{\sigma}, t)}{\partial t} = & -\frac{N_d}{d^2} \left[\sum_{nm} l_{mn}(\tilde{\sigma}) \exp \left[\frac{\beta}{2d} (\mu_m(\tilde{\sigma}) - \mu_n(\tilde{\sigma})) \right] P(\tilde{\sigma}, t) - \right. \\ & \left. \sum_{nm} l_{mn}(\tilde{\sigma}') \exp \left[\frac{\beta}{2d} (\mu_m(\tilde{\sigma}') - \mu_n(\tilde{\sigma}')) \right] P(\tilde{\sigma}', t) \right]. \end{aligned} \quad (2.12)$$

where $\tilde{\sigma}'$ refers to a mesoscopic configuration identical to $\tilde{\sigma}$, except a concentration unit, here identified as $\frac{1}{N_d}$, is exchanged between two compartments c_n and c_m . Moreover, the function $l_{mn}(\tilde{\sigma})$, which we call mobility ², is defined as,

$$l_{mn}(\tilde{\sigma}) \propto \exp \left[\frac{\beta}{2} (g_n^a(\tilde{\sigma}) + g_n^r(\tilde{\sigma}) + g_m^a(\tilde{\sigma}) + g_m^r(\tilde{\sigma})) \right]. \quad (2.13)$$

¹ $\langle f(i \in n) f(j \in m) \rangle_{\sigma/\tilde{\sigma}} \sim \langle f(i \in n) \rangle_{\sigma/\tilde{\sigma}} \langle f(j \in m) \rangle_{\sigma/\tilde{\sigma}}$

²This naming is based on the conventional understanding that mobility refers to the capacity for boundaries to move within a system. In our case study, this mobility depends on the material property known as the average chemical potential of active and repressive modifiers.

As studying the exact solution of the master equation 2.12 is not possible, we extend it and derive a Fokker-Planck equation that outlines the temporal changes in the concentration fluctuations of the system around a stationary state. To this end, by using a Kramers - Moyal expansion (to the second order in $\frac{1}{N_d}$), one can rewrite the master equation as (Appendix B),

$$\frac{\partial P(\tilde{\sigma}, t)}{\partial t} = \frac{1}{d^2} \left[- \sum_{n,m} \frac{\partial}{\partial c_n} l_{nm}(\tilde{\sigma}) \beta (\mu_m(\tilde{\sigma}) - \mu_n(\tilde{\sigma})) P(\tilde{\sigma}, t) + \sum_{n,m} \frac{1}{N_d} \frac{\partial^2}{\partial c_n \partial c_m} (l_{nm}(\tilde{\sigma}) P(\tilde{\sigma}, t)) \right]. \quad (2.14)$$

Whereas the Fokker-Planck equation is equivalent to a Langevin formalism, one can write down the dynamical equation for the concentration variable using *Ito* calculus as [94],

$$\frac{\partial c_n}{\partial t} = \frac{\beta}{d^2} \sum_m l_{nm}(\tilde{\sigma}) (\mu_m(\tilde{\sigma}) - \mu_n(\tilde{\sigma})) + \zeta_n(\tilde{\sigma}, t). \quad (2.15)$$

and ζ_n is a multiplicative Gaussian noise as a function of the coarse-grained configuration, with the following moments,

$$\begin{aligned} \langle \zeta_n(\tilde{\sigma}, t) \rangle &= 0, \\ \langle \zeta_n(\tilde{\sigma}, t), \zeta_m(\tilde{\sigma}, t') \rangle &= \frac{2}{N_d d^2} l_{nm}(\tilde{\sigma}) \delta(t - t'). \end{aligned} \quad (2.16)$$

Equation 2.15 is equivalent to the conventional Cahn-Hilliard equation with noise and definite expressions for the mobilities and the moments of the noise [95,96]. The connection with the well-known Cahn-Hilliard will be more clear by expressing coarse-grained chemical potentials in the Eq. 2.15 as the functional derivative of mesoscopic free energy, which for the continuous description reads,

$$\partial_t c = \nabla \cdot \left(l(\tilde{\sigma}) \nabla \frac{\delta F}{\delta c} \right) + \zeta(\tilde{\sigma}, t). \quad (2.17)$$

Note that using a proper normalization, one can neglect the explicit effect of the mesoscopic length scales d, N_d . The total free energy of the mesoscopic system reads,

$$F = E - TS. \quad (2.18)$$

Here, we provide a map that resembles the microscopic description of the system where energy, E , corresponds to the Hamiltonian 2.6, and the second term of the free energy corresponds to the configurational entropy of the system [97],

$$S = k_B \sum_n [c_n \ln(c_n) + (1 - c_n) \ln(1 - c_n)]. \quad (2.19)$$

Putting all together, the total free energy is

$$F = - \sum_n T [c_n \ln(c_n) + (1 - c_n) \ln(1 - c_n)] + \frac{1}{2} \sum_{nm} \chi_{nm} c_n (1 - c_m) + h(\tau) \sum_n c_n. \quad (2.20)$$

where we define,

$$\chi_{nm} = [J_{nm}^{aa} + J_{nm}^{rr} - 2J_{nm}^{ar}]. \quad (2.21)$$

that is the Flory-Huggins interaction parameter, and when the magnitude of the attractive interactions between *aa* and *rr* pairs are larger than those between the *ar* pairs, the system will tend to phase separate and maximize the number of *aa* pairs and *rr* pairs in the two coexisting phases. The continuum format of the free energy 2.20 is obtained by noting that,

$$\chi_{nm} c_n (1 - c_m) = \frac{1}{2} \chi_{nm} [(c_n - c_m)^2 - c_n^2 - c_m^2 + 2c_n]. \quad (2.22)$$

The difference in concentrations between two compartments can be converted to a gradient, and following a spatial continuous description ($c_n \rightarrow c(\vec{r})$), the total free energy becomes,

$$F = \int d\vec{r} f[c(\vec{r})] + \frac{\epsilon}{2} |\nabla c(\vec{r})|^2. \quad (2.23)$$

where $\epsilon = \chi/2$ and $\chi = \sum_m \chi_{nm}$, and the nonlinear local part of the free energy per unit volume, f , is,

$$f[c] = T [c \ln(c) + (1 - c) \ln(1 - c)] + \frac{\chi}{2} c(1 - c) + h(\tau)c. \quad (2.24)$$

Fig 2.16 shows the dependency of the free energy density on the various parameters. Slow changes in h impose asymmetry in the free energy, and the competition between active and repressive modifiers that are imposed by the χ parameter will eventually be tilted towards the active ones (Fig 2.16. B). Considering a fixed temperature as the temperature of a living environment for mammalian tissues, we can create a phase diagram to outline and summarize the phase behavior of the mixture.(Fig 2.16. C). Taking into account the free energy of Eq. 2.20, the phase boundary, referred to as the binodal, is established by the common tangent of the free energy at compositions c' and c'' , which correspond to the two equilibrium phases,

$$\left(\frac{\delta F}{\delta c} \right)_{c'} = \left(\frac{\delta F}{\delta c} \right)_{c''}. \quad (2.25)$$

The solid line in the phase diagram, Fig. 2.16. C shows the solution of this criterium. Furthermore, the curve corresponding to the inflection points of the free energy, which can be found by equating the second derivatives, depicted by the dashed lines, is a boundary between stable and metastable regions. Any perturbation in histone modifier concentrations in this region will grow, eventually leading to a steady state. To gain a more precise grasp of how concentration perturbations will disperse within the system, it could be helpful to

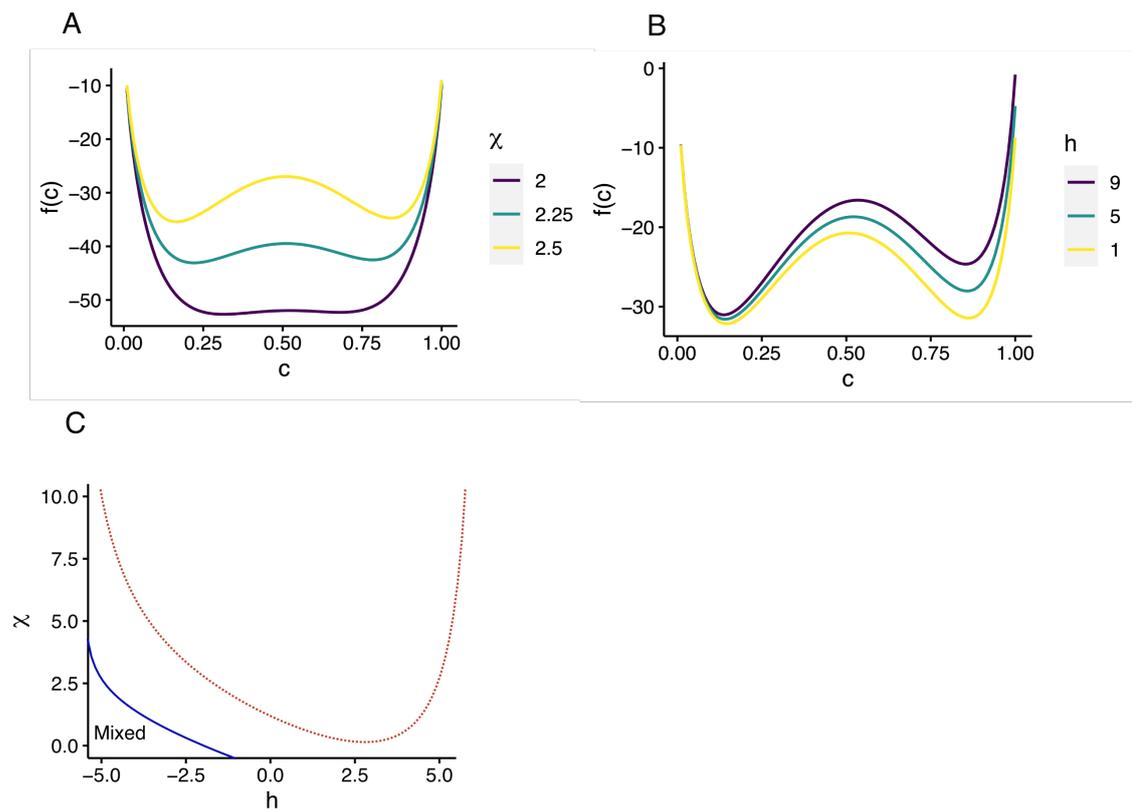


Figure 2.16: (A) Free energy density for $h = 1$. Changing (increasing) interaction χ leads to phase separation from a convex free energy (one well) to a non-convex (double well) structure, implying phase separation between two histone modifiers. (B) Free energy density for $\chi = 2.5$ and different h that resembles a decrease in the binding affinity of PRC2 and concentration. In all analyses, we keep the temperature as the temperature for mammalian cell lines, 37°C . (C) Phase diagram: The solid line shows the binomial criteria where complete phase separation appears in the system, while the dashed line represents the spinodal decomposition.

linearize equation 2.17 by expanding $\frac{\delta f}{\delta c}$ about some uniform initial composition, c_0 . Then substituting $c(r, t) = e^{(ik \cdot r - \omega(k)t)} + c_0$ in Eq. 2.17, where the first term is a minor fluctuation, amplification factor $\omega(k)$ should be tracked in terms of the wave number k to determine which frequencies will be amplified. The variation in the concentrations causes variations in chemical potentials and mobilities. However, in order to track the fluctuations to the lowest order, one may use the unperturbed values of the mobilities, $l(\tilde{\sigma}) = \bar{l}$, and obtain,

$$\omega(k) = \bar{l}k^2 \left(\frac{\epsilon}{2}k^2 + \frac{\delta^2 f}{\delta c_0^2} \right). \quad (2.26)$$

According to the classical definition [98], the spinodal region is characterized by negative values of $\frac{\delta^2 f}{\delta c_0^2}$. From Eq. 2.26, we see that $\omega(k)$ can be negative in the spinodal region for sufficiently small k ; thus, there are exponentially growing solutions of Eq. 2.17 in this region. The most rapidly growing such solution occurs at $k_c^2 = \frac{1}{2\epsilon} \left| \frac{\delta^2 f}{\delta c_0^2} \right|$.

Finally, in order to solve Eq. 2.17 explicitly, we use a sigmoidal ansatz as the initial profile for the repressive modifier, H3K27me3, and track its behavior following the studied dynamic in this section,

$$c_n = c' \frac{1}{1 + e^{-x+x_0}} + c'' \quad (2.27)$$

where c' and c'' are the minima of the free energy, and x_0 is a random variable that embedded different lengths of different CGIs with a defined mean. By substituting it into Eq. 2.17, and finding the steady state solution considering $h(\tau) = h_0$, and then using this solution and applying the concentration modification due to loss of PRC2, $h(\tau) < h_0$, we find the steady state solution once more to define the corresponding profile of repressive histone modifiers at the new age. We implemented this procedure in a simulation, computed the corresponding concentration profiles, and repeated it for different simulation ages coming from various $h(\tau)$, Fig 2.17(A). We also calculated the correlation ratios accordingly, Fig 2.17(B). Using a system size the same as the average length of CpG islands (1000 bps), we observe similar behavior from the correlation ratios, a particular gaining correlation around 500 bps during ageing. Furthermore, our simulation results predict the expansion of low methylated regions and the increase of methylation in these regions, Fig 2.17(C).

Although the results of this model qualitatively satisfy all essential criteria and the critical observations from data, to our knowledge, there is no experimental study on the mutual effect of the decrease in the concentration of PRC2 and H3K27me3 and DNA methylation changes during ageing. To adequately assess this model, ChIP-seq experiments need to be carried out to investigate the impact of ageing on DNMT3, H3k27me3, and H3k4me3 levels in the presence as well as knockout of PRC2. Only after that can adjustments be made to the free energy associated with changes in PRC2, as well as the associated increases in methylation within the region of low methylation and the rate at which the region expands in length.

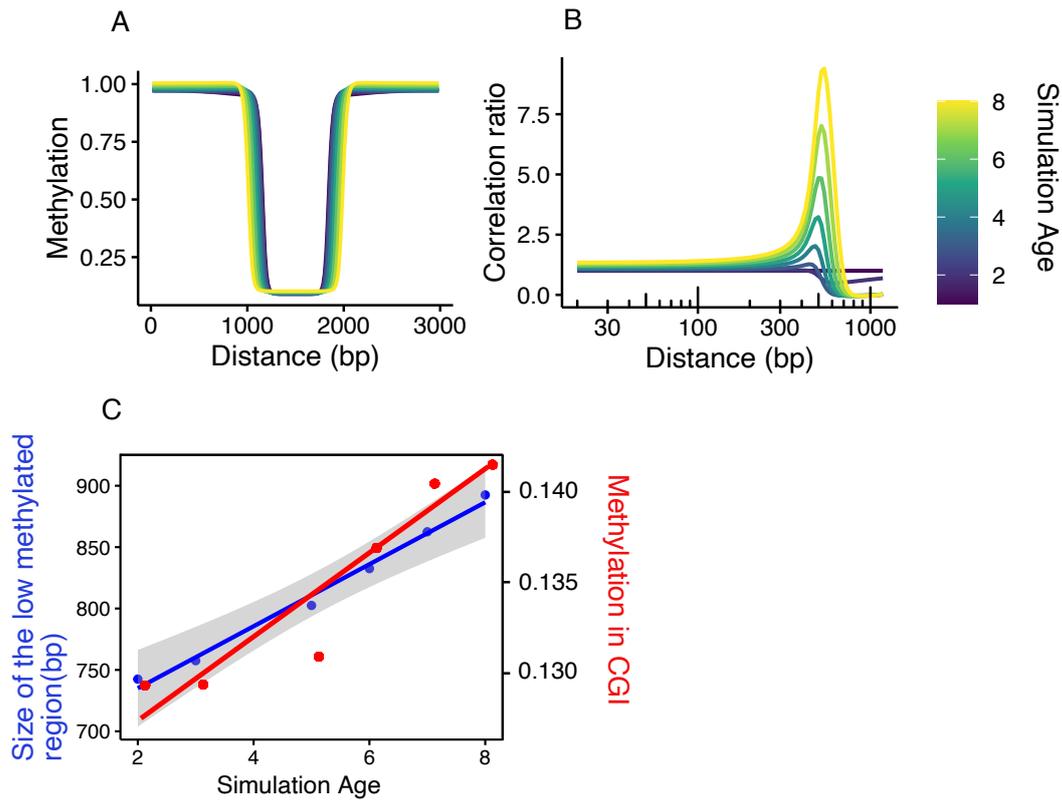


Figure 2.17: Predictions of the phase field model (Eq.2.17). (A) occupancy profile of repressive histone modifiers that resembles methylation values in CpG islands during ageing. (B) Correlation ratios for the simulated profiles show the striking patterns we expected, considering the average length of simulated CGIs. (C) Gain of methylation and expansion of the low methylated regions align with other observations from the Petkovich dataset. Different colors are associated with varying ages of simulation.

However, a few recent efforts align with the predictions we presented in Fig 2.17. Moqri et al. [99] recently reported a correlation between an age-dependent gain of methylation in low methylated regions and lower PRC2 bindings in these regions for somatic tissues. Yang et al. [75] also reported a local loss of H3K27me3 coverage in a ChIP seq experiment following their occupancy in broader regions and a decrease in coverage of PRC2 components such as EZH2 in a two-state ageing system that is young 11 weeks, and old 80 weeks mouse liver cells. However, a systematic whole-genome examination of all PRC2 targets in the context of the ageing methylome is still lacking.

2.2.4 Application to single-cell data

Discovering new perspectives on age-related biological changes will come from going beyond the misleading uniformity of bulk data and delving into the diversity offered by single-cell analysis. [100, 101]. That is because bulk methylation analysis provides an average signal across all cells, which may obscure significant differences between individual cells. However, many DNAm age associations are potentially confounded by changes in cell composition, particularly in blood where proportions of white blood cells are known to vary with age [102]. Additionally, within a given tissue or cell type, it is unknown whether DNAm ageing is a cell-intrinsic process, such that all cells age at the same rate, or whether the ageing process is inherently heterogeneous and, as such, a cell population phenomenon. This underscores the need for better understanding and the relevance of developing single-cell age predictors and analysis. Due to the recent development in single-cell technologies, it is possible to profile DNAm in single cells genome-wide. [103, 104]. However, the data sets generated by these methods differ greatly from those generated by bulk approaches in some aspects. Firstly, single-cell DNAm data is almost entirely binary, i.e., methylation values are either 0 or 1 for any given cytosine. Secondly, the data sets are very sparse, typically $> 90\%$ missing values. Furthermore, the genomic coverage is essentially random, such that a dataset of a few hundred cells will contain no or only very few genomic loci with information from every single cell, which represents major challenges for generating age predictors using current approaches that rely on deep coverages at a consistent set of CpG sites [105], see Fig.2.18.

Within this section, we will analyze a large dataset of single blood cells from mice of varying ages, addressing the pertinent issues. Additionally, we will utilize our correlation analysis method from the previous section to a certain extent.

Key sequencing experiment

An extensive collection of 1055 peripheral blood cells from mice at 4 ages (10, 36, 77, and 101 weeks of age) have been prepared by our collaborators at the Bebraham Institute. After sorting these cells, they performed single-cell *M&T - seq*. That is, they generated paired single-cell methylomes and transcriptomes (see Fig.2.19. A), from which 853 cells passed

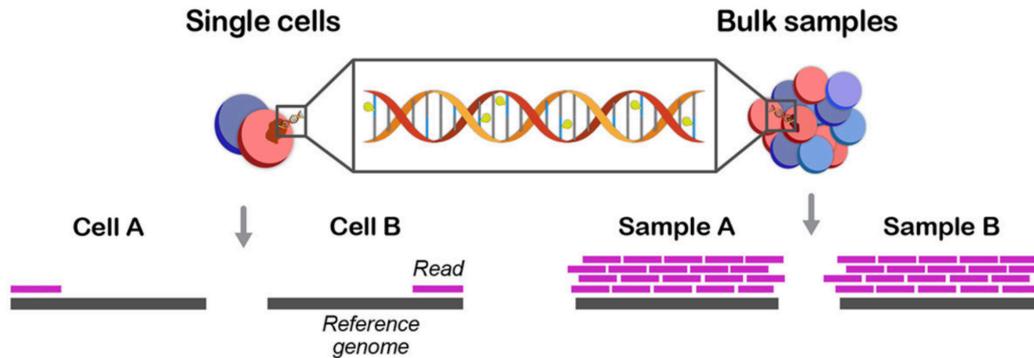


Figure 2.18: Schematic representation of the distinction between single-cell and bulk methylation sequencing outputs. With bulk approaches (right), read coverage is high and consistent between samples. In single cells (left), read coverage is low (often 1) and inconsistent between single cells, resulting in limited, distinct methylome profiles. Figure is adapted with permission from [84] (License Number 5659301508019).

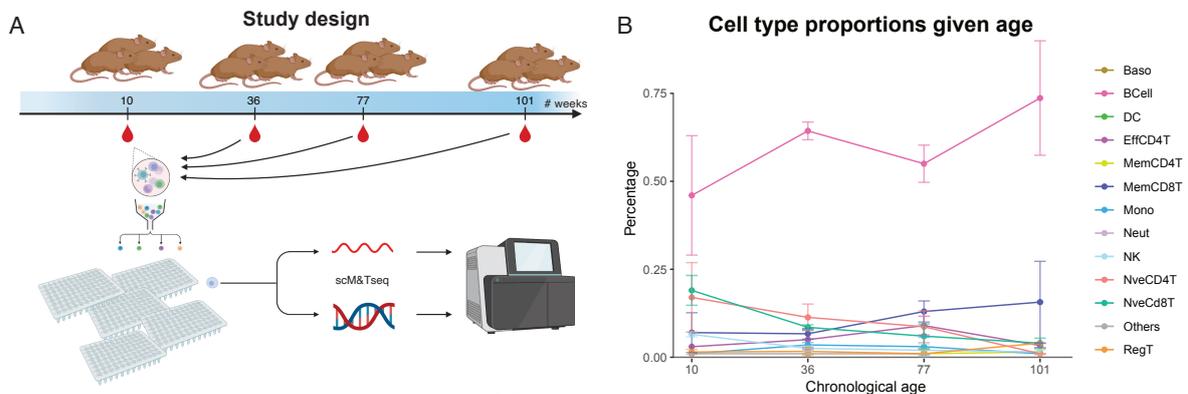


Figure 2.19: (A) Illustration of the single-cell sequencing data set collected for this study. (B) Cell type composition overview stratified by chronological age [105] (Courtesy of Marc Jan Bonder).

quality control filters ¹. Fig.2.19.B also shows a cell-type annotation based on the single-cell transcriptome data. Although we focus on DNA methylation analyses in this thesis, we consider these cell type-specificities for interpreting our analyses when needed.

DNA methylation analysis

Following the central hypothesis of this thesis that ageing is a result of a collective process that can be reflected in systematic changes of two-point correlation functions, we tried to translate our previous analysis to this single-cell data set. Fig. 2.20.A displays the connected

¹On the DNAm side, we removed cells with a low read depth (removing cells with $< 1M$ reads), removed cells with a low unique number of mapping reads (cells with $< 50,000$ uniquely aligned reads are removed), and high non-CpG methylation levels (cells with non-CpG meth $> 20\%$ are removed). For more details on quality control criteria, refer to Sec. 2.2.1.

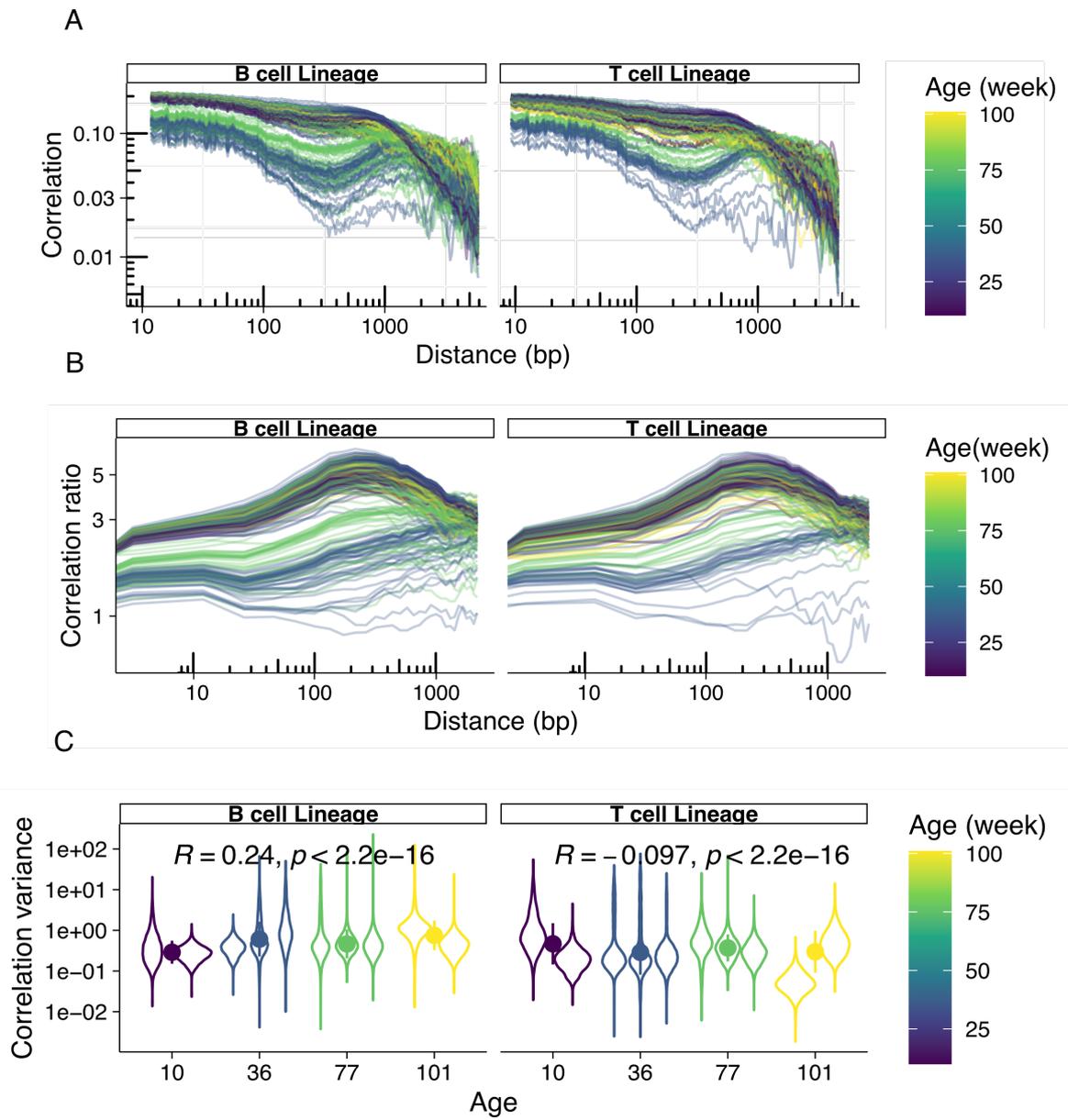


Figure 2.20: (A) Genome-wide connected two-point correlation functions between DNA methylation marks at different ages for single-cell mouse blood cells with two-cell lineages. (B) Genome-wide correlation ratios, considering the first age of the experiment as the reference. Each line corresponds to a single cell. Strong heterogeneity dominates age differences across cells. (C) Increase of correlation variance over age.

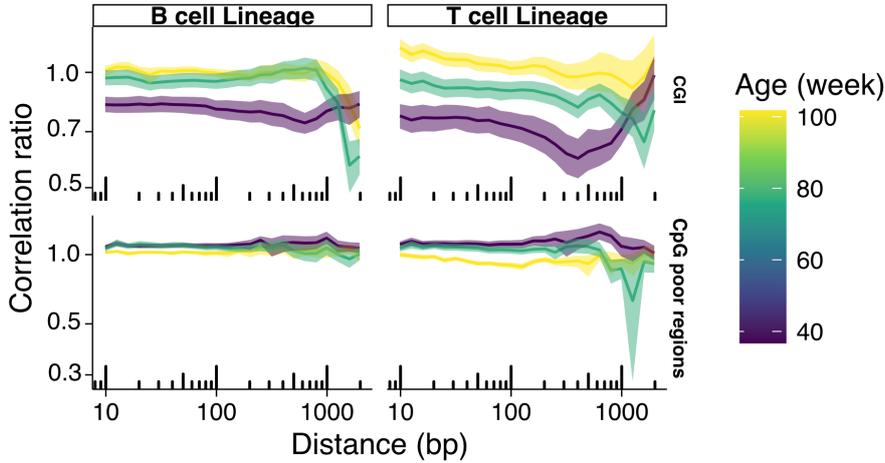


Figure 2.21: Correlation functions in pseudo bulk single-cell data show systematic changes during ageing when cells of the same age are grouped together.

two-point correlation functions on a genome-wide scale for the two most populated lineages of blood cells as depicted in Fig.2.19. B: B cell lineage and T cell lineage. Here, a very strong heterogeneity across cells would dominate any age differences previously observed for the *bulk* Petkovich data set. Subsequently, we examined the correlation ratios, Eq. 2.4, for this dataset, and as it is shown in Fig. 2.20. B, although the stereotypical behavior of gaining correlation around 500 bps is roughly detectable, there is no clear trend of ageing due to the strong variability. Moreover, these correlation variabilities increase with age, specifically for B cells, as shown in Fig. 2.20. C. Despite the strong cell-to-cell heterogeneity, we found that average DNAm correlations changed systematically with age when cells with the same age pooled together as a *pseudobulk* data set. Fig.2.21 shows this behavior for two cell lineages filtered by the density of CpG regions. In order to examine the varying impact of different genomic regions on the ageing process in more detail, we also examined alterations in DNA methylation correlations based on distance within various genomic features, see Fig.2.22. As depicted in this figure and in line with our expectations, CpG islands are protected from unspecific changes, which can be seen in the behavior of the correlation ratios. In particular, this region’s general short-distance gain of correlation is evident. Nevertheless, unspecific ageing changes also predominantly affect gene bodies and enhancers, particularly those involving developmental genes and Polycomb target sites, and the characteristic length scale of this effect is preserved.

Taken together, we observed ageing as a characteristic behavior at the level of tissues or general groups of cells. At the same time, it is not well defined at the level of single cells. This can be further validated by examining low-dimensional representations of single-cell correlation functions across genomic annotations. A suitable dimension reduction technique for this purpose is the Uniform Manifold Approximation and Projection (UMAP). UMAP is a machine learning tool that aids in visualizing and comprehending extensive, high-dimensional

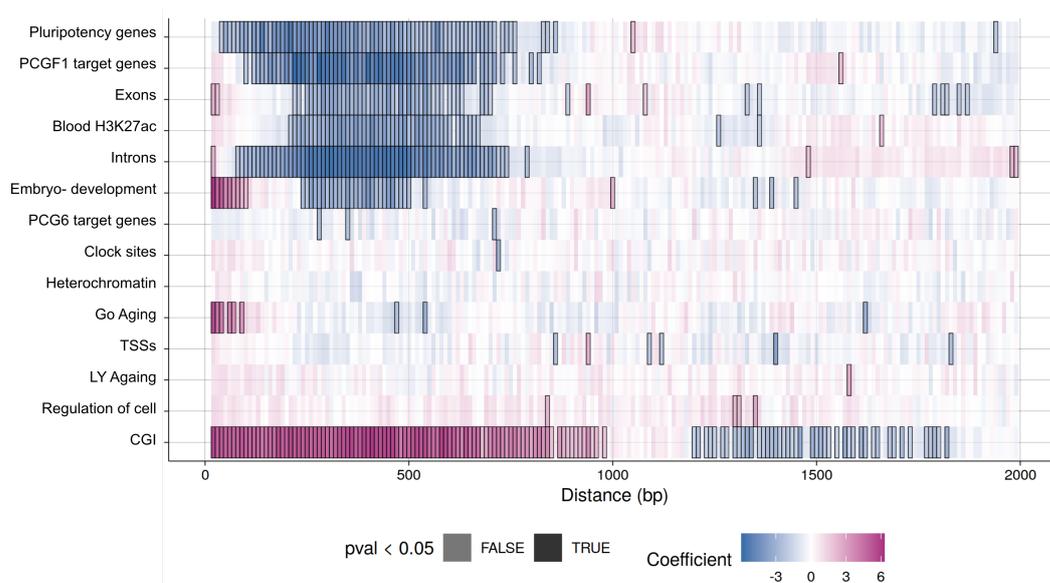


Figure 2.22: Distance contribution of different genomic regions in multiplicative changes of correlation function beyond the correlated noise. Blue shades correspond to loss of correlation, and red corresponds to correlation gain.

datasets, like our single-cell correlation data. To achieve this purpose, UMAP that is a non-linear and topology-preserving algorithm assumes that the original data is uniformly distributed on a Riemann manifold. By learning this manifold, it preserves topology, ensuring that points close together in two dimensions also maintain proximity in higher dimensions.

Fig.2.23 represents a low dimensional representation of correlations. There is a noticeable overall trend of ageing in specific genomics regions like CGIs, promoters, and polycomb target genes; this trend can also be seen in the heatmap Fig.2.22. However, cells are generally evenly dispersed in this pattern, making it difficult to identify distinct clusters of cells with different DNAm ages. This suggests that individual cells within a tissue may age at different rates, which could be associated with variations in cell function. However, one may ask if this variability is genuinely biological or comes from the technical basis in individual annotations. To answer this question, we represent another UMAP of the same data, considering total coverage in different genomic regions that could be one of the most important sources of variability. As is shown in Fig. 2.24, there is no spectacular technical bias, indicating that age variability across single cells within one chronological age is biologically genuine.

In the upcoming section, we present similar findings using a different approach, specifically utilizing machine learning methods. This analysis was primarily carried out by our collaborators at EMBL and ETH [105], and the specifics are not covered in this thesis. Nonetheless, we provide a brief explanation of the technique as it will aid in understanding the age diversity we observed in single cells using DNAm correlation analysis.

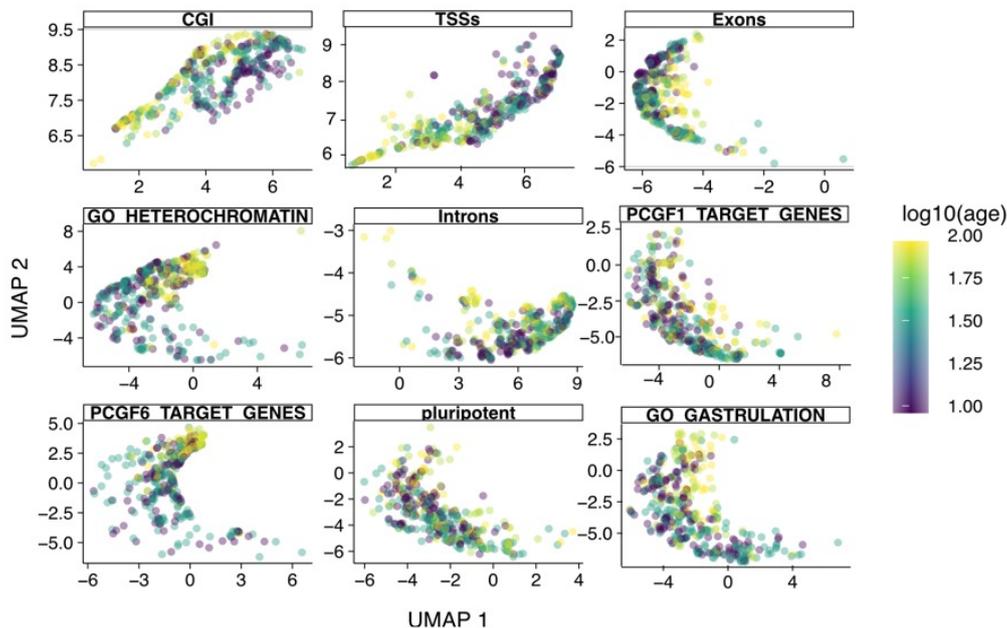


Figure 2.23: Low dimensional representations (UMAP) of single-cell correlation functions across genomic annotations.

Single-cell DNAm age prediction using machine learning

The main difficulties in evaluating the chronological age of individual cells are the sparse and binarized methylation patterns they possess. This makes it difficult to apply traditional regression approaches, which depend on consistent coverage of CpG sites across samples (see section 2.1.2). To this end, a ranked intersection algorithm has been designed to overcome the intrinsic sparsity of single-cell methylomes. In short, the top age-correlating sites were present in the intersection of the single-cell dataset, and the other two bulk datasets for blood tissue, Petkovich et al., 2017 and Thompson et al., 2018 have been selected. Next, and to address the challenge of binarized methylation, it's been hypothesized that the methylation levels observed in densely covered CpG sites within bulk sequencing data could serve as an approximation for the probability of binary methylation at these specific CpG sites in single-cell datasets. That is, if one measures a bulk methylation level of 0.7 at a single CpG and picks a random single cell from that tissue, one can assume that there is a 70% chance this particular cell will be methylated at this locus. Then the most likely age is computed by multiplying the probabilities across loci and selecting the maximum. To demonstrate the ability of our model to assign DNAm age to a broader range of cell types, we also generate

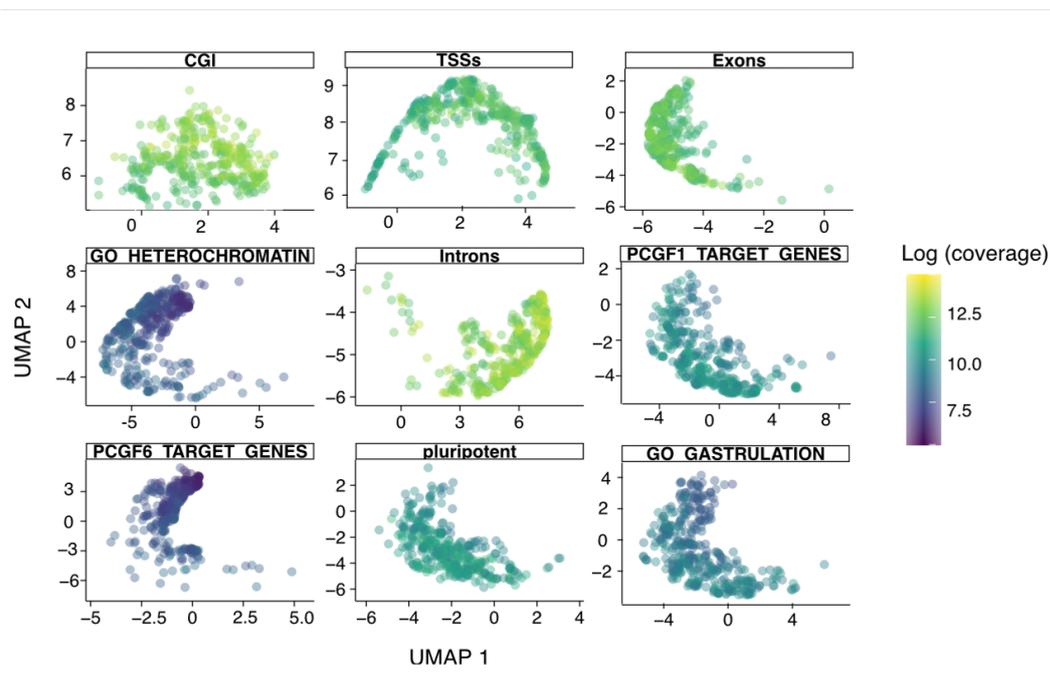


Figure 2.24: Low dimensional representations (UMAP) of single-cell correlation functions across genomic annotations. There is no spectacular technical bias in individual annotations coming from coverage differences.

a similar ageing model for single-cell liver DNAm, by considering liver bulk datasets as the intersection references: Reizel et al. study, 2015, Thompson et al., 2018, and Meer et al., 2018. The results of these models are depicted in Fig.2.25. As shown in this figure, we first evaluated the model’s performance at the pseudo-bulk level by combining cells from the same donor. This was done to determine the influence of batch effects and various sequencing methods. Our findings showed an MAE¹ of 10 weeks in blood and an MAE of 6.5 weeks in liver. However, single-cell predictions result in a broader spectrum; the MAE increases to 26 weeks in blood and 16 weeks in the liver when predicting actual single cells. Furthermore, a comparison of our liver and blood predictor showed that epigenetic ageing signals are stronger (more accurate) in the liver as compared to blood, possibly reflecting the more complex cell type distributions in blood. Despite the promising results of this work, answering the fundamental question of whether cells age at different rates remains challenging, and further integrated studies need to be done.

2.3 Summary and discussion

In this chapter, we investigated the collective epigenetic processes underlying DNA methylation modifications during ageing, and whether DNA methylation age is measurable in single cells. We found that the DNA methylation ageing clock is the result of a collective process

¹Mean absolute error (MAE), the sum of absolute errors divided by the sample size

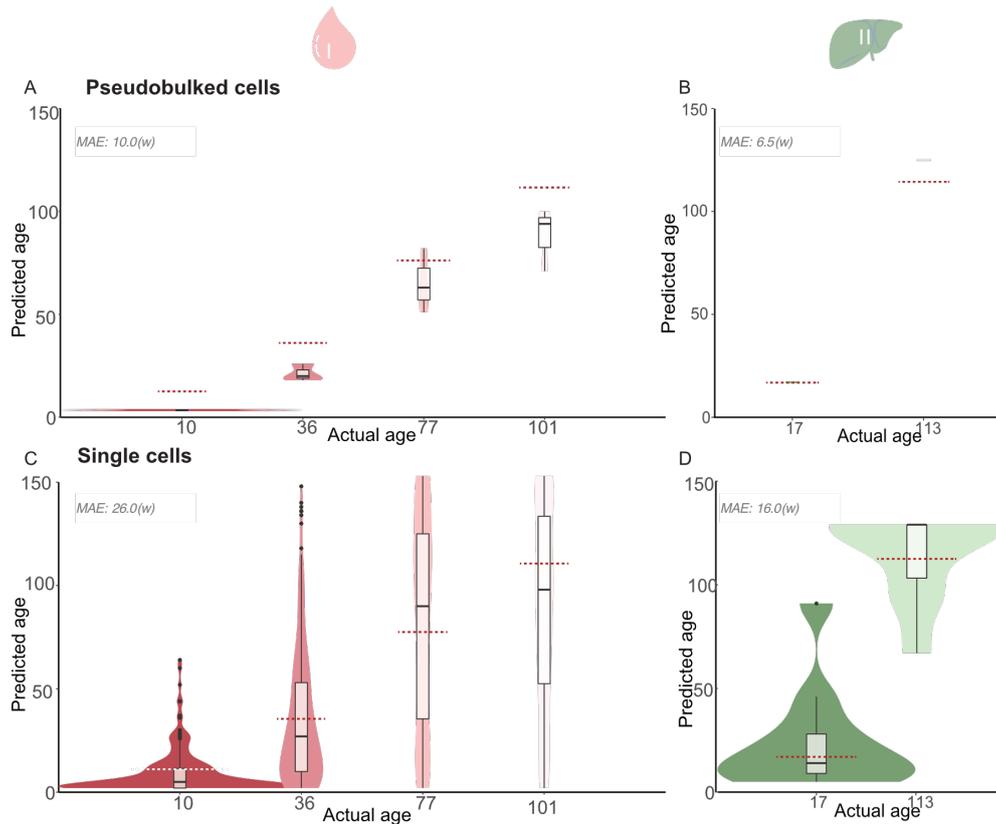


Figure 2.25: (A-B) Predicted DNAm age of pseudo bulked (A) blood and (B) liver single-cell data. C-D) Predicted DNAm age of the single cell (C) blood and (D) liver data [105] (Courtesy of Marc Jan Bonder).

that is reflected in systematic changes in two-point correlation functions. To quantify these systematic changes over time, we calculated the ratios of correlations at a given distance and consecutive time points, which is a measure of changes in correlation beyond uncorrelated noise. Our analysis demonstrates that these changes increase over time around a characteristic distance of $500bp$. To reveal the genomic origin of this striking pattern of ageing, we analyzed multiplicative changes in correlations separately for different genomic regions classified by the density of CpGs. We found that the gain of correlations at a distance of $500bp$ is specific to regions with high CpG density, termed CpG islands. As the average distance of sites inside a CpG island to the boundary is $500bp$, we hypothesized that the increase in correlations at this distance is associated with DNA methylation dynamics at the boundaries of CpG islands (shores). This is supported by our observation that the position of the peak in multiplicative changes in correlations scales linearly with the size of CpG islands. We also found that these changes in DNA methylation correlations characterize ageing in a broad range of tissues. Our analysis also revealed that during the ageing process, there is an increase in the size of low methylated regions that are mainly linked to CpG islands.

At the same time, the methylation levels inside these islands show a rise as well. We studied the crosstalk between modifications to DNA methylation and histones, which are the proteins responsible for compacting DNA. We discovered that the ageing indicators identified through correlation ratios are specific to CpG islands with histone modifiers associated with gene activation (H3K4me3) and suppression (H3K27me3), termed bivalent CpG islands. Considering a decrease in the concentration of H3K27me3 due to the depletion in polycomb enzymes essential for establishing H3K27me3, we developed an asymmetric interactive two-component phase field model that can reconstruct DNA methylation profiles over time. To this end, we started with a microscope system description and systematically coarse-grained it. While our model meets all necessary criteria and addresses critical observations from the data, additional experiments are required to fine-tune the parameters. Specifically, we need to investigate how changes in the concentration of PRC2, H3K27me3, and DNA methylation interact during the ageing process.

Furthermore, we tried to study ageing dynamics at the level of single cells by investigating their collective behavior and through machine learning models. Conducting both approaches, we faced solid cell-to-cell heterogeneity within a group of cells with the same chronological age, making biological age prediction challenging. Nevertheless, we found how genuine this heterogeneity is by minimizing any technical biases that could affect this study. Further studies need to be conducted for a comprehensive understanding of single-cell ageing.

Chapter 3

Epigenetic reprogramming during regeneration

3.1 Introduction

In Chapter 2, we examined the alterations in epigenetic modifications that occur during ageing. We explored how the interaction among various layers of modifications gives rise to collective behavior that can be utilized to predict biological age. While the epigenetic changes related to aging generally occur slowly over a prolonged period, there are instances, as discussed in Chapter 1, when these modifications can be altered rapidly. One crucial example is during regeneration, where differentiated cells display extraordinary adaptability by reprogramming into various cell types and removing DNA methylation marks through passive and active demethylation processes. In this chapter, we will investigate the collective behavior in liver cells' reprogramming during regeneration, focusing on DNA methylation and chromatin accessibility. Our analysis reveals that DNA methylation levels remain consistent throughout the regeneration process, indicating that they do not differentiate between different stages of regeneration. However, the structure of the chromatin is significantly dynamic. To better understand these modifications, we propose a theoretical framework that elucidates the mechanisms underlying epigenetic changes and how the system remembers past injuries during regeneration.

3.1.1 Epigenetic regulation in liver regeneration

Throughout adult life, tissues maintain cellular function and constant cell number through robust homeostatic mechanisms, maintaining the fragile equilibrium between proliferation and differentiation. The rate of cellular proliferation depends on the turnover requirement of the tissue [106,107]. In the mammalian system, the intestine and the skin are amongst the organs with the highest cellular turnover, such that they take advantage of specific adult stem cell compartments that are able to generate all the cell types of the resident tissue in order to support homeostasis [108]. In contrast to the organs with high cellular turnover

during homeostasis, the adult liver, which has a remarkable regenerative capacity, does not contain a defined stem-cell compartment to maintain homeostasis. However, after acute liver injury (e.g., toxic injury, alcohol, viral infection, and even fatty diet), regeneration can be accomplished by proliferating quiescent and mature cells, which re-enter the cell cycle and exhibit remarkable plasticity¹ by reprogramming into different progenitors [109]. In order to understand this property, let us briefly discuss the liver's structure and functionality.

The liver is the largest organ inside the body and undertakes numerous essential functions., such as regulating homeostasis and metabolism. It has a highly organized architecture and contains several cell types, including hepatocytes, cholangiocytes (also named ductal cells), endothelial cells, Kupffer cells, and stellate cells. Most of the metabolic functions are carried out by hepatocytes, which are epithelial cell types that account for the greater part of the organ size (more than 60 %). Hepatocytes manage the processing and uptake of nutrients, manufacture serum proteins and blood clotting factors, and break down pharmaceutical drugs and toxins. Cholangiocytes, categorized as biliary epithelial cells, shape bile duct tubules and are responsible for altering and moving bile (released by hepatocytes) along the biliary tree into the small intestine. [110]. Both hepatocytes and cholangiocytes start to proliferate to restore the size and functionality of the liver upon injury. This can be done by acquiring a bipotent liver progenitor capable of generating both liver epithelial cell types [21, 111]. Studying these regulations in liver cell-fate decisions required understanding the epigenetic mechanisms in response to environmental cues that regulate chromatin structure and gene transcription needed for reprogramming the genome. Yet, this area is not fully understood, and the molecular mechanisms through which adult (Quiescent) cells depart from their lineage-restricted state and react to damage are mostly unidentified [21]. Using a mouse model, it has been found that DNMT1 plays a crucial role in maintaining the identity and function of the hepatocytes, such that deletion induced liver fibrosis and inflammation and resulted in hepatocyte senescence due to aberrant methylation changes [112]. Moreover, Arid1a, a DNA interacting subunit of the SWI/SNF chromatin remodeling complex has been shown to regulate hepatocyte plasticity through increased chromatin accessibility. Such permissive states facilitate the chromatin binding of the transcriptional machinery (e.g., YAP/TEAD) that promotes the establishment of liver progenitor identity [113]. Similarly, Cholangiocytes undergo epigenetic remodeling mediated by the methylcytosine dioxygenase TET1 that oxidizes the repressive DNA mark 5-methylcytosine into 5-hydroxymethylcytosine (5hmC). 5hmC can act as a stable epigenetic mark associated with gene activation or represent an intermediate of complete de-methylation after further oxidation catalyzed by TET proteins (see Fig. 1.1). It has been found that TET1 was lowly expressed in homeostatic adult cholangiocytes and was up-regulated following DDC-mediated liver injury in vivo [21]. Up-regulation of TET1 also facilitates the expression of components and targets of signaling

¹Here, plasticity is defined as the ability of a cell to acquire novel features or adopt alternative fates in a tissue-specific, controlled manner in response to distinct context-dependent intracellular or extracellular cues [106].

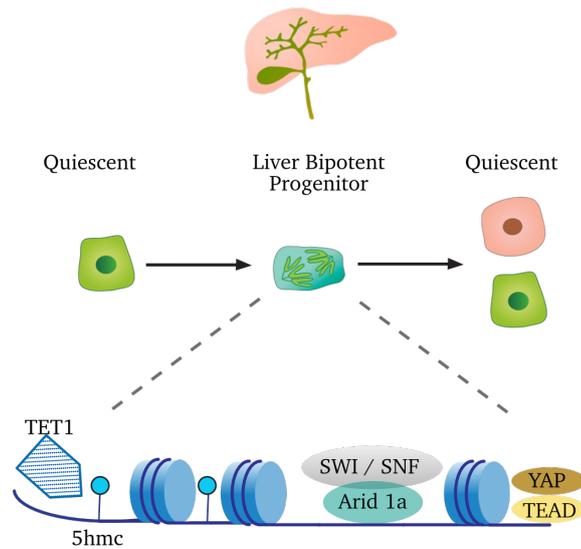


Figure 3.1: Some epigenetic modifications that allow cell-fate changes into bipotent liver progenitors. Upon injury, both adult (Quiescent) hepatocytes and cholangiocytes can de-differentiate into bipotent liver progenitors, giving rise to both liver epithelial cell types. These modifications include increased chromatin accessibility mediated by Arid1a and the chromatin remodeling complex SWI/SNF [113] and oxidation of 5-methylcytosine (5mC) into 5-hydro methylcytosine (5hmC) mediated by the methylcytosine dioxygenase TET1 [21]. Such permissive states facilitate the chromatin binding of the transcriptional machinery (e.g., YAP/TEAD) that promotes the establishment of liver progenitor identity.

pathways required for liver regeneration, including YAP. The epigenetic modifications that we reviewed in this section are depicted in Fig.3.1. Similar to the ageing system introduced in Chapter 2, the spatiotemporal structure of epigenome during regeneration undergoes a complex process and has been affected by many degrees of freedom. However, despite its complexity, we are again interested in understanding the emergent properties, such as collective behaviors, that can be studied through an effective statistical model. With this in mind, we will approach our research questions and give an overview in the next section.

3.2 Research questions and overview of this chapter

Solving the cellular reprogramming puzzle is impossible without a comprehensive study of the different spatiotemporal scales of the epigenome. In Chapter 2, we delved into investigating epigenetic modifications over an extended temporal scale, allowing us to explore the underlying mechanisms of the aging process. Nonetheless, it is worth noting that the epigenome can also undergo fast alterations, such as those occurring during tissue regeneration following injuries. The liver stands out as an exceptional organ endowed with a remarkable regenerative capacity. The examination of epigenetic modifications during the process of liver regeneration holds the potential to unveil the molecular underpinnings that propel this regenerative phenomenon. Thus, in this chapter, our primary objective is to address the following query:

What is the mechanistic basis underlying DNA methylation during the regeneration of the liver? Trying to answer this question, we narrow our focus to investigate whether liver epithelial cells preserve a memory of the damage-repair process or if they revert to their initial state upon recovery. For this purpose and in collaboration with the group of Dr. Meritxell Huch (MPI-CBG), we drew on a single-cell experiment conducted for different phases of liver regeneration, including pre-injury, proliferation (response to the toxin diet/injury), and recovery and termination. In the upcoming sections, we conduct a thorough examination of this dataset. After making critical observations related to the memory effect within the system, which emerges from our analysis of connected correlation functions, we proceed to employ a minimal coarse-grained model aligned with the bistable nature of chromatin structure. We leverage concepts from non-Markovian processes and the generalized Langevin equation to quantify the behavior inferred from these correlation functions.

3.3 Results

3.3.1 Sequencing experiment

In this study, our primary emphasis is on the regeneration of ductal cells within the adult mouse liver in response to tissue damage, a condition that typically poses a challenge to the liver's regenerative capabilities. To induce such damage, a diet enriched with 0.1% 3,5-diethoxycarbonyl-1,4-dihydrocollidine (DDC), which selectively eliminates hepatocytes, was administered to the liver. Consequently, neighboring ductal cells are compelled to undergo proliferation in order to compensate for the inflicted damage [21]. Then, ductal cells were collected at different time points from priming (when cells sense the damage but without proliferation) to proliferation to recovery and termination state (fully regenerated tissue). Following cell collections, a single-cell sequencing experiment known as scNMT-seq was carried out for each time point (see Fig. 3.2). This experiment was designed to simultaneously examine chromatin accessibility, DNA methylation, and transcriptome data. Since DNA methylation and other epigenomic factors, such as chromatin accessibility, are not operating in isolation but are intricately interconnected [114], the ability to profile various epigenetic features alongside gene expression is crucial for obtaining a more comprehensive understanding of epigenetic interdependencies [33]. Note that the standard protocol for detecting methylation marks primarily involves bisulfite sequencing, as introduced in Sec.1.2. However, for the purpose of parallel analysis of chromatin accessibility, a slight modification is applied. To simultaneously assess chromatin accessibility and DNA methylation, a GpC methyltransferase is used to mark accessible regions before conducting bisulfite sequencing (BS-seq). This alteration enables the differentiation between these two epigenetic states. In mammalian cells, CpG dinucleotides are typically highly methylated, while cytosine, followed by adenine, cytosine, or thymine (referred to as CpH), is methylated at a much lower rate. Therefore, the use of a GpC methyltransferase to label accessible chromatin allows for the

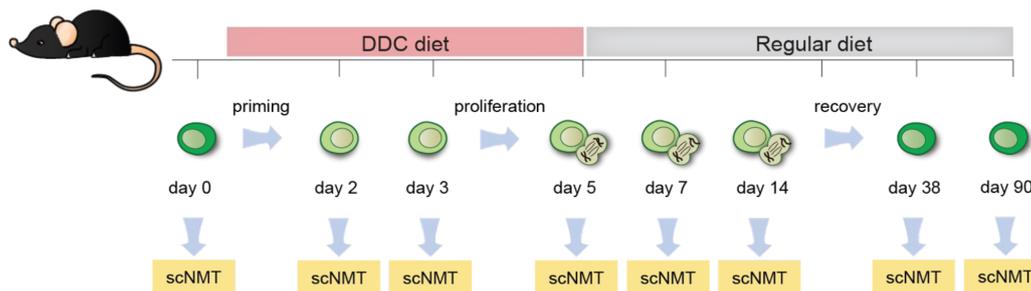


Figure 3.2: Experimental design, courtesy of German Belenguer at MPICBG, for studying the regeneration of ductal cells. Genomic DNA from undamaged or DDC-damaged livers undergoes a scNMT sequencing experiment at each time point.

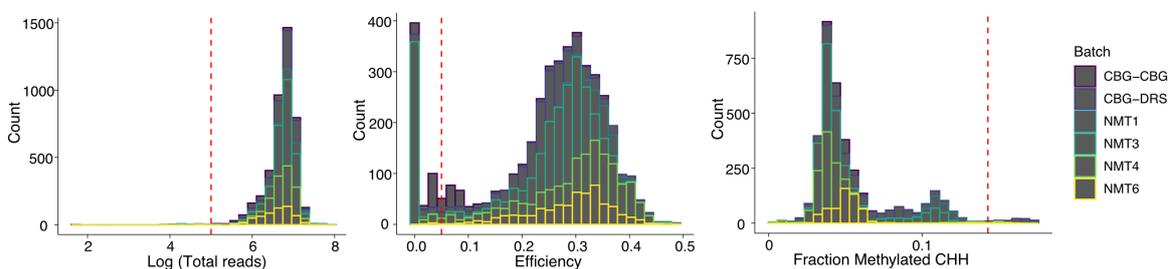


Figure 3.3: Quality control of the bisulfite sequencing component of the scNMT-seq experiment. We retained cells meeting specific criteria, including a total read count exceeding 100000, mapping efficiency surpassing 5%, and CHH methylation levels below 15%. The dashed red lines serve as visual references denoting these criteria. Various colors are used to denote different batches within the experiment.

simultaneous capture of CpG methylation details [33]. A portion of this experiment, specifically covering the initial three time points, had been conducted and independently published before this research [21]. In the case of the new dataset, we obtained the raw sequencing data directly from the sequencing machine. Our responsibility included performing preliminary analyses, encompassing quality control, trimming, and aligning the Bisulfite-Seq data to the reference genome, all preceding any statistical measurements. A concise overview of this workflow can be found in Appendix C. Here, we proceed with analyzing the aligned dataset. Our processed dataset comprises substantially aligned 4291 cells originating from six distinct experiment batches. We initiated the quality control process, retaining cells meeting specific criteria: less than 15% CHH methylation, more than 5% mapping efficiency, and a total read count exceeding 100,000. Consequently, we opted to exclude approximately 12% of the total cell population from further analysis. Fig.3.3 illustrates the various quality control measures applied to this dataset. Once the data is prepared, it is possible to proceed with the statistical analysis. This begins by examining the changes in the overall methylation and chromatin accessibility depicted in Fig.3.4. It is revealed that the global changes in methy-

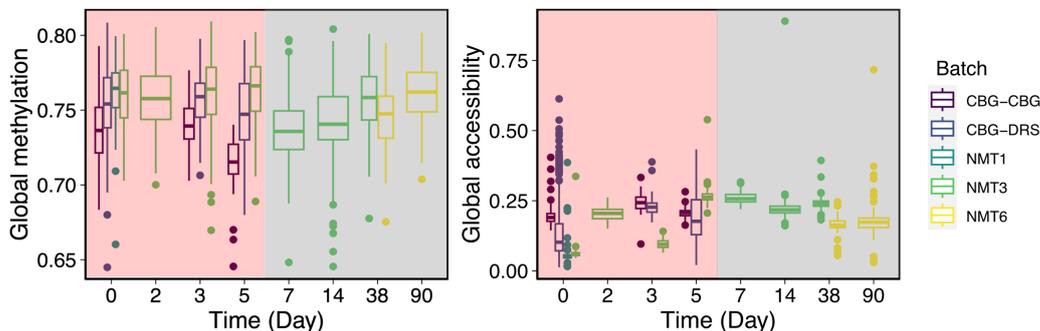


Figure 3.4: Global DNA methylation levels and global chromatin structure are retained during regeneration: The mean of methylation and accessibilities are depicted for different experiment batches. Plots are split into two sets of red and gray colors associated with various stages of the regeneration experience during and after the DDC diet, respectively.

lation levels before, during, and after regeneration are insignificant. Only slight differences in global accessibility levels are observed. However, the heterogeneity between different experimental batches is more significant than any changes in accessibility. Although epigenetic modifications are essential for cell reprogramming and gene transcription, monitoring global behavior does not provide much insight. Consequently, the focus shifts towards examining the collective behavior of cells in local regions. This objective will be pursued in the following section, where regeneration is viewed as a collective epigenetic process.

3.3.2 Collective epigenetic processes during regeneration

In order to challenge our hypothesis that the reprogramming of cells during liver regeneration is a result of collective behavior, we first asked whether there is any local association between DNAm and DNA accessibility, which can be summarised in the connected cross-correlation of both quantities. The cross-correlation function quantifies the correlation between DNAm and DNA accessibility at sites a given distance apart. Analogously to the Eq.2.3, the cross-correlation function for a given distance, j is defined as,

$$CC(j) = \langle m_i a_{i+j} \rangle - \langle m_i \rangle \langle a_{i+j} \rangle. \quad (3.1)$$

where a_{i+j} is the accessibility state at position $i + j$. Calculating this quantity for our processed dataset introduced in the previous section first reveals the negative relation between DNA methylation and DNA accessibility, characterized in Fig. 3.5. It becomes evident that the regeneration process greatly influences the cross-correlation between methylation and accessibility. After liver injury from DDC poisoning, the correlation between methylation and accessibility is decreased. To be more precise, the local compaction of the chromatin

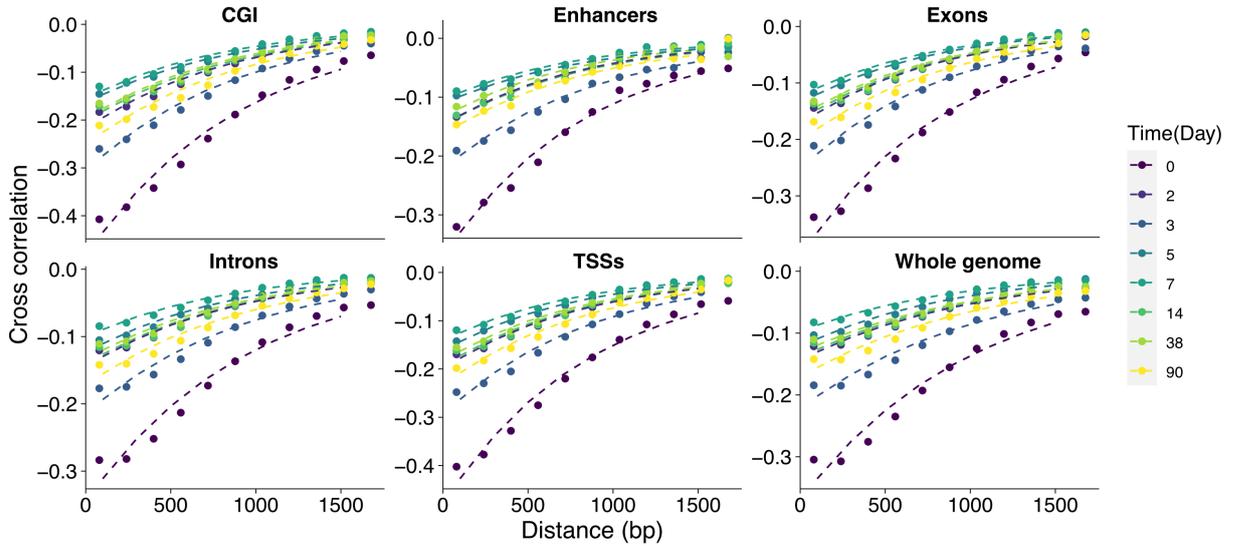


Figure 3.5: Connected cross-correlation between DNA methylation and accessibility for the whole genome as well as different local features along the genome. The inverse relationship between DNA methylation and accessibility remains similar for different genomic features. However, following liver injury caused by DDC poisoning, the association between methylation and accessibility is disrupted. Although it gradually returns to initial levels during the recovery period, it never fully recovers to the same level as day 0 throughout the duration of the experiment. Dashed lines are exponential fits to the data using R’s nls function.

is less correlated with methylation values after regeneration. Although, as it is observed in Fig. 3.5, it slowly returns to baseline levels during recovery, it never fully recovers to the same level as before the poisoning throughout the experiment. These patterns also remain consistent across various genomic regions and the entire genome. Based on this observation, our hypothesis is that the system’s “*history*” possibly affects the changes in the interplay of DNA methylation and chromatin accessibility. Thus, we continue by examining the autocorrelations of methylation and accessibility separately in order to test our hypothesis and determine its underlying mechanisms. According to Fig. 3.6, the autocorrelation of methylation, which represents the correlation between methylated sites (as shown in Eq. 2.3), decay for all time points during the experiment. Moreover, this spatial correlation between methylated sites is not altered throughout the experiment. In contrast, when examining the autocorrelation of accessibilities, which represents the correlation between accessible sites, we only observe the correlations decay along the genome for their envelope, which can be achieved by reducing the number of bins used in calculating the correlation along the genome. Without this adjustment, we observe oscillations with a period of approximately 150 base pairs, which corresponds to the length of the DNA wrapped around a single nucleosome. This indicates that the chromatin structure is mainly preserved at the nucleosome level on a short-length scale, that is, the size of the nucleosome and inter-nucleosomal DNA. However,

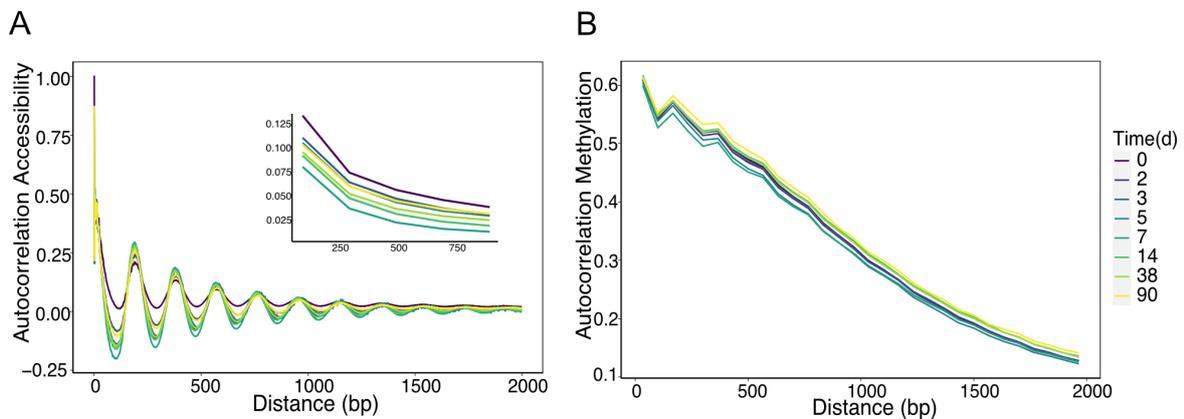


Figure 3.6: Autocorrelation of (A) accessible cytosines and (B) methylated cytosines along the genome. Both plots show decaying behavior; there are no changes detected between the different time points of the regeneration process for the methylation; however, the structure of chromatin is disturbed during regeneration due to the autocorrelation of accessibilities. Although no clear trends were detected here, it seems that the structure of the chromatin will first be looser/less compact after sensing the poison and starting the regeneration, and it will slowly compact and relax to the initial state at late time points of the experiment. The inset of autocorrelation for accessibilities shows the same measurement only with lower bins. Both autocorrelations for accessibility and methylation have been normalized such that the area underlying the curve, which in the case of correlation functions is dependent on sequencing coverage, is equal to 1.

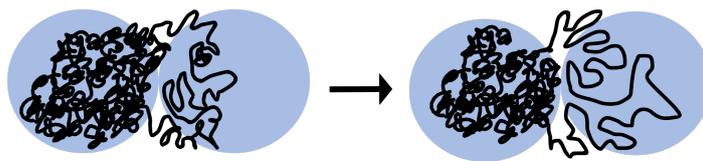


Figure 3.7: Schematic representation of liquid-liquid phase separation in chromatin. The purpose of the blue droplet shapes is to visually indicate the two main compartments of chromatin: the more compacted and less compacted regions. This organization is not static but rather adaptable to environmental changes, such as injury. For example, based on our findings in Fig. 3.6, the chromatin tends to become looser/less compact during the regeneration process. However, whether it returns to its pre-injury structure is uncertain.

deviations from this pattern can be observed on larger scales from the beginning of the experiment. This is shown in the inset of Fig. 3.6. Although we do not see a clear trend in the alteration of the chromatin structure from the first time point of the experiment, from Fig. 3.6, we can interpret that the chromatin structure will first be looser after sensing the poison and gradually recover back to its initial state before poisoning at the later time points of the experiment.

Therefore, one can conclude that chromatin is organized at multiple scales, and during the process of regeneration, these levels are rearranged both spatially and temporally. To investigate the effects of perturbations such as poisoning (injury) on this system and to interpret the changes we observe in the spatial correlation for accessible sites (Fig. 3.6), we utilize a coarse-grained model. In the following section, we combine the evidence of liquid-liquid phase separation of chromatin from literature along with our own observations shown in Fig. 3.6. Using a minimal free energy approach, we develop a dynamic equation for the accessibility of the DNA. We then employ a field theory approach to solve this equation and determine the corresponding autocorrelations. In the discussion section, we acknowledge that while the model’s results align with the observed behavior of correlation functions, further validation and refinement can only be achieved through experimental investigation.

3.3.3 Towards a biophysical mechanism of regeneration

Chromatin exhibits structural dynamics over multiple scales in space and time. In fact, chromatin dynamics are central to its function as they provide biochemical access for the cellular machinery to the underlying genomic DNA that facilitates gene expression, DNA replication, repair, and recombination. Due to the flexible nature of chromatin that makes these dynamics possible, such as its ability to fuse and rapidly exchange components with its surroundings, it has been believed that chromatin has a liquid-like behavior [115, 116]. Moreover, the development of new technologies like whole-genome chromosome conformation capture (Hi-C) reveals that chromatin is, in general, compacted inside the cell nucleus while it has two main compartments, the mostly transcriptionally active (less compacted/ euchromatin) and the mostly transcriptionally inactive (more compacted/ heterochromatin)

compartment, which is known as phase separation. Considering the liquid-like behavior of these compartments and being membrane-free, this phenomenon is called liquid-liquid phase separation in physics terminology [116–120]. This chromatin compartmentalization must be dynamically orchestrated and amenable to acute regulation for processes like injury and subsequent reprogramming/regeneration, see Fig. 3.7.

With this in mind, and in order to write a dynamical equation for chromatin accessibilities and to investigate the corresponding correlations, we start with a general but effective free energy for the coarse-grained field of accessibilities (structure of chromatin), φ , that fulfills the dynamic of liquid-liquid phase separation of chromatin,

$$F = \int d^3x \left(f[\varphi(x)] + \frac{1}{2} |\nabla\varphi(x)|^2 \right). \quad (3.2)$$

This free energy is similar to Eq.2.23, in chapter 2, where we discussed DNA methylation behavior during aging based on an interactive two-component phase field model. We have already learned that the free energy density has to have non-convex regions for phase separation to occur, such as the Flory-Huggins free energy with a significant enough interaction parameter that we used in Chapter 2. Here, we use a variation of a so-called Ginzburg-Landau free energy density. [121]. A combination of a quadratic term and a biquadratic term generates a concave interval in the free energy density, and the total free energy follows,

$$F = \int d^3x \left(\frac{1}{2} a_0 \varphi(x)^2 + \frac{\lambda}{4!} \varphi(x)^4 + \frac{1}{2} |\nabla\varphi(x)|^2 \right). \quad (3.3)$$

Where a_0 and λ are arbitrary constants parameters named as the so-called mass of the model and coupling coefficient, respectively [122]. Given the availability of this free energy, we can proceed with the dynamical equation of the accessibility field using the Mori-Zwanzig formalism [123]. This approach is necessary because the interactions of multiple unknown components, which we discussed in Section 3.1.1, have an impact on chromatin structure. However, by utilizing the Mori-Zwanzig formalism, we can effectively project all these components onto the space of methylation and accessibility, as discussed in Section 1.3.2. The cost of employing this formalism is the emergence of memory effects in the system. As we see later in this section, the memory effect significantly influences the observed changes in the spatial correlations of the accessibilities. With this in mind, and considering methylation, $m(x, t)$, as an orthogonal fast-changing variable compared to accessibility (see Sec. 1.3.2 for more detail), we can express the generalized Langevin equation for the accessibilities as,

$$\frac{\delta\varphi(x, t)}{\delta t} = - \int_0^t d\tau K_\Gamma(t - \tau) \frac{\delta F}{\delta\varphi} + m(x, t). \quad (3.4)$$

in which we substitute F from Eq. 3.3,

$$\frac{\delta\varphi(x, t)}{\delta t} = - \int_0^t d\tau K_\Gamma(t - \tau) \left((\nabla^2 + a_0^2)\varphi(x, \tau) + \frac{\lambda}{3!} \varphi(x, \tau)^3 \right) + m(x, t). \quad (3.5)$$

Here, K_Γ is a memory kernel that remains stationary during each attempt of the sequencing experiment (regeneration state) since the level of injury is assumed to be constant at each sequencing experiment. However, throughout the entire duration of the regeneration process, that is, from one sequencing experiment to another, the level of the injury itself changes, as shown in Fig. 3.2, which we highlight using Γ . This Γ distinguishes between different regeneration states, and we anticipate that it is in the same order as the inverse time scale of the experiment [124, 125]; such that in our case study, $\Gamma \rightarrow \infty$ represents the pre-injury state. Therefore, we consider the following ansatz for the memory kernel,

$$K_\Gamma(t) = \Gamma^2 e^{-\Gamma^2 t}. \quad (3.6)$$

The statistical behavior of the fast variables (methylation), $m(x, t)$, in Eq. 3.4, can be described as,

$$\langle m(x, t)m(x', t') \rangle = K_\Gamma(t - t')C(x - x'). \quad (3.7)$$

and also,

$$\langle m(x_1, t_1) \dots m(x_{2n}, t_{2n}) \rangle = \sum_{\substack{\text{possible pair} \\ \text{combinations}}} \langle m(x_1, t_1)m(x_2, t_2) \rangle \langle m(x_n, t_n)m(x_l, t_l) \rangle \dots \quad (3.8)$$

Eq. 3.7 represents the fluctuation-dissipation relation of the second kind [126]. It establishes a connection between the memory kernel in the Generalized Langevin Equation and the fast variable. In the same equation, $C(x - x')$ is used to describe the spatial correlation for methylation. Based on Fig. 3.6, we learn that $C(x - x')$ is an exponentially decaying function with distance. To solve Eq. 3.5, we use a perturbative series in λ [124, 127],

$$\varphi(x, t) = \varphi^{(0)}(x, t) + \lambda \varphi^{(1)}(x, t) + \lambda^2 \varphi^{(2)}(x, t) + \dots \quad (3.9)$$

Substituting the Eq. 3.9 into Eq. 3.5, and after equating terms with equal powers in λ , we get,

$$\begin{aligned} \left[\frac{\delta}{\delta t} + (\nabla^2 + a_0^2) \right] \varphi^{(0)}(x, t) &= m(x, t), \\ \left[\frac{\delta}{\delta t} + (\nabla^2 + a_0^2) \right] \varphi^{(1)}(x, t) &= -\frac{1}{3!} \int_0^t d\tau K_\Gamma(t - \tau) \left(\varphi^{(0)}(x, \tau) \right)^3, \\ \left[\frac{\delta}{\delta t} + (\nabla^2 + a_0^2) \right] \varphi^{(2)}(x, t) &= -\frac{1}{2!} \int_0^t d\tau K_\Gamma(t - \tau) \left(\varphi^{(0)}(x, \tau) \right)^2 \varphi^{(1)}(x, \tau), \end{aligned} \quad (3.10)$$

and so on. Using the retarded Green function technique [122, 124, 127], one can write the solutions of the above equations in a Fourier representation as,

$$\begin{aligned}\varphi^{(0)}(k, t) &= \int_0^t dt' \int d^3 k' G(k - k', t - t') m(k', t'), \\ \varphi^{(1)}(k, t) &= -\frac{1}{3!} \int_0^t dt_1 \int d^3 k_1 G(k - k_1, t - t_1) \int_0^{t_1} d\tau K_\Gamma(t_1 - \tau) \\ &\quad \left(\int_0^{t_1} dt' \int d^3 k' G(k_1 - k', t_1 - t') m(k', t') \right)^3.\end{aligned}\quad (3.11)$$

where the Green function is $G(k, t) = e^{-t(k^2 + a_0^2)}$. Note in Eq. 3.11 the initial condition $\varphi(k, 0) = 0$ has been used to simplify the solution, which may seem to differ from the initial accessibilities from the experiment. However, any trace of an arbitrary initial condition should die out with $t \rightarrow \infty$, the time scale we study the chromatin structure at the end of each sequencing experiment. Using Eq.3.7, Eq. 3.10 and Eq. 3.11, the zeroth order of the connected correlation, $\langle \varphi(x, t) \varphi(x', t) \rangle^{(0)} \equiv \langle \varphi \varphi \rangle^{(0)} \equiv D(x - x'; t, t)$, can be computed in a Fourier space as,

$$\begin{aligned}D(k; t, t') &= C(k) \int_0^t d\tau \int_0^{t'} d\tau' G(k; t - \tau) G(k; t' - \tau') K_\Gamma(\tau - \tau') \\ &= C(k) \int_0^t d\tau \int_0^{t'} d\tau' \exp\left(-(\tau - \tau')(k^2 + a_0^2)\right) K_\Gamma(\tau - \tau').\end{aligned}\quad (3.12)$$

At the limit $t = t' \rightarrow \infty$, we have $D(k; t, t') = \frac{C(k)}{k^2 + a_0^2}$. To analyze the correlation's dependence on memory, one should consider the first correction to the correlations. However, the analytical solution for this correction, obtained from Eq. 3.11, is very complex. To simplify the procedure, we employ the Feynman diagram, which can serve as a graphical representation for computing higher-order corrections. Using Feynman diagrams, we represent the Green function, G , with a line and methylation, m , with a cross graphically. In this formalism, Eq. 3.9 can be expressed as,

$$\varphi = \text{---}\times + \text{---}\begin{array}{c} \nearrow + \\ \times \\ \searrow + \end{array} + \text{---}\begin{array}{c} \nearrow \times \\ \times \\ \searrow + \end{array} + \dots$$

To compute the correlations using this method, it is important to take into account that when the average is taken over the fast variables (m), all combinations of crosses are combined together in all possible ways as a result of the Wick-decomposition property, as described in Eq. 3.8. Therefore, the two-point correlation function, up to the first correction (one-loop level), can be expressed as follows,

$$\langle \varphi \varphi \rangle = \text{---} \times \text{---} + \text{---} \circ \times \text{---} + \text{---} \times \circ \text{---}$$

The first diagram here gives the zeroth order of the correlations as we calculated in Eq. 3.12, and with the combination of the last two terms, the right-hand side of the above diagram can be written as,

$$\begin{aligned} RHS = \langle \varphi \varphi \rangle^{(0)} &- \frac{\lambda}{2} \int d^3k \int_0^{t_1} dt \int_0^t d\tau \quad C(k)G(k_1; t_1 - t)D(k; \tau, \tau)D(k_2; t_2, \tau)K_\Gamma(t - \tau) \\ &- \frac{\lambda}{2} \int d^3k \int_0^{t_2} dt \int_0^t d\tau \quad C(k)G(k_2; t_2 - t)D(k; \tau, \tau)D(k_1; t_1, \tau)K_\Gamma(t - \tau) \end{aligned} \quad (3.13)$$

Substituting the Green function, Eq. 3.6, and Eq. 3.12 in the Eq. 3.13, one obtains the final result in the limit $t_1 = t_2 \rightarrow \infty$ as,

$$RHS = \langle \varphi \varphi \rangle^{(0)} - \lambda Q(\Gamma) \left(\frac{1}{k_2^2 - a_0^2} + \frac{1}{k_1^2 - a_0^2} \right) \int d^3k \frac{C(k)}{k^2 + a_0^2} \quad (3.14)$$

where $Q(\Gamma)$ is a complex function¹ that represents the dependencies of the connected correlations to the memory. Based on this outcome and referring back to the sequencing experiment, it is clear that the connected correlation for accessibilities varies from the connected correlation for the pre-injury state when $\Gamma \rightarrow \infty$. The variation between connected correlations at different states of the regeneration is also because of the presence of different Γ s in different sequences. In other words, when dealing with numerous degrees of freedom projected onto a two-dimensional space, alterations in temporal correlations will impact spatial correlations, thereby demonstrating the influence of memory within the system. Note that our primary focus here was to present a formal framework to explain the variations in the connected correlation of accessibilities in an injured system without relying on data fitting. To fully comprehend the trends in the alteration of accessibilities correlation (Fig. 3.6), which directly reflects the structure of the chromatin, it is necessary to combine our analysis with subsequent experiments that specifically investigate the topological structure of the chromatin, such as Hi-C experiments. With that, we summarise our investigations in this chapter and discuss potential future studies related to epigenetic changes during liver regeneration in the next section.

¹

$$\begin{aligned} Q(\Gamma) &\equiv \left(\frac{((72\Gamma^4 + 9\beta_1^2 - \beta_2^2)\Gamma^4 + \beta_2^2(\beta_1^2 - \beta_2^2))}{(4\Gamma^4 + (\beta_1 - \beta_2)^2)(4\Gamma^4 + (\beta_1 + \beta_2)^2)} \right), \\ \beta_1 &= \Gamma \sqrt{2(k_1^2 + a_0^2) - \Gamma^2}, \\ \beta_2 &= \Gamma \sqrt{2(k_2^2 + a_0^2) - \Gamma^2}. \end{aligned} \quad (3.15)$$

3.4 Summary and discussion

In this chapter, we have explored the underlying collective epigenetic mechanism responsible for the liver's cellular reprogramming during the regeneration process, specifically by studying the connection between chromatin structures and past injuries. Our research has revealed that the interplay between methylation modification and accessibilities of chromatin is strongly dynamic after facing an injury and during the regeneration process. By considering the various factors that may impact these changes in the interplay, we effectively projected their influence on fast methylation variables and relatively slow chromatin structure. This led to the emergence of memory within the system. By representing chromatin behavior with a coarse-grained free energy and using perturbation analysis, we have explored the connected correlations that support our experimental findings. We have observed that the chromatin structure varies at different stages of regeneration and does not return to the same structure as before the injury. To further investigate trends in the topological changes of chromatin, we recommend integrating our studies with subsequent experiments that directly examine chromatin structures, such as Hi-C maps. Additionally, inspired by the heterogeneity among single cells discussed in Chapter 2, we propose exploring the possibility of distinct inherent repair predispositions among different subpopulations of liver cells as a future direction for this research.

Chapter 4

Conclusion

“You cannot undo what has been done,” the tree replied. “But if you truly desire to embrace life, then live it as if every day was your last, loving deeply and unre-servedly, and cherishing each moment as it passes.”

– Rumi, *Masnavi*

Chemical modifications of the DNA and chromatin, collectively termed epigenetics, are key players in regulating the identity and behavior of cells. In this thesis, by applying statistical physics tools, we have explored the emergent collective processes following the changes in epigenetic modifications at different time scales. In particular, we studied the mechanistic basis of temporal changes in DNA methylation, a chemical modification of the DNA, and their crosstalk with other epigenetic modifiers during biological ageing (Chapter 2) and regeneration after injury (Chapter 3).

In Chapter 2, we highlighted the strong correlation between these modifications and an individual’s biological age, and we demonstrated that machine learning models can accurately predict biological age by monitoring alterations in individual modifications. Nevertheless, it is important to note that these machine learning models, while effective in predicting age, typically rely on extensive sets of variables and are primarily descriptive in nature. Furthermore, the underlying mechanisms driving these methylation-based age predictors remain poorly understood. In order to understand the mechanistic basis underlying DNA methylation during ageing, we shifted our focus from tracking the changes of individual methylation marks to their collective behavior. More precisely, we assessed the spatial correlation function among methylation marks and tracked the systematic alterations observed in these correlations over time. In particular, we computed the ratios of correlations at specific distances and successive time points, which measure changes in correlation beyond uncorrelated noise. Our analysis revealed a progressive increase in these changes over time, with a distinct pattern emerging at an approximate distance of 500 base pairs. We also found that the augmentation of correlations at the 500 base pairs is primarily characteristic of the dynamic at the boundary of regions rich in CpG sites, commonly referred to as CpG islands.

Our analysis also unveiled an age-dependent increase in the size of low methylated regions that are mainly linked to CpG islands. Concurrently, the methylation levels within these islands displayed an upward trend. To understand the origins of this phenomenon, we delved into the interplay between DNA methylation and histone modifications, which are responsible for DNA compaction. Our investigations pinpointed that the aging indicators identified through correlation ratios are specifically associated with CpG islands bearing histone modifiers related to both gene activation (H3K4me3) and suppression (H3K27me3), known as bivalent CpG islands. Taking into account the reduction in H3K27me3 concentration resulting from the depletion of polycomb enzymes crucial for its establishment, we developed an asymmetric interactive two-component phase field model that can reconstruct DNA methylation profiles over time. To this end, we started with a microscope system description and systematically coarse-grained it. While our model adheres to the necessary criteria and addresses critical observations from the data, additional experiments are imperative to fine-tune the model parameters. Specifically, we propose future investigations into how variations in the concentrations of PRC2, H3K27me3, and DNA methylation interact during the aging process.

Furthermore, in collaboration with Ferdinand von Meyenn's group at ETH Zürich, Wolf Reik's group at the Babraham Institute and AltosLabs, and Oliver Stegle's division at DKFZ and EMBL Heidelberg, we embarked on the exploration of aging dynamics at the single-cell level. We encountered substantial cell-to-cell heterogeneity within a cohort of cells of the same age, making predicting biological age challenging. Nevertheless, we tried to minimize the potential technical biases affecting the study's results.

In addition to the slow epigenetic dynamics during ageing, epigenetic marks are modifying rapidly upon active and passive processes during reprogramming after injury. In Chapter 3, we studied the underlying collective epigenetic mechanism responsible for the liver's cellular reprogramming during regeneration. For this purpose, we drew on a single-cell experiment conducted by the group of Meritxell Huch at MPI-CBG for different phases of liver regeneration, including pre-injury, proliferation (response to the toxin diet/injury), and recovery and termination.

Our research has unveiled the profound impact of the regeneration process on the intricate interplay between methylation modifications and chromatin accessibility. To comprehend the multifaceted factors influencing these dynamics, we systematically assessed their effects on the rapid methylation variable and the comparatively slower-changing chromatin structure. This analysis revealed the emergence of memory within the system, where past influences persisted.

In our investigation of chromatin behavior through a coarse-grained free energy representation and perturbation analysis, we explored the interconnected correlations that underpin our experimental findings. Notably, we observed that the chromatin structure undergoes variations at different stages of the regeneration process and does not revert to its pre-injury

configuration.

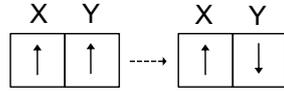
To delve deeper into the trends regarding the topological changes in chromatin, we suggest integrating our studies with subsequent experiments that directly examine chromatin structures, such as Hi-C maps. Furthermore, taking inspiration from the discussions on single-cell heterogeneity in Chapter 2, we propose further exploration of the potential existence of distinct inherent repair predispositions among different subpopulations of liver cells.

Appendices

Appendix A

Multiplicative changes in correlation functions

We consider DNA as a linear arrangement with spins placed at each location, representing the methylation level. To clarify, if a spin points upwards, it indicates a methylated site, whereas a downward spin corresponds to an unmethylated site. As we are interested in uncorrelated changes, we will consider a scenario where our lattice consists of only two sites, and during the process of ageing, one spin will be flipped following the Bernoulli trial, as depicted below,



To investigate how the correlation function will be modified, we use expectation value or mean representation of an arbitrary physical quantity $O(\vec{\sigma})$, which can be given by,

$$\langle O(\vec{\sigma}, t) \rangle = \sum_{\vec{\sigma}} P(\vec{\sigma}, t) O(\vec{\sigma}). \quad (\text{A.1})$$

where $P(\vec{\sigma}, t)$ is the probability for a certain configuration $\vec{\sigma} = (\sigma_1, \sigma_2, \dots, \sigma_N)$, $\sigma_i = \uparrow, \downarrow$, at time t . Starting from a system that is in an initial configuration (\uparrow, \uparrow) with probability $P(\vec{\sigma}, t) \equiv P$, we modify the system following the Bernoulli trial till we have the first downward spin, and we will define it as the new age of the system, $t' = k$ (for $k = \mathbb{N} = \{1, 2, 3, \dots\}$). Hence, the probability of finding the system within the new configuration is $P(\vec{\sigma}', t') = (1 - P)^{k-1} P$.

The average methylation status at a given site in our introduced system at the new age t' can be written as,

$$\langle m'_Y \rangle = (1 - P)^{k-1} P m_Y = (1 - P)^{k-1} \langle m_Y \rangle. \quad (\text{A.2})$$

Using Eq. A.1 and Eq. A.2, the two-point correlation function at a new age and a given

distance $C(d, t')$ can be written as,

$$C(d, t') = (1 - P)^{k-1} \langle m_X m_Y \rangle - (1 - P)^{k-1} \langle m_X \rangle \langle m_Y \rangle = (1 - P)^{k-1} C(d, t). \quad (\text{A.3})$$

That is, the uncorrelated changes in the correlation function come from multiplicative changes.

Appendix B

From the Master equation to the Fokker-Planck equation

The master equation is an integro-differential equation, which is typically challenging to solve. As a result, it is commonly approximated as a partial differential equation. This approximation is accomplished through the Kramers-Moyal expansion technique. To begin with, consider the master equation as,

$$\frac{\partial P(\sigma, t)}{\partial t} = \int_{-\infty}^{\infty} (\omega(\sigma', \sigma)P(\sigma', t) - \omega(\sigma, \sigma')P(\sigma, t)) d\sigma'. \quad (\text{B.1})$$

We modify the parameterization of the integral to incorporate an initial state and a variation from that initial state, $r \equiv \sigma - \sigma'$,

$$\frac{\partial P(\sigma, t)}{\partial t} = \int_{-\infty}^{\infty} (\omega(r, \sigma - r)P(\sigma - r, t) - \omega(-r, \sigma)P(\sigma, t)) dr. \quad (\text{B.2})$$

The idea of the Kramers-Moyal expansion is now to Taylor expand the first term in the integral of this equation around σ , and we get,

$$\begin{aligned} \frac{\partial P(\sigma, t)}{\partial t} &= \int_{-\infty}^{\infty} \left(\omega(r, \sigma)P(\sigma, t) + \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} r^n \frac{\partial^n}{\partial \sigma^n} (\omega(r, \sigma)P(\sigma, t)) - \omega(-r, \sigma)P(\sigma, t) \right) dr \\ &= \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \frac{\partial^n}{\partial \sigma^n} a_n(\sigma)P(\sigma, t). \end{aligned} \quad (\text{B.3})$$

Where we define, $a_n(\sigma) \equiv \int_{-\infty}^{\infty} r^n \omega(r, \sigma) dr \equiv \int_{-\infty}^{\infty} (\sigma - \sigma')^n \omega(\sigma, \sigma') d\sigma'$.

The Fokker-Planck equation is the special case of the Eq. B.3 when $n = 2$. According to the Puwala theorem, the Eq. B.3 will either converge at the first term or the second term. If the expansion continues beyond the second term, it is necessary for the equation's solution to have an infinite number of terms in order to interpret it as a probability density function [128].

Appendix C

Preliminary analysis of sequencing experiments

Analysis of raw sequence reads was accomplished identically to [21]. Beginning with the raw FastQ files generated by an Illumina sequencer, an initial quality control step is crucial. This is vital because poor-quality calls potential adapter contamination can have adverse effects, potentially leading to reduced mapping efficiency or, if mapped, introducing errors in methylation calls and misalignment. These errors could make the base call accuracy approach 50% akin to a completely random outcome. To remove these undesirable reads, we use Trim Galore (Version 0.6.5, Babraham Inst.) with the default parameters for two genome strands of a cell in a single-end mode. With this, we automatically remove 1) base calls with a Phred score (a measure of the expected error rate per base pair) of 20 or lower, 2) short sequences after trimming (< 20 bp), and 3) unmethylated cytosine residues that come from adaptor contamination. Fig. C.1 displays an example of Trim Galore's output. The trimmed reads were aligned to the mouse assembly GRCm39 using Bismark [36] (v0.22.3, parameters: `- non_directional`). Afterward, reads underwent deduplication using `deduplicate_bismark`, which randomly chose an alignment for positions covered more than once. Following deduplication, methylation calls were extracted from the mapping output using the Bismark methylation extractor (v0.14.4) with the specified parameters: `-bedGraph -buffer 10G -parallel 4 -CX`. Lastly, methylation calls need to be distinguished for the actual methylated CpG sites, and the ones that are labeled for GpC sites represent the accessible sites along the chromatin. For this purpose, we use `Bismark coverage2cytosine` (v0.12.1, parameters: `- nome`). Further statistical analysis, such as secondary quality control and average DNase and accessibility levels, was performed by custom pipeline in R.

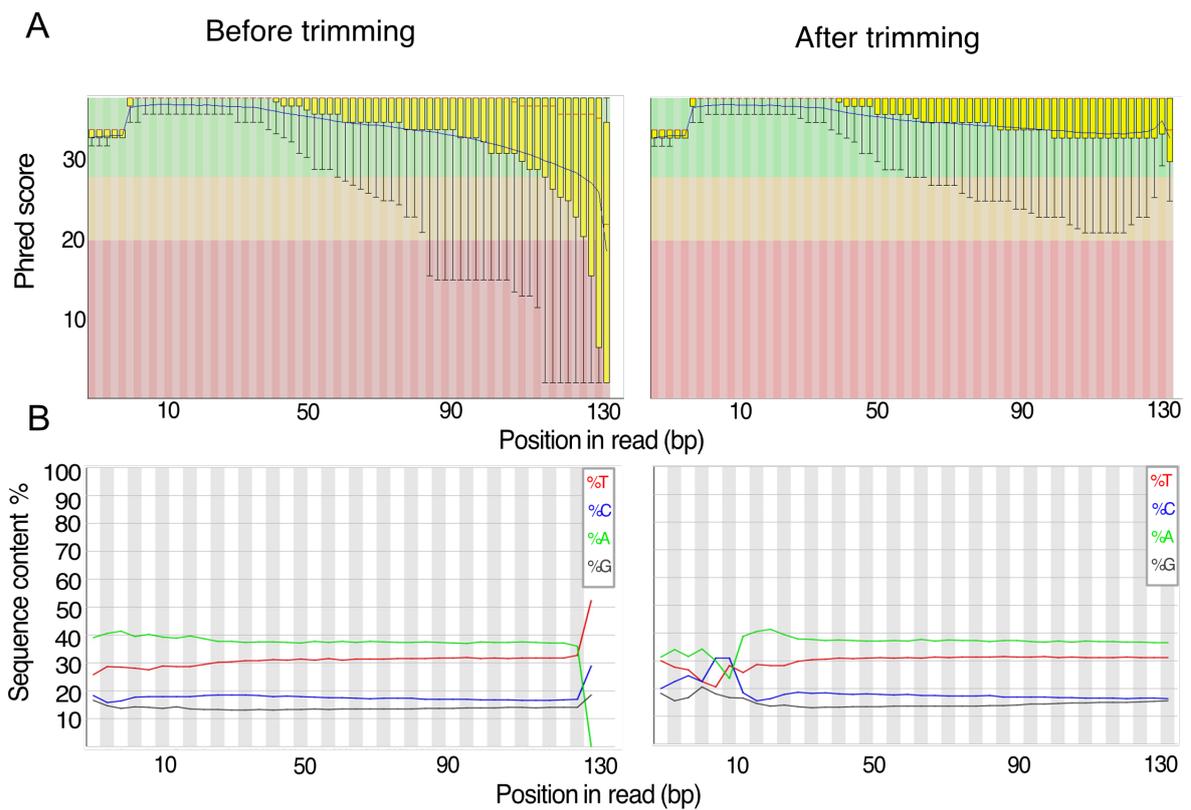


Figure C.1: An example of raw data trimming for a single cell from the scNMT-seq experiment introduced in section 3.2. Up: low-quality base calls (indicated by the red zone) are trimmed off from the end of reads. Bottom: An increase in cytosine and thymine content toward the end of the reads, attributed to adaptor contamination. This contamination is subsequently eliminated through the trimming process.

List of Figures

1.1	Illustration of two major epigenetic mechanisms (i) DNA methylation occurs when methyl groups tag DNA and repress (through the effect of DNMT3) or activate (through the effect of TET) genes. Removing the effect of methylation (demethylation) happens in different steps, which have been shown in the cycle at the right. (ii) Nucleosome or histone modification occurs when the binding of epigenetic factors (methylation, ubiquitination, phosphorylation, acetylation) to histone tails alters the extent to which DNA is wrapped around histones and the availability of genes in the DNA to be activated.	4
1.2	Illustration of a bivalent CpG island. H3K4me3 usually affects the center of the CpG islands, while H3K27me3 usually affects the shores in the case of bivalent CGI.	5
1.3	Sketch of the bisulfite sequencing experiments, adapted by permission from [38], (License Number 5659270261727).	7
1.4	Sketch of ChIP-sequencing experiments, adapted from [39]	8
1.5	Schematic visualization of the projectors (blue arrows) in the space of the observables.	10
2.1	The scheme enumerates the hallmarks of ageing, adapted from [52](with licence CC-BY-ND).	14
2.2	Relation of DNA Methylation and Chromatin Marks. (A) Average density profiles for H3K4me3 and H3K27me3 around TSSs. (B) Heatmaps showing the signal difference of H3K27me3 and H3K4me3 between KO and control cells surrounding TSSs analyzed in seven clusters based on their different binding properties. Green, decreased; yellow, increased signal intensity, compared to WT control. (C) Aggregating the density of DNA methylation along with the density of other chromatin marks within annotated CpG islands (CGIs). Figures A, B are adapted from [71], and C is adapted from [72].	16
2.3	Schematic describes the interplay between different epigenetic modifiers at bivalent CpG islands (CGI). H3K4me3 affects the island's center, while H3k27me3 affects the shores, and both repress the methyltransferase (DNMT3). PRC2, in interaction with H3K27me3, regulates the shores.	17

2.4	Illustration of DNA methylation ageing clocks: (A) Depicts the age-related methylation trajectory of eight individual CpGs, divided into four with negative coefficients (blue) and four with positive coefficients (red). The colors reflect the rate of change with age, with darker shades indicating faster rates and stronger weights in the epigenetic clock. (B) Displays the methylation values of these eight clock CpGs (columns) in three individuals of different chronological ages (rows). The shading of the filled circles represents the fractional methylation (ranging from 0.0 to 1.0). The box color signifies the coefficient of change over age (slope of the line, top). In each colored box, the numerical product of the methylation value and its corresponding coefficient is presented. The summation of these products, combined with the intercept learned during the construction of the clock, results in an epigenetic age approximation that corresponds closely with chronological age. Figures are adapted from [82].	19
2.5	Universal clock: Every point on the graph corresponds to a tissue sample identified by the mammalian species index. For example, 1 denoting primates, 2 indicating elephants, 3 representing cetaceans, and so on. Figure is adapted from [83].	20
2.6	Illustration of the data from Petkovich et al., we used for this study. In the table, we show typical information one gets from sequencing experiments. . .	22
2.7	Methylation density during a 32-month time-course sequencing experiment for chromosome 1 and on a broad genomic region. For visualization, the genome is re-scaled into bins with 100,000 base pairs.	23
2.8	Connected two-point correlation functions across the genome between DNA methylation marks at different ages for mouse blood cells, which shows a clear ageing signal (Data from Petkovich et.al).	24
2.9	Characteristic behavior of genome-wide multiplicative changes in connected two-point correlation functions as quantified in correlation ratios, Eq.2.4 (Data from Petkovich, et.al). There is a striking increase in the correlations around 500 bps.	25
2.10	(A) CpG islands have a higher concentration of CpGs along the genome, and the average length of this regain is 1000 bps. (B) The increase in correlations observed at a 500bp distance is unique to regions characterized by high CpG density, known as CpG islands (CGIs). The distance of 500bp can be understood as the average distance between sites within a CpG island and its boundary. Correlations decrease over time in regions with low CpG density (CpG-poor regions), but genomic proximity does not influence this decay. . .	25

2.11	Position of the peak in multiplicative changes of correlations scales with the size of CpG islands, which indicates that an increase in correlations at a distance of 500bp is associated with DNA methylation dynamics in the boundaries of CpG islands. The position of this peak aligns with the half-length scale of each group of CpG islands with different lengths (500, 1000, 1500, 2000), as indicated by the vertical line for better observation.	26
2.12	Generic changes in DNA methylation correlations across tissue. A clear trend of ageing and a characteristic gain of correlation (bump) around 500 bps is evident (indicated by the vertical lines). Data is collected from Stubbs et al. [86], except for blood from Petkovich et al.	27
2.13	Correlation ratios for three groups of CpG islands based on the enrichment of H3k4me3 and H3k27me3. Ageing signals are specific to bivalent CGIs, where one can find both of these modifiers together.	28
2.14	During ageing, expansion of the lowly methylated region goes along with an increase of methylation in this region.	29
2.15	Microscopic model that accounts for the exchange dynamics and interactions between different histone modifiers. These interactions are long-ranged due to the chromatin structure in three dimensional space.	29
2.16	(A) Free energy density for $h = 1$. Changing (increasing) interaction χ leads to phase separation from a convex free energy (one well) to a non-convex (double well) structure, implying phase separation between two histone modifiers. (B) Free energy density for $\chi = 2.5$ and different h that resembles a decrease in the binding affinity of PRC2 and concentration. In all analyses, we keep the temperature as the temperature for mammalian cell lines, $37^\circ C$. (C) Phase diagram: The solid line shows the binomial criteria where complete phase separation appears in the system, while the dashed line represents the spinodal decomposition.	35
2.17	Predictions of the phase field model (Eq.2.17). (A) occupancy profile of repressive histone modifiers that resembles methylation values in CpG islands during ageing. (B) Correlation ratios for the simulated profiles show the striking patterns we expected, considering the average length of simulated CGIs. (C) Gain of methylation and expansion of the low methylated regions align with other observations from the Petkovich dataset. Different colors are associated with varying ages of simulation.	37

2.18	Schematic representation of the distinction between single-cell and bulk methylation sequencing outputs. With bulk approaches (right), read coverage is high and consistent between samples. In single cells (left), read coverage is low (often 1) and inconsistent between single cells, resulting in limited, distinct methylome profiles. Figure is adapted with permission from [84] (License Number 5659301508019).	39
2.19	(A) Illustration of the single-cell sequencing data set collected for this study. (B) Cell type composition overview stratified by chronological age [105] (Courtesy of Marc Jan Bonder).	39
2.20	(A) Genome-wide connected two-point correlation functions between DNA methylation marks at different ages for single-cell mouse blood cells with two-cell lineages. (B) Genome-wide correlation ratios, considering the first age of the experiment as the reference. Each line corresponds to a single cell. Strong heterogeneity dominates age differences across cells. (C) Increase of correlation variance over age.	40
2.21	Correlation functions in pseudo bulk single-cell data show systematic changes during ageing when cells of the same age are grouped together.	41
2.22	Distance contribution of different genomic regions in multiplicative changes of correlation function beyond the correlated noise. Blue shades correspond to loss of correlation, and red corresponds to correlation gain.	42
2.23	Low dimensional representations (UMAP) of single-cell correlation functions across genomic annotations.	43
2.24	Low dimensional representations (UMAP) of single-cell correlation functions across genomic annotations. There is no spectacular technical bias in individual annotations coming from coverage differences.	44
2.25	(A-B) Predicted DNAm age of pseudo bulked (A) blood and (B) liver single-cell data. C-D) Predicted DNAm of the single cell (C) blood and (D) liver data [105] (Courtesy of Marc Jan Bonder).	45
3.1	Some epigenetic modifications that allow cell-fate changes into bipotent liver progenitors. Upon injury, both adult (Quiescent) hepatocytes and cholangiocytes can de-differentiate into bipotent liver progenitors, giving rise to both liver epithelial cell types. These modifications include increased chromatin accessibility mediated by Arid1a and the chromatin remodeling complex SWI/SNF [113] and oxidation of 5-methylcytosine (5mC) into 5-hydro methylcytosine (5hmC) mediated by the methylcytosine dioxygenase TET1 [21]. Such permissive states facilitate the chromatin binding of the transcriptional machinery (e.g., YAP/TEAD) that promotes the establishment of liver progenitor identity.	49

3.2	Experimental design, courtesy of German Belenguer at MPICBG, for studying the regeneration of ductal cells. Genomic DNA from undamaged or DDC-damaged livers undergoes a scNMT sequencing experiment at each time point.	51
3.3	Quality control of the bisulfite sequencing component of the scNMT-seq experiment. We retained cells meeting specific criteria, including a total read count exceeding 100000, mapping efficiency surpassing 5%, and CHH methylation levels below 15%. The dashed red lines serve as visual references denoting these criteria. Various colors are used to denote different batches within the experiment.	51
3.4	Global DNA methylation levels and global chromatin structure are retained during regeneration: The mean of methylation and accessibilities are depicted for different experiment batches. Plots are split into two sets of red and gray colors associated with various stages of the regeneration experience during and after the DDC diet, respectively.	52
3.5	Connected cross-correlation between DNA methylation and accessibility for the whole genome as well as different local features along the genome. The inverse relationship between DNA methylation and accessibility remains similar for different genomic features. However, following liver injury caused by DDC poisoning, the association between methylation and accessibility is disrupted. Although it gradually returns to initial levels during the recovery period, it never fully recovers to the same level as day 0 throughout the duration of the experiment. Dashed lines are exponential fits to the data using R's nls function.	53
3.6	Autocorrelation of (A) accessible cytosines and (B) methylated cytosines along the genome. Both plots show decaying behavior; there are no changes detected between the different time points of the regeneration process for the methylation; however, the structure of chromatin is disturbed during regeneration due to the autocorrelation of accessibilities. Although no clear trends were detected here, it seems that the structure of the chromatin will first be looser/less compact after sensing the poison and starting the regeneration, and it will slowly compact and relax to the initial state at late time points of the experiment. The inset of autocorrelation for accesibilities shows the same measurement only with lower bins. Both autocorrelations for accessibility and methylation have been normalized such that the area underlying the curve, which in the case of correlation functions is dependent on sequencing coverage, is equal to 1.	54

3.7	Schematic representation of liquid-liquid phase separation in chromatin. The purpose of the blue droplet shapes is to visually indicate the two main compartments of chromatin: the more compacted and less compacted regions. This organization is not static but rather adaptable to environmental changes, such as injury. For example, based on our findings in Fig. 3.6, the chromatin tends to become looser/less compact during the regeneration process. However, whether it returns to its pre-injury structure is uncertain.	55
C.1	An example of raw data trimming for a single cell from the scNMT-seq experiment introduced in section3.2. Up: low-quality base calls (indicated by the red zone) are trimmed off from the end of reads. Bottom: An increase in cytosine and thymine content toward the end of the reads, attributed to adaptor contamination. This contamination is subsequently eliminated through the trimming process.	69

Bibliography

- [1] F. Olmeda. Non equilibrium physics of single-cell genomics (2022)
- [2] J. Viña, C. Borrás, and J. Miquel. Theories of ageing. *IUBMB life*, **59**(4-5): 249–254 (2007)
- [3] S. E. C. Bafei and C. Shen. Biomarkers selection and mathematical modeling in biological age estimation. *npj Aging*, **9**(1): 13 (2023)
- [4] B. B. Misra. The chemical exposome of human aging. *Frontiers in Genetics*, **11**: 574936 (2020)
- [5] D. Stauffer. Introduction to statistical physics outside physics. *Physica A: Statistical Mechanics and Its Applications*, **336**(1-2): 1–5 (2004)
- [6] J. C. Xavier, K. R. Patil, and I. Rocha. Systems biology perspectives on minimal and simpler cells. *Microbiology and Molecular Biology Reviews*, **78**(3): 487–509 (2014)
- [7] J. Zhang, Q. Nie, and T. Zhou. Revealing dynamic mechanisms of cell fate decisions from single-cell transcriptomic data. *Frontiers in genetics*, **10**: 1280 (2019)
- [8] E. Jablonka and M. J. Lamb. The changing concept of epigenetics. *Annals of the New York Academy of Sciences*, **981**(1): 82–96 (2002)
- [9] B. Hwang, J. H. Lee, and D. Bang. Single-cell rna sequencing technologies and bioinformatics pipelines. *Experimental & molecular medicine*, **50**(8): 1–14 (2018)
- [10] Scientific report 7/2015 12/2018, max planck institute for the physics of complex systems
- [11] Scientific report 1/2019 9/2021, max planck institute for the physics of complex systems
- [12] European research council, understanding collective mechanisms of cell fate regulation using single-cell genomics
- [13] K. Grace, J. Salvatier, A. Dafoe, B. Zhang, and O. Evans. When will ai exceed human performance? evidence from ai experts. *Journal of Artificial Intelligence Research*, **62**: 729–754 (2018)
- [14] P. C. Bressloff. *Stochastic processes in cell biology*, volume 41. Springer (2014)

- [15] G. Felsenfeld. A brief history of epigenetics. *Cold Spring Harbor perspectives in biology*, **6**(1): a018200 (2014)
- [16] F. Olmeda, T. Lohoff, S. J. Clark, L. Benson, F. Krüger, W. Reik, and S. Rulands. Inference of emergent spatio-temporal processes from single-cell sequencing reveals feedback between de novo dna methylation and chromatin condensates. *bioRxiv*, pages 2020–12 (2021)
- [17] M. Wells and L. Steiner. Epigenetic and transcriptional control of erythropoiesis. *Frontiers in Genetics*, **13**: 805265 (2022)
- [18] R. M. Kohli and Y. Zhang. Tet enzymes, tdg and the dynamics of dna demethylation. *Nature*, **502**(7472): 472–479 (2013)
- [19] P. W. Hill, R. Amouroux, and P. Hajkova. Dna demethylation, tet proteins and 5-hydroxymethylcytosine in epigenetic reprogramming: an emerging complex story. *Genomics*, **104**(5): 324–333 (2014)
- [20] M. A. Hahn, P. E. Szabó, and G. P. Pfeifer. 5-hydroxymethylcytosine: a stable or transient dna modification? *Genomics*, **104**(5): 314–323 (2014)
- [21] L. Aloia, M. A. McKie, G. Vernaz, L. Cordero-Espinoza, N. Aleksieva, J. van den Ameele, F. Antonica, B. Font-Cunill, A. Raven, R. Aiese Cigliano, et al. Epigenetic remodelling licences adult cholangiocytes for organoid formation and liver regeneration. *Nature cell biology*, **21**(11): 1321–1333 (2019)
- [22] M. Kim and J. Costello. Dna methylation: an epigenetic mark of cellular memory. *Experimental & molecular medicine*, **49**(4): e322–e322 (2017)
- [23] S. C. Zheng, M. Widschwendter, and A. E. Teschendorff. Epigenetic drift, epigenetic clocks and cancer risk. *Epigenomics*, **8**(5): 705–719 (2016)
- [24] T. O. Tollefsbol. *Epigenetics of aging*. Springer (2009)
- [25] C. Caruso. *Centenarians: An Example of Positive Biology*. Springer (2019)
- [26] T. O. Tollefsbol. *Handbook of epigenetics: the new molecular and medical genetics*. Academic Press (2017)
- [27] D. Kumar and R. Jothi. Bivalent chromatin protects reversibly repressed genes from irreversible silencing. *bioRxiv*, pages 2020–12 (2020)
- [28] H. Kimura. Histone modifications for human epigenome analysis. *Journal of human genetics*, **58**(7): 439–445 (2013)
- [29] A. Harikumar and E. Meshorer. Chromatin remodeling and bivalent histone modifications in embryonic stem cells. *EMBO reports*, **16**(12): 1609–1619 (2015)

-
- [30] J. P. Thomson, P. J. Skene, J. Selfridge, T. Clouaire, J. Guy, S. Webb, A. R. Kerr, A. Deaton, R. Andrews, K. D. James, et al. CpG islands influence chromatin structure via the cpg-binding protein cfp1. *Nature*, **464**(7291): 1082–1086 (2010)
- [31] M. Ku, R. P. Koche, E. Rheinbay, E. M. Mendenhall, M. Endoh, T. S. Mikkelsen, A. Presser, C. Nusbaum, X. Xie, A. S. Chi, et al. Genomewide analysis of prc1 and prc2 occupancy identifies two classes of bivalent domains. *PLoS genetics*, **4**(10): e1000242 (2008)
- [32] Y. Li and T. O. Tollefsbol. Dna methylation detection: bisulfite genomic sequencing analysis. *Epigenetics Protocols*, pages 11–21 (2011)
- [33] S. J. Clark, R. Argelaguet, C.-A. Kapourani, T. M. Stubbs, H. J. Lee, C. Alda-Catalinas, F. Krueger, G. Sanguinetti, G. Kelsey, J. C. Marioni, et al. scnmt-seq enables joint profiling of chromatin accessibility dna methylation and transcription in single cells. *Nature communications*, **9**(1): 781 (2018)
- [34] enseqlopedia
- [35] H. Li and R. Durbin. Fast and accurate long-read alignment with burrows–wheeler transform. *Bioinformatics*, **26**(5): 589–595 (2010)
- [36] F. Krueger and S. R. Andrews. Bismark: a flexible aligner and methylation caller for bisulfite-seq applications. *bioinformatics*, **27**(11): 1571–1572 (2011)
- [37] Babraham bioinformatics training, methylation course
- [38] B. Owens. The single life. *Nature*, **491**(7422): 27 (2012)
- [39] e biotechnologia.pl
- [40] D. Oh, J. S. Strattan, J. K. Hur, J. Bento, A. E. Urban, G. Song, and J. M. Cherry. Cnn-peaks: Chip-seq peak detection pipeline using convolutional neural networks that imitate human visual inspection. *Scientific reports*, **10**(1): 7933 (2020)
- [41] P. C. Hohenberg and B. I. Halperin. Theory of dynamic critical phenomena. *Reviews of Modern Physics*, **49**(3): 435 (1977)
- [42] R. Qin and H. Bhadeshia. Phase field method. *Materials science and technology*, **26**(7): 803–811 (2010)
- [43] H. Meyer, T. Voigtmann, and T. Schilling. On the dynamics of reaction coordinates in classical, time-dependent, many-body processes. *The Journal of chemical physics*, **150**(17) (2019)

- [44] D. S. Lemons and A. Gythiel. Paul langevin, 1908 paper on the theory of brownian motion, [Sur la théorie du mouvement brownien, *Cr Acad. Sci. (Paris)* 146, 530–533 (1908)]. *American Journal of Physics*, **65**(11): 1079–1081 (1997)
- [45] H. Mori. Transport, collective motion, and brownian motion. *Progress of theoretical physics*, **33**(3): 423–455 (1965)
- [46] R. Zwanzig. Memory effects in irreversible thermodynamics. *Physical Review*, **124**(4): 983 (1961)
- [47] H. Meyer, T. Voigtmann, and T. Schilling. On the non-stationary generalized langevin equation. *The Journal of chemical physics*, **147**(21) (2017)
- [48] H. Meyer, T. Voigtmann, and T. Schilling. On the ” generalized generalized langevin equation”. *arXiv preprint arXiv:1706.00658* (2017)
- [49] R. Takaki, L. Jawerth, M. Popović, and F. Jülicher. Theory of rheology and aging of protein condensates. *arXiv preprint arXiv:2303.18028* (2023)
- [50] T. Kirkpatrick and D. Thirumalai. Colloquium: Random first order transition theory concepts in biology and physics. *Reviews of Modern Physics*, **87**(1): 183 (2015)
- [51] J.-P. Bouchaud. Weak ergodicity breaking and aging in disordered systems. *Journal de Physique I*, **2**(9): 1705–1713 (1992)
- [52] C. López-Otín, M. A. Blasco, L. Partridge, M. Serrano, and G. Kroemer. The hallmarks of aging. *Cell*, **153**(6): 1194–1217 (2013)
- [53] A. I. Yashin and S. M. Jazwinski. *Aging and health: a systems biology perspective*. Karger (2014)
- [54] X. Zhu, Z. Chen, W. Shen, G. Huang, J. M. Sedivy, H. Wang, and Z. Ju. Inflammation, epigenetics, and metabolism converge to cell senescence and ageing: the regulation and intervention. *Signal transduction and targeted therapy*, **6**(1): 245 (2021)
- [55] C. J. Steves, T. D. Spector, and S. H. Jackson. Ageing, genes, environment and epigenetics: what twin studies tell us now, and in the future. *Age and ageing*, **41**(5): 581–586 (2012)
- [56] S. Horvath and K. Raj. Dna methylation-based biomarkers and the epigenetic clock theory of ageing. *Nature Reviews Genetics*, **19**(6): 371–384 (2018)
- [57] A. E. Kane and D. A. Sinclair. Epigenetic changes during aging and their reprogramming potential. *Critical reviews in biochemistry and molecular biology*, **54**(1): 61–83 (2019)

-
- [58] B. Vanyushin, L. Nemirovsky, V. Klimenko, V. Vasiliev, and A. Belozersky. The 5-methylcytosine in dna of rats: tissue and age specificity and the changes induced by hydrocortisone and other agents. *Gerontology*, **19**(3): 138–152 (1973)
- [59] I. Hernando-Herraez, B. Evano, T. Stubbs, P.-H. Commere, M. Jan Bonder, S. Clark, S. Andrews, S. Tajbakhsh, and W. Reik. Ageing affects dna methylation drift and transcriptional cell-to-cell variability in mouse muscle stem cells. *Nature communications*, **10**(1): 4361 (2019)
- [60] M. F. Fraga, E. Ballestar, M. F. Paz, S. Ropero, F. Setien, M. L. Ballestar, D. Heine-Suñer, J. C. Cigudosa, M. Urioste, J. Benitez, et al. Epigenetic differences arise during the lifetime of monozygotic twins. *Proceedings of the National Academy of Sciences*, **102**(30): 10604–10609 (2005)
- [61] K. Seale, S. Horvath, A. Teschendorff, N. Eynon, and S. Voisin. Making sense of the ageing methylome. *Nature Reviews Genetics*, **23**(10): 585–605 (2022)
- [62] S. E. Johnstone, V. N. Gladyshev, M. J. Aryee, and B. E. Bernstein. Epigenetic clocks, aging, and cancer. *Science*, **378**(6626): 1276–1277 (2022)
- [63] G. Zimmer-Bensch and H. Zempel. Dna methylation in genetic and sporadic forms of neurodegeneration: lessons from alzheimer, related tauopathies and genetic tauopathies. *Cells*, **10**(11): 3064 (2021)
- [64] S. Orjuela, H. R. Parker, S. Sajibu, F. Cereatti, M. Sauter, F. Buffoli, M. D. Robinson, and G. Marra. Disentangling tumorigenesis-associated dna methylation changes in colorectal tissues from those associated with ageing. *Epigenetics*, **17**(6): 677–694 (2022)
- [65] A. Tovy, J. M. Reyes, M. C. Gundry, L. Brunetti, H. Lee-Six, M. Petljak, H. J. Park, A. G. Guzman, C. Rosas, A. R. Jeffries, et al. Tissue-biased expansion of dnmt3a-mutant clones in a mosaic individual is associated with conserved epigenetic erosion. *Cell Stem Cell*, **27**(2): 326–335 (2020)
- [66] J. E. Sandoval, Y.-H. Huang, A. Muise, M. A. Goodell, and N. O. Reich. Mutations in the dnmt3a dna methyltransferase in acute myeloid leukemia patients cause both loss and gain of function and differential regulation by protein partners. *Journal of Biological Chemistry*, **294**(13): 4898–4910 (2019)
- [67] D. Sun, M. Luo, M. Jeong, B. Rodriguez, Z. Xia, R. Hannah, H. Wang, T. Le, K. F. Faull, R. Chen, et al. Epigenomic profiling of young and aged hscs reveals concerted changes during aging that reinforce self-renewal. *Cell stem cell*, **14**(5): 673–688 (2014)
- [68] S.-J. Yi and K. Kim. New insights into the role of histone changes in aging. *International Journal of Molecular Sciences*, **21**(21): 8241 (2020)

- [69] A. Meissner, T. S. Mikkelsen, H. Gu, M. Wernig, J. Hanna, A. Sivachenko, X. Zhang, B. E. Bernstein, C. Nusbaum, D. B. Jaffe, et al. Genome-scale dna methylation maps of pluripotent and differentiated cells. *Nature*, **454**(7205): 766–770 (2008)
- [70] R. Weigert, S. Hetzel, N. Bailly, C. Haggerty, I. A. Ilik, P. Y. K. Yung, C. Navarro, A. Bolondi, A. S. Kumar, C. Anania, et al. Dynamic antagonism between key repressive pathways maintains the placental epigenome. *Nature Cell Biology*, **25**(4): 579–591 (2023)
- [71] T. Gu, X. Lin, S. M. Cullen, M. Luo, M. Jeong, M. Estecio, J. Shen, S. Hardikar, D. Sun, J. Su, et al. Dnmt3a and tet1 cooperate to regulate promoter epigenetic landscapes in mouse embryonic stem cells. *Genome biology*, **19**: 1–15 (2018)
- [72] A. Jeltsch, J. Broche, and P. Bashtrykov. Molecular processes connecting dna methylation patterns with dna methyltransferases and histone modifications in mammalian genomes. *Genes*, **9**(11): 566 (2018)
- [73] H. Wu, A. C. D. Alessio, S. Ito, K. Xia, Z. Wang, K. Cui, K. Zhao, Y. Eve Sun, and Y. Zhang. Dual functions of tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature*, **473**(7347): 389–393 (2011)
- [74] P. Heyn, C. V. Logan, A. Fluteau, R. C. Challis, T. Auchynnikava, C.-A. Martin, J. A. Marsh, F. Taglini, F. Kilanowski, D. A. Parry, et al. Gain-of-function dnmt3a mutations cause microcephalic dwarfism and hypermethylation of polycomb-regulated regions. *Nature genetics*, **51**(1): 96–105 (2019)
- [75] N. Yang, J. R. Ocean, D. P. Melters, C. Shi, L. Wang, S. Stransky, M. E. Doyle, C.-Y. Cui, M. Delannoy, J. Fan, et al. A hyper-quiescent chromatin state formed during aging is reversed by regeneration. *Molecular Cell*, **83**(10): 1659–1676 (2023)
- [76] S. Horvath. Dna methylation age of human tissues and cell types. *Genome biology*, **14**(10): 1–20 (2013)
- [77] D. A. Petkovich, D. I. Podolskiy, A. V. Lobanov, S.-G. Lee, R. A. Miller, and V. N. Gladyshev. Using dna methylation profiling to evaluate biological age and longevity interventions. *Cell metabolism*, **25**(4): 954–960 (2017)
- [78] M. V. Meer, D. I. Podolskiy, A. Tyshkovskiy, and V. N. Gladyshev. A whole lifespan mouse multi-tissue dna methylation clock. *Elife*, **7**: e40675 (2018)
- [79] T. R. Robeck, Z. Fei, A. T. Lu, A. Haghani, E. Jourdain, J. A. Zoller, C. Z. Li, K. J. Steinman, S. DiRocco, T. Schmitt, et al. Multi-species and multi-tissue methylation clocks for age estimation in toothed whales and dolphins. *Communications Biology*, **4**(1): 642 (2021)

-
- [80] G. S. Wilkinson, D. M. Adams, A. Haghani, A. T. Lu, J. Zoller, C. E. Breeze, B. D. Arnold, H. C. Ball, G. G. Carter, L. N. Cooper, et al. Dna methylation predicts age and provides insight into exceptional longevity of bats. *Nature communications*, **12**(1): 1615 (2021)
- [81] S. Horvath, A. T. Lu, A. Haghani, J. A. Zoller, C. Z. Li, A. R. Lim, R. T. Brooke, K. Raj, A. Serres-Armero, D. L. Dreger, et al. Dna methylation clocks for dogs and humans. *Proceedings of the National Academy of Sciences*, **119**(21): e2120887119 (2022)
- [82] A. E. Field, N. A. Robertson, T. Wang, A. Havas, T. Ideker, and P. D. Adams. Dna methylation clocks in aging: categories, causes, and consequences. *Molecular cell*, **71**(6): 882–895 (2018)
- [83] A. T. Lu, Z. Fei, A. Haghani, T. R. Robeck, J. A. Zoller, C. Z. Li, R. Lowe, Q. Yan, J. Zhang, H. Vu, et al. Universal dna methylation age across mammalian tissues. *BioRxiv*, pages 2021–01 (2021)
- [84] A. Trapp, C. Kerepesi, and V. N. Gladyshev. Profiling epigenetic age in single cells. *Nature Aging*, **1**(12): 1189–1201 (2021)
- [85] B. H. Chen, R. E. Marioni, E. Colicino, M. J. Peters, C. K. Ward-Caviness, P.-C. Tsai, N. S. Roetker, A. C. Just, E. W. Demerath, W. Guan, et al. Dna methylation-based measures of biological age: meta-analysis predicting time to death. *Aging (Albany NY)*, **8**(9): 1844 (2016)
- [86] T. M. Stubbs, M. J. Bonder, A.-K. Stark, F. Krueger, F. von Meyenn, O. Stegle, and W. Reik. Multi-tissue dna methylation age predictor in mouse. *Genome biology*, **18**(1): 1–14 (2017)
- [87] S. Reynal. *Phase transitions in long-range spin models: the power of generalized ensembles*. Dissertation, Université de Cergy Pontoise (2005)
- [88] K. Khoo and H. Sy. Mean-field renormalization group study of the long-range ising model. *Journal of Physics: Condensed Matter*, **13**(1): 101 (2001)
- [89] J. O. Haerter, C. Lövkvist, I. B. Dodd, and K. Sneppen. Collaboration between cpg sites is needed for stable somatic inheritance of dna methylation states. *Nucleic acids research*, **42**(4): 2235–2244 (2014)
- [90] P. B. Singh, S. N. Belyakin, and P. P. Laktionov. Biology and physics of heterochromatin-like domains/complexes. *Cells*, **9**(8): 1881 (2020)
- [91] B. Tolhuis, I. Muijrrers, E. de Wit, H. Teunissen, W. Talhout, B. van Steensel, and M. van Lohuizen. Genome-wide profiling of prc1 and prc2 polycomb chromatin binding in drosophila melanogaster. *Nature genetics*, **38**(6): 694–699 (2006)

- [92] S. Park, G. W. Kim, S. H. Kwon, and J.-S. Lee. Broad domains of histone h3 lysine 4 trimethylation in transcriptional regulation and disease. *The FEBS journal*, **287**(14): 2891–2902 (2020)
- [93] M. Plapp. Phase-field models. In *Handbook of Crystal Growth*, pages 631–668. Elsevier (2015)
- [94] Q. Bronchart, Y. Le Bouar, and A. Finel. New coarse-grained derivation of a phase field model for precipitation. *Physical review letters*, **100**(1): 015702 (2008)
- [95] J. W. Cahn and J. E. Hilliard. Free energy of a nonuniform system. i. interfacial free energy. *The Journal of chemical physics*, **28**(2): 258–267 (1958)
- [96] C. M. Elliott and Z. Songmu. On the cahn-hilliard equation. *Archive for Rational Mechanics and Analysis*, **96**: 339–357 (1986)
- [97] S. Safran. *Statistical thermodynamics of surfaces, interfaces, and membranes*. CRC Press (2018)
- [98] P. M. Chaikin, T. C. Lubensky, and T. A. Witten. *Principles of condensed matter physics*, volume 10. Cambridge university press Cambridge (1995)
- [99] M. Moqri, A. Cipriano, D. Nachun, T. Murty, G. de Sena Brandine, S. Rasouli, A. Tarkhov, K. A. Aberg, E. van den Oord, W. Zhou, et al. Prc2 clock: a universal epigenetic biomarker of aging and rejuvenation. *BioRxiv*, pages 2022–06 (2022)
- [100] R. Bahar, C. H. Hartmann, K. A. Rodriguez, A. D. Denny, R. A. Busuttil, M. E. Dollé, R. B. Calder, G. B. Chisholm, B. H. Pollock, C. A. Klein, et al. Increased cell-to-cell variation in gene expression in ageing mouse heart. *Nature*, **441**(7096): 1011–1014 (2006)
- [101] C. G. Bell, R. Lowe, P. D. Adams, A. A. Baccarelli, S. Beck, J. T. Bell, B. C. Christensen, V. N. Gladyshev, B. T. Heijmans, S. Horvath, et al. Dna methylation aging clocks: challenges and recommendations. *Genome biology*, **20**: 1–24 (2019)
- [102] A. E. Jaffe and R. A. Irizarry. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome biology*, **15**(2): 1–9 (2014)
- [103] S. Gravina, X. Dong, B. Yu, and J. Vijg. Single-cell genome-wide bisulfite sequencing uncovers extensive heterogeneity in the mouse liver methylome. *Genome biology*, **17**(1): 1–8 (2016)
- [104] C. Luo, C. L. Keown, L. Kurihara, J. Zhou, Y. He, J. Li, R. Castanon, J. Lucero, J. R. Nery, J. P. Sandoval, et al. Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. *Science*, **357**(6351): 600–604 (2017)

-
- [105] M. J. Bonder, S. Clark, F. Krueger, S. Luo, J. Agostinho de Sousa, A. M. Hashtroud, T. M. Stubbs, A.-K. Stark, S. Rulands, O. Stegle, et al. Single cell dna methylation ageing in mouse blood. *bioRxiv*, pages 2023–01 (2023)
- [106] L. Aloia, M. A. McKie, and M. Huch. Cellular plasticity in the adult liver and stomach. *The Journal of Physiology*, **594**(17): 4815–4825 (2016)
- [107] A. S. Alvarado and S. Yamanaka. Rethinking differentiation: stem cells, regeneration, and plasticity. *Cell*, **157**(1): 110–119 (2014)
- [108] S. J. Morrison and A. C. Spradling. Stem cells and niches: mechanisms that promote stem cell maintenance throughout life. *Cell*, **132**(4): 598–611 (2008)
- [109] A. Miyajima, M. Tanaka, and T. Itoh. Stem/progenitor cells in liver development, homeostasis, regeneration, and reprogramming. *Cell stem cell*, **14**(5): 561–574 (2014)
- [110] V. L. Gadd, N. Aleksieva, and S. J. Forbes. Epithelial plasticity during liver injury and regeneration. *Cell Stem Cell*, **27**(4): 557–573 (2020)
- [111] X. Han, Y. Wang, W. Pu, X. Huang, L. Qiu, Y. Li, W. Yu, H. Zhao, X. Liu, L. He, et al. Lineage tracing reveals the bipotency of sox9+ hepatocytes during liver regeneration. *Stem Cell Reports*, **12**(3): 624–638 (2019)
- [112] K. Kaji, V. M. Factor, J. B. Andersen, M. E. Durkin, A. Tomokuni, J. U. Marquardt, M. S. Matter, T. Hoang, E. A. Conner, and S. S. Thorgeirsson. Dnmt1 is a required genomic regulator for murine liver histogenesis and regeneration. *Hepatology*, **64**(2): 582–598 (2016)
- [113] W. Li, L. Yang, Q. He, C. Hu, L. Zhu, X. Ma, X. Ma, S. Bao, L. Li, Y. Chen, et al. A homeostatic arid1a-dependent permissive chromatin state licenses hepatocyte responsiveness to liver-injury-associated yap signaling. *Cell Stem Cell*, **25**(1): 54–68 (2019)
- [114] R. E. Thurman, E. Rynes, R. Humbert, J. Vierstra, M. T. Maurano, E. Haugen, N. C. Sheffield, A. B. Stergachis, H. Wang, B. Vernot, et al. The accessible chromatin landscape of the human genome. *Nature*, **489**(7414): 75–82 (2012)
- [115] K. Maeshima, S. Ide, K. Hibino, and M. Sasai. Liquid-like behavior of chromatin. *Current opinion in genetics & development*, **37**: 36–45 (2016)
- [116] S. Palikyras and A. Papantonis. Modes of phase separation affecting chromatin regulation. *Open biology*, **9**(10): 190167 (2019)
- [117] G. Bajpai, D. A. Pavlov, D. Lorber, T. Volk, and S. Safran. Mesoscale phase separation of chromatin in the nucleus. *Biophysical Journal*, **118**(3): 549a (2020)

- [118] J. C. Hansen, K. Maeshima, and M. J. Hendzel. The solid and liquid states of chromatin. *Epigenetics & Chromatin*, **14**(1): 50 (2021)
- [119] K. Rippe. Liquid–liquid phase separation in chromatin. *Cold Spring Harbor perspectives in biology*, **14**(2): a040683 (2022)
- [120] A. A. Hyman, C. A. Weber, and F. Jülicher. Liquid-liquid phase separation in biology. *Annual review of cell and developmental biology*, **30**: 39–58 (2014)
- [121] P. C. Hohenberg and A. P. Krekhov. An introduction to the ginzburg–landau theory of phase transitions and nonequilibrium patterns. *Physics Reports*, **572**: 1–42 (2015)
- [122] U. C. Täuber. *Critical dynamics: a field theory approach to equilibrium and non-equilibrium scaling behavior*. Cambridge University Press (2014)
- [123] C. Hijón, P. Español, E. Vanden-Eijnden, and R. Delgado-Buscalioni. Mori–zwanzig formalism as a practical computational tool. *Faraday discussions*, **144**: 301–322 (2010)
- [124] G. Menezes and N. Svaiter. Stochastic quantization of $(\lambda\phi^4)$ d scalar theory: Generalized langevin equation with memory kernel. *Physica A: Statistical Mechanics and its Applications*, **374**(2): 617–630 (2007)
- [125] S. H. Lim, J. Wehr, and M. Lewenstein. Homogenization for generalized langevin equations with applications to anomalous diffusion. In *Annales Henri Poincaré*, volume 21, pages 1813–1871. Springer (2020)
- [126] G. Jung and F. Schmid. Fluctuation–dissipation relations far from equilibrium: a case study. *Soft matter*, **17**(26): 6413–6425 (2021)
- [127] G. Parisi, Y. S. Wu, et al. Perturbation theory without gauge fixing. *Sci. sin.*, **24**(4): 483–496 (1981)
- [128] T. D. Frank. *Nonlinear Fokker-Planck equations: fundamentals and applications*. Springer Science & Business Media (2005)

Versicherung

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht. Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt. Die Arbeit wurde in Dresden am Max-Planck-Institut für Physik komplexer Systeme unter der Betreuung von Prof. Steffen Rulands.

Aida Mohammadzadehhashtroud

Dresden, December 2023