Aus dem Biomedizinischen Centrum

der Ludwig-Maximilians-Universität München,

Lehrstuhl Molekularbiologie

Vorstand: Prof. Dr Peter. B Becker

# A novel perspective on the *in vivo* chromatin landscape of *Saccharomyces cerevisiae*

Dissertation

zum Erwerb des Doktorgrades der Naturwissenschaften

an der Medizinischen Fakultät der

Ludwig-Maximilians-Universität München

vorgelegt von

**Daan George Frederik Verhagen**

aus

Venlo, Niederlande

2021

Mit Genehmigung der Medizinischen Fakultät der Universität München

Betreuer: Prof. Dr. Peter B. Becker

Zweitgutachter: Prof. Dr. Andreas Ladurner

Dekan: Prof. Dr. med. Thomas Gudermann

Tag der mündlichen Prüfung: 03.05.2022

LMU
LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

Promotionsbüro
Medizinische Fakultät

MMRS

# Affidavit

Verhagen, Daan George Frederik

Guardinistraße 171
81375, Munich, Germany

I hereby declare, that the submitted thesis entitled:

**A novel perspective on the *in vivo* chromatin landscape of *Saccharomyces cerevisiae***

is my own work. I have only used the sources indicated and have not made unauthorized use of services of a third party. Where the work of others has been quoted or reproduced, the source is always given.
I further declare that the submitted thesis or parts thereof have not been presented as part of an examination degree to any other university.

**Utrecht, 02.07.2022**
Place, date

**Daan George Frederik Verhagen**
Signature doctoral candidate

## TABLE OF CONTENTS

# I. ABSTRACT

Chromatin is the defining feature of the eukaryotic genome and is essential for survival. The basic units of chromatin are abundantly present and are referred to as nucleosomes. These nucleosomes are often described and visualized as 'beads on a string'. However, unlike beads on a string, nucleosomes are not randomly scattered across the chromatin, but are generally well positioned in a stereotypical pattern. This pattern is characterized by a nucleosome free region (NFR), followed by a regular pattern – or array – of nucleosomes which are typically phased relative to a defined genomic location. The organization of nucleosomes allows for controlled gene expression and is therefore regarded as a key regulatory element. A major influence on the position of nucleosomes are ATP dependent chromatin remodelers. These remodelers may work synergistically or antagonistically and often take redundant roles in positioning nucleosomes throughout the chromatin landscape. Mutations or deletions of these remodelers can lead to genome instability, loss of array regularity or diminished phasing, thus increasing heterogeneity in the chromatin landscape. Besides remodelers, other minor factors such as the inherent DNA sequence and general regulatory factors (GRFs) influence the chromatin landscape. Moreover, the chromatin landscape across independent (healthy) cells and genomic locations may be altered or vary significantly as a result of the cell cycle stage or genome replication. Due to the plethora of different factors that can be of influence, it is difficult to dissect how the chromatin landscape is shaped.

In this thesis, we investigate generation and heterogeneity of the chromatin landscape in *S. cerevisiae*. To this end, we initially developed a novel sequencing approach that allows us to probe the location of nucleosomes on a single chromatin fiber. We show that this approach provides several unique insights allowing us to determine cell-to-cell heterogeneity and which factors attribute to array regularity. Moreover, we provide a detailed description of experiment procedures and analyses that result in optimal reproducibility, accuracy and throughput.

We applied our novel approach to initially investigate the rDNA loci *in vivo* and utilized our high throughput and accuracy to scrutinize previous observations on these loci. Our results find a similar dichotomy in the chromatin landscape, but do not agree with the transcriptionally independent character proposed previously. As our approach could uniquely detect the heterogeneity of individual chromatin fibers, we investigated the underlying regularity of arrays that would otherwise be classified as irregular using traditional techniques. We find that despite a lack of phasing, nearly all arrays display a high regularity in wild type cells. Similarly, previous findings suggested that lack of ISWI and CHD remodelers would result in unphased arrays, but could not definitively provide evidence for a lack of regularity. Here, we show direct evidence that deletion of these remodelers results in unphased and irregular nucleosomal arrays. We further investigate *in vitro* findings suggesting a 'clamping' model in which remodeler activity results in a density independent array spacing. Comparison of computational modeling with our *in vivo* findings do not support this model. In support of previous findings by our lab, we identify that INO80 is a bona fide spacing remodeler. We further provide preliminary evidence that this novel approach can be used to detect strain variants in order to sensitively detect global changes in nucleosome occupancy.

In the final chapter of this thesis, we investigate the spacing mechanism of the ISWI remodeler utilizing our sequencing approach in combination with an *in vitro* chromatin array. Previous findings in our lab have demonstrated the ability to investigate the kinetics of the remodeler using this array in

combination with a restriction based assay. However, these results do not provide details on the remodeling process and how nucleosomes are moved on a single fiber. To this end, we first designed a novel DNA template containing multiple Widom 601 sequences and linkers optimized for our sequencing approach. These optimizations would allow for accurate determination of nucleosome positions before and during remodeling. Our findings suggest that these arrays can be efficiently chromatinized by salt gradient dialysis (SGD) and that remodeling can be performed as previously described. In our preliminary findings we validate that the optimizations to the DNA template have a potential to improve accuracy in determining the nucleosome position.

Taken together, in this thesis I document the technological advancements and novel, previously unobtainable, insights that characterize the *S. cerevisiae* chromatin landscape. Furthermore, these findings lay the foundation for future studies to apply an analogous approach on more complex, multicellular organisms and identify how chromatin regulates life.

## II. ZUSAMMENFASSUNG

Eukaryotische Zellen verpacken und schützen ihre Erbinformationen mit Hilfe von Chromatin. Chromatin ist für diese Zellen überlebenswichtig. Nukleosome sind die Grundbausteine des Chromatins. Sie sind entlang der DNA wie Perlen auf einer Schnur positioniert. Die Nukleosome sind jedoch nicht etwa zufällig entlang des Genoms verstreut. In der Nähe der Startstelle für die Transkription befindet sich im Allgemeinen eine nukleosomenfreie Region (NFR). Diese wird gefolgt von vielen Nukleosomen, die in sehr regelmäßigen Abständen angeordnet sind. Wenn man dieses sog. Nukleosomen-*‚Array‘* in verschiedenen Zellen kartiert, bemerkt man, dass es sich an erstaunlich ähnlichen Stellen befindet. Man spricht in dem Zusammenhang von einem phasierten *‚Array‘*. Diese stereotype Organisation von Nukleosomen ermöglicht die Kontrolle der Genexpression und wird daher als ein wichtiges regulatorisches Element angesehen.

Einen großen Einfluss auf die Position von Nukleosomen haben ATP-abhängige Chromatin-*‚Remodeling‘*-Enzyme. Diese Enzyme können synergistisch oder antagonistisch zueinander wirken und übernehmen häufig redundante Rollen bei der Positionierung von Nukleosomen im Genom. Mutationen oder Deletionen dieser Enzyme können zu Instabilität des Genoms, zum Verlust der Regelmäßigkeit der Nukleosomen-*‘Arrays‘* oder zu verminderter Phasierung führen, wodurch die Heterogenität in der Chromatinlandschaft erhöht wird. Neben *‚Remodelern‘* beeinflussen andere Faktoren wie die DNA-Sequenz und Allgemeine Regulatorische Faktoren (GRFs) die Nuklesomenlandschaft. Die Nukleosomenlandschaft kann zwischen verschiedenen Zellytpen oder in Abhängigkeit des Zellzyklus variieren. Aufgrund der Vielzahl unterschiedlicher Einflussfaktoren ist es aber schwierig, die Biogenese der Chromatinlandschaft zu analysieren.

In dieser Arbeit untersuchen wir die Biogenese und Heterogenität der Chromatinlandschaft in *Saccharomyces cerevisiae*. Zu diesem Zweck haben wir zunächst einen neuartigen Sequenzierungsansatz entwickelt, der es uns ermöglicht, die Lage von Nukleosomen auf einer einzelnen Chromatinfaser zu untersuchen. Wir zeigen, dass dieser Ansatz einzigartige Einblicke in die Nukleosomenorganisation der DNA bietet. Er ermöglicht es uns, die Heterogenität von Zelle zu Zelle zu bestimmen. Auch können wir feststellen, welche Enzymfaktoren die Regelmäßigkeit des *‚Arrays‘* hervorrufen. Wir beschreiben detailliert die Versuchsabläufe und Analysen, die zu einer optimalen Reproduzierbarkeit, Genauigkeit und hohem Durchsatz führen.

Zuerst wendeten wir unsere neuartige Technik an, um den rDNA-Locus zu untersuchen. Unsere Ergebnisse bestätigen eine zuvor beschriebene Dichotomie in der Chromatinlandschaft dieses Locus. Sie stimmen jedoch nicht mit dem vorgeschlagenen transkriptionell unabhängigen Charakter einzelner rDNA-Wiederholungseinheiten überein. Des Weiteren nutzten wir unsere Technik um die Heterogenität von Nukleosomen-‚Arrays‘ zu bestimmen. Hier fokussierten wir uns u.a. auf diejenigen Regionen im Genom, die mit traditionellen Techniken als unregelmäßig klassifiziert wurden. Wir stellten fest, dass trotz fehlender Phasierung fast alle dieser *‚Arrays‘* in Wildtypzellen eine hohe Regelmäßigkeit aufweisen. Zellen, denen ISWI- und CHD-Remodeler fehlen, besitzen hingegen ein tatsächlich unregelmäßige *‚Arrays‘*, die nicht auf fehlende Phasierung zurückgeführt werden können. Des Weiteren untersuchten wir, inwiefern die Reduzierung der Nukleosomendichte die Arbeit von *‚Remodeling‘* Enzymen beeinflusst. Nach der ‚Klemmen‘ Theorien sollte eine Reduzierung der Dichte die Nukleosomenabstände nicht beeinflussen. Unsere Ergebnisse, gepaart mit Computermodellierungen, deuten aber auf einen nur schwachen Beitrag dieser molekularen

'Klemmen' hin. Unsere Ergebnisse stützen hingegen frühere Ergebnisse unseres Labors, dass INO80 ein Remodeler ist, der einen regelmäßigen Abstand zwischen Nukleosomen einstellen kann. Mit Hilfe einer Variante unserer Technik können wir zwei verschiedene Hefestämme miteinander mischen und dadurch gleichzeitig analysieren. Vorläufige Ergebnisse zeigen, dass wir so globale Veränderungen der Nukleosomendichte höchst sensitiv detektieren können.

Im letzten Kapitel dieser Arbeit untersuchen wir unter Verwendung unseres Sequenzierungsansatzes, wie ISWI-Remodeler regelmäßige Abstände zwischen Nukelsomen *in vitro* einstellt. Frühere Ergebnisse aus unserem Labor haben gezeigt, dass die *'Remodeling'* Kinetik mit einem *in vitro* rekonstituierten Nukleosomen-*'Array'* in Kombination mit Restriktionsendonukleasen untersucht werden kann. Diese Ergebnisse liefern jedoch keine Details zum Remodellierungsprozess und wie Nukleosomen auf einer einzelnen Faser bewegt werden. Zu diesem Zweck entwarfen wir zunächst eine neuartige DNA-Sequenz, die mehrere Widom 601-Nukleosompositionierungssequenzen enthält, die für unseren Sequenzierungsansatz optimiert wurden. Diese Optimierungen sollen eine genauere Bestimmung der Nukleosomenpositionen vor und während des *'Remodelings'* ermöglichen. Unsere Ergebnisse zeigen, dass diese *'Arrays'* durch Salzgradientendialyse (SGD) effizient chromatinisiert werden können und dass das ISWI diese *'Arrays'* als Substrat verwenden kann. In unseren vorläufigen Ergebnissen bestätigen wir, dass die Optimierungen der DNA Sequenz das Potenzial haben, die Genauigkeit bei der Bestimmung der Nukleosomenpositionen zu verbessern.

Insgesamt dokumentiere ich in dieser Arbeit die technologischen Fortschritte und neuartigen Einsichten in die Chromatinlandschaft von *S. cerevisiae*. Darüber hinaus legen diese Ergebnisse die Grundlage für zukünftige Studien, um mit Varianten unserer Technik Chromatin komplexerer, mehrzelliger Organismen zu untersuchen, in der Hoffnung, besser zu verstehen, wie Chromatin das Leben der Zelle reguliert.

## III. LIST OF FIGURES

# IV. LIST OF ABBREVIATIONS

| | |
|---|---|
| 5mC | 5-methyl Cytosine |
| A | Adenine |
| Abf1 | ARS binding factor one |
| ARS | Autonomously Replicating Sequence |
| ATAC | Assay for Transposase-Accessible Chromatin |
| ATP | Adenosine triphosphate |
| bp | basepair(s) |
| C | Cytosine |
| Chd1 | Chromodomain-helicase-DNA-binding protein one |
| ChIP | Chromatin Immunoprecipitation |
| Chr | Chromosome |
| CpG | Cytosine-Guanine pair |
| DNA | Deoxyribonucleic acid |
| DNase | Deoxyribonuclease |
| dNTP | Deoxyribonucleotide triphosphate |
| ds | Double strand |
| DTT | Dithiothreitol |
| E. coli | Escherichia coli |
| EDTA | Ethylenediaminetetraacetic acid |
| FA | Formaldehyde/formalin |
| G | Guanine |
| $G_1$-phase | Gap 1 phase |
| $G_2$-phase | Gap 2 phase |
| GAL | Galactose |
| gDNA | genomic deoxyribonucleic acid |
| GLU | Glucose |
| GpC | Guanine-Cytosine pair |
| GRF | General Regulatory Factor |
| H1 | Histone 1 |
| H2A | Histone 2A |
| H2A.Z | Histone 2A variant Z |
| H2B | Histone 2B |
| H3 | Histone 3 |
| H4 | Histone 4 |
| HD | Histone Depleted |
| INO80 | Inositol requiring 80 |
| ISWI | Imitation Switch |
| kb | kilobase |
| L | Liter |
| m6A | methyl-6-Adenine |
| MEFSIMO | Methylation Footprinting (followed by) Single Molecule |
| Mg2+ | Magnesium 2+ |
| MgCl | Magnesium Chloride |
| mL | Milliliter |

| | |
|---|---|
| mM | Millimolar |
| MNase | Micrococcal Nuclease |
| M-phase | Mitotic phase |
| MQ | Milli-Q water |
| MUM | Methyl-Unmethyl-Methyl group |
| N+1/+X | Nucleosome +1/+X |
| NCP | Nucleosome Core Particle |
| NDR | Nucleosome Depleted Region |
| NFR | Nucleosome Free Region |
| nM | Nanomolar |
| NOME-seq | Nucleosome Occupancy and Methylation sequencing |
| NPD | Nanopore Daan (experimental nomenclature) |
| NRL | Nucleosome Repeat Length |
| N-terminal | Amino-terminal |
| NTR | N-terminal Region |
| Occ | Occupancy |
| PAGE | Polyacrylamide Gel Electrophoresis |
| PCR | Polymerase Chain Reaction |
| PIC | Pre-initiation Complex |
| PMSF | Phenylmethylsulfonyl fluoride |
| Pol I/II | Polymerase I/II |
| rDNA | ribosomal deoxyribonucleic acid |
| RES | Restriction Enzyme Site |
| RNA | Ribonucleic Acid |
| RSC | Remodel the Structure of Chromatin |
| *S. cerevisiae* | Saccharomyces cerevisiae |
| *S. pombe* | Schizosaccharomyces pombe |
| SAH | S-(5'-Adenosyl)-L-Homocysteine |
| SAM | S-adenosylmethionine |
| SDS | Sodium Dodecyl Sulfate |
| Seq | Sequencing |
| SGD | Salt Gradient Dialysis |
| SHL | Super Helical Loop |
| SLIM-ChIP | Short-fragment-enriched, Low-input, indexed MNase ChIP |
| SMAC-seq | Single-Molecule long-read Accessible Chromatin sequencing |
| SMRT-seq | Single Molecule Real Time sequencing |
| SNP | Single Nucleotide Polymorphism |
| S-phase | Synthesis phase |
| SWI/SNF | Switch/Sucrose Non-Fermenting |
| T | Thymine |
| TSS | Transcription Start Site |
| TTS | Transcription Termination Site |
| UAF | Upstream Activating Factor |
| UMU | Unmethylated-Methylated-Unmethylated group |
| URS | Unique Restriction-enzyme Sites |
| VLL | Variable Linker Length |

| | |
|---|---|
| w/v | weight/volume |
| WT | Wild Type |
| Zn2+ | Zinc 2+ |
| µg | microgram |
| µl | microliter |

# 1. INTRODUCTION

First identified in the late 19th century by W. Flemming, the DNA of eukaryotic organisms is packaged into a macro-molecular structure referred to as chromatin (Flemming, 1882). Here, he discovered that a certain substance in the cell nucleus would stain when a dye was added, and hence referred to it as 'chromatin' from the Greek word *chroma*. Subsequent discoveries over the next century, identified that histones (by A. Kossel in 1911) and nucleic acids (by F. Miescher in 1871) were the main components that made up chromatin. Seven decades later, the term 'nucleosome' was proposed as the basic repeating unit of chromatin (Oudet et al., 1975). These nucleosomes were observed as 'beads on a string' by electron microscopy, coating the entire length of chromatin fibers (Olins and Olins, 1974; Woodcock et al., 1976). Together with nuclease digestion- and crosslinking data, it was determined that these nucleosomes consist of approximately 200 bp of DNA and four distinct histone dimers (Hewish and Burgoyne, 1973; Kornberg, 1974; Kornberg and Thomas, 1974). Two decades later, a high-resolution crystal structure revealed a detailed insight into the nucleosome (Luger et al., 1997). This structure consisted of the canonical histone dimers (H2A, H2B, H3 and H4) wrapped with 146 bp of DNA. In the genome, these nucleosomes are connected by linker DNA, which can be of varying length depending on the cell, cell type or organism. The role of nucleosomes, DNA and chromatin in eukaryotic organisms is a well-studied, but often poorly understood field, that is pertinent to understanding life in general.

## 1.1. Fundamental features of chromatin

Chromatin is a complex, dynamic and heavily regulated part of any eukaryotic organism. Despite its surrounding complexity, it is defined by a few key features. Here, I will first describe these features and their role.

### 1.1.1. The canonical nucleosome and its variants

A canonical nucleosome consists of two distinct components. The nucleosome core particle (NCP) and a cell-type and/or species variable DNA linker of 10-80 bp. The NCP itself consists of eight, evolutionary conserved, histones (H2A, H2B, H3 and H4, Figure 1.1.1A). Each of the eight histones have a similar 'histone fold' motif that consist of three α-helices which facilitate the dimerization of H2A to H2B and H3 to H4 (Arents and Moudrianakis, 1995). The H3/H4 dimers may further form tetramers which subsequently interact with two H2A/H2B dimers to form the canonical histone octamer of the NCP. 145-147 bp of DNA wraps in approximately 1.65, left-handed, super-helical turns around the octamer (Davey et al., 2002; Luger et al., 1997). The center of the NCP is defined by a pseudo-twofold axis and is often referred to as the nucleosome dyad. Contacts between the DNA and the histone octamer occur at regular intervals through direct DNA-protein interactions (Davey et al., 2002).

Starting from the nucleosome dyad, the interactions between DNA and histone octamer are defined as super helical location (SHL) 0 to ±7 in 10 bp steps. Sites where the major groove of DNA interacts with the octamer are defined as SHL ±1 to ±7 and minor groove interactions are defined as SHL ±0.5 to ±6.5. Notably, interactions are not energetically identical at each SHL (Hall et al., 2009). Strongest interactions are found at the dyad, whereas weaker interactions are found at the entry/exit site or at SHL ±2.5 to facilitate (asymmetric) unwrapping or as binding sites for remodeling enzymes (Farnung et al., 2017; Kono et al., 2018; Liu et al., 2017; Zofall et al., 2006).

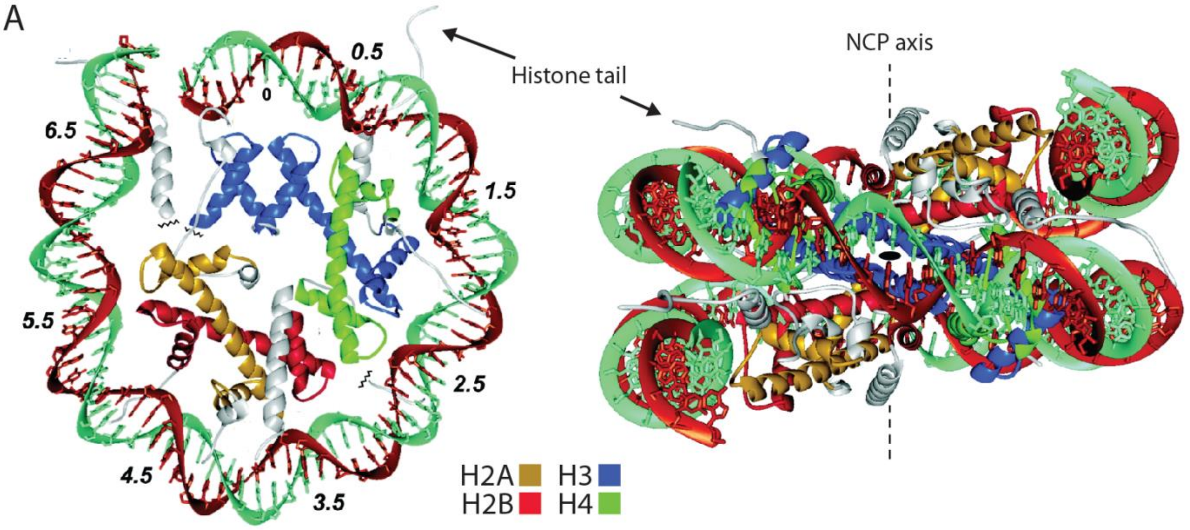


**Figure 1.1.1 High resolution structure of the nucleosome core particle (NCP)**
***(A)*** *Left: Top view of the nucleosome core particle consisting of human alpha-satellite DNA and recombinant Xnopus laevis histones. Only one of each histone and half of the DNA is shown. Each component is colored individually. Values on outside represent super-helical loops starting from the dyad (0). Right: Side view of all eight histones and full length of DNA is shown. NCP axis shows the pseudo two-fold axis. Protruding from the sides are the histone tails, of which one is indicated by arrows. Figure adapted from Davey et al., 2002.*

The characteristic disc-like shape of the NCP is approximately 5.5 nm tall and 11 nm wide. Protruding from the main body of the NCP are unstructured 'histone tails' from the N-terminal of histones H3, H4 and H2B and C-terminal of histone H2A. Such tails may be post-translationally modified to provide targeted binding sites for other proteins, stimulate the activity of bound proteins or can result in the (re-)structuring of higher order chromatin (reviewed in Peterson and Laniel, 2004).

Canonical histones can be exchanged for histone variants. In *S. cerevisiae* it is often observed that the canonical H2A is replaced by the H2A.Z variant. Deletion of this non-essential histone results in defective transcriptional activation and silencing. In centromeres, histone H3 is replaced by Cse4, which is essential and results in cell cycle arrest when inactivated (Stoler et al., 1995). Lastly, in transcriptionally inactive chromatin, referred to as heterochromatin, we often find histone H1. The H1 histone binds 20 bp of linker DNA, bringing the NCP entry/exit sites together, thereby stabilizing the nucleosome and aiding in the silencing of the DNA (Hamiche et al., 1996). Together with the NCP, this structure is called the chromatosome (Simpson, 1978). These histones are less conserved compared to the canonical histones, but are abundant in the heterochromatin of higher organisms. Organisms such as *S. cerevisiae* have little to no inactive or gene-poor DNA. The homolog of histone H1, Hho1, is therefore in low abundance and deletion does not affect the cell in any meaningful way (Patterton et al., 1998).

### 1.1.2. Levels of chromatin structure

The most basic and well-known structure of the chromatin landscape is the 10-nm fiber (Figure 1.1.2A). This structure consists of several NCPs connected by linker DNA and is often also referred to as the 'beads on a string' model or as a nucleosomal array (Olins and Olins, 1974; Woodcock et al., 1976). How the 10-nm fiber is organized into a higher-order structure such as the 30-nm fiber, is debated and may be salt concentration dependent. Earlier observations based on electron microscopy

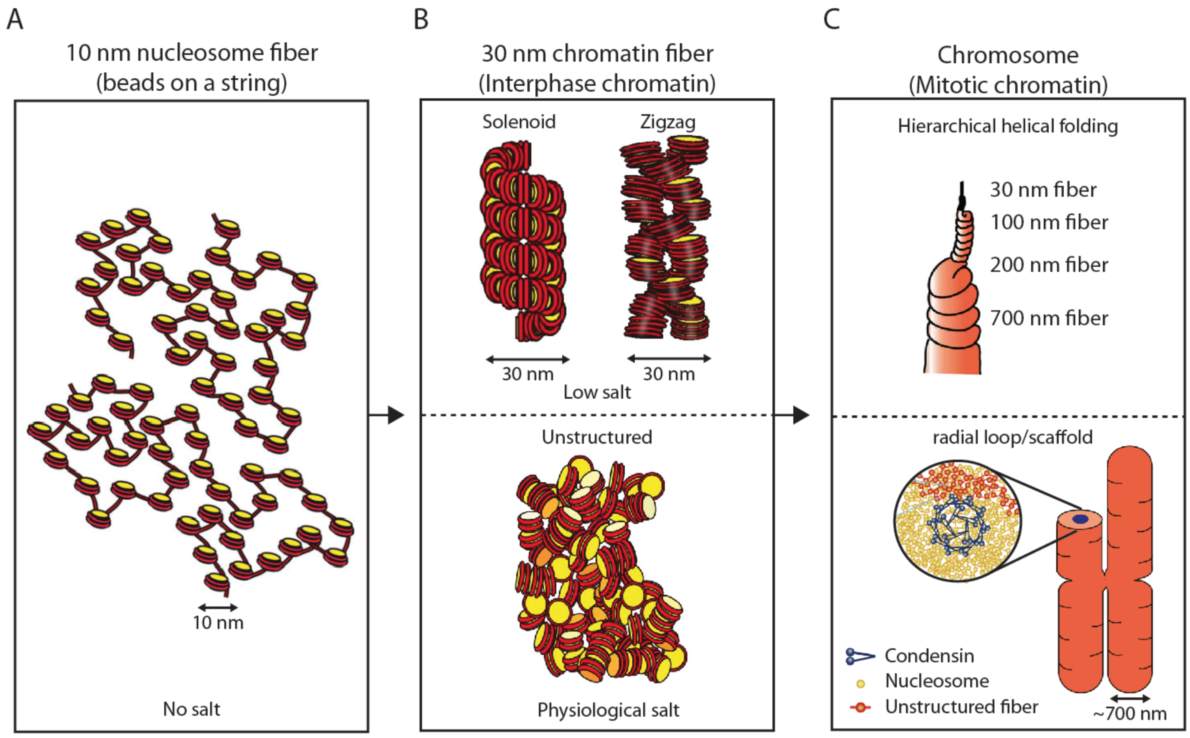


**Figure 1.1.2 Levels of chromatin structure**
***(A)*** *The 10 nm nucleosome fiber is the lowest level of chromatin organization. Shown are two individual fibers consisting of nucleosomes (yellow) linked by DNA (red) under no salt conditions.* ***(B)*** *The subsequent level of chromatin organization is the 30 nm fiber, found during interphase. Under low salt conditions (<1mM* $MgCl_2$ *or <100mM NaCl) it is suggested that the 30 nm fiber can form either (or both) of two distinct fiber organizations: a solenoid or zigzag. At physiological salt conditions it is argued that an unstructured organization is favored.* ***(C)*** *The highest level of chromatin organization is the chromosome during the mitotic phase. One sister chromatid is approximately 700 nm in width. Two models with varying supporting evidence have been proposed. A hierarchical model in which progressively shapes larger chromatin structures, and a radial loop/scaffold model which relies on condensing as a core around which unstructured 10 nm fibers form. All figures in (A-C) adapted from (Maeshima et al., 2014).*

suggested a solenoid structure, where nucleosomes are neatly organized side-by-side (Finch and Klug, 1976). Alternatively, a zigzag structure was proposed in which nucleosomes are also organized neatly, but further arrange as a helical ribbon (Woodcock et al., 1984). Both of these observations are under scrutiny as more recent findings suggest that such 30-nm fibers only occur under *in vitro* low-salt conditions or in highly specific cell types where they may play a role in gene silencing (reviewed in Hansen, 2012; Scheffer et al., 2011; Woodcock, 1994). More recent models posit a model in which the 10-nm fibers are dynamically and irregularly folded into a higher order chromatin structure that resembles a liquid-like structure (Maeshima et al., 2010, 2014, 2019)(Figure 1.1.2B).

Regardless, the (lack of) chromatin organization at non-mitotic cell cycle stages is overshadowed by the well-known organization of chromosomes during interphase, which can be seen by typical light microscopy. The underlying mechanism by which this organization occurs is also debated. One model, referred to as the ‘hierarchical helical folding model’ proposes that the interphase chromosome is progressively shaped in four distinct steps. These steps start with a 30 nm fiber folding into a 100 nm fiber and subsequently into a 200 nm fiber until a typical chromosome is formed (Belmont and Bruce, 1994; Belmont et al., 1987; Sedat and Manuelidis, 1977). Alternatively, a ‘radial loop/scaffold model’ posits that irregularly folded 10-nm fibers organize around a central condensing core (Figure 1.1.2C).

This core serves as a scaffold along which the 10-nm fiber attaches in an irregular pattern (Mirny et al., 2019; Nishino et al., 2012).

### 1.1.3. The typical chromatin landscape organization

Despite the debate on the formation of a ~700-nm chromosome that is visible by light microscopy, it is clear that this does not happen without some form of regulation. Similarly, the organization of nucleosomes on the 10-nm fiber is not random. How these nucleosomes are distributed and positioned results in a highly typical chromatin landscape. First indications of an organized distribution

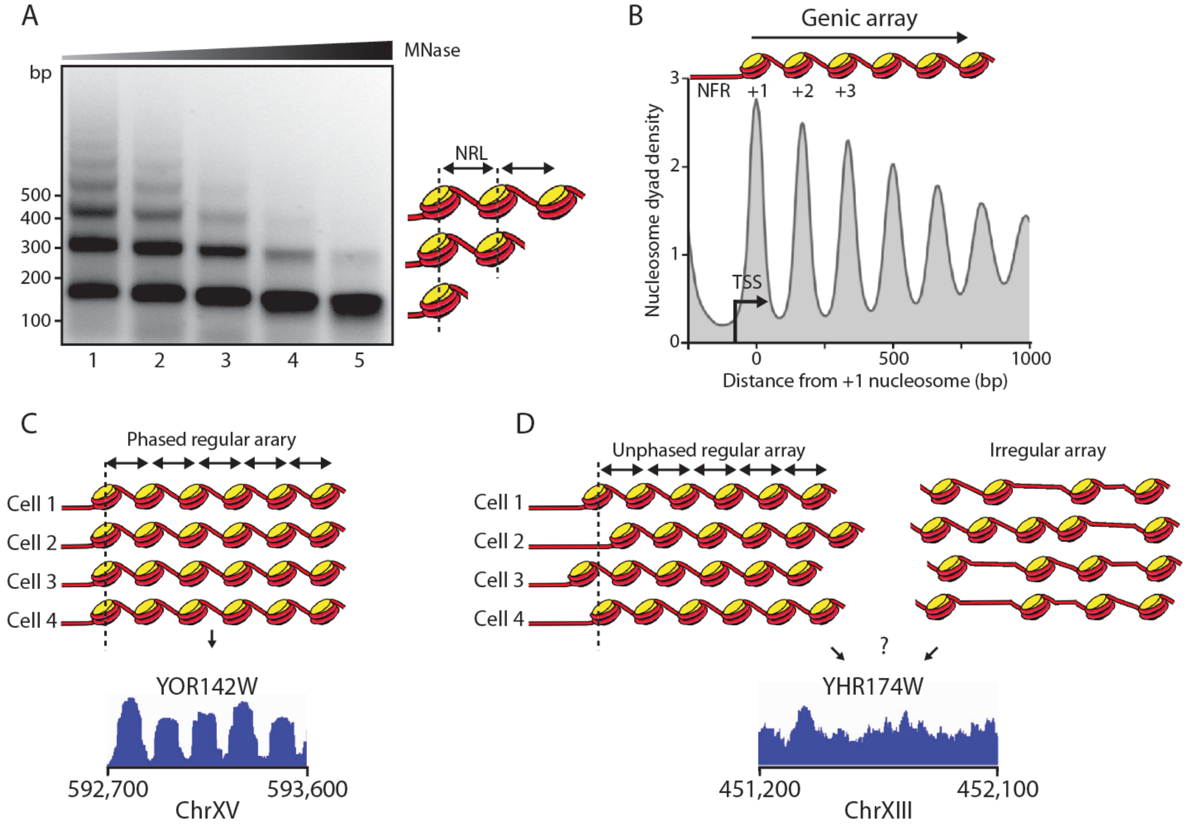


**Figure 1.1.3 Nucleosome regularity, phasing and spacing on the chromatin landscape**
***(A)*** *Typical result of a limited micrococcal nuclease (MNase) assay. Digestion increases in accordance with increased enzyme concentration when chromatin is incubated for similar durations. DNA is isolated and analyzed by agarose gel electrophoresis and ethidium bromide staining. At lower concentrations (lane 1-3) trinucleosome fragments can be observed (octamer and DNA, yellow and red, resp.). At elevated concentrations, only mononucleosomal bands remain (lane 5). Note the continuous decrease in mononucleosomal band size inherent to MNase digestion. Based on the location of the bands, a global nucleosome repeat length (NRL) can be estimated. Nucleosome fiber adapted from Maeshima et al., 2014.* ***(B)*** *Result of an MNase-seq experiment showing the typical distribution of a genic nucleosome array when data is aligned to the +1 nucleosome. Upstream of the +1 nucleosome we find a nucleosome free region (NFR), which typically contains the TSS. The decrease in peak amplitude implies a global variability of regularity within the gene body. MNase-seq composite plot adapted from Singh and Mueller-Planitz, 2021.* ***(C)*** *A theoretically perfect phased array has the first nucleosome at a defined location (dashed line) and all subsequent nucleosomes equally spaced downstream. Such perfect organization does not exist in nature, but close approximations can be found. Shown is MNase-seq data for the YOR142W gene. Nearly all nucleosomes are at defined locations as indicated by the distinct peaks. In S. cerevisiae it is relatively common to find an organization as shown here. Figure adapted from Singh and Mueller-Planitz, 2021.* ***(D)*** *A less common occurrence is shown for the YHR174W gene. MNase-seq data suggests cannot discern the locations of the nucleosomes. From this type of data, it is unknown whether the underlying chromatin architecture is unphased but regular, or irregular in general. Figure adapted from Singh and Mueller-Planitz, 2021.*

were observed by the digestion of chromatin using endonucleases. These endonucleases digest preferentially linker DNA, resulting in individual nucleosomes (Clark and Felsenfeld, 1971; Hewish and Burgoyne, 1973; Noll, 1974). When the digestion is performed for a limited time or with a limited endonuclease concentration, equally spaced DNA bands were observed after gel electrophoresis (Figure 1.1.3A). These so-called ‘DNA ladders’ suggest that there is an equal distance between nucleosomes. By observing such DNA ladders, it was later deduced that the nucleosomal repeat length (NRL) for *S. cerevisiae* is on average 165 bp. These 165 bp consists of 147 bp of nucleosomal DNA and an 18 bp linker. However, the NRL varies between species, cell types and even transcriptional states within the same cell (Chereji et al., 2018; Lantermann et al., 2010).

Over time, the digestion of chromatin using nucleases has been refined. Currently, a micrococcal nuclease (MNase) isolated from *Staphylococcus aureus* is used (reviewed in Telford and Stewart, 1989). This enzyme, in combination with sequencing methods is used to determine the location of individual nucleosomes on the chromatin (MNase-seq). Utilizing the positional information of single nucleosomes, each one can be aligned to the transcription start site (TSS) of individual genes (Figure 1.1.3B). Combining the information of all locations and all genes, results in so-called composite plots.

From these composite plots generated by utilizing e.g. MNase-seq data, it was identified that upstream of the +1 nucleosome a nucleosome depleted/free region (NDR/NFR, hereafter referred to as NFR) is formed. The NFR width is demarcated by an upstream -1 nucleosome. Downstream of the +1 nucleosome we find a regular repeating pattern of nucleosomes. The distances between the peaks defines the NRL of a gene or locus. Globally this NRL is similar to what was observed in gel electrophoresis experiments. What is moreover apparent from these composite plots is that the underlying regularity is phased, i.e. the array starts at a defined location, typically close to the TSS. This observation holds true for the genome wide average, but is not necessarily true for individual genes. When probing the location of nucleosomes at individual genes, we find that the profile can be similar to the global composite plot. This suggests that this gene has a regular and phased array (Figure 1.1.3C). In stark contrast, it is impossible to define the state of an array for genes that do not display such profile due to the limitations of MNase-seq (Figure 1.1.3D). In this thesis I will provide a novel approach (Chapter 1) to elucidate this issue (Chapter 2).

## 1.2. Establishment and role of nucleosome arrays

The nucleosomal organization in eukaryotes is not random, but is cell and/or species specific. Many factors regulate and influence the position and thus shape the chromatin landscape. Understanding how they function is key in understanding eukaryotic life. Here, I will provide an overview of the factors that shape and maintain the chromatin landscape.

### 1.2.1. Sequence intrinsic determinants for nucleosome positioning

DNA, one of the fundamental features of chromatin, has been studied *in vitro* and *in vivo* to determine whether its intrinsic properties can influence the position of nucleosomes (Figure 1.2.1B, step 1). Certain sequences wrap around a histone octamer with higher efficiency compared to others. For example, it was established that poly(dA-dT) and poly(dG-dC) dinucleotides wrap around the octamer much easier compared to stretches of poly(dA)-poly(dT) and poly(dG)-poly(dT) (Rhodes, 1979; Simpson and Kunzler, 1979). More specifically, by utilizing 177 different DNA fibers, it was identified that the minor groove of the DNA that faced inwards to the histone octamer, preferentially consisted of AAA/TTT or AAT/ATT nucleotides (Satchwell et al., 1986). More recently, it was shown that *in vitro* reconstituted salt gradient dialyzed (SGD) chromatin repels nucleosomes when lengthy poly(dA)-poly(dT) stretches were used. The intrinsic ability to keep DNA free of nucleosomes may be especially important in organisms such as *S. cerevisiae* that have poly(dA)-poly(dT) stretches at NFRs (Anderson and Widom, 2001; Iyer and Struhl, 1995; Zhang et al., 2011).

### 1.2.2. Passive regulation of nucleosome position by general regulatory factors (GRFs)

Non-ATPase dependent factors further aide in shaping the chromatin landscape. General regulatory factors (GRFs) such as Abf1, Rap1 and Reb1 have specific DNA binding motifs, often located in the NFR (Figure 1.2.1B, step 2). Due to the lack of an ATPase domain, they physically exclude the possibility of nucleosome assembly or transcription factor binding (Gutin et al., 2018; Wu et al., 2018). Indeed, depletion of the (often) essential GRFs results in a shift of nucleosomes over the binding site (Kubik et al., 2015, 2018). GRFs recruit and physical interact with nucleosome remodelers such as RSC (Wu et al., 2018). *In vitro* results indicate that Abf1 in combination with remodelers such as CHD, RSC and ISWI improve the position of the +1 nucleosome to near native states (Krietenstein et al., 2016). Thus, GRFs play an important role in establishing the chromatin landscape.

### 1.2.3. Active regulation of nucleosome position by ATP-dependent nucleosome remodelers

ATP-dependent nucleosome remodelers can be subdivided into four families, classified by their ATPase subunit: ISWI, CHD, INO80 and SWI/SNF. Their roles include nucleosome spacing, sliding, phasing, eviction and histone variant switching. Common across all the remodelers is a catalytic subunit with two RecA-like ATPase lobes. The mechanism by which the RecA-like lobes move nucleosomes is generally similar and described as an ‘inchworming’ mechanism. Briefly, cyclic binding of ATP changes the conformation of lobes. This drives the translocation of DNA by 1-2 bp for each hydrolyzed ATP. However, the exact details remain obscured and may be different for each remodeler (reviewed in Clapier et al., 2017). The ATPase subunits are able to function independently (albeit in slightly different methods), but typically fulfill their role as part of a multi-subunit complex *in vivo* (Dechassa et al., 2012; McKnight et al., 2011). Depending on the type and number of associated subunits, specific role(s) for each complex may vary (Table 1).

**Table 1 Overview of *S. cerevisiae* remodeler families, features and functions.** *The ATPase is the catalytic subunit. Together with any additional subunits form a complex.*

| *Subfamily* | *Complex* | *ATPase* | *Additional subunit(s)* | *Function* |
|---|---|---|---|---|
| ***ISWI*** | *ISW1a* | *Isw1* | *1* | *Sliding, spacing* |
| | *ISW1b* | *Isw1* | *2* | *Sliding* |
| | *ISW2* | *Isw2* | *1* | *Sliding, spacing* |
| ***CHD*** | *CHD1* | *Chd1* | *-* | *Sliding, spacing* |
| ***INO80*** | *INO80* | *Ino80* | *15* | *H2A.Z removal, sliding, spacing* |
| | *SWR1* | *Swr1* | *14* | *H2A.Z incorporation* |
| ***SWI/SNF*** | *SWI/SNF* | *Swi2/Snf2* | *8 – 14* | *Sliding, eviction* |
| | *RSC* | *Sth1* | *15 – 16* | *Sliding, eviction* |

The ISWI family in *S. cerevisiae* is represented by two homologs: Isw1 and Isw2. The homologs form three remodeling complexes: ISW1a, ISW1b and ISW2. The ISWI remodelers are well-known for their sliding and spacing activity. *In vitro*, ISW1a and ISW2 sliding results in the centering of a nucleosome on the DNA. ISW1b does not have such centering activity (Stockdale et al., 2006). The ability to center nucleosomes is related to the capability to equally space nucleosomes. Indeed, when ISW1a and ISW2 are incubated with *in vitro* SGD chromatin they generate regular arrays upon the addition of ATP (Krietenstein et al., 2016). *In vivo*, deletion of the catalytic subunit *ISW1* results in a reduction of regular nucleosomal arrays. Deletion of *ISW2* has little to no effect on the global array phenotype (Gkikopoulos et al., 2011; Ocampo et al., 2016). However, these studies also identified that deletion of *CHD1*, the sole member of the CHD family has a drastic effect on the regularity of nucleosomal arrays. In a context where *ISW1*, *ISW2* and *CHD1* remodelers were deleted, array regularity was worse compared to a deletion of only *ISW1* and *CHD1* (Gkikopoulos et al., 2011; Ocampo et al., 2016; Singh et al., 2021). These findings indicate that there is likely a redundant function between the remodelers, which may not directly be clear from the deletion of a single remodeler. Moreover, the redundancy suggests that the state of the chromatin landscape is highly important and does not depend on a single remodeler. Lastly, the ISWI/CHD remodelers have been suggested to function as a ‘clamp’ between nucleosomes (Figure 1.2.1B step 4). This clamping mechanism ensures fixed NRLs, despite variable nucleosome densities on the DNA (Lieleg et al., 2015). The ability to keep nucleosomes relatively close to one another, and prevent large ‘gaps’ in an array, may be important to prevent unwanted or ‘cryptic’ transcription (Figure 1.2.1A). This process results in assembly of the transcription initiation machinery to bind and initiate transcription at a different location than the defined TSS. Indeed, deletion of *ISWI* and *CHD1* remodelers was shown to increase (anti)sense transcription (Smolle et al., 2012)

The SWI/SNF family consists of the SWI/SNF and RSC remodeler. The remodelers each have a different, but highly homologous, catalytic subunit. Additional non-catalytic subunits of each complex can be shared amongst the two remodelers whilst others are unique to either remodeler. Based on global yeast expression levels in *S. cerevisiae,* the RSC remodeler is highly abundant and found in approximate 10-fold excess compared to the SWI/SNF remodeler (Ghaemmaghami et al., 2003). Moreover, the RSC catalytic subunit is essential whereas SWI/SNF is not, but certain subunits of the latter complex are (Cairns et al., 1996). These remodelers are most known for their capability to slide nucleosomes away from the promotor NFR or evict nucleosomes (Kubik et al., 2019). Due to its high abundance and broad function, depletion of RSC results in a global shift of the +1 nucleosome into the NFR (Ganguli et al., 2014). It is suggested that this shift results in a reduction of transcription of the affected genes (Kubik et al., 2018). Additionally, It has been suggested that the RSC remodeler and ISW2/ISW1a are in a ‘tug of war’ over the position of the +1 nucleosome (Parnell et al., 2015). The lower abundant SWI/SNF complex is mostly found at highly expressed genes where it cooperates with RSC (Figure 1.2.1B, step 2 and 3). Here, their combined function results in widened NFRs, presumably to aide in robust transcription (Rawal et al., 2018).

Last of the four remodeling families is INO80, which consists of INO80 and SWR1 complexes with Ino80 and Swr1 as the catalytic subunits. INO80 remodelers have the two characteristic RecA-like lobes, but are separated by a long insertion in this family. Besides their sliding activity, it has been suggested that the INO80 family remodelers fulfill a role in the incorporation or eviction of H2A.Z. The SWR1 complex exchanges H2A with H2A.Z (Mizuguchi et al., 2004), whilst the INO80 complex is suggested (although

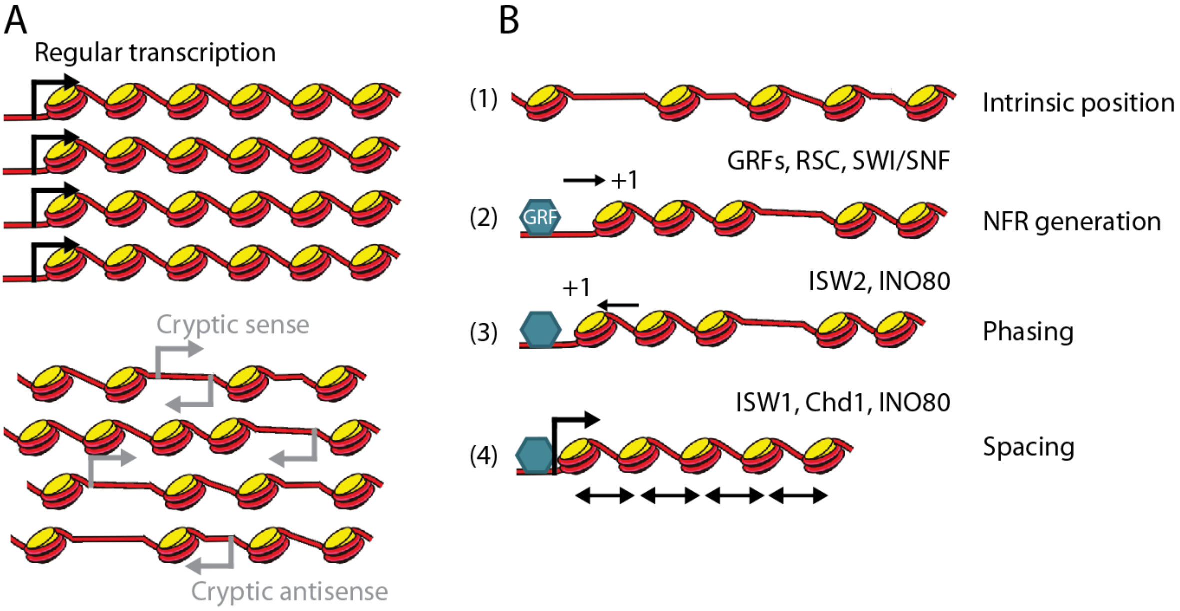


**Figure 1.2.1 Role and establishment of nucleosomal arrays**
***(A)*** *Regular and phased arrays generated by nucleosome remodelers favor regular transcription from the native transcription start site (TSS). Irregular spacing of arrays may lead to cryptic transcription in the sense or antisense direction at non-native TSSs.* ***(B)*** *Positioning of nucleosomes and the generation of a regular array depends on many factors. (1) DNA sequence can impose a favored position on the position of nucleosomes. (2) The intrinsic position is overwritten initially by several factors. Generation of a nucleosome free region (NFR) by GRFs + RSC and SWI/SNF shift the +1 nucleosome away and into the gene body. (3) A tug of war occurs between RSC and ISW2/INO80 which determines the ultimate location of the +1 nucleosome and sets the correct TSS. (4) Spacing remodelers such as ISW1, Chd1 and INO80 generate a regular array downstream of the +1 nucleosome. Figure adapted from Singh and Mueller-Planitz, 2021.*

later contested) to perform the opposite exchange (Brahma et al., 2017; Papamichos-Chronakis et al., 2011; Wang et al., 2016; Watanabe et al., 2013). Additionally, the INO80 complex has been shown to slide nucleosomes to the center of DNA *in vitro* and generate longer NRLs compared to ISWI and CHD remodelers (Azmi et al., 2017; Udugama et al., 2011). Recent findings in our lab have indicate that INO80 has a clear effect on array regularity and spacing *in vivo* and that depletion leads to shorter NRLs (Singh et al., 2021). These findings suggest that INO80 functions as a spacing remodeler in addition to its previously mentioned roles. Despite being of the same family, SWR1 has not been reported to have any significant sliding activity in *S. cerevisiae*, although short shifts within the bounds of the nucleosomal DNA have been reported (Ranjan et al., 2015; Singh et al., 2019)

## 1.3. Visualization of the chromatin landscape through conventional and novel sequencing techniques

I have outlined the typical characteristics of chromatin and the nucleosome landscape (Figure 1.3.1A). These invaluable insights have been made possible through decades of studies utilizing a wide variety of techniques. Despite best efforts, many of these techniques provide a snapshot of the chromatin based on the average over many observations of different cells. Recent technological advances have made it possible to visualize the chromatin landscape on a single-cell level, by utilizing an alternative approach. I will now discuss certain specific techniques that have allowed us to study chromatin up until now. Then, I will discuss the recent advances and how I intend to use them in this thesis.

### 1.3.1. Conventional methods in determining the chromatin landscape

The previously mentioned MNase-seq method is presumably the most utilized technique to study the nucleosome landscape at the time of writing. The functional enzyme in this technique is the micrococcal nuclease (MNase) that is isolated from *Staphylococcus aureus* (Axel, 1975; Dingwall et al., 1981). Importantly, this enzyme exhibits both endo- and exonuclease activity and preferentially digests non-nucleosomal DNA. Initially, MNase was paired with microarrays (MNase-chip) to provide high resolution maps of nucleosomes (Lee et al., 2007). Not much later, MNase was combined with next-generation sequencing techniques (MNase-seq) to improve throughput and improve coverage (Schones et al., 2008). Despite the broad use of MNase to probe the chromatin landscape, it is important to understand that digestion with this enzyme exhibits different kinetics dependent on the concentration. Moreover, the enzyme has a positive bias for certain regions in the genome, such as the poly(dA)-poly(dT)-rich NFR, and may digest DNA of 'fragile' nucleosomes making them more difficult to detect (Chung et al., 2010; Mieczkowski et al., 2016; Xi et al., 2011). Managing this bias may be possible by implementing an orthogonal chemical cleavage technique that does not use the MNase enzyme (Chereji et al., 2018). An additional highly similar technique uses a DNase I endonuclease in combination with next-generation sequencing (DNase-seq). Similar to MNase, DNase I has an intrinsic bias, and degrades certain DNA sequences and/or shapes more efficiently than others (Lazarovici et al., 2013). Taken together, MNase-seq, DNase-seq and chemical cleavage all provide information on the position of individual nucleosomes or open chromatin by cutting the DNA and subsequent sequencing.

With a similar goal to the techniques described above, the assay for transposase-accessible chromatin sequencing (ATAC-seq) technique instead maps the accessible (i.e. non-nucleosomal) instead of the inaccessible (i.e. nucleosomal) DNA. To this end, a hyperactive Tn5 transposase is used to insert sequencing adapters into the accessible regions of the chromatin (Buenrostro et al., 2015). In essence, the transposase cuts within the accessible DNA, adds a pre-loaded sequencing adapter, after which the cut DNA gets amplified by PCR and sequenced. Together this provides a map of the open chromatin and thus by inverse, the location of inaccessible regions. Due to the reliance on accessible DNA, ATAC-seq results can be contaminated with 20-80% mitochondrial DNA reads. Improved methods implementing additional washing or CRISPR/Cas9 can prevent this issue (Corces et al., 2017; Montefiori et al., 2017).

The localization of chromatin-bound proteins such as histone octamers, but also RNA polymerases or GRFs can be further detected by using a chromatin immuno-precipitation sequencing approach (ChIP-seq). Here, instead of relying on an enzyme to digest the (in)accessible DNA, samples are sonicated to physically shear the chromatin, but MNase may added to reduce non-nucleosomal DNA (Albert et al., 2007). Next, the chromatin is incubated with an antibody specific to the protein of interest (e.g. histone H2A) and immunopurified. The DNA within the resulting eluate is isolated and sequenced using next-generation sequencing. Several alterations to the standard ChIP-seq protocol currently exist to increase resolution such as ChIP-exo and SLIM-ChIP (Gutin et al., 2018; Rhee and Pugh, 2012)

First steps to a non-destructive sequencing method were made when bisulfite sequencing (BS-seq) was combined with the induced methylation of accessible cytosines. This approach measures the nucleosome occupancy by methylome sequencing (NOME-seq) and uses a commercially available methyltransferase enzyme (M.CviPI). This enzyme non-destructively methylates accessible cytosines

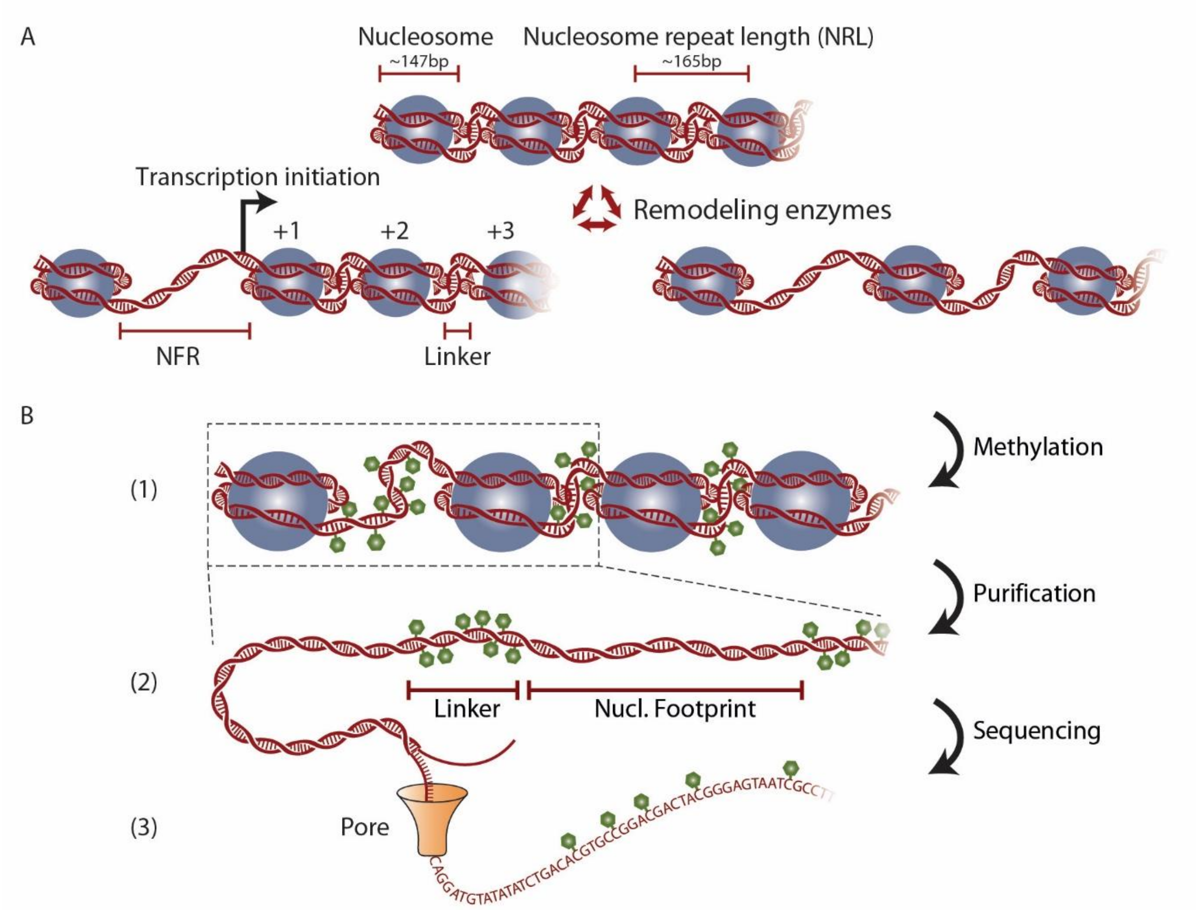


**Figure 1.3.1 Concept of single-molecule sequencing of methylated DNA using the Nanopore platform**

***(A)** General concept of the sequencing strategy used throughout this thesis, referred to as methylation footprinting followed by single-molecule sequencing (MEFSIMO-seq). The chromatin landscape may be organized in many different ways. At the center of these changes are often remodeling enzymes creating nucleosome free regions (NFRs), altering linker distances which affect the nucleosome repeat length (NRL), or evicting nucleosomes. **(B)** Chromatin goes through three distinct steps during the MEFSIMO-seq protocol. (1) Accessible DNA is methylated (green hexagons). (2) In a subsequent step, DNA is purified, leaving a methylated region which corresponds to e.g. a linker or NFR and an unmethylated region. This unmethylated region is referred to as a nucleosome footprint. (3) The purified DNA is directly sequenced utilizing the Oxford Nanopore Technologies (ONT) platform, which (after computational analysis) can directly detect (modified) nucleotides.*

in a GpC context (Fides et al., 2018; Xu et al., 1998). Briefly, the methyltransferase enzyme methylates preferentially accessible DNA (i.e., non-nucleosomal). In a subsequent bisulfite conversion step, unmethylated cytosines are deaminated converting them into uracil. Importantly, methylated cytosines are not affected by this conversion. After a PCR amplification step, the uracil is converted into thymine. In order to efficiently perform the PCR conversion, a fragmentation/sonication step has to be implemented after methyltransferase treatment which shears DNA into smaller fragments. This approach has been used to detect nucleosome positions for specific loci and genome wide (Kelly et al., 2010, 2012).

### 1.3.2. Single-molecule long-range sequencing approaches

Inherent to the techniques outlined above is a mandatory step with an unavoidable result: fragmentation of DNA. Whether this is by the inherent nuclease activity of the enzyme or the sequencing technique, the resulting data returns short fragments. A clear disadvantage of short reads is that single-cell information is mostly lost. After the fragmentation, it is near impossible to determine which combination of short fragments originated from a single cell. Methods to apply DNase-seq and ATAC-seq on single cells have been reported, but can be experimentally laborious or low throughput (Cusanovich et al., 2018; Lai et al., 2018; Mulqueen et al., 2019). To overcome these limitations, an approach is required that combines the ability to retain the integrity of DNA, whilst providing a platform that allows the sequencing unfragmented DNA.

Methyltransferase enzymes such as those used in BS-seq and NOME-seq, have the benefit of non-destructively methylating accessible chromatin. In essence, this leaves a footprint of inaccessible chromatin regions, e.g., nucleosomes (Figure 1.3.1B, step 1). Currently, several commercially available methyltransferase enzymes can be utilized for such application. The previously mentioned M.CviPI methyltransferase modifies cytosines in a GpC context, whilst the M.SssI methyltransferase modifies cytosines in a CpG context. Alternatively, EcoGII methyltransferase methylates single adenines, but cannot be utilized in combination with a bisulfite conversion approach. However, the advent of novel sequencing platforms such as that of Oxford Nanopore Technologies (ONT), do not require a bisulfite conversion.

Under development since 1995, the ONT sequencing platform has matured into a commercially viable approach more recently. The principle of the ONT platform (referred to as nanopore sequencing hereafter) relies on two key biological proteins. Firstly, *Mycrobacterium smegmatis* porin A (MspA) is an interconnected octamer resembling a tube or goblet that is suspended in a lipid membrane (Wendell et al., 2009). When immersed in a conducting agent and a voltage is applied, an electric current can be measured due to the flow of ions through the pore. When DNA is passes through the pore, the flow of ions is modulated (at microsecond and pico-ampere levels) due to the specific characteristics of each nucleotide (Branton et al., 2008; Stoddart et al., 2009). For use in nanopore sequencing, the MspA pore was modified in order to improve translocation of DNA (Derrington et al., 2010). Secondly, the phi29 DNA polymerase that can control the rate at which DNA passes through the pore (Cherf et al., 2012; Manrao et al., 2012; Wendell et al., 2009). This polymerase captures the purified double-stranded DNA and feeds single-stranded DNA into the pore (Figure 1.3.1B, step 2 and 3). Taken together, DNA of any length can be added to the ONT platform. As there is no fragmentation, the information that is obtained from a single read (which can span over many kilobases) is per definition from a single cell. A single-stranded DNA enters the pore, thereby altering the ion flow, which is directly recorded. The absolute current and current change can be computationally deduced to inform on the DNA sequence (Clarke et al., 2009). Importantly, DNA modifications of accessible regions by e.g., methyltransferase enzymes alter the size and shape of the affected nucleotides. These modified nucleotides directly impact the current through the pore, thereby allowing the detection of modified nucleotides (Laszlo et al., 2013; Ni et al., 2019; Rand et al., 2017; Shim et al., 2013; Simpson et al., 2017). This results in a comprehensive map of accessible regions within a single DNA molecule that originates from a single cell (Figure 1.3.1A-B).

The advantages nanopore sequencing has over more conventional techniques are clear when it comes to read length and single-cell capabilities. However, a big limitation of nanopore sequencing is the relatively low read accuracy when compared to short-read sequencing. Increased read accuracy can be accomplished through computational improvements and/or pore chemistry. For example, an orthogonal technique by Pacific Biosciences (PacBio), termed single-molecule real time (SMRT) sequencing, utilizes both strands of the DNA molecule by circularizing the individual strands in an approach referred to as SMRT-bell. Each SMRT-bell is sequenced multiple times, resulting in higher accuracy when sequencing errors are random (Wenger et al., 2019). More recent releases of the ONT platform (R9.4, used throughout this thesis) use a mutated *Escherichia coli* CsgG pore, which further improve accuracies and pore stability (Carter and Hussain, 2017; Goyal et al., 2014).

### 1.3.3. A novel perspective on the chromatin landscape

With the recent rise of techniques that can map the chromatin landscape by modifying accessible regions comes a new way to look at nucleosome occupancy. Conventional approaches are often referred to as 'yield methods', due to their limited capabilities in mapping either nucleosomes (e.g. MNase-seq, ChIP-seq or chemical cleavage) or accessible regions (ATAC-seq). The resulting fragments derived from techniques such as e.g. MNase-seq correspond to the position of nucleosomes (Figure 1.3.2A). Mapping such fragments to the genome provides total number of nucleosomes per location. This approach results in a relative occupancy, and only informs on the number of nucleosomes at one position, relative to another. A further limitation of these yield methods is the inability to compare occupancy of individual cells over large(r) distances. Due to the fragmentation, contextual information within a cell is lost.

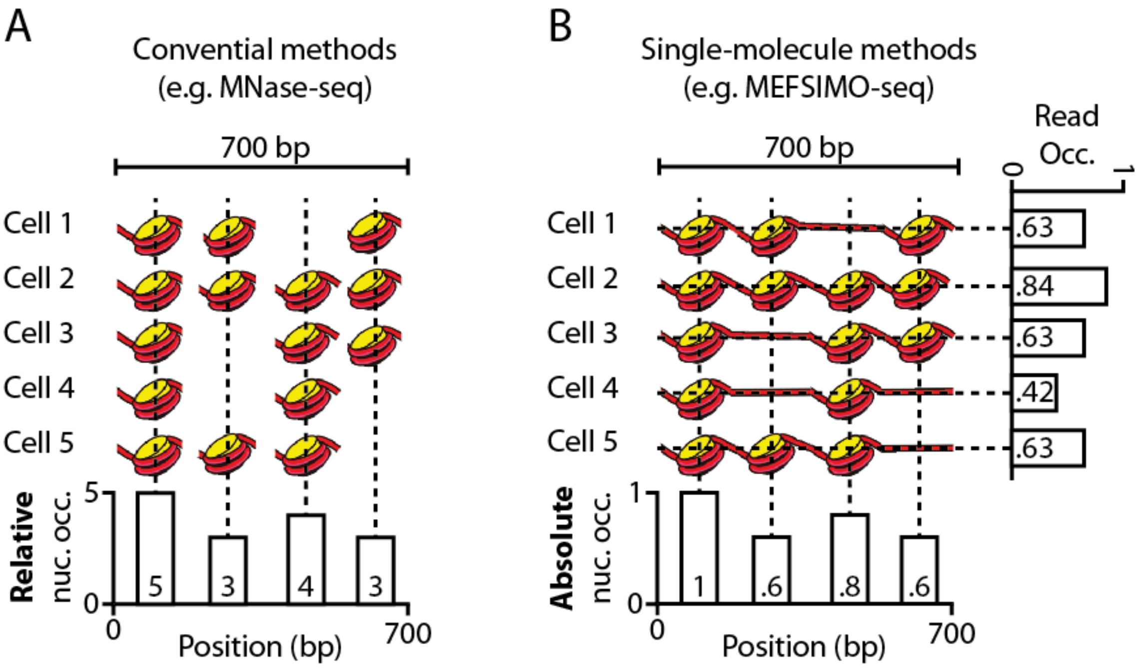


**Figure 1.3.2 Conventional yield-based methods versus single-molecule methods resulting in different occupancy readouts**

***(A)*** *Conventional methods rely on fragmentation and digestion of non-nucleosomal DNA by e.g. MNase. This process results in individual nucleosomal DNA fragments. These fragments are processed by short range sequencing methods (e.g. Illumina). This type of approach informs on the relative nucleosome occupancy (vertical dashed line) for the remaining DNA fragments. No information is retained from regions lacking a nucleosome. Contextual information between individual nucleosomes within a single cell is lost, thereby lacking any single-molecule information.* ***(B)*** *Single molecule methods such as MEFSIMO-seq do not rely on fragmentation. With this approach, cellular context of multiple nucleosomes is retaind. As no fragmentation has occurred, the absolute occupancy can be determined for each site (vertical dashed line). In addition, the read occupancy for individual fibers originating from a single cell can be determined (horizontal dashed line).*

To overcome this limitation, previous studies have implemented a novel sequencing platform, such as Nanopore sequencing, to detect the nucleosome position without fragmentation, and can thus determine the absolute occupancy (Oberbeckmann et al., 2019). Utilizing such an approach, the absolute nucleosome occupancy for any given position can be calculated as no fragmentation occurs during the experimental procedure. In addition, this approach allows the ability to visualize read heterogeneity and measure the read occupancy (Figure 1.3.2B). Employing methods that inform on the absolute occupancy and cell-to-cell heterogeneity can for example be used to directly elucidate differences between a lack of phasing or spacing (Figure 1.1.3D). Indeed, recent studies have started to visualize heterogeneity at the single-molecule level. For example, CTCF binding at a single genomic locus can be observed per cell and may influence neighboring array regularity (Stergachis et al., 2020). Others have shown a clear dichotomy of accessibility within the rDNA locus (Shipony et al., 2020). These and other studies provide a novel insight into the chromatin landscape.

### 1.4. Aims of this thesis

In this thesis, I first outline the development and characterization of a single-molecule sequencing technique that utilizes the ONT platform. Although recent publications have provided first insights into the chromatin landscape from a single-molecule perspective, experimental conditions to modify the accessible DNA vary widely. It is not clear what experimental conditions result in an optimal sample for sequencing. Moreover, well-defined parameters by which read-length, modification of accessible DNA and reproducibility are measured are not provided. In chapter 2.1, I investigated several methyltransferase enzymes and determine optimal conditions for enzymatic activity. I further describe how we established conditions that result in long reads, to take maximum advantage of the nanopore sequencing platform. I lastly describe the key parameters by which a sample can be judged before and after sequencing to acquire high reproducibility.

In Chapter 2.2, I utilize the sequencing approach to gain novel insights into the *S. cerevisiae* chromatin landscape. We utilize ultra-long reads over the rDNA locus to probe transcriptional dependency between loci of over 10.000 bp. We further investigate how heterogeneous the chromatin landscape truly is by utilizing single-molecule reads to visualize heterogeneity of individual nucleosomal arrays. I provide direct evidence which suggest that nearly all nucleosomal arrays have relatively high regularity, despite composite plots suggesting otherwise. Moreover, we probe the role and mechanism of several nucleosome remodelers, transcription and other factors to identify how they affect and regulate the chromatin landscape.

# 2. RESULTS

## 2.1. Chapter 1: Development and characterization of methylation footprinting followed by single-molecule sequencing (MEFSIMO-seq)

### 2.1.1. Background

**Abstract**

Our current understanding of the chromatin landscape relies on methods that require, or result in, the fragmentation of chromatin in order to map the location of nucleosomes and other bound factors. Such approaches fail to provide a detailed insight in cell-to-cell variability and contextual information on a single chromatin fiber. Emerging techniques such as Nanopore sequencing by Oxford Nanopore Technologies, provide the capabilities to sequence chromatin as a whole and thus eliminate the need for fragmentation. Our approach utilizes this technology and applies exogenous methyltransferases to probe accessible regions within the chromatin, thereby leaving a ‘footprint’ of nucleosomes. Together, we refer to this method as MEFSIMO-seq (methylation footprinting followed by single-molecule sequencing). In this chapter, we investigate several methyltransferase enzymes and find that the M.SssI CpG methyltransferase is currently most suitable for *in vivo* and *in vitro* applications. Moreover, we describe experimental conditions to ensure optimal methylation of solely accessible DNA, and prevention of undesired nuclease fragmentation. Lastly, we provide a set of observables by which experimental reproducibility can be determined. Taken together, we demonstrate that MEFSIMO-seq provides a stable and reproducible platform that allows us to map the chromatin landscape.

**Introduction**

In this chapter we establish MEFSIMO-seq, which allows us to probe, and ultimately map, the position of individual nucleosomes within the chromatin landscape. This novel approach is utilized to study the chromatin landscape in greater depth by elucidating cell-to-cell heterogeneity and single-molecule information. Prior to the advent of single-molecule long-range sequencing techniques such as MEFSIMO-seq, sequencing techniques relied on the digestion of accessible (i.e. non-nucleosomal) DNA. Techniques such as micrococcal nuclease sequencing (MNase-seq), DNase I hypersensitive sequencing (DNase-seq) and assay for transposase-accessible sequencing (ATAC-seq), fragment DNA to obtain the location of a nucleosome. Different approaches probe the nucleosome occupancy by methylome sequencing (NOME-seq) and utilizes a GpC methyltransferase (M.CviPI) to label accessible DNA (Kelly et al., 2012). This approach lacks the bias introduced by nucleases, but does require a bisulfite conversion (which includes PCR amplification), limiting the long-range capabilities drastically. Based on NOME-seq, single-molecule long-read accessible chromatin mapping sequencing assay (SMAC-seq) was developed. Whilst utilizing the same methyltransferase approach as NOME-seq, it applies Nanopore sequencing to obtain multi-kilobase sequence reads (Shipony et al., 2020). During the course of this study, several analogous techniques were published that similarly rely on the modification of accessible DNA using (combinations of) exogenous methyltransferases such as EcoGII ($m^6A$), M.CviPI (GpC), M.SssI (CpG) and Hia5 ($m^6A$) (Abdulhay et al., 2020; Lee et al., 2020; Liu et al., 2020; Oberbeckmann et al., 2019; Stergachis et al., 2020; Wang et al., 2019). Such techniques are in their fundamental principles similar to MEFSIMO-seq.

Despite the abundance of analogous techniques published during the course of this study, MEFSIMO-seq addresses several caveats not directly inherent in other approaches. We investigated several opportunities to assess quality of the biological sample before continuing with the modification of DNA. These steps provide high quality and consistent DNA integrity ensuring multi-kilobase reads. We describe quality controls and conditions that result in optimal methylation of accessible regions, thereby providing a best-case scenario for nucleosome detection. Lastly, we compare and provide parameters by which samples can be compared and any variability across samples be defined.

### 2.1.2. Establishing methylation using *in vitro* nucleosomal arrays

The ability to probe and visualize the chromatin landscape hinges on the optimal methylation of accessible regions. An increase in nucleotides that can be modified within an accessible region, will more accurately describe the size and location of such region. It simultaneously improves the accuracy of identifying and locating e.g. nucleosomes, by demarcating the borders of an inaccessible region. Whether a single enzyme or a cocktail of enzymes is used to methylate the DNA, incomplete methylation will reduce the accuracy in determining the position of a feature such as a nucleosome. Intuitively, utilizing multiple methyltransferase enzymes simultaneously, such as a CpG (M.SssI, 5mC modification) and GpC (M.CviPI, 5mC modification), will increase spatial resolution. However, optimal buffer conditions for either enzyme may (and do) differ, resulting in potential incomplete methylation by either enzyme (Figure 2.1.1A). Alternatively, enzymes such as EcoGII/Hia5 modify not only cytosines in a CpG/GpC dinucleotide conformation, but modify adenines ($m^6A$) indiscriminately. Utilization of these enzymes results in a higher density of modified nucleotides, but may limit the ultimate resolution due to other technical limitations (see Discussion related to this chapter). Here we explore different methyltransferase enzymes and report the conditions that will ensure optimal modification of accessible regions.

To initially rapidly test the methylation of DNA we utilized a naked (i.e. non-chromatinized) plasmid from which the backbone and insert were previously separated by restriction digest (Ludwigsen et al., 2018). This insert consists of 25, equally spaced Widom 601 sequences separated by a 50bp linker (Lowary and Widom, 1998). We devised of a method to test the methylation efficiency by utilizing methylation sensitive restriction enzymes, HpaII and HaeIII (CpG and GpC sensitive, respectively). We initially utilized restriction enzymes as it provides interpretable results quickly, and does not rely on the sequencing and subsequent computational analyses. We incubated the backbone and insert with either GpC (M.CviPI) or CpG (M.SssI) methyltransferase enzymes, utilizing their recommended buffers. After heat quench of the methylation reaction at several timepoints, we supplemented the reaction with the restriction enzyme (for details see Methods). We find that methylation provides full protection against the activity of the restriction enzymes within 180 minutes (Figure 2.1.1B). The results further indicate that we can probe the efficiency of methylation *in vitro* by a rapid restriction digest.

Next, we sought to find optimal conditions to methylate chromatinized genomic DNA, since the ultimate goal was to map the chromatin landscape *in vivo* (see Chapter 2). We realized that a heat quench could perturb the integrity of nucleosomes, revealing previously inaccessible restriction sites and thereby resulting in inaccurate findings. We therefore decided to quench the methylation reaction by the addition of S-(5’-Adenosyl)-L-homocysteine (SAH). The methylation of cytosines by M.CviPI and M.SssI relies on the addition of co-factor S-adenosylmethionine (SAM). SAM is bound by the

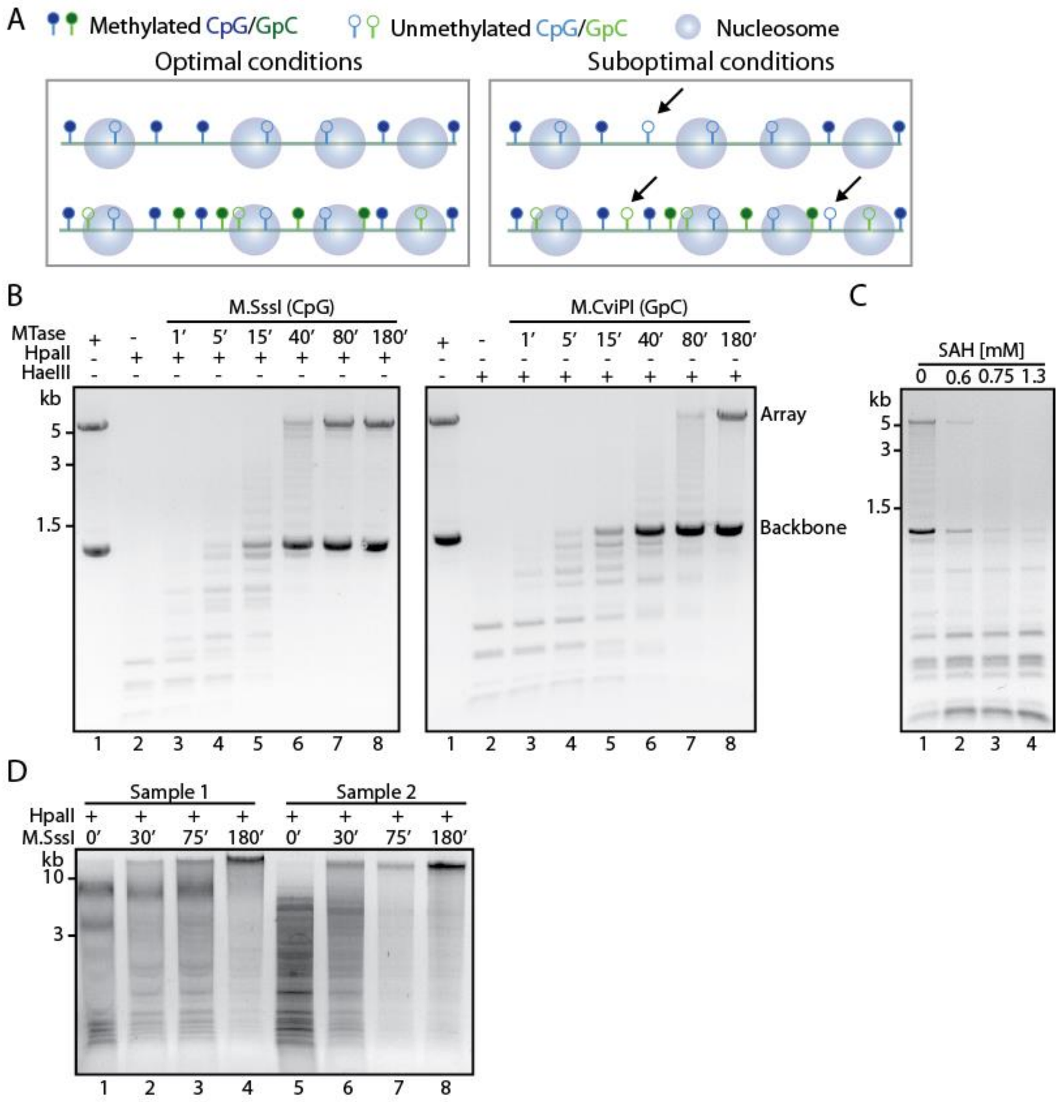


**Figure 2.1.1 Restriction-based assay to rapidly test methylation**

***(A)*** *Cartoon outline describing theoretical optimal and suboptimal conditions with a single or cocktail of enzymes. Optimal conditions will lead to a scenario in which nucleosome positions can be approximated most efficiently. Suboptimal conditions may occur due to reduced efficiency by buffer composition. This may lead to unmethylated sites (arrows) within accessible regions, complicating the approximation of nucleosomes or other factors.* ***(B)*** *Methylation of a naked in vitro array using two different methyltransferases protects against restriction enzyme digestion. Array and backbone remain intact after prolonged incubation (lane 8 versus 1). When methyltransferases have not fully modified all sites, digestion results in smaller fragments (lanes 2-5). Gel is representative of three replicates.* ***(C)*** *Addition of SAH can prevents methylation from occurring when added in 10-fold excess over 130-160 µM SAM (lane 4). Samples were incubated with one unit M.SssI methyltransferase together with indicated concentrations of SAH. Methylation was allowed to occur for 60 minutes after which HpaII was added. Digestion is largely prohibited when no SAH was added (lane 1). Gel is representative of two replicates.* ***(D)*** *Two samples (0.05 g yeast nuclei) of chromatinized genomic DNA were treated with 200 units M.SssI methyltransferase up to 180 minutes. At indicated timepoints, samples were quenched with SAH (3.2mM) and heat (65C), after which HpaII was added for 60 minutes. Methylation protects against digestion (lanes 4 and 8). Insufficient methylation results in (partial) digestion (lanes 1-3 and 5-7).*

methyltransferase which in turn results in the donation of a methyl group to the fifth atom of cytosine, resulting in a 5-methylcytosine (5mC). During the conversion to 5mC, SAH is formed by the removal of a methyl group from SAM, which acts as a strong inhibitor. Our results utilizing the *in vitro* array and backbone indicate that the addition SAH in a ten-fold concentration excess prevents any methylation from occurring (Figure 2.1.1C). Methylation was allowed to occur when no SAH was added (lane 1), which ultimately protects against HpaII digestion. Conversely, addition of ten-fold excess of SAH, rapidly prevents the methylation from occurring (lane 4), thereby preventing protection against HpaII digestion.

The restriction enzyme digest assay was next applied to probe methylation efficiency of the M.SssI methyltransferase on *in vivo* chromatin (Figure 2.1.1D). After extended incubation with the enzyme, we find that it provides protection against HpaII digestion for two independent samples. We do note some residual digestion at final timepoints (lanes 4 and 8). This may be due to cut sites located within fragile nucleosomes, transiently bound factors or an incomplete methylation. Due to the lack of a non-digested control within this experimental setup, we are unable to validate how the fragment distribution is when lacking HpaII. However, later experiment would suggest that undigested chromatin has a distinct high molecular weight band with little to no degradation (Figure 2.1.3A). Regardless, we conclude that methylation of accessible DNA markedly prevents digestion by the HpaII restriction enzyme. However, restriction- and agarose-based assays do not inform on the relative degree of methylation between enzymes. Moreover, these enzymes do not probe each CpG/GpC site that can be methylated by a methyltransferase enzyme. More quantitative analyses are required to directly address these issues.

### 2.1.3. Optimizing experimental methyltransferase conditions for *in vivo* applications

Detection of 5mC is possible by an enzyme-linked immunosorbent assay (ELISA). This plate-based assay non-discriminately and rapidly detects methylated cytosines by utilizing an Anti-5mC monoclonal antibody. A secondary HRP-conjugated antibody provides color after addition of a developer and is measured by absorbance at 410 nm (See Methods for full details). We first incubated 12.5 µg of naked gDNA with a single methyltransferase enzyme or a cocktail (40 units total, regardless of single/cocktail experiment) of both enzymes simultaneously (Figure 2.1.2B). For these reactions, we utilized a buffer that was deemed suitable for both enzymes based on previous findings by Shipony et al. We find that naked DNA samples reach a maximum absorbance within twenty minutes of the time course. (Figure 2.1.2B). This suggests that full methylation of all potential sites was obtained early within the time course. Although unlikely, the stable absorbance level at later time points could be due to a lack of SAM (or conversely, an abundance of SAH) or active enzyme in the reaction, preventing full methylation. However, excess SAM (600 µM) and enzyme (40 units) were used during these assays to circumvent these caveats. We next performed the same experiment on chromatinized gDNA. Similar to the naked DNA we observe a stable absorption level early in the time course (Figure 2.1.2A). However, due to the presence of potential fragile nucleosomes or nucleosome unwrapping, we cannot exclude that only highly accessible DNA was methylated.

Despite keeping DNA and enzyme concentrations similar across experiments, we observed different maximum absorbance levels within the experiments. Both naked and chromatinized DNA experiments suggests that a higher fraction of CpG sites, compared to GpC sites, were methylated. Notably, the *S. cerevisiae* genome contains a higher occurrence of GpCs (approx. 350.000 compared to 450.000 for CpG and GpC, respectively) suggesting that the buffer composition derived from Shipony et al. may not be optimal for the GpC methyltransferase. The cocktail of methyltransferases resulted unsurprisingly in the highest degree of methylation. Our preliminary findings thus suggest that a cocktail can improve total amount of methylated sites, but may be sub-optimal due to buffer composition. Taken together, the ELISA data can be used to rapidly probe the methylation degree. When utilizing a cocktail of methyltransferase enzymes, buffer conditions may need to be altered to allow optimal activity of all enzymes.

Based on the observations from our preliminary ELISA assays, we decided to utilize the single CpG methyltransferase and optimize conditions in the context of this enzyme. We hypothesized that we could better optimize conditions for a single enzyme. Moreover, existing workflows in the context of CpG modifications were already established, making it attractive to continue with this methyltransferase (See Discussion related to this chapter). We initially performed methylation of native genomic DNA with a CpG methyltransferase using a buffer used in similar applications (Darst et al., 2012). After performing standard library preparation procedures and sequencing (see Methods) we observed that methylation continued to increase with time (21% to 55%, 30 to 180 minutes, respectively. See Appendix Supplementary table 2 and Supplementary table 3 NPD03), seemingly without an end point (Figure 2.1.2C). Based on previous observations we identified that the M.SssI enzyme in the absence of $Mg^{2+}$ exhibited processive activity. Conversely, the presence of $Mg^{2+}$ results in a distributive activity of the enzyme (Matsuo et al., 1994). We hypothesized that the processive activity of the enzyme in the absence of $Mg^{2+}$ either evicts nucleosomes or is not hindered by its presence, thereby continuously increases the degree of methylation.

Due to the lack of a stable end point in our methylation reaction, an alternative approach was required to efficiently modify – and not artificially create or expand – native accessible regions. One approach is the addition of a crosslinking reagent such as formaldehyde during the preparation of nuclei. Nucleosomes, but in addition all other bound factors, are hereby immobilized on the DNA. The processive enzyme will no longer be able to evict, or pass through nucleosomes. We observed a stable end point of the reaction in our ELISA assay experiments and hypothesized that the buffer utilized here, which included 10mM $MgCl_2$, would prevent the methylation of inaccessible DNA. To this end we further explored conditions utilizing a buffer with 10 mM $MgCl_2$ and measured the absorbance over time using the ELISA assay. We find that middle (50 units) and high concentrations (150 units) of enzyme lead to similar absorption degrees after approximately 60 minutes (Figure 2.2.1B). Sequencing results reveal that high enzyme concentrations reflect global methylation levels (~20-25%) similar to that found in previous studies (Oberbeckmann et als., 2019) (Figure 2.1.2D). Importantly, we do not observe a continuous decrease in occupancy in these samples. We note that we observe a fraction of reads that are undermethylated (<10% methylated), suggesting partially incomplete methylation (see Appendix Supplementary table 2 and Supplementary table 3 NPD08). However, the observations made for high enzyme concentrations are consistent with later datasets (See Chapter 2.1.5).

Next, we performed a comparison between formaldehyde-crosslinked samples and samples that were left untreated during the preparation of nuclei. Briefly, the addition of crosslinking may be important to identify transiently bound factors (Chapter 2.2.12) or elucidating temporal differences in the chromatin landscape (Chapter 2.2.14). In order to provide a platform that is broadly applicable to different samples and applications, we aimed to obtain comparable data despite upstream sample treatment. We initially split a 1 L culture of log-phase growing *S. cerevisiae* in two batches, where we processed one with formaldehyde and left the other untreated. This initial experiment used a wild-type (WT) strain in which histones H3 and H4 can be selectively depleted and is therefore referred to as ‘WT-like’ when histones are not depleted (see Methods yeast strain list yFMP358). Nuclei

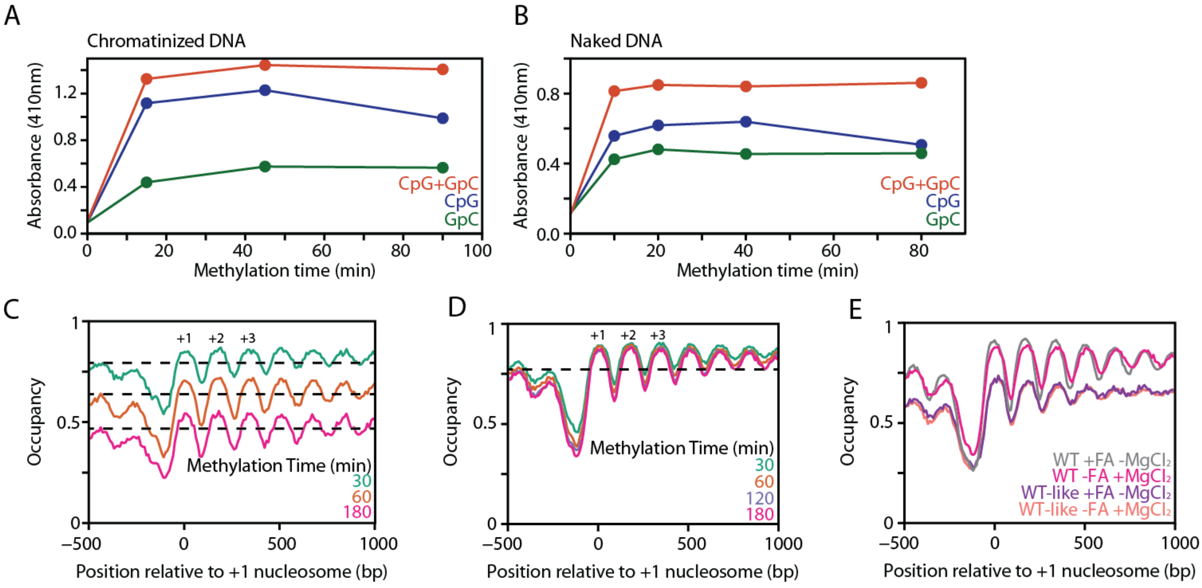


**Figure 2.1.2 Methyltransferase buffer conditions are critical to obtain an end-point reaction**

***(A)*** *Incubation of chromatinized genomic DNA (yFMP012) with a single (40 units) or cocktail of methyltransferase enzymes (M.SssI and M.CviPI, 20 units each). Isolated DNA (100 ng) was subjected to a 5mC ELISA assay (Zymo research) and absorbance (410 nm) was measured. Points are averages of two measurements within the same technical replicate. Value at timepoint zero is a single measurement within the time course. Differences between samples and points indicate relative methylation degrees.* ***(B)*** *Same ELISA assay as in* ***(A)*** *performed on naked genomic DNA.* ***(C)*** *Composite signal from raw MEFISMO-seq data after methylation of genomic DNA with M.SssI methyltransferase. Samples were incubated using a buffer based on Darst et al., lacking* $MgCl_2$*. Samples were incubated for indicated times before quenched with SAH. Dashed lines indicated global occupancy values for each timepoint. Y-axis displays global occupancy corresponding to 1 – mean site accessibility. Positions of +1/+2/+3 nucleosomal peaks are indicated above traces. For additional details, see Appendix Supplementary table 1, NPD03, WT0_XX samples.* ***(D)*** *Composite signal from raw MEFSIMO-seq data after methylation of genomic DNA with M.SssI methyltransferase. Samples were incubated using a buffer supplemented with 10mM* $MgCl_2$ *for indicated times before quenched with SAH. Positions of +1/+2/+3 nucleosomal peaks are indicated above traces. Dashed line corresponds to global occupancy of samples. Samples originate from NPD08 experiment (WT1_150_XX). See Appendix Supplementary table 1 for additional details.* ***(E)*** *Composite signal from raw MEFSIMO-seq data after methylation of genomic DNA with M.SssI methyltransferase. During preparation, samples fixed with formaldehyde (1%; +FA) or left untreated (-FA). During methylation, samples were incubated with identical buffers, but lacking 10mM* $MgCl_2$ *when left untreated. WT samples are biological replicates sequenced on separate dates. WT –FA sample is a copy of 180 minute sample in* ***(D)****. WT-like samples are similar to WT, but genotypically different (yFMP358, see Appendix strain table or Appendix Supplementary table 1 samples WT1_HDGAL_XX_90 of NPD11 experiment). These strains contain genic deletions for both copies of histone H3 and H4. Instead, a single copy is expressed under a Galactose promotor. WT-like samples were split during log-phase growth and either FA fixed or left untreated. WT-like samples were sequenced within same experiment.*

preparation and isolation was performed in parallel using identical procedures. From earlier observations, we determined that the addition of $MgCl_2$ was unfavorable for formaldehyde treated samples (Oberbeckmann et al., 2019). We therefore performed methylation using identical buffers, but in the absence or presence of 10 mM $MgCl_2$, depending on the upstream procedure. After sequencing and downstream processing, we calculated the average methylation per CpG site over the entire genome. The results using the WT-like strain indicate similar average CpG methylation degrees (39% vs 38% for with and without FA treatment, resp.) as well as genome wide average composite signals (Figure 2.1.2E). However, under non-depleted conditions, this WT-like strain natively exhibits slightly increased methylation compared to previous observations.

To further confirm that the addition of FA does not alter global methylation or composite signal, we compared two independent biological WT replicates (see Methods yeast strain list yFMP012). These samples, of which one was FA crosslinked, were prepared and sequenced independently on different dates. Similar to the WT-like data, the WT data suggest high similarities in terms of average site methylation (27% vs 23% for with and without FA treatment, resp.) and composite signal (Figure 2.1.2E). We do note that the biological independent WT strains differ to some extent, but this degree of variability is within experimental variability (Figure 2.2.1F). Based on our observations, we find that utilizing these controlled conditions will yield consistent data on the global chromatin landscape.

### 2.1.4. Nuclease activity affects long-range sequencing capabilities, mostly at highly accessible regions

The procedure we follow to isolate nuclei from yeast cells may lead to the residual presence of nucleases (See Methods). The activity of (most) nucleases requires the presence of metal ions, often $Mg^{2+}$ or $Zn^{2+}$. During the establishment of conditions that result in a stable end-point reaction, we identified that the inclusion of $MgCl_2$ in our methyltransferase reaction buffer was essential. Despite the favorable conditions for the methyltransferase that this addition resulted in, we noticed that on few occasions this resulted in a slight degradation of genomic DNA. Since our MEFSIMO-seq approach relies on long-range sequencing data to interpret the chromatin landscape, we sought to identify conditions that would result in genomic DNA unaffected by nucleases. Additionally, we describe the tools and observables to determine a potential nuclease digestion issue and provide an alternative approach that prevents nuclease activity.

A rough estimate on the quality of genomic DNA can be made by comparing the fragment pattern by standard agarose gel. After isolation of nuclei, we incubate samples over a period of time with the methyltransferase reaction buffer. Individual timepoints are quenched and genomic DNA is subsequently isolated (see Methods). For timepoints up to 90 minutes we observe fragments of >48.5 kb in length for a wild-type (WT) and a deletion strain (*isw1Δ, isw2Δ, chd1Δ,* referred to as TKO hereafter). After 150 minutes of incubation, we note a shift towards shorter fragment lengths for the TKO strain (Figure 2.1.3A, lanes 5-8). Additionally, we utilized a strain (yFMP358/yFMP359, see strain list) in which we can reduce global occupancy by selective inactivation of histone H3 and H4 generation (referred to as histone depleted or HD. See Appendix Supplementary table 1 NPD10). This results in a higher degree of accessible regions which are therefore more sensitive to nucleases. We observed a degradation of the high molecular weight fragments by agarose gel analysis in some of the HD strains (Figure 2.1.3A lanes 9-24). This degradation could be suggestive of nuclease activity as the effect was more severe in the HD strains.

Next, we repeated the experiment with an additional biological replicate of a depleted HD strain grown in glucose (2% w/v; yFMP358/359. See Appendix Supplementary table 1 WT_HDGLU_(FA)_30/90 samples from NPD11). We compared it to an aliquot of the sample that was formaldehyde treated during nuclei preparation. We argued that the crosslinking together with the absence of $MgCl_2$ in the buffer would abolish any degradation due to nucleases. Indeed, crosslinking seemingly prevented any (nuclease-induced) degradation (Figure 2.1.3B lanes 2 and 4). Degradation of the non-formaldehyde treated samples was severe (Figure 2.1.3B lanes 1 and 3). Based on these preliminary observations (and in line with previous findings by Oberbeckmann et al.), we concluded that nucleases could affect genomic DNA integrity. Importantly, we find that formaldehyde

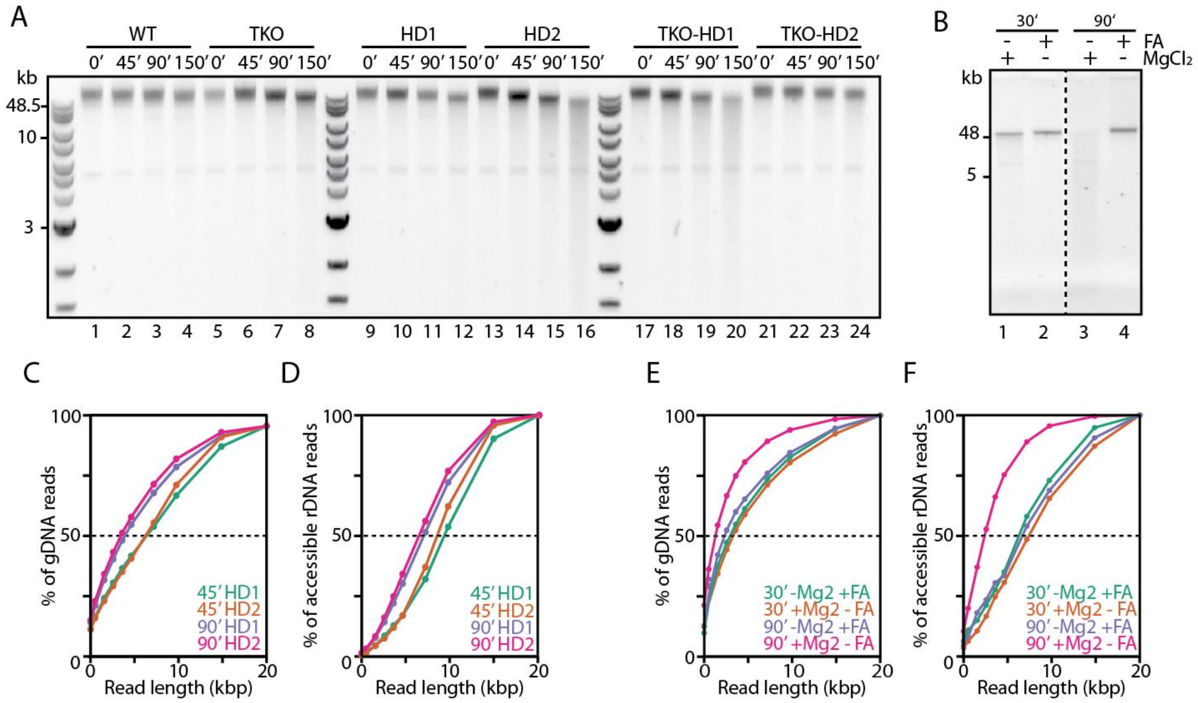


**Figure 2.1.3 Nuclease activity results in genomic DNA degradation and affects mostly highly accessible regions such as active rDNA loci**

***(A)*** *Chromatinized genomic DNA from wild-type (WT), triple knock-out (TKO), histone-depleted (HD) or TKO-HD strains. Biological replicates are indicated by the value after the strain indication. Values below strain indicate methyltransferase incubation time before quenching. Samples were all treated and processed identically downstream of methylation. Degradation of genomic DNA can be observed in certain samples (e.g. lanes 16 and 20). Unlabeled lanes are 1kb extended DNA ladders (NEB).* ***(B)*** *Comparison between samples with and without formaldehyde fixation during preparation of nuclei. Samples were methylated using identical conditions for indicated times. Buffers differed only by in-/exclusion of 10mM $MgCl_2$. Severe degradation is observed for samples without crosslinking (lane 3Such severe degradation is observed rarely, however. See Appendix Supplementary table 1-3, HDGAL_XX_30/90 samples from NPD11 for additional details.* ***(C)*** *Cumulative read length distribution of all genomic DNA reads of samples in* ***(A)****. Dashed line indicates median read length. A shift in read length is mostly observed for longer reads in the sample with severe degradation (pink line). Samples originate from NPD10 experiment (WT1/2_HD_45/90), see Appendix Supplementary table 3 for median genomic DNA values).* ***(D)*** *Cumulative read length distribution of accessible (>50% methylated) RDN37-1 reads of samples in* ***(A)****. Dashed line indicates median read length. A clear shift is observed for the sample with severe degradation (pink line).* ***(E)*** *Similar cumulative read length distribution of all genomic DNA reads as in* ***(C)****, but for samples in* ***(B)****.* ***(F)*** *Similar cumulative read length distribution of accessible RDN37-1 reads as in* ***(D)****, but for samples in* ***(B).***

crosslinking and the absence of $MgCl_2$ could prevent this degradation, at least in this scenario where degradation was severe without it. We do note that such severe degradation occurs rarely and is variable between samples of the same strain (e.g. Figure 2.1.3A lanes 9-16). This leaves the possibility open that DNA degradation is induced due to other (external) factors, and may not be directly due to nuclease activity.

We next sought to determine whether the perceived degradation by agarose gel truly originates from nuclease activity. To this end, we investigated features in the data that could reveal nuclease degradation. As stated earlier, nucleases favor accessible regions. Long reads of high accessibility and devoid of nucleosomes can be found at the rDNA locus of *S. cerevisiae* (Merz et al., 2008; Shipony et al., 2020). Moreover, we utilized strains in which we selectively reduced the global occupancy, thereby creating more nuclease sensitive regions. We hypothesized that under normal nucleosome density, accessible rDNA read length would be affected more severely compared to the genome wide average.

To this end, we first calculated the cumulative distribution of reads lengths globally for HD strains. Earlier, we already observed a slight degradation at later timepoints for these samples (Figure 2.1.3A, lane 9 – 16). We find that the degradation observed by agarose gel, correlates with a shift in cumulative read length distribution globally and does so as well for accessible rDNA reads (Figure 2.1.3C-D. See Appendix Supplementary table 3 NPD10). However, the effect is limited, as only mild degradation was observed between the 45 and 90 minute timepoints.

We performed the same analysis on the samples in which we saw a high degradation from 30 to 90 minutes when not formaldehyde-fixated (Figure 2.1.3B, lane 1 & 3. See Appendix Supplementary table 3 NPD11). Globally, we find that the read length distribution is affected when degradation is observed (Figure 2.1.3E). However, this effect may be damped due to the relatively high nucleosome density in these samples. For accessible rDNA reads, the degradation affects the read length distribution more than that is observed globally (Figure 2.1.3F). Samples that do not show a degradation over time, do not exhibit an altered rDNA read length distribution. Based on these findings, we conclude that nucleases are able to affect genomic DNA integrity, especially when highly accessible, and can result in shorter read lengths. However, severe degradation as observed in Figure 2.1.3B is only observed and analyzed for this single sample. Such degradation is generally not favorable and should be avoided by preparing new samples and/or performing formaldehyde crosslinking. The effect of nucleases on genomic DNA and specifically rDNA may require a closer look in the future if such degradation occurs more frequently under certain (mutant) conditions.

We note that there appears to be a discrepancy between genomic and rDNA median read lengths. Theoretically, these values should be identical as there is no know inherent bias for enhanced sequencing of this specific locus. However, as the read length is extrapolated from the mapped data it could alternatively be referred to as aligned read length. Read length is not determined by the actual length of DNA passing through a nanopore. We speculate that genomic reads may be artificially/computationally fragmented during mapping to provide an optimal alignment of our *S. cerevisiae* W303 data to the *S. cerevisiae* S288C reference genome. The W303 genome is known to contain ~9500 SNPs, and could thus prevent accurate mapping of all reads (more details in chapter 2.2.13) Currently, this issue remains speculative, but could be resolved by mapping our W303 reads to a W303 reference genome. For all further analyses we used the S288C (SacCer3) reference genome due to the wide availability of annotated loci and features such as NFRs, GRFs, TSSs, TTSs etc.

### 2.1.5. Observables by which consistency and quality of data can be determined

The observation that DNA and read length can be affected by nucleases is not favored in a technique that relies on long-range information. However, the absolute length of DNA may be variable across samples and experiments without necessarily affecting e.g. global occupancy, localizing nucleosomes or defining heterogeneity. We next aimed to define a set of observables by which samples can be compared in order to ensure reproducibility.

First, we determined the global occupancy measured over several WT and TKO biological replicates, all of which were treated near identically but on separate occasions. To this end, we calculate the total amount of methylated CpGs at each site over the sum of all methylated and unmethylated sites at that same location. An identical analysis was performed in a previous publication to determine global occupancy map for the *S. cerevisiae* genome (Oberbeckmann et al., 2019). During this analysis we discard any CpG site which cannot be unambiguously referred to as methylated or unmethylated. We find a global occupancy of 74±4% and 72±2% for WT and TKO, respectively (Figure 2.1.4A). This observation is akin to previously reported values by Oberbeckmann et al., despite determined from a different yeast variant (S288C by Oberbeckmann et al. versus W303 used here) and utilizing a slightly different approach. From a limited set of observations, we find that occupancy in formaldehyde (FA) crosslinking samples is largely similar to that in WT/TKO strains (Figure 2.1.4A FA samples). Thus, independent of strain or nuclei preparation, global occupancy is consistent throughout multiple samples and can be used to rapidly determine consistency.

Secondly, we investigated the occupancy per read as this observation can compare specific genomic regions if required. Here, we only consider reads that have a minimal length (≥1000 bp). Now, we can identify which, and how frequently reads are over- or under-methylated by isolating reads with a total methylation degree above 90% or below 10%. We applied these parameters as under-methylation could be indicative of incomplete methylation of accessible regions. This could be due to experimental conditions such as short incubation times, low enzyme concentration or non-spheroplasted nuclei. Conversely, over-methylation may be due to nucleosome eviction or dead cells. With few exceptions (3 out of 35 WT/TKO samples, Supplementary table 2) samples have low degrees (≤5%) of over-methylation, in agreement with our ELISA readings that suggest a steady end point. We do observe a larger fraction of samples (13 out of 35) exhibiting undermethylation (≥20%), and find that (chronologically) earlier timepoints and/or sequencing runs tend to have higher degrees of under-methylation. This is likely due to the slight increased concentration of methyltransferase enzyme in later experiments or improved spheroplasting. Without filtering any reads, we find that the variability is quite large. Applying the filter that removes over- and under-methylated reads reduces the variability and equalizes the means per sample (Figure 2.1.4B, Filtered versus Unfiltered). However, variability is still quite significant, likely due to the high accessibility heterogeneity in the general chromatin landscape and during cell cycle stages.

Thirdly, we calculate the methylation per read and plot the distribution, as it can reveal the degree of reads that fall within the filter parameters and whether these are an independent population. Under-methylated reads can clearly be identified as an independent population, which could justify the filter application when such reads are strictly undesired (Figure 2.1.4C, Filtered versus Unfiltered). We have not observed an impact by the in- or exclusion of these reads, and therefore do not filter in subsequent analyses. Taken together, average read methylation, fraction of over- and under-methylation and read methylation distribution can be used to compare consistency between samples.

Fourthly, we investigated the distribution of accessible and inaccessible regions. These regions are highly important for downstream analyses as they directly inform on the predicted location of nucleosomes and other bound factors (Chapter 2.2.5). Moreover, they are key in determining heterogeneity on a per read basis (Chapter 2.2.11). The distribution of inaccessible regions (hereafter referred to as nucleosome footprint) within gene bodies is highly similar between the WT samples. In WT strains we observe mono-/di-/tri-nucleosomal footprints of 164±4bp, 327±4bp and 485±8bp in length, respectively (Figure 2.1.4D, left). The length of inaccessible regions (hereafter referred to as linkers, includes NFRs) is 100±4bp (Figure 2.1.4D, right). We do not detect a significant difference in either nucleosomal footprint length or linker lengths when comparing genome wide to solely genic

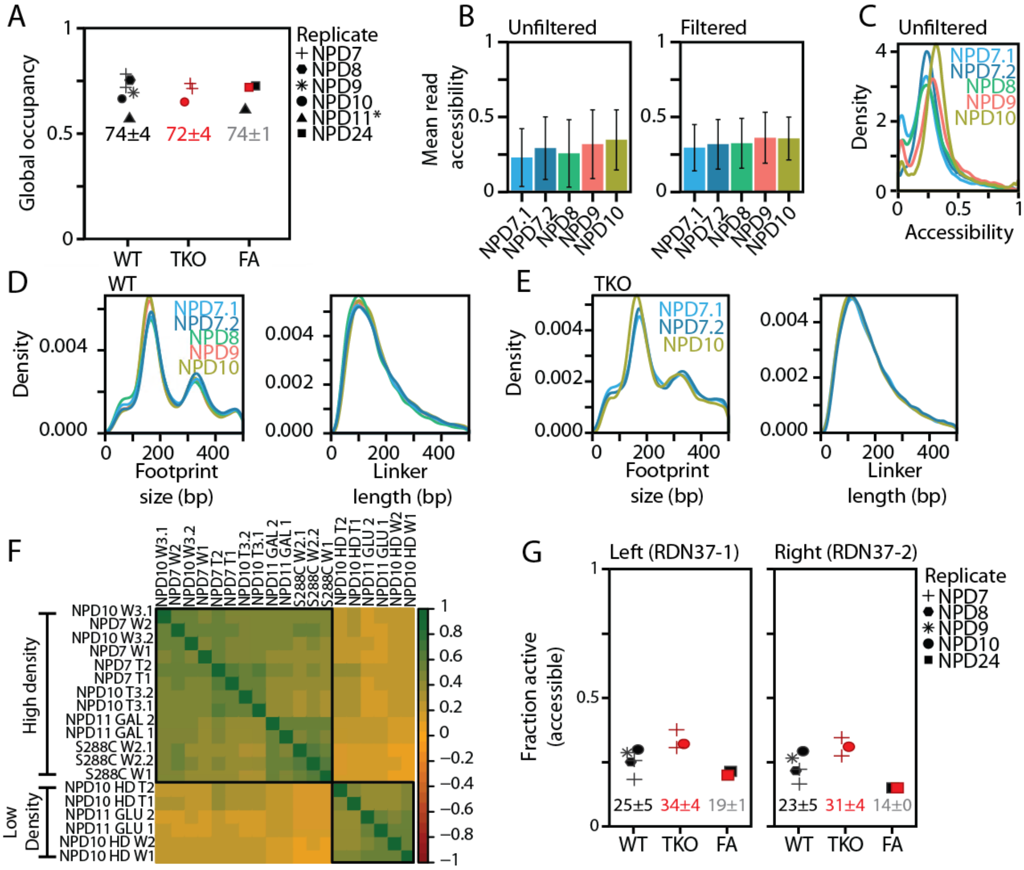


**Figure 2.1.4 Controlled experimental conditions result in consistency amongst biological replicates**

***(A)*** *Overview of global occupancy values measured from unfiltered MEFSIMO-seq data for the indicated sequencing experiments labelled as NPD and with different symbols. Values are calculated for wild-type (WT), triple knock-out (TKO) and formaldehyde (FA, either WT or TKO indicated by black and red, resp.) samples and are indicated below data points. Asterisk indicates a non-depleted WT-like sample in the yFMP358 background and was considered an outlier and not included in mean calculations. Samples used are final timepoints in their respective sequencing experiments. Values for individual samples can be found in Appendix Supplementary table 1.* ***(B)*** *Mean read methylation per sequencing run for WT samples of independent sequencing experiments. Unfiltered data represents all raw and mapped MEFSIMO-seq reads. Filtered data excludes reads with low (<10%) and high (>90%) methylation. Whiskers indicate standard deviations.* ***(C)*** *Read length distribution of unfiltered and filtered MEFSIMO-seq reads for WT samples from of independent sequencing experiments. A distinct population of undermethylated reads can be observed in unfiltered reads. These reads mainly consists of lowly (<10%) methylated reads and are absent when filtering is applied.* ***(D)*** *Left: Size of nucleosome footprint as measured by the distance between two methylated sites, separated by minimally one unmethylated sites. Independent sequencing experiments for WT samples are used as comparison. Peak densities indicate mono- di- and tri-nucleosomal fragments. Right: Size of linkers within genic regions as measured by the distance between two unmethylated sites, separated by minimally one methylated site.* ***(E)*** *Same as in* ***(D)*** *but for independent sequencing experiments of TKO strains.* ***(F)*** *Correlation between global occupancy per gene (n=1803) with a coverage of ≥6 and a gene length of >1000 bp. For each gene per sample, the mean methylation was calculated between the +1 nucleosome – 100 bp and TTS. Samples plotted are from S288C strain, non-histone-depleted (GAL; High density), histone depleted (HD or GLU; Low density) and multiple W303 WT (W) or TKO (T) strains. Samples sequenced by me are labelled with NPD. Samples labeled as S288C originate from Oberbeckmann et al. NPD10 W/T samples are 45- and 90-minute timepoints (3.1 and 3.2, resp.). Hierarchical clustering (black borders) order was applied using “complete” method. Original script by Dr. Michael Wolff, figure adapted and replotted by me.* ***(G)*** *Fraction of active DNA loci within different biological replicates and mutants similar to those in* ***(A)****. rDNA locus was identified as active when the mean accessibility over (part of) the particular locus was >70%. Fraction active was calculated by taking sum of active reads divided by total reads over the particular locus. Values below data points are mean and standard deviation. Mean values for individual samples can be found in Appendix Supplementary table 3.*

regions. Nucleosome footprint distribution is less defined in TKO strains, indicating more heterogeneity in the location of nucleosomes (Figure 2.1.4E, left). The distribution of nucleosomal footprints is akin to that observed when performing a limited MNase digest. Length of linkers is largely similar between WT and TKO, presumably due to the large spread observed in WT already (Figure 2.1.4E, right). The length and distribution of nucleosome footprints and linkers is thus a useful indication of consistency between samples. It describes the state of key observables used for many downstream analyses.

Next, we investigated whether the mean methylation degree of genic loci correlates well across biological replicates. To this end, we first calculate the mean methylation for each individual genic locus within a sample (+1 nucleosome – 100 bp to TTS). Next, we correlated the genic values from one sample to that of other samples. Lastly, the mean correlation value between each sample was calculated and used as an indication of consistency (Figure 2.1.4E). The datasets include several high density strains including *S. cerevisiae* variant strain S288C (BY4741), WT-like strains (genotypically nearly identical to W303) and TKO strains. In addition, we also include several independent biologically replicated histone depleted (HD) samples. We find a relatively high correlation between all strains that exhibit high nucleosome density (global occupancy ~70-75%), with a correlation coefficient of 0.4 – 0.6. These samples cluster together when a typical hierarchical clustering is applied. This indicates that despite fluctuations in global occupancy, the mean methylation per gene is relatively similar (Figure 2.1.4E). Histone-depleted strains form a separate cluster and do not correlate well to regular occupancy strains. Moreover, their correlation amongst each other is relatively low, suggesting a wider variety in occupancy between genes of different samples. These observations for HD strains suggest that: (1) the nucleosome depletion does not reduce nucleosome density equally over all genes, some genes might be more/less affected than others, and (2) it is seemingly randomly lowered across HD strains. With these finding we further validate high reproducibility amongst a wide variety of samples based on the occupancy of each gene.

As a last validation of reproducibility, we determined the fraction of active (i.e. highly accessible) rDNA reads within each sample as this should be consistent within a strain. As discussed later in chapter 2, and previously shown by similar methods in Shipony et al., the rDNA locus exhibits two mutually exclusive populations. Although this view is contested, we identify the population with high accessibility as transcriptionally active as nucleosomes are evicted by the Pol I transcription machinery, leaving the locus devoid (Conconi et al., 1989; Dammann et al., 1995; Jones et al., 2007). The fraction of active rDNA loci was determined by the amount of reads that (partially) cover the locus exhibiting >70% accessibility over the total amount of reads of said locus. We performed this analysis both for the RDN37-1 and RDN37-2 loci (left and right, respectively), which should exhibit identical characteristics as their sequences are copies in the reference genome. Our analysis indicates little variability across biological replicates of the same strain (Figure 2.1.4G). For WT samples we find that both left and right rDNA loci are active in 23-25±5% of cases. We observe slight mean increase in fraction of active loci in the TKO strain to 31-34±4%, but across replicates this value does not deviate more than seen in WT. FA-treated samples exhibit a lowered fraction of active reads. This reduced accessibility could be due to the crosslinking of Pol I to DNA in active reads, reducing the overall mean accessibility. Regardless, across the limited FA-treated replicates, we do not observe any differences. Taken together, samples exhibit a similar degree of active rDNA loci when nuclei are prepared in similar fashion, demonstrating the reproducibility across samples. However, care should be taken when investigating growth rates or working with non-log-phase growing cells. Under these circumstances, values may be different and more variable as rDNA transcription is coupled to growth rates (reviewed in Dai and Zhu, 2020).

### 2.1.6. Discussion (related to this chapter)

#### 2.1.6.1. Alternative DNA modifying enzymes and detection methods

MEFSIMO-seq utilizes a single methyltransferase enzyme (M.SssI) in order to modify accessible cytosines found in a CpG context. During the exploratory phase of development, we have investigated the possibility to utilize a cocktail of enzymes, namely M.SssI and M.CviPI (Figure 2.1.2A). We concluded that optimizing conditions surrounding a single enzyme would be most beneficial. Curiously, out of the commercially available enzymes, M.SssI has the lowest theoretical resolution due to the relative low abundance of CpG sites in the genome. GpC sites, or single adenine sites are much more abundant and can be modified using commercially available enzymes. A key factor in the decision to use the M.SssI enzyme, was the ability to accurately detect the 5mC modification in a CpG context. The Nanopolish software was used to detect modified nucleotides and deemed most optimal for detection of CpG sites (Simpson et al., 2017). Recent developments have opened new avenues by implementation of better training data that allows the detection of CpG and GpC modified sites simultaneously (https://github.com/jts/nanopolish). In addition, alternative software such as Tombo and Deepsignal have been developed which can detect modified nucleotides as well (Ni et al., 2019; Stoiber et al., 2017). These alternatives may be an improvement, or can be used in tandem with existing pipelines to increase accuracy. A recent study has already implemented the detection of modified CpG, GpC and adenine sites simultaneously using Tombo software (Shipony et al., 2020). Alternative DNA modifications may improve the resolution or detection accuracy. The methyltransferases mentioned earlier, generate 5mC and $m^6A$ modifications. However, 4mC, 5-hydroxymethyluracil and cytidine deamination are known nucleotide modifications that have been detected natively (Kawasaki et al., 2017; Salter et al., 2016; Timinskas et al., 1995). It would be interesting to investigate the capabilities and implementation of such modifying enzymes for even higher accuracy in detecting (in)accessible regions. These findings re-open the possibilities to further improve MEFSIMO-seq.

#### 2.1.6.2. Steps towards ultra long-range sequencing

The multi-kilobase reads produced by the Nanopore sequencing platform is several orders of magnitude greater than conventional sequencing approaches. In our results we routinely find median read lengths between 5-10 kilobases (See Appendix Supplementary table 3). We however also observe a significant degree of variability between samples. This variability is likely introduced by the presence of active nucleases (Figure 2.1.3D). Many analyses performed throughout this thesis are not affected by this variability. However, future endeavors may require consistently long reads, or reads of even greater length. For example, exploring the transcriptional activity of >2 rDNA loci simultaneously would be possible if read lengths increase. Alternatively, one could probe chromatin organization as identified by chromatin conformation capture techniques at the single-cell level.

We explored the implementation of formaldehyde crosslinking and found that this likely reduces nuclease activity, thereby reducing variability. However, formaldehyde crosslinking has several caveats which may limit the ultimate read length. One limitation is the reversal of the crosslinked sample, which requires extended incubation at 60°C. The elevated temperatures may induce breaks or nicks, which ultimately result in a limited read length. Moreover, the act of crosslinking on its own may induce single-stranded DNA breaks (Grafstrom et al., 1983). Formaldehyde crosslinking furthermore induces DNA modifications such as N6-hydroxymethyldeoxyadenosine and N2-hydroxymethyldeoxyguanosine (Kawanishi et al., 2014). Whilst these modifications may not directly impact read length, they could affect detection of modified bases using existing software. To improve consistency and median read length simultaneously, one would require an approach with the least amount of perturbations by external forces.

Preparation of spheroplasts from *S. cerevisiae* cells may be alternatively or additionally be improved to reduce the presence of nucleases. Our current approach does not explicitly prevent the carryover of cytoplasmic nucleases during the nuclei purification stage. Implementing multiple Ficoll gradients during the isolation of nuclei may reduce cytoplasmic contaminations. Shorter incubations with higher concentration of methyltransferases might improve read lengths. This reduced incubation time can help to reduce the effect of any nucleases that may be present. Reducing the total amount of nuclei to a minimal required amount and incubation in a larger reaction volume would further dilute nucleases. Ultimately, the read length may be physically constrained by the sequencing platform and commercially available protocols. For example, current recommended library preparation protocols (provided by Oxford Nanopore Technologies) require several purification steps using AMPure XP beads. These inevitably result in shearing of long fragments due to washing and elution. Designing a custom library preparation protocol may improve final read lengths, but would need extensive testing as to maintain yields. In conclusion, our results suggest relatively high consistency between samples in terms of read length and are in line with other reports. However, improvements and alterations such as the ones suggested above are required if longer reads are desired.

#### 2.1.6.3. Alternative long-range sequencing approaches

The Oxford Nanopore Technologies sequencing platform provides excellent results with regards to throughput, ease of use and quality of data. However, the biological nature of the sequencing approach, i.e. an unwinding enzyme and protein pore, imposes certain limitations (Laszlo et al., 2013). Moreover, software-based modified nucleotide detection developed for use with Nanopore sequencing, has its own limitations. For example, current nanopolish software is trained on either fully methylated or fully unmethylated (5mC) data. The two different methylation states cannot be distinguished when they occur simultaneously within a 6-mer. This results in a single methylation state when two CpG sites are within 10bp of each other (Simpson et al., 2017). Our data consists of significant amounts of grouped sites, and lack of additional training data has not decreased the amount over the duration of this project. Additionally, Nanopore sequencing detects only a single DNA strand, discarding the complementary strand. This may result in ambiguity as e.g. stalling can affect accurate detection. Current theoretical raw read accuracies are approximately 95%, whereas more traditional methods such as short read Illumina sequencing having >99.9% accuracy.

Improving the output and detection accuracy of (modified) bases may require an alternative sequencing approach altogether. As an alternative to Nanopore sequencing, Pacific BioSciences (PacBio) utilizes a different approach to sequence long reads. Here, instead of relying on a biological protein pore, single-stranded DNA diffuses into a physical sequencing unit referred to as a zero-mode waveguide (ZMW). Within the ZMW, a polymerase attaches fluorescently labelled nucleotides that produce a light pulse when incorporated. Each of the four nucleotides has a distinct label, allowing the detection of the sequence (Eid et al., 2009). Modified nucleotides are detected due to a prolonged incorporation time between nucleotides (Rhoads and Au, 2015). The addition of circular consensus sequencing (CCS) further improves basecalling accuracies to similar levels as short read Illumina sequencing (Travers et al., 2010; Vollger et al., 2020). The PacBio approach is not without its own limitations. Read length is on average shorter and output (amount of sequence nucleotides) is less compared to the Oxford Nanopore Technologies platform. Thus, MEFSIMO-seq in combination with the Nanopore sequencing platform has revealed features of the chromatin landscape that were previously unobtainable (see Chapter 2.2). However, pushing beyond the current limits and exploring other (larger) genomes, may require the addition, or shift to a different sequencing platform.

## 2.2. Chapter 2: Mapping the *in vivo* landscape by single-molecule sequencing

### 2.2.1. Background

#### Abstract

The chromatin landscape is essential for proper gene regulation and cellular homeostasis. This landscape relies on many factors which can shape individual chromatin fibers over long distances on a cell-to-cell level. Yet, our understanding of the chromatin organization often relies on fragmented samples and bulk processing. Here, we utilize a technique that probes accessible chromatin regions by labeling with an exogenous CpG methyltransferase, creating a footprint of inaccessible regions. This procedure is followed by direct sequencing using the Oxford Nanopore Technologies (ONT) platform to acquire (methylation) information on intact, multi-kilobase DNA fibers. We refer to this approach as methylation footprinting followed by single-molecule sequencing (MEFSIMO-seq) throughout this thesis. We leverage this approach to refute transcriptional independence of the rDNA loci. We also find that the nucleosome landscape is even more regular than previously observed. Furthermore, we are able to directly show that chromatin remodelers affect array regularity genome wide and further investigate *in vitro* hypotheses regarding their mechanisms. Together, our results provide novel insights into the chromatin landscape and the factors that shape it.

#### Introduction

The global architecture of the chromatin landscape in eukaryotes is well defined. Nucleosomes are, often actively, placed at regular intervals over genic regions. Here they are aligned to the transcription start site (TSS), in a process referred to as phasing. The primary nucleosome, closest to the TSS, is assigned as the +1 nucleosome and downstream nucleosomes are assigned +2, +3 etc. Upstream of the +1 nucleosome of active genes a nucleosome free region (NFR) is often found, which is both passively and actively kept devoid of nucleosomes. This region is therefore accessible for factors such as the transcription machinery and other general regulatory factors. Inactive genes often have an inaccessible NFR occupied by a nucleosome, and thereby prevent transcription factors from binding. The NFR is demarcated by the -1 nucleosome, which together with the +1 nucleosome define the NFR width. This width is modulated by factors such as nucleosome remodelers and wider NFRs are often associated with higher transcription.

The role of nucleosome remodelers is key in establishing phased and regular arrays. Deletion of remodelers from the ISWI- and CHD families results in a detrimental reduction of phasing and regularity (Gkikopoulos et al., 2011; Ocampo et al., 2016). Recent findings show a further reduction when the INO80 remodeler is depleted (Singh et al., 2021). Despite the high regularity imposed by these remodelers as observed by a genome wide average, heterogeneity is inevitable in any living organism. Certain genes tend to have a more regular nucleosomal array distribution than others. The position of nucleosomes is often determined by enzymatic digestion of non-nucleosomal DNA by micrococcal nuclease (MNase). This digestion results in fragmentation of the DNA, but leaves nucleosomal DNA of which the origin can be determined after MNase-seq. MNase-seq can efficiently map the average position of nucleosome positions but due to fragmentation no information on the long-range organization is retained. Formally, the possibility remains that regular arrays are prevalent throughout the genome, but are obscured by a lack of phasing or inherent heterogeneity. Such observations require a technique which can assess the array architecture on a cell-to-cell basis, ideally whilst retaining the original physical organization. Recent developments in sequencing techniques by ONT and PacBio, together with exogenous methylation of accessible DNA, have allowed single-molecule insights into the chromatin landscape (Abdulhay et al., 2020; Lee et al., 2020; Liu et al., 2020; Oberbeckmann et al., 2019; Shipony et al., 2020; Stergachis et al., 2020; Wang et al., 2019). Using MEFSIMO-seq, analogous to previously published approaches, we further dissect the chromatin landscape and the role of nucleosome remodelers.

Nucleosome remodelers are not the sole factors influencing the chromatin landscape. Recent findings show that the RNA Pol II transcription has a profound effect on nucleosomal arrays *in vivo* (Singh et al., 2021). Moreover, nucleosome integrity and occurrence may be affected by transcriptional activity resulting in subnucleosomes and/or ‘fragile’ nucleosomes (Kulaeva et al., 2010; Ramachandran et al., 2017). Regions such as the rDNA locus, have been suggested to be completely devoid of any nucleosomes when actively transcribed by RNA Pol I (Conconi et al., 1989). The accessibility of the promoter together with the position +1 nucleosome may stimulate transcription. Moreover, the factors associated with the promotor determine how transcription of genes is regulated (Rossi et al., 2021). Interplay between transcription and chromatin architecture is widely studied, but often poorly understood. Here, we probe how transcription is correlated to chromatin architecture.

### 2.2.2. MEFSIMO-seq discerns chromatin features with high reproducibility

Methylation Footprinting followed by SIngle-MOlecule sequencing (MEFSIMO-seq) is analogous to previously published techniques such as nanoNOME, SMAC-seq, Fiber-seq, MeSMLR-seq, ODM-seq and SAMOSA (Abdulhay et al., 2020; Lee et al., 2020; Liu et al., 2020; Oberbeckmann et al., 2019; Shipony et al., 2020; Stergachis et al., 2020; Wang et al., 2019). Similarly, the MEFSIMO-seq approach relies on the preferential methylation of accessible cytosines using a CpG-specific 5mC methyltransferase (M.SssI). Utilizing the Oxford Nanopore Technologies platform allows us to sequence multi kilobase single-molecule chromatin fibers. Detection of methylated CpG sites is handled by the previously established software tool nanopolish (Simpson et al., 2017). Taken together, MEFSIMO-seq results in a comprehensive map of the genome where accessible regions are methylated and nucleosomes leave an unmethylated inaccessible footprint (Figure 2.2.1A).

As previous methods have demonstrated, the CpG-specific 5mC methyltransferase preferentially methylates accessible DNA. However, the activity of the enzymes is dependent on magnesium, where the addition of magnesium results in a distributive, rather than processive activity (Matsuo et al., 1994). Without the addition of magnesium or formaldehyde for crosslinking, we identified that the enzyme would methylate inaccessible CpG-sites over time. To establish a robust end-point where exclusively accessible sites are methylated, we performed time course experiments with different concentrations of the methyltransferase (MTase) enzyme, M.SssI. Genomic DNA was methylated for increasing lengths of time, and aliquots were taken at set intervals. The gDNA of individual timepoints was isolated and the degree of methylation was determined using an ELISA assay. At low (25U; referred to as WT_25_XX, where XX indicates time before quenching the reaction) enzyme concentration, methylation degree continued to increase and saturation was not observed. At middle to high (50U and 150U, respectively. Referred to as WT_50_XX or WT_150_XX.) enzyme concentrations, we observed little to no increase in absorbance after 60 minutes (Figure 2.2.1B). The results suggest that methylation of accessible regions was saturated under these conditions by the methyltransferase enzyme.

We performed the methylation on samples that were not crosslinked during the preparation. We hypothesized that exclusion of crosslinking would provide a snapshot of the true nucleosome landscape. Transient or other weakly bound factors will not leave an inaccessible footprint. Stable bound factors, such as nucleosomes, will leave a footprint by preventing the methyltransferases access to the DNA. Crosslinking does have advantages, such as preventing endo- and exonucleases

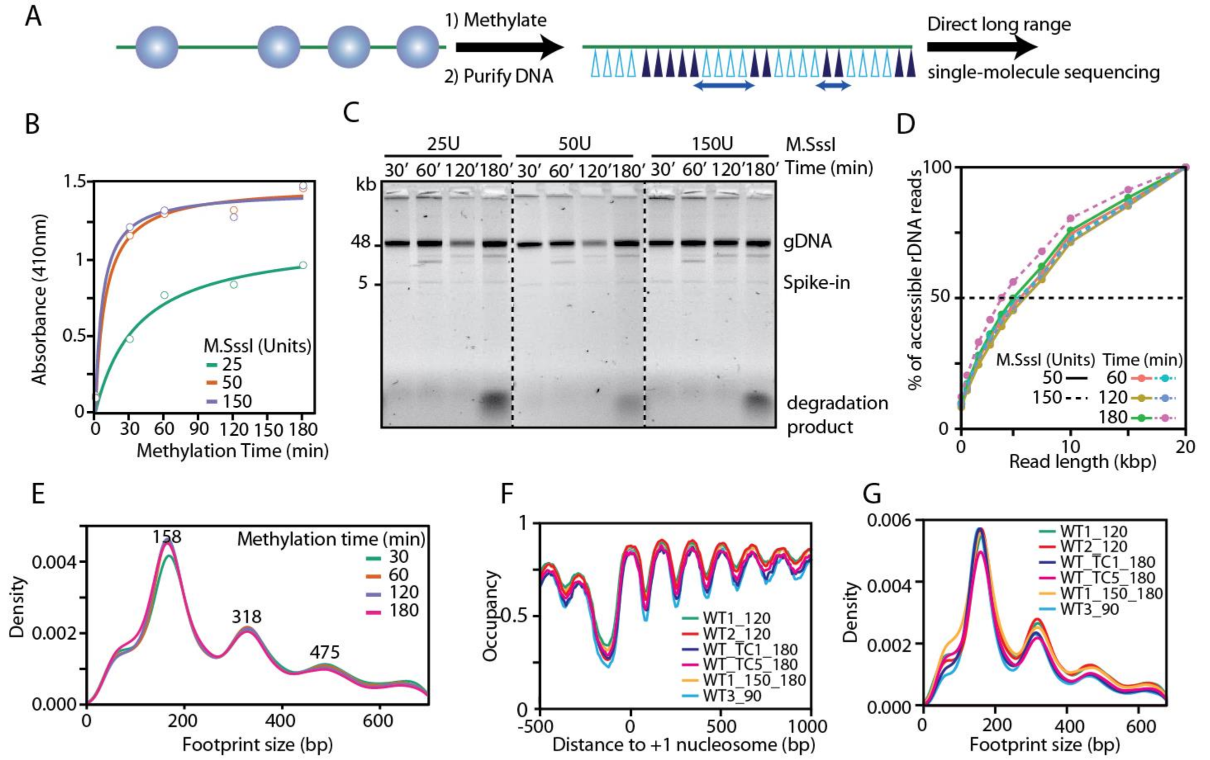


**Figure 2.2.1 Tightly controlled experimental conditions result in highly reproducible MEFSIMO-seq data**

***(A)** Cartoon outline of MEFSIMO-seq. DNA (green) and histone octamers (blue) together represent a chromatin fiber. The individual chromatin fibers are methylated by a CpG-specific 5mC methyltransferase (M.SssI). Next, the DNA is purified, resulting in accessible methylated regions referred to as linkers. Conversely, inaccessible unmethylated regions represent the nucleosome footprint. Lastly, the DNA is directly sequenced using the ONT platform and modified cytosines are detected using Nanopolish software. **(B)** Incubation of equal amounts (25 µg) of chromatin with low, medium and high (25U, 50U, 150U, resp. See Appendix run NPD08) concentration units of M.SssI. gDNA is subjected to an ELISA assay which detects 5mC. Quantification of 5mC methylation is measured by absorbance at 410nm. Points are average from two measurements. **(C)** gDNA integrity of samples incubated with low, medium or high MTase concentrations over a 180 minute timecourse. Naked and assembled chromatin spike-ins were utilized throughout this experiment. A degradation product of (presumably) naked chromatin can be observed at the 180′ timepoint. Identical samples to **(B)**. **(D)** Cumulative distribution of global read length of all gDNA to validate gDNA integrity. Dashed horizontal line indicates median read length of samples (See Appendix Supplementary table 3 NPD08 for median values). Solid traces are samples incubated with 50U and dashed traces are incubated with 150U of methyltransferase **(E)** Nucleosome footprints (i.e. inaccessible regions) are near identical over the incubation period with 150U of M.SssI indicating that the methyltransferase enzyme does not gain access to the inaccessible regions over time. peaks of mono-, di- and trinucleosomal peaks are noted above the respective peaks. **(F)** Genome wide composite plots from three independent biological replicates (WT/WT1, WT2 and WT3) over four different sequencing runs (NPD07-NPD10). See Appendix Supplementary table 1 and 2 for further details on global occupancy levels. **(G)** Nucleosome footprint (i.e. inaccessible region) distribution comparison of the samples described in **(F)**.*

from digesting the DNA (see Chapter 2.1.4), or when capturing transient factors is essential for the experiment (see Chapter 2.2.12). However when crosslinking is performed, methylation should be performed in a buffer that lacks magnesium, as the inclusion prevents complete methylation of accessible regions (Oberbeckmann et al., 2019).

After performing the methylation, we validated the DNA integrity by gel electrophoresis as this is typically a first indication for stable DNA fragment length (Figure 2.2.1C). We observe a stable, high molecular weight band, for all three samples (25U, 50U and 150U) incubated with different concentrations of methyltransferase enzyme over the timecourse (30 to 180 minutes). A non-native chromatinized DNA spike-in of 5kb in length was added at timepoint zero and shows no degradation over time. However, after 60 minutes into the timecourse, we added an additional non-native spike-in of 'naked' DNA (~15 kb; pFMP503). We observe a degradation product at the final timepoints of each sample, potentially the result from nuclease degradation of the naked DNA spike-in. However, we have not been able to directly confirm that the degradation product seen here originates from this spike in. Other similar samples that were not supplemented with this spike-in do not show this degradation. As the samples do not show indications of affected gDNA integrity, we continue to sequencing.

Utilizing the acquired sequencing data, we first probed the integrity of accessible rDNA reads over the timecourse to determine global sample integrity using the sequencing data. We find that the median fragment length of accessible rDNA reads is near identical between the three latest timepoints at middle and high enzyme concentrations (Figure 2.2.1D; See Appendix Supplementary table 3 NPD08). In line with the ELISA assay performed earlier, we find that a low MTase concentration (25 U) results in a maximum global occupancy of 89% and an undermethylation of 54%. These observations clearly suggest that this sample was not optimally methylated and is therefore not included in subsequent analyses. The sample incubated with 50 U and 150 U of MTase displays an 82% and 77% global occupancy in addition to an 26% and 16% undermethylation at the final timepoint. This suggests that the increased MTase concentration largely alleviates any potential undermethylation and results in a global occupancy akin to that of other biological replicates (For all values see Appendix Supplementary table 1-2, NPD08). These sequencing results, in combination with the observations by gel electrophoresis, suggest that nucleases did not affect gDNA integrity and that high (and potentially also middle) MTase concentrations result in fully methylated DNA of accessible regions. For these formerly mentioned reasons, we decided to continue the analyses with the high MTase sample (150 U), as these conditions match closest to any previous and subsequent biological replicates.

The goal of incubating chromatin with methyltransferases is to methylate accessible regions, thereby allowing the identification of inaccessible regions. These inaccessible regions are most likely DNA that is wrapped around a (stable) histone octamer and is typically considered to be 146bp in length (Luger et al., 1997). The footprint left by the inaccessible regions (also referred to as the nucleosome footprint) is defined by measuring the distance between two methylated sites, with at least one unmethylated site separating them. We find that, in the sample incubated with 150 U of methyltransferase, mononucleosomal footprints are on average 158 bp wide (Figure 2.2.1E). The larger than expected footprint is likely due to the genome wide distribution of CpG sites resulting in a limited resolution. A lack of CpG sites directly before and after the nucleosome entry- and exit site prevents the detection of the true nucleosome footprint. We also observe peaks at 318 bp and 475 bp, which correspond to di- and trinucleosomal fragments. Due to the limited resolution and/or sequencing error, not every region between nucleosomes (i.e. linkers) have a detectable CpG site. Di- and trinucleosomal fragments (318 bp and 475 bp, respectively) provide a better estimate of the actual nucleosomal DNA length and linkers, presumably due to a diminishable resolution error at longer lengths.

Lastly, we validated the robustness of our protocol by comparing results of four separate MEFSIMO-seq experiments (NPD07-NPD10) spanning three biologically independent replicates (WT/WT1, WT2, WT3) of wild-type (WT) *S. cerevisiae* cells. For all samples, we performed similar quality controls as described before and experimental conditions were near-identical to those described earlier. Values for the individual samples can be found in Appendix Supplementary table 2. We find minor variations of mean global occupancy (Appendix Supplementary table 1) and genome wide average composite signal between sequencing experiments are highly similar (Figure 2.2.1F). Moreover, we observe that mono-, di- and trinucleosomal peaks are present at near identical locations across the evaluated samples (Figure 2.2.1G). Taken together, these findings suggest a high reproducibility across all samples and can reliably be utilized for downstream analyses.

### 2.2.3. MEFSIMO-seq faithfully captures the chromatin landscape *in vitro* and *in vivo*

Having validated the technical aspects to ensure optimal experimental throughput, we explored the capabilities of our sequencing approach. First, we validated the capabilities of detecting individual nucleosomes on an *in vitro* salt gradient dialysis (SGD) assembled array. This array consists of 12, unmodified Widom 601 sequences and a subsequent 13 modified 601 sequences (Lowary and Widom, 1998). The modification of the latter 13 601 sequences replaces the native restriction enzyme site with a unique one (see Appendix 6.1 plasmid map for pFMP232). Each of the 601 sequences is equally spaced by a 50 bp. The 601 sequences provide a high affinity binding sites for histone octamers, resulting in stably bound nucleosomes at regular intervals. We find that our approach detects full-length arrays with clear inaccessible regions, where nucleosomes are known to be present (Figure 2.2.2A). Most likely due to the slight modification of the latter 13 Widom 601 sequences, affinity for histone octamers is slightly reduced. This results in a mild decrease in the peak-trough ratio, suggesting less defined nucleosome positions. We do note that the 601 sequence is a high affinity binding region, and could therefore prevent methylation of nucleosomal DNA in ways native gDNA cannot.

Next, we investigated the position and distribution of individual information points (i.e. CpG sites). We find that the 5' region of the inaccessible region (i.e. nucleosome footprint) has a lower information density compared to the 3' region (Figure 2.2.2B, Bottom). In addition, as the sequencing platform and the subsequent bioinformatical pipeline can introduce ambiguity, we observe certain sites that do not provide non-ambiguous information on the methylation state (e.g. position 250, read 1). Such sites provide no information on the state of a CpG site and are therefore not plotted. On average, we detect approximately 18% of sites as ambiguous when analyzing the *in vitro* arrays. Taken together, the *in vitro* data highlights the long range sequencing capabilities of MEFSIMO-seq. Conversely, it also displays limitations, such as lack of information due to ambiguity or low CpG density. Using the *in vitro* arrays, we can clearly distinguish individual nucleosomes on single-molecule reads when CpG distribution and density is adequate. Improvements and alterations to this array can be found in Chapter 3.

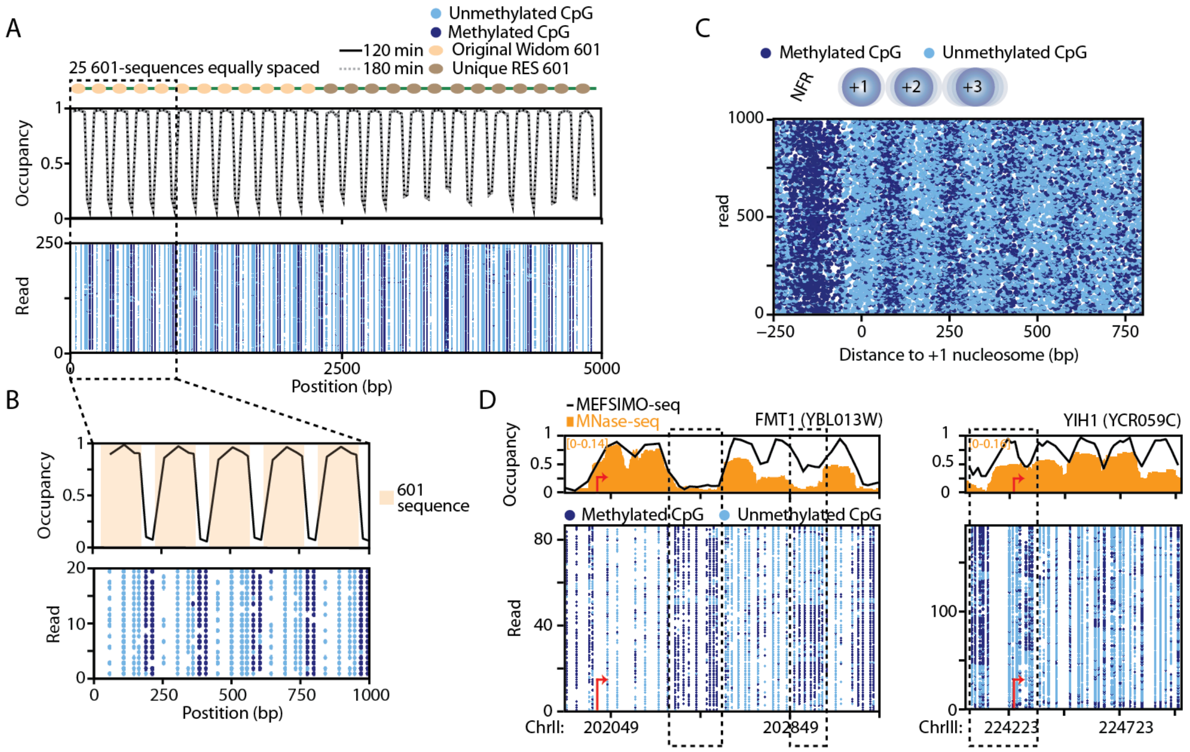


**Figure 2.2.2** ***In vitro* and *in vivo* MEFSIMO-seq data faithfully captures the chromatin landscape**
***(A)*** *Chromatinized spike-in from NPD08 experiment. Spike in consists of 25 601-sequences separated by 50bp linkers. First 12 601-sequences are of the original Widom sequence (Lowary and Widom, 1998), whereas the latter 13 have a unique restriction enzyme site (RES) that replaces the native AluI site (Ludwigsen et al., 2018). Top: composite signal of MEFSIMO-seq data. Dashed and solid lines represent two independent timepoints. Bottom: Raw MEFSIMO-seq data of 250 randomly selected reads. Methylated CpG and unmethylated CpG sites are represented as dark-blue and light-blue dots, respectively.* ***(B)*** *Top: Composite signal of first 1000bp from chromatinized spike-in. Widom 601-sequences are highlighted. Bottom: Raw MEFSIMO-seq data of 20 randomly selected reads. Methylated CpG and unmethylated CpG sites are represented as dark-blue and light-blue dots, respectively.* ***(C)*** *Raw MEFSIMO-seq data of 1000 randomly selected in vivo genic reads aligned to the +1 nucleosome. Nucleosome free region (NFR) and predicted average positions of +1 to +3 nucleosomes are indicated on top.* ***(D)*** *Top: Composite signal (black line) of MEFSIMO-seq data overlayed with MNase-seq data (147 bp extended dyads, orange) of FMT1 and YIH1 gene. Bottom: Raw MEFSIMO-seq reads computationally sorted by reads that are most similar in CpG modification using a Manhattan clustering method. Data is compiled from samples: WT_TC1_180, WT_TC5_180, WT3_90, WT1_150_180, W1_120min, W2_120min. Phenotype in compiled data is representative individual samples (individual samples not shown). Transcription start site is indicated by red arrow. Areas within dashed boxes indicate regions of interest/heterogeneity within the gene. MNase-seq data generated by Dr. Ashish Singh.*

Our aim was to visualize the *in vivo* nucleosome landscape and thus we performed MEFSIMO-seq on yeast chromatin. Utilizing unfiltered MEFSIMO-seq data, we plot a snapshot of the global *in vivo* WT chromatin landscape (Figure 2.2.2C). When aligning a random sample of 1000 reads to the predicted +1 position of the respective genes, we find relatively well-defined inaccessible regions at regular intervals (Figure 2.2.2C, light-blue regions). However, the position and size of each region can vary, due to inherent heterogeneity across genes and cells stimulated by nucleosome remodelers (reviewed in (Clapier et al., 2017; Prajapati et al., 2020; Singh and Mueller-Planitz, 2021). Similarly, upstream of the +1 nucleosome we detect the nucleosome free region (NFR). The NFR is typically highly accessible, however, its size can vary significantly due to sequence, transcriptional activity or remodeler activity (Krietenstein et al., 2016; Lai et al., 2018; Lorch et al., 2014; Mahloogi and Behe, 1997). These findings suggest that MEFSIMO-seq results in the visualization of single-molecule chromatin fibers which correspond well to known chromatin features.

We next explored the capability of MEFSIMO-seq to inform on the (ir)regularity of individual genes instead of genome wide. We find that the ability of MEFSIMO-seq to call the average nucleosome positions correlates well with MNase-seq data when comparing two different genes (Figure 2.2.2D, Top) (Singh et al., 2021). Location of each peak, i.e. the most likely position of a nucleosome, is highly similar between MNase-seq and MEFSIMO-seq. However, we do find that the amplitude of peaks is different between the two types of data. Such differences are indicative of different underlying data which may reveal previously unidentified features.

To this end we can utilize the underlying single-molecule capabilities of MEFSIMO-seq to better understand features or reveal heterogeneity that would be indiscernible otherwise (Figure 2.2.2D, Bottom). In agreement with MNase-seq data of the FMT1 gene, we observe a fully accessible region approximately 200 bp downstream of the TSS. Utilizing the MEFSIMO-seq data, we can exclude that this region is present due to MNase bias (Figure 2.2.2D, left shaded area). Instead, this region is presumably nucleosome free or harboring a fragile nucleosome. MEFSIMO-seq data can further be used to identify heterogeneity as distinct populations within a single gene. For example, further into the gene body of FMT1, individual reads contain either an accessible or inaccessible region, suggesting two distinct population of this single gene (Figure 2.2.2D, middle shaded area). Other genes may exhibit similar distinct populations, but at a different location. For example, based on the composite signal, YIH1 gene has a typical array without any discernable features. However, single-molecule data reveals two population of reads, defined by a distinct accessibility pattern flanking both sides of the TSS (Figure 2.2.2D, right shaded area). Such heterogeneity within these two exemplary genes cannot be discerned by the MNase-seq composite signal. Thus, utilizing the MEFSIMO-seq data, we can recapitulate MNase-seq data and identify the likeliest positions of nucleosomes within a gene. Importantly, the novel single-molecule information can identify further nuances and heterogeneity, unobtainable by yield based methods such as MNase-seq.

Taken together, we demonstrate the technical advances of MEFSIMO-seq has made to analogous sequencing techniques. Through carefully controlled conditions, we can probe the chromatin landscape and detect the location of (in)accessible regions. Compared to other sequencing techniques, the approach can reveal single-molecule insights of cell-to-cell heterogeneity. Elucidating the factors that work in favor or against the establishment of heterogeneity will be addressed in this study.

### 2.2.4. rDNA loci are not transcriptionally independent and are associated with UAF30 when accessible

We demonstrated the capability of MEFSIMO-seq to study *in vivo* chromatin genome wide. As a stepping stone, we first investigated the known features of the rDNA locus and aimed to further elucidate debated features. The rDNA locus in *S. cerevisiae* has been shown to exhibit binary chromatin accessibility states that are dependent on their transcriptional activity. Each of the approximately 100-200 copies are ~9.1kb in length, making a 1-2Mb region referred to as the RDN1 locus. As the exact number of copies within the RDN1 locus can vary cell to cell, the SacCer3 genome has only two copies annotated as RDN37-1 and RND37-2, each separated by non-transcribed spacers (NTS1 and NTS2). The individual loci consist each of a 35S unit transcribed by Pol I. The resulting 35S pre-rRNA transcript is processed into 25S, 18S and 5.8S rRNAs. These transcripts are represented within a single RDN37 locus as RDN25, RDN18 and RDN58 and separated by non-transcribed spacers

ITS1 and ITS2 (reviewed in Venema & Tollervey, 1999). Upstream of the RDN37 locus is a 5S unit transcribed by Pol III on the opposite strand, and a non-transcribed replication origin autonomously replicating sequence (ARS). Transcription of the 35S unit results in a region devoid of nucleosomes making the DNA highly accessible. Conversely, transcriptionally silent loci tend to have a high degree of assembled nucleosomes and are therefore deemed inaccessible. However, this dichotomy in accessibility is contested as it has been suggested that nucleosomes remain present at transcribed rDNA loci (Conconi et al., 1989; Dammann et al., 1995; Jones et al., 2007).

Recent findings using an analogous sequencing method clearly exhibited two mutually exclusive states of the 35S region (Shipony et al., 2020). With our observations using MEFSIMO-seq we were able to confirm these observations for a single 35S locus (Figure 2.2.3A). We find that reads of the RND37-1 locus are either highly methylated (accessible) or unmethylated (inaccessible). Moreover, utilizing the long-range sequencing capabilities of MEFSIMO-seq, we are also able to determine the state of reads spanning both RDN37 loci. We find that the RDN37 loci can have an inverse accessibility compared to their neighbor (mixed accessibility). Based on the distribution of methylation states when plotting either left or right (RDN37-1 or RDN37-2 locus, respectively), we define a locus as fully methylated when it exhibits >50% methylation (Figure 2.2.3B). Our calculations suggest that typically for WT yeast, approximately 30±4% of reads are in the accessible state (Figure 2.2.3C, Left Active). Similar observations and values (33±4%) are found for the neighboring RDN37-2 locus (Figure 2.2.3C, Right Active).

As MEFSIMO-seq was able to capture reads spanning both the RDN37-1 and RDN37-2 locus simultaneously, we asked whether the accessibility (and therefore transcriptional activity) of one locus was independent of its neighbor. Previous findings characterized the activity of rDNA loci as random, suggesting that they are independent (Dammann et al., 1995; French et al., 2008). Utilizing our approach, we were able to scrutinize these results utilizing MEFSIMO-seq's high throughput and accuracy. For each sample, we generated a contingency table, determining the absolute amount of reads that fall into the four states described in Figure 2.2.3A. Next, we determined the transcriptional independence of each locus by calculating the probability of both rDNA loci being active over the marginal probability distribution of one locus being active (calculations performed by Dr. Michael Wolff, see Figure 2.2.3E legend for formula). These probability observations reveal that neighboring rDNA loci are between 25-90% away from independence, depending on the sample probed (Figure 2.2.3E). This independence suggests that the transcriptional state of one locus affects the neighboring transcription state. We next performed a statistical Fisher's exact using the contingency tables test to test the null hypothesis stating that the rDNA loci are transcriptionally independent. The Fisher's exact test results in a rejection ($p<0.05$) of the null hypothesis for all tested WT samples (Figure 2.2.3F). These results indicate that, in contrast to previously published findings, the rDNA loci are not transcriptionally independent (Dammann et al., 1995; French et al., 2008). Thus, based on the probability calculations supported by the Fisher's exact test, we find that the transcriptional state of one rDNA locus can influence the activity of its neighboring locus.

We next focused on a recently identified anti-correlated region upstream of the 35S TSS (Figure 2.2.3G, purple region) (Shipony et al., 2020). This anti-correlated region is found upstream of the TSS of both rDNA loci. Briefly, when the 35S gene (RDN37-1) is methylated, the anti-correlated region is unmethylated and vice-versa. Based on the previous study, it was hypothesized that this region was

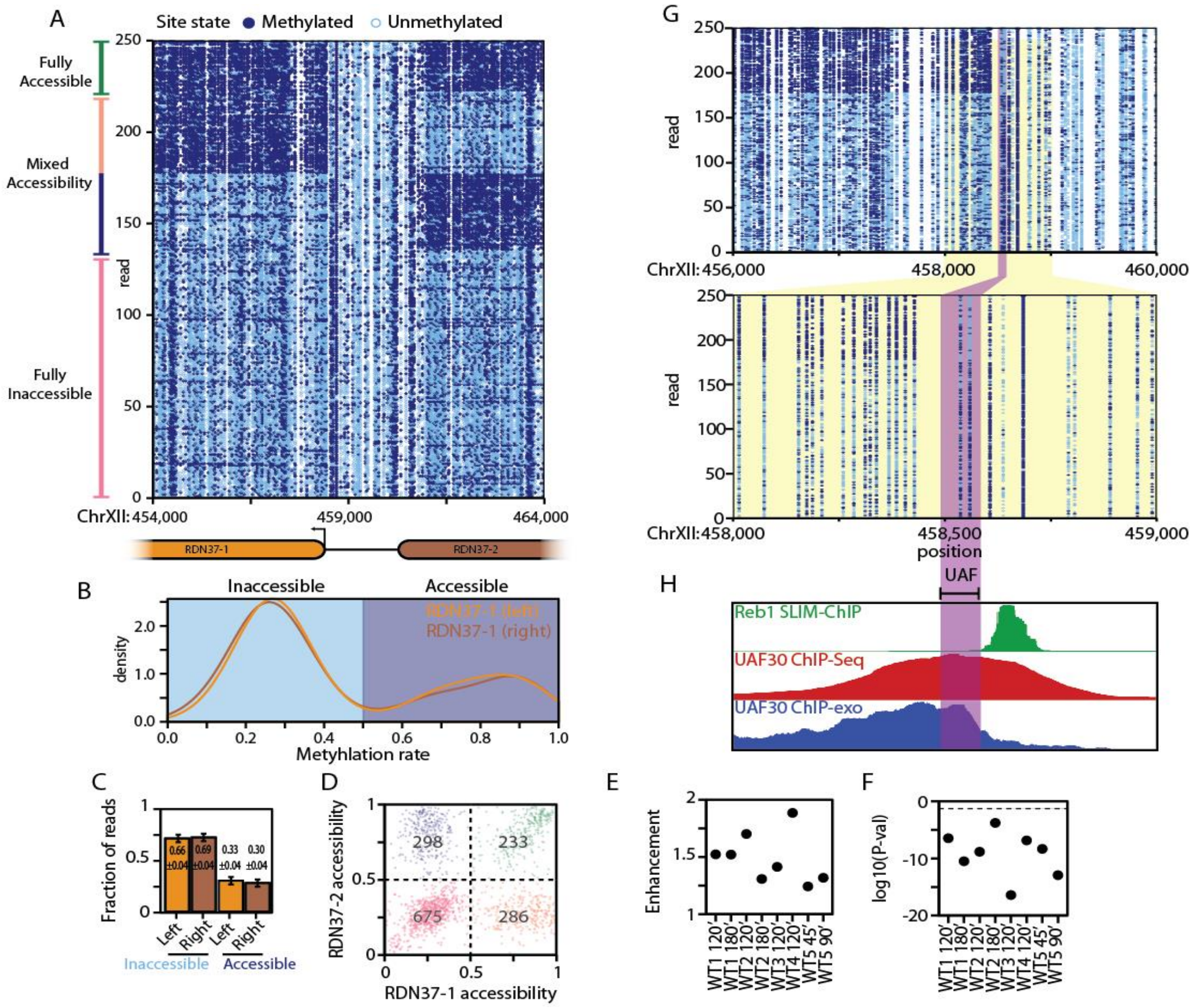


**Figure 2.2.3 rDNA loci are not transcriptionally independent and the UAF-complex binds upstream at the anti-correlated region**

***(A)*** *The rDNA loci (RDN37-1 and RDN37-2) exhibit a binary methylation state, either largely methylated (dark-blue; accessible) or unmethylated (light-blue; inaccessible). Four distinct (translational) populations spanning two rDNA loci can be observed (green, orange, blue and pink).* ***(B)*** *Distribution of accessible (dark-blue, >50% methylation rate) and inaccessible (light-blue, <50% methylation rate) reads for RDN37-1 (left) and RDN37-2 (right) locus.* ***(C)*** *Fraction of accessible compared to inaccessible (active and inactive, resp.) reads per sample. Error bars represent standard deviation. Values are derived from WT strain (yFMP013) and samples: W1_120min, W2_120min, WT3_90, WT1_150_180, WT_TC1_180, WT_TC5_180 (NPD07-NPD10).* ***(D)*** *Amount of reads for each population as determined by a cutoff of 50% methylation for each rDNA locus. Values for a single biological sample (WT3_90, NPD10) are inserted as a representative example for distribution.* ***(E)*** *Calculating the probability whether the transcriptional state of one locus affects the state of a neighboring locus (enhancement) To this end we calculate the conditional probability distribution of one locus being active whilst the neighboring locus is active as well. This probability is divided by the marginal probability distribution of one locus being active whilst ignoring the neighboring locus. To measure this observable, we utilize the following equation:* $r = \frac{P(locus\ 2\ active \mid locus\ 1\ active)}{P(locus\ 2\ active)}$*.* ***(F)*** *To statistically test whether the two loci are transcriptionally independent we utilize the contingency tables generated in* ***(D)*** *and perform a Fisher's exact test. Values are converted on a log10 base scale to better represent the significance. The log10 base threshold of p=0.05 is indicated by the dashed line. Analyses performed by Dr. Michael Wolff* ***(G)*** *An anticorrelated region upstream of either rDNA (here shown is RDN37-1) locus can be detected using MEFSIMO-seq as previously reported (Shipony et al., 2020). The highlighted area in yellow is enlarged in the bottom panel. The predicted binding region of the UAF-complex (60 to 155 bp upstream of the RDN37-1 TSS) is highlighted in purple and overlaps with the anticorrelated region.* ***(H)*** *The anticorrelated region and predicted UAF-complex binding site (highlighted in purple) both overlap with the peak of UAF30 ChIP-seq and ChIP-exo data (Iida and Kobayashi, 2019; Rossi et al., 2021). Reb1-SLIM-ChIP data does not overlap with the anticorrelated region (Gutin et al., 2018).*

the binding site of transcription factor Reb1. However, this study also identified that Reb1 ChIP-exo data did not overlap with this region. We hypothesized whether the UAF-complex could be responsible for this anticorrelation as it was shown to target the upstream activating sequence (UAS) of the rDNA promoter (60-155bp upstream of the 35S TSS, Keys et al., 1996). Furthermore, the complex is required for Pol I activity *in vivo* and targets the complex to the promoter region (Hontz et al., 2008; Knutson et al., 2020; Smith et al., 2018). The complex consists of six subunits, Uaf30, Rrn5, Rrn9, Rrn10 and histones H3 and H4. Absence of subunit Uaf30 prevents efficient binding of the complex to the rDNA promoter (Hontz et al., 2008). We took Uaf30 ChIP-seq data and found that it overlaps with the inaccessible region found at the promoter region upstream of the 35S locus (Iida and Kobayashi, 2019) (Figure 2.2.3H). In concordance to the Uaf30 ChIP-seq data, we also find that high-resolution Uaf30 ChIP-exo data overlaps with the same inaccessible region (Rossi et al., 2021). Similar to the previous finding, we do not observe an overlap of the inaccessible region and Reb1 binding (Gutin et al., 2018; Shipony et al., 2020). Based on the ChIP-seq and high-resolution ChIP-exo data, we conclude that the UAF-complex binds to the promoter region of active 35S loci resulting in an anticorrelated accessibility.

### 2.2.5. The *S. cerevisiae* chromatin landscape is more regular than previously observed

We next turned our attention back to the genome wide chromatin landscape. Whilst at a single locus of the rDNA we observed a binary accessibility state, typical genic loci have broad heterogeneity associated with them. Sequencing methods such as MNase-seq, have demonstrated this heterogeneity by plotting composite averages of all known genes aligned to the +1 nucleosome in wild-type cells. These findings showed a decrease in array regularity in the direction of transcription (Figure 1.1.3B). Moreover, genome browser shots of individual genes highlight the different chromatin structures that may occur (Figure 1.1.3C-D). Utilizing our MEFSIMO-seq data we observe a decaying regularity downstream of the +1 nucleosome, similar to the heterogeneity that can be observed utilizing MNase-seq (Figure 2.2.1H). Different however, is the underlying ability of MEFSIMO-seq to capture the cell-to-cell heterogeneity of individual genes (Figure 2.2.4A). For example, we currently do not know if genes with low regularity scores (Singh et al., 2021), may be the result of ill-phased arrays despite the presence of nucleosome remodelers. Such genes could in theory be highly regular, but have a dynamic alignment point (i.e. the +1 nucleosome) frequently found at different locations. To elucidate the differences between arrays with and without ill-phased arrays, we utilized the single-molecule capabilities of MEFSIMO-seq.

First, we identified whether a nucleosome was present at the predicted +1 nucleosome position. For each read we probed the CpG site(s) within a ±20 bp window around the predicted position of the +1 nucleosome (Chereji et al., 2018) (Figure 2.2.4B). The mean methylation state of the site(s) within this window was calculated and was assigned to each (partial) read and locus (See Methods). Based on this approach we observe two distinct states for the +1 nucleosome which we defined as N+1 present or N+1 shifted. A third state in which we cannot define the methylation state unambiguously within the window (e.g. two sites within the window are of different methylation status), was filtered out in this approach (2±0.1%). The average amount of CpG sites found within the window is 1.02±0.01, suggesting that only a small fraction of windows contain multiple CpG sites that could result in ambiguity. We further observed that increasing or decreasing the window size can influence output and ambiguity. We find that a ±5 bp window will result in less ambiguity (0±0%) and only a single CpG site per window (1.0±0), as the chance for finding multiple CpG sites with different methylation states is reduced. However, the overall probability of finding a CpG site within this smaller window is also decreased, which directly impacts the output. Conversely, utilizing a larger window (e.g. ±50 bp) results in more ambiguous windows (13±1%) and an average higher amount of sites per window (1.13±0.02), but more throughput of reads. Despite the differences in output and ambiguity, we did not observe genome wide differences in downstream applications and therefore continued with a window size of ±20 bp (Figure 2.2.4C). After categorizing reads based on the state of the +1 nucleosome, we find that 19±4% of all reads were defined as N+1 shifted (averaged over sequencing runs NPD07-NPD10, samples: W1_120, W2_120, WT_TC1_180, WT_TC5_180, WT1_150_180, WT3_90).

To determine the regularity and phasing of an array on a per-read basis we first required to identify the dyad position of all nucleosomes. To this end we took raw MEFSIMO-seq data and computationally fitted nucleosome dyads in the center of inaccessible regions. We required each inaccessible region to be minimally 147±10 bp in width. If an inaccessible region can theoretically hold more than a single nucleosome (e.g. 300 bp wide), the nucleosome dyads are distributed at equidistance from the outer limits and from each other (Figure 2.2.4D). We note that due to the density and distribution of CpG sites in the genome, the exact location of a nucleosome dyad cannot be determined. Regions with a higher CpG density, or better distribution, will therefore naturally provide better approximations, although physical limitations apply (see Chapter 3). We however argued that this stochastic approach is best suited for this type of data and allows us to predict the nucleosome dyad position. Further potential improvements are found in the discussion.

Having defined the methylation state within the window for each read and added the ability to identify the nucleosome dyad, we next separated the reads to find differences in regularity between the two N+1 states (Figure 2.2.4E). From the single-molecule plots we observe that regularity appears to be higher for reads with a N+1 present. Reads defined as ‘N+1 shifted’ clearly have methylated CpG sites at the 0 position. Downstream regularity however is not directly apparent when compared to N+1 present reads. Utilizing the analysis to approximate the nucleosome dyads described earlier (Figure 2.2.4D), we can fit nucleosomes in inaccessible regions (Figure 2.2.4F). Visual inspection indeed suggests an improved regularity for reads where an N+1 is present (Figure 2.2.4F; Top). However, underlying array regularity may be missed due to the lack of phasing in reads with a shifted N+1 (Figure 2.2.4F; Bottom).

Following the separation of reads based on methylation state and a nucleosome fitting, we performed a computational phasing of individual reads (See Methods for details). This allows us to disregard the genomic +1 dyad location and instead observe arrays as if they were perfectly aligned. To this end we took the nucleosome dyad position we calculated before, and aligned it to the hypothetical zero coordinate. To prevent artificial regularity that may be introduced by fitting multiple dyads in large inaccessible regions, we only computationally phase reads in which the inaccessible region at the +1 nucleosome fits a single dyad. We note that this step may exclude up to 75% of reads globally and thus can reduce coverage of single loci drastically if raw sequencing output is low or not combined with other (biologically similar) samples. Regardless, using the current conditions with a single CpG methyltransferase enzyme, this phasing still results in the precise alignment of the predicted +1 nucleosome dyad, thereby largely nullifying the native heterogeneity of the +1 nucleosome found in

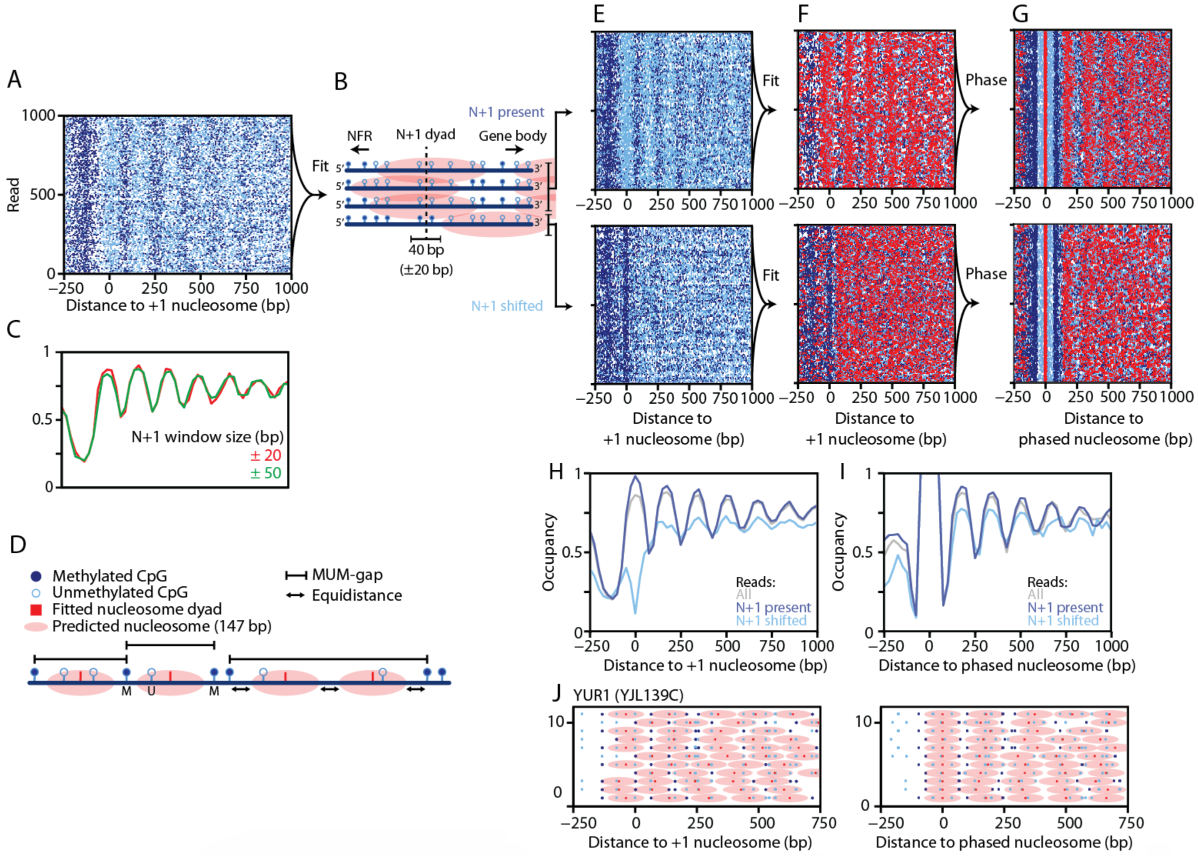


**Figure 2.2.4 The *S. cerevisiae* nucleosome landscape shows high regularity when single-molecule reads are computationally phased**

***(A)*** *Unfiltered MEFSIMO-seq data of wild-type S. cerevisiae. Each read represents an uninterrupted single-molecule. A subset of 1000 randomly selected reads are shown from a single biological sample (WT3_90 NPD10. See Appendix Supplementary table 1).* ***(B)*** *Cartoon outline of computational strategy for separating reads with phased and shifted +1 nucleosomes. Methylation state of CpG site(s) 40 bp (±20 bp) surrounding the predicted +1 nucleosome dyad (N+1 dyad, Chereji et al., 2018) are monitored. Individual reads with either unambiguously methylated or unmethylated site(s) in this window are separated into different categories based on their methylation state (N+1 phased and N+1 shifted, unmethylated and methylated, respectively).* ***(C)*** *Global composite plots of all reads after determination of N+1 state using N+1 window size of ±20 bp or ±50 bp. Effect on global composite plots is negligible. A larger window size results in more reads filtered (2±0.1% versus 13±1% for 20 and 50 bp windows, respectively).* ***(D)*** *Cartoon outline of computational strategy for fitting nucleosomes at inaccessible (i.e. unmethylated) regions. The MUM-gap is the distance between two methylated (M) CpG sites, separated by minimally one unmethylated (U) site. The nucleosome dyad is fitted at the center of an MUM-gaps of minimally 147±10 bp. Dyads are fitted at equidistant from MUM-gap limits and each other when multiple nucleosomes can be fitted (right side of cartoon).* ***(E)*** *Filtered MEFSIMO-seq data of 1000 randomly selected reads are plotted for each N+1 state (top and bottom).* ***(F)*** *Fitting of nucleosome dyads as illustrated in* ***(B)*** *using the reads plotted in* ***(D)****.* ***(G)*** *Computational phasing of reads categorized in* ***(E)****. The fitted nucleosome dyad closest to the predicted +1 nucleosome dyad is set to coordinate 0. Only reads for which footprint of the phased +1 nucleosome only can fit a single nucleosome (<180bp). For more details see Methods.****(H-I)*** *genome wide average composite signal of unphased and computationally phased MEFSIMO-seq data for all, N+1 present and N+1 shifted reads (grey, dark-blue and light-blue, respectively).* ***(J)*** *MEFSIMO-seq data for YUR1 gene before and after computational phasing (left and right, respectively). Methylated and unmethylated CpG sites are indicated as dark- and light-blue dots, respectively. Fitted nucleosome dyads are indicated as red dots. Manually inserted representations of the nucleosome footprints are shown as light-red ovals and represent a nucleosome footprint over an area of 146 bp.*

cells (Figure 2.2.4G; Top). Visual observations suggest that reads defined as N+1 shifted, are now showing an improved regular downstream array when compared to the pre-phasing state (Figure 2.2.4F-G; Bottom).

Next, we further investigated how the position of the +1 nucleosome affects the average array regularity by utilizing composite plots. Based on the initial observations, we hypothesized that computational phasing would result in high array regularity compared to non-phased reads. We first compared reads before nucleosome fitting or computational phasing and find that, compared to all reads, reads within the N+1 aligned state, have a slightly higher regularity (Figure 2.2.4H, dark-blue). In line with our initial visual observations, reads in which the N+1 was shifted, show a drastic reduction in regularity (Figure 2.2.4H, light-blue). Next, we compared genome wide composite plots of computationally phased reads. We find a similar slight improvement in N+1 present reads when compared to all reads (Figure 2.2.4I, grey and dark-blue, resp.). Interestingly, N+1 shifted reads improve drastically and show a high degree of regularity after computational phasing (Figure 2.2.4I, light-blue). Despite the markedly increased regularity in N+1 shifted reads, it remains reduced when compared to all or N+1 present reads. Based on these observations we find that (1) the array regularity in WT *S. cerevisiae* is already highly optimized and (2) the actual regularity of unphased arrays are still highly regular despite composite plots suggesting otherwise. We conclude that sequencing techniques which provide a global average (such as MNase-seq), accurately portray the genome wide array pattern. In addition to previously identified array regularity, we here show that the underlying regularity of the nucleosome landscape is more regular than initially observed and that nearly all genic arrays are highly regular. The precise position of the +1 nucleosome is not a determining factor of proper downstream array regularity. However, lack of a native well-positioned +1 nucleosome (such as those in N+1 present reads) does affect the downstream regularity to an extent, as evidenced by the reduced regularity of N+1 shifted reads after computational phasing.

Lastly, we aimed to utilize our MEFSIMO-seq data to determine the actual regularity of individual genes which contain a fraction of reads with a shifted +1 nucleosome. Moreover, we aimed to visualize that the improved regularity we observe is not only apparent when plotting a composite signal, but also for individual genes. One example is the YUR1 gene located on chromosome 10 in *S. cerevisiae*. MNase-seq as well as our MEFSIMO-seq data would suggest that this gene has a low regularity based on the array regularity score (Singh et al., 2021). We find that the +1 nucleosome dyad is often located at a different location than predicted (i.e. not at coordinate 0). After computationally phasing the reads that span this gene, we find a regular array in the underlying data (Figure 2.2.4J). We do note that this particular gene has a relative low coverage, despite being compiled from data over several biological replicates. This issue may arise for other genes as well due to the filtering step during phasing as described earlier. However, this limited data does represent well how computational phasing can reveal underlying array regularity at a single locus. Taken together, genome composite plots, genome wide single-molecule plots and those of individual genes suggest the composition of arrays is highly regular in WT *S. cerevisiae*.

### 2.2.6. Transcription destroys array regularity, but does not result in hexasomes in the gene body

We next wondered whether transcriptional activity could be correlated to the state of the +1 nucleosome we defined earlier. Transcribed genes typically tend to have a well-positioned +1 nucleosome, an open NFR and regular downstream phasing to aide in transcription activation.

Conversely, inactive genes may be regular, but have unphased +1 nucleosomes within the NFR (Wang et al., 2019). Moreover, it has been suggested that proper positioning of the +1 nucleosome may be aided by the transcription machinery (Struhl and Segal, 2013; Vasseur et al., 2016). However, these observations are contested by other publications suggesting a destructive effect by transcription (Baldi et al., 2018; Singh et al., 2021).

To further elucidate the effect of transcription on array regularity, we utilized our data on the N+1 shifted ratio for each gene. We compared the N+1 shifted ratio to the transcriptional rate of genes. We hypothesized that highly transcribed genes tend to have a higher ratio of shifted +1 nucleosomes. This observation would be in line with recent findings that transcription destroys arrays (Singh et al., 2021). To this end, we took 4tu-seq data acquired from Barras et al. and RNA Pol II ChIP-seq data from Ocampo et al. 4tU-seq data examines RNA processing kinetics informing on the transcriptional activity (Barrass et al., 2015). Alternatively, RNA Pol II ChIP-seq data quantifies the presence RNA Pol II at the gene, which informs directly on the transcriptional activity (Ocampo et al., 2016). In order to compare transcriptional activity to the N+1 shifted ratio, we split genes in four quartiles based in their transcriptional activity. Utilizing these RNA Pol II ChIP-seq data, we find a significantly higher ratio of N+1 shifted reads for highly expressed genes (quartile 4) when compared to lowly expressed genes (quartile 1) (Figure 2.2.5A). Near identical results were observed for the 4tU-seq dataset from Barras et al. These datasets suggest a correlation between transcription rate and a shifted +1 nucleosome. We took alternative published 4tU-seq data to further validate our findings and performed the same analysis (Xu et al., 2017). Curiously, utilizing this dataset we find no significant difference between the lowest (1) and highest (4) transcriptional quartiles.

To investigate the discrepancy between datasets, we performed a correlation analysis. We find that the datasets from Barras et al., and Ocampo et al. correlate relatively well, despite being different techniques (r=0.51). In contrast, the two previously mentioned datasets correlate significantly worse with the Xu et al., dataset, suggesting that this dataset may be faulty (r=0.27 and r=0.32, respectively). Based on the datasets that agree with each other, we conclude that highly transcribed genes tend to have a higher ratio of shifted +1 nucleosomes when compared to lowly transcribed genes. Transcription thus affects the position of the +1 nucleosome.

We next investigated whether transcriptional activity has an effect on the regularity of arrays. As mentioned earlier, recent findings would suggest that RNA Pol II destroys arrays. In addition, it was suggested that highly transcribed genes have wider NFRs compared to lowly transcribed genes (Weiner et al., 2010). We observe mild to no improvement in regularity for the top 25% transcribed genes when compared to the bottom 25% transcribed genes when Xu et al., 4tU-seq data is used (Figure 2.2.5B, left). These trends remain similar when comparing top- and bottom 5% or 10% (Figure 2.2.5B, right). Utilizing the Xu et al., dataset, we do not observe a widened NFR for highly transcribed genes further suggesting some sort of defect in the dataset. In contrast, RNA Pol II ChIP-seq data and 4tU-seq from Barras et al., do suggests that highly transcribed genes tend to have worse array regularity (Figure 2.2.5B). Moreover, we find that these datasets show a wider NFR at highly transcribed genes. We therefore conclude that RNA Pol II has a negative impact on array regularity as previously suggested.

Besides the role of transcription on array regularity and NFR width, it has furthermore been shown that RNA Pol II transcription can lead to subnucleosomal fragments e.g. hexasomes. Highly transcribed genes are suggested to have such hexasomes or partially unwrapped nucleosomal DNA at the +1 position or in the gene body (Kulaeva et al., 2010; Ramachandran et al., 2017). In order to address these observations, we first determined the location and size of each footprint from 250 bp upstream to 600 bp downstream of the predicted +1 nucleosome (Figure 2.2.5C). We find that our mononucleosomal footprints are most common throughout our data and are, as previously stated, approximately 158 bp. Interestingly, at the +1 position we find two distinct nucleosome footprint populations. One population appears to be of 'regular' mononucleosomal size and is centered at the zero coordinate. Another population of mononucleosome is found slightly upstream, and exhibits a footprint of approximately 175 bp. This population could be nucleosomes with an additional factor

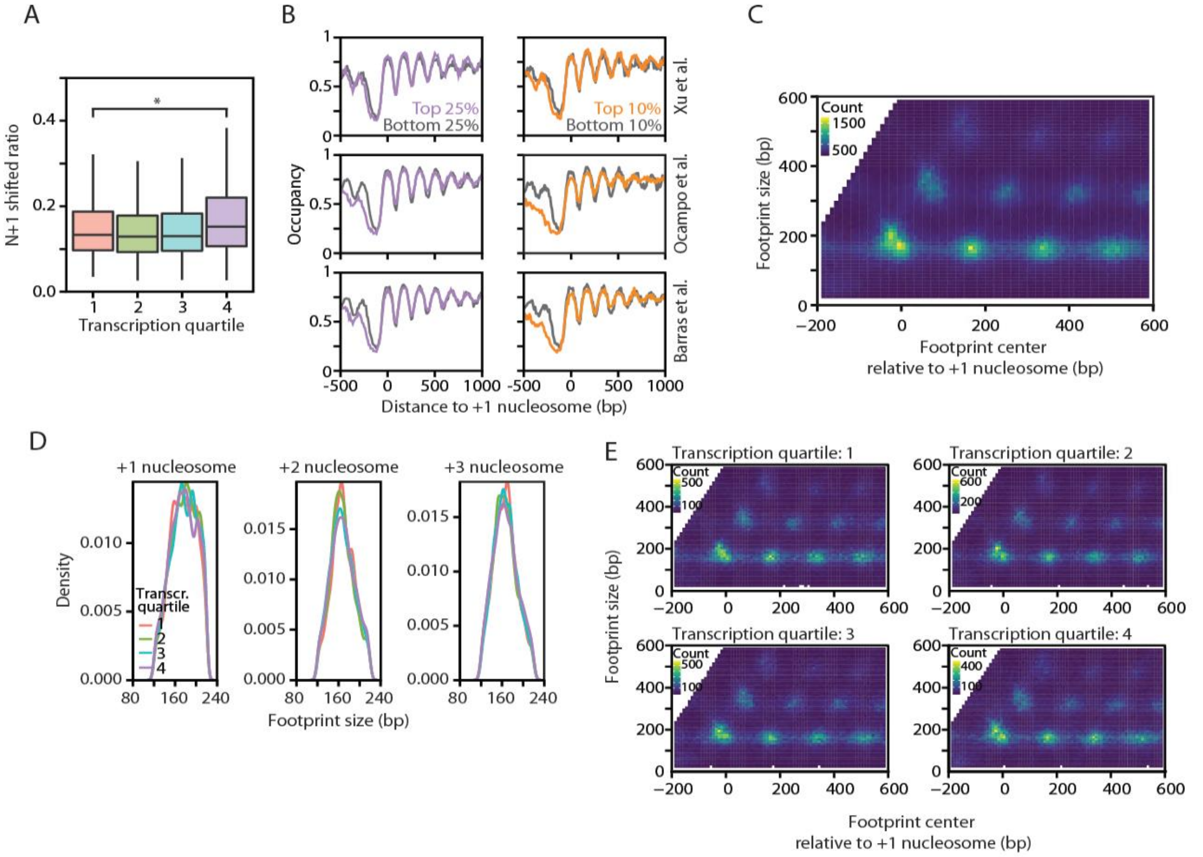


**Figure 2.2.5 Transcription destroys array regularity, but does not result in hexasomes in the gene body**

***(A)*** *N+1 shifted ratio per transcriptional quartile from low (1) to high (4) based on Ocampo et al. RNA Pol II ChIP-seq data. Boxplots visualize median and 25th and 75th percentile. Whiskers extend from 25th and 75th percentile to lowest/highest value no further than 1.5 times the interquartile range. A pairwise t-test was performed to identify a statistical difference between low (1) and highly (4) transcribed genes (marked by asterisk). (n=6; runs NPD07-NPD10, samples: W1_120, W2_120, WT_TC1_180, WT_TC5_180, WT1_150_180, WT3_90).* ***(B)*** *Genome wide composite plots of top and bottom quartiles (left) or centiles (right). Comparison between three independent publications shows differences at highly transcribed genes (purple and orange lines).* ***(C)*** *Nucleosome footprint locations within the gene body. The footprint size is measured by the distance between two methylated sites with minimally a single unmethylated site separating them (see Figure 2.2.4B). Location of footprint is relative to the predicted +1 nucleosome for each gene. Counts indicate the number of footprint dyads at a particular position.* ***(D)*** *Nucleosome footprint size for genes based on transcriptional activity from low (1) to high (4) based on Ocampo et al. RNA Pol II ChIP-seq data. Footprints are assigned +1/+2/+3 nucleosome when the dyad is <220bp and within a 150bp window around the predicted peak position of each nucleosome (coordinates 0, 166, 232) (n=3).* ***(E)*** *Nucleosome footprint locations within the gene body separated by transcriptional quartiles based on Ocampo et al. RNA Pol II ChIP-seq data. Counts for each individual quartile are shown as inserts and indicate the number of footprint dyads at a particular position.*

bound. However, contrary to the previous observations, our data does not suggest the presence of a distinct population of smaller (subnucleosomal) footprints at the +1 or into the gene body, even in the quartile of genes with the highest transcription.

We hypothesized that the footprint observed at the +1 nucleosome could be a nucleosome with a bound factor, e.g. part of the transcription machinery. To investigate this possibility we split the genes in quartiles based on RNA Pol II ChIP-seq. For the genes in each quartile, we calculated the footprint size distribution within a 200 bp window around the predicted locations of the +1, +2 and +3 nucleosome. Based on our observations, we see no difference in footprint size distribution when comparing highly transcribed to lowly transcribed genes (quartiles 1 to 4) (Figure 2.2.5D). No differences are observed in footprint distribution for each nucleosome when utilizing Ocampo et al., 4tU-seq data either.

We argued that the distribution of the +1 nucleosome in Figure 2.2.5D may be too broad to observe two distinct populations. To this end we visualized the footprints of all nucleosomes as we did in Figure 2.2.5C, but for each transcriptional quartile. Similar to Figure 2.2.25D, we find an enlarged distribution of footprints remains present around the +1 nucleosome for all transcriptional quartiles (Figure 2.2.5E). We therefore conclude that (1) highly transcribed genes do not have a tendency to have hexasomes at the +1 or in the gene body and (2) the +1 nucleosome footprint is not enlarged by a component of the transcription machinery. These findings suggest that a factor other than that of the transcription machinery is bound to certain +1 nucleosomes.

### 2.2.7. Inducible promoters associated with STM and RPG themes are overrepresented in genes with shifted +1 nucleosomes

The promoters of genes transcribed by the RNA Pol II machinery can be broadly distributed into two categories, inducible and constitutively active. Recent findings have defined these promoters into four distinct themes based on their architecture (Rossi et al., 2021). Here, findings suggest that roughly two-thirds (>4000) of all promoters have an architecture corresponding to constitutive, but low, gene expression. These promoters are associated to the TFO and UNB architectural theme. Briefly, the TFO promotors contain a sequence-specific transcription factor (ssTF) or other cofactor bound (hence TFO) between the -1 and +1 nucleosome. UNB themed promotors showed no identified binding events aside from a potential pre-initiation complex (PIC; hence UNB for unbound). In addition RP and STM themed promoters are associated with ribosomal proteins (hence RP) and inducible genes. The STM theme is SAGA-dominated but also includes co-factor complexes Mediator and Tup1, which when co-occuring at promoters are associated with highly expressed genes. The SAGA co-factor complex furthermore facilitates assembly of the pre-initiation complex which further regulates gene expression.

We hypothesized that RP and STM themed promoters were overrepresented in genes that exhibited a high ratio of shifted +1 nucleosomes as RNA Pol II can destroy arrays. Conversely, constitutively active and lowly expressed promoters of the TFO and UNB theme would have mostly well positioned +1 nucleosomes. To this end we took the N+1 shifted ratio calculated previously and assigned the promoter theme to each associated gene. We next split the data in equal quartiles based on the N+1 shifted ratio of individual genes. We found that indeed TFO and UNB themed promoters were mildly overrepresented in genes with a well positioned +1 nucleosome, but did not significantly differ across

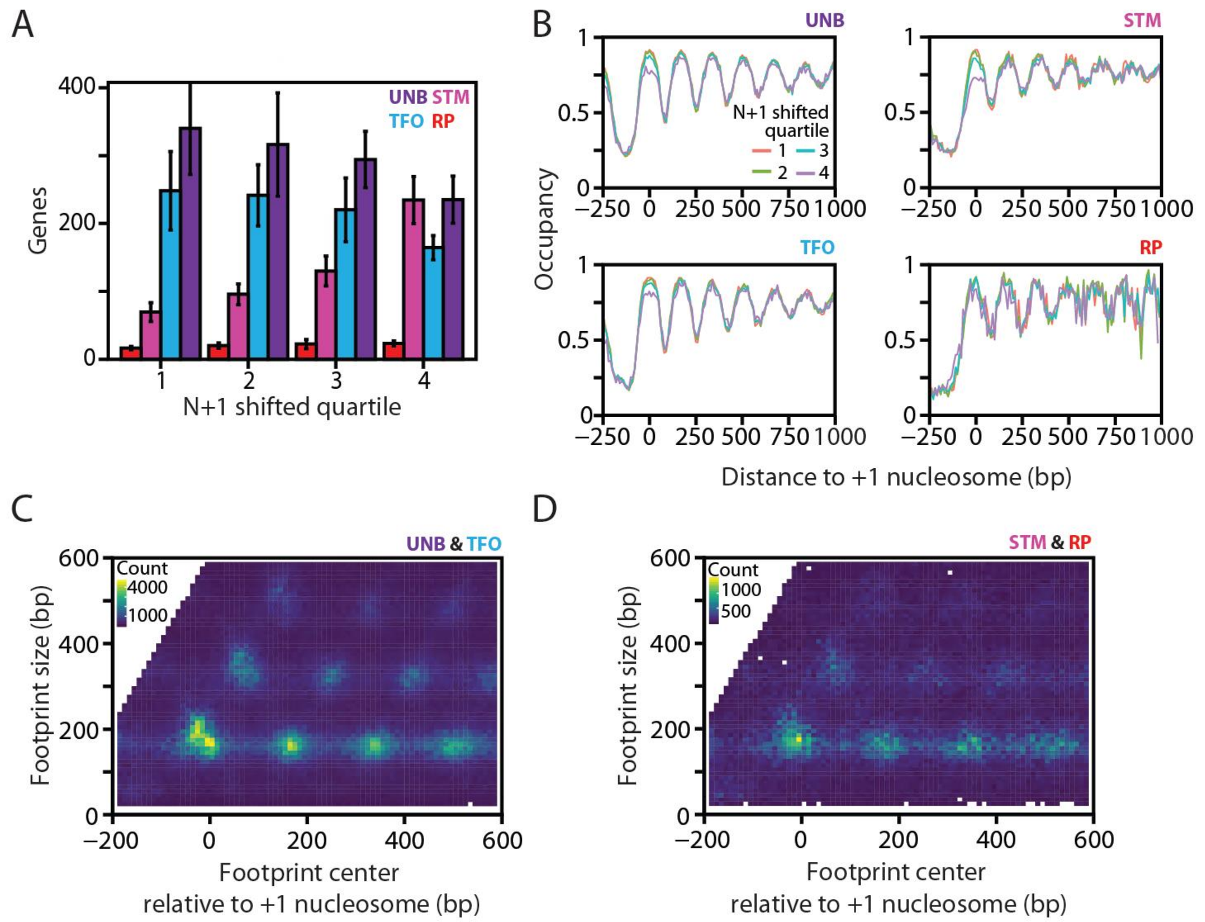


**Figure 2.2.6 Inducible promoters associated with STM and RP themes are overrepresented in genes with shifted +1 nucleosomes**

***A)*** *Total count of genes from low (1) to high (4) shifted N+1 ratio for each promoter theme (UNB, TFO, STM and RP) based on Rossi et al. classification. Error bars represent standard deviation over biological replicates (n=3).* ***(B)*** *Genome wide composite plots for genes within each promoter theme separated by low (1) to high (4) shifted N+1 ratio.* ***(C)*** *Nucleosome footprint location for genes associated with UNB and TFO themed promoters (N=2382). A larger footprint, akin to a second population, can be seen near the predicted +1 nucleosome position (coordinate 0) indicative of a bound factor enlarging the footprint size.* ***(D)*** *Similar analysis as performed in* ***(C).*** *Nucleosome footprint location for genes associated with STM and RPG themed promoters (N=546). No additional distinct population can be observed close to the +1 nucleosome.*

quartiles. RP, and especially STM themed promoters, were clearly overrepresented in genes with a high ratio of shifted +1 nucleosomes (Figure 2.2.6A). Taken together, the data further supports a negative role for RNA Pol II with regards to array regularity consistently linking the state of the +1 nucleosome to transcription. Here, we further define RP and STM themed promoters to be associated with a shifted +1 nucleosome. Thus, transcription affects the position of the +1 nucleosome.

Next, we wondered whether the different themed promoters not only influenced the position of the +1 nucleosome, but thereby also the downstream array. We earlier identified that reads with a shifted +1 nucleosome still had regular arrays, but not to the extent of the genome wide pattern (Figure 2.2.4G). We find that genes with the highest ratio (quartile 4) of N+1 shifted reads have a lesser phased +1 nucleosome (Figure 2.2.6B). This is especially apparent for STM themed promoters where we find the majority of genes having a high ratio of shifted +1 nucleosome. Similarly, but to a lesser extent we find the same for UNB/TFO themed promoters. Akin to earlier observations, an unphased (i.e. shifted) +1 nucleosome results in reduced downstream array regularity, which is most apparent in genes with an STM themed promoter. Based on these observations, it appears that the overall chromatin architecture of the genes within the same promoter theme is similar, but is influenced by the position of the +1 nucleosome. In agreement with the negative role for RNA Pol II on array regularity, we do find that the overall quality of arrays is less for genes with an STM/RP themed promoter.

Highly transcribed genes did not preferentially have larger +1 nucleosome footprints (Figure 2.2.5E). We hypothesized that STM/RP themed promoters (associated with inducibility and high transcription) would not have a population of enlarged footprints. In contrast, UNB themed promoters are not bound by any of the STM co-factors but instead only had a PIC present. TFO themed promoters also lack any STM co-factors, but are often bound by an insulator such as Reb1, Rap1 or Abf1 close to the +1 nucleosome. The +1 nucleosome of genes with a UNB/TFO themed promoter were therefore hypothesized to contain reads where the +1 nucleosome has an additional factor bound. Indeed, we observe two distinct populations of the +1 nucleosome footprint for UNB/TFO themed promoters (Figure 2.2.6C). STM/RP themed promoters lack a second enlarged footprint at the +1 nucleosome (Figure 2.2.6D). Moreover, the distribution of footprints is broader, consistent with a reduction in array regularity for highly transcribed genes. We conclude that constitutive and insulated genes characterized by the UNB/TFO themed promoters can have an enlarged +1 footprint, likely due to the binding of another factor.

### 2.2.8. ISWI/CHD nucleosome remodelers shape the genomic landscape downstream of the +1 nucleosome

We identified that arrays are largely regular despite scenarios in which they show a lack of phasing (Figure 2.2.4H-I). Next, we wondered how regularity would be affected in a strain which reduces array regularity globally. To this end, we utilized a strain lacking ISWI/CHD remodelers (*isw1Δ, isw2Δ, chd1Δ,* hereafter referred to as TKO) that had been previously generated (Tsukiyama et al., 1999). The ISWI/CHD remodelers are well-known for their ability to establish nucleosome arrays and aide in nucleosome assembly (Fei et al., 2015; Ito et al., 1997). Moreover, sliding activity of these remodelers was shown to result in centering of mononucleosomes, suggesting a role in the creation of regularly spaced arrays. Further *in vitro* studies showed that this activity indeed results in phased and regular arrays (Krietenstein et al., 2016). Later, genome wide *in vivo* results validated these findings, and have shown a reduced regularity when these remodelers are deleted in *S. cerevisiae* (Gkikopoulos et al., 2011; Ocampo et al., 2016; Singh et al., 2021). Further *in vivo* evidence in a mammalian system showed that SNF2H deletion (the ATPase of ISWI complex), results in reduced array regularity and phasing (Barisic et al., 2019). However, it remains unclear whether deletion of these remodelers results in a reduced array regularity simply due to a lack of phased arrays. Limited MNase data suggests that regular arrays are prevalent despite a lack of remodelers (Lieleg et al., 2015). This leaves the possibility that regular arrays are generated simply due to the limited organizational possibilities. Indeed, it was shown that a purely statistical positioning theory could result in regular arrays (Kornberg and Stryer, 1988; Rube and Song, 2014). To identify how the depletion of ISWI/CHD remodelers affects individual arrays *in vivo*, we utilized MEFSIMO-seq data to determine array regularity on a per-read basis.

We first validated that our MEFSIMO-seq data recapitulates the composite signals observed in MNase-seq for wild-type (WT) and TKO cells (Figure 2.2.7A). We find that the regularity is severely affected in the TKO strains as previously reported. Single molecule plots of the unfiltered MEFSIMO-seq data further indicates a strong reduction in regularity (Figure 2.2.7D; First panel). Fitting of nucleosome

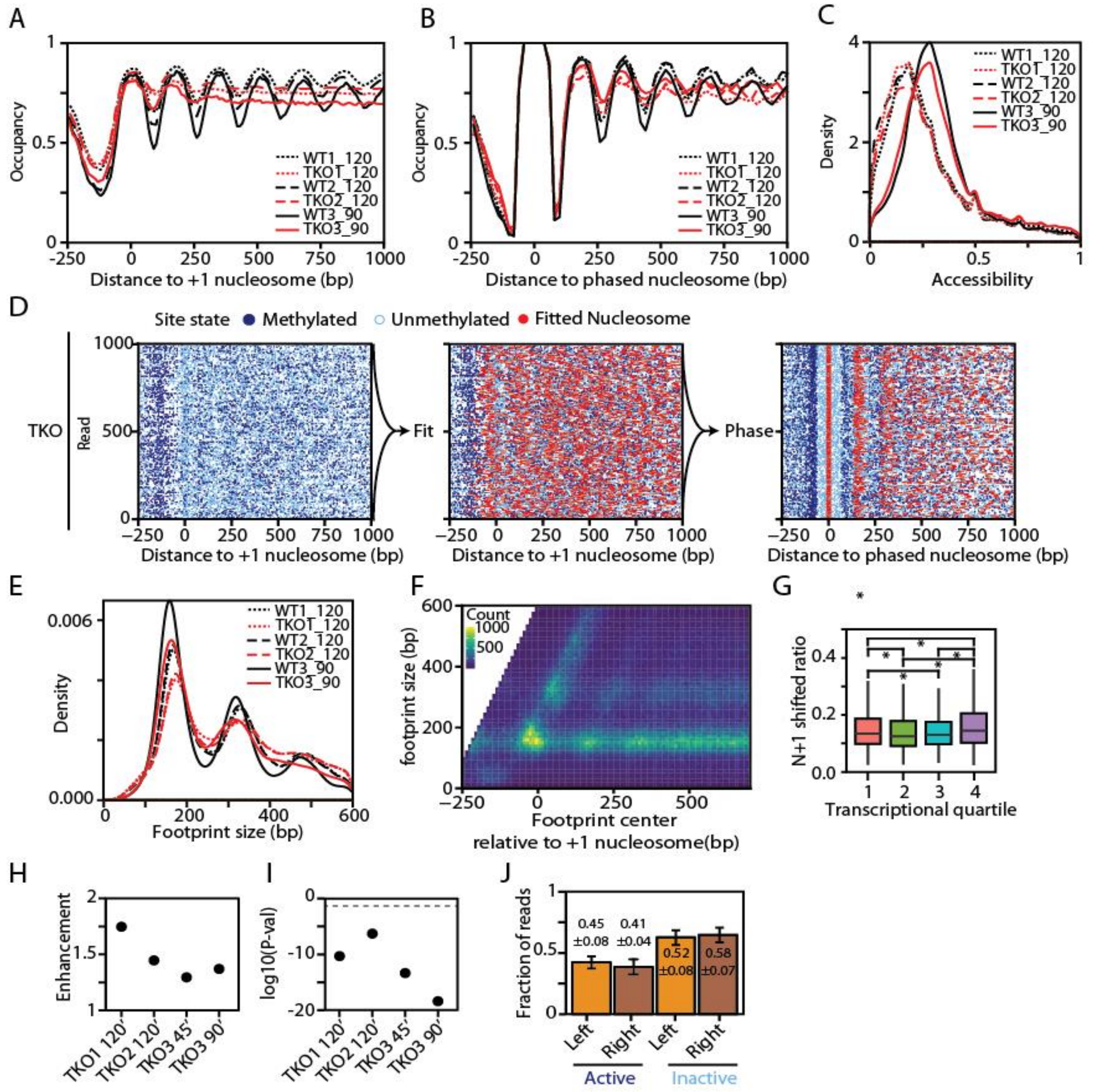


**Figure 2.2.7 The ISWI- and CHD-remodelers are essential for genic nucleosome regularity, but do not affect assembly levels genome wide and locus specifically**

***(A)*** *Genome wide composite plot comparison of WT and TKO data. Genes are aligned to predicted +1 nucleosome position in WT based on chemical mapping. More details on samples see Appendix Supplementary table 1-2 NPD07 and NPD10.* ***(B)*** *Computational phased composite plots for WT and TKO strains. +1 nucleosome of individual reads are phased and fixed to 0 coordinate. Reads are filtered that can not fit a single nucleosome dyad within the footprint closest to the predicted +1 nucleosome.* ***(C)*** *Accessibility (i.e. methylation) distribution comparison of all reads in WT or TKO strain. Data is representative of all biological replicates.* ***(D)*** *MEFSIMO-seq data of 1000 random unfiltered reads from S. cerevisiae TKO (isw1Δ, isw2Δ, chd1Δ) strain which are subsequently fitted with nucleosomes and then computationally phased.* ***(E)*** *Nucleosome footprint size distribution comparison for WT and TKO strains. Footprints exclusively found within gene bodies (TSS to TTS) are compared.* ***(F)*** *Center location for all mono- di- and tri-nucleosomal footprints. Counts indicate the number of footprint dyads at a particular position.* ***(G)*** *Ratio of reads with N+1 shifted or phased in TKO strain from low (1) to high (4) transcription rate based on Ocampo et al. RNA Pol II ChIP-seq DKO data (n=3)* ***(H)*** *The degree to which rDNA loci of individual samples are removed from transcriptional independence can be measured as an observable. For a detailed description see Figure 2.2.3E* ***(I)*** *Log10 base converted P-values of Fisher's exact test of transcriptional independence of rDNA loci in TKO strain. Log10 base converted P-value of 0.05 is indicated as dashed line.* ***(J)*** *Ratio of rDNA reads of left (RDN37-1) and right (RDN37-2) loci being fully active (≥50% accessible) or inactive (<50% inaccessible) in TKO strain (See NPD07 and NPD10 TKO1_120, TKO2_120, WT3_45 and WT3_90 samples of NPD07 and NPD10 in Appendix Supplementary table 3)* dyads in the TKO data does not provide a clear visual regularity as observed in WT (Figure 2.2.7D; Second panel). The computational phasing, which visually improved array regularity in WT, only has a mild effect in TKO (Figure 2.2.7D; Third panel). Visual observation does suggest a mildly improved +2 nucleosome regularity. However, the TKO strains do not appear to exhibit a regular array, even when computationally phased.

To further test whether computational phasing could reveal an underlying regularity in TKO strains we investigated the composite signal over all single-molecule reads. The irregular arrays observed by unphased composite plots (Figure 2.2.7A), could be due to different underlying phenotypes indistinguishable by methods such as MNase-seq. Arrays may be regular, and merely unphased, as suggested by previous MNase-seq studies on the TKO strain (Ocampo et al., 2016; Singh and Mueller-Planitz, 2021). Alternatively, arrays may be both unphased *and* irregular due to the lack of remodelers. Either scenario is formally a possibility when utilizing yield-based methods, such as MNase-seq. We utilized the computationally phased data (Figure 2.2.7D; Right panel) and calculated the composite signal over all reads (Figure 2.2.7B). We find that the computational phasing does not improve arrays to a similar extent as observed in WT. Compared to the unphased TKO composite plots, we do observe a slight improvement (Figure 2.2.7A-B; Red traces). This slight improvement after computational phasing might be due to the remaining remodelers in the cell such as INO80 (further addressed in 2.2.9). However, we conclude that cells lacking ISWI/CHD remodelers not only result a lack of phasing, but also exhibit little regularity downstream.

Next, we used our data to identify a potential change in global occupancy between WT and TKO strains. These remodelers have been implicated in nucleosome assembly *in vitro*, but *in vivo* data supporting assembly is slim (Fyodorov and Kadonaga, 2002; Lusser et al., 2005; Yadav and Whitehouse, 2016). We previously determined that the mean global occupancy levels do not differ significantly between WT and TKO cells (Figure 2.1.4A). In addition to this observation, we calculate the distribution of read occupancy as this can reveal alternative populations of reads within the data. For both WT and TKO strains we note a shift between sequencing runs similar to that seen between WT strains (Figure 2.2.7C and Figure 2.1.4B-C; NPD07 (x2) and NPD10, samples WT1/TKO1, WT2/TKO2 and WT3/TKO3, see Appendix Supplementary table 2 for mean read accessibility values). However, WT and TKO strains treated equally within the same sequencing run do not exhibit a different or additional distribution. This indicates that the differences are introduced by experimental variability and not strain background. Based on these observations, we conclude that deletion of ISWI/CHD remodelers does not detectably affect global occupancy. We will address this question in further detail later (See Chapter 2.2.13).

We then investigated how the phasing and regularity defect in TKO strains affects the position and distribution of nucleosomes. We hypothesized that the lack of phasing and regularity would impact both the nucleosome footprint distribution, as well as the location of footprint centers in the genome. Indeed, we find that the nucleosome footprint distribution is not as well defined compared to the WT strain (Figure 2.2.7E). This suggests that nucleosomes are no longer at ‘fixed’ distances from one another, further indicating irregular arrays. In support of these observations, we find that the center of individual mono-nucleosomal footprints is no longer at well-defined locations, but is smeared over a much broader region, indicating a lack of phasing (Figure 2.2.7F).

We next hypothesized that the lack of phasing and regularity would impact the ratio of reads with a shifted +1 nucleosome. Curiously, we do not find a difference in the ratio of reads with a shifted +1 nucleosome when compared to WT (16±5% versus 19±4%, TKO and WT, respectively). This lack of a differences suggests that, despite being unphased, the +1 nucleosomes does not shift more than half a nucleosome (Figure 2.2.4B). Similar to the WT data, we observe a significant difference between lowly and highly transcribed genes using RNA Pol II ChIP-seq data acquired from a DKO strain (*isw*1Δ, *chd1*Δ; Ocampo et al., 2016). In addition, we find that highly and lowly (quartile 1 and 4, respectively) transcribed genes have a higher ratio of shifted +1 nucleosomes compared to genes in the middle quartiles (quartile 2 and 3; Figure 2.2.7G).

Lastly, we asked whether the remodelers deleted in a TKO strain affect the independency of the 35S rDNA loci. The deleted ISWI/CHD remodelers are known to catalyze nucleosome assembly and formation of regular arrays. Thus, we hypothesized that a potential assembly defect could increase the ratio of fully accessible reads over the rDNA locus. Compared to the WT strain, the TKO strain exhibits a mildly higher fraction of loci with >50% methylation compared to wild-type (33±6% and 43±7%, WT and TKO, respectively. TKO observations based on four samples from NPD07 and NPD10 TKO1_120, TKO2_120, TKO3_45 and TKO3_90. See Appendix Supplementary table 3 for values) (Figure 2.2.7J). We attribute this mild increase to experimental variability and the comparatively low number of biological replicates. Both left and right rDNA loci retain a near identical fraction of reads with high accessibility (45±8% and 41±4%, left and right, respectively) suggesting further that the TKO strain does not change the overall observations we made in WT. Based on these observations we validated whether the TKO strain would influence the transcriptional dependence we found in WT. Our observations suggested that the TKO strains are similarly removed from independence as WT (Figure 2.2.7H). A Fisher's exact test was performed on the contingency tables and found a similar statistical significance (Figure 2.2.7I). We conclude that, despite the deletion of three nucleosome remodelers, the rDNA loci are not transcriptionally independent.

### 2.2.9. INO80 is a bona fide spacing remodeler

Besides the ISWI- and CHD-remodelers, INO80 has been shown to alter chromatin states *in vitro* by remodeling nucleosomes and fine-tuning the positioning of the +1 nucleosome (Krietenstein et al., 2016; Udugama et al., 2011). Additionally, INO80 facilitates the exchange of H2A.Z-H2B dimers (typically enriched at the -1 and +1 nucleosomes) by replacing them with H2A-H2B dimers (Brahma et al., 2017; Raisner et al., 2005). More recently, the chromatin remodeling role of INO80 has been demonstrated *in vivo* (Singh et al., 2021). We therefore wondered whether the residual regularity observed at the start of computationally aligned reads in a TKO strain, could be due to residual INO80 remodeling. To answer this question, we performed similar analyses as described above on a TKO strain where INO80 can be selectively depleted (hereafter referred to as TKO-INO80) by the anchor away technique. We hypothesized that if INO80 is a true nucleosome remodeler acting on the +1 nucleosome, we would see a further reduction in regularity.

We performed MEFSIMO-seq on TKO-INO80 strains and performed the nucleosome dyad fitting and computational phasing (Figure 2.2.8A). Compared to a TKO strain, we do not observe a clear pattern in the raw data or when nucleosome dyads are fitted. Computational phasing furthermore does not provide a direct indication that regular arrays are present in the underlying data.

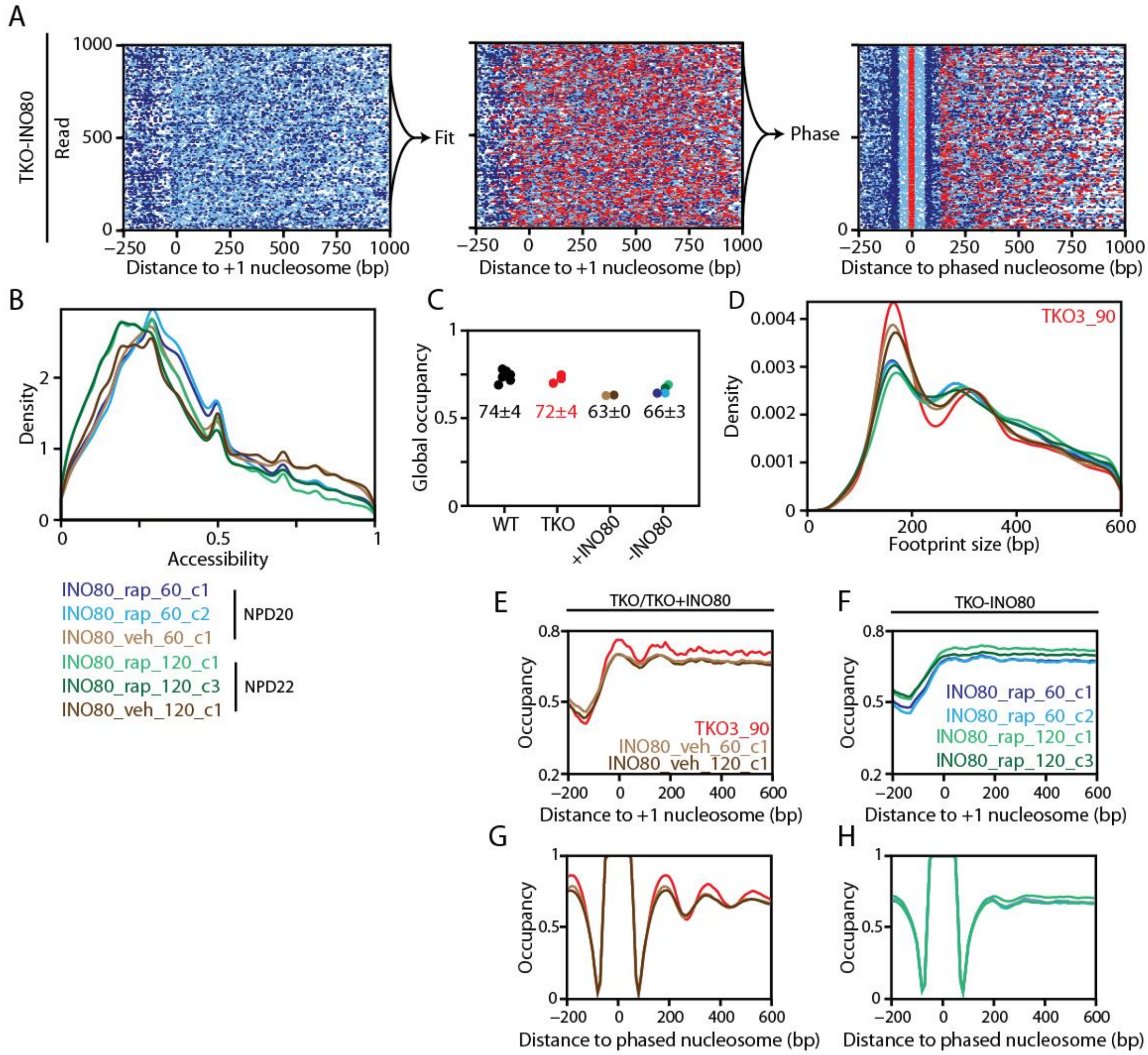


**Figure 2.2.8 INO80 is a nucleosome spacing remodeler and helps establish regularity at 5' of a gene**

***(A)*** *Raw MEFSIMO-seq data of 1000 random reads aligned to the predicted +1 nucleosome position. Nucleosomes are fitted within inaccessible regions (MUM-gaps). Subsequent computational phasing aligns all +1 nucleosomes to the 0 coordinate.* ***(B)*** *The read accessibility distribution of TKO±INO80 strains treated with a vehicle agent (+veh, controls TKO+INO80) or 1mg/ml rapamycin (+rap, depletion TKO-INO80) for 2 hours. Plotted is the distribution of a multiple biological replicate (clone 1-3) over two sequencing runs (NPD20/22). For global read statistics see Appendix Supplementary table 2. Sample legend is below figure.* ***(C)*** *Global occupancy for WT, TKO INO80 control and depleted samples. For WT and TKO strains, see individual samples in Figure 2.1.4A. Samples ±INO80 are colored as indicated in* ***(B)****. Mean and SD values are below individual datapoints. For values on global occupancy see Appendix Supplementary table 1.* ***(D)*** *Distribution of the nucleosome footprint (MUM-gap) for INO80 control and depleted samples. In addition, a single true TKO sample is added (TKO3_90, NPD10. For other TKO samples see Figure 2.2.7C). Sample line colors are identical to* ***(B)****.* ***(E-F)*** *Unfiltered composite plots of MEFSIMO-seq reads of TKO, TKO INO80 control strains and TKO INO80 depleted strains.* ***(G-H)*** *Computationally phased reads of filtered MEFSIMO-seq data. Only reads with a mononucleosome at the predicted +1 nucleosome position are plotted.*

Next, we verified that the depletion of INO80 does not alter the genome wide accessibility or the accessibility distribution (Figure 2.2.8B). We compared several TKO-INO80 strains to control TKO+INO80 strains (NPD20/22, Rap for depletion, Veh for control. For more details see Appendix Supplementary table 1-2). This latter strain is isogenic to the former, but no selective depletion by the anchor away system was performed, allowing for a proper comparison. We find that depletion of

INO80 does not alter the read occupancy density. Global occupancy is not different between isogenic strains (Figure 2.2.8C, 63±2% versus 66±3% for control and depletion, respectively). We do observe a reduced global occupancy for the TKO with which we can selectively deplete INO80 compared to the true TKO strain (yFMP333-335 to yFMP014, resp.), potentially due to the altered genotypical background or growth conditions (See Methods section 4.2.12 for growth conditions and Yeast strain list for genotype).

To further define the effect of INO80 depletion we investigated the nucleosome footprint distribution (Figure 2.2.8D). In our non-depleted control samples, we observe better defined mono- and dinucleosomal peaks compared to the rapamycin depleted samples. A defined tri-nucleosomal peak is not present in either control or depleted samples, similar to that observed for TKO strains. These differences suggest that depletion of INO80 results in reduced array regularity, similar to that seen when comparing WT and TKO. When the control sample is compared to a true TKO strain, we do find that mono- and dinucleosomal peaks are less defined. This altered pattern (and thus regularity) is potentially due to the effect of a C-terminal GFP and FRB tagging of the INO80 protein (See yeast strains yFMP333-335). Moreover, a mild growth defect and slightly altered composite signals after MNase-seq were observed for these strains as well (anecdotal evidence based on internal communication). These observations suggest that indeed INO80 affects genome regularity, in line with previous findings.

When genome wide composite plots are compared between TKO+INO80 strains and a 'true' TKO strain, we observe a reduction in regularity (Figure 2.2.8E). The TKO strain has slightly better defined +1 and +2 nucleosomal peaks compared to the TKO+INO80 strain, similar to that observed in the footprint distribution. Again, this reduction in regularity seen in the TKO+INO80 strain background is potentially due to the C-terminal GFP and FRB tag and growth defect. Depletion of INO80 leads to a further reduction in regularity, most noticeable at the start of the gene at the +1 nucleosome (Figure 2.2.8F). Based on these observations, it appears that +1 nucleosome phasing is further reduced in the absence of INO80. We do note that the lack of a +1 nucleosome is in contrast to MNase-seq data (Singh et al., 2021). Here, the +1 nucleosome remains clearly present when composite plots are generated. We attribute this difference to a potential MNase bias, which preferentially isolates the +1 nucleosome during these experiments. MEFSIMO-seq does not have such a bias, and would potentially provide the true occupancy of the +1 nucleosome (more in Discussion related to this chapter).

Lastly, we investigated if computational phasing would reveal any remaining underlying regularity in these strains. To this end, we initially performed the computational phasing on the control TKO+INO80 and true TKO strains. We observe a similar trend for TKO and TKO+INO80 strains where regularity improves after phasing (Figure 2.2.8G). Similar to the earlier observations in unphased samples, we do note a reduction in regularity for the TKO+INO80 strains when compared to a true TKO. Computational phasing of the TKO-INO80 strains does not reveal any noticeable regularity downstream of the +1 nucleosome (Figure 2.2.8H). From these observations and with supporting evidence from MNase-seq (Singh et al., 2021), we conclude that INO80 is indeed a true nucleosome spacing remodeler.

### 2.2.10. Genic nucleosome arrays display a translational and directional symmetry

Our observations by computationally phasing reads to the +1 nucleosome made us wonder whether this particular nucleosome holds a special role in array regularity. Precise positioning of the +1 nucleosome is the result of a tug-of-war between RSC-, ISW2- and INO80-remodelers (Krietenstein et al., 2016; Kubik et al., 2019; Parnell et al., 2015). The final position is a major factor for proper transcription and is therefore carefully regulated. Furthermore, barriers such as the NFR, PIC or the transcription machinery could further aide in a well-positioned +1 nucleosome to which downstream nucleosomes align. Conversely, the presence of H2A.Z containing nucleosomes at the promoter region of Pol II transcribed genes has been suggested to be less stable, although this observation has been contested (reviewed in Marques et al., 2010). Thus, the position, intrinsic properties and factors that control it could provide a particular role for the +1 nucleosome in an array.

We initially investigated whether regularity of arrays would differ by phasing reads to different genic nucleosomes. By computationally phasing wild-type MESIMO-seq data not solely to the first nucleosome, but to up to the fourth nucleosomes, we can distinguish if the regularity differs between them. To this end, we identified the dyad of all inaccessible regions (MUM-gaps) of mononucleosomal size and assigned them to the closest predicted nucleosome. We assigned each dyad to a nucleosome position (e.g. +1 nucleosome) based on its position within an aligned nucleosomal array. For example, a dyad at 15 bp downstream of the predicted +1 nucleosome would be referred to as the +1 nucleosome. Conversely, a dyad 180 bp downstream, would be assigned as a +2 nucleosome. From biological observations made in the past, we hypothesized that the +1 nucleosome could result in better regularity compared to other (i.e. downstream) nucleosomes (Struhl and Segal, 2013). However, from a physics perspective, no differences were to be expected if arrays consist out of merely canonical nucleosomes. Our findings indicate that, in line with the physics perspective, the pattern downstream of the +1 nucleosome was near identical to downstream patterns of other nucleosomes (Figure 2.2.9A). Based on this observation we find that WT cells exhibit a translational symmetry, implying that when phasing is shifted by one or more nucleosomes, regularity is not affected.

Next, we investigated whether this translational symmetry is also present in strains depleted of ISWI/CHD remodelers (TKO), or when INO80 is further depleted (TKO-INO80). Such strains lack many tools to properly position and regulate nucleosomes. As previously stated, these remodelers affect genome wide regularity, but are closely involved in regulating +1 positioning and composition. Therefore, deletion or depletion of these remodelers could result in a diminished/absent translational symmetry. Surprisingly, similar to the WT strain, near identical patterns were observed irrespective of the aligned nucleosome, albeit with a faster decay in regularity (Figure 2.2.9B-C). Thus, when challenging the system by deletion/depletion of these remodelers, the position (and potentially composition) of the +1 nucleosome is altered. However, a translational symmetry is retained. This implies that the intrinsic properties and positional regulation of the +1 nucleosome do not affect downstream regularity differently than other nucleosomes.

From the observations that suggested a translational symmetry, we observed a directional symmetry emerging in both directions from the computationally phased nucleosome (coordinate 0). This was especially apparent when arrays were aligned to downstream nucleosomes such as the +3 and +4 nucleosomes. We therefore wondered whether, in addition to a translational symmetry, such arrays also display a directional symmetry. This observation would be in contrast to the 5’-3’ asymmetry that was proposed to stem from the transcription by the Pol II machinery (Struhl and Segal, 2013). Previous computational modeling experiments identified a symmetry emanating from the gene center by allowing a fixed amount of nucleosomes between the 5’ and 3’ NFR (Vaillant et al., 2010). The modeled data corresponded well with *in vivo* data (Lee et al., 2007). However, due to a lack of single-molecule datasets that are not averaged over many cells, one could not directly test this symmetry. Using our

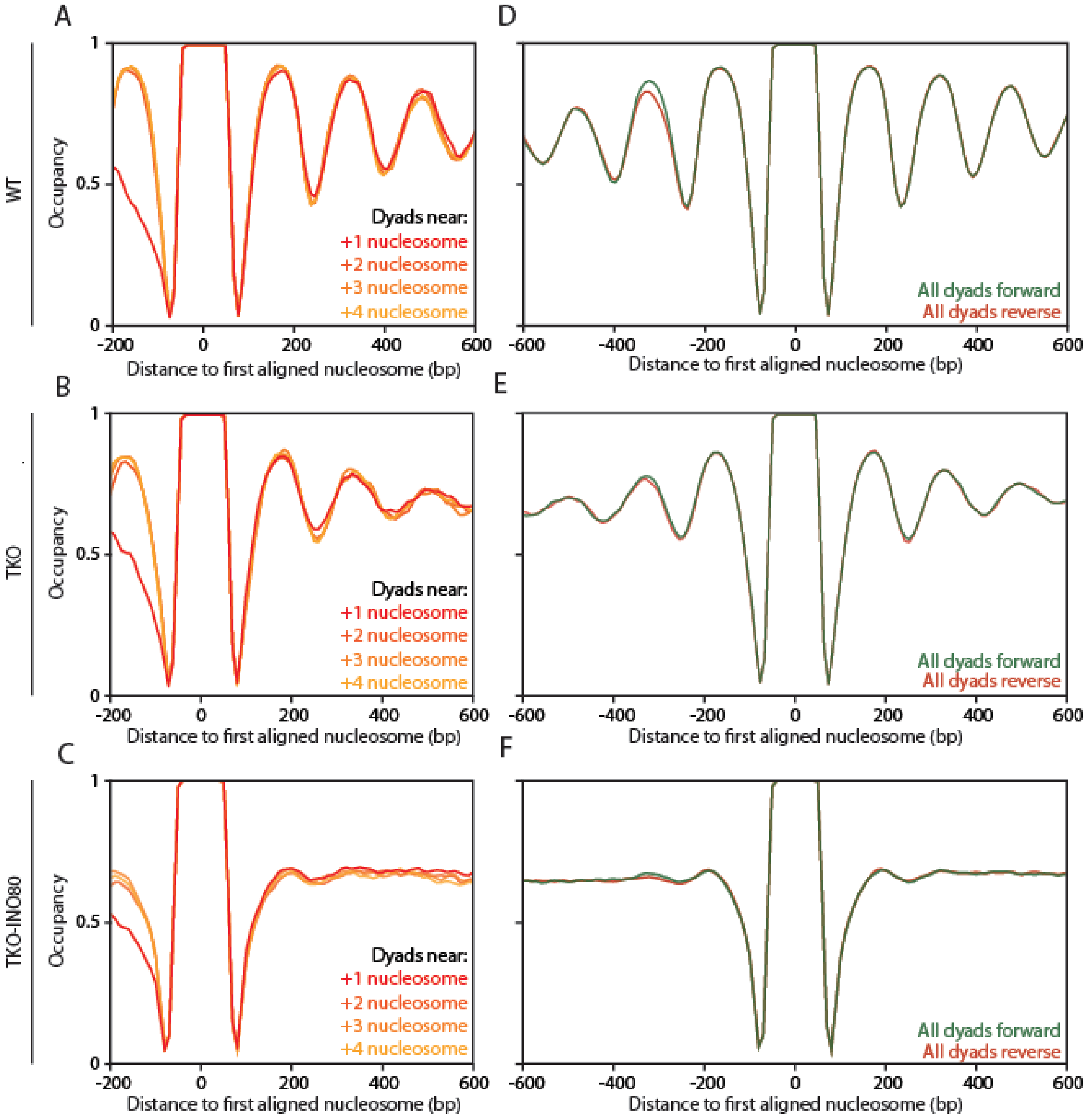


**Figure 2.2.9 Genic arrays display both a translational and directional symmetry not mediated by nucleosome remodelers**

***(A-C)*** *Computational phasing of +1 to +4 nucleosomes in order to visualize translational symmetry. Mononucleosomal dyads closest to the predicted site of a genic nucleosome are treated similarly and aligned to coordinate 0. Plots and data generated by Dr. Michael Wolff.* ***(D-E)*** *Combined dyads of mononucleosomal gaps from all reads are computationally phased. All dyads in forward direction always have the 3’ nucleosome at relative higher coordinates. All dyads in reverse direction have the 3’ nucleosomes at relative lower coordinates. Reads are cut off at the +1 nucleosome and 250 bp upstream of the transcription termination site (TTS). Plots and data generated by Dr. Michael Wolff.*

MEFSIMO-seq data, we now can test whether a directional symmetry is inherent to nucleosome arrays.

To investigate the directional symmetry, we performed an alignment of nucleosomes identical to that performed in the translational symmetry analyses. However, for this analysis we did not discriminate between the positions of dyads within the array and compiled all data into a single observable. To accurately determine a directional symmetry, we excluded the NFR as this region is highly variable in nucleosome content and would skew the symmetry. We performed the analysis for WT, TKO and TKO-INO80 strains by aligning all dyads contained in the reads in the forward (5'-3') direction (Figure 2.2.9D-E, green traces). As suggested earlier, we observe an apparent directional symmetry emanating from the center. To identify if reads were directionally symmetrical, we performed the same analysis, but calculated the composite signal in the 3'-5' direction (Figure 2.2.9D-E, red traces). We hypothesized that if arrays were truly directionally symmetrical, the composite signals should have a high degree of overlap. Indeed, overlapping the forward and reverse composites data, we observe identical patterns, irrespective of the direction. These findings show that there is a clear unidirectional symmetry in genic arrays unbiased by the Pol II machinery. Moreover, the directional symmetry is not affected by the absence or presence of nucleosome remodelers.

### 2.2.11. *In vivo* data does not support a clamping mechanism by the ISWI/CHD remodelers

We next investigated the mechanism by which the ISWI/CHD remodelers define the distance between nucleosomes. Previous *in vitro* findings using SGD assembled chromatin at variable nucleosome density resulted in clusters of tightly packed nucleosomes when incubated with ISWI- and/or CHD-remodelers (Lieleg et al., 2015). These *in vitro* observations suggested that these remodelers exhibited a density independent 'clamping' activity to fix the distance between nucleosomes (Figure 2.2.10A). The data was suggested to be in contrast to a density dependent mechanism, in which nucleosome spacing would simply decrease equally when the density of nucleosomes increases (Möbius and Gerland, 2010). We aimed to validate this 'clamping' mechanism *in vivo* by utilizing MEFSIMO-seq data. Utilizing this approach allows us to identify clusters of nucleosomes by identifying the position of nucleosome dyads which could be suggestive of clamping on a per-read basis.

To challenge the clamping activity of these remodelers, we utilized a strain where we could induce a histone depleted (HD) state. We chose this approach as the native nucleosome density is relatively high in *S. cerevisiae* (70% – 75% globally). This high density may not lend itself to detecting a difference in clamping when comparing a WT and TKO strain. Moreover, simulations have shown that an average regularity of nucleosome clusters is obtained despite the influence of remodelers (Rube and Song, 2014; Vaillant et al., 2010). We achieved a reduced global nucleosome occupancy by employing a previously generated yeast strain (Mann and Grunstein, 1992; Singh et al., 2021). This strain has the two genic copies of histones H3 and H4 deleted, and instead contains a plasmid with a copy of the H3 and H4 gene under a galactose inducible promotor. Depletion of the histones can be achieved by switching to a 2% glucose-containing medium for 120 minutes (see Methods) (Figure 2.2.10B). We hypothesized that under HD conditions, ISWI- and CHD-remodelers could form clusters of nucleosomes, similar to *in vitro* clamping observations at low nucleosome densities. Conversely, deletion of these remodelers (TKO+HD) would result in reduced clustering, perhaps reverting to a density dependent model (Figure 2.2.10A)

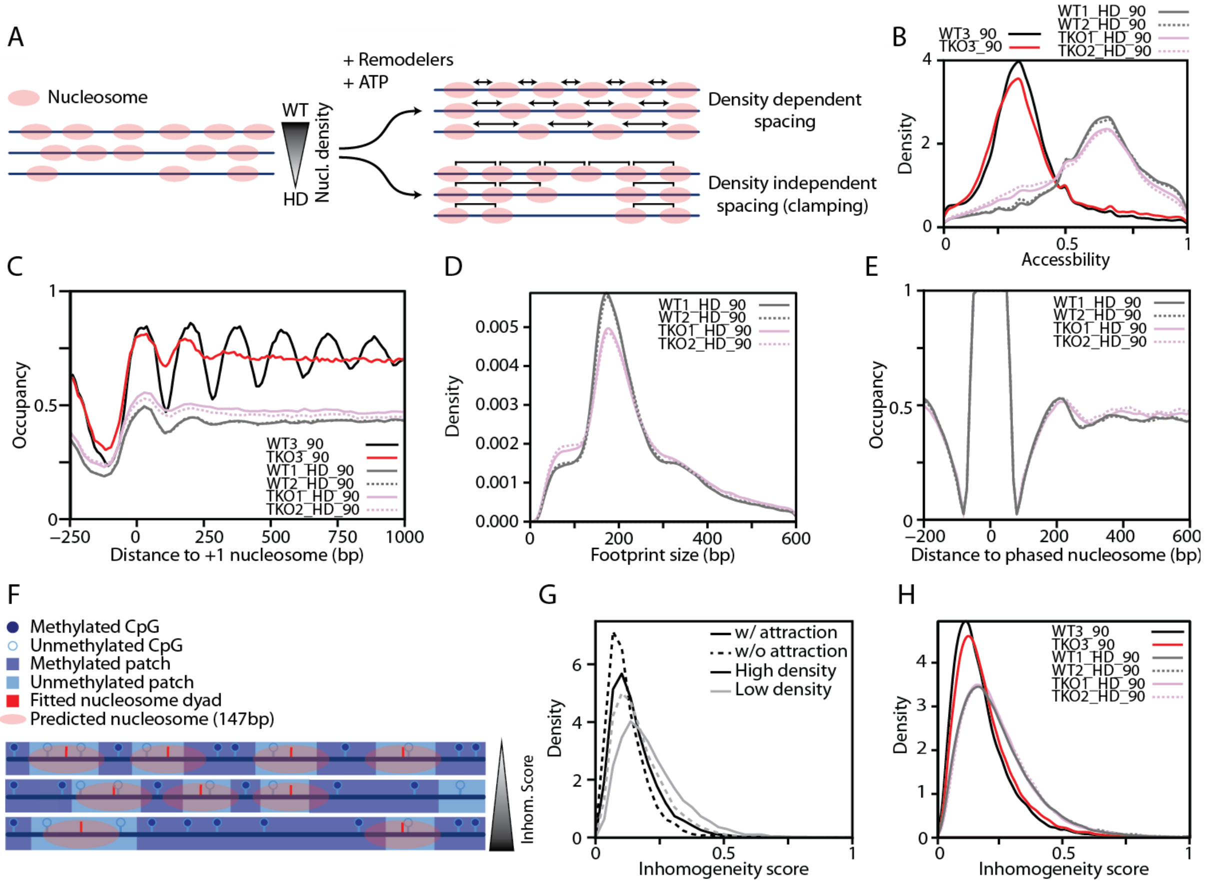


**Figure 2.2.10 Clamping mechanism for nucleosome remodelers is not supported by *in vivo* data**

***(A)*** *Cartoon outline of two proposed mechanisms nucleosome remodelers can create regularity in vitro. Nucleosome organization is randomly distributed at wild-type (WT) and histone-depleted (HD) nucleosome densities after SGD assembly. The addition of remodelers and ATP may create regularly spaced nucleosomes by a density dependent or independent manner. Figure adapted from Lieleg et al., 2015.* ***(B)*** *Distribution of read accessibility for regular density WT and TKO strains, compared to HD samples in which all remodelers are present (HD) and a histone-depleted strain in a TKO (isw1Δ, isw2Δ, chd1Δ) background (TKO+HD). Dashed lines for HD and TKO+HD indicate biological replicates. For mean values see Appendix Supplementary table 2 sequencing run NPD10.* ***(C)*** *Genome wide average composite plots of MEFSIMO-seq data for WT and TKO strains at regular and low density. All reads are aligned to the predicted +1 nucleosome position for each gene of a WT non-HD strain (Chereji et al., 2018).* ***(D)*** *Nucleosome footprint distribution for HD and TKO+HD strains. Dashed lines indicate biological replicates.* ***(E)*** *Genome wide average composite plots after computational phasing of filtered MEFSIMO-seq data for HD and TKO+HD strains. Plots consist of phased reads for which the first phased nucleosome originates from an inaccessible region of mononucleosomal size. Dashed lines indicate biological replicates.* ***(F)*** *Computational strategy for calculating inhomogeneity score. (Un)methylated CpG sites are extended into 'patches' that intersect at equidistance from either site. The underlying (non-CpG) nucleotides are assigned the methylation status of the closest unambiguous CpG site. A sliding window (500 bp) counts all unmethylated nucleotides in each window. The variance of each window per GBR is calculated and divided over the maximum inhomogeneity score of that GBR:* $Inhomogeneity\ sc. = \frac{Number\ variance\ of\ GBR}{Num.\ var.\ of\ max.\ inhom.\ of\ same\ GBR}$*. Reads with a 'regular' array are least inhomogeneous, whereas large regions of accessibility result in high inhomogeneity.* ***(G)*** *In silico simulations of inhomogeneity scores for low (grey) and high (black) nucleosome density. Nucleosome clamping is simulated by modulating the attraction between nucleosomes. Simulations with attraction (solid traces) mimic WT scenarios, whereas without attraction (dashed traces) mimics ISWI/CHD deleted conditions. Data generated by Dr. Michael Wolff and Matthias Hanke* ***(H)*** *Inhomogeneity score density distribution of filtered MEFSIMO-seq data. Reads with an arbitrarily set bad resolution ratio (BRR) > 0.25 are filtered out. The BRR is determined based on the sum of badly resolved regions (i.e. regions with >150 bp between unambiguous CpG sites), over the GBR read length:* $BRR = \frac{\Sigma\ badly\ resolved\ regions}{GBR\ read\ length}$*. For a detailed description see Chapter 3.4 of Matthias Hanke's thesis. Data generated by Dr. Michael Wolff and Matthias Hanke.*

We first validated that HD and TKO+HD strains resulted in a similar phenotypical composite plot as observed. Indeed, we find that genome wide composite plots show that regularity is severely reduced in HD and TKO+HD cells (Figure 2.2.10C). Similar to TKO cells, phasing of the +1 nucleosome is abrogated and downstream regular arrays are largely absent. We do note a slightly improved phasing and regularity for HD strains compared to TKO+HD, which is in line with previous findings. This minor improvement could be due to a density independent clamping activity of the nucleosome remodeler. Alternatively, this slight improvement could be the result of equilibration between linker lengths in a density dependent manner and aided by remodelers. Regardless, the HD strain demonstrates that under these conditions, arrays are not phased properly.

Next, we were interested how the nucleosome footprint size distribution looked like in the HD and TKO+HD strain. Despite the lack of phasing and regularity in composite plots, we could not rule out the clamping hypothesis. We hypothesized that strong clamping would lead to a nucleosome footprint distribution similar to WT conditions, as this observable does not rely on phased arrays. In contrast, we observe little to no dinucleosomal peaks in the HD strain (Figure 2.2.10D). However, the TKO+HD strain shows a further broadening of the mononucleosomal peak and a lack of a dinucleosomal peak. This could be suggestive of a clamping activity, albeit weakly.

We further hypothesized that if clamping would (strongly) induce clusters of nucleosomes, we could detect them as regular arrays after computational phasing. Minimally, if clamping activity were weak, we would observe an improved array regularity. We therefore phased all mononucleosomal gaps, irrespective of its genic location, in a similar fashion as performed during the directional symmetry analyses (Chapter 2.2.10). The phasing analysis resulted in a mild nucleosomal peak downstream of the phased nucleosome (Figure 2.2.10E). In contrast to the proposed hypothesis however, we observe only a very mild increased regularity for HD samples when compared to TKO+HD samples and could be indicative of some form of clamping (Figure 2.2.10E). This finding is in line with the broad distribution of the nucleosomal footprints and the mildly broader distribution seen for TKO+HD. Based on the observations made above, we argue that clamping activity may formally be possible, but is only weakly clustering nucleosomes.

To further probe for evidence on nucleosome clamping, we employed a different computational strategy to detect any clamping activity on a per-read basis. This analysis was conceived and carried out by Dr. Michael Wolff and Matthias Hanke (both under supervision of Prof. Dr. Ulrich Gerland at the time of the analysis). We selected genes where the region between the TSS and TTS (referred to as gene body read or GBR) passes a certain ‘resolution ratio’. This ratio is a combination of the optimal density and distribution of CpG sites within the GBR. This parameter provides best case scenarios in terms of nucleotide resolution for downstream analyses. Next, non-ambiguous CpG sites were converted into so-called ‘patches’. This conversion results in an assigned hit state (i.e. methylated or unmethylated) for each nucleotide akin to the closest CpG site (Figure 2.2.10F). Then, within a window of 500 bp, each unmethylated nucleotide is counted. This window is displaced in one bp increments and the variance of all windows within a GBR is calculated. This value is divided by the number variance of a read with a maximally inhomogeneous configuration of the same length and occupancy (See Methods). Together, these calculations produce an observable we refer to as the ‘inhomogeneity score’ of a read. Arrays with high regularity will result in a low inhomogeneity score, whereas sparsely distributed nucleosomes or regions of low occupancy result in high inhomogeneity scores.

Initially we performed this inhomogeneity score analysis *in silico* to test whether this score would indeed identify clamping. To this end we utilized data where clamping can be simulated by modulating the attraction between nucleosomes (Möbius et al., 2013). In a scenario without attraction (simulating a TKO scenario) at native nucleosome density (high density), we observe little to no difference when compared to a simulation with attraction (akin to a WT scenario) (Figure 2.2.10G, High density). As expected, this suggests that at high nucleosome densities, clamping is difficult to detect or discern between attraction scenarios (i.e. WT vs TKO). However, at histone depleted conditions (low density), we observe a higher inhomogeneity score when the attraction is induced to mimic a scenario where remodelers are present (Figure 2.2.10G, Low density). This suggests that the inhomogeneity score can detect clamped nucleosome clusters using *in silico* models.

Finally, we performed the inhomogeneity score analysis on *in vivo* MEFSIMO-seq data of WT and TKO strains. Similar to the simulated data, at high nucleosome densities we detect no changes in the inhomogeneity score within sequencing runs of WT and TKO strains (Figure 2.2.10H). In contrast to the *in silico* observations, under HD and TKO+HD conditions, no differences in inhomogeneity score are detected (Figure 2.2.10H). Contrary to our hypothesis, previous *in vitro* findings and simulated *in silico* data, we find little support in favor of the clamping activity proposed in previous *in vitro* studies.

### 2.2.12. GRFs binding reduces array regularity at flanking regions

Aside from the tight regulation of nucleosome position and spacing by nucleosome remodelers, other factors referred to as general regulatory factors (GRFs) are known to organize nucleosomes. GRFs can serve as barriers to which nucleosome arrays are aligned and are often found at gene promoters (Gutin et al., 2018). Consistent with their role as barriers, depletion leads to higher nucleosome occupancy within the NFR and a reduction in well-phased +1 nucleosomes (Kubik et al., 2015, 2018). Based on these observations it is clear that GRFs play an important role in genome wide chromatin organization. However, the absence of a GRF at a specific locus can only be determined by performing a depletion (e.g. by means of anchor-away experiments) of the GRF of interest. These studies could not determine what the effect of unbound GRFs was under wild-type conditions. Here we aimed to study the difference in chromatin architecture in the presence or absence of GRF under wild-type conditions.

In order to determine the effect of GRF binding under wild-type conditions we initially compared the genomic landscape surrounding three GRFs: Abf1, Rap1 and Reb1. We performed MEFSIMO-seq on isolated nuclei that were prepared from log-phase yeast with and without formaldehyde crosslinking. We hypothesized that the addition of the formaldehyde crosslinking could capture more (if not all) bound GRFs. Our average genome wide results indicated that the addition of crosslinking did not affect the chromatin landscape. Moreover, samples did not differ from one another significantly by several parameters, despite the presence (or absence) of formaldehyde during the preparation of nuclei. Consistent with previously reported findings, we observe that Rap1, and to a lesser degree Abf1, show an increased occupancy at the reported binding site (Figure 2.2.11A). In contrast, without the addition of formaldehyde, binding of GRFs was not apparent. Similar across the two samples is the observation that Abf1 and Reb1 have a regular array emanating from the binding site. Despite the observed binding of Rap1 in formaldehyde crosslinked samples, there is no discernable pattern emanating from the binding site in either sample.

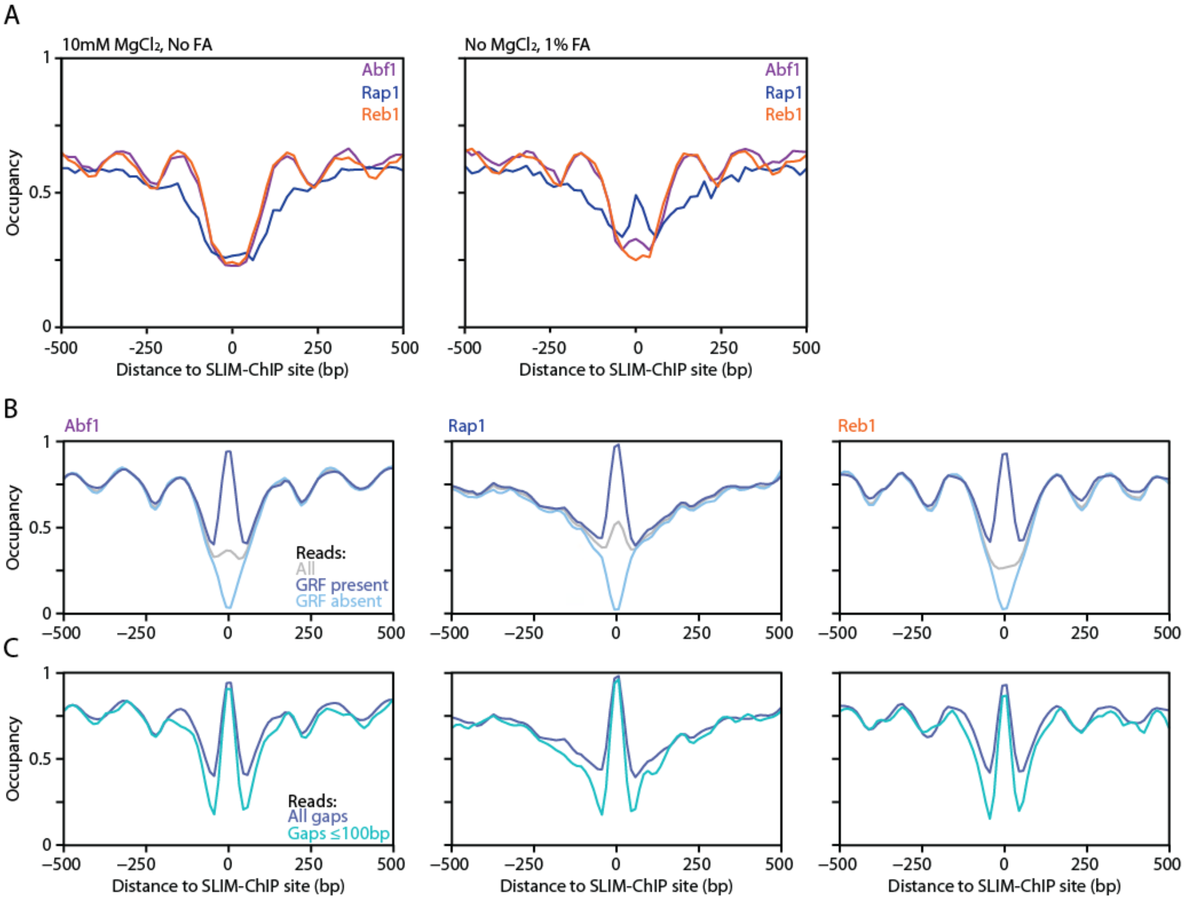


**Figure 2.2.11 GRF binding reduces array regularity at flanking regions**
***(A)*** *Left: Genome wide composite signal of arrays flanking GRF binding sites, Abf1, Rap1 and Reb1. Preparation of nucleosomes was performed without the addition of formaldehyde crosslinking (No FA). During methylation, buffer was supplemented with 10mM* $MgCl_2$*. Right: Same genome wide composite signal of GRFs. Nuclei preparation was performed with the addition of 1% Formaldehyde and methylation was performed without the addition of* $MgCl_2$*.* ***(B)*** *Separating individual reads based on the methylation status at and around the GRF binding site. Sites with inaccessible CpG sites (i.e. unmethylated) are categorized as GRF present. Conversely, accessible sites (i.e. methylated) are categorized as GRF absent. Reads within each category are utilized to generate genome wide composite plots of the region flanking the binding site. The two categories are compared to an unfiltered genome wide average. Data acquired from formaldehyde crosslinked samples.* ***(C)*** *Reads categorized as GRF present are further filtered based on the size of the inaccessible region in which the GRF binding site is located. Inaccessible regions smaller or equal than 100bp are filtered and plotted separately (Gaps ≤ 100bp). Data is overlayed with reads where GRF is present (All gaps).*

We next utilized the single-molecule capabilities of MEFSIMO-seq to examine the effects of GRF binding on nucleosome regularity. We hypothesized that the presence of a GRF, would increase array regularity flanking the binding site. To this end, we took all unfiltered MEFSIMO-seq data and probed the methylation state of CpG site(s) in a window of 20 bp surrounding the GRF binding site. Similar to our approach in assessing the state of the +1 nucleosome, we categorized reads with only unambiguous and unmethylated site(s) within the windows as GRF present. Conversely, reads in which the CpG site(s) within the window were methylated are categorized as GRF absent. Curiously, our data suggest that the presence of a GRF results in a reduction of array regularity for both types of samples (Figure 2.2.11B).

Next, we wondered whether the reads we categorized as GRF present, could have a nucleosome bound instead of a GRF. The presence of a nucleosome instead of a GRF could result in a reduced array regularity as the barrier effect of a GRF may be much stronger. Moreover, GRFs can facilitate nucleosome remodeler recruitment, aiding in the establishment of a regular array. We isolated reads in which the inaccessible region was no larger than 100 bp. This would largely exclude the possibility of a canonical nucleosome being present. Moreover, it most likely includes only bound GRFs as the binding motif of Abf1, Rap1 and Abf1 is between 10-20 bp in width (Gutin et al., 2018). The array regularity of reads with a GRF bound versus the pre-filtered reads do not shown improvements, nor resemble regularity of reads without a GRF bound (Figure 2.2.11C). These findings further suggest a decrease in regularity upon binding. Our findings thus indicate that, contrary to the current understanding, GRF binding has an (initially) negative effect on array regularity. We do not exclude the possibility that GRF binding requires remodeler recruitment before a regular array is attained (more in Discussion related to this chapter).

### 2.2.13. Yeast strain variants can be uniquely identified by MEFSIMO-seq

Throughout this thesis I utilized the W303-1A yeast *S. cerevisiae* strain that was originally created to study polymerase correlated recombination rates (Thomas and Rothstein, 1989). However, for more practical reasons, such as the availability of a TKO and anchor away background, we decided to use this as our model organism. The W303 strain is closely related to the commonly used S288C yeast strain and is practically identical with only approximately 0.5 variations per kb. In total, there exist ~9500 variations (SNPs or indels) affecting the protein sequence of ~700 genes (Matheson et al., 2017). We identified that the absolute nucleosome occupancy values of the W303 strain was akin to those previously published data utilizing an S288C strain (Oberbeckmann et al., 2019; Figure 2.1.4A). However, we aimed to verify that indeed these differences were negligible on a global scale, with the possibility to compare occupancy differences on a local (i.e. genic) scale.

To sensitively investigate the differences between yeast strain variants we decided to develop and perform a 'mix-n-match' experiment. This strategy aims to physically combine two yeast variants in a single tube, whereas previously we would prepare and sequence each variant as an independent sample. The added benefit of such an approach would be that we can directly compare mutants generated in different variants. This annuls any experimental variations such as reaction volume, enzyme concentration, contaminations etc. further improving consistency and reproducibility. However, the requirement for a successful comparison is that individual reads originating from either variant can be uniquely traced back to its source.

To explore the feasibility of this mix-n-match approach, we initially performed an *in silico* experiment by taking previously acquired W303 (this thesis) and S288C (Oberbeckmann et al., 2019) data. Equal amount of reads from either dataset were combined into a single dataset to mimic an *in vivo* sample. Next, this dataset, and the individual datasets, were mapped to a custom reference genome in which the S288C and W303 reference genomes were combined back-to-back. We find that the individual datasets map fairly accurately to their respective genome of origin, even when the lowest (Q1) read mapping quality (mapQ) filter is applied (Figure 2.2.12A). In contrast, lack of a read quality filtering step (Q0) results in many reads being mismatched to the wrong genome, mapped to both genomes, or have generally poor read quality. To this end, we apply a medium (Q5) mapping quality filter to remove the bulk of mismatched or poor quality reads. More stringent filtering steps based on

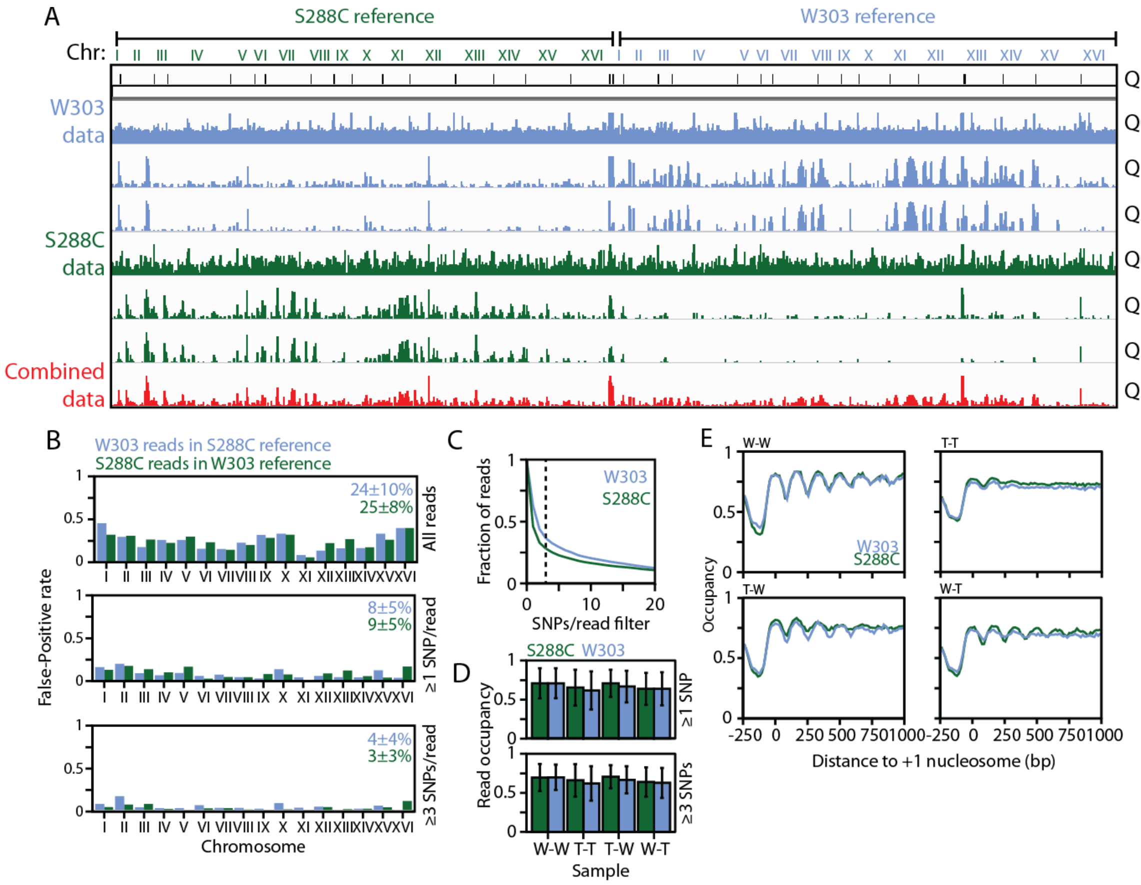


**Figure 2.2.12 Simultaneous mapping of yeast variants using MEFSIMO-seq**

***(A)** The S288C (SacCer3) genome was combined with the W303 genome to create a new combined genome. In vivo W303 data (personal, blue signal), S288C data (Dr. Elisa Oberbeckmann, green signal) or a combination (red signal) acquired through MEFSIMO-seq was mapped to the combined genome. No (Q0), low (Q1) or medium (Q5) MAPQ read filtering (Software: Samtools) was applied to select reads with good mapping quality. **(B)** As the origin of each read is known during the in silico analysis, false-positive rates could be calculated for each chromosome. Top shows the false-positive rate for all reads passing a Q5 read filtering without applying SNP filter. Middle shows filtered reads which contain ≥1 SNP site. Bottom are reads with ≥3 SNP sites per read. Values inside plots are mean false-positive rates for all reads per chromosome within opposite reference genomes. Errors are standard deviation between false-positive rates over all chromosomes. **(C)** Reduction of total fraction of Q5 filtered reads upon filtering in vivo data by SNP occurrence per read. Fraction of reads mapped to the corresponding reference genome is plotted. Dashed line indicates the fraction of reads when applying a ≥3 SNP/read filter. **(D)** In vivo data of a preliminary mix-n-match experiment where WT (W) and TKO (T) samples are combined from W303 (blue) and S288C (green) strains in four combinations. Reads were mapped to the combined reference genome, a read quality filter of Q5 was applied and reads were selected which contain ≥3 SNPs. The mean occupancy for each read mapping to a distinct reference genome is calculated. Error bars represent standard deviation within the sample (n=1). **(E)** Genome wide composite signal of selected reads from **(C)**.*

SNP/indel occurrence will be applied in downstream analysis and are thus not necessary here. This approach also allows for more flexible filtering parameters later whilst preventing the discarding of potential true-positive reads. Lastly, the *in silico* combined dataset unsurprisingly maps to both sides of the reference genome as it contains reads originating from both genomes. We utilized this *in silico* dataset for further filtering optimizations and read selection.

To identify how many reads are mapped correctly we were able to measure the false-positive rate after processing, since the origin of each read is known. False-positives are reads which are mapped to the correct genomic location, but of the wrong reference genome. We find that there is a relatively high false-positive rate for chromosomes (e.g. ChrI and ChrX) with low SNP/indel correlated to the chromosome size (Figure 2.2.12B, Top) (Matheson et al., 2017). Conversely, chromosome XI, which has a high degree of SNP/indels compared to its size, has a relatively low false-positive rate. This suggests that reads typically map to the correct genome when a SNP/indel is present within the read. Indeed, when reads are removed that do not contain any SNPs/indel, the false-positive rate is reduced by 50-75% for all chromosomes (Figure 2.2.12B, Middle). Mean false-positive rate decreases from 25±10% to 8±5% when reads without any SNPs are filtered. Filtering for reads with minimally three SNPs, reduces the false-positive rate further to 4±4% (Figure 2.2.12B, Bottom). This indicates that the largest source of reads mapped to the incorrect genome originate from reads lacking a SNP/indel. However, despite ideal circumstances *in silico*, certain chromosomes still have approximately 10-15% of reads mapped incorrectly. Taken together, the *in silico* results from Figure 2.2.12A-B show that it is technically feasible to perform a mix-n-match experiment.

Next, we prepared *in vivo* samples which contain both S288C and W303 cells. Prior to nuclei preparation, we pooled log-phase growing samples of the same $OD_{600}$ in four different combinations (WT-WT, WT-TKO, TKO-WT, TKO-TKO, for details see Methods). These combinations further allowed us to probe any potential assembly defect of the TKO strain. We subsequently processed them using the standard experimental procedures in order to methylate accessible DNA. After sequencing, the data was mapped to the combined genome created for the *in silico* analysis. Since read information on the origin of strain is unavailable here, we relied on the inherent mapping to accurately assign each read to the correct reference genome. Using existing tools, we split the mapped reads per genome in order to process them in parallel but separately (for details see Methods). Our *in silico* analysis suggested that false-positive rates decrease when only considering reads with SNPs. To this end we filter reads that do not overlap with the location of a SNP, similar to our approach *in silico*. Our *in vivo* findings suggest that approximately 40-55% of reads do not overlap with a SNP site. Curiously, we find that more reads of the S288C reference genome are removed when filtering for at least one SNP (55% versus 40%, S288C versus W303, respectively). This is potentially due to a bias in assigning non-SNP reads to the S288C reference genome. More stringent filtering results in a further decay with increased SNP occurrence (Figure 2.2.12C). Utilizing this filtering we can select reads which contain ≥3 SNPs. Such reads most likely mapped to the correct reference genome based on our *in silico* analysis.

Using the filtering steps described above, we next investigated whether the mean methylation differs between yeast variants and mutants. When applying a filter that removes reads containing less than three SNP sites per read, we observe a mean read occupancy of 67±3% across all S288C and W303 variants/mutants (Figure 2.2.12D). Within a sample we observe a typical read occupancy standard deviation of approximately 22%. Mean read occupancy difference within samples of the same mix-n-match experiment is at most 4% (Figure 2.2.12D, W-T sample). These variabilities are similar to the spread we observed in previous experiments (Figure 2.1.4A). Mean read occupancy values between variants and mutants remain within 1%, independent of altered SNP filtering stringency. Composite plots of WT/TKO combinations (filtered for ≥3 SNPs) indicate that the reads are typically mapped correctly as the TKO samples exhibit typical genome irregularity (Figure 2.2.12E).

Thus, based on this preliminary experiment, we conclude that the occupancy between these two variant yeast strains is within 2±2% using a ≥3 SNP filter. Increasing (≥10 SNP) or decreasing (≥1 SNP) the SNP filter stringency does not alter the variability (2±2% for both). Experimental procedures may have suggested otherwise, but utilizing this sensitive mix-n-match approach we can refute these observations directly. Additional biological replicates or fine-tuned filtering parameters may be necessary to conclusively answer this question. Moreover, better mapping and lower false-positives may benefit from optimized experimental conditions better suited for retaining ultra long reads (see Discussion related to this chapter).

### 2.2.14. The role of chromatin remodelers during the cell cycle

The experiments performed above are all based on log-phase growing cells. This implies that at the time of nuclei preparation, cells can be found at any stage of the cell cycle. Throughout the cell cycle however, chromatin structure and landscape undergo many changes in order to multiply. For example, during the replication process, nucleosomes are rapidly disassembled and subsequently reassembled in a process referred to as histone recycling (reviewed in Stewart-Morgan et al., 2020). Indeed, accessibility during S-phase (during which replication occurs) was observed to be higher based on limited MNase digest. This suggests that histone recycling alters the chromatin landscape, which is corrected by factors such as nucleosome remodelers that are implicated in re-establishing the chromatin landscape in the wake of replication (Smolle et al., 2012; Vasseur et al., 2016). It was further observed that nucleosomes are better positioned during gap phases (G1 and G2) compared to M- and S-phase (Deniz et al., 2016). Globally, only minimal differences in array regularity and NRL were observed when the individual cell cycle stages were directly compared using the existing data (Singh et al., 2021 supplementary data). We therefore wondered whether MEFSIMO-seq could be utilized to directly analyze how the chromatin landscape changes throughout the cell cycle. Moreover, we could sensitively measure the global nucleosome occupancy change over time, and directly examine the role of remodelers after replication (Figure 2.2.13A).

We initially performed a cell cycle arrest of WT and TKO cells using the mating pheromone alpha factor and ensured proper arrest. To this end, we deleted the *BAR1* gene, which codes for a protease that cleaves and inactivates alpha factor. Deletion of the *BAR1* gene from WT and TKO cells increases sensitivity to the mating pheromone (MacKay et al., 1988). Briefly, a large volume of log-phase growing cells was split in two equal parts and 1 mg/mL alpha factor was added to one part. Both parts were further incubated for 90 minutes after which they were treated with azide and formaldehyde to inhibit ATP-hydrolysis and fix the cells, respectively. Each part was analyzed by light microscopy and FACS in order to determine full arrest of alpha factor treated samples (Figure 2.2.13B-C). We find that cells are efficiently arrested at G1 phase when using these conditions for both WT and TKO strains.

Next, cells were methylated using standard MEFSIMO-seq procedures in order to investigate the chromatin landscape of arrested cells and detect any potential global occupancy differences. We find that WT arrested samples exhibit slightly elevated global occupancy values with respect to previous biological replicates (80±2% versus 74±4% for other WT samples, see Appendix Supplementary table 1 WT1/2_120_a samples of NPD13). These elevated values could be due to the slight undermethylation observed in these samples (~20%). Comparatively, WT unarrested samples tend to have a high degree of undermethylated reads (30% and 44% for WT1_120 and WT2_120, respectively. See Appendix Supplementary table 2, sequencing run NPD13) resulting in high global occupancy levels and therefore making a direct comparison difficult. TKO samples exhibiting <20% undermethylation all have similar global occupancy levels (76%-79%, TKO1_120, TKO1_120_a, TKO2_120), suggesting that arrest does not alter the global occupancy more than the standard deviation seen in other WT and TKO samples (Figure 2.1.4A). These preliminary findings based on a single sequencing run and a limited number of samples, do not suggest that arrest has an effect on the global occupancy of cells.

We validated previous findings indicating that arrest does not alter array regularity by plotting the composite signal of all TKO samples with <20% undermethylation. For WT samples, we plot all data, despite a high degree of undermethylation for unarrested samples in order to make the comparison using this limited dataset. Due to the undermethylation degree, we see a reduced peak-trough ratio for unarrested WT samples when compared to arrested WT samples (Figure 2.2.13D). These differences in undermethylation and peak-trough ratio make a direct comparison difficult and would thus need future replication. We hypothesized that arrays could be improved in arrested TKO samples

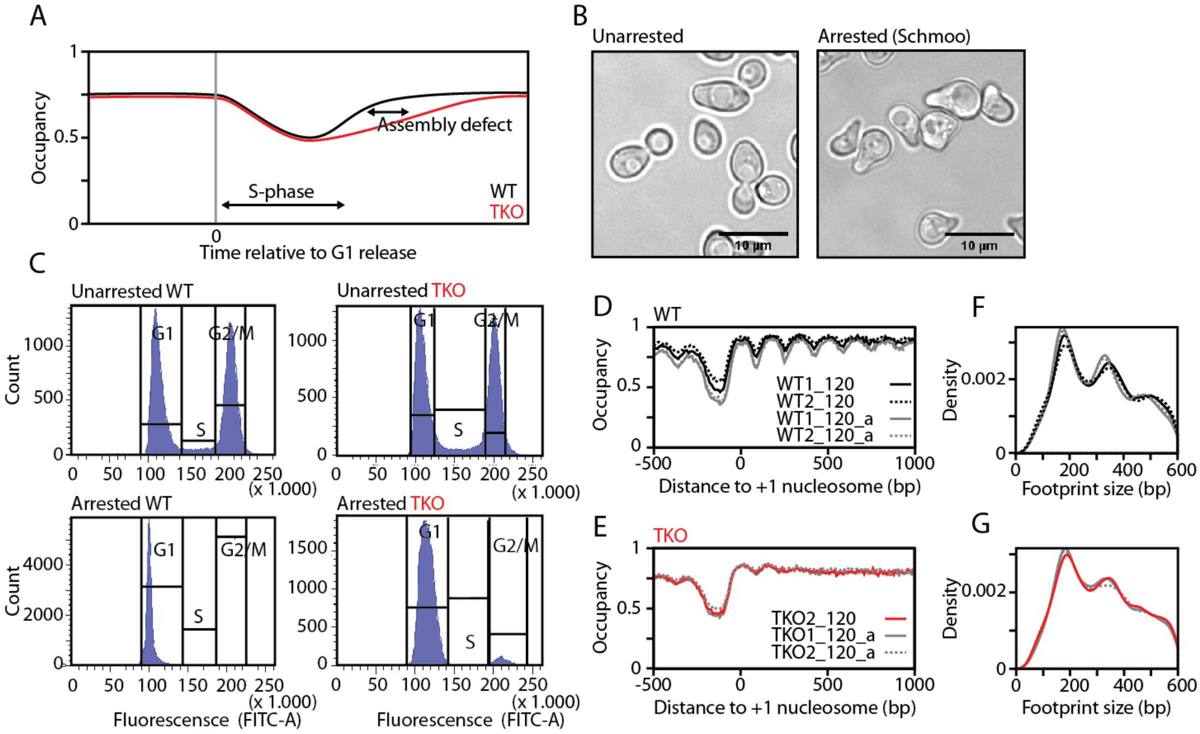


**Figure 2.2.13 Cell cycle arrest of yeast strains**

***(A)*** *Cartoon depicting the global occupancy change over time. During the cell cycle arrest in G1-phase, WT and TKO cells have an approximate global occupancy of 70-75%. Cells are released at timepoint 0 and re-enter the cell cycle where they go into S-phase. During this period, replication may reduce global occupancy at equal rate between WT and TKO. However, the re-assembly of nucleosome could be affected in the TKO strains missing important remodeling enzymes implicated in the assembly of nucleosomes. By taking multiple timepoints after the release of cells into S-phase, we can determine the global occupancy differences.* ***(B)*** *Typical microscopy capture of unarrested cells and arrested cells. During the arrest in G1-phase, all cells exhibit a typical 'schmoo' phenotype where a small protrusion is observed. A more heterogeneous phenotype is observed in unarrested cells. Black bars indicate 10 µm. Images acquired by Petra Vizjak and Jessica Furtmeier.* ***(C)*** *FACS data acquired from an unarrested and arrested population of cells of both WT and TKO cells. When unarrested, cells are observed at all cell cycles phases (G1, S and G2/M). After cell cycle arrest at G1-phase, a dominant population of cells in G1 phase (86-91%) is observed based on FACS analyses. Cell cycle is determined by total fluorescence of DNA. Cells in G2/M phase have a higher FITC-A fluorescent intensity due to the amplification of DNA after replication. FACS experiment and quantification performed by LMU BMC flow cytometry core facility.* ***(D-E)*** *Comparison of composite plots for arrested (grey) and unarrested WT (black) and TKO (red) samples. Biological replicates are shown using dashed lines. For global occupancy values of individual samples see Appendix Supplementary table 1, NPD13.* ***(F-G)*** *Nucleosome footprint size of arrested and unarrested samples identical to* ***(D-E)****.*

as this allows an extended time for the remaining (i.e. non-deleted) remodelers to organize arrays before replication. Alternatively, constant replication in unarrested samples could negatively impact array regularity. However, TKO samples exhibit highly similar array patterns across all samples, suggesting that an arrest does not improve regularity or that replication affects the global array regularity in unarrested samples (Figure 2.2.13E).

In addition to the genome wide composite plots, we compared the footprint size and distribution of arrested and unarrested samples for both WT and TKO as this could further suggest a change in regularity. We observe a shift of 6-7 bp and a broadening of mono- and dinucleosomal footprints for unarrested WT strains, likely caused by the undermethylation of these samples (Figure 2.2.13F). Mononucleosomal footprints of arrested samples are of similar size compared to previously analyzed samples (161 bp and 164 bp for WT1_120_a and WT2_120_a, respectively). This broadening of the distribution could be indicative of global reduction in array regularity, but is most likely due to the previously mentioned undermethylation of these samples. For TKO strains, we do not detect a similar shift as seen in WT samples (Figure 2.2.13G), suggesting that arrays did not improve upon cell cycle arrest. We do note a slight broadening of the peak for the TKO2_120_a sample, which has again likely to do with the elevated undermethylation in this sample compared to the other TKO samples and can also be observed in the composite signal.

Based on the current analyses we propose several alterations to future experiments. Cell cycle arrest of WT and TKO may be performed in a strain that has the native *BAR1* gene. This alternative strategy requires more alpha factor mating pheromone to be added, but better resembles strain conditions as used before. Moreover, these strains can be used to perform a ‘release’ experiment. After the arrest in G1, cells are washed, and resuspended in fresh growing media to remove alpha factor. This ensures that cells can return to normal growth rates, thus being ‘released’ into the cell cycle. Then, at short intervals after release, aliquots of cells can be collected and fixed. This may provide an insight in the occupancy and nucleosome assembly changes throughout the S-phase when WT and TKO cells are compared (Figure 2.2.13A).

### 2.2.15. Discussion (related to this chapter)

The chromatin landscape of *S. cerevisiae* is defined by many factors that tightly control nucleosome positioning. One key feature is the high regularity of genic arrays. These arrays are largely established by nucleosome remodelers, but may be influenced by other factors such as transcription and other general regulatory factors. Despite the plethora of factors that aim to control the chromatin landscape, heterogeneity is inevitable. Certain factors may work in conjunction, whilst others have antagonistic properties. Disentangling the role of these factors often relied on techniques that provide an average view of the cell population. Moreover, these techniques did not allow the visualization of the chromatin landscape at the single-cell level and contextual information between nucleosomes on the same DNA molecule is also lost. These limitations (which are well-known and understood) are due to the utilization of enzymes such as micrococcal nucleases (MNases) and amplification by PCR. MNases digest preferentially non-nucleosomal DNA, resulting in a loss of context, whereas PCR amplification results in a loss to probe cell-to-cell variations. Techniques that utilize such approaches have been key understanding the chromatin landscape and how certain factors shape it. However, the true chromatin landscape, its heterogeneity and cell-to-cell variations remain elusive.

Here, we report the utilization of MEFSIMO-seq, which aims to define the chromatin landscape – and thereby its heterogeneity – at single-molecule resolution. This approach probes the position of nucleosomes by means of a methyltransferase enzyme (MTase). Instead of digestion of non-nucleosomal DNA with MNases, the MTase methylates CpG sites of these regions. Therefore, there is no digestion of the chromatin, and contextual information remains intact. Furthermore, MEFSIMO-seq utilizes the Oxford Nanopore Technologies (ONT) platform to sequence (long) DNA fragments and directly detect methylated CpG sites. This approach does not necessitate amplification by PCR, cell-to-cell information is therefore retained as each strand of DNA originates from a different cell. We find that MEFSIMO-seq faithfully recapitulates the chromatin landscape as defined by others, whilst providing additional, previously unobtainable, information. This combination has allowed us to probe the chromatin landscape in new ways and has led to a better understanding of how it is shaped.

#### 2.2.15.1. Transcriptional dependence of rDNA loci

We capitalized on the ability of MEFSIMO-seq to directly sequence reads up to 20 kb in length. This allowed us to probe the accessibility of both rDNA loci simultaneously, despite their large size (Figure 2.2.3A). Based on previous findings, we correlate high accessibility of a single rDNA locus to transcriptional accessibility. We find that rDNA loci in both WT and TKO cells are not transcriptionally independent as previously suggested (Figure 2.2.3D, Figure 2.2.7I)(Dammann et al., 1995; French et al., 2008). Future experiments employing the deletion of transcriptional regulator components such as those from UAF-complex, may provide further support of these new findings. For example, deletion of *UAF30* prevents efficient docking of the UAF-complex, thereby potentially affecting improper transcriptional regulation by Pol I and the transcription machinery, in turn leading to a random or transcriptionally independent regulation.

Would transcriptional dependency be favored over independency? One model could posit that the transcriptional activity creates a positive feedback, thereby stimulating neighboring rDNA loci. This feedback allows quick generation of large amounts of rDNA. During cell cycle stages in which rDNA is required in large amounts, this model would aide in the critical supply. An alternative model could posit that certain fractions of the approximately 200 rDNA repeats are simultaneously highly active or inactive. Previous studies have shown that the rDNA is subdivided into ‘subdomains’ (Dauban et al., 2019). Potentially, certain subdomains are mostly active and thereby accessible. Moreover, geometry within the subdomains changes throughout the cell-cycle. Perhaps certain geometrical conformations stimulate transcription and/or positive feedback.

MEFSIMO-seq could theoretically disentangle transcriptional dependence controlled by subdomains and/or geometry. However, the limitations are both physical and technical. Despite the capabilities of sequencing ultra-long reads by the ONT platform, there is a physical limit of DNA length during experimental handling. During the methylation procedure and library preparation, gDNA will inevitably fragment. This prevents the sequencing of entire chromosomes which would allow probing of all rDNA repeats in context. Additionally, we are limited technically in probing many rDNA reads simultaneously. The currently utilized reference genome SacCer3, has two rDNA loci annotated at chromosome 12. After alignment of MEFSIMO-seq data, we can therefore only determine the accessibility of two loci at once. Manually extending the SacCer3 reference genome with more rDNA loci is possible, but will result in physical limitations mentioned before. Key improvements to further elucidate the rDNA architecture therefore lay in improved approaches that limit fragmentation.

#### 2.2.15.2. The wild-type chromatin landscape is more regular than previously observed

Despite best efforts by techniques such as MNase-seq, differences in phasing within a gene will lead to an average signal that suggests irregularity. Utilizing MEFSIMO-seq data, we can computationally align individual reads, excluding phasing related heterogeneity. We find that nearly all nucleosomal arrays in wild-type cells have a relatively high regularity (Figure 2.2.4G).

We noticed that regularity improved significantly for N+1 shifted reads after computational phasing, but never reached regularity similar to the global pattern. What may cause this slight reduction in regularity for these particular reads? One explanation may be of technical nature. When computationally aligning individual reads, we align the nucleosome dyad found closest to the predicted +1 nucleosome. When the actual +1 nucleosome is shifted downstream of the TSS, we have a high likelihood of aligning the read to the correct nucleosome. Due to the NFR region upstream of the +1 nucleosome, it is unlikely we find a nucleosome dyad closer to the predicted +1 nucleosome position. However, when the actual +1 nucleosome is shifted upstream of the TSS, one may find the closest dyad to be the +2 nucleosome. Reads are then computationally aligned to the ‘incorrect’ nucleosome. The regularity of the array may therefore be affected. However, this explanation is largely made obsolete by the translational symmetry found in Figure 2.2.9A. Here we find that array regularity is identical starting from any genic nucleosome.

An alternative explanation could be cell-cycle related. Potentially, the relatively small number of reads categorized as N+1 shifted, may originate from cells in e.g. S-phase. During DNA replication, nucleosomes are evicted and quickly re-assembled after which they are carefully repositioned by nucleosome remodelers (Alabert and Groth, 2012). The subpopulation of reads in which the N+1 is shifted, may be not fully positioned yet after replication. In order to validate these claims, one could perform a cell cycle arrest experiment in which cells are arrested in e.g. $G_1$-phase. Globally, no differences are found in array regularity. However, categorization of reads based on N+1 state, could result in a reduced ratio of reads with a shifted N+1. Moreover, reads that fall in this category, could show similar regularity to global patterns as they are unlikely to be affected by replication.

#### 2.2.15.3. Transcription alters array regularity, but not nucleosome composition

We find that highly transcribed genes tend to have worse array regularity. These observations are in agreement with previous findings in our lab suggesting RNA Pol II transcription negatively impacts array regularity (Figure 2.2.5B). This observation however, is dependent on the transcriptional activity attributed to each gene. We find that 4tU-seq data generated by Xu et al., does not support, nor is well correlated to data supporting a negative role for RNA Pol II in array regularity. Moreover, we observe no clear enlarged NFR, which is typically associated with highly transcribed genes. We argue that this data may not accurately represent genic transcription rates and should be used with caution. Future analyses could utilize the transcriptional rate information for genes and perform a computational phasing on highly transcribed genes which have a shifted +1 nucleosome (Figure 2.2.4B). If indeed RNA Pol II destroys arrays, such reads should not be restored to the same degree as observed globally (Figure 2.2.4I).

MEFSIMO-seq data does not detect a smaller nucleosome footprint at highly transcribed genes, despite findings in support of nucleosome unwrapping due to transcription (Figure 2.2.5D) (Kulaeva et al., 2010; Ramachandran et al., 2017). What may prevent MEFSIMO-seq from detecting smaller nucleosome footprints? We have optimized conditions in which the MTase methylates all accessible (i.e. non-nucleosomal) DNA. Partially unwrapped DNA may still protect against MTases from entering and methylating sites in this portion of nucleosomal DNA. Alternatively, DNA sequence bias may favor CpG-poor DNA at nucleosomal entry- and exit-sites, thereby making it unable to detect subnucleosomes accurately. Increasing resolution by utilizing different/multiple MTases, may aide in better detection of subnucleosomes or unwrapped DNA

We moreover find that highly transcribed genes do not correlate with an enlarged nucleosome footprint at the +1 nucleosome (Figure 2.2.6C). In contrast, constitutive and lowly expressed genes with a UNB/TFO themed promoter tend to have a population of enlarged +1 footprints. What may cause this enlarged footprint at these promoters? The largest portion of genes are associated to the unbound (‘UNB’) theme and represent 2474 genes. These promoters lack evidence of any binding aside from the pre-initiation complex (PIC) or a nucleosome. Our approach mainly detects nucleosomes which – when of mono-nucleosomal size – are approximately 158 bp by our measurements. The PIC is a multi-subunit 1.5 MDa complex but surprisingly makes up only a relatively small footprint of approximately 50bp (Krebs et al., 2017). Added to the size of a nucleosome footprint, it may be in agreement with the ~200 bp footprint found in UNB/TFO themed promoters.

#### 2.2.15.4. Depletion of INO80 leads to almost completely abrogated array regularity

Depletion of INO80 in a strain that lacks ISW/CHD1 (referred to as TKO) shows a further reduction in array regularity. Computational phasing only improves regularity by a negligible amount (Figure 2.2.8G). We therefore conclude, with supporting MNase-seq data from our lab, that INO80 is a bona-fide spacing remodeler.

When performing a limited MNase digest, the TKO-INO80 strain shows mono-/di- and tri-nucleosomal peaks, suggesting that clusters of equally spaced nucleosomes still present, despite lacking four remodelers. The lack of any array regularity from composite plots of MEFSIMO-seq data would suggest otherwise. Does the lack of a +1 nucleosome peak in the TKO-INO80 strain reflect the true biological average or is it a simple artifact? One argument explaining the significant differences in composite plots between MNase-seq and MEFSIMO-seq could be the way data is displayed. MNase-seq composite plots are based on the dyad density, whereas MEFSIMO-seq shows the occupancy (Chereji and Clark, 2018). Plotting both approaches as occupancy would potentially account for some degree in the discrepancy. However, it is unlikely that both datasets will match completely after replotting the data and this does not fully explain the missing +1 nucleosome in MEFSIMO-seq data.

MEFSIMO-seq, as well as other techniques such as ODM-seq, calculate the true occupancy (Oberbeckmann et al., 2019). In contrast, MNase-seq and similar techniques map the (average) nucleosome position and compare relative occupancies. Additionally, MNase-seq may overemphasize the +1 nucleosomal peak due to MNase digestion bias. The +1 nucleosome is located near the NFR, which is hypersensitive to MNase digestion. Excising a mono-nucleosome originating from the +1 position is therefore more likely than from e.g. the gene body. The result would be an overrepresentation of mono-nucleosomes which is translated in a high(er) peak in composite plots. However, findings in recent studies have shown that the MNase bias does not affect peak height and positions when compared to a technique similar to MEFSIMO-seq (Oberbeckmann et al., 2019). These comparisons were performed in wild-type cells however, and not in mutant strains. Differences between the ‘true’ (i.e. MEFSIMO-seq) and overemphasized (i.e. MNase-seq) +1 peak could be enlarged with each additional depletion of a nucleosome remodeler. Further arguing an overemphasis of the +1 nucleosomal peak is the composite signal of wild-type MNase-seq versus MEFSIMO-seq. Genome wide composite signals of MNase-seq have a prominent +1 nucleosomal peak, downstream of which the peak heights -and therefore the relative occupancy- declines (for an example see Singh et al., 2021, Fig. 1A). In contrast, MEFSIMO-seq data shows an equal occupancy for all nucleosomes (Figure 2.2.1H). Arguing the technical point that suggests MEFSIMO-seq can measure the accurate +1 peak height in relation to other nucleosomes may be possible. To validate the results, one would require an extensive comparison with matching MNase-seq data, potentially with multiple MNase digestion levels as each digestion degree can influence the peak heights.

### 2.2.15.5. The nucleosome clamping mechanism is not supported by *in vivo* data

In favor over a density dependent spacing, the nucleosome clamping mechanism was suggested to ensure equal spacing between nucleosomes in vitro (Lieleg et al., 2015). By decreasing the global occupancy, we were able to challenge this mechanism *in vivo*. We find no evidence of clamping, suggesting an alternative mechanism by which nucleosome remodelers equally space nucleosomes (Figure 2.2.10H).

How are nucleosomes equally spaced *in vivo* when the clamping mechanism is not supported? An attractive alternative hypothesis is the density dependent mechanism. This mechanism posits that spacing between nucleosomes increases linearly and evenly with decreasing occupancy (Figure 2.2.10A). Formally, this mechanism remains a possibility as our inhomogeneity calculations cannot distinguish between density dependent spacing at high and low nucleosome occupancy. Under histone depleted conditions, genome wide composite plots are similar to a truly random organization of nucleosomes (Singh et al., 2021). However, the position of the +2 and +3 nucleosomes remains at roughly similar distances from the +1 as under wild-type conditions. This suggests there is some activity that favors distances shorter than predicted from a density dependent spacing mechanism.

We observe only mild differences between HD and TKO-HD strains, nowhere to the extent we see between WT and TKO (Figure 2.2.10C, E). The slight differences we observe between HD and TKO-HD could be due to the ‘ruler-mechanism’. This mechanism posits that the distance between two nucleosomes is set by interaction of remodelers and nucleosomes, specifically the HAND-SANT-SLIDE domain of ISW1a (Yamada et al., 2011). In contrast to our inhomogeneity score analyses performed on HD and TKO-HD samples (Figure 2.2.10H), we do not detect any differences between the samples. This minor discrepancy between the two approaches leaves the possibility that the clamping mechanism holds true *in vivo*. The inhomogeneity score analysis may not be a suitable approach as the accuracy would increase with elevated resolution (i.e. a cocktail or m6A methyltransferases). Future experiments implementing such approaches could directly utilize the inhomogeneity score analyses to further investigate potential clamping *in vivo.*

The difficulties in elucidating the mechanism by which remodelers enforce regular spacing may be resolved with increased resolution of techniques such as MEFSIMO-seq. Our current approach only probes CpG sites in the genome and thereby provides a limited resolution on the location of nucleosomes. Additionally, ambiguity in methylation calling further reduces resolution. Certain regions – low in CpG sites – are therefore difficult to probe for (small) differences upon nucleosome deletion under HD conditions. The advent of techniques that utilize multiple methyltransferases simultaneously may aide in elucidating the mechanism by which remodelers space nucleosomes (Lee et al., 2020; Stergachis et al., 2020).

#### 2.2.15.6. Reduced array regularity upon GRF binding

Our preliminary data suggests that binding of GRFs does not improve regularity of flanking arrays (Figure 2.2.11B-C). This finding is in contrast to observations that show a decrease in regularity upon GRF deletion (Kubik et al., 2015). What model could cause reduced array regularity upon GRF binding? GRFs are known to work in conjunction with nucleosome remodelers and even physically interact with the RSC remodeler, a member of the SWI/SNF subgroup (Gavin et al., 2002; Wu et al., 2018). However, it is currently unknown whether GRFs directly recruit RSC, if RSC binds DNA which is made accessible by GRFs, or if GRFs are even required (Lorch and Kornberg, 2015). One model would thus be conceivable in which GRFs are bound, but remodeling has not yet taken place. Lacking this activity, arrays flanking the binding site are not yet organized as a regular array. This observation would be consistent with a global decrease in array regularity upon GRF deletion, as no remodeler can be recruited/targeted to the region. Moreover, it is consistent with our data in which array regularity is worse when observing reads with small accessible regions (Figure 2.2.11C). Based on the available data, we propose a model in which GRFs bind and subsequently recruit/target remodelers which organize the nucleosomal arrays flanking the binding site.

#### 2.2.15.7. Mix-n-match approach to determine differences between yeast strains variants

*In silico* analysis utilizing a mix of two strain variants showed that it was technically feasible to utilize MEFSIMO-seq in order to accurately determine read reference origins. When applied *in vivo*, we found no difference in global occupancy between a WT and TKO strain, nor a difference between W303 and S288C in general. These preliminary findings can further be improved by applying several improvements and alterations. For example, mapping accuracy increases when read length increases. When a read does not contain a SNP, it can be mapped equally well to each reference genome. Such reads are deleted by applying a filter. Longer reads increase the chance of a SNP being present within the read and thus decrease ambiguous mapping. After performing the methylation of chromatin, isolation of long DNA can be optimized by DNA size selection using e.g. BluePippin. However, DNA length is further affected by the standard library preparation of ONT. This preparation contains several bead purification steps, which can shear and fragment the DNA. Alternative DNA sequencing kits from ONT, may circumvent this issue and are advertised to reliably generate data with a median read length of >50kb (kit: SQK-ULK001). However, such sequencing kits require additional changes during DNA isolation to optimize DNA length, and do not support multiplexing using barcode ligation. High throughput of multiple samples is therefore not recommended when utilizing such approaches.

As the S288C and W303 strains are highly similar based on our preliminary data, future questions using this mix-n-match approach could gain further insight into the chromatin landscape. Dissecting the effect of deletion or mutations of a single subunit from a multi-subunit complex is often difficult to discern. Such comparisons can be made by utilizing a sensitive approach described here as day to day, or even sample to sample variations are taken out of the equation. Care should be taken however that each combination is tested (e.g. Figure 2.2.12C-D), to assure that mutations do not affect the individual strains differently.

## 2.3. Chapter 3: Elucidating the spacing mechanism of the ISWI remodeler *in vitro*

In this final chapter of my thesis, I will discuss the preliminary results regarding an *in vitro* assay to study and dissect the role of nucleosome remodelers. The concepts and hypotheses outlined below remain partially unanswered due to a lack of time. In addition, parts of this chapter are adapted from a master thesis project of Jessica Furtmeier, whom I supervised, and have therefore not been fully developed. Here, I will describe the steps toward establishing this assay.

### 2.3.1. Background

#### Introduction

The best studied function of the ISWI remodeler is its ability to slide nucleosomes. *In vitro* assays have definitively shown that a single nucleosome, located at the edge of a short stretch of DNA, will get slid towards the center after incubation with ISWI and ATP (Clapier and Cairns, 2012; Stockdale et al., 2006). These observations further suggested an essential function for this remodeler: nucleosome spacing. This process imposes a regular distance between neighboring nucleosomes. Indeed, further *in vitro* studies utilizing salt-gradient dialyzed (SGD) chromatin have shown that ISWI can generate regular arrays (Krietenstein et al., 2016; Vary et al., 2003). Additional structural analysis of the ISWI remodeler suggested that specifically the HSS domain of ISWI determines the distance between two nucleosomes, acting as a ‘protein ruler’ (Yamada et al., 2011). General sliding activity is not dependent on this HSS domain, nor on the N-terminal region (NTR). Deletion of the HSS domain or non-conserved residues of the NTR was shown to support nucleosome sliding, albeit slower compared to the full-length protein (Mueller-Planitz et al., 2013).

The role of the ISWI remodeler and its domains is well documented. Many of the *in vitro* assays dissecting the spacing mechanism of this remodeler rely on mono-nucleosomal substrates. Such approaches cannot test spacing per definition. Alternative approaches using SGD chromatin rely on MNase to determine average nucleosome positions, thus losing the ability to probe individual chromatin fibers. The use of chromatinized arrays has the benefit of creating a controlled environment to study remodelers (Logie and Peterson, 1997; Ludwigsen et al., 2013; Mueller-Planitz et al., 2013). However, such approaches can only determine the position of a single nucleosome and do not inform on individual fibers either. Elucidating the spacing mechanism of remodelers *in vivo* is complicated due to the high redundancy between remodelers (Kubik et al., 2019). However, recent advances have shown methods to drastically reduce this redundancy (Singh et al., 2021).

Despite the availability of several different approaches to study nucleosome remodelers, several caveats remain. Here, we take steps to create an *in vitro* chromatinized array, to study how ISWI slides nucleosomes and shapes the chromatin landscape. We combine this assay with MEFSIMO-seq to provide direct information on the position of individual nucleosomes in the context of neighboring nucleosomes, on a single DNA fiber.

### 2.3.2. Optimizing spatial resolution of *in vitro* arrays for optimal nucleosome detection by MEFSIMO-seq

In order to improve the spatial resolution of an *in vitro* array we started off using a previously published nucleosomal array consisting of 25, equally spaced, Widom 601 sequences (Mueller-Planitz

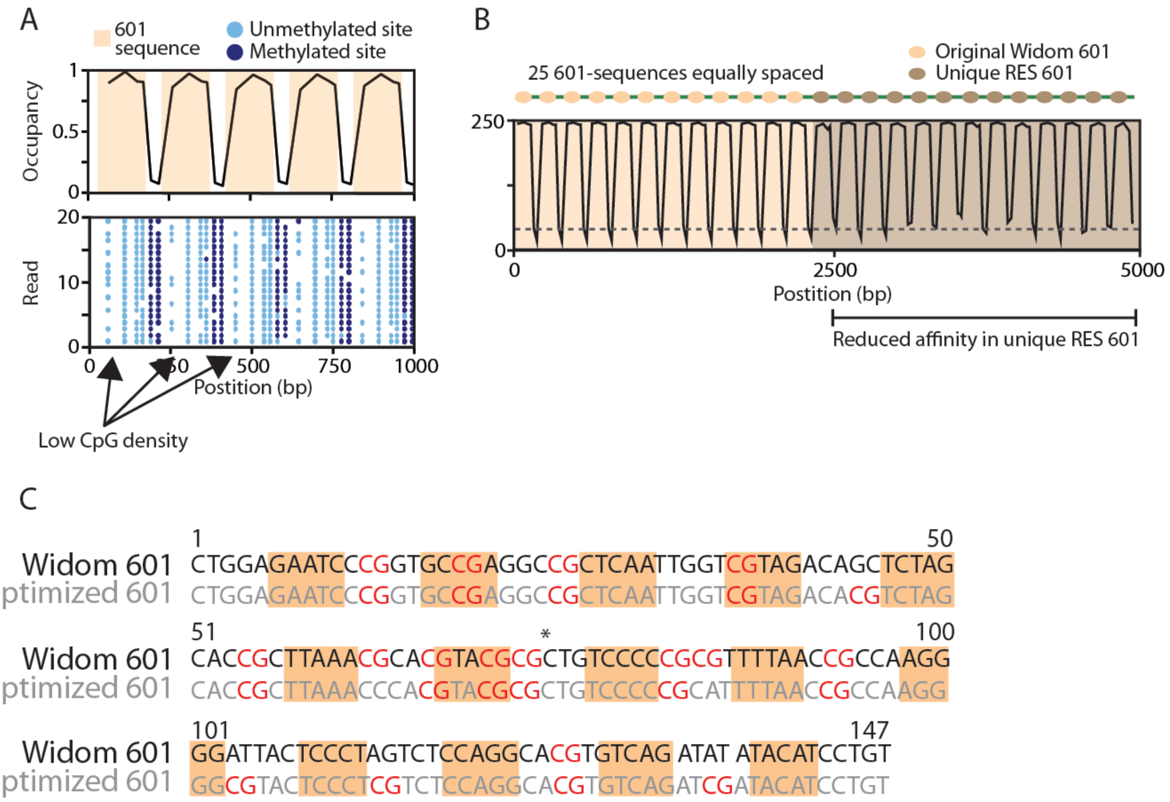


**Figure 2.3.1 Optimization of Widom 601 to increase spatial CpG resolution**
***(A)*** *CpG distribution in the reference 601 sequence is sub-optimal due to CpG density and distribution. This results in regions (here 5' of sequence) to have a low resolution (indicated by arrows for first three sequences). Figure taken from Chapter 2 Figure 2.2.2.* ***(B)*** *Affinity may be reduced by the slight alterations compared to the original Widom 601 sequence in the reference array. This results in less-defined nucleosomal occupancy as indicated by the lower peak-trough ratio in the latter half of the array. Difference is highlighted by dashed line. Figure taken from Chapter 2 Figure 2.2.2.* ***(C)*** *Comparison of reference Widom 601 sequence and optimized 601 sequence. CG sites are highlighted in red. Minor grooves facing inward marked in orange. Nucleosomal dyad is marked by asterisk. Relative nucleotide positions are indicated on top of sequence.*

et al., 2013). Briefly, this Widom 601 sequence has a high affinity for histone octamers during salt gradient dialysis. This ensures stably bound nucleosomes at well-defined locations. Earlier, we have shown that MEFSIMO-seq faithfully recapitulates the predicted organization of nucleosomes and linkers on this array (Figure 2.2.2A-B). However, throughout the array, low CpG density prevented the exact determination of a nucleosome edge, especially within the first half of each 601 sequence (Figure 2.3.1A). Moreover, despite their 50 bp length, linker regions essentially only had two points of information. To more accurately define the location of nucleosomes (and thereby linkers) we decided to increase CpG density within the Widom 601 sequence and linkers. The increased CpG density would result in a higher spatial resolution to map nucleosomes as more points of information (i.e. CpG sites) would be available throughout the array.

One side effect is to lose the high affinity for histone octamers of the original 601 sequence when making modifications. Our previous findings suggested that already minor modifications (alteration of a 6 bp long native restriction site) could affect the affinity and/or positioning markedly (Figure 2.3.1B). Studies characterizing the 601 sequence determined that nucleotides at the minor groove facing inwards required flexible TA dinucleotides as DNA distortion at these regions is most challenging (Chua et al., 2012). Indeed, addition of TA dinucleotides at minor grooves facing inwards improved nucleosome stability compared to the original 601 sequence. An additional caveat was that CpG dinucleotides should not be in close proximity to one another. The computational model that we use to detect CpG methylation cannot distinguish k-mers which have multiple (un)methylated CpGs (Simpson et al., 2017). In practice, this results in the grouping of CpG sits that are within 10 bp of each other and a loss of spatial resolution. Based on these limitations, we optimized the original 601 sequence and linkers to have an increased CpG density. We took care avoiding critical regions such as minor grooves facing inwards and having too many CpGs in close proximity (Figure 2.3.1C).

### 2.3.3. Structural reorganization of arrays to improve *in vitro* studies on remodeler sliding and spacing

Next, we decided to alter the general organization of the array. In previous studies using this array, the effect a remodeler had on the position of a nucleosome was determined by accessibility of a unique restriction site within the 601 sequence (Lieleg et al., 2015; Ludwigsen et al., 2013; Mueller-Planitz et al., 2013). Thus, only a single nucleosome was probed for movement. Here, we intended to utilize the array to determine the position of all nucleosomes, and how remodelers influence the spacing and position. Based on *in vivo* and *in vitro* observations, we hypothesized that regular arrays could be generated upon addition of a remodeler such as ISWI (Ito et al., 1997; Krietenstein et al., 2016; Längst et al., 1999; Ocampo et al., 2016; Singh et al., 2021). To observe the generation of a regular array, we argued that this effect could be best determined when the starting organization is irregular. The reference array had equally spaced 601 sequences, linked by 50 bp of DNA. Here, the starting organization would be regular as well as the predicted outcome. Additionally, we wanted to utilize the array to examine the density independent ‘clamping’ mechanism of the ISWI remodeler (Lieleg et al., 2015). Based on the suggested model, we argued that we could directly test it by utilizing an array with large linkers and comparing it to the reference array. Taken together, we utilized the optimized 601 sequence and implemented them in two types of arrays: (1) an array with significant increase in linker length to 99 bp (referred to as 99bp) between ten, 601 sequences and (2) an array with variable linker lengths (referred to as VLL) between twelve, 601 sequences (Figure 2.3.2A). Optimized arrays were synthesized by Genscript and provided as plasmids cloned into pUC18 (See Appendix for plasmid maps, plasmids stored as pFMP541 and pFMP542).

These new arrays were designed in such a fashion that they could be used as standalone arrays, fused together by molecular cloning, or fused with one half of the reference array (referred to as URS for the unique restriction sites within this array). I, together with master student Jessica Furtmeier, performed standard restriction enzyme digestion cloning to create two new arrays. First, we took the backbone of the reference 601 array and combined the 10-mer 99bp array with the 12-mer VLL (Figure 2.3.2B, Top). Second, we excised the non-URS part of the reference array and inserted the 99bp array (Figure 2.3.2B, Bottom). A third combination is possible by inserting the VLL array into the excised reference array, due to time limitations, this has not been done yet. These combinations result in arrays of 4.8 kb and 4.9 kb for the 99bp+URS and 99bp+VLL combination respectively (See Appendix for plasmid maps, new plasmids stored as pFMP546 and pFMP547). Each of the full length arrays can be excised from the plasmid by digestion with XbaI and EcoRI. All 601 sequences within the array contain a BsiWI restriction site which can be used during the validation of SGD assembly (Figure 2.3.2C). Correct array lengths were initially validated by digestion with EcorI and XbaI (Figure 2.3.2D).

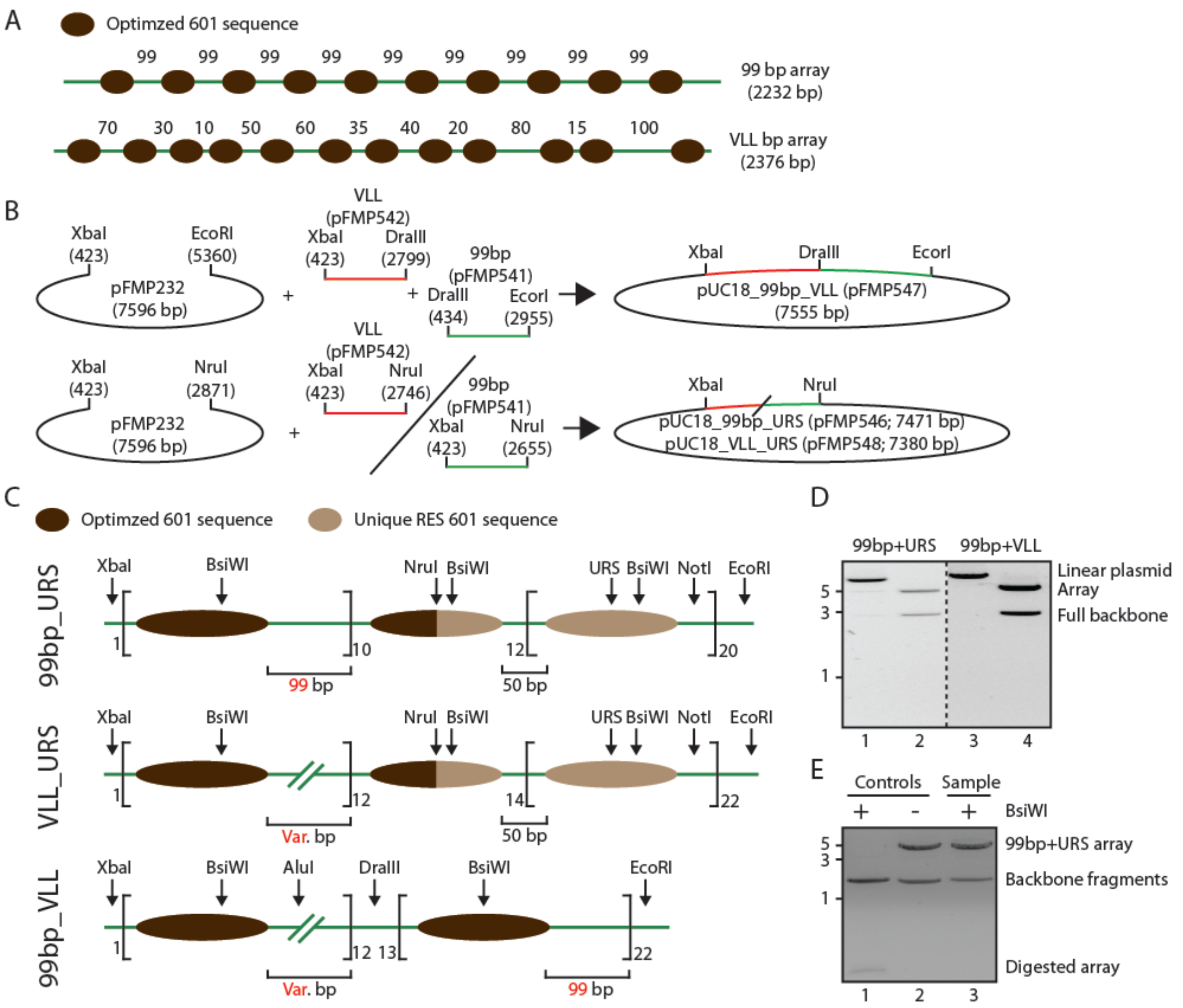


**Figure 2.3.2 Cloning strategy and SGD assembly of optimized arrays**

*__(A)__ Cartoon outline of the 99 bp array and the VLL bp array. Nucleosomes are represented by dark-blue ovals. Linker lengths are indicated between the nucleosomes. __(B)__ Restriction enzyme cloning strategy for different combinations of the array. Top: A combination of the 99 bp array and the VLL bp array. The pFMP232 is utilized as a backbone in both newly synthesized and optimized inserts. This results in a plasmid containing both referred to as 99bp_VLL. Bottom: the pFMP232 plasmid containing the twelve URS 601 sequences is used as a backbone for the insertion of either the VLL bp or 99 bp array. This results in a plasmid containing the 99bp+URS or VLL+URS arrays. Utilized restriction enzymes are annotated together with their site between brackets. Plasmid names are indicated as pFMPXXX. Total plasmid sizes are indicated in bp. __(C)__ Schematic depiction of the three different arrays which can be generated using the strategy in __(B)__. Nucleosomes consist of 147bp. Linker lengths are highlighted in red and can be variable as illustrated in __(A)__. Values flanking the square brackets indicate the number of nucleosomes within. Commonly used restriction enzymes are indicated and point toward the approximate restriction location. URS refers to unique restriction site and varies for each 601 sequence. For details on exact restriction sites see (Ludwigsen et al., 2018) or plasmid map in Appendix. __(D)__ 99bp+URS and VLL+URS plasmids were checked for correct insertion lengths after cloning. Lane 1 and 3 show XbaI linearized plasmid to check plasmid length. Lane 2 and 3 show XbaI+EcoRI digested plasmid. This results the excision of the array (approx. 5.0kb) and the pUC18 backbone (approx. 2.5kb) __(E)__ BsiWI digestion of unassembled (control) or assembled 99bp+URS array (sample) to validate SGD assembly. Unassembled arrays were used as controls in lane 1-2 to indicate the digestion of the array when not no nucleosomes are present. Backbone fragments were digested with BsaI prior to BsiWI digestion. Assembled array in lane 3 were subjected to similar BsiWI digestion. The SGD assembled DNA contains some traces of (partially assembled) backbone which is considered a contamination, but does not impact downstream analyses or experiments. Experiment performed by Jessica Furtmeier, figure adapted by me*

Next, we assembled chromatin by standard salt gradient dialysis (SGD) as previously described (Ludwigsen et al., 2018, see Methods). We validated that the 99bp+URS array is fully assembled at each 601 site by utilizing a restriction enzyme present within the 601 sequence. The BsiWI enzyme is only able to cut when no nucleosome is present at the 601 site. The unassembled control array incubated with BsiWI enzyme results in the full digestion of the 99bp+URS array. No digestion was observed after SGD assembly, suggesting that nucleosomes were assembled efficiently on the array and no sites were generally un-assembled (Figure 2.3.2E). We do note the presence of a BsaI digested backbone in the assembled samples. This is likely partially assembled DNA and is typically regarded as undesired and can be eliminated by performing the assembly with a lower histone:DNA ratio. As we will not perform sensitive kinetic assays, we continue with these assembled arrays as they will not interfere with preliminary downstream analyses.

### 2.3.4. Characterization of optimized *in vitro* arrays

Next, we validated the ability of these arrays to be remodeled by the ISWI remodeler. Due to time and priority constraints, we continued only with the 99bp+URS array. We performed a standardized remodeling assay using the KpnI restriction enzyme (Figure 2.3.2A) (Ludwigsen et al., 2018). The KpnI restriction enzyme site is located on the middle 601 sequence of the URS region of the array. Due to an error in the design, no unique restriction site is present in the 99bp part of the array. Future iterations of the array may be redesigned to include such a site. To initiate remodeling, we incubated the array with ISWI and ATP for up to 60 minutes at 26˚C (see Methods for details). We observe that the ISWI remodeler acts rapidly after the addition of ATP, sliding minimally the one nucleosome, which in an unremodeled state blocks the KpnI restriction site (Figure 2.3.2B). We note that during the remodeling, we observe additional DNA fragments, potentially due to star activity of the restriction enzyme or mutations in the DNA (see Discussion related to this chapter). These fragments are not observed when the same unassembled DNA is digested.

Based on the quality controls we performed earlier (Figure 2.3.2D-E) we were confident that these assays are fully chromatinized array. To test how well the modified 601 sequence positions nucleosomes, we next added 99bp+URS (excised from pFMP546, see Appendix Plasmid maps) unremodeled arrays as a spike-in during a MEFSIMO-seq experiment. Single-molecule sequencing of this array allowed us to simultaneously validate the increased spatial CpG resolution (99bp part) and have a direct comparison to the reference array (URS part). The addition of the array as a spike-in to *in vivo* samples allows us to simultaneously sequence the *in vivo* data, whilst retaining the ability to acquire a high coverage of *in vitro* arrays. Our initial findings validate that the nucleosomes are able to assemble on all 601 sequences (Figure 2.3.2C, top). We observe nearly 100% occupancy at the center of the optimized 601 sequence, suggesting that all analyzed reads (>50.000) have a nucleosome present. We could not confirm the 99bp+URS sequence utilizing traditional Sanger sequencing due to the strong repetitiveness of the underlying sequence. We therefor resorted to utilize the long-read nanopore sequencing approach to confirm the sequence identity. We validated that the sequence of individual reads, in addition to the consensus sequence over all reads is similar with respect to the designed reference sequence. Individual reads >4.5 kb, as well as the consensus sequence, have >90% homology to the reference sequence based on NCBI basic local alignment search tool (BLAST). Using this procedure we found no secondary mutations, suggesting that the plasmid was successfully cloned. When performing an analysis where we predict the position of each nucleosome, we find that the

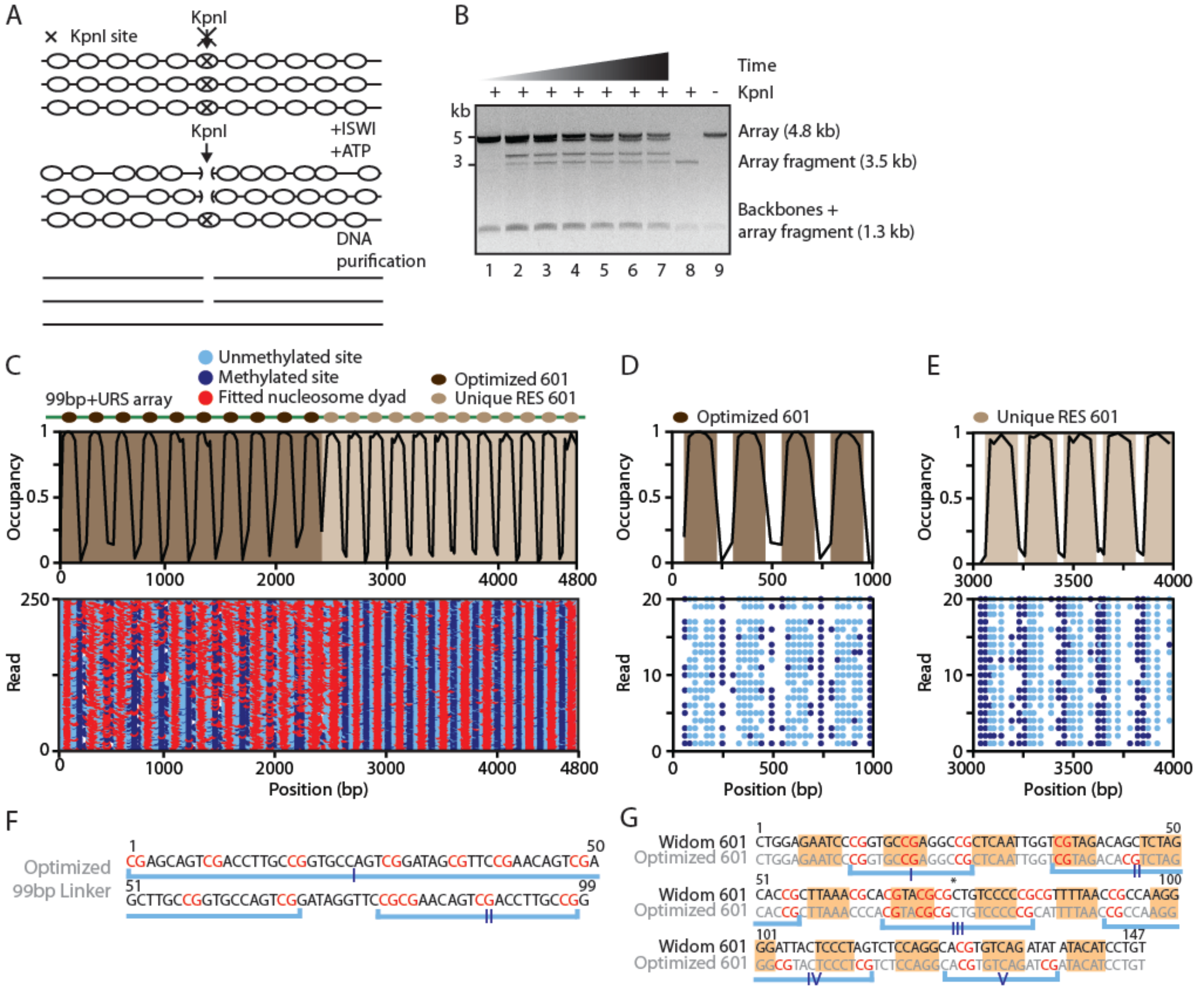


**Figure 2.3.3 Characterization of optimized 99bp+URS array**

***(A)*** *Schematic overview of the remodeling assay using KpnI. Before remodeling, nucleosome occludes the KpnI restriction site. When ISWI and ATP are added, nucleosome sliding results in the ability for KpnI to access the restriction site. DNA is purified and subsequently analyzed by agarose gel. Figure adapted from Mueller-Planitz et al., 2013.* ***(B)*** *Remodeling assay performed using the 99bp+URS array. Upon the addition of ISWI and ATP (lane 1-7) remodeling occurs rapidly as indicated by the ability of KpnI to cut the array (lane 2-7). KpnI digestion of the array results in two fragments of 3.5 and 1.3kb each. In this experiment, pUC18 backbone is further digested in two equal size fragments using BsaI resulting in two fragments of ~1.3kb each. A positive control of naked array DNA is cut with KpnI to show the expected fragment lengths (lane 8). A negative, undigested, naked DNA control is shown in lane 9. Samples are taken at timepoints between 0 and 60 minutes.* ***(C)*** *MEFSIMO-seq composite plot and 250 random single-molecule reads of 99bp+URS array (top and bottom, resp.). Optimized and unique RES 601 sequences are indicated by dark- and light-brown ovals and shaded area in the composite plots. Fitted nucleosome dyads (red circles) are less well defined in the optimized region, likely due to lower efficiency of the 601 sequence or low resolution within the linkers. Unmethylated and methylated sites are indicated as light- and dark-blue, respectively.* ***(D-E)*** *MEFSIMO-seq composite plot and 20 random single-molecule reads for short regions of the 99bp+URS array. 601 regions are indicated by dark- and light-brown shaded areas. Light- and dark-blue circles indicate unmethylated and methylated CpG sites, respectively.* ***(F)*** *Theoretical nucleotide sequence of 99bp linker. Groups of CpG sites after computational analysis are indicated by light-blue brackets. Groups are indicated by dark-blue roman numerals. CpG sites are highlighted in red.* ***(G)*** *Reference Widom 601 nucleotide sequence compared to optimized 601 sequence. Light-blue brackets indicate groups of CpG sites after computational analysis. Individual groups are indicated by dark-blue roman numerals. CpG sites are highlighted in red. Minor grooves facing inward marked in orange. Nucleosomal dyad is marked by asterisk. Relative nucleotide positions are indicated on top of sequence.*

nucleosome dyad position is not well-defined at a fixed location in the 99bp part of the array (Figure 2.3.2C, bottom). Prediction of the position for each nucleosome is more difficult compared to the latter half of the array. This reduced localization precision could be due to a low(er) nucleosome affinity of our modified 601 sequence or a decrease in uniquely identifiable CpG sites within the linker.

To further elucidate whether the increased CpG density within the linker actually increases overall spatial resolution, we compared either half of the array. In contrast to our prediction, we find that the CpG resolution within the linker of the 99bp part of the array is reduced when compared to the latter half (Figure 2.3.2D-E). The linkers in the 99bp part of the array only contain maximally two points of information (Figure 2.3.2D, dark-blue dots). Comparatively, the URS part of the array has typically two to three points of information in half the distance (Figure 2.3.2E, dark-blue dots). Upon closer inspection of single-molecule data, we indeed find that the increased CpG sites within the linker actually decreased the resolution. This is due to the first three quarters of the 99bp linker only having a single point of information. This single point of information is due to grouped methylation status of nine CpG sites (Figure 2.3.2F, group I). The latter quarter of the linker contains the second point of information, and consists of four CpG sites (Figure 2.3.2F, group II). The high degree of CpG sites per group is due to a design error. The Nanopolish software used to detect methylated CpG dinucleotides groups them when they are within ten basepairs of one another (Simpson et al., 2017). Our interpretation of this limitation was that the cytosine of a CpG dinucleotide should be ten base pairs away from the cytosine of the neighboring CpG site. In our attempt to optimize CpG density, this interpretation deemed detrimental to the resolution as we left ten bases between cytosines and not CpG sites. Interestingly though, based on these *in vitro* samples, it appears that a distance of nine bases is also sufficient to prevent grouping of sites (Figure 2.3.2F, transition from group I to II). Moreover, the reduced accuracy in predicted positions of nucleosomes in the 99bp half of the array (Figure 2.3.2C, bottom) is likely due to the lower resolution in the linker and not due to the reduced affinity of the modified 601 sequence.

Next, we took the single-molecule information and identified whether the alterations to the optimized 601 sequence improved the resolution. In contrast to the linker, we do observe an increase of uniquely detectable sites within the 601 sequence. We find that the optimized 601 sequence has five information points distributed equally over the 601 sequence (Figure 2.3.2D, light-blue dots). In comparison, the original 601 sequence has only two to three information points, mostly located in the 5’ half (Figure 2.3.2E, light-blue dots). The alterations to the 601 sequence have thus led to a mild increase in resolution and a noticeably better distribution of information points. However, these five points of information are less than we expected based on our theoretical design. Similar to the optimized 99bp linker, our assumption of the minimal separation of CpG sites were deemed problematic. We find that each of the information points within the 601 sequence consists of minimally two grouped CpG sites (Figure 2.3.2G, group I-V). Future designs should increase the distance between CpG sites to prevent the grouping during analysis. Alternatively, new training data or improved chemistry may reduce the ten basepair limitation (see Discussion related to this chapter).

Based on the altered resolution within the linker and the 601 sequence, we were interested how the overall resolution of the newly synthesized array was affected. To this end, we calculated the average CpG group size of the entire array and compared it to the reference array. The group size is a good indication of the resolution, as under theoretical ideal circumstances it is one. A group size of one would suggest that each CpG site is a unique and individual point of information. The theoretical opposite would be that the entire array would be a single group, providing only a single point of information for the entire array. Our analysis indicates that the 99bp+URS group size is larger compared to the reference array. Presumably due to the large groups of CpGs in the linker regions. We observe a mild increase from an average of 3.2 to 3.6 CpG sites per group for reference array and 99bp+URS array, respectively. The mean number of groups per array decreased mildly as well, from approximately 142 to 131 for the reference array and the 99bp+URS array, respectively. Taken together, the data suggest that the overall resolution of the newly synthesized array was actually decreased by these presumed optimizations. However, these arrays would be suitable for exploratory experiments and provide useful insights that can be taken into account when designing a new array.

### 2.3.5. Discussion (related to this chapter)

#### 2.3.5.1. Future optimizations to the CpG distribution and density

Based on the preliminary data outlined in this chapter, we find that the optimizations can positively impact the results by MEFSIMO-seq. These findings give rise to new ideas on future optimizations to the sequence. For example, the optimized 601 sequence initially was designed to contain 15 unique CpG sites, throughout the 147 bp sequence. Sequencing results show that we detect approximately five distinct CpG sites/groups unambiguously within this stretch. This suggests, that despite the careful distribution of CpG sites, certain sites are still grouped together. A clear example of grouping is the CpG sites surrounding the CpG dyad (Figure 2.3.1C). We left these sites untouched as previous findings suggested that this region is in the top three of important locations for a strong octamer interaction (Chua et al., 2012). On further examination, other sites may have been too close together in the reference sequence (first three CpG sites). Resolution may have benefitted from the deletion of the first CpG site and moving it upstream towards the nucleosome entry site. However, nucleosome entry and exit sites may already have weak interactions between DNA and octamer. Further CpG additions could make these regions even stiffer leading to an increase in DNA unwrapping or difficulty in assembly (Ngo et al., 2015).

The current computational limitation prevents the ability to distinguish individual CpG sites when they are within a ten basepair proximity. Alternative training data may circumvent this limitation. However, the limitation lies within the ability to create a large enough library where CpG sites are spaced at varying intervals with varying nucleotides as spacers. Moreover, it would theoretically require the combination of all iterations of modified versus unmodified CpG sites. These combinations should be known prior to sequencing to accurately validate the resulting training data. As such a library is not available or hard to create, current training data relies on fully methylated DNA. Alternatively, a prediction model could be implemented where, based on current knowledge and data, more combinations can be added and subsequently verified. Neither of the above mentioned datasets are available, but as the ambition of multiple research groups to accurately detect modified sites increases, such tools may become available in the near future.

#### 2.3.5.2. Elucidating the mechanism of ISWI remodeling utilizing a resolution-optimized array

Optimizations to the 601 sequence and linkers were initially performed to further elucidate the spacing and sliding mechanism of the ISWI remodeler. These preliminary results provide a first insight on the actual feasibility. A major hurdle will be the localization of nucleosomes with high precision. This hurdle is three-fold: Firstly, acquiring an *in vitro* array with a theoretically optimal CpG distribution may be difficult to design and produce (see above). Secondly, software and computational limitations group CpG sites together despite best efforts, thereby reducing the resolution. Moreover, these limitations currently prevent a resolution finer than ten basepairs (Simpson et al., 2017). Thirdly, even if software and computational limitations improve and allow a higher resolution, the density of CpG sites can (and should) not be increased. Too high of a CpG density severely impedes the affinity of the 601 sequence to histone octamers and may bias nucleosome sliding.

Alternative approaches may be an outcome to overcome the limitations described above. For example, utilization of a different or additional methyltransferase such as EcoGII (m6A) or M.CviPI (5mC) aide in the overall resolution by modifying more accessible nucleotides. Software and computational limitations may be circumvented by applying bisulfite conversion of methylated cytosines (potentially in combination with a cocktail of M.SssI and M.CviPI enzymes, similar to NOME-seq and/or nanoNOME (Kelly et al., 2012; Lee et al., 2020)). This conversion abrogates the requirement to call methylated sites directly, and instead compares converted nucleotides to the reference sequence. A completely different sequencing strategy could also be implemented. PacBio SMRT-sequencing may be highly useful for sequencing an *in vitro* array, especially in combination with the SMRTbell technology. This technology circularizes the array, and subsequently sequences each strand multiple times. This assures high accuracy, especially for highly repetitive sequences such as these *in vitro* arrays (Flusberg et al., 2010; Rhoads and Au, 2015). Moreover, resolution may not be limited by CpG density as the physical and computational approaches do not constrain detection of neighboring modified sites.

Taken together, we have shown that optimizations could improve detection of nucleosomes and aide in better understanding of the remodeling mechanism. However, as with any approach, caveats exist and more analyses need to be performed to identify whether our method can provide further insights. The current preliminary data provides a foundation to design future arrays and/or sequencing strategies.

# 3. GENERAL DISCUSSION AND OUTLOOK

A discussion relevant to each chapter is presented at the end of the respective chapter. Here, I will summarize and discuss the general aspects of my thesis and provide an outlook into future experiments and scientific directions.

In the first results chapter, we demonstrate the development and careful characterization of MEFSIMO-seq. We developed this approach to overcome several limitations of more traditional sequencing techniques. MEFSIMO-seq can provide novel insights into cell-to-cell heterogeneity of intact, long-range nucleosomal arrays. To this end, we explore several methyltransferase enzymes to identify their usefulness and efficiency. We provide detailed experimental conditions that allow optimal methylation of accessible DNA both *in vitro* and *in vivo.* Biochemical and bioinformatical tools were used to identify potential issues that would prevent high quality data. Lastly, we describe a set of key observables by which the data can be compared to ensure reproducibility.

The second chapter of this thesis utilizes MEFSIMO-seq to study the *in vivo* chromatin landscape of *S. cerevisiae.* We exploit the single-molecule long-range sequencing capabilities of MEFSIMO-seq to scrutinize the current understanding of the multi-kilobase rDNA locus. We find contradictory results concerning transcriptional regulation and identify the binding region of the UAF complex. Our data furthermore indicates that chromatin regularity is prevalent, even in regions that traditionally would be characterized irregular. We investigate the role of several key chromatin remodelers (Isw1, Isw2, Chd1 and Ino80). Our findings provide direct evidence that deletion/depletion of these proteins results in genome wide irregularity and a lack of phasing, and provide additional *in vivo* evidence that the INO80 remodeler can space nucleosomes. Contrary to *in vitro* experiments, our data does not support the 'clamping' mechanism by which remodelers organize nucleosomal arrays. Lastly, preliminary results indicate that MEFSIMO-seq can be used to accurately assign yeast variants and to discern differences in chromatin organization. Taken together, this sequencing approach allows for rapid characterization novel of features of the *in vivo* chromatin landscape.

In the third and last chapter of this thesis we explore whether *in vitro* arrays in combination with single-molecule sequencing can be used to study remodeler functioning in depth. To this end, we redesigned arrays to allow for optimal detection of nucleosomal positions. We find that our optimizations of the Widom 601 sequence retain the ability to accurately position octamers, whilst increasing the experimental resolution. Our single-molecule data furthermore provides valuable insights into the newly designed array, allowing the implementation of further improvements in future iterations. Utilizing the new arrays can aide to better understand how remodelers shape chromatin by elucidating how they alter chromatin at single-molecule resolution *in vitro.*

## 3.1. Future iterations and alternatives to the MEFSIMO-seq approach

Throughout this thesis I have referred to several methods and approaches that could theoretically significantly improve the spatial resolution of MEFSIMO-seq. Current limitations are due to the sparseness of CpG sites and the computational limitation of the Nanopolish software (Simpson et al., 2017). Alternative methods include the utilization of alternative m6A methyltransferases such as EcoGII (NEB) or Hia5 (Stergachis et al., 2020). These enzymes methylate all accessible adenines instead of only cytosines in a CpG context as used here, greatly increasing the number of modifieable sites.

Alternatively, in Chapter 1, we investigated the use of multiple methyltransferases simultaneously, specifically a CpG and GpC methyltransferase (M.SssI and M.CviPI, respectively). This approach has been implemented in previous studies, but as demonstrated may be suboptimal due to experimental or computational limitations (Shipony et al., 2020). Despite the caveats identified in this thesis, and the currently inherent computational limitations, approaches using alternative or multiple enzymes theoretically aide the spatial resolution of MEFSIMO-seq directly. A higher density of methylated sites will provide improved determination of nucleosomal position and thus the chromatin landscape and cell-to-cell heterogeneity.

Instead of the Nanopore sequencing platform utilized by the current version of MEFSIMO-seq, future experiments may be performed using the PacBio platform. In practice, the methylation of samples would remain identical for both platforms. An adaptation in library preparation and parts of the computational pipeline would need to be implemented in order to handle the new information. However, the conceptual output of either platform would be similar, as both contain information on the state of individual nucleotides. One advantage of PacBio over Nanopore is the ability to perform circular consensus sequencing (CCS). This approach sequences the Watson and Crick strand multiple times, improving output accuracy (Travers et al., 2010). Moreover, the software is not limited by training data to detect methylated nucleotides, but instead relies on the incorporation time (Rhoads and Au, 2015). However, the advantages of Nanopore sequencing are its high throughput in terms of sequencing depth and ease of use. These two features allow for rapid turnover between experiments, whilst (potentially in combination with the optimizations mentioned above), creating large datasets.

### 3.2. Elucidating the heterochromatin landscape

The yeast *Schizosaccharomyces pombe* is an established single celled model organism to study the role of heterochromatin (reviewed in Cam and Whitehall, 2016). Heterochromatin is a repetitive and condensed form of chromatin and is typically found in higher eukaryotes, but not in *S. cerevisiae.* The role of heterochromatin is often related to gene regulation and maintenance of chromosome integrity (Pidoux and Allshire, 2004). In *S. pombe*, the heterochromatin is found at centromere, telomere and mating type regions. Due to its repetitiveness and A/T richness, traditional sequencing methods have difficulties mapping their short sequencing reads accurately to the reference genome. For example, nucleosome density at centromeres was suggested to be sparse (Moyle-Heyrman et al., 2013). However, observation may be skewed due to the low mapping accuracy (16% uniquely and 90% non-uniquely mappable). In contrast to traditional methods, MEFSIMO-seq lends itself exceptionally well to study such types of chromatin. Firstly, the repetitive DNA in combination with the long-range sequencing capabilities retain the ability to map reads accurately to the reference genome. Secondly, the methylation footprinting allows direct detection of the position and density of nucleosomes in the heterochromatin regions.

The ability to easily manipulate *S. pombe* genetically can further be used to study how heterochromatin is formed and how heterochromatin-deficient mutants have altered chromatin landscapes. Earlier observations already suggested that histone methyltransferases, histone deacetylases and chromo-domain proteins are important for heterochromatin formation (Garcia et al., 2010; Sadaie et al., 2008; Thon and Verhein-Hansen, 2000). However, such studies all lack the same kind of information only possible by using single-molecule sequencing techniques such as MEFSIMO-seq. It would therefore be highly interesting to see how such mutants affect heterochromatin assembly and/or organization.

### 3.3. Chromatin changes during cell-cycle and aging

In this thesis, we have generally utilized log-phase growing yeast cells, which are freshly inoculated, typically no more than 24 hours before. Under wild-type conditions, such cells duplicate approximately every 90 minutes. Within the log-phase growing culture, individual cells may have undergone a single to dozens of multiplications. We performed preliminary experiments utilizing strains in which we arrested the cell cycle at G1 phase (Chapter 2.2.14). Here we did not find any noticeable differences between arrested and un-arrested cells. However, during this initial set of experiments, we did not gather the necessary data to investigate the occupancy change throughout the cell cycle. Nor did we investigate if nucleosome remodelers influence the kinetics of the nucleosome (dis)assembly *in vivo* (Figure 2.2.13A). Future experiments using MEFSIMO-seq should be performed to directly test the hypothesis whether nucleosome assembly is affected in TKO cells. Not only does MEFSIMO-seq allow to uniquely detect these changes in occupancy/assembly, it could also directly inform on which regions get assembled first and how the regularity is established thereafter.

Changes in nucleosome occupancy or chromatin landscape also occur throughout the so-called replicative lifespan of yeast. For instance, instable rDNA loci may form extra-chromosomal circles (ERCs), which are correlated with reduced age of a single cell (Sinclair and Guarente, 1997). Conversely, caloric restriction may increase the replicative lifespan of cells in a Gcn4-dependent manner (Mittal et al., 2017). Chromatin landscape related studies which indicate that gene promoters tend to be closed at elevated ages, also suggest a role for transcription factors and nucleosome remodelers (Hendrickson et al., 2018). Interestingly, histone depletion has been associated with aging cells despite the increase in histone transcript levels (Dang et al., 2009; Lesur and Campbell, 2004). In future experiments, MEFSIMO-seq can be utilized to accurately determine the change in nucleosome occupancy during aging. Importantly, this approach can be applied to directly detect the (increased) cell-to-cell heterogeneity throughout the aging process. Further studies using remodeler-depleted strains may shed even further light on the role of remodelers during the aging process.

### 3.4. Novel prospects on *in vitro* remodeling

The third and final chapter of this thesis utilizes an *in vitro* chromatinized array consisting of CpG optimized linkers and Widom 601 sequences. We demonstrated that utilization of this array can aide in elucidating the method of regular array generation by remodelers using MEFSIMO-seq. However, the current iteration of optimized arrays can still benefit from further optimizations to improve sequencing resolution. Despite these limitations, the array would still be highly suitable to study remodelers using non-sequencing approaches. For example, atomic force microscopy (AFM) in collaboration with the Lipfert lab (LMU, Munich, Germany) can be used to directly visualize the state of arrays before and/or after remodeling. AFM is capable of detecting the position of individual nucleosomes and goes well beyond the diffraction limit of (typical) fluorescence microscopy (Konrad et al., 2021). AFM cannot capture the dynamics of nucleosome sliding. However, high-speed AFM is able to detect real-time sliding and could be performed in collaboration with the Dalal lab (NCI, Bethesda, USA). This type of AFM can image dynamic processes at 15 – 25 frames per second, allowing for the identification of changes over time (Shibata et al., 2010). However, high-speed AFM has a lower resolution compared to regular AFM, making the detection of individual nucleosomes potentially difficult. Regardless, AFM can provide a powerful tool in tandem with MEFSIMO-seq to study the function of nucleosome remodelers in the context of a chromatinized array *in vitro*.

# 4. MATERIALS AND METHODS

## 4.1. Materials

### 4.1.1. Yeast strain list

| ***Strain*** | ***Common name*** | ***Genotype*** | ***Source*** | ***Method*** |
|---|---|---|---|---|
| yFMP012 | BY4741/ S288C WT | BY4741 MATa; his3Δ 1; leu2Δ 0; met15Δ 0; ura3Δ 0 | Euroscarf | - |
| yFMP013 | W1588-4C/ W303 WT | W303-1A MATa ade2-1 his3-11,15 leu2-3,112 trp1-1 ura3-1 can1-100 rad5 | (Tsukiyama et al., 1999) | - |
| yFMP014 | YTT227/ W303 TKO | MATa ade2-1 his3-11,15 leu2-3,112 trp1-1 ura3-1 can1-100 RAD5+ isw1Δ:ADE2 isw2Δ::LEU2 chd1Δ::TRP1 | (Tsukiyama et al., 1999) | - |
| yFMP266/ yFMP267/ yFMP268 | HHY170 TKO/ Pol II AA | MATa ade 2-1 trp1-1 can1-100 leu2-3,112 his3-11,15 ura3 GAL psi+ tor1-1 fpr1::NAT RPL13A-2xFKBP12::TRP1 RPO21-FRB::kanMX6 isw1::ADE2 isw2::LEU2 chd1::TRP1 RAD5 | (Singh et al., 2021) | - |
| yFMP333/ yFMP334/ yFMP335 | HHY168 TKO Ino80-GFP-FRB/ Ino80 AA | MATa ade2-1 trp1-1 can1-100 leu2-3,112 his3-11,15 ura3 tor1-1 fpr1Δ::NAT RPL13A-2×FKBP12::TRP1 isw1Δ::ADE2 isw2Δ::LEU2 chd1Δ::TRP1 rad5-G535R INO80-GFP-FRB::hphMX6 | (Singh et al., 2021) | - |
| yFMP377/ yFMP378 | S288C TKO | MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 isw1::kanMX isw2::kanMX chd1::HYG | (Singh et al., 2021) | - |
| yFMP464/ yFMP465 | S288C WT bar1Δ | S288C MATa ade2-1 his3-11,15 leu2-3,112 trp1-1 ura3-1 can1-100 rad5 bar1Δ::HIS3 | This study | Transformation in yFMP012 |
| yFMP356/ yFMP357 | TKO HD | MATa ade2 can1 his3 leu2 lys2 trp1 ura3 hht1-hhf1Δ::LEU2 hht2-hhf2Δ::kanMX3 isw1::ADE2 isw2::LEU2 chd1::TRP1 pFMP519[P(GAL10)-HHT2, P(GAL1)-HHF2] | (Singh et al., 2021) | - |
| yFMP358/ yFMP359 | HD | MATα ade2 can1 his3 leu2 lys2 trp1 ura3 hht1-hhf1Δ::LEU2 hht2-hhf2Δ::kanMX3 pFMP519[P(GAL10)-HHT2, P(GAL1)-HHF2] | (Singh et al., 2021) | - |
| yFMP466 | W303 WT bar1Δ | W303-1A MATa ade2-1 his3-11,15 leu2-3,112 trp1-1 ura3-1 can1-100 rad5 bar1Δ::HIS3 | This study | Transformation in yFMP013 |
| yFMP467 | W303 TKO bar1Δ | W303-1A MATa ade2-1 his3-11,15 leu2-3,112 trp1-1 ura3-1 can1-100 but rad5, isw1Δ::ADE2 isw2Δ::LEU2 chd1Δ::TRP1 bar1Δ::HIS3 | This study | Transformation in yFMP014 |

### 4.1.2. Bacterial strains

| Strain | Manufacturer |
|---|---|
| BL21-Gold (DE3) | Agilent Technologies |
| DH5α | NEB 5-alpha |

### 4.1.3. Plasmid list

| Plasmid | Common name | Description | Source |
|---|---|---|---|
| pFMP128 | Histone H2A | Histone H2A | - |
| pFMP129 | Histone H2B | Histone H2B | - |
| pFMP186 | Histone H3 | Histone H3 | - |
| pFMP187 | Histone H4 | Histone H4 | - |
| pFMP210 | ISWI FL 1-1027 | Full length drosophila ISWI | - |
| pFMP232/ pFMP233 | 25mer array | 25 601 sequences with 50 bp linkers. 13 unique restriction sites (URS) and 12 original 601 sequences | (Mueller-Planitz et al., 2013) |
| pFMP541 | pUC18-10mer_99bp | 10mer CpG optimized 601 sequence with 99 bp linkers | This study |
| pFMP542 | pUC18-12mer_VLL | 12mer CpG optimized 601 sequence with variable linker lengths | This study |
| pFMP546 | pUC18-99bp_URS | Insertion of 99 bp array from pFMP541 into pFMP232 | This study |
| pFMP547 | pUC18-99bp_VLL | Insertion of 99 bp array from pFMP541 pFMP542 | This study |

### 4.1.4. Oligo list

| Oligo | 5’-3’ sequence | Purpose | Purification |
|---|---|---|---|
| oFMP1235 | ATCGCCTAAAATCATACCAAAATAAAAAGAGTGTCTAGAAGGGTCATATATGGCTTAACTATGCGGCAT | Bar1 cloning upstream overlap | HPLC |
| oFMP1236 | ACTATATATTTGATATTTATATGCTATAAAGAAATTGTACTCCAGATTTCCTCCTTACGCATCTGTGC | Bar1 cloning downstream overlap | HPLC |
| oFMP1237 | CGTAACTATTGCAACGAAATGCTTTTGAAG | Bar1 KO control forward primer | Desalt |
| oFMP1238 | TTGTTCAGGCATTTATACGTTTGGTTAGTTCAG | Bar1 KO control reverse primer | Desalt |
| oFMP1239 | ACCTTGGGCGCATTCAATAGC | Bar1 KO control internal primer | Desalt |
| oFMP1240 | GTGCCTCATCCAAAGGCG | HIS insert control primer | Desalt |

### 4.1.5. Enzymes and kits

| Description | Manufacturer (Catalogue number) |
|---|---|
| 5-mC DNA ELISA kit | Zymo Research (D5325/D5326) |
| Apyrase | NEB (M0398S) |
| BamHI-HF | NEB (R3136S) |
| Blunt/TA ligase master mix | NEB (M0367L) |
| BsaI_HF v2 | NEB (R3733S) |
| Bromophenol blue | Sigma (B0126) |
| BsiWI-HF | NEB (R3553S) |
| DraIII-HF | NEB (R3510S) |
| EcoGII (m6A methyltransferase) | NEB (M0603S) |
| HaeIII | NEB (R0108S) |
| HpaII | NEB (R0171S) |
| Ligation Sequencing kit | ONT (SQK-LSK109) |
| M.CviPI (GpC methyltransferase) | NEB (M0227L) |
| M.SssI (CpG methyltransferase) | NEB (M0226M) |
| NEBNext® compation module for ONT® ligation sequencing | NEB (E7180S) |
| NEBNext® FFPE DNA repair mix | NEB (M6630L) |
| NEBNext® quick ligation module | NEB (E6056L) |
| NEBNext® Ultra™ II end repair/dA-tailing module | NEB (E7546L) |
| NotI-HF | NEB (R3189S) |
| NruI-HF | NEB (R3192S) |
| NucleoBond Xtra Midi | Macherey-Nagel (740410.50) |
| NucleoSpin Gel and PCR Cleanup | Macherey-Nagel (740609.250) |
| NucleoSpin Plasmid EasyPure | Macherey-Nagel (740727.250) |
| OneTaq® DNA polymerase | NEB (M0482L) |
| PCR-Free barcoding expansion 1-12/13-24 | ONT (EXP-NBD104/EXP-NBD114) |
| Phusion® high-fidelity DNA polymerase | NEB (M0530L) |
| Proteinase K | Bioline (BIO-37039) |
| Qubit Assay Tubes | Thermo Fisher Scientific (Q32856) |
| Qubit dsDNA HS Assay kit | Thermo Fisher Scientific (Q32854) |
| RNase A | Sigma (R4875) |
| SalI-HF | NEB (R3138S) |
| T4 DNA Ligase | NEB (M0202L) |
| XbaI | NEB (R0145S) |
| Zymolyase-100T | Gerbu Biotechnik (07665) |

### 4.1.6. Chemicals and consumables

| Description | Manufacturer (Catalog number) |
|---|---|
| 1 kb DNA Ladder | NEB (N3232S) |
| 100 bp DNA Ladder | NEB (N3231S) |
| 1kb extend DNA ladder | NEB (N3239S) |
| Agarose Universal | Bio&Sell (BS20.46.500) |
| Agencourt AMPure XP beads | Beckmann Coulter (A63882 |

| | |
|---|---|
| Amicon Ultrafiltration devices (30.000 MWCO) | Merck (UFC803024/UFC903024) |
| Alpha factor | GenScript (59401-28-4) |
| Ampicillin | Roth (K029.2) |
| Aprotinin | Genaxxon Bioscience (M6361.0010) |
| Arginine | BD Biosciences (214010) |
| ATP | Sigma (A2754-1G) |
| Bacto Agar | BD Biosciences (211820) |
| Bacto Peptone | Life technologies (211820) |
| Bromophenol Blue | Sigma (B0126-25G) |
| BSA | Sigma (A9418) |
| Calcium Chloride | Sigma (C3306) |
| Chloroform | VWR Chemicals (22711.324) |
| Complete Protease Inhibitor Cocktail tablet | Sigma (11836145001) |
| Costar Spin-X centrifuge tube filters | Sigma (CLS8162-96EA) |
| Coverslips | Roth (0657.2) |
| Cryobox | Kisker Biotech (R034-7) |
| CutSmart | NEB (B7204S) |
| DAPI | Sigma (DUO82040) |
| Difco Yeast Nitrogen Base | BD Biosciences (291920) |
| Dipotassium phospate | VWR (1.05099.1000) |
| DMSO | Sigma (D2438) |
| dNTPs | NEB (N0447S) |
| Drop out powder components (Ade, Ala, Asn, Asp, PABA, Cys, Glu, Gln, Gly, Ile, Myo-Inositol, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr, Ura, Val) | Sigma (A8626, A7627, A9256, A5040, C1276, G1251, G3126, G8790, I2752, I5125, L8000, L5626, M9625, P2126, P0380, S4500, T8625, T0254, T3754, U0750, V0500) |
| DTT (Dithiothreitol) | Life technologies (R0861) |
| EDTA | Pan Reac Appli Chem (131669.1210) |
| EGTA | Roth (3054.3) |
| EtOH 100% - high quality | Sigma (32205-2.5L-M) |
| EtOH 96% - low quality | CLN GmbH (N-1196.9025) |
| Ficoll PM400 | Sigma (F4375) |
| Formaldehyde | Sigma (47608) |
| Galactose | Sigma (G0625) |
| Glass beads | Roth (N030.1) |
| Gel loading dye, Purple (6x) no SDS | NEB (B7025S) |
| Glucose | VWR (1.08342.1000) |
| Glycerol | VWR (1.04092.2500) |
| Glycoblue | Thermo Fisher Scientific (AM9515) |
| Glycogen | Sigma (10901393001) |
| Guanidine hydrochloride | Merck (369079) |
| Haemocytometer | Fischer Scientific (11314052) |
| HEPES | VWR Chemicals (1.10110.1000) |
| Histidine | VWR (1.04351.0100) |
| Isopropanol (2-Propanol) | Sigma (34863-2.5L-M) |

| | |
|---|---|
| Leupeptin | Genaxxon (M6100.0100) |
| Lithium acetate | Sigma (L6883) |
| Low-melt agarose | Biozym (850070) |
| Lysine | Sigma (L8662) |
| Magnesium Chloride | VWR (25108.295) |
| Magnetic Rack | GE Healthcare (28948964) |
| MaXtract High Density | Qiagen (129073) |
| MinION flow cell (R9.4.1) | ONT (FLO-MIN106D) |
| NEBuffer 2 | NEB (B7002S) |
| IGEPAL (NP-40) | Sigma (I8896) |
| Orange G | Sigma (O-1625) |
| PEG (Polyethylene glycol) 3350 | Sigma (P-3640) |
| Pepstatin | Genaxxon (M6539.0100) |
| Peptone | Life technologies (211820) |
| Phenol:Chloroform:Isoamyl alcohol (25:24:1) | Roth (A156.1) |
| phosphoenolpyruvate | Molekula (5541-93-5) |
| PMSF (Phenylmethanesulfonylfluoride) | Sigma (P7626) |
| Polyethylene glycol | Sigma (25322-68-3) |
| Potassium acetate | VWR (104820) |
| Potassium chloride | Sigma (P9541) |
| Potassium dihydrogen phosphate | Labochem International (LC5067.1) |
| Potassium hydroxide | Merck (5033-5000) |
| Protease Inhibitor Cocktail | Sigma (P8215) |
| pyruvate kinase-lactate dehydrogenase mixture | Sigma Aldrich (P0294) |
| Rapamycin | Hölzel Diagnostika (R-5000) |
| S-adenosylhomocysteine (SAH) | Merck (A9384) |
| S-adenosylmethionine (SAM) | NEB (B9003S) |
| Sodium hydroxide | Neolab (LC-4994.2) |
| Sodium acetate | Merk (106268) |
| Sodium dodecyl sulfate (SDS) | Serva Electrophoresis (20765.03) |
| Sodium chloride | Serva (30183.01) |
| Sorbitol | Serva Electrophoresis (35230.02) |
| Sucrose | Merck (1.07687.1000) |
| SYTOX Green | Invitrogen (S7020 |
| TBE | VWR (J885-4L) |
| Tris ultrapure | Diagonal (A1086.1000) |
| Triton X-100 | Sigma (T8787) |
| Tween-20 | Sigma (P9416) |
| Ultra Pure Salmon Sperm DNA Solution | Life technologies (15632011) |
| Urea | Life technologies (15505027) |
| Whatman blotting paper | VWR (588-3148) |
| Yeast Extract | BD Biosciences (212750) |
| β-Mercaptoethanol | Sigma (M6250) |

| | |
|---|---|
| Sodium azide | Merck (106268) |
| Uracil | Sigma (U0750-100G) |
| Titriplex VI | Merck (108435) |
| Sodium dihydrogenphosphate | Merck (106346) |
| Slide-a-lyzer™ MINI dialysis device 7k MWCO 0.1mL | ThermoFisher Scientific (69560) |
| Slide-A-Lyzer™ MINI Dialysis Device, 3.5K MWCO, 0.5 mL | Life Technologies (88400) |
| Spectra/Por 1 dialysis tubing 6 – 8kD MWCO | Repligen (132645) |
| Spectra/Por 3 dialysis tubing 3.5kD MWCO | Repligen (132720) |

### 4.1.7. Buffers and solutions

| **Description** | **Components** |
|---|---|
| 5x SDS-PAGE loading dye | 250 mM Tris-HCl (pH 6.8), 10% (w/v) SDS, 0.1% (w/v) bromophenol blue, 50% (v/v) glycerol, 0.5 M DTT |
| 10X TE buffer | 100 mM Tris-HCl, 1 mM EDTA, pH 7.5 or 8.0 |
| 10x M.SssI buffer | 100 mM Tris-HCl pH 7.9, 500 mM NaCl, 100 mM MgCl2, 1000ug/mL BSA. Add 10mM DTT fresh when using |
| 10X MNase digestion buffer | 150 mM Tris-HCl pH 7.5, 500 mM NaCl, 14 mM CaCl2, 2 mM EGTA, 2 mM EDTA, 50 mM β-Mercaptoethanol |
| 10x STOP buffer | 50mM Tris-HCl pH 8.0, 4% SDS, 100mM EDTA pH 8.0 |
| 1X LiOAc buffer | 0.1 M LiOAc, 10 mM Tris-HCl pH 8.0, 1 mM EDTA |
| 20x regenerating system | 120 mM phosphoenolpyruvate (Molekula), 310 U/mL pyruvate kinase–lactate dehydrogenase mixture (Sigma), 20 mM DTT in 1x remodeling buffer |
| 5x remodeling buffer stock | 125 mM Hepes–KOH pH 7.6, 5 mM MgCl2, 0.5 mM EDTA, 250 mM NaCl, 50% glycerol. Store at -20C |
| 5x TB buffer pH 8.3 | 445 mM Tris base, 445 mM boric acid |
| 5x TBE buffer pH 8.3 | 445 mM Tris base, 445 mM boric acid, 10 mM EDTA pH 8 |
| 6x Blue gel loading buffer | 0.1% Bromophenol Blue, 60% glycerol in TE buffer |
| 6x Orange G gel loading buffer | 0.1% Orange G, 60% glycerol in TE buffer |
| Coomassie Blue staining | 0.1% (w/v) Bromophenol Blue R, 50% (v/v) ethanol, 10% (v/v) acetic acid |
| CpG+GpC buffer | 50mM Tris-HCl pH 7.6, 50mM NaCl, 10mM MgCl2, 10mM DTT |
| Dialysis buffer | 15mM Tris-HCl pH 7.4, 150mM NaCl and 1mM DTT |
| EX50 buffer | 10 mM Hepes–KOH pH 7.6, 50 mM KCl, 1.5 mM MgCl2, 0.5 mM EGTA. |
| Ficoll buffer | 18% Ficoll, 20 mM KH2PO4 pH 6.8, 1 mM MgCl2, 0.25 mM EGTA, 0.25 mM EDTA |
| GF buffer | 50mM Hepes-KOH pH 7.6, 0.2mM EDTA, 20mM KOAc, 10mM DTT |
| HisA buffer | 50mM Tris-HCl pH 7.4, 300mM NaCl |
| HisB buffer | 50mM Tris-HCl pH 7.4, 300mM NaCl, 400mM imidazole pH7.6 |
| High salt buffer | 10 mM Tris–HCl pH 7.6, 2 M NaCl, 1 mM EDTA pH 8, 0.01% NP-40, 1 mM DTT |
| Low salt buffer | 10mMTris–HCl pH 7.6, 50mMNaCl, 1mM EDTA pH 8, 0.01% NP-40, 1 mM DTT |
| MonoS A buffer | 15mM Tris-HCl pH 7.4, 1mM DTT |
| MonoS B buffer | 15mM Tris-HCl-pH 7.4, 2M NaCl, 1mM DTT |
| No salt buffer | 10 mM Tris–HCl pH 7.6, 1 mM EDTA pH 8, 0.01% NP-40, 1 mM DTT |
| PEG solution | 50% w/v PEG 3350, 1x TE buffer, 100mM LiOAc |

| | |
|---|---|
| Precipitation buffer | 10 mM Tris–HCl pH 7.6, 7 mM MgCl2 |
| Preincubation solution | 0.7 M ß-mercaptoethanol, 28 mM EDTA pH 8.0 |
| Proteinase K solution | 50 mM Tris-HCl pH 8.0, 10 mg/mL proteinase K |
| Quenching solution | 0.4% SDS, 20 mM EDTA |
| Refolding buffer | 10 mM Tris-Cl pH 7.5, 2 M NaCl, 1 mM EDTA, 5 mM β-Mercaptoethanol |
| RNase solution | 50 mM Tris-HCl pH 8.0, 0.4 mg/mL RNaseA |
| SDS-PAGE fixing solution | 10% (v/v) acetic acid, 50% (v/v) ethanol |
| SDS-PAGE running buffer | 2.5 mM Tris, 19.2 mM glycine, 0.01% SDS, pH 8.3 |
| Single-stranded carrier DNA | 10 mg/ml salmon sperm DNA, 10 mM Tris-HCl, 1 mM EDTA pH 8.0, 1mM EDTA |
| SYTOX solution | 10 mL 1x TE buffer, 1 ul SYTOX (1:10.000 dilution) |
| Sorbitol-β-ME | 1 M sorbitol, 5 mM β-Mercaptoethanol |
| Unfolding buffer | 7 M guanidine hydrochloride, 20 mM Tris-Cl pH 7.5, 10 mM DTT |
| Vehicle solution | 90% ethanol, 10% Tween-20 |

### 4.1.8. Growth media

| **Description** | **Components** |
|---|---|
| YPAD (full media) | 1 g/L KH2PO4, 10 g/L Yeast Extract, 20 g/L peptone, 20 g/L glucose or galactose, 100 mg/L adenine |
| YNB (synthetic media) | 6.7 g/L Yeast Nitrogen Base, 1.6 g/L amino acid dropout-mix (-His, -Leu, -Ura, -Trp), 20 g/L glucose or galactose, pH 5.4. Optional supplement with 84 mg/L His / Trp / Ura, 168 mg/L Leu |
| LB | 10g/L Tryptone, 5g/L yeast extract, 10g/L NaCl |

## 4.2. Experimental methods

### 4.2.1. PCR amplification

Amplification of DNA was performed by utilizing OneTaq® DNA polymerase or Phusion DNA polymerase. 5x OneTaq® standard reaction buffer or 5x high-fidelity (HF) buffer was used, respectively. Primer design took into account that annealing temperatures were between 60-65 °C or between 52-58 °C for Phusion® and OneTaq®, respectively. Primers were designed using online NEB Tm calculator. Amplification was performed in Eppendorf Mastercycler nexus GX2. PCR reaction mixture for both enzymes was prepared in 0.1 mL PCR tubes and was as follows:

| Component | Volume | Final conc. |
|---|---|---|
| Phusion® polymerase | 0.5 µL | 1 unit |
| 5x HF buffer | 5 µL | 1x |
| 10 mM dNTP | 0.5 µL | 0.4 mM |
| 10 µM Fwd. primer | 1.25 µL | 0.5 µM |
| 10 µM Rev. primer | 1.25 µL | 0.5 µM |
| Template DNA | 1-3 µL | Variable |
| MQ water | To 25 µL | |

| Component | Volume | Final conc. |
|---|---|---|
| OneTaq® polymerase | 0.25 µL | 1.25 units |
| 5x OneTaq® buffer | 10 µL | 1x |
| 10 mM dNTP | 1 µL | 0.2 mM |
| 10 µM Fwd. primer | 1 µL | 0.2 µM |
| 10 µM Rev. primer | 1 µL | 0.2 µM |
| Template DNA | 1-3 µL | Variable |
| MQ water | To 50 µL | |

PCR was performed as follows:

| Stage | Temperature | Duration | Cycle |
|---|---|---|---|
| Initial denature | 95 °C | 30 sec | 1 cycle |
| Denaturation | 95°C | 30 sec | 30 cycles |
| Annealing | Var. °C | 60 sec | |
| Extension | 68 °C | 60 sec/kb | |
| Final extension | 68 °C | 300 sec | 1 cycle |
| Hold | 10 °C | ∞ | |

### 4.2.2. DNA separation by gel electrophoresis

Agarose gel was made by adding 0.8 – 1% (w/v) agarose to 1x TBE buffer. When working with genomic DNA 0.6 – 0.7% (w/v) agarose was added to 1x TBE buffer. Mixture was heated in microwave until just boiling. After mixture was cooled so it could be handled with gloves, 0.1 µg/mL ethidium bromide was added in a fume hood. Mixture was poured into gel casting tray, desired amount of combs were added and then left to solidify. DNA was mixed with 1x NEB gel loading dye, purple, no SDS and loaded into wells. Gel was electrophoresed at 5 – 10 V/cm until desired separation was achieved. Visualization was performed using Peqlab Vilber Gel Documentation system. Raw images were saved as 16-bit TIF format.

### 4.2.3. DNA purification after gel electrophoresis

Desired DNA fragments were physically cut out of agarose gel and transferred to separate Eppendorf tube. NucleoSpin Gel and PCR clean-up kit was subsequently utilized to isolate DNA. Extraction was performed according to manufacturer's protocol.

### 4.2.4. SDS-PAGE

In order to separate proteins by mass, Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was used. Samples were denatured by heating to 95 °C for 5 minutes. To each sample 1x SDS-PAGE loading dye was added and samples were loaded on appropriate gel (typically 10% or 8 –

16% gradient) with a relevant marker to indicate molecular weights. Gels were ran in 1x SDS-PAGE buffer at 100V until desired separation was achieved. Gel was immersed in SDS-PAGE fixing solution for 15 minutes and stained using Coomassie Blue staining solution for 15 minutes or until bands appeared. Excess stain was briefly washed in MilliQ water and gel was imaged using ChemiDoc gel imaging system.

### 4.2.5. *E. coli* cultivation

Unless stated otherwise, *E. coli* cells were grown at 37 °C on solid or in liquid LB media containing the desired antibiotic (typically Ampicillin). Liquid growth was performed in <MACHINE> shaking at 130 rpm. E. coli grown on solid LB media was grown in 37 °C incubator.

### 4.2.6. Plasmid isolation from *E. coli*

A single (transformed) *E. coli* colony was taken from an LB plate and was grown for up to 24 hours in 7-10 mL or 200-250 mL media containing relevant antibiotic. Subsequent mini- or midi-prep was performed using NucleoSpin Plasmid EasyPure Mini kit or NucleoBond Xtra Midi kit, respectively. Isolation of plasmid DNA using either kit was performed according to manufacturer's protocol.

### 4.2.7. Transformation of DNA into *S. cerevisiae*

*S. cerevisiae* cells were grown in YPAD medium at 37 °C overnight. The subsequent morning, a small aliquot was diluted to $OD_{600}$ 0.2 in fresh YPAD medium and grown until $OD_{600}$ ~1.0 (measured using Thermo Scientific GENESYS 20 spectrophotometer). Cells were spun down at 3000g for 5 minutes at 4 °C, washed in 10 mL 1x LiOAc buffer, spun down again at 3000g for 5 minutes at 4 °C, and finally resuspended in 0.5 mL 1x LiOAc buffer and kept on ice. In a clean Eppendorf tube, 100 µL of washed cells, 10 µg of purified PCR product and 10 µL of single stranded carrier DNA was combined and mixed by vortexing for 30 seconds. Mixture was incubated for 30 minutes at room-temperature. After incubation 43 µL of DMSO was added and mixed well. Cells were heat shocked for 15 minutes at 42 °C and subsequently rapidly cooled on ice for 3 minutes. Next, cells were spun down at 1000g for 3 minutes at 4 °C, supernatant was decanted and cells were washed in 100 µL milli-Q water and spun down again at 1000g for 3 minutes at 4°C. Supernatant was discarded once more, and cells were resuspended in 200 µL TE buffer before spreading on appropriate plates. Plates were incubated at 30 °C until colonies started forming.

### 4.2.8. Transformation of DNA into competent DH5α *E. coli*

Previously prepared chemically competent DH5α cells (NEB) were thawed on ice (cell stock maintained by Andrea Schmid). To the thawed cells, 10-100 ng of plasmid DNA was added and further incubated on ice for 10 minutes. Cells were heat shocked for 60 seconds at 42 °C, and immediately transferred back to ice for 5 minutes. To the cells, 250 µL LB media was added and subsequently transferred to an incubator at 1000 rpm for 45 minutes at 37 °C. Typically, a control experiment was performed which was treated identically, but to which no DNA was added. Cells were transferred to LB plates with appropriate antibiotic, evenly spread out over the media and placed in 37 °C incubator.

### 4.2.9. *BAR1* gene deletion in *S. cerevisiae*

Deletion of *BAR1* gene from yeast was performed in both S288C and W303 strains of the MATa type. The *BAR1* gene was replaced by the *HIS3* gene which acts as a selection marker. HPLC purified oligonucleotides with 50 bp homology upstream of the BAR1 start codon and downstream of the stop codon were designed (oFMP1235 and oFMP1236). The oligonucleotides furthermore have a 19 bp homology to the upstream and downstream region of the HIS3 gene on the pRS-plasmid. First, the selection marker was PCR amplified using OneTaq DNA polymerase. PCR product was separated using gel electrophoresis, isolated and purified using the Nucleospin Gel and PCR clean-up kit according to manufacturer's protocol. The purified PCR product should contain a 50 bp overhang to the up- and downstream regions of the *BAR1* gene, linked by the *HIS3* gene. Quantification of the yield and concentration of the PCR product was performed using DeNovix DS-11+ Spectrophotometer.

Transformation of 1 µg of PCR product was performed according to the protocol described earlier. Validation of BAR1 deletion was performed by PCR amplification using three sets of oligonucleotides. The first set validates the change in DNA length due to the replacement of *BAR1* by HIS3. The two oligonucleotides bind upstream and downstream of the original *BAR1* stop- and start codon (oFMP1237 and oFMP1238). The second set validates the insertion of the HIS3 gene by utilizing an oligonucleotide that has a homology with the centre of the HIS3 gene (oFMP1237 and oFMP1240). A third set is a negative control and should not get PCR amplified and utilizes an oligonucleotide which has homology with the centre of the BAR1 gene (oFMP1237 and oFMP1239).

### 4.2.10. Nuclei preparation from *S. cerevisiae*

Nuclei were largely prepared as described earlier but using a slightly modified version utilized in our lab (Almer and Hörz, 1986; Singh et al., 2021). Cells were grown overnight in the required volume of YPAD complete media to $OD_{600}$ 0.8 – 1.2. The optional addition of 1% formaldehyde was performed to cross-link cells for 20 minutes at room temperature whilst shaking at 130 rpm. Cross-linking was quenched by the addition of 250 mM of glycine for 5 minutes at room temperature whilst shaking at 130 rpm. Next, cells were spun down using the Heraeus Cryofuge 6000i at 4000 rpm for 30 minutes at 4 °C. Cell pellet was checked to confirm it was relatively tight, if not, cells were spun down an additional 15 minutes using the same conditions. After centrifugation, supernatant was carefully decanted and pellet was resuspended in 45 mL of cold milli-Q water. Resuspended pellet was transferred to pre-tared 50 mL falcons and spun down at 4000 rpm for 5 minutes at 4 °C using the ThermoFisher Scientific TX-1000 rotor. Supernatant was decanted and 'wet weight' of pellet was measured and noted. Per gram of wet weight, 2 volumes of preincubation solution was used to resuspend the pellet. Cells were incubated for 30 minutes at 30 °C whilst shaking at 130 rpm. Cells were subsequently pelleted at 4000 rpm for 5 minutes at 4 °C. Supernatant was carefully decanted in appropriate container and resuspend in 40 mL pre-chilled 4 °C 1 M sorbitol. Cells were pelleted again at 4000 rpm for 5 minutes at 4 °C. Supernatant was decanted in appropriate container and cells were resuspended in 5 volumes sorbitol-β-ME solution per gram wet weight. The $OD_{600}$ was measured from a 1:100 dilution in water and noted before the start of zymolyase addition. Per gram of wet weight, 100 µL of 2% zymolyase 100T was added to digest the cell wall. Cells were transferred for 20-25 minutes at 30 °C whilst shaking at 130 rpm. After 20-25 minutes $OD_{600}$ was measured and a decrease in absorbance of 80-90% was expected. The spheroplasts were pelleted at 4000 rpm for 8 minutes at 4 °C. Supernatant was decanted in appropriate container and spheroplasts were washed in 40 mL pre-chilled 4 °C 1 M sorbitol.

Spheroplasts were spun down at 4000 rpm for 8 minutes at 4 °C after which supernatant was decanted. Cleaned spheroplasts were now resuspended in 7 mL Ficoll buffer per gram wet weight and aliquotted into desired amounts (typically 0.5 gram). Finally, spheroplasts (further referred to as nuclei) were spun down using Beckman Coulter JA 20.1 rotor at 12.000g for 30 minutes at 4 °C. Supernatant was decanted and pellet was pre-frozen in EtOH/dry ice mixture before being stored at -80C.

### 4.2.11. Mix-n-match sample preparation

W303 and S288C strains were grown separately overnight in YPAD complete media to an $OD_{600}$ of 0.8 – 1.2. Equal number of cells from each strain were combined in a clean container. Subsequent preparation of nuclei was performed as described earlier.

### 4.2.12. Protein depletion by the anchor-away technique

Strains carrying the correct tags that allow the application of the anchor-away technique were grown overnight in YPAD complete media to an $OD_{600}$ of 0.2 – 0.3. For this thesis, we depleted INO80, using previously generated strains (yFMP333/334/334, Singh et al., 2021). A rapamycin stock solution was prepared by dissolving 1 mg/mL rapamycin in a vehicle solution (can be stored indefinitely). To the cells, the dissolved rapamycin was added to a final concentration of 1 µg/mL. As a control, the vehicle solution can be added to an aliquot of the cells. Cells were incubated for 120 minutes at 30 °C whilst shaking at 130 rpm. Subsequent preparation of nuclei was performed as described earlier.

### 4.2.13. Alpha-factor arrest of *S. cerevisiae*

Strain carrying a *bar1Δ* were grown overnight in YPAD complete media to an $OD_{600}$ of 0.2 – 0.3 to ensure a subsequent arrest. A non-arrested control was prepared by aliquoting 20% of the volume into a clean vessel. To the large volume, 50 ng/mL alpha factor was added. Both vessels were incubated for 60 minutes at 30 °C whilst shaking at 130 rpm. After 60 minutes, an additional 50 ng/mL alpha factor was added and cells continued to incubate for 30-60 minutes at 30 °C whilst shaking at 130 rpm. From each vessel, a 10 mL aliquot was taken to be utilized for light microscopy imaging and FACS analysis. The small aliquot was spun down at 4000 rpm for 5 minutes at 4 °C. Supernatant was discarded and pellet was resuspended in 70% EtOH, 50mM Tris-HCl pH 8.0 and stored at 4 °C. The arrested sample and control was mixed with 15 mM sodium azide and 2 mM EDTA and incubated for 3 minutes at room temperature whilst shaking at 130 rpm. Samples were subsequently crosslinked by addition of 1% formaldehyde for 20 minutes at room temperature whilst shaking at 130 rpm. Crosslinking was quenched by the addition of 250 mM glycine for 5 minutes. Subsequent preparation of nuclei was performed as described earlier.

### 4.2.14. Sample preparation for FACS analysis

Aliquots reserved during alpha-factor arrest of *S. cerevisiae* were placed on ice. 1 – $2x10^7$ cells were transferred to a clean Eppendorf tube and pelleted at 4000 rpm for 5 minutes at 4 °C. Pelleted cells were resuspended in 1 mL with 70% EtOH and 50 mM Tris-HCl pH 8.0. At this point, cells can be stored at 4 °C for several months. Before further use, cells were pelleted at 4000 rpm for 5 minutes at 4 °C and resuspended in 520 µL of RNase solution and incubated minimally four hours, but preferably overnight, at 750 rpm, 37 °C. Cells were next pelleted at 4000 rpm for 5 minutes at 4 °C. Supernatant was decanted, pellet was resuspended in 220 µL proteinase K solution (200µl 50mM Tris pH 8 + 20µl

Proteinase K 10mg/ml in 50% glycerol, 10mM Tris pH 7.5, 25mM $CaCl_2$) and incubated at 750 rpm for 30 minutes at 50 °C. Cells were pelleted by centrifugation at 4000 rpm for 5 minutes at 4 °C. Lastly, resuspended cells were briefly sonicated using Bioruptor Pico for 1 cycle of 15 seconds on and 30 seconds off. To 15 µL of the sonicated samples, 285 µL SYTOX solution was added. Samples were subsequently handed over to the BMC FACS facility for analysis where they analysed the samples using the BD FACS Canto™ II Cell Analyzer.

### 4.2.15. Yeast growth conditions for histone depletion

For histone depletion (HD) experiments, previously generated strains were utilized which have both genic copies of histone H3 and H4 deleted and are transformed with a plasmid containing a galactose inducible H3 and H4 copy (Singh et al., 2021). Cells were grown to $OD_{600}$ 0.8 – 1.2 in YNB synthetic media supplemented with 2% galactose. The cells were pelleted at 3000g for 5 minutes at 20 °C using Beckman Coulter JLA 8.1000 rotor. Pellets were resuspended in pre-warmed to 20-25 °C YNB synthetic media supplemented with 2% glucose. Resuspension volume was adjusted so $OD_{600}$ was approximately 0.5 after 180 minute incubation at 30 °C whilst shaking at 130 rpm. Subsequent preparation of nuclei was performed as described earlier.

### 4.2.16. Methylation footprinting using methyltransferases

Prepared nuclei were thawed on ice for 20 minutes before proceeding with subsequent steps. Nuclei were resuspended in 10 mL 1x M.SssI buffer per gram of nuclei by vortexing. The typical wet weight for a methylation reaction with two timepoints was 0.5g. Resuspended nuclei were incubated on ice and subsequently spun down at 4000 rpm for 6 minutes at 4 °C using Heraus Multifuge X3R. Supernatant was discarded and pellet was resuspended in 1 mL 1x M.SssI buffer per gram of nuclei and placed on ice. Empty Slide-A-Lyzer™ MINI Dialysis Device, 3.5K MWCO, 0.5 mL tubes were pre-hydrated (using the associated 15 mL tube) in 14 mL of 1x M.SssI buffer and placed to the side to reach room temperature. Nuclei DNA concentration was approximated by 1:20 and 1:40 dilution in 1x M.SssI buffer and measured using ThermoFisher Scientific Qubit 3.0 fluorometer. For each methylation reaction the volume that contains approximately 25 µg of DNA was utilized and transferred to a clean LoBind tube. To this volume 20µl (0.5 mg/mL final conc.) RNase A was added, 10 µL (200 units, 20.000U/ml, NEB) of M.SssI methyltransferase and cold 1x M.SssI buffer to 397.5 µL. Each reaction was gently mixed by flicking the tube and placed on ice for 5 minutes. To the 14 mL in which the Slide-A-Lyzer™ tubes were hydrated, 87.5 µL SAM (200 µM final conc., 32mM stock from NEB) was added. Next, the methylation reaction was added to the 0.5 mL pre-hydrated dialysis device. To each dialysis device containing the methylation reaction a micro stir bar (VWR 5x2 mm) was added and the reaction was stirred for 2 minutes to reach room temperature. To ensure stirring, the entire tube and dialysis device was placed in close proximity to an IKA REO magnetic stirrer. Stirring was set to the lowest speed to keep nuclei in suspension and may vary depending on proximity and strength of magnets in stirring device. To start the methylation, 2.5 µL (200 µM final conc.) of SAM was added to the methylation reaction (400 µL total) and incubated for 30-45 minutes at room temperature whilst slowly stirring. An additional 200 µM SAM and 50 units of M.SssI were supplemented (typically after extracting an aliquot) after 30 and 60 minute incubation. 190 µL aliquots were taken at halfway (typically 45 – 60 minutes) and at end of the timecourse (typically 90 – 120 minutes) and transferred to a clean Eppendorf tube. Note, more timepoints, including one before adding SAM to start the reaction, can be taken. However, the described conditions are optimized for a maximum of three timepoints (which would ensure sufficient purified gDNA in downstream applications), upscaling may be necessary when requiring more timepoints or higher downstream gDNA yields. To the timepoint aliquots 15 mL (2mM final concentration, typically ~10-fold over SAM concentration) of SAH (26 mM stock, Merck) was added and tubes were placed at 65 °C for 15 minutes to quench the methylation reaction. After this, samples were placed at 25 °C with mild shaking to continue RNase A activity until all aliquots and timepoints were taken. To all samples, 25 µL (final conc. 1x) STOP buffer (10x stock) and 25 µL (final conc. 1 mg/mL) proteinase K solution was added to quench any enzymatic activity and deproteinize the DNA for a total volume of 255 µL. Samples were incubated at 37 °C for 120 minutes. When formaldehyde crosslinked nuclei were utilized, incubation was performed at 60 °C overnight. Isolation of DNA is described separately.

### 4.2.17. gDNA integrity control

The integrity of genomic DNA from prepared nuclei can be validated before performing a methylation footprint experiment. This experiment is identical to the methylation footprinting experiment, but does not utilize expensive enzymes or consumables. Briefly, nuclei were washed, prepared and measured as described earlier. In a clean Eppendorf tube, 25 µg of DNA (based on approximate measurements by Qubit) was combined with 200 µM of SAM, 0.5 µg/µL RNase A and filled to 400 µL with 1x M.SssI buffer. Samples were incubated at room temperature and supplemented with 200 µM SAM after 30 and 60 minute incubation. Aliquots were taken at appropriate times and transferred to a clean Eppendorf tube where 2 mM SAH was added and placed at 65 °C for 15 minutes to quench the methylation reaction. After all timepoints had been collected, 1x STOP buffer and 1 µg/µL proteinase K was added. Samples were incubated at 37 °C for 2 hours, or when formaldehyde crosslinked samples were used, overnight at 60 °C. Subsequent isolation of DNA is described separately.

### 4.2.18. Genomic DNA isolation

Genomic DNA isolation as described here aims to retain the DNA integrity as much as possible. Steps were performed using wide bore pipette tips when possible and vortexing was kept to a minimum. Deproteinized DNA was combined with 1 volume of phenol:chloroform:isoamyl and vortexed for 3 seconds. Samples were carefully transferred to MaxTract High Density tubes and spun at 16.000g for 5 minutes at room temperature. Upper layer was transferred to clean DNA LoBind tubes and 0.2 M NaCl was added with 2.5 volumes of 100% EtOH next. Samples were placed on ice for 30 minutes. Next, samples were spun down at 15.000g for 30 minutes at 4 °C. Supernatant was decanted and 1 mL 70% EtOH was used for washing without disturbing the pellet. Sample was spun again at 15.000g for 10 minutes at 4 °C after which supernatant was discarded and tube was placed upside down on absorbent material for 1 – 2 minutes. Excess EtOH was removed by placing opened tubes at 37 °C for 5 – 10 minutes. Pelleted DNA was resuspended in 50 µL TE buffer by gently flicking the closed tube. Sample was placed at 4 °C overnight to allow optimal resuspension before utilizing the DNA.

### 4.2.19. Nanopore library preparation and sequencing

Methylated and isolated DNA was subsequently utilized for Oxford Nanopore Technologies library preparation using the SQK-LSK109 ligation sequencing kit. For multiplexing up to 24 samples, native barcode expansion kits (EXP-NBD104/114) were used. The ligation sequencing kit in combination with the native barcode expansion kit was utilized according the manufacturer’s protocol. Finalized library was loaded onto MinION R9.4.1 flow cells. For detailed settings and subsequent analysis see ‘Bioinformatical methods’.

### 4.2.20. Salt gradient dialysis assembly and quality controls of *in vitro* arrays

Assembly of *in vitro* arrays was performed according to previously published methods (Ludwigsen et al., 2018). Please refer to the published data for a detailed description of the method. Briefly, a test assembly was performed first to determine optimal molar DNA:octamer ratio. The molar ratio of octamer to array is varied between 0.7 and 1.5 by keeping the concentration of DNA stable. The total volume for each test assembly was 50 µL in 2M NacL and 10 mM Tris-HCl pH 7.6. Test assemblies were transferred to dialysis tubes and placed into a Styrofoam float to keep suspended in 2 L of high salt buffer. Over a period of 24 hours, 1.8 L of no salt buffer exchanged the high salt buffer, starting off slow and increasing the exchange rate after 6 hours. After the 24 hour period, dialysis tubes were transferred to a 1 L beaker of low salt buffer and dialysed for another 2-3 hours.

Arrays are next purified by $MgCl_2$ purification. Contents of the dialysis tube are spun at max speed for 15 minutes at 4 °C to remove any precipitates. A 200 ng aliquot (input fraction) was transferred to a clean Eppendorf tube for later analysis. 1 volume of precipitation buffer was added to the remaining sample and mixed gently. Samples were incubated on ice for 15 minutes and were next spun down at max speed for 15 minutes at 4 °C. Supernatant was transferred to a clean Eppendorf tube (supernatant fraction) and pellet resuspended in 1x TE buffer such that the final concentration is 200 – 300 ng (pellet fraction). All fractions were stored at 4 °C.

Three quality controls were performed to validate optimal assembly by salt gradient dialysis. (1) A native gel was prepared utilizing 0.7% low melt agarose and 0.2x TB buffer without ethidium bromide. Equal amounts of all fractions was mixed with 1x blue gel loading buffer and separated by gel electrophoresis until blue front nearly ran off the gel. Gel was finally stained in 0.2x TB buffer containing 5 µg/mL ethidium bromide for 30 minutes and subsequently destained in 0.2x TB buffer. (2) Digestion of 200 ng from all fractions with NotI restriction enzyme was performed. To the fractions, 2.3 µL of NotI enzyme was added in a total volume of 15 uL of EX50 buffer. Samples were incubated for 180 minutes at 26 °C. To each sample, orange G gel loading buffer was added and samples were ran similar to described in quality control (1) except for utilizing a 1.1% native gel. (3) Lastly, 250 ng per sample was digested using 2 µL BsiWI restriction enzyme in 20 µL 1x remodeling buffer for 60 minutes at 26 °C. Samples were deproteinized for 180 minutes at 37 °C by addition of 0.4% SDS, 20 mM EDTA and 0.6 mg/mL proteinase K in a total volume of 30 uL. DNA was precipitated by addition of 1/10 volume 3M NaOAC pH 5.2, and 2.5 volumes of 100% EtOH. Samples were placed on ice for 60 minutes and subsequently spun down at max speed for 15 minutes at 4 °C. Supernatant was discarded and pellet was resuspended in 8ul 10 mM Tris-HCl pH 8.0. 1x purple gel loading buffer was added to the samples and subsequently ran on 1% universal agarose in 0.5x TBE buffer containing 0.5 µg/mL ethidium bromide. All gels were visualized using Peqlab Vilber Gel Documentation system.

After performing the test assembly and quality controls, a preparative assembly was performed using the optimal conditions as identified by the test assembly. Procedures were identical to the test assembly, but were scaled up to produce large(r) volumes.

### 4.2.21. Remodeling of in *vitro arrays*

Remodeling of the *in vitro* array was performed as described previously (Ludwigsen et al., 2018). Briefly, SGD 12.5 ng/µL assembled arrays was added to 1x remodeling buffer, 1x regeneration system, 1mM ATP-$Mg^{2+}$, 0.2 g/L BSA on ice in a total volume of 60 µL. An appropriate amount of quench tubes containing 5 µL of 10U/uL apyrase were prepared and set aside at 4 °C. Reaction mixture was equilibrated at 26 °C for 2 minutes before adding 0.5 µM purified full-length drosophila ISWI and briefly flicking the tube to mix. After desired amount of timepoints, aliquots were taken and directly transferred to the quench tubes.

To the samples, 1mg/mL proteinase K (final conc.) was added and incubated at 55 °C for 3 hours or overnight at 37 °C. DNA was purified by adding 1 volume of phenol:chloroform:isoamyl vortexing for 10 seconds and spun down at 16.000g for 5 minutes at room temperature. Optionally, MaxTract tube were used. Upper layer was transferred to clean Eppendorf tube, 0.2 M NaCl and 2.5 volumes of 100% EtOH were added to precipitate the DNA. Samples were pelleted at max speed for 20 minutes at 4 °C. Supernatant was discarded and pellet was resuspended in 10-20 µl TE buffer.

### 4.2.22. Methylation of remodeled *in vitro* arrays

Methylation of *in vitro* arrays is largely similar to the protocol for *in vivo* chromatin. The remodeling reaction was performed in 20 µL instead of 60 µL, but quenched in 5 µL apyrase as described above. To each quenched aliquot 320 µM SAM and 5 µL NEB2 buffer was added. Volume was adjusted to 39 µL with milli-Q water and 1 uL M.SssI (4U/µL) was added to start methylation. Sample was incubated for 120 minutes at 26 °C. Methylation reaction was quenched by the addition of 3.2 mM SAH and placed at 65 °C for 15 minutes. An optional HpaII digestion was added to validate full methylation of accessible DNA. To the 45 µL of sample, 4 ul of 10x Cutsmart buffer was added and 1 µL HpaII. Restriction digestion was performed for 60 minutes at 26 °C. Arrays were deproteinized by the addition of 0.4% SDS, 20 mM EDTA and 1 mg/mL Proteinase K. DNA was precitipated by addition of 0.3 M NaOAC and 3 volumes of 100% EtOH. Samples were spun down max speed for 15 minutes at 4 °C. Supernatant was discarded and sample was resuspended in 10 µL TE buffer.

### 4.2.23. 5-mC ELISA quantification

Quantification of methylation for array or genomic DNA was performed by Zymo Research 5-mC DNA ELISA kit according to manufacturer's protocol. Absorbance was measured using Biotek Powerwave HT microplate reader. Raw values were processed using R.

### 4.2.24. Purification of full-length drosophila ISWI

Purification of the full-length ISWI protein was performed as previously described (Forné et al., 2012; Harrer et al., 2018; Ludwigsen et al., 2017). Briefly, the full-length drosophila ISWI was N-terminally tagged with a $HIS_6$-TEV tag and was expressed in *E. coli* BL21-Gold (DE3) cells. Cells were cultured in LB media, supplemented with ampicillin at 37 °C at 130 rpm to an $OD_{600}$ 0.6 – 0.8. To the cells, 0.2 M of IPTG was added and cells were transferred to 18 °C incubator for overnight induction. Cells were harvested, resuspended and lysed using an LM10 microfluidizer. Affinity purification was performed utilizing a nickel resin (GE healthcare HisTrap 1 mL) using a gradient of HisA to HisB buffer. Next, the sample was transferred to a 6-8K MWCO bag and the $His_6$-TEV tag was cleaved overnight at 4 °C by

TEV protease (1:80 (w/v)) in dialysis buffer. Protein was further purified by passing the solution over a pre-equilibrated (HisA:HisB buffer 10:1) $Ni^{2+}$-affinity column. Final purification was performed by Mono S 5/50 GL ion exchange column using a gradient MonoS A:MonoS B buffer and subsequent size exclusion by Superdex200 INCREASE 10/300 GL or HiLoad Superdex200 16/600 column using GF buffer. Purified proteins were concentrated to 3 – 10 mg/mL using Amicon Ultrafiltration (30.000/50.000 MWCO), aliquotted into appropriate volumes and stored at -80 °C.

### 4.2.25. Purification of histones and octamer assembly

Expression, purification and assembly of Drosophila histone H2A, H2B, H3 and H4 was performed as described previously (Klinker et al., 2014). Briefly, BL21-Gold (DE3) cells were transformed with the selected expression plasmid and grown to $OD_{600}$ 0.6 – 0.8 in LB media, supplemented with ampicillin at 37 °C at 130 rpm. Induction was induced by the addition of 1mM IPTG for 2 hours. Cells were harvested and stored at -80 °C. Pellet was resuspended in 1x SAU buffer, 200 mM NaCl 5 mM β-Mercaptoethanol, 7 M Urea and protease inhibitors (1 mM PMSF, 1 mg/mL Aprotinin, 1 mg/mL Leupeptin and 0.7 mg/mL Pepstatin) in a total volume adjusted 35 mL with milli-Q water. Cells were lysed using LM10 microfluidizer and subsequently spun down at 19.000 rpm for 30 minutes at 4 °C. Supernatant was extracted and further clarified using 0.45 µM filters. Histones were further purified using a cation/anion exchange which utilizes a HiTrap Q HP column stacked on top of an SP column. After removal of the HiTrap Q HP column, histones are eluted from the SP column in 1 mL fractions using an NaCl gradient in 1x SAU buffer. Pooled fractions containing histones were dialysed using Spectra/por 3.5kD MWCO in cold water overnight at 4 °C and stored at -80 °C.

Approximately 1 mg of the individual histones was first lyophilized and subsequently resolubilized in 1/1.2 mL unfolding buffer. Histones were combined in a Spectra/Por 6 – 8 kD MWCO dialysis tube in a 1.2:1.2:1:1 ratio for H2A, H2B, H3 and H4, respectively. Stoichiometry was checked by performing a standard SDS-PAGE gel on a small aliquot. The combined histones were dialysed three times into 2 L refolding buffer of which the first two times were 2 hours and the last overnight. After dialysis, samples were briefly spun for max speed for 5 minutes at 4 °C to remove precipitates. Supernatant was loaded onto a Superdex 200 HiLoad 16/600 120 mL size exclusion chromatography column. Individual fractions were analysed using standard SDS-PAGE gel analysis to determine which fractions had the octamers. These fractions were pooled and concentrated using Amicon Ultrafiltration 30.000 MWCO filters. Octamers were stored at -80 °C.

## 4.3. Bioinformatical methods

### 4.3.1. Running a Nanopore sequencing experiment

Prior to loading an R9.4.1 MinION flow cell it was inserted into a MinION sequencing device (MIN-101B) and connected to a Windows 10 computer running the latest MinKNOW software (software downloaded from: https://community.nanoporetech.com/downloads). A standard 'flow cell check' was performed to validate the quality of the flow cell. A typical MinION flow cell should have ≥800 nanopores available for direct sequencing.

After the R9.4.1 MinION flow cell was loaded according to manufacturer's protocol, it was inserted into a MinION sequencing device and connected to a computer running the latest MinKNOW software. Sequencing was initiated using standard parameters. Local basecalling was turned off and output files set to FAST5 only. Data was saved on a local high capacity (1TB) SSD drive. Raw FAST5 files were used for further processing.

### 4.3.2. Basecalling, demultiplexing using Guppy software

Processing of raw FAST5 files was performed using Guppy software after sequencing was finalized and was performed in two distinct steps:

First, basecalling was performed using the latest GPU version of Guppy software installed on a local machine (Guppy software downloaded from: https://community.nanoporetech.com/downloads). This machine ran an up-to-date Ubuntu 16.04 LTS install and was fitted with a NVIDIA GTX 1080 for GPU accelerated basecalling. For basecalling in recent versions of Guppy a high accuracy configuration was used (`guppy_basecaller --config dna_r9.4.1_450bps_hac.cfg`).

Secondly, after basecalling was finalized, samples were demultiplexed using the Guppy software. The utilized barcoding kits defined and default configuration files were used (`guppy_barcoder --config configuration.cfg --barcode_kits “EXP-NBD104 EXP-NBD114”).`

Please refer to the latest Guppy protocol for more settings and features (latest version: https://community.nanoporetech.com/protocols/Guppy-protocol/v/gpb_2003_v1_revv_14dec2018)

### 4.3.3. Alignment and detection of modified bases

Basecalled and demultiplexed files were transferred to the high performance computing (HPC) system ran by Dr. Tobias Straub at the biomedical center (BMC). FASTQ files within each individual barcode folder were concatenated into a single file, labelled with the corresponding barcode. Using Conda, a virtual Python 2.7 environment was set up which is used throughout the following steps. Within this environment the following software was installed: Pysam (https://anaconda.org/bioconda/pysam), Numpy (https://anaconda.org/anaconda/numpy) and Pandoc (https://anaconda.org/conda-forge/pandoc). In order to perform the alignment Minimap2 (version 2.10, https://github.com/lh3/minimap2) was used using settings tuned for nanopore data (`minimap2 -ax map-ont`). An appropriate reference genome was used to align the reads against. Typically we used the SacCer3 R64-1-1 (https://www.ncbi.nlm.nih.gov/assembly/GCF_000146045.2/). To this we manually added a set of unique ‘chromosomes’ which corresponded to the spike-in fragment(s). Filtering for reads was performed using Samtools (version 1.9, https://github.com/samtools/samtools). For a typical sequencing run we filtered for reads with quality above 30 (`samtools view -F 2052 -q 30 -b -S`). Samtools was further used for sorting and indexing. Next, two custom python scripts written by Dr. Michael Wolff were used to isolate read information from the BAM and FASTQ files from the individual barcodes (get_read_infos_from_bam.py and get_read_infos_from_fastq.py, respectively). The resulting files were concatenated into a single file for easy access of all information later.

Detection of modified bases was performed using Nanopolish software (version 11.1, https://github.com/jts/nanopolish). I modified a script (‘Nanopolish_loop_only.sh’) written by Dr. Michael Wolff to parallelize the process on the HPC system. In essence, for each barcode of the sequencing run a separate instance of Nanopolish was initiated. Within each instance the Nanopolish software first performed an indexing step on the raw FAST5 files and the concatenated FASTQ file (`nanopolish index -d $FAST5_files $FASTQ_file`). Next, the software was utilized to detect methylated CpG sites using the built-in call-methylation function (`nanopolish call-methylation`). Finally, two unmodified python scripts were used to isolate the methylation read statistics and site statistics from the nanopolish output file (‘calc_read_stats.py‘ and ‘calc_site_stats.py’, respectively).

### 4.3.4. Basic analysis of sequencing run using R

The analysis performed next was largely written by Dr. Michael Wolff and Dr. Maryam Khatami, but was modified by me in order to run efficiently on the HPC. The full script is saved as ‘nanopore_analysis.R’. In short, this analysis results in the generation of several files which are key for most downstream analyses. The functions within this analysis are sourced from other locations and were also written by Dr. Michael Wolff and Dr. Maryam Khatami, but in some instances were slightly modified by me to improve speed. Firstly, the analysis produces a methylation data (meth_data) file for each chromosome (including the unique spike-in ‘chromosomes’) which informs on the methylation location, sequence, log likelihood ratio etc. of each individually called CpG site/group. Secondly, a single file for read statistics (read_stats) is generated which informs on the unique read name, length, total CpG sites, location etc. Thirdly, a site statistics (site_stats) file is generated which informs on the state of an CpG site on average by compiling the total coverage of a site, the percentage methylated, and the ambiguity rate.

The analysis was performed on the HPC in order to generate the files (and more) described above but can be performed locally when hardware allows. After a completed analysis, the generated files were used to create a HTML file to provide a quick overview of many basic observables of the sequencing run. Note that changes in the reference genome and/or spike in may require modifications in the original analysis script to generate data correctly.

### 4.3.5. Generation of aligned composite plots

Typical alignment of data was performed according to the +1 nucleosome location based on H3Q85C chemical mapping dataset (Chereji et al., 2018). Briefly, site_stats data was used to assign the locus name to the CpG sites within the defined boundaries of a +1 nucleosome (typically 1000 bp upstream and 2000 bp downstream of the +1 nucleosome for each locus). Next, the original coordinates of the CpG sites were redefined relative to the +1 nucleosome of the locus. Thus, the coordinate of a CpG site exactly at the +1 nucleosome dyad set to zero (0). Data was typically saved as “av_gene_ind_samples.RData” and can be found as an output from the basic analysis described above. Data was generated using the “calc_average_gene_ind_samples” function. A similar function was created (calc_average_gene_ind_samples_GRFs) from the existing script to generate composite plots for general regulatory factors (GRFs). Instead of chemical mapping data, position of GRFs was acquired from SLIM-ChIP data (Gutin et al., 2018).

### 4.3.6. Generation of single-molecule plots

Single molecule plots were generated in a two-step process. First, meth_data was utilized to align each individual CpG site/group to the +1 nucleosome in a similar fashion as done when creating composite plots. Importantly, each CpG site was kept uniquely identifiable to the original read and locus. The script to produce this dataset is referred to as “align_meth_data_Np1”. Second, individual reads were generated using the “plot_individual_reads” function.

### 4.3.7. Determination of +1 nucleosome localization

The localization of a +1 nucleosome was determined by the state of CpG site(s) within a window around the nucleosome dyad (based on chemical mapping data). Briefly, “calc_n1_overlaps” function was utilized to  perform several steps. First, the CpG site(s) within a defined window (typically 20 bp) around the +1 nucleosome dyad were assigned to that specific locus. To this end the findOverlapPairs function in the GenomicRanges package of Bioconductor was used (https://bioconductor.org/packages/release/bioc/html/GenomicRanges.html). Second, the mean hit state of the site(s) within this window was calculated. If all sites were unmethylated, they were assigned a -1 state, if methylated, a 1 state. If sites within the window were of variable methylation state was between -1 and 1. Third, the mean hit state of the +1 nucleosome was assigned back to the CpG sites that fell within 500 bp upstream of the gene start and the gene end site. This led to the assignment of a locus name to each CpG site per read and the mean hit state of the +1 nucleosome for that particular read. Lastly, the dataset was split into two based on the mean hit state of the +1 nucleosome for quick access in subsequent analyses. In this application, reads with an ambiguous mean hit state at the +1 nucleosome (i.e. not -1 or 1) were filtered out. The original meth_data dataset can be used to access these reads and sites.

### 4.3.8. Computational phasing and nucleosome dyad calculations of single-molecule reads

Alignment to the +1 nucleosome was performed on a per-read basis using the “re-alignment_NpX_cluster” script. Meth_data was loaded and optionally randomly down sampled to increase output speed. Data was aligned to Np1 as described previously for single-molecule plots. Nucleosome footprints (referred to as MUM-gaps in script) were calculated using a parallelizable “calc_ss_gap” script written by Matthias Hanke or a slower (but easier to run) “calc_mm_gap” script written by Dr. Maryam Khatami. Similar to aligning CpG sites, nucleosome footprints were aligned to the +1 nucleosome of each gene. The coordinate of the dyad for each nucleosome footprint was calculated using the calc_dyad_position. Nucleosome footprints were defined as 147 bp with a 10 bp overlay. Aligned meth_data was subsequently re-aligned (i.e. computationally phased) to the desired nucleosome using realign_meth_data_NpX. Here NpX could be dynamically altered to the desired nucleosome to which the read should be phased. The first nucleosome of an array was defined by the dyad position closest to the +1 nucleosome dyad based on chemical mapping, but no further upstream than 147 bp. Dyad coordinates for each nucleosome footprint was corrected for the shift due to computational phasing. Single molecule plots were generated using ”plot_individual_reads_with_nucl” script which takes the computationally phased meth_data and the dyad coordinates. Of note, dyad coordinates could be generated for all types of data providing the nucleosome footprints (i.e. MUM-gaps) were calculated.

### 4.3.9. rDNA analyses

The methylation state of either of the two RDN1 copies was individually calculated using the “rDNA_meth_calc” script (typically referred to as ‘left’ and ‘right’ locus). The script was designed to retain the information of the read name so subsequent analyses could match two RDN1 copies of the same read together. Based on the combined methylation state of the two neighboring RDN1 copies, single-molecule plots were ordered and generated. The "rDNA_test" script, written by Michael Wolff, was used to test the transcriptional independency between two neighboring RDN1 loci. Briefly, the probability of independence for locus 1 or 2 (i.e. left and right, resp.) was compared against the independent state of the alternative locus as such:

$$r = \frac{P(locus\ 2\ active \mid locus\ 1\ active)}{P(locus\ 2\ active)}$$

The script was further used to generate contingency tables of the combined methylation state of both rDNA loci on a single read. These contingency tables were used to perform a Fisher's exact test to statistically test the null hypothesis.

### 4.3.10. Separation of yeast strain variants from MEFSIMO-seq data

Raw MEFSIMO-seq FAST5 data was basecalled and demultiplexed as described earlier. Mapping of reads was performed using same Minimap2 settings, except to a reference genome which combined the S288C and W303 genome and a reduced. Samtools was used to filter reads with quality below 1 (`samtools -q 1)` to allow for further read filtering later based on SNP content. Detection of CpG using Nanopolish was identical to described earlier. Next, bamtools (version 2.5.1, https://github.com/pezmaster31/bamtools) was used to separate BAM files per reference chromosome (`bamtools split -in *.bam -reference`). Samtools was used to merge the BAM files originating from each reference genome back together into a single BAM file. Bedtools (version 2.28.0, https://github.com/arq5x/bedtools2) was used to convert BAM files to FASTQ files (`bedtools bamtofastq`). The FASTQ files of each genome were used to map to a single SacCer3 reference genome to have a direct comparison between locations and SNPs/indels using minimap2 as described previously. The data was subsequently processed as if it were a typical sequencing run as described earlier. The individual data generated for the S288C and W303 reference genomes was processed using the basic analysis of sequencing runs.

Comparison of data was performed in R. *In silico* data was used to determine false-positive rates by calculating the total amount of reads mapped to the correct genome. For further downstream analyses, such as the determination of global occupancy between samples, reads were selected with contain ≥X of SNPs per read. To this end we utilize the location of all SNPs between S288C and W303 genomes as identified by a previous publication (Matheson et al., 2017). We performed an analysis on all reads associated to the S288C and W303 genome using the GenomicRanges R package and the FindOverlapPairs function. This allowed us to detect and quantify the reads that contain a SNP site on a per genome basis. Next, an arbitrary cutoff for SNP sites per read can be used to filter reads (typically ≥3 SNPs per read were required). The read information obtained after filtering can be further used to isolate CpG site- and methylation information from other datasets acquired during sequencing.

### 4.3.11. Inhomogeneity score calculations

The calculations to determine the inhomogeneity score were defined by Matthias Hanke and Dr. Michael Wolff. A detailed description can be found in the master thesis of Matthias Hanke. Briefly, the inhomogeneity score calculations were created to reduce the influence of nucleosome occupancy on the number variance of a read. First, each nucleotide was given the hit state of the closest, non-ambiguous, CpG site. Regions of interest with an arbitrarily defined 'bad resolution ratio' (BRR) were discarded. The BRR for each region was calculated in two steps: (1) calculating the sum of distances between non-ambiguous CpG sites which are separated by a certain threshold (typically 150 bp) and (2) dividing it by the length of the region of interest. A typical BRR threshold cutoff of 0.25 was utilized which essentially removes reads where 25% of distances between CpG sites is longer than the threshold. Next, the number variance was determined by utilizing a sliding window of 150 bp in 1 bp increments over an area of interest (e.g. a gene body). The variance between each window over the area is what defines the number variance. The inhomogeneity score was finally determined by taking the number variance of a read and dividing this by the maximum variance for that same window whilst considering the read length and occupancy:

$$inhomogeneity\ score = \frac{number\ variance_{read}}{\max\,(number\ variance)_{read\ length,\ \ occupancy}}$$

The scripts used to calculate these inhomogeneity scores can be found in “gbr_av_comparison” and utilized intermediate datasets generated by Dr. Michael Wolff using the “multipleanalyses” script. Filtering of reads was performed by applying a ‘bad resolution ratio’ for each. Each region of interest (e.g. a gene body) was analyzed for the presence of, and distance between CpG sites within a 150 bp sliding window.

# 5. BIBLIOGRAPHY

Abdulhay, N.J., McNally, C.P., Hsieh, L.J., Kasinathan, S., Keith, A., Estes, L.S., Karimzadeh, M., Underwood, J.G., Goodarzi, H., Narlikar, G.J., et al. (2020). Massively multiplex single-molecule oligonucleosome footprinting. Elife *9*, 1–23.

Alabert, C., and Groth, A. (2012). Chromatin replication and epigenome maintenance. Nat. Rev. Mol. Cell Biol. *13*, 153–167.

Albert, I., Mavrich, T.N., Tomsho, L.P., Qi, J., Zanton, S.J., Schuster, S.C., and Pugh, B.F. (2007). Translational and rotational settings of H2A.Z nucleosomes across the Saccharomyces cerevisiae genome. Nature *446*, 572–576.

Almer, A., and Hörz, W. (1986). Nuclease hypersensitive regions with adjacent positioned nucleosomes mark the gene boundaries of the PHO5/PHO3 locus in yeast. EMBO J. *5*, 2681–2687.

Anderson, J.D., and Widom, J. (2001). Poly(dA-dT) Promoter Elements Increase the Equilibrium Accessibility of Nucleosomal DNA Target Sites. Mol. Cell. Biol. *21*, 3830–3839.

Arents, G., and Moudrianakis, E.N. (1995). The histone fold: A ubiquitous architectural motif utilized in DNA compaction and protein dimerization. Proc. Natl. Acad. Sci. U. S. A. *92*, 11170–11174.

Axel, R. (1975). Cleavage of DNA in Nuclei and Chromatin with Staphylococcal Nuclease. Biochemistry *14*, 2921–2925.

Azmi, I.F., Watanabe, S., Maloney, M.F., Kang, S., Belsky, J.A., MacAlpine, D.M., Peterson, C.L., and Bell, S.P. (2017). Nucleosomes influence multiple steps during replication initiation. Elife *6*, 1–23.

Baldi, S., Krebs, S., Blum, H., and Becker, P.B. (2018). Genome-wide measurement of local nucleosome array regularity and spacing by nanopore sequencing. Nat. Struct. Mol. Biol. *25*.

Barisic, D., Stadler, M.B., Iurlaro, M., and Schübeler, D. (2019). Mammalian ISWI and SWI/SNF selectively mediate binding of distinct transcription factors. Nature.

Barrass, J.D., Reid, J.E.A., Huang, Y., Hector, R.D., Sanguinetti, G., Beggs, J.D., and Granneman, S. (2015). Transcriptome-wide RNA processing kinetics revealed using extremely short 4tU labeling. Genome Biol. *16*, 1–17.

Belmont, A.S., and Bruce, K. (1994). Visualization of G1 chromosomes: A folded, twisted, supercoiled chromonema model of interphase chromatid structure. J. Cell Biol. *127*, 287–302.

Belmont, A.S., Sedat, J.W., and Agard, D.A. (1987). A three-dimensional approach to mitotic chromosome structure: evidence for a complex hierarchical organization. J. Cell Biol. *105*, 77–92.

Brahma, S., Udugama, M.I., Kim, J., Hada, A., Bhardwaj, S.K., Hailu, S.G., Lee, T.-H., and Bartholomew, B. (2017). INO80 exchanges H2A.Z for H2A by translocating on DNA proximal to histone dimers. Nat. Commun. *8*, 15616.

Branton, D., Deamer, D.W., Marziali, A., Bayley, H., Benner, S.A., Butler, T., Di Ventra, M., Garaj, S., Hibbs, A., Huang, X., et al. (2008). The potential and challenges of nanopore sequencing. Nat. Biotechnol. *26*, 1146–1153.

Buenrostro, J.D., Wu, B., Litzenburger, U.M., Ruff, D., Gonzales, M.L., Snyder, M.P., Chang, H.Y., and Greenleaf, W.J. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. Nature *523*, 486–490.

Cairns, B.R., Lorch, Y., Li, Y., Zhang, M., Lacomis, L., Erdjument-Bromage, H., Tempst, P., Du, J., Laurent, B., and Kornberg, R.D. (1996). RSC, an essential, abundant chromatin-remodeling complex. Cell *87*, 1249–1260.

Cam, H.P., and Whitehall, S. (2016). Analysis of heterochromatin in schizosaccharomyces pombe. Cold Spring Harb. Protoc. *2016*, 920–927.

Carter, J.M., and Hussain, S. (2017). Robust long-read native DNA sequencing using the ONT CsgG Nanopore system. Wellcome Open Res. *2*, 1–11.

Chereji, R. V., and Clark, D.J. (2018). Major Determinants of Nucleosome Positioning. Biophys. J. *114*, 2279–2289.

Chereji, R. V., Ramachandran, S., Bryson, T.D., and Henikoff, S. (2018). Precise genome-wide mapping of single nucleosomes and linkers in vivo. Genome Biol. *19*, 19.

Cherf, G.M., Lieberman, K.R., Rashid, H., Lam, C.E., Karplus, K., and Akeson, M. (2012). Automated forward and reverse ratcheting of DNA in a nanopore at 5-Å precision. Nat. Biotechnol. *30*, 344–348.

Chua, E.Y.D., Vasudevan, D., Davey, G.E., Wu, B., and Davey, C.A. (2012). The mechanics behind DNA sequence-dependent properties of the nucleosome. Nucleic Acids Res. *40*, 6338–6352.

Chung, H.R., Dunkel, I., Heise, F., Linke, C., Krobitsch, S., Ehrenhofer-Murray, A.E., Sperling, S.R., and Vingron, M. (2010). The effect of micrococcal nuclease digestion on nucleosome positioning data. PLoS One *5*.

Clapier, C.R., and Cairns, B.R. (2009). The Biology of Chromatin Remodeling Complexes. Annu. Rev. Biochem. *78*, 273–304.

Clapier, C.R., and Cairns, B.R. (2012). Regulation of ISWI involves inhibitory modules antagonized by nucleosomal epitopes. Nature *492*, 280–284.

Clapier, C.R., Iwasa, J., Cairns, B.R., and Peterson, C.L. (2017). Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. Nat. Rev. Mol. Cell Biol.

Clark, R.J., and Felsenfeld, G. (1971). Structure of chromatin. Nature *229*.

Clarke, J., Wu, H.C., Jayasinghe, L., Patel, A., Reid, S., and Bayley, H. (2009). Continuous base identification for single-molecule nanopore DNA sequencing. Nat. Nanotechnol. *4*, 265–270.

Conconi, A., Widmer, R.M., Koller, T., and Sogo, J.M. (1989). Two different chromatin structures coexist in ribosomal RNA genes throughout the cell cycle. Cell *57*, 753–761.

Corces, M.R., Trevino, A.E., Hamilton, E.G., Greenside, P.G., Sinnott-Armstrong, N.A., Vesuna, S., Satpathy, A.T., Rubin, A.J., Montine, K.S., Wu, B., et al. (2017). An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat. Methods *14*, 959–962.

Cusanovich, D.A., Hill, A.J., Aghamirzaie, D., Daza, R.M., Pliner, H.A., Berletch, J.B., Filippova, G.N., Huang, X., Christiansen, L., DeWitt, W.S., et al. (2018). A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility. Cell *174*.

Dai, X., and Zhu, M. (2020). Coupling of Ribosome Synthesis and Translational Capacity with Cell Growth. Trends Biochem. Sci. *45*, 681–692.

Dammann, R., Lucchini, R., Koller, T., and Sogo, J. (1995). Transcription in the Yeast rRNA Gene Locus: Distribution of the Active Gene Copies and Chromatin Structure of Their Flanking Regulatory Sequences. Mol. Cell. Biol. *15*, 10.

Dang, W., Steffen, K.K., Perry, R., Dorsey, J.A., Johnson, F.B., Shilatifard, A., Kaeberlein, M., Kennedy, B.K., and Berger, S.L. (2009). Histone H4 lysine 16 acetylation regulates cellular lifespan. Nature *459*, 802–807.

Darst, R.P., Nabilsi, N.H., Pardo, C.E., Riva, A., and Kladde, M.P. (2012). DNA methyltransferase accessibility protocol for individual templates by deep sequencing (Elsevier Inc.).

Dauban, L., Kamgoué, A., Wang, R., Léger-Silvestre, I., Beckouët, F., Cantaloube, S., and Gadal, O. (2019). Quantification of the dynamic behaviour of ribosomal DNA genes and nucleolus during yeast Saccharomyces cerevisiae cell cycle. J. Struct. Biol. *208*, 152–164.

Davey, C.A., Sargent, D.F., Luger, K., Maeder, A.W., and Richmond, T.J. (2002). Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. J. Mol. Biol. *319*, 1097–1113.

Dechassa, M.L., Hota, S.K., Sen, P., Chatterjee, N., Prasad, P., and Bartholomew, B. (2012). Disparity in the DNA translocase domains of SWI/SNF and ISW2. Nucleic Acids Res. *40*, 4412–4421.

Deniz, Ö., Flores, O., Aldea, M., Soler-López, M., and Orozco, M. (2016). Nucleosome architecture throughout the cell cycle. Sci. Rep. *6*, 1–11.

Derrington, I.M., Butler, T.Z., Collins, M.D., Manrao, E., Pavlenok, M., Niederweis, M., and Gundlach, J.H. (2010). Nanopore DNA sequencing with MspA. Proc. Natl. Acad. Sci. U. S. A. *107*, 16060–16065.

Dingwall, C., Lomonossoff, G., and Laskey, R. (1981). High sequence specificity of micrococcal nuclease. Methods *9*.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., et al. (2009). Real-time DNA sequencing from single polymerase molecules. Science (80-. ). *323*, 133–138.

Farnung, L., Vos, S.M., Wigge, C., and Cramer, P. (2017). Nucleosome–Chd1 structure and implications for chromatin remodelling. Nature.

Fei, J., Torigoe, S.E., Brown, C.R., Khuong, M.T., Kassavetis, G.A., Boeger, H., and Kadonaga, J.T. (2015). The prenucleosome, a stable conformational isomer of the nucleosome. Genes Dev. *29*, 2563–2575.

Fides, L., Theresa, K., and Jones, P. (2018). DNA methylation protocols (3rd edition). Methods Mol. Biol. *3rd editio*, 267–284.

Finch, J.T., and Klug, A. (1976). Solenoidal model for superstructure in chromatin. Proc. Natl. Acad. Sci. U. S. A. *73*, 1897–1901.

Flemming, W. (1882). Zellsubstanz, Kern und Zelltheilung.

Flusberg, B.A., Webster, D.R., Lee, J.H., Travers, K.J., Olivares, E.C., Clark, T.A., Korlach, J., and Turner, S.W. (2010). Direct detection of DNA methylation during single-molecule, real-time sequencing. Nat. Methods *7*, 461–465.

Forné, I., Ludwigsen, J., Imhof, A., Becker, P.B., and Mueller-Planitz, F. (2012). Probing the Conformation of the ISWI ATPase Domain With Genetically Encoded Photoreactive Crosslinkers and Mass Spectrometry. Mol. Cell. Proteomics *11*, M111.012088.

French, S.L., Osheim, Y.N., Schneider, D.A., Sikes, M.L., Fernandez, C.F., Copela, L.A., Misra, V.A., Nomura, M., Wolin, S.L., and Beyer, A.L. (2008). Visual Analysis of the Yeast 5S rRNA Gene Transcriptome: Regulation and Role of La Protein. Mol. Cell. Biol. *28*, 4576–4587.

Fyodorov, D. V., and Kadonaga, J.T. (2002). Dynamics of ATP-dependent chromatin assembly by ACF. Nature *418*, 896–900.

Ganguli, D., Chereji, R. V., Iben, J.R., Cole, H.A., and Clark, D.J. (2014). RSC-dependent constructive and destructive interference between opposing arrays of phased nucleosomes in yeast. Genome Res. *24*, 1637–1649.

Garcia, J.F., Dumesic, P.A., Hartley, P.D., El-Samad, H., and Madhani, H.D. (2010). Combinatorial, site-specific requirement for heterochromatic silencing factors in the elimination of nucleosome-free regions. Genes Dev. *24*, 1758–1771.

Gavin, A.C., Bösche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.M., Cruciat, C.M., et al. (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature *415*, 141–147.

Ghaemmaghami, S., Huh, W.K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K., and Weissman, J.S. (2003). Global analysis of protein expression in yeast. Nature *425*, 737–741.

Gkikopoulos, T., Schofield, P., Singh, V., Pinskaya, M., Mellor, J., Smolle, M., Workman, J.L., Barton, G.J., and Owen-Hughes, T. (2011). A Role for Snf2-Related Nucleosome-Spacing Enzymes in Genome-Wide Nucleosome Organization. Science (80-. ). *333*, 1758–1760.

Goyal, P., Krasteva, P. V., Van Gerven, N., Gubellini, F., Van Den Broeck, I., Troupiotis-Tsaïlaki, A., Jonckheere, W., Péhau-Arnaudet, G., Pinkner, J.S., Chapman, M.R., et al. (2014). Structural and mechanistic insights into the bacterial amyloid secretion channel CsgG. Nature *516*, 250–253.

Grafstrom, R.C., Fornace, A.J., Autrup, H., Lechner, J.F., and Harris, C.C. (1983). Formaldehyde damage to DNA and inhibition of DNA repair in human bronchial cells. Science (80-. ). *220*, 216–218.

Gutin, J., Sadeh, R., Bodenheimer, N., Joseph-Strauss, D., Klein-Brill, A., Alajem, A., Ram, O., and Friedman, N. (2018). Fine-Resolution Mapping of TF Binding and Chromatin Interactions. Cell Rep. *22*, 2797–2807.

Hall, M.A., Shundrovsky, A., Bai, L., Fulbright, R.M., Lis, J.T., and Wang, M.D. (2009). High-resolution dynamic mapping of histone-DNA interactions in a nucleosome. Nat. Struct. Mol. Biol. *16*, 124–129.

Hamiche, A., Schultz, P., Ramakrishnan, V., Oudet, P., and Prunell, A. (1996). Linker histone-dependent DNA structure in linear mononucleosomes. J. Mol. Biol. *257*, 30–42.

Hansen, J.C. (2012). Human mitotic chromosome structure: What happened to the 30-nm fibre? EMBO J. *31*, 1621–1623.

Harrer, N., Schindler, C.E.M., Bruetzel, L.K., Forné, I., Ludwigsen, J., Imhof, A., Zacharias, M., Lipfert, J., and Mueller-Planitz, F. (2018). Structural Architecture of the Nucleosome Remodeler ISWI Determined from Cross-Linking, Mass Spectrometry, SAXS, and Modeling. Structure *26*, 282-294.e6.

Hendrickson, D.G., Soifer, I., Wranik, B.J., Kim, G., Robles, M., Gibney, P.A., and McIsaac, R.S. (2018). A new experimental platform facilitates assessment of the transcriptional and chromatin landscapes of aging yeast. Elife 1–31.

Hewish, D.R., and Burgoyne, L.A. (1973). Chromatin sub-structure. The digestion of chromatin DNA at regularly spaced sites by a nuclear deoxyribonuclease. Biochem. Biophys. Res. Commun. *52*, 504–510.

Hontz, R.D., French, S.L., Oakes, M.L., Tongaonkar, P., Nomura, M., Beyer, A.L., and Smith, J.S. (2008). Transcription of Multiple Yeast Ribosomal DNA Genes Requires Targeting of UAF to the Promoter by Uaf30. Mol. Cell. Biol. *28*, 6709–6719.

Iida, T., and Kobayashi, T. (2019). RNA Polymerase I Activators Count and Adjust Ribosomal RNA Gene Copy Number. Mol. Cell *73*, 645-654.e13.

Ito, T., Bulger, M., Pazin, M.J., Kobayashi, R., and Kadonaga, J.T. (1997). ACF, an ISWI-containing and ATP-utilizing chromatin assembly and remodeling factor. Cell *90*, 145–155.

Iyer, V., and Struhl, K. (1995). Poly(dA:dT), a ubiquitous promoter element that stimulates transcription via its intrinsic DNA structure. EMBO J. *14*, 2570–2579.

Jones, H.S., Kawauchi, J., Braglia, P., Alen, C.M., Kent, N.A., and Proudfoot, N.J. (2007). RNA polymerase I in yeast transcribes dynamic nucleosomal rDNA. Nat. Struct. Mol. Biol. *14*, 123–130.

Kawanishi, M., Matsuda, T., and Yagi, T. (2014). Genotoxicity of formaldehyde: Molecular basis of DNA damage and mutation. Front. Environ. Sci. *2*, 1–8.

Kawasaki, F., Beraldi, D., Hardisty, R.E., McInroy, G.R., van Delft, P., and Balasubramanian, S. (2017). Genome-wide mapping of 5-hydroxymethyluracil in the eukaryote parasite Leishmania. Genome Biol. *18*, 1–8.

Kelly, T.K., Miranda, T., Liang, G., Berman, B., Lin, J., Tana, A., and Jones, P. (2010). H2A.Z Maintenance During Mitosis Reveals Nucleosome Shifting on Mitotically Silenced Genes. Mol. Cell *39*.

Kelly, T.K., Liu, Y., Lay, F.D., Liang, G., Berman, B.P., and Jones, P. a (2012). Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. Genome Res. 2497–2506.

Keys, D.A., Lee, B.S., Dodd, J.A., Nguyen, T.T., Vu, L., Fantino, E., Burson, L.M., Nogi, Y., and Nomura, M. (1996). Multiprotein transcription factor UAF interacts with the upstream element of the yeast RNA polymerase I promoter and forms a stable preinitiation complex. Genes Dev. *10*, 887–903.

Klinker, H., Haas, C., Harrer, N., Becker, P.B., and Mueller-Planitz, F. (2014). Rapid purification of recombinant histones. PLoS One *9*.

Knutson, B.A., Smith, M.L., Belkevich, A.E., and Fakhouri, A.M. (2020). Molecular Topology of RNA Polymerase I Upstream Activation Factor. Mol. Cell. Biol. 1–17.

Kono, H., Sakuraba, S., and Ishida, H. (2018). Free energy profiles for unwrapping the outer superhelical turn of nucleosomal DNA. PLoS Comput. Biol. *14*, 1–21.

Konrad, S.F., Vanderlinden, W., Frederickx, W., Brouns, T., Menze, B.H., De Feyter, S., and Lipfert, J. (2021). High-throughput AFM analysis reveals unwrapping pathways of H3 and CENP-A nucleosomes. Nanoscale *13*, 5435–5447.

Kornberg, R. (1974). Chromatin Structure : A Repeating Unit of Histones and DNA. Science (80-. ). *184*, 868–871.

Kornberg, R.D., and Stryer, L. (1988). Statistical distributions of nucleosomes: nonrandom locations by a stochastic mechanism. Nucleic Acids Res. *16*.

Kornberg, R.D., and Thomas, J.O. (1974). Chromatin Structure: Oligomers of the Histones. Science (80-. ). *184*, 865–868.

Krebs, A.R., Imanci, D., Hoerner, L., Gaidatzis, D., Burger, L., and Schübeler, D. (2017). Genome-wide Single-Molecule Footprinting Reveals High RNA Polymerase II Turnover at Paused Promoters. Mol. Cell *67*, 411-422.e4.

Krietenstein, N., Wal, M., Watanabe, S., Park, B., Peterson, C.L., Pugh, B.F., and Korber, P. (2016). Genomic Nucleosome Organization Reconstituted with Pure Proteins. Cell *167*, 709-721.e12.

Kubik, S., Bruzzone, M.J., Jacquet, P., Falcone, J.L., Rougemont, J., and Shore, D. (2015). Nucleosome Stability Distinguishes Two Different Promoter Types at All Protein-Coding Genes in Yeast. Mol. Cell *60*, 422–434.

Kubik, S., O'Duibhir, E., de Jonge, W.J., Mattarocci, S., Albert, B., Falcone, J.L., Bruzzone, M.J., Holstege, F.C.P., and Shore, D. (2018). Sequence-Directed Action of RSC Remodeler and General Regulatory Factors Modulates +1 Nucleosome Position to Facilitate Transcription. Mol. Cell *71*.

Kubik, S., Challal, D., Bruzzone, M.J., Dreos, R., Mattarocci, S., Bucher, P., Libri, D., and Shore, D. (2019). Opposing chromatin remodelers control transcription initiation frequency and start site selection. Nat. Struct. Mol. Biol. *26*.

Kulaeva, O.I., Hsieh, F.K., and Studitsky, V.M. (2010). RNA polymerase complexes cooperate to relieve the nucleosomal barrier and evict histones. Proc. Natl. Acad. Sci. U. S. A. *107*, 11325–11330.

Lai, B., Gao, W., Cui, K., Xie, W., Tang, Q., Jin, W., Hu, G., Ni, B., and Zhao, K. (2018). Principles of nucleosome organization revealed by single-cell micrococcal nuclease sequencing. Nature 1.

Längst, G., Bonte, E.J., Corona, D.F. V, and Becker, P.B. (1999). Nucleosome Movement by CHRAC and ISWI without Disruption or trans-Displacement of the Histone Octamer. Cell *97*, 843–852.

Lantermann, A.B., Straub, T., Strålfors, A., Yuan, G.-C., Ekwall, K., and Korber, P. (2010). Schizosaccharomyces pombe genome-wide nucleosome mapping reveals positioning mechanisms distinct from those of Saccharomyces cerevisiae. Nat. Struct. Mol. Biol. *17*, 251–257.

Laszlo, A.H., Derrington, I.M., Brinkerhoff, H., Langford, K.W., Nova, I.C., Samson, J.M., Bartlett, J.J., Pavlenok, M., and

Gundlach, J.H. (2013). Detection and mapping of 5-methylcytosine and 5-hydroxymethylcytosine with nanopore MspA. Proc. Natl. Acad. Sci. U. S. A. *110*, 18904–18909.

Lazarovici, A., Zhou, T., Shafer, A., MacHado, A.C.D., Riley, T.R., Sandstrom, R., Sabo, P.J., Lu, Y., Rohs, R., Stamatoyannopoulos, J.A., et al. (2013). Probing DNA shape and methylation state on a genomic scale with DNase I. Proc. Natl. Acad. Sci. U. S. A. *110*, 6376–6381.

Lee, I., Razaghi, R., Gilpatrick, T., Molnar, M., Gershman, A., Sadowski, N., Sedlazeck, F.J., Hansen, K.D., Simpson, J.T., and Timp, W. (2020). Simultaneous profiling of chromatin accessibility and methylation on human cell lines with nanopore sequencing. Nat. Methods.

Lee, W., Tillo, D., Bray, N., Morse, R.H., Davis, R.W., Hughes, T.R., and Nislow, C. (2007). A high-resolution atlas of nucleosome occupancy in yeast. Nat. Genet. *39*, 1235–1244.

Lesur, I., and Campbell, J.L. (2004). The Transcriptome of Prematurely Aging Yeast Cells Is Similar to That of Telomerase-deficient Cells. Mol. Biol. Cell *15*.

Lieleg, C., Ketterer, P., Nuebler, J., Ludwigsen, J., Gerland, U., Dietz, H., Mueller-Planitz, F., and Korber, P. (2015). Nucleosome spacing generated by ISWI and CHD1 remodelers is constant regardless of nucleosome density. Mol. Cell. Biol. *35*, 1588–1605.

Liu, M., Lu, Y., Yang, B., Chen, Y., Radda, J.S.D., Hu, M., Katz, S.G., and Wang, S. (2020). Multiplexed imaging of nucleome architectures in single cells of mammalian tissue. Nat. Commun. *11*, 1–14.

Liu, X., Li, M., Xia, X., Li, X., and Chen, Z. (2017). Mechanism of chromatin remodelling revealed by the Snf2-nucleosome structure. Nature *544*, 440–445.

Logie, C., and Peterson, C.L. (1997). Catalytic activity of the yeast SWI/SNF complex on reconstituted nucleosome arrays. EMBO J. *16*, 6772–6782.

Lorch, Y., and Kornberg, R.D. (2015). Chromatin-remodeling and the initiation of transcription. Q. Rev. Biophys. *48*, 465–470.

Lowary, P.., and Widom, J. (1998). New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. J. Mol. Biol. *276*, 19–42.

Ludwigsen, J., Klinker, H., and Mueller-Planitz, F. (2013). No need for a power stroke in ISWI-mediated nucleosome sliding. EMBO Rep. *14*, 1092–1097.

Ludwigsen, J., Pfennig, S., Singh, A.K., Schindler, C., Harrer, N., Forné, I., Zacharias, M., and Mueller-Planitz, F. (2017). Concerted regulation of ISWI by an autoinhibitory domain and the H4 N-terminal tail. Elife *6*, 1–24.

Ludwigsen, J., Hepp, N., Klinker, H., Pfennig, S., and Mueller-Planitz, F. (2018). Remodeling and Repositioning of Nucleosomes in Nucleosomal Arrays. Mol. Mot. Methods Protoc. *1805*, 349–370.

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251–260.

Lusser, A., Urwin, D.L., and Kadonaga, J.T. (2005). Distinct activities of CHD1 and ACF in ATP-dependent chromatin assembly. Nat. Struct. Mol. Biol. *12*, 160–166.

MacKay, V.L., Welch, S.K., Insley, M.Y., Manney, T.R., Holly, J., Saari, G.C., and Parker, M.L. (1988). The Saccharomyces cerevisiae BAR1 gene encodes an exported protein with homology to pepsin. Proc. Natl. Acad. Sci. U. S. A. *85*, 55–59.

Maeshima, K., Hihara, S., and Eltsov, M. (2010). Chromatin structure: Does the 30-nm fibre exist in vivo? Curr. Opin. Cell Biol. *22*, 291–297.

Maeshima, K., Imai, R., Tamura, S., and Nozaki, T. (2014). Chromatin as dynamic 10-nm fibers. Chromosoma *123*, 225–237.

Maeshima, K., Ide, S., Hibino, K., and Sasai, M. (2016). Liquid-like behavior of chromatin. Curr. Opin. Genet. Dev. *37*, 36–45.

Maeshima, K., Ide, S., and Babokhov, M. (2019). Dynamic chromatin organization without the 30-nm fiber. Curr. Opin. Cell Biol. *58*, 95–104.

Mann, R.K., and Grunstein, M. (1992). Histone H3 N-terminal mutations allow hyperactivation of the yeast GAL 1 gene in vivo. EMBO J. *11*, 3297–3306.

Manrao, E.A., Derrington, I.M., Laszlo, A.H., Langford, K.W., Hopper, M.K., Gillgren, N., Pavlenok, M., Niederweis, M., and Gundlach, J.H. (2012). Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. Nat. Biotechnol. *30*, 349–353.

Marques, M., Laflamme, L., Gervais, A.L., and Gaudreau, L. (2010). Reconciling the positive and negative roles of histone H2A.Z in gene transcription. Epigenetics *5*, 267–272.

Matheson, K., Parsons, L., and Gammie, A. (2017). Whole-Genome Sequence and Variant Analysis of W303, a Widely-Used Strain of Saccharomyces cerevisiae. G3; Genes, Genomes, Genet. *7*, 2219–2226.

Matsuo, K., Silke, J., Gramatikoff, K., and Schaffner, W. (1994). The CpG-specific methylase SssI has topoisomerase activity in the presence of Mg2+. Nucleic Acids Res. *22*, 5354–5359.

McKnight, J.N., Jenkins, K.R., Nodelman, I.M., Escobar, T., and Bowman, G.D. (2011). Extranucleosomal DNA Binding Directs Nucleosome Sliding by Chd1. Mol. Cell. Biol. *31*, 4746–4759.

Merz, K., Hondele, M., Goetze, H., Gmelch, K., Stoeckl, U., and Griesenbeck, J. (2008). Actively transcribed rRNA genes in S. cerevisiae are organized in a specialized chromatin associated with the high-mobility group protein Hmo1 and are largely devoid of histone molecules. Genes Dev. *22*, 1190–1204.

Mieczkowski, J., Cook, A., Bowman, S.K., Mueller, B., Alver, B.H., Kundu, S., Deaton, A.M., Urban, J.A., Larschan, E., Park, P.J., et al. (2016). MNase titration reveals differences between nucleosome occupancy and chromatin accessibility. Nat. Commun. *7*.

Mirny, L.A., Imakaev, M., and Abdennur, N. (2019). Two major mechanisms of chromosome organization. Curr. Opin. Cell Biol. *58*, 142–152.

Mittal, N., Guimaraes, J.C., Gross, T., Schmidt, A., Vina-Vilaseca, A., Nedialkova, D.D., Aeschimann, F., Leidel, S.A., Spang, A., and Zavolan, M. (2017). The Gcn4 transcription factor reduces protein synthesis capacity and extends yeast lifespan. Nat. Commun. *8*.

Mizuguchi, G., Shen, X., Landry, J., Wu, W.H., Sen, S., and Wu, C. (2004). ATP-Driven Exchange of Histone H2AZ Variant Catalyzed by SWR1 Chromatin Remodeling Complex. Science (80-. ). *303*, 343–348.

Möbius, W., and Gerland, U. (2010). Quantitative test of the barrier nucleosome model for statistical positioning of nucleosomes up- and downstream of transcription start sites. PLoS Comput. Biol. *6*.

Möbius, W., Osberg, B., Tsankov, A.M., Rando, O.J., and Gerland, U. (2013). Toward a unified physical model of nucleosome patterns flanking transcription start sites. Proc. Natl. Acad. Sci. U. S. A. *110*, 5719–5724.

Montefiori, L., Hernandez, L., Zhang, Z., Gilad, Y., Ober, C., Crawford, G., Nobrega, M., and Sakabe, N.J. (2017). Reducing mitochondrial reads in ATAC-seq using CRISPR/Cas9. Sci. Rep. *7*, 1–9.

Mueller-Planitz, F., Klinker, Henrikeeller-Planitz, F., Ludwigsen, J., and Becker, P.B. (2013). The ATPase domain of ISWI is an autonomous nucleosome remodeling machine. Nat. Struct. Mol. Biol. *20*, 82–89.

Mulqueen, R.M., DeRosa, B.A., Thornton, C.A., Sayar, Z., Torkenczy, K.A., Fields, A.J., Wright, K.M., Nan, X., Ramji, R., Steemers, F.J., et al. (2019). Improved single-cell ATAC-seq reveals chromatin dynamics of in vitro corticogenesis. BioRxiv.

Ngo, T.T.M., Zhang, Q., Zhou, R., Yodh, J.G., and Ha, T. (2015). Asymmetric unwrapping of nucleosomes under tension directed by DNA local flexibility. Cell *160*, 1135–1144.

Ni, P., Huang, N., Zhang, Z., Wang, D.P., Liang, F., Miao, Y., Xiao, C. Le, Luo, F., and Wang, J. (2019). DeepSignal: Detecting DNA methylation state from Nanopore sequencing reads using deep-learning. Bioinformatics *35*, 4586–4595.

Nishino, Y., Eltsov, M., Joti, Y., Ito, K., Takata, H., Takahashi, Y., Hihara, S., Frangakis, A.S., Imamoto, N., Ishikawa, T., et al. (2012). Human mitotic chromosomes consist predominantly of irregularly folded nucleosome fibres without a 30-nm chromatin structure. EMBO J. *31*, 1644–1653.

Noll, M. (1974). Subunit structure of chromatin. Nature *251*.

Oberbeckmann, E., Wolff, M., Krietenstein, N., Heron, M., Ellins, J.L., Schmid, A., Krebs, S., Blum, H., Gerland, U., and Korber, P. (2019). Absolute nucleosome occupancy map for the Saccharomyces cerevisiae genome. Genome Res. 1–49.

Ocampo, J., Chereji, R. V., Eriksson, P.R., and Clark, D.J. (2016). The ISW1 and CHD1 ATP-dependent chromatin remodelers compete to set nucleosome spacing in vivo. Nucleic Acids Res. *44*, 4625–4635.

Olins, A.L., and Olins, D.E. (1974). Spheroid chromatin units (v bodies). Science (80-. ). *183*, 330–332.

Oudet, P., Gross-Bellard, M., and Chambon, P. (1975). Electron Microsopic and Biochemical Evidence that Chromatin Structure Is a Repeating Unit. Cell *4*, 281–300.

Papamichos-Chronakis, M., Watanabe, S., Rando, O.J., and Peterson, C.L. (2011). Global regulation of H2A.Z localization by the INO80 chromatin-remodeling enzyme is essential for genome integrity. Cell *144*, 200–213.

Parnell, T.J., Schlichter, A., Wilson, B.G., and Cairns, B.R. (2015). The chromatin remodelers RSC and ISW1 display functional and chromatin-based promoter antagonism. Elife *2015*, 1–21.

Patterton, H.G., Landel, C.C., Landsman, D., Peterson, C.L., and Simpson, R.T. (1998). The biochemical and phenotypic characterization of Hho1p, the putative linker histone H1 of Saccharomyces cerevisiae. J. Biol. Chem. *273*, 7268–7276.

Peterson, C.L., and Laniel, M.A. (2004). Histones and histone modifications. Curr. Biol. *14*, 546–551.

Pidoux, A.L., and Allshire, R.C. (2004). Kinetochore and heterochromatin domains of the fission yeast centromere. Chromosom. Res. *12*, 521–534.

Raisner, R.M., Hartley, P.D., Meneghini, M.D., Bao, M.Z., Liu, C.L., Schreiber, S.L., Rando, O.J., and Madhani, H.D. (2005). Histone variant H2A.Z Marks the 5′ ends of both active and inactive genes in euchromatin. Cell *123*, 233–248.

Ramachandran, S., Ahmad, K., and Henikoff, S. (2017). Transcription and Remodeling Produce Asymmetrically Unwrapped Nucleosomal Intermediates. Mol. Cell *68*, 1038-1053.e4.

Rand, A.C., Jain, M., Eizenga, J.M., Musselman-Brown, A., Olsen, H.E., Akeson, M., and Paten, B. (2017). Mapping DNA methylation with high-throughput nanopore sequencing. Nat. Methods *14*, 411–413.

Ranjan, A., Wang, F., Mizuguchi, G., Wei, D., Huang, Y., and Wu, C. (2015). H2A histone-fold and DNA elements in nucleosome activate SWR1-mediated H2A.Z replacement in budding yeast. Elife *4*, 1–11.

Rawal, Y., Chereji, R. V, Qiu, H., Ananthakrishnan, S., Govind, C.K., Clark, D.J., and Hinnebusch, A.G. (2018). SWI / SNF and RSC cooperate to reposition and evict promoter nucleosomes at highly expressed genes in yeast. Genes Dev. 1–16.

Rhee, H.S., and Pugh, B.F. (2012). ChiP-exo method for identifying genomic location of DNA-binding proteins with near-single-nucleotide accuracy. Curr. Protoc. Mol. Biol. 1–14.

Rhoads, A., and Au, K.F. (2015). PacBio Sequencing and Its Applications. Genomics, Proteomics Bioinforma. *13*, 278–289.

Rhodes, D. (1979). Nucleosome cores reconstituted from poly (dA-dT) and the octamer of histones. Nucleic Acids Res. *6*.

Rossi, M.J., Kuntala, P.K., Lai, W.K.M., Yamada, N., Badjatia, N., Mittal, C., Kuzu, G., Bocklund, K., Farrell, N.P., Blanda, T.R., et al. (2021). A high-resolution protein architecture of the budding yeast genome. Nature.

Rube, H.T., and Song, J.S. (2014). Quantifying the role of steric constraints in nucleosome positioning. Nucleic Acids Res. *42*, 2147–2158.

Sadaie, M., Kawaguchi, R., Ohtani, Y., Arisaka, F., Tanaka, K., Shirahige, K., and Nakayama, J. (2008). Balance between Distinct HP1 Family Proteins Controls Heterochromatin Assembly in Fission Yeast. Mol. Cell. Biol. *28*, 6973–6988.

Salter, J.D., Bennett, R.P., and Smith, H.C. (2016). The APOBEC Protein Family: United by Structure, Divergent in Function. Trends Biochem. Sci. *41*, 578–594.

Satchwell, S.C., Drew, H.R., and Travers, A.A. (1986). Sequence Periodicities in Chicken Nucleosome Core DNA rotational sequencing. J. Mol. Biol. *191*, 659–675.

Scheffer, M.P., Eltsov, M., and Frangakis, A.S. (2011). Evidence for short-range helical order in the 30-nm chromatin fibers of erythrocyte nuclei. Proc. Natl. Acad. Sci. U. S. A. *108*, 16992–16997.

Schones, D.E., Cui, K., Cuddapah, S., Roh, T.Y., Barski, A., Wang, Z., Wei, G., and Zhao, K. (2008). Dynamic Regulation of Nucleosome Positioning in the Human Genome. Cell *132*, 887–898.

Sedat, J., and Manuelidis, L. (1977). A direct approach to the structure of eukaryotic chromosomes. Cold Spring Harb. Symp. Quant. Biol. *42*, 331–350.

Shibata, M., Yamashita, H., Uchihashi, T., Kandori, H., and Ando, T. (2010). High-speed atomic force microscopy shows dynamic molecular processes in photoactivated bacteriorhodopsin. Nat. Nanotechnol. *5*, 208–212.

Shim, J., Humphreys, G.I., Venkatesan, B.M., Munz, J.M., Zou, X., Sathe, C., Schulten, K., Kosari, F., Nardulli, A.M., Vasmatzis, G., et al. (2013). Detection and quantification of methylation in DNA using solid-state nanopores. Sci. Rep. *3*, 1389.

Shipony, Z., Marinov, G.K., Swaffer, M.P., Sinott-Armstrong, N.A., Skotheim, J.M., Kundaje, A., and Greenleaf, W.J. (2020). Long-range single-molecule mapping of chromatin accessibility in eukaryotes. Nat. Methods.

Simpson, R.T. (1978). Structure of the Chromatosome, a Chromatin Particle Containing 160 Base Pairs of DNA and All the Histones. Biochemistry *17*, 5524–5531.

Simpson, R., and Kunzler, P. (1979). Chromatin and core particles formed from the inner histones and synthetic polydeoxyribonucleotides of defined sequence. Nucleic Acids Res. *6*.

Simpson, J.T., Workman, R.E., Zuzarte, P.C., David, M., Dursi, L.J., and Timp, W. (2017). Detecting DNA cytosine methylation using nanopore sequencing. Nat. Methods *14*, 407–410.

Sinclair, D.A., and Guarente, L. (1997). Extrachromosomal rDNA circles - A cause of aging in yeast. Cell *91*, 1033–1042.

Singh, A.K., and Mueller-Planitz, F. (2021). Nucleosome positioning and spacing: from mechanism to function. J. Mol. Biol. 166847.

Singh, A.K., Schauer, T., Pfaller, L., Straub, T., and Mueller-Planitz, F. (2021). The biogenesis and function of nucleosome arrays. BioRxiv.

Singh, R.K., Fan, J., Gioacchini, N., Watanabe, S., Bilsel, O., and Peterson, C.L. (2019). Transient Kinetic Analysis of SWR1C-Catalyzed H2A.Z Deposition Unravels the Impact of Nucleosome Dynamics and the Asymmetry of Histone Exchange. Cell Rep. *27*, 374-386.e4.

Smith, M.L., Cui, W., Jackobel, A.J., Walker-Kopp, N., and Knutson, B.A. (2018). Reconstitution of RNA Polymerase I Upstream Activating Factor and the Roles of Histones H3 and H4 in Complex Assembly. J. Mol. Biol. *430*, 641–654.

Smolle, M., Venkatesh, S., Gogol, M.M., Li, H., Zhang, Y., Florens, L., Washburn, M.P., and Workman, J.L. (2012). Chromatin remodelers Isw1 and Chd1 maintain chromatin structure during transcription by preventing histone exchange. Nat. Struct. Mol. Biol. *19*, 884–892.

Stergachis, A.B., Debo, B.M., Haugen, E., Churchman, L.S., and Stamatoyannopoulos, J.A. (2020). Single-molecule regulatory architectures captured by chromatin fiber sequencing. Science (80-. ). *368*, 1449–1454.

Stewart-Morgan, K.R., Petryk, N., and Groth, A. (2020). Chromatin replication and epigenetic cell memory. Nat. Cell Biol. *22*, 361–371.

Stockdale, C., Flaus, A., Ferreira, H., and Owen-Hughes, T. (2006). Analysis of nucleosome repositioning by yeast ISWI and Chd1 chromatin remodeling complexes. J. Biol. Chem. *281*, 16279–16288.

Stoddart, D., Heron, A.J., Mikhailova, E., Maglia, G., and Bayley, H. (2009). Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore. Proc. Natl. Acad. Sci. U. S. A. *106*, 7702–7707.

Stoiber, M., Quick, J., Egan, R., Eun Lee, J., Celniker, S., Neely, R., Loman, N., Pennacchio, L., and Brown, J. (2017). De novo Identification of DNA Modifications Enabled by Genome-Guided Nanopore Signal Processing. BioRxiv 094672.

Stoler, S., Keith, K.C., Curnick, K.E., and Fitzgerald-Hayes, M. (1995). A mutation in CSE4, an essential gene encoding a novel chromatin- associated protein in yeast, causes chromosome nondisjunction and cell cycle arrest at mitosis. Genes Dev. *9*, 573–586.

Struhl, K., and Segal, E. (2013). Determinants of nucleosome positioning. Nat. Struct. Mol. Biol. *20*, 267–273.

Telford, D.J., and Stewart, B.W. (1989). Micrococcal nuclease: Its specificity and use for chromatin analysis. Int. J. Biochem. *21*, 127–138.

Thomas, B.J., and Rothstein, R. (1989). Elevated recombination rates in transcriptionally active DNA. Cell *56*, 619–630.

Thon, G., and Verhein-Hansen, J. (2000). Four chromo-domain proteins of Schizosaccharomyces pombe differentially repress transcription at various chromosomal locations. Genetics *155*, 551–568.

Timinskas, A., Butkus, V., and Janulaitis, A. (1995). Sequence motifs characteristic for DNA [cytosine-N4] and DNA [adenine-N6] methyltransferases. Classification of all DNA methyltransferases. Gene *157*, 3–11.

Travers, K.J., Chin, C.S., Rank, D.R., Eid, J.S., and Turner, S.W. (2010). A flexible and efficient template format for circular consensus sequencing and SNP detection. Nucleic Acids Res. *38*.

Tsukiyama, T., Palmer, J., Landel, C.C., Shiloach, J., and Wu, C. (1999). Characterization of the imitation switch subfamily of ATP-dependent chromatin-remodeling factors in Saccharomyces cerevisiae. Genes Dev. *13*, 686–697.

Udugama, M., Sabri, A., and Bartholomew, B. (2011). The INO80 ATP-Dependent Chromatin Remodeling Complex Is a Nucleosome Spacing Factor. Mol. Cell. Biol. *31*, 662–673.

Vaillant, C., Palmeira, L., Chevereau, G., Audit, B., D'Aubenton-Carafa, Y., Thermes, C., and Arneodo, A. (2010). A novel strategy of transcription regulation by intragenic nucleosome ordering. Genome Res. *20*, 59–67.

Vary, J.C., Gangaraju, V.K., Qin, J., Landel, C.C., Kooperberg, C., Bartholomew, B., and Tsukiyama, T. (2003). Yeast Isw1p Forms Two Separable Complexes In Vivo. Mol. Cell. Biol. *23*, 80–91.

Vasseur, P., Tonazzini, S., Ziane, R., Camasses, A., Rando, O.J., and Radman-Livaja, M. (2016). Dynamics of Nucleosome Positioning Maturation following Genomic Replication. Cell Rep. *16*, 2651–2665.

Venema, J., and Tollervey, D. (1999). Saccharomyces cerevisiae. Annu. Rev. Genet. 1750–1751.

Vollger, M.R., Logsdon, G.A., Audano, P.A., Sulovari, A., Porubsky, D., Peluso, P., Wenger, A.M., Concepcion, G.T., Kronenberg, Z.N., Munson, K.M., et al. (2020). Improved assembly and variant detection of a haploid human genome using single-molecule, high-fidelity long reads. Ann. Hum. Genet. *84*, 125–140.

Wang, F., Ranjan, A., Wei, D., and Wu, C. (2016). Comment on "A histone acetylation switch regulates H2A.Z deposition by the SWR-C remodeling enzyme." Science (80-. ). *353*, 358a.

Wang, Y., Wang, A., Liu, Z., Thurman, A., Powers, L.S., Zou, M., Hefel, A., Li, Y., Zabner, J., and Au, K.F. (2019). Single-molecule long-read sequencing reveals the chromatin basis of gene expression. Genome Res. 533158.

Watanabe, S., Radman-Livaja, M., Rando, O.J., and Peterson, C.L. (2013). A histone acetylation switch regulates H2A.Z deposition by the SWR-C remodeling enzyme. Science (80-. ). *340*, 195–199.

Weiner, A., Hughes, A., Yassour, M., Rando, O.J., and Friedman, N. (2010). High-resolution nucleosome mapping reveals transcription-dependent promoter packaging. Genome Res. *20*, 90–100.

Wendell, D., Jing, P., Geng, J., Subramaniam, V., Lee, T.J., Montemagno, C., and Guo, P. (2009). Translocation of double-stranded DNA through membrane-adapted phi29 motor protein nanopores. Nat. Nanotechnol. *4*, 765–772.

Wenger, A.M., Peluso, P., Rowell, W.J., Chang, P.C., Hall, R.J., Concepcion, G.T., Ebler, J., Fungtammasan, A., Kolesnikov, A., Olson, N.D., et al. (2019). Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. Nat. Biotechnol. *37*, 1155–1162.

Woodcock, C.L. (1994). Chromatin fibers observed in situ in frozen hydrated sections. Native fiber diameter is not correlated with nucleosome repeat length. J. Cell Biol. *125*, 11–19.

Woodcock, C.L.F., Safer, J.P., and Stanchfield, J.E. (1976). Structural repeating units in chromatin. I. Evidence for their general occurrence. Exp. Cell Res. *97*, 101–110.

Woodcock, C.L.F., Frado, L.L.Y., and Rattner, J.B. (1984). The higher-order structure of chromatin: Evidence for a helical ribbon arrangement. J. Cell Biol. *99*, 42–52.

Wu, A.C.K., Patel, H., Chia, M., Moretto, F., Frith, D., Snijders, A.P., and van Werven, F.J. (2018). Repression of Divergent Noncoding Transcription by a Sequence-Specific Transcription Factor. Mol. Cell *72*, 942-954.e7.

Xi, Y., Yao, J., Chen, R., Li, W., and He, X. (2011). Nucleosome fragility reveals novel functional states of chromatin and poises genes for activation. Genome Res. *21*, 718–724.

Xu, M., Kladde, M.P., Van Etten, J.L., and Simpson, R.T. (1998). Cloning, characterization and expression of the gene coding for a cytosine-5-DNA methyltransferase recognizing GpC. Nucleic Acids Res. *26*, 3961–3966.

Xu, Y., Bernecky, C., Lee, C.T., Maier, K.C., Schwalb, B., Tegunov, D., Plitzko, J.M., Urlaub, H., and Cramer, P. (2017). Architecture of the RNA polymerase II-Paf1C-TFIIS transcription elongation complex. Nat. Commun. *8*.

Yadav, T., and Whitehouse, I. (2016). Replication-Coupled Nucleosome Assembly and Positioning by ATP-Dependent Chromatin-Remodeling Enzymes. Cell Rep. *15*, 715–723.

Yamada, K., Frouws, T.D., Angst, B., Fitzgerald, D.J., DeLuca, C., Schimmele, K., Sargent, D.F., and Richmond, T.J. (2011). Structure and mechanism of the chromatin remodelling factor ISW1a. Nature *472*, 448–453.

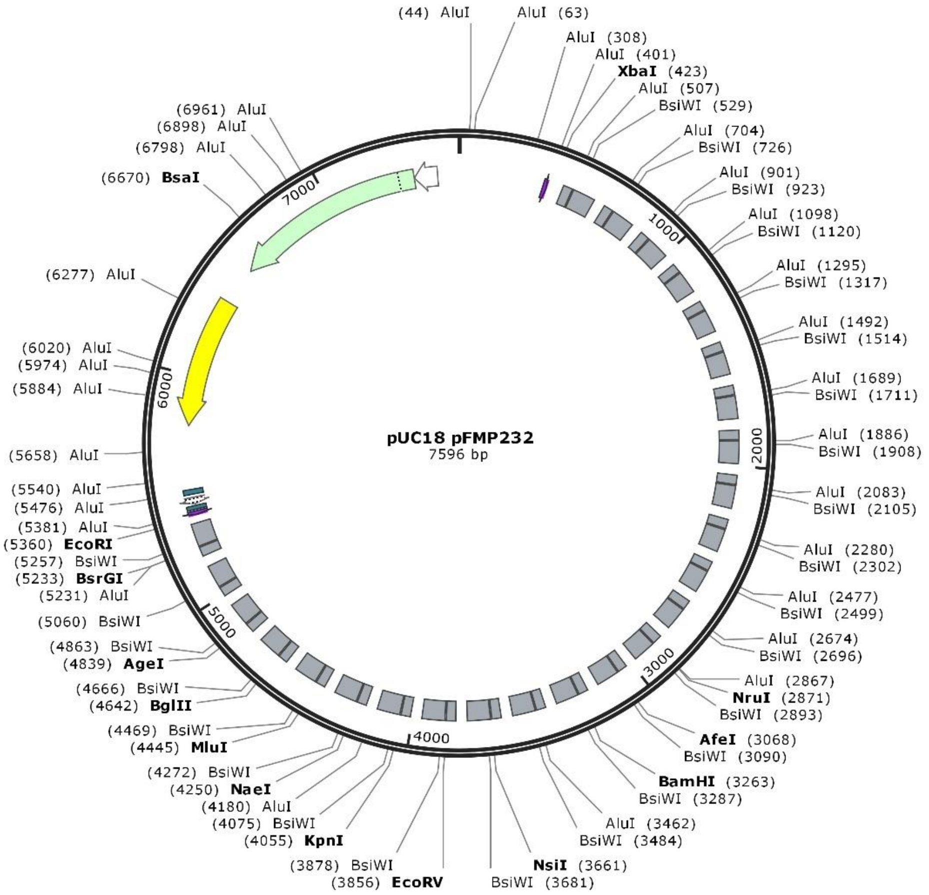
Zhang, Z., Wippo, C.J., Wal, M., Ward, E., Korber, P., and Pugh, F. (2011). A Packing Mechanism for Nucleosome Organization Reconstituted Across a Eukaryotic Genome. Science (80-. ). *332*, 977–980.

Zofall, M., Persinger, J., Kassabov, S.R., and Bartholomew, B. (2006). Chromatin remodeling by ISW2 and SWI/SNF requires DNA translocation inside the nucleosome. Nat. Struct. Mol. Biol. *13*, 339–346.

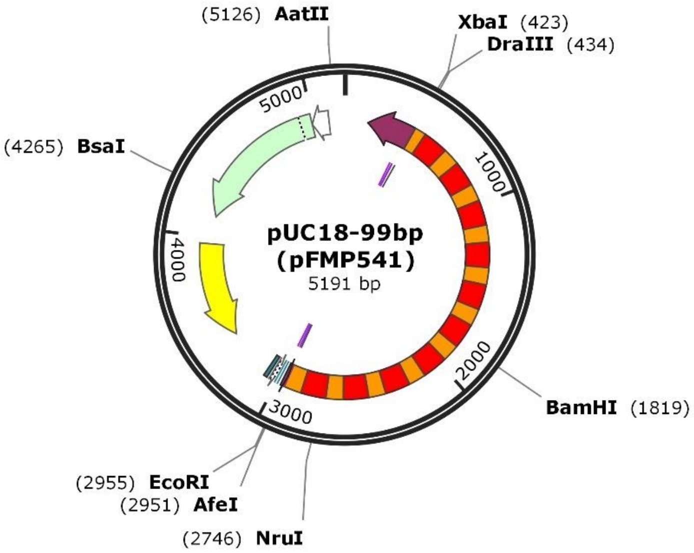
# 6. APPENDIX

## 6.1. Plasmid maps

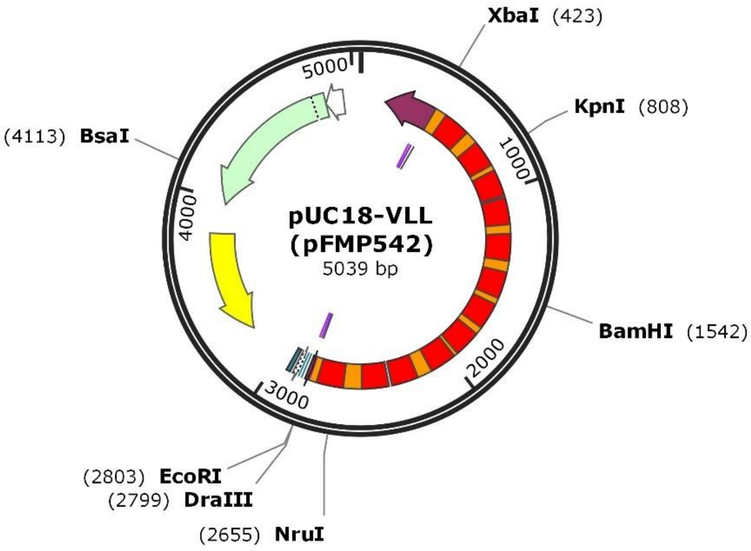
Plasmid maps showing the features and common restriction enzyme site locations. A brief summary of colors is given below each plasmid map. First plasmid map has brief description of common features found in all subsequent maps.



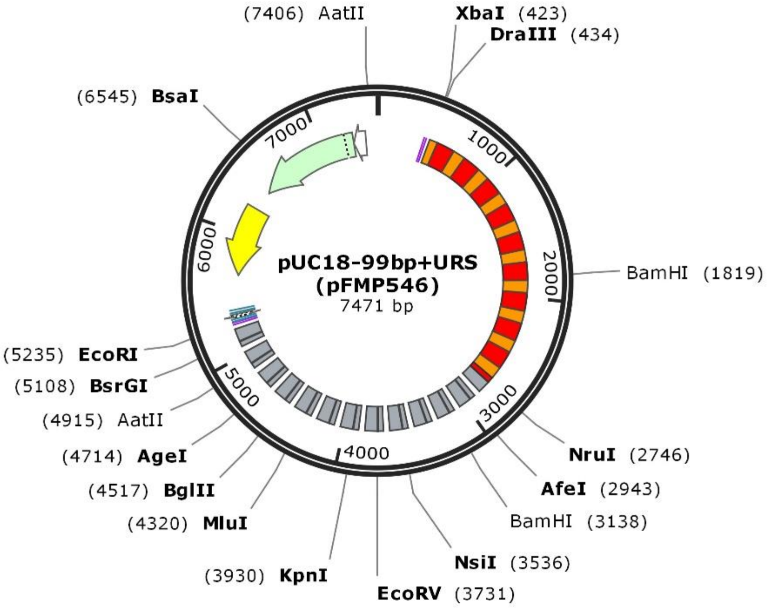
*Plasmid map of pFMP232 containing the reference 25-mer reference Widom 601 array with 50 bp linkers (Mueller-Planitz et al., 2013). Total length of plasmid is defined below the centre plasmid name. Original backbone was derived from a pUC18 plasmid. A reversed (but otherwise identical) version of this plasmid is stored as pFMP233. Individual reference Widom 601 sequences are indicated as grey blocks. One half of the array has unique restriction sites (URS) within the reference Widom 601 sequence. URS are labelled in bold, non-unique restriction sites are in regular font. Values in brackets indicate the location of each restriction site. Green feature is the AmpR gene which confers ampicillin resistance. Upstream white feature is the AmpR promotor. Yellow feature is the high copy number origin of replication. Other common features can be found described and labelled in reference pUC18 plasmid map (not provided here)*



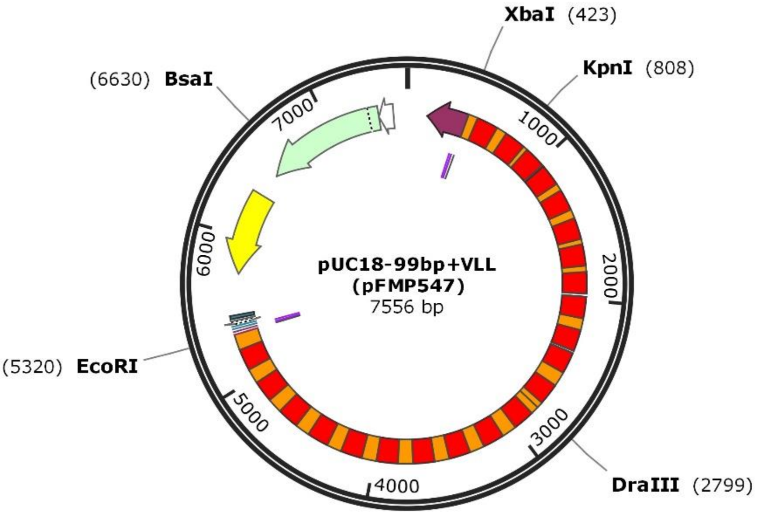
*Plasmid map of pFMP541 showing the distribution of modified 601 sequences (red) separated by 99 bp linkers (orange). Restriction sites for cloning and isolation are shown flanking the map. The pFMP541 plasmid has not been validated by long range sequencing. The 99 bp array section (DraIII to EcorI) has been successfully sequenced and validated as part of the pFMP546 plasmid.*



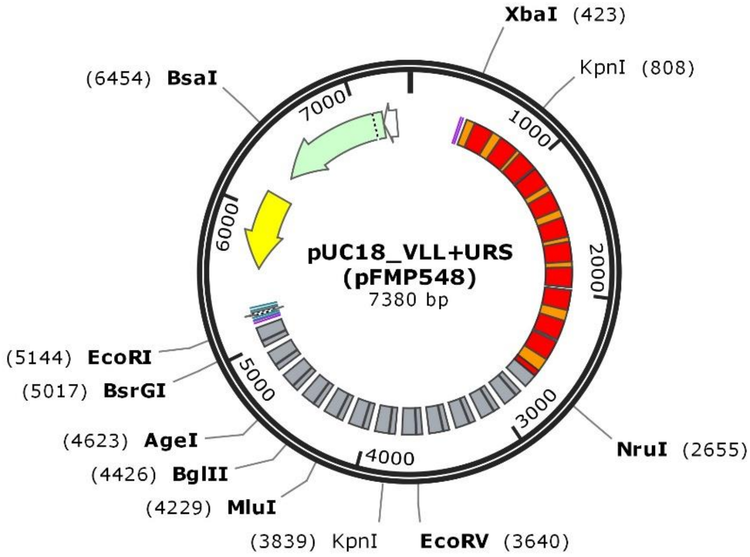
*Plasmid map of pFMP542 showing the distribution of modified 601 sequences (red) separated by variable linker lengths (VLL; orange). Restriction sites for cloning and isolation are shown flanking the map. The pFMP542 plasmid has not been sequence validated by long range sequencing.*



*Plasmid map of pFMP546 showing the distribution of modified 601 sequences (red) separated by 99 bp linkers (orange). Downstream of the NruI restriction site are original Widom 601 sequences follow (grey) and are separated by 50 bp linkers (uncoloured). pFMP546 is a combination of pFMP541 and pFMP232. Restriction sites for cloning and isolation are shown flanking the map. The pFMP546 plasmid has been successfully sequenced and validated (See Chapter 2.3.3)*



*Plasmid map of pFMP546 showing the distribution of modified 601 sequences (red) separated by variable linker lengths (VLL). Downstream of the DraIII restriction site are modified 601 sequences (red) identical to those found in the former half. 99 bp linkers (orange) separate the modified 601 sequences. pFMP547 is a combination of pFMP541 and pFMP542. Restriction sites for cloning and isolation are shown flanking the map. The pFMP547 plasmid has not been sequence validated by long range sequencing. The 99 bp array part of the plasmid (DraIII to EcorI) has been validated as part of the pFMP546 plasmid.*



*Theoretical plasmid map for pFMP548 showing the distribution of modified 601 sequences (red) separated by variable linker lengths (VLL). Downstream of the NruI restriction site are original Widom 601 sequences (grey) separated by 50 bp linkers (uncoloured). pFMP548 is a theoretical combination of pFMP232 and pFMP542. Restriction sites for cloning and isolation are shown flanking the map. Note that this plasmid has not been generated yet and is described for future reference.*

## 6.2. Sequencing run statistics

### 6.2.1. Site and run statistics

For each sample in each sequencing run (indicated by NPDXX) the total sequenced amount of megabases (Mbases) and unfiltered mean (aligned) read length is indicated (for median read lengths see Supplementary table 3). Biological replicates of WT or TKO strains are defined by a number after the strain type (i.e. WT1 and WT2 are biological replicates). Additional values in the sample name may indicate methylation time (i.e. time incubated with methyltransferase before quench) or units of methyltransferase used. The yeast strain used for each sample is indicated by a yFMP value and additional details pertaining to the genotype are found in the yeast strain list under Materials and Methods. A brief description for samples is added for clarification on experimental conditions. Unless otherwise state, samples are prepared and methylated according to materials and methods section 4.2.16. In addition, for each sample we calculate the site coverage (i.e. total non-ambiguous CpGs per site), site fraction ambiguous (i.e. fraction CpGs per sites that cannot be called methylated or unmethylated) and site accessibility (i.e. fraction of methylated sites over the sum of non-ambiguous sites), which can be used to calculate the absolute occupancy by subtracting the accessibility from one (Figure 1.3.2B). Samples that are not representative due to alternative experimental conditions are greyed out. Samples within a timecourse for which the site accessibility of the later timepoints is increased with >10% are denoted with a section sign (§) in the 'Notes on quality criteria' column. Such samples may not (yet) be in saturating conditions. Samples in which we saw a degradation pattern resulting in a loss of the high molecular weight band (based on gel electrophoresis experiments, for example see Figure 2.1.3B) are denoted with a pound sign (#) in the 'Notes on quality criteria' column. For description on the '&' sign see Supplementary table 2, for description on the '+' sign see Supplementary table 3

**Supplementary table 1: Site and sample information**

| Sample Name | Seq. run | Strain | Brief description | Mbases total | Mean length | Site mean coverage | Site fraction ambiguous | Site accessibility | Notes on quality criteria |
|---|---|---|---|---|---|---|---|---|---|
| WT0_30 | NPD03 | yFMP013 | WT sample methylated for 30 min using buffer lacking Mg2+. Not representative for typical conditions | 104 | 1756 | 5.7 | 0.34 | 0.210 | §, & |
| WT0_60 | NPD03 | yFMP013 | WT sample methylated for 60 min using buffer lacking Mg2+. Not representative for typical conditions | 156 | 1732 | 8.3 | 0.33 | 0.380 | § |
| WT0_180 | NPD03 | yFMP013 | WT sample methylated for 120 min using buffer lacking Mg2+. Not representative for typical conditions | 90 | 3115 | 4.8 | 0.35 | 0.550 | § |
| **TKO1_120min** | NPD07 | yFMP014 | 120 min M.SssI methylation using standard conditions | 1050 | 3744 | 59.5 | 0.31 | 0.274 | |
| **TKO1_60min** | NPD07 | yFMP014 | 60 min M.SssI methylation using standard conditions | 533 | 5763 | 30.0 | 0.32 | 0.273 | |
| **TKO2_120min** | NPD07 | yFMP014 | 120 min M.SssI methylation using standard conditions | 950 | 4104 | 53.4 | 0.31 | 0.253 | |
| **TKO2_60min** | NPD07 | yFMP014 | 60 min M.SssI methylation using standard conditions | 568 | 4989 | 31.7 | 0.32 | 0.249 | |
| **WT1_120min** | NPD07 | yFMP013 | 120 min M.SssI methylation using standard conditions | 485 | 4758 | 27.5 | 0.32 | 0.220 | &, + |
| **WT1_60min** | NPD07 | yFMP013 | 60 min M.SssI methylation using standard conditions | 328 | 5312 | 18.4 | 0.32 | 0.217 | &, + |
| **WT2_120min** | NPD07 | yFMP013 | 120 min M.SssI methylation using standard conditions | 1098 | 4566 | 62.9 | 0.31 | 0.265 | + |
| **WT2_60min** | NPD07 | yFMP013 | 60 min M.SssI methylation using standard conditions | 694 | 5513 | 39.5 | 0.32 | 0.247 | + |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **WT1_150_120** | NPD08 | yFMP013 | Sample methylated with 150 units of M.SssI for 120 min | 204 | 4334 | 11.1 | 0.30 | 0.224 | & |
| **WT1_150_180** | NPD08 | yFMP013 | Sample methylated with 150 units of M.SssI for 180 min | 197 | 3268 | 11.1 | 0.29 | 0.232 | |
| **WT1_150_30** | NPD08 | yFMP013 | Sample methylated with 150 units of M.SssI for 30 min | 187 | 5322 | 10.1 | 0.30 | 0.179 | §, & |
| **WT1_150_60** | NPD08 | yFMP013 | Sample methylated with 150 units of M.SssI for 60 min | 164 | 5866 | 8.9 | 0.30 | 0.206 | §, &, + |
| **WT1_25_120** | NPD08 | yFMP013 | Sample methylated with 25 units of M.SssI for 120 min | 169 | 3783 | 9.2 | 0.30 | 0.106 | & |
| **WT1_25_180** | NPD08 | yFMP013 | Sample methylated with 25 units of M.SssI for 180 min | 229 | 3728 | 12.8 | 0.29 | 0.114 | & |
| **WT1_50_120** | NPD08 | yFMP013 | Sample methylated with 50 units of M.SssI for 120 min | 120 | 4354 | 6.6 | 0.30 | 0.164 | & |
| **WT1_50_180** | NPD08 | yFMP013 | Sample methylated with 50 units of M.SssI for 180 min | 248 | 4074 | 13.8 | 0.29 | 0.176 | §, & |
| **WT1_50_30** | NPD08 | yFMP013 | Sample methylated with 50 units of M.SssI for 30 min | 164 | 5862 | 8.8 | 0.31 | 0.119 | §, & |
| **WT1_50_60** | NPD08 | yFMP013 | Sample methylated with 50 units of M.SssI for 60 min | 186 | 5132 | 10.2 | 0.30 | 0.145 | §, & |
| **WT2_50_30** | NPD08 | yFMP013 | Sample methylated with 50 units of M.SssI for 30 min. Later timepoints show degraded gDNA | 185 | 7024 | 10.1 | 0.30 | 0.132 | §, & |
| **WT2_50_60** | NPD08 | yFMP013 | Sample methylated with 50 units of M.SssI for 60 min. Later timepoints show degraded gDNA | 277 | 6261 | 15.6 | 0.29 | 0.166 | §, & |
| **WT_TC1_120** | NPD09 | yFMP013 | Timecourse (TC) experiment. Sample methylated for 120 min | 137 | 7296 | 7.2 | 0.32 | 0.262 | |
| **WT_TC1_180** | NPD09 | yFMP013 | Timecourse (TC) experiment. Sample methylated for 180 min | 165 | 6505 | 8.8 | 0.31 | 0.283 | |
| **WT_TC1_60** | NPD09 | yFMP013 | Timecourse (TC) experiment. Sample methylated for 60 min | 121 | 7543 | 6.4 | 0.32 | 0.240 | |
| **WT_TC5_120** | NPD09 | yFMP013 | Timecourse (TC) experiment. Sample methylated for 120 min | 162 | 4815 | 8.7 | 0.31 | 0.208 | |
| **WT_TC5_180** | NPD09 | yFMP013 | Timecourse (TC) experiment. Sample methylated for 180 min | 239 | 4624 | 12.9 | 0.31 | 0.252 | |
| **WT_TC5_60** | NPD09 | yFMP013 | Timecourse (TC) experiment. Sample methylated for 60 min | 93 | 7067 | 4.9 | 0.32 | 0.188 | §, & |
| **TKO1_HD_45** | NPD10 | yFMP356 | Histone Depleted (HD) TKO sample methylated for 45 min | 415 | 7952 | 23.0 | 0.31 | 0.516 | |
| **TKO1_HD_90** | NPD10 | yFMP356 | Histone Depleted (HD) TKO sample methylated for 90 min | 413 | 6795 | 23.0 | 0.30 | 0.547 | |
| **TKO2_HD_45** | NPD10 | yFMP357 | Histone Depleted (HD) TKO sample methylated for 45 min | 447 | 7791 | 25.0 | 0.30 | 0.529 | |
| **TKO2_HD_90** | NPD10 | yFMP357 | Histone Depleted (HD) TKO sample methylated for 90 min | 444 | 6793 | 24.7 | 0.30 | 0.563 | |
| **TKO3_0** | NPD10 | yFMP014 | Unmethylated TKO control experiment | 234 | 6060 | 12.9 | 0.31 | 0.053 | §, & |
| **TKO3_45** | NPD10 | yFMP014 | Sample methylated for 45 min. Referred to as T3.1 | 503 | 9197 | 27.9 | 0.31 | 0.300 | |
| **TKO3_90** | NPD10 | yFMP014 | Sample methylated for 90 min. Referred to as T3.2 | 500 | 7583 | 27.7 | 0.31 | 0.325 | |
| **WT1_HD_45** | NPD10 | yFMP358 | Histone Depleted (HD) WT sample methylated for 45 min. | 554 | 6392 | 28.0 | 0.30 | 0.556 | |
| **WT1_HD_90** | NPD10 | yFMP358 | Histone Depleted (HD) WT sample methylated for 90 min. | 525 | 4501 | 24.9 | 0.30 | 0.590 | |
| **WT2_HD_45** | NPD10 | yFMP359 | Histone Depleted (HD) WT sample methylated for 45 min. | 497 | 6902 | 31.1 | 0.31 | 0.560 | |
| **WT2_HD_90** | NPD10 | yFMP359 | Histone Depleted (HD) WT sample methylated for 90 min. | 441 | 5010 | 29.5 | 0.30 | 0.588 | |
| **WT3_0** | NPD10 | yFMP013 | Unmethylated WT control experiment | 451 | 9464 | 25.1 | 0.31 | 0.036 | §, & |
| **WT3_45** | NPD10 | yFMP013 | Sample methylated for 45 min. Referred to as W3.1 | 475 | 9717 | 26.3 | 0.32 | 0.277 | § |
| **WT3_90** | NPD10 | yFMP013 | Sample methylated for 90 min. Referred to as W3.2 | 434 | 9167 | 23.9 | 0.32 | 0.311 | |
| **WT1_HDGAL_30** | NPD11 | yFMP358 | GAL grown HD strain methylated for 30 min | 610 | 7234 | 33.3 | 0.32 | 0.268 | §, & |
| **WT1_HDGAL_90** | NPD11 | yFMP358 | GAL grown HD strain methylated for 90 min | 941 | 2980 | 52.5 | 0.31 | 0.389 | |
| **WT1_HDGAL_FA_30** | NPD11 | yFMP358 | GAL grown HD strain formaldehyde (FA) treated methylated for 30 min. No Mg2+ in buffer during methylation. | 104 | 5478 | 5.3 | 0.35 | 0.309 | § |
| **WT1_HDGAL_FA_90** | NPD11 | yFMP358 | GAL grown HD strain formaldehyde (FA) treated methylated for 90 min. No Mg2+ in buffer during methylation. | 116 | 5055 | 5.9 | 0.35 | 0.379 | |
| **WT1_HDGLU_30** | NPD11 | yFMP358 | Glucose (GLU) grown HD strain methylated for 30 min | 875 | 4578 | 48.8 | 0.31 | 0.474 | § |
| **WT1_HDGLU_90** | NPD11 | yFMP358 | Glucose (GLU) grown HD strain methylated for 90 min | 917 | 1763 | 51.3 | 0.30 | 0.557 | # |
| **WT1_HDGLU_FA_30** | NPD11 | yFMP358 | GLU grown HD strain FA treated methylated for 30 min. No Mg2+ in buffer during methylation | 217 | 5396 | 11.3 | 0.34 | 0.396 | § |
| **WT1_HDGLU_FA_90** | NPD11 | yFMP358 | GLU grown HD strain FA treated methylated for 30 min. No Mg2+ in buffer during methylation | 190 | 5145 | 9.9 | 0.34 | 0.497 | |
| **TKO1_120** | NPD13 | yFMP467 | Unarrested TKO bar1Δ strain methylated for 120 min | 46 | 1960 | 2.4 | 0.39 | 0.228 | |
| **TKO1_120_a** | NPD13 | yFMP467 | Arrested TKO bar1Δ strain methylated for 120 min | 25 | 1396 | 6.8 | 0.39 | 0.236 | |
| **TKO2_120** | NPD13 | yFMP467 | Unarrested TKO bar1Δ strain methylated for 120 min | 61 | 1781 | 3.1 | 0.39 | 0.193 | & |
| **TKO2_120_a** | NPD13 | yFMP467 | Arrested TKO bar1Δ strain methylated for 120 min | 154 | 1838 | 7.4 | 0.39 | 0.211 | |
| **WT1_120** | NPD13 | yFMP466 | Unarrested WT bar1Δ strain methylated for 120 min | 47 | 2280 | 2.4 | 0.39 | 0.158 | |
| **WT1_120_a** | NPD13 | yFMP466 | Arrested WT bar1Δ strain methylated for 120 min | 148 | 1784 | 7.3 | 0.39 | 0.201 | + |
| **WT2_120** | NPD13 | yFMP466 | Unarrested WT bar1Δ strain methylated for 120 min | 47 | 2166 | 2.4 | 0.38 | 0.133 | & |
| **WT2_120_a** | NPD13 | yFMP466 | Arrested WT bar1Δ strain methylated for 120 min | 39 | 1987 | 3.1 | 0.40 | 0.198 | |
| **INO80_rap_60_c1** | NPD20 | yFMP333 | Rap depleted INO80 strain clone 1 methylated for 60 min | 587 | 3893 | 32.8 | 0.33 | 0.357 | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **INO80_rap_60_c2** | NPD20 | yFMP334 | Rap depleted INO80 strain clone 2 methylated for 60 min | 483 | 5328 | 26.7 | 0.33 | 0.357 | |
| **INO80_veh_60_c1** | NPD20 | yFMP333 | Vehicle (veh) treated (non-depleted) INO80 strain (clone 1) methylated for 60 min | 385 | 4912 | 21.5 | 0.33 | 0.371 | |
| **PolII_rap_60_c1** | NPD20 | yFMP266 | Rap depleted Pol II strain clone 1 methylated for 60 min | 564 | 3644 | 31.6 | 0.33 | 0.548 | |
| **PolII_veh_60_c1** | NPD20 | yFMP266 | Veh treated Pol II strain clone 1 methylated for 60 min | 524 | 3761 | 29.2 | 0.33 | 0.461 | |
| **PolII_rap_120_c1** | NPD20 | yFMP266 | Rap depleted Pol II strain clone 1 methylated for 120 min | 594 | 1882 | 32.8 | 0.32 | 0.507 | |
| **PolII_veh_120_c1** | NPD20 | yFMP266 | Veh treated Pol II strain clone 1 methylated for 120 min | 651 | 1949 | 35.7 | 0.33 | 0.419 | |
| **Ino80_rap_120_c1** | NPD22 | yFMP333 | Rap depleted INO80 strain clone 1 methylated for 60 min | 1768 | 4899 | 99.2 | 0.33 | 0.309 | |
| **Ino80_rap_120_c3** | NPD22 | yFMP335 | Rap depleted INO80 strain clone 3 methylated for 60 min | 826 | 5706 | 46.5 | 0.32 | 0.328 | |
| **Ino80_veh_120_c1** | NPD22 | yFMP333 | Vehicle (veh) treated (non-depleted) INO80 strain (clone 1) methylated for 120 min | 2013 | 3844 | 113.0 | 0.32 | 0.368 | |
| **PolII_rap_120_c1** | NPD22 | yFMP266 | Rap depleted Pol II strain clone 1 methylated for 120 min | 1592 | 1390 | 87.9 | 0.32 | 0.483 | # |
| **PolII_rap_120_c3** | NPD22 | yFMP268 | Rap depleted Pol II strain clone 3 methylated for 120 min | 977 | 2479 | 55.1 | 0.32 | 0.527 | # |
| **PolII_veh_120_c1** | NPD22 | yFMP266 | Veh treated Pol II strain clone 1 methylated for 120 min | 1442 | 5134 | 80.7 | 0.33 | 0.335 | |
| **PolII_rap_120_Ash** | NPD22 | yFMP266 | Rap depleted Pol II strain methylated for 120 min. Nuclei generated by Dr. Ashish Singh | 1698 | 4510 | 95.9 | 0.33 | 0.407 | |
| **PolII_veh_120_Ash** | NPD22 | yFMP266 | Veh treated Pol II strain methylated for 120 min. Nuclei generated by Dr. Ashish Singh | 1236 | 6844 | 69.4 | 0.33 | 0.376 | |
| **TKO6_FA_120** | NPD24 | yFMP014 | FA crosslinked TKO sample methylated for 120 minutes. No Mg2+ in buffer during methylation | 599 | 8411 | 30.4 | 0.39 | 0.269 | |
| **WT6_FA_120** | NPD24 | yFMP013 | FA crosslinked WT sample methylated for 60 minutes. No Mg2+ in buffer during methylation | 1065 | 7733 | 55.1 | 0.38 | 0.248 | |
| **WT6_FA_60** | NPD24 | yFMP013 | FA crosslinked WT sample methylated for 120 minutes. No Mg2+ in buffer during methylation | 541 | 7220 | 27.9 | 0.38 | *0.189 | |

### 6.2.2. Read statistics

The statistics for each read (i.e. single molecule) of ≥1000 bp or longer within a sample are calculated (Figure 1.3.2B). We calculate the read accessibility over an entire read by taking the total methylated CpG sites over the total non-ambiguous CpG sites per read. Reads ≥1000 bp with ≤10% or ≥90% accessibility are regarded as under- or overmethylated, respectively. Samples which have a fraction of undermethylated reads >20% are denoted with an ampersand sign (&) in the 'Notes on quality criteria' column. Such samples may not have reached full methylation. For description on the '§' and '#' sign, see Supplementary table 1, for description on the '+' sign see Supplementary table 3

#### Supplementary table 2: Read information per sample

| Sample Name | Seq. run | Read mean accessibility | Read fraction overmethylated | Read fraction undermethylated | Notes on quality criteria |
|---|---|---|---|---|---|
| **WT0_30** | NPD03 | 0.210 | 0.010 | 0.390 | §, & |
| **WT0_60** | NPD03 | 0.370 | 0.080 | 0.160 | § |
| **WT0_180** | NPD03 | 0.540 | 0.230 | 0.020 | § |
| **TKO1_120min** | NPD07 | 0.300 | 0.016 | 0.106 | |
| **TKO1_60min** | NPD07 | 0.303 | 0.019 | 0.132 | |
| **TKO2_120min** | NPD07 | 0.276 | 0.007 | 0.078 | |
| **TKO2_60min** | NPD07 | 0.274 | 0.011 | 0.111 | |
| **WT1_120min** | NPD07 | 0.236 | 0.003 | 0.220 | &, + |
| **WT1_60min** | NPD07 | 0.235 | 0.004 | 0.220 | &, + |
| **WT2_120min** | NPD07 | 0.295 | 0.009 | 0.071 | + |
| **WT2_60min** | NPD07 | 0.275 | 0.008 | 0.116 | + |
| **WT1_150_120** | NPD08 | 0.218 | 0.001 | 0.262 | & |
| **WT1_150_180** | NPD08 | 0.269 | 0.002 | 0.157 | |
| **WT1_150_30** | NPD08 | 0.190 | 0.001 | 0.323 | §, & |

| | | | | | |
|---|---|---|---|---|---|
| **WT1_150_60** | NPD08 | 0.219 | 0.002 | 0.254 | §, &, + |
| **WT1_25_120** | NPD08 | 0.096 | 0.000 | 0.663 | & |
| **WT1_25_180** | NPD08 | 0.119 | 0.000 | 0.539 | & |
| **WT1_50_120** | NPD08 | 0.161 | 0.001 | 0.397 | & |
| **WT1_50_180** | NPD08 | 0.207 | 0.001 | 0.262 | §, & |
| **WT1_50_30** | NPD08 | 0.124 | 0.001 | 0.548 | §, & |
| **WT1_50_60** | NPD08 | 0.153 | 0.001 | 0.427 | §, & |
| **WT2_50_30** | NPD08 | 0.003 | 0.586 | 0.280 | §, & |
| **WT2_50_60** | NPD08 | 0.006 | 0.425 | 0.275 | §, & |
| **WT_TC1_120** | NPD09 | 0.292 | 0.006 | 0.161 | |
| **WT_TC1_180** | NPD09 | 0.315 | 0.005 | 0.131 | |
| **WT_TC1_60** | NPD09 | 0.267 | 0.006 | 0.186 | |
| **WT_TC5_120** | NPD09 | 0.440 | 0.130 | 0.159 | |
| **WT_TC5_180** | NPD09 | 0.475 | 0.206 | 0.122 | |
| **WT_TC5_60** | NPD09 | 0.209 | 0.005 | 0.293 | §, & |
| **TKO1_HD_45** | NPD10 | 0.529 | 0.018 | 0.029 | |
| **TKO1_HD_90** | NPD10 | 0.572 | 0.044 | 0.019 | |
| **TKO2_HD_45** | NPD10 | 0.541 | 0.021 | 0.029 | |
| **TKO2_HD_90** | NPD10 | 0.586 | 0.048 | 0.018 | |
| **TKO3_0** | NPD10 | 0.053 | 0.002 | 0.950 | §, & |
| **TKO3_45** | NPD10 | 0.321 | 0.007 | 0.065 | |
| **TKO3_90** | NPD10 | 0.362 | 0.027 | 0.049 | |
| **WT1_HD_45** | NPD10 | 0.581 | 0.033 | 0.035 | |
| **WT1_HD_90** | NPD10 | 0.618 | 0.058 | 0.021 | |
| **WT2_HD_45** | NPD10 | 0.580 | 0.030 | 0.033 | |
| **WT2_HD_90** | NPD10 | 0.623 | 0.058 | 0.018 | |
| **WT3_0** | NPD10 | 0.036 | 0.000 | 0.986 | §, & |
| **WT3_45** | NPD10 | 0.297 | 0.002 | 0.118 | § |
| **WT3_90** | NPD10 | 0.344 | 0.020 | 0.072 | |
| **WT1_HDGAL_30** | NPD11 | 0.292 | 0.013 | 0.275 | §, & |
| **WT1_HDGAL_90** | NPD11 | 0.438 | 0.019 | 0.042 | |
| **WT1_HDGAL_FA_30** | NPD11 | 0.325 | 0.002 | 0.051 | § |
| **WT1_HDGAL_FA_90** | NPD11 | 0.396 | 0.002 | 0.026 | |
| **WT1_HDGLU_30** | NPD11 | 0.495 | 0.026 | 0.092 | § |
| **WT1_HDGLU_90** | NPD11 | 0.584 | 0.041 | 0.025 | # |
| **WT1_HDGLU_FA_30** | NPD11 | 0.410 | 0.002 | 0.032 | § |
| **WT1_HDGLU_FA_90** | NPD11 | 0.511 | 0.003 | 0.016 | |
| **TKO1_120** | NPD13 | 0.237 | 0.000 | 0.133 | |
| **TKO1_120_a** | NPD13 | 0.243 | 0.000 | 0.151 | |
| **TKO2_120** | NPD13 | 0.204 | 0.000 | 0.215 | & |
| **TKO2_120_a** | NPD13 | 0.220 | 0.000 | 0.192 | |
| **WT1_120** | NPD13 | 0.170 | 0.000 | 0.307 | |
| **WT1_120_a** | NPD13 | 0.212 | 0.000 | 0.185 | + |
| **WT2_120** | NPD13 | 0.144 | 0.000 | 0.440 | & |
| **WT2_120_a** | NPD13 | 0.212 | 0.000 | 0.196 | |
| **INO80_rap_60_c1** | NPD20 | 0.393 | 0.015 | 0.015 | |
| **INO80_rap_60_c2** | NPD20 | 0.398 | 0.012 | 0.012 | |
| **INO80_veh_60_c1** | NPD20 | 0.423 | 0.025 | 0.012 | |
| **PolII_rap_60_c1** | NPD20 | 0.611 | 0.100 | 0.008 | |

| | | | | | |
|---|---|---|---|---|---|
| **PolII_veh_60_c1** | NPD20 | 0.543 | 0.083 | 0.016 | |
| **PolII_rap_120_c1** | NPD20 | 0.556 | 0.076 | 0.008 | |
| **PolII_veh_120_c1** | NPD20 | 0.466 | 0.046 | 0.021 | |
| **Ino80_rap_120_c1** | NPD22 | 0.347 | 0.008 | 0.030 | |
| **Ino80_rap_120_c3** | NPD22 | 0.365 | 0.014 | 0.031 | |
| **Ino80_veh_120_c1** | NPD22 | 0.418 | 0.026 | 0.014 | |
| **PolII_rap_120_c1** | NPD22 | 0.512 | 0.074 | 0.026 | # |
| **PolII_rap_120_c3** | NPD22 | 0.580 | 0.100 | 0.014 | # |
| **PolII_veh_120_c1** | NPD22 | 0.383 | 0.016 | 0.020 | |
| **PolII_rap_120_Ash** | NPD22 | 0.482 | 0.080 | 0.022 | |
| **PolII_veh_120_Ash** | NPD22 | 0.466 | 0.085 | 0.024 | |
| **TKO6_FA_120** | NPD24 | 0.289 | 0.000 | 0.029 | |
| **WT6_FA_120** | NPD24 | 0.284 | 0.000 | 0.021 | |
| **WT6_FA_60** | NPD24 | 0.215 | 0.000 | 0.091 | |

### 6.2.3. rDNA and genome statistics

For each sample the fraction of accessible (≥50% methylated) rDNA reads are calculated for each rDNA locus (RDN37-1 and RDN37-2). We calculate the median rDNA and genome lengths for comparisons and to detect the effect of nuclease activity. In addition, we calculate the mean coverage of all sites over the NFR (coordinates -225 to -75 relative to the +1 nucleosome dyad) and mean coverage of all sites over the gene body region (GBR; coordinates 0 to 1000 relative to the +1 nucleosome). Site coverage over the rDNA is calculated based on the RDN37-1 locus. The rDNA:GBR ratio indicates the fold coverage with which the rDNA loci is sequenced compared to the mean GBR coverage. Samples in which the NFR site coverage is >8% less compared to GBR site coverage are denoted with a plus sign (+) in the 'Notes on quality criteria' column. Cutoff is determined based on site coverage difference detected in bottom 10% of all samples. This discrepancy could indicate a potential loss of NFR coverage due to nuclease activity. For description on the '§' and '#' sign, see Supplementary table 1, for description on the '&' sign see Supplementary table 2.

**Supplementary table 3: Global statistics on rDNA, NFRs and GBRs**

| Sample Name | Seq. run | rDNA fraction accessible left (RDN37-1) | rDNA fraction accessible right (RDN37-2) | rDNA median length | Genome median length | Mean site coverage NFR | Mean site coverage GBR | Mean site coverage rDNA | rDNA:GBR ratio | Notes on quality criteria |
|---|---|---|---|---|---|---|---|---|---|---|
| **TKO1_120min** | NPD07 | 0.45 | 0.40 | 4691 | 2540 | 76.5 | 84.5 | 4706 | 55.7 | |
| **TKO1_60min** | NPD07 | 0.44 | 0.40 | 7414 | 3622 | 37.8 | 41.1 | 3365 | 81.8 | |
| **TKO2_120min** | NPD07 | 0.57 | 0.51 | 4600 | 2737 | 71.7 | 76.8 | 2699 | 35.1 | |
| **TKO2_60min** | NPD07 | 0.57 | 0.51 | 5989 | 2704 | 41.3 | 44.6 | 2563 | 57.5 | |
| **WT1_120min** | NPD07 | 0.35 | 0.29 | 5027 | 2710 | 32.9 | 37.0 | 3204 | 86.6 | &, + |
| **WT1_60min** | NPD07 | 0.29 | 0.27 | 5584 | 2760 | 22.0 | 24.6 | 2202 | 89.5 | &, + |
| **WT2_120min** | NPD07 | 0.26 | 0.24 | 6026 | 3155 | 73.7 | 84.4 | 8165 | 96.7 | + |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **WT2_60min** | NPD07 | 0.27 | 0.24 | 7163 | 3446 | 45.8 | 52.0 | 6255 | 120.4 | + |
| **WT1_150_120** | NPD08 | 0.29 | 0.25 | 5830 | 1626 | 14.8 | 15.0 | 1244 | 82.8 | & |
| **WT1_150_180** | NPD08 | 0.32 | 0.28 | 3993 | 1373 | 15.1 | 15.2 | 1060 | 69.7 | |
| **WT1_150_30** | NPD08 | 0.21 | 0.19 | 6176 | 2000 | 12.2 | 13.1 | 1002 | 76.3 | §, & |
| **WT1_150_60** | NPD08 | 0.26 | 0.22 | 5386 | 2219 | 10.5 | 11.6 | 952 | 82.2 | §, &, + |
| **WT1_25_120** | NPD08 | 0.09 | 0.06 | 5190 | 1414 | 12.9 | 12.9 | 845 | 65.6 | & |
| **WT1_25_180** | NPD08 | 0.10 | 0.09 | 5290 | 1508 | 17.0 | 17.4 | 1343 | 77.1 | & |
| **WT1_50_120** | NPD08 | 0.21 | 0.18 | 6027 | 1704 | 9.0 | 9.0 | 678 | 75.4 | & |
| **WT1_50_180** | NPD08 | 0.25 | 0.22 | 5037 | 1650 | 18.3 | 18.7 | 1431 | 76.4 | §, & |
| **WT1_50_30** | NPD08 | 0.08 | 0.06 | 6652 | 2306 | 10.5 | 11.3 | 1005 | 88.6 | §, & |
| **WT1_50_60** | NPD08 | 0.14 | 0.12 | 5702 | 1927 | 12.0 | 13.1 | 1145 | 87.6 | §, & |
| **WT2_50_30** | NPD08 | 0.07 | 0.07 | 7434 | 3830 | 12.6 | 13.5 | 970 | 72.1 | §, & |
| **WT2_50_60** | NPD08 | 0.12 | 0.12 | 7058 | 3738 | 19.4 | 20.6 | 1357 | 65.7 | §, & |
| **WT_TC1_120** | NPD09 | 0.34 | 0.29 | 7118 | 3750 | 8.8 | 9.5 | 841 | 88.5 | |
| **WT_TC1_180** | NPD09 | 0.36 | 0.32 | 6338 | 3298 | 10.7 | 11.7 | 948 | 81.2 | |
| **WT_TC1_60** | NPD09 | 0.31 | 0.27 | 6802 | 3664 | 7.8 | 8.5 | 675 | 79.8 | |
| **WT_TC5_120** | NPD09 | 0.27 | 0.24 | 7053 | 2208 | 12.7 | 12.3 | 979 | 79.5 | |
| **WT_TC5_180** | NPD09 | 0.34 | 0.31 | 6203 | 2203 | 18.5 | 18.2 | 1376 | 75.6 | |
| **WT_TC5_60** | NPD09 | 0.21 | 0.18 | 7318 | 3063 | 6.0 | 6.6 | 582 | 88.9 | §, & |
| **TKO1_HD_45** | NPD10 | 0.87 | 0.87 | 8955 | 7985 | 31.3 | 32.2 | 1135 | 35.2 | |
| **TKO1_HD_90** | NPD10 | 0.91 | 0.90 | 7878 | 6455 | 31.6 | 32.2 | 1048 | 32.6 | |
| **TKO2_HD_45** | NPD10 | 0.88 | 0.87 | 9383 | 8209 | 33.9 | 34.6 | 1177 | 34.0 | |
| **TKO2_HD_90** | NPD10 | 0.91 | 0.91 | 6243 | 6084 | 33.9 | 34.7 | 1110 | 32.0 | |
| **TKO3_0** | NPD10 | 0.04 | 0.04 | 7738 | 3715 | 17.0 | 17.8 | 882 | 49.6 | §, & |
| **TKO3_45** | NPD10 | 0.38 | 0.35 | 11349 | 9964 | 35.4 | 37.6 | 3348 | 88.9 | |
| **TKO3_90** | NPD10 | 0.41 | 0.39 | 10460 | 7688 | 35.6 | 37.7 | 3223 | 85.4 | |
| **WT1_HD_45** | NPD10 | 0.86 | 0.86 | 8554 | 6312 | 35.9 | 38.1 | 2607 | 68.4 | |
| **WT1_HD_90** | NPD10 | 0.89 | 0.89 | 5779 | 3378 | 32.6 | 34.2 | 2005 | 58.7 | |
| **WT2_HD_45** | NPD10 | 0.85 | 0.87 | 9121 | 6414 | 41.0 | 42.6 | 2428 | 56.9 | |
| **WT2_HD_90** | NPD10 | 0.89 | 0.90 | 6802 | 3807 | 39.3 | 41.1 | 2000 | 48.6 | |
| **WT3_0** | NPD10 | 0.01 | 0.01 | 10698 | 9979 | 32.5 | 34.2 | 2347 | 68.6 | §, & |
| **WT3_45** | NPD10 | 0.35 | 0.35 | 10768 | 10380 | 33.9 | 35.8 | 2765 | 77.3 | § |
| **WT3_90** | NPD10 | 0.38 | 0.38 | 10425 | 9854 | 31.1 | 32.7 | 2468 | 75.6 | |
| **WT1_HDGAL_30** | NPD11 | 0.31 | 0.31 | 8580 | 4139 | 43.9 | 46.5 | 2951 | 63.4 | §, & |
| **WT1_HDGAL_90** | NPD11 | 0.50 | 0.52 | 5046 | 1693 | 69.9 | 74.9 | 2422 | 32.4 | |
| **WT1_HDGAL_FA_30** | NPD11 | 0.48 | 0.45 | 7468 | 3470 | 7.6 | 7.9 | 377 | 47.8 | § |
| **WT1_HDGAL_FA_90** | NPD11 | 0.58 | 0.54 | 7671 | 2820 | 8.3 | 8.7 | 391 | 45.0 | |
| **WT1_HDGLU_30** | NPD11 | 0.66 | 0.67 | 6774 | 3269 | 65.1 | 68.9 | 3656 | 53.1 | § |
| **WT1_HDGLU_90** | NPD11 | 0.78 | 0.81 | 2441 | 1141 | 68.4 | 73.2 | 1392 | 19.0 | # |
| **WT1_HDGLU_FA_30** | NPD11 | 0.62 | 0.57 | 7609 | 4190 | 16.0 | 16.6 | 851 | 51.2 | § |
| **WT1_HDGLU_FA_90** | NPD11 | 0.80 | 0.79 | 7317 | 3120 | 13.9 | 14.4 | 738 | 51.4 | |
| **TKO1_120** | NPD13 | 0.51 | 0.36 | 2649 | 996 | 3.7 | 4.0 | 68 | 16.9 | |
| **TKO1_120_a** | NPD13 | 0.65 | 0.54 | 1300 | 661 | 10.2 | 11.0 | 138 | 12.5 | |
| **TKO2_120** | NPD13 | 0.53 | 0.31 | 2058 | 1165 | 5.0 | 5.2 | 86 | 16.4 | & |
| **TKO2_120_a** | NPD13 | 0.57 | 0.39 | 2272 | 1106 | 11.9 | 12.3 | 161 | 13.1 | |
| **WT1_120** | NPD13 | 0.22 | 0.12 | 2554 | 1338 | 3.8 | 4.1 | 135 | 33.1 | |
| **WT1_120_a** | NPD13 | 0.43 | 0.32 | 1237 | 950 | 10.6 | 11.8 | 362 | 30.8 | + |
| **WT2_120** | NPD13 | 0.18 | 0.12 | 2385 | 1311 | 3.8 | 4.1 | 114 | 28.0 | & |
| **WT2_120_a** | NPD13 | 0.49 | 0.36 | 2189 | 1300 | 5.1 | 5.2 | 110 | 20.9 | |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **INO80_rap_60_c1** | NPD20 | 0.74 | 0.66 | 5519 | 2629 | 45.1 | 48.1 | 2078 | 43.2 | |
| **INO80_rap_60_c2** | NPD20 | 0.82 | 0.75 | 7778 | 3338 | 37.0 | 39.2 | 1927 | 49.1 | |
| **INO80_veh_60_c1** | NPD20 | 0.78 | 0.74 | 6800 | 3083 | 29.2 | 31.1 | 1712 | 55.1 | |
| **PolII_rap_60_c1** | NPD20 | 0.82 | 0.81 | 4782 | 2268 | 43.0 | 45.5 | 2378 | 52.2 | |
| **PolII_veh_60_c1** | NPD20 | 0.78 | 0.73 | 5037 | 2262 | 40.3 | 42.4 | 1893 | 44.6 | |
| **PolII_rap_120_c1** | NPD20 | 0.75 | 0.71 | 2576 | 1165 | 46.5 | 48.9 | 925 | 18.9 | |
| **PolII_veh_120_c1** | NPD20 | 0.64 | 0.62 | 2490 | 1174 | 51.5 | 53.9 | 696 | 12.9 | |
| **Ino80_rap_120_c1** | NPD22 | 0.69 | 0.61 | 5783 | 3155 | 135.8 | 144.0 | 6491 | 45.1 | |
| **Ino80_rap_120_c3** | NPD22 | 0.70 | 0.62 | 8647 | 3503 | 64.2 | 66.9 | 2826 | 42.2 | |
| **Ino80_veh_120_c1** | NPD22 | 0.75 | 0.70 | 5066 | 2496 | 154.9 | 164.4 | 6896 | 41.9 | |
| **PolII_rap_120_c1** | NPD22 | 0.69 | 0.67 | 2168 | 871 | 121.2 | 130.9 | 2221 | 17.0 | # |
| **PolII_rap_120_c3** | NPD22 | 0.79 | 0.77 | 3576 | 1587 | 74.4 | 79.9 | 3491 | 43.7 | # |
| **PolII_veh_120_c1** | NPD22 | 0.73 | 0.69 | 5364 | 3085 | 111.4 | 117.8 | 4757 | 40.4 | |
| **PolII_rap_120_Ash** | NPD22 | 0.80 | 0.76 | 5088 | 2991 | 127.0 | 136.9 | 8606 | 62.9 | |
| **PolII_veh_120_Ash** | NPD22 | 0.82 | 0.78 | 7752 | 4619 | 92.0 | 98.5 | 7299 | 74.1 | |
| **TKO6_FA_120** | NPD24 | 0.44 | 0.42 | 9828 | 5778 | 44.2 | 46.3 | 3206 | 69.3 | |
| **WT6_FA_120** | NPD24 | 0.45 | 0.38 | 8704 | 5761 | 75.4 | 82.0 | 8328 | 101.5 | |
| **WT6_FA_60** | NPD24 | 0.31 | 0.24 | 8145 | 5279 | 38.5 | 41.7 | 3997 | 96.0 | |

## 7. ACKNOWLEDGEMENTS

I’d like to thank Prof. Dr. Felix Müller-Planitz for his daily supervision and mentorship. I’ve learned more than I could ever imagine in these last years. Your attention to detail and perseverance have been formative on my perspective and approach to scientific -and non-scientific- problems. Thanks for always providing highly detailed and constructive feedback.

I would further like to thanks my “Doktorvater”, Prof. Dr. Peter Becker, for his supervision on my thesis and his valuable feedback. As head of the Molecular Biology department, you have created an excellent environment for scientists of all stages to prosper. I am lucky to have ended up in such a place that allowed me to grow uninhibited.

Many thanks go out to my thesis advisory committee: Prof. Dr. Peter Becker, Prof. Dr. Ulrich Gerland and Prof. Dr. Felix Müller-Planitz. The yearly meetings we had were often critical for setting out long-term goals and keeping goals realistic but ambitious.

To all former and present members of the Müller-Planitz lab: Ashish, Petra and Kripi, I will never forget the fruitful daily discussions we could freely have and the fun after-work talks. These have been imperative for my scientific, mental and psychological wellbeing. I’d also like to thank Jessica, who has been patient with my supervision during her bachelor and master thesis in the lab, and has directly contributed to this thesis. Also, many thanks to Sylvia and Madhura, who have often provided key technical support for me to continue experiments, without delay.

My thanks further extend beyond the lab to my collaborators: Dr. Michael Wolff, Dr. Maryam Khatami and Matthias Hanke. Your contributions have helped me elucidate many scientific questions by performing often difficult and time-consuming computational analyses. Your combined knowledge has also aided me greatly in learning to perform such analyses myself and propelling our findings to a next level.

Thanks to the SFB IRTG1064 and Dr. Elizabeth Schroeder-Reiter for a great graduate program at the BMC. Also, many thanks to the IMPRS-LS graduate program and Dr. Hans Joerg Schaeffer, Dr. Ingrid Wolf, Maximiliane Reif and Dr. Marta Cipińska.

Many thanks go out to my friends in Germany: Michael, Vera, David, Niklas, Alan, Denise, Juri, Elli, Tobias, Lara. You all have made my time in abroad especially great. Without you as friends these years wouldn’t have flown by. A special thanks to Michael and David, who have been there for me from day one.

Not to forget, many thanks to Annick and Geert. You have always made me happy to be back in The Netherlands, even though visits were short and sometimes infrequent. Last but not least, thanks go out to my family: Peer, Maureen, Amber, Job, Peter, Frans, Opa en Oma. Thanks for being there and your unconditional support.