



HAL
open science

Coupling between Sequence-Mediated Nucleosome Organization and Genome Evolution

Jérémy Barbier, Cédric Vaillant, Jean-Nicolas Volff, Frédéric Brunet, B. Audit

► **To cite this version:**

Jérémy Barbier, Cédric Vaillant, Jean-Nicolas Volff, Frédéric Brunet, B. Audit. Coupling between Sequence-Mediated Nucleosome Organization and Genome Evolution. *Genes*, 2021, 12 (6), pp.851. 10.3390/genes12060851 . hal-03454763

HAL Id: hal-03454763

<https://hal.science/hal-03454763v1>

Submitted on 29 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Review

Coupling between Sequence-Mediated Nucleosome Organization and Genome Evolution

Jérémy Barbier ^{1,2} , Cédric Vaillant ² , Jean-Nicolas Volff ^{1,*} , Frédéric G. Brunet ¹  and Benjamin Audit ^{2,*} 

¹ Institut de Génomique Fonctionnelle de Lyon, Univ Lyon, CNRS UMR 5242, Ecole Normale Supérieure de Lyon, Univ Claude Bernard Lyon 1, F-69364 Lyon, France; jeremy.barbier@ens-lyon.fr (J.B.); frederic.brunet@ens-lyon.fr (F.G.B.)

² Laboratoire de Physique, Univ Lyon, ENS de Lyon, CNRS, F-69342 Lyon, France; cedric.vaillant@ens-lyon.fr

* Correspondence: jean-nicolas.volff@ens-lyon.fr (J.-N.V.); benjamin.audit@ens-lyon.fr (B.A.)

Abstract: The nucleosome is a major modulator of DNA accessibility to other cellular factors. Nucleosome positioning has a critical importance in regulating cell processes such as transcription, replication, recombination or DNA repair. The DNA sequence has an influence on the position of nucleosomes on genomes, although other factors are also implicated, such as ATP-dependent remodelers or competition of the nucleosome with DNA binding proteins. Different sequence motifs can promote or inhibit the nucleosome formation, thus influencing the accessibility to the DNA. Sequence-encoded nucleosome positioning having functional consequences on cell processes can then be selected or counter-selected during evolution. We review the interplay between sequence evolution and nucleosome positioning evolution. We first focus on the different ways to encode nucleosome positions in the DNA sequence, and to which extent these mechanisms are responsible of genome-wide nucleosome positioning *in vivo*. Then, we discuss the findings about selection of sequences for their nucleosomal properties. Finally, we illustrate how the nucleosome can directly influence sequence evolution through its interactions with DNA damage and repair mechanisms. This review aims to provide an overview of the mutual influence of sequence evolution and nucleosome positioning evolution, possibly leading to complex evolutionary dynamics.

Keywords: DNA sequence-encoded nucleosome ordering; nucleosome depleted regions; DNA sequence mutation; chromatin evolution



Citation: Barbier, J.; Vaillant, C.; Volff, J.-N.; Brunet, F.G.; Audit, B. Coupling between Sequence-Mediated Nucleosome Organization and Genome Evolution. *Genes* **2021**, *12*, 851. <https://doi.org/10.3390/genes12060851>

Academic Editor: Manuel A. Garrido-Ramos

Received: 1 May 2021

Accepted: 27 May 2021

Published: 1 June 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

To fit in the nucleus of each cell, eukaryotic DNA needs to be highly compacted. This compaction is achieved by the formation of a protein-DNA complex called chromatin [1]. The first level of compaction consists of the wrapping of ~146 bp of DNA around an octamer of four core histone proteins (H2A, H2B, H3 and H4), forming a nucleosome [2]. In the nucleosome, the DNA is wrapped almost twice around the core histone octamer (a tetramer of (H3-H4)₂ flanked by two dimers of H2A-H2B), with contact points between DNA and the histone proteins every ~10 bp [3,4]. The mid-point of the complexed DNA is called the dyad, and serves as a reference to specify nucleosome positions. The nucleosome repeat length (NRL), that represents the distance between two consecutive nucleosome dyads, ranges from 155 bp in fission yeast [5] to about 240 bp in echinoderm sperm [6]. Taking into account the length of DNA wrapped in each nucleosomes, there is thus a high density of nucleosome in living cells regardless of the cell type or organism, with at least two third of the genome participating in a nucleosome. Nucleosomes come in several forms. Core histones may carry post-translational modifications (PTMs), such as methylation, acetylation or phosphorylation occurring mostly in the N-terminal tail of histones (e.g., tri-methylation of histone H3 lysine 9, also known as H3K9me3). Histone cores may also contain histone variants, which are alternative histone proteins encoded by genes that appeared throughout the evolution of Eukaryotes [4,7,8]. PTMs and histone variants are

associated with different chromatin states of genome compaction and genome regulation and have thus received most of the attention in chromatin biology studies. Nevertheless, the precise position of nucleosomes on the DNA is also of great importance [1]. Indeed, the accessibility of DNA to non-histone chromatin factors like transcription and replication factors is modulated by nucleosome occupancy, with nucleosomal DNA being considerably less accessible to these factors than the naked “linker” DNA between nucleosomes. From a collective perspective, the position of nucleosomes relative to each other is also associated to chromatin state, probably in relation to higher order chromatin compaction. Indeed, actively transcribed genomes where chromatin needs to be open and accessible tend to have shorter NRL (ranging from 160 to 189 bp in yeast, embryonic stem cells and tumour cells for example) than transcriptionally inactive genomes (NRL ranging from 190 to 240 in chicken erythrocytes and echinoderm sperm for example) [9]. This distinction has also been made within the human genome, where the NRL of active genes is way shorter (178 bp) than the NRL of repressed or heterochromatic non-coding sequences (206 bp) [10]. However, there are exceptions to this rule. For example, in higher eukaryotes, telomeric DNA is packaged in nucleosomes with a NRL 20–40 bp shorter than the NRL of bulk nucleosome [11]. This has been observed in vertebrates [12–16] but also in sea urchin [16], and several plant species [17–19]. The position of nucleosomes on the DNA and relative to each other is thus crucial for genetic functions, because it modulates the efficiency of trans-acting factors such as the transcription machinery [1,2,20]. Nucleosomal positioning on DNA depends on various factors, including DNA sequence effects, competition for DNA such as with transcription factors, and remodeling by ATP-dependent enzyme [21]. Notably, the DNA sequence has an important contribution to nucleosomal positioning at the genome scale [1,10,21,22]. Nucleosome positions are thus to some significant extent a sequence-encoded feature that have a functional role in genomes (as modulator of the accessibility to DNA). As other sequence-encoded functional features (such as genes), nucleosome positions can then be selected during evolution. In other words, sequences could be selected not for their direct coding properties as genes, but for their abilities to favor or impair nucleosome formation at specific loci, directly impacting their accessibility to external regulatory factors. Selection of sequences for their nucleosomal affinity has been described in several species such as yeasts [23,24] but also in more complex organisms like maize [25,26] or human [27–29]. Note that the repositioning of nucleosomes according to the evolution of sequences can also occur in a neutral scenario leading to possible drifts of nucleosome positions [30]. Interestingly, the nucleosome itself also shapes the evolution of sequences by interacting with DNA damage and repair mechanisms, leading to biased mutational patterns inside and around nucleosomes [31]. Here, we will review some of the findings about these mechanisms, focusing first on how nucleosome positions are encoded in the DNA sequence, then on how sequence nucleosomal properties have been selected during evolution, and finally, on how the nucleosome directly modulates mutational patterns. This provides an opportunity to discuss the mutual feedback between the evolution of DNA sequence and chromatin organization at a genomic scale.

2. How Is Nucleosome Positioning Encoded in the DNA Sequence?

2.1. DNA Sequence Does Influence Nucleosome Positioning

Using SELEX (Systematic Evolution of Ligands by EXponential enrichment) experiments on synthetic and genomic DNA with the core histone proteins as ligands, it was shown that the DNA sequence does influence the affinity of a DNA fragment for histones up to a 5000-fold range [32–35]. In such experiments, an excess of DNA fragments of variable sequence compete for a ligand. The DNA-ligand complexes are then extracted, DNA fragments are purified, amplified and brought back into competition with the same ligand, a process repeated several times to purify sequences with the highest affinities for the ligand of interest. Lowary and Widom used this approach with synthetic DNA fragments and core histone proteins as ligands to select from a random set of sequences the ones with the highest affinities for the nucleosome [33]. It revealed the existence of sequences with un-

expectedly high affinity for the histone octamer. Similar experiments were also performed with fragments extracted from genomic DNA. It showed that their affinity for histones had a much narrower range than random DNA fragments [32,35]. These experiments clearly indicate that the DNA sequence matters on how easily a nucleosome can be formed and so where nucleosomes are intrinsically positioned along chromosomes. The sequence-encoded nucleosome positioning can therefore be seen as a basal “ground state” that can be “remodeled” in vivo by the site-specific recruitment and (energy consuming) action of trans-acting factors to establish at proper times and positions an “epigenetic” reversible nucleosome positioning pattern, either permissive or repressive for genome activity. As demonstrated by Parmar et al. [36] when considering a composite model of nucleosome positioning that accounts for both sequence effects and ATP-dependent remodelers and as evidenced by experiments [37], sequence effects are indeed sufficiently strong to control the first steps of the relaxation dynamics of the nucleosomal array after strong perturbation, i.e., in a transient phase of non or weak activity of remodelers. Strikingly, nucleosomal pattern in germ cells where remodelers activity is reduced has been shown to be mostly controlled by the DNA sequence [38]. In vitro nucleosome reconstitution experiments on the yeast genome further demonstrated that ATP was required to obtain a nucleosome positioning pattern that deviate from the sequence encoded pattern and resemble the native pattern [39]. All these results suggest that the primary sequence is a parameter that needs to be taken into account in nucleosome positioning studies, even if sequence effects can be refined or even overridden in vivo by other factors such as ATP-dependent remodelers.

Technical progresses made it possible to decipher DNA sequence-mediated effects genome-wide, mainly with experiments such as MNase-seq, in which the chromatin is digested with an enzyme (the micrococcal nuclease, MNase) that cuts and digests the naked linker DNA between nucleosomes [40–42]. After histone removal, the remaining DNA can be sequenced with high-throughput sequencing techniques, and the alignment of the reads on the reference genome provides information about the genome-wide positioning of nucleosomes [10,42–47]. Such genome-wide mapping of nucleosomes has been established in vivo in various species, including yeast [43,47,48], human [10,44,45], fly [49], plants [25,50], mouse [51], and the nematode *Caenorhabditis elegans* [52], but also in vitro [10,53]. The availability of such experimental data has been reviewed by Teif [54]. Comparison of in vivo and in vitro nucleosome maps revealed a high consistency between in vitro and in vivo genome-wide positioning of nucleosomes [10,53,55]. These results showed that the sequence effects are relevant even in vivo in the presence of external factors influencing nucleosomal positioning. Indeed, the sequence-directed nucleosome positioning is directly observed from in vitro data, because chromatin is reconstituted from DNA and histones only, without any other external factors such as remodelers found in vivo. Accordingly, models established from in vitro genome-wide reconstitution of chromatin predict rather well in vivo nucleosome positioning [22,53,55–60], corroborating the hypothesis that the DNA sequence plays a major role among the different factors influencing the position of nucleosomes [61]. During the past 40 years, attempts to describe the sequence-directed nucleosomal positioning showed that one needs to consider two types of mechanisms (Figure 1): (i) positioning mechanisms where DNA motifs at specific location accommodate DNA wrapping in the nucleosome, for example by favoring certain dinucleotides at contact points between DNA and histones; and (ii) inhibiting mechanisms, with sequences such as poly(dA:dT) preventing nucleosome formation [1].

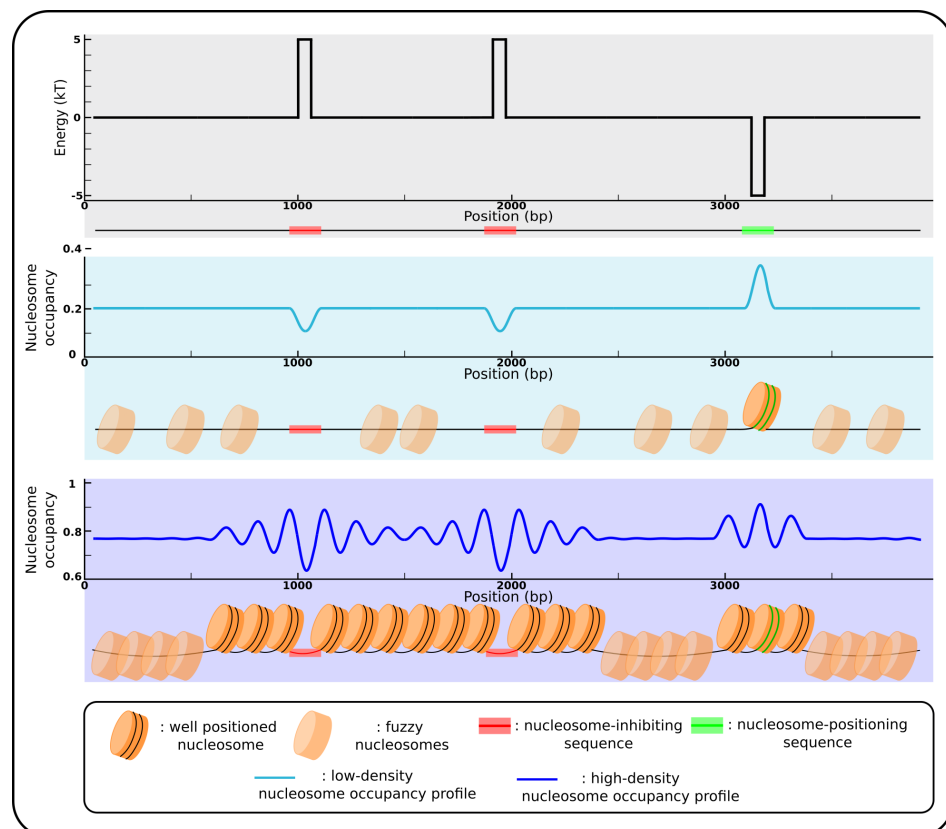


Figure 1. Nucleosomal positioning by sequence motifs. (**Top** panel) Landscape of the energy needed to bend the DNA fragment into the nucleosome depending on the sequence around the nucleosome dyad position (x -axis); the hypothetical landscape present two high energy peaks corresponding to two nucleosome-inhibiting sequence motifs (red), and a low energy well at a nucleosome-positioning sequence motif (green). (**Mid** panel) Nucleosome occupancy profile corresponding to the energy landscape for a low density of nucleosomes. Nucleosomes tend to avoid the inhibiting sequences (as represented by the minima in sky blue curve), and the only preferential nucleosome localisation is at the positioning sequence (peak in sky blue curve). (**Bottom** panel) Nucleosome occupancy profile to corresponding the energy landscape for a high nucleosome density. In this case, nucleosomes still avoid inhibiting sequences (minima in the blue curve), and a global positioning appears between and beside these nucleosomal barriers (oscillations in the blue curve), as a “parking” phenomenon resulting from the non overlapping property of nucleosomes (statistical positioning) (see Section 2.3). Nucleosome positioning also appears beside the well-positioned nucleosome formed on positioning sequence, according to the “anchor-positioning” model described in Section 2.3. Transparent nucleosomes represent fuzzy positioning, meaning that at these loci, nucleosomes have no preferential locations and can be formed more or less anywhere on the DNA.

2.2. Sequence Motifs with 10 Base Pair Periodicity as Nucleosome Positioning Signals

In the 1980s, the analysis of 32 coding and non-coding sequences (representing about 36,000 nucleotides) that were known to fold in chromatin-like structures (i.e., nucleosomes) exhibited a periodicity of ~ 10.5 base pair (bp) in the distribution of dinucleotides along their sequences [62]. Dinucleotides GG, TA, TG and TT were found to be the strongest contributors to this observed periodicity. In other words, in sequences that fold in chromatin-like structures, dinucleotides GG, TA, TG and TT tend to be regularly spaced by 10 or 11 bp whereas other dinucleotides are more randomly positioned. Interestingly, no 10.5 bp periodicity was found for prokaryotic sequences. Further analysis showed a symmetry in the phasing of the preferential positioning of complementary dinucleotides within the 10.5 bp periodicity [63]. An explanation proposed for these observations was about the affinity of the DNA sequence for histone core. It was suggested that sequence periodicity

and their symmetries facilitates the bending of the DNA molecule around the nucleosome core histones proteins [62,63]. It was even expected that it would be possible to predict nucleosome positioning from these sequence properties.

The “periodicity model” successfully predicted the curved shape of a 423 bp DNA restriction fragment containing a strong periodicity of AA and TT dinucleotides [64]. Sequence-encoded bending of DNA was explored in several studies [65–68], from which nucleosomal DNA bending tables were derived. In the nucleosome, A/T-rich sequences are preferred where the minor groove is facing inward, and G/C-rich sequences where it is facing outward of the structure [67]. In addition, homopolymers tend to be excluded from the nucleosome, especially from the dyad position [65–68]. Finally, it was observed that linker DNA regions between nucleosomes are cut poorly by DNase I enzyme, that is known to cut poorly in homopolymers, probably revealing their strong occurrence in linker DNA [68], in accordance with the previous observation.

The sequence periodicities described here facilitate the bending of DNA around the histone octamer to form a nucleosome. Such sequences could have a positioning effect. During the course of evolution, some selective pressure could have acted on genomes to select those sequences at specific loci where the presence of a nucleosome is necessary. Periodicities associated to nucleosomal sequences have been found in several species, in chicken, but also in yeast, human and worm [53,56,61,68–70]. However, among genomic sequences, even the most powerful positioning sequences only have a weak positioning power [33]. Sequences optimized for wrapping into the nucleosome, like the sequence of the clone 601 established by Lowary and Widom in their SELEX experiment on artificial DNA [33], are not found in genomic DNA. In addition, the global positioning power of genomic DNA is not much higher than that of random DNA sequences [33]. Thus, positioning sequences and their periodicities in the dinucleotide distributions fail to explain the genome-wide sequence-encoded nucleosomal positioning [33]. However, periodic distribution of sequence motifs is not the only way to encode nucleosome position.

2.3. Sequence-Encoded Nucleosome Depleted Regions and Statistical Positioning

In yeast, it has been showed that promoters are enriched in what are called nucleosome-depleted regions (NDRs) [43]. In several yeast species, these NDRs are found both in vivo and in vitro, indicating that they are directly encoded in the DNA sequence, mainly through poly(dA:dT) sequences that are known to inhibit nucleosome formation [1]. The strength of the depletion depends mainly on the length and purity of the poly(dA:dT) sequence [1], allowing a fine tune regulation of gene expression in yeast [71]. Positioning of nucleosomes can arise from these NDRs, following a statistical positioning model [72,73], where nucleosomes stack against a fixed object (either a NDR or a highly positioned nucleosome) that serves as an anchor, forming an array of positioned nucleosomes (Figure 1). The closer a nucleosome is to the anchor, the better it is positioned. Thus, counter-intuitively, sequence-encoded nucleosome positioning could arise not from positioning sequences but rather from anti-positioning sequences that anchor the position of nucleosomal arrays. In the case of yeast promoters, if NDRs are observed both in vivo and in vitro, arrays of nucleosomes are only observed in vivo, on the side of the transcribed units [74]. In this case, the in vivo nucleosomal organization results from the combination of the sequence effect (mainly specifying the NDRs and probably the +1 nucleosomes) and the ATP-dependent chromatin remodelers (for the ordering of nucleosomes). Another type of arrays of nucleosomes relying only on sequences have been observed in yeast, where nucleosomes are confined between sequence-encoded NDRs when these NDRs are close to one another [55,57]. Indeed, when two NDRs are close enough to each other, constraints appear on the nucleosomal positioning, mainly because of the exclusion interaction between nucleosomes since two nucleosomes cannot superimpose. For example, if two sequence-encoded NDRs are separated by a distance of about 300 bp (~2 nucleosomes), and one nucleosome is formed between the NDRs, it can be formed quite anywhere along the 300 bp. However, if 2 nucleosomes are formed, taking about 147 bp each, then the possi-

bilities are greatly reduced and preferential positioning appears. Sequence-encoded arrays of nucleosomes can thus result from sequence-encoded NDRs and a high density of nucleosomes. This “statistical positioning between NDRs” model was experimentally validated with atomic force microscopy (AFM) visualization of nucleosome positioning along a DNA fragment bounded by two sequence-encoded NDRs separated by a two-nucleosomes long distance [55,75]. When either one or two nucleosomes were reconstituted on this fragment, single nucleosomes were observed anywhere between the barriers, but as predicted, the position of nucleosome pairs were very constrained.

In human, part of the genome-wide nucleosomal positioning follows this scenario of statistical positioning between NDRs [28,76]. Indeed, a physical model of nucleosome formation based on sequence-dependent bending properties of the DNA double helix revealed about 1.6 million nucleosome-inhibiting energy barriers (NIEBs) along the human genome. These NIEBs correspond to NDRs, both among *in vivo* and *in vitro* data. In both conditions, when NIEBs are close enough to each other (about four nucleosomes or less), a constrained positioning of nucleosomes is observed, just as described above in yeast. The *in vitro* observation indicates that this positioning is not dependent of the action of remodelers, but relies only on the sequence-encoded NIEBs/NDRs and high density of nucleosomes. *in vitro* map of nucleosomes also showed that a nucleosome-favoring sequence flanked by two nucleosome-detering sequences can form what is called a “container” site in which a nucleosome is trapped [10]. Taken alone, each of these sequences do not have any significant positioning or anti-positioning power, but taken together, they form a highly positioned nucleosome at a specific locus. These container sites were also found in the *in vivo* nucleosomes maps, where they can serve as anchors to form nucleosomal arrays by stacking of the other nucleosomes against the well positioned one. The situation is similarly found at the promoters of yeast genome: a fixed object (here, a highly positioned nucleosome, a NDR in yeast) serves as an anchor for regularly spaced nucleosomal arrays. The difference is that the formation of the array is not associated with transcription as in yeast. However, these arrays are also only observed *in vivo*, indicating that if the anchor is sequence-encoded, the action of remodelers is needed to fluidify the movement of nucleosomes and allow statistical positioning. Note that isolated NIEBs can also serve as anchors: two to three positioned nucleosomes have been observed on their borders in human, both *in vivo* and *in vitro* [28,76], illustrating that the “stacking against an anchor” model does not always need the activity of remodelers.

2.4. Predicting Nucleosomal Positioning from Sequences

Nucleosome occupancy encoded in the sequence can presumably be predicted through sequence-based modeling. This was achieved using mainly two types of approaches: bioinformatic models relying on machine learning [22,53,56,58], and physical models relying on energy calculations [55,57,59,60,77]. The general idea of the bioinformatic models is to detect, genome-wide, the sequence features associated with nucleosomal positioning. For example, the model detailed in [53] is based on an *in vitro* map of yeast nucleosomes. From this map, the sequence preferences for nucleosomes are extracted to establish a probabilistic model that assigns a score to each 147 bp fragment. This score is based on the 5-mers observed along the sequence of the fragment. From the score landscape, and taking into account the impossibility to superimpose two nucleosomes, nucleosomal positioning can be predicted. This approach reproduced well experimental mapping of nucleosomes [53]. A simpler approach has been developed in [22], in which the over 2000 parameters of [53] are reduced down to only 14 parameters. It was even claimed that a model taking into account only the GC content and poly(dA:dT) sequences is sufficient to achieve good predictions of nucleosome occupancy [22]. The GC content is tightly correlated to nucleosome occupancy [27,28]. It was in fact argued that the observation that the genomic GC content of Eukarya is way less variable than that of Bacteria and Archaea corroborates this observation. It was linked to the high level of conservation of histones between organisms, whereas nucleoid-associated proteins are more variable,

possibly allowing wider range for genomic GC content between species [78]. The physical modeling approach was considered independently by different groups [55,57,60,79]. It is based on intrinsic bending properties of the DNA and thus, its ability to be wrapped around histone octamers. The idea is to compute the energy needed to deform all 147 bp DNA fragments from their intrinsic conformation to the helical conformation adopted in the nucleosome, based on tabulated sequence-dependent elastic parameters. This provides an energy landscape for the formation potential of nucleosomes along the genome. The dynamic assembly of histone octamers along the DNA chain is then modeled as a fluid of rods of finite extension (the DNA wrapping length around the octamer), binding and moving in the nucleosome formation potential and respecting the exclusion relationship between nucleosomes. The nucleosome occupancy profile can then be deduced given a temperature and a chemical potential allowing to fix the average nucleosome density to the experimentally determined value. Nucleosome occupancy based on our implementation of the model [55,57] fits well the experimental occupancy data in yeasts, in the nematode *C. elegans* and the fly *D. melanogaster* [55,59,80], and in human [28,76].

3. Nucleosome Positioning during Evolution

3.1. Nucleosome Position as a Darwinian Feature

Nucleosome occupancy influences the binding of transcription factors by controlling the accessibility to DNA [25]. The modulation of nucleosome occupancy is thus a critical feature for gene transcription regulation. Indeed, the distribution of nucleosomes around genes was associated with transcription levels in several species, including yeast [81], human [10,44], mouse [51], drosophila [49], and plants such as the thale cress [50], rice [50] and maize [26]. For example, highly expressed genes are associated with a more pronounced nucleosome depletion at their promoter than lowly expressed genes. The transcriptional changes during cell life processes such as differentiation, reprogramming, stress or even aging are associated with changes in nucleosome occupancy [82–85]. Modifying the nucleosome organization at some loci is thus expected to have either a positive or a negative impact on the fitness of an individual [86]. As nucleosome positions are at least partially sequence-encoded (Section 2), this strongly suggests that natural selection on DNA sequence could have an impact on the nucleosomal positioning. In other words, mutations could be selected or counter-selected, not for their direct effect on coding sequences, but for their influence on the position of nucleosomes at some specific loci, indirectly influencing features under selection such as gene expression. Following this hypothesis, natural selection could favor nucleosome inhibiting sequences where sequences need to be constantly available to transcription factors (at the regulating sequences of constitutive genes for example). It could also favor certain nucleosomal organization on the body of genes according to the basal level of transcription needed. The latter possibility question the compatibility between the nucleosomal and the genetic codes, to allow encoding of both a protein sequence and the nucleosomal organization in the same sequences. This compatibility has been explored by Eslami-Mossallam et al. [87], revealing the possibility of multiplexing genetic and mechanical information along a single sequence. Indeed, it is achievable to change the nucleosomal organization on the body of a gene without changing the protein(s) associated with the gene, thanks to the redundancy of the genetic code [87].

3.2. Nucleosome Positioning and the Evolution of Gene Regulation

In yeast, “growth genes” are identified as genes almost constantly expressed during growth, often associated with the metabolic pathways used in *ideal* growth conditions. In contrast, “stress genes” are genes expressed only in certain specific conditions, for example to respond to an environmental change. At the nucleosomal level, differences have been observed between growth and stress genes. The prediction of the nucleosomal organization at the promoter of these different types of genes in two yeast species, *Candida albicans* and *Saccharomyces cerevisiae*, showed that on average growth genes exhibit an intrinsically open chromatin at their promoter, when stress genes harbor a more closed patterns [23].

The experimental confirmation of the predicted organizations, both in vitro and in vivo, demonstrated that they are encoded directly in both genomes. Thus, in these two yeasts, we have two distinct sequence-encoded nucleosomal patterns associated with the two modes of gene expression. These two species display major metabolism differences when grown in a high glucose environment: *C. albicans* that grows mainly using respirative metabolism is identified as an aerobic yeast, as oppose to *S. cerevisiae* that grows mainly using fermentative metabolism, identified as an anaerobic yeast. From an evolutionary standpoint, orthologous genes associated with respiration are growth genes in the former, that switched to stress genes in the latter during the evolution of yeasts. By comparing the nucleosomal organization at the promoter of these genes in these two species, it was shown that they exhibit an intrinsically open chromatin in *C. albicans*, and a closed chromatin in *S. cerevisiae* [23]. This pattern was also observed in 10 other yeast species for which the nucleosome occupancy was predicted genome-wide from the DNA sequence. These results were confirmed experimentally with the direct comparison of experimental nucleosome positioning and gene expression data in the same 10 yeast species [24]. It showed that gain or loss of poly(dA:dT) tracts are associated with modifications of the nucleosomal organization at several phylogenetic branch points [24]. For example, the promoters of mitochondrial ribosomal protein (mRP) genes have lost their poly-A-like sequences in anaerobic yeasts, changing the chromatin organization on these genes from an open conformation (in aerobic yeasts) to a closed one (in anaerobic yeasts) [23,24]. These experiments show that in the course of yeast evolution, nucleosomes located at the promoter of genes have been repositioned, notably through the modification of the DNA sequence, and it was associated to a major change in yeast metabolisms, such as the switch from an aerobic to an anaerobic metabolism. This is a very good example of sequence selection not acting directly on coding properties, but for their affinity to nucleosomes, allowing a fine tuning of gene regulation from growth expression to stress expression pattern.

A similar dichotomy is present in multi-cellular organisms, such as maize, in the form of constitutive genes that are expressed regardless of the cell type, versus tissue-specific genes that are expressed only in some specific cell types. Sequences selected for nucleosomal positioning have been observed in this species [25,26]. In maize, the expression level between tissues show only minor differences in constitutive genes which contrast with tissue-specific genes that show higher differences. This difference shows that tissue-specific genes have higher transcriptional plasticity than constitutive genes. It was proposed that the sequence-encoded nucleosomal organization of each gene controls its transcriptional plasticity instead of directly its level of expression [25,26]. Indeed, the level of expression can change between cell types and conditions, particularly for tissue-specific genes. If the level of expression was directly sequence-encoded through nucleosomal positioning, transcriptional plasticity could not be achieved, since the gene sequence is the same in each cell and condition. In maize, the prediction from sequences of the nucleosomal organization of different set of genes showed that constitutive genes have the lowest sequence-encoded global nucleosome occupancy, while tissue-specific genes have the highest [26]. Compared to tissue-specific genes, constitutive genes have bigger and stronger NDRs at their transcription start site (TSS) as well as longer distances between both their 5' NDR and TSS, and their 3' NDR and transcription termination site. All these predicted features have been confirmed experimentally with MNase experiments. These two types of genes have different nucleosomal organization resulting in different transcriptional plasticity. In maize, it was also observed that the sequence of constitutive genes has a lower GC content than the sequence of tissue-specific genes, both in introns and exons where it is mainly driven by different codon usage. This likely illustrates selective pressures acting on the nucleosome positioning. The redundancy of the genetic code, allowing the multiplexing of genetic and structural informations [87], is used in this species to promote AT-rich codons in constitutive genes and GC-rich codons in tissue-specific genes, to reduce the GC content of the former and raise the GC content of the latter. This leads to differences in maize genes

nucleosomal organization, with a reduced occupancy on constitutive genes, associated with lower transcriptional plasticity. In contrast, the nucleosome occupancy is higher in tissue-specific genes, and associated with higher transcriptional plasticity. This interplay between nucleosome and transcriptional plasticity has also been observed in several other species such as *C. elegans* and *S. cerevisiae*. In *C. elegans*, a time-course of MNase digestion showed that the AT content in the promoter influences nucleosome stability [88]. In this type of experiments, various levels of chromatin digestion are obtained using different concentrations of MNase or different digestion times, providing information about the stability of nucleosomes [52,88,89]. Fragile nucleosomes are identified as nucleosomes only apparent in low-digestion data, as they are more easily destabilized by the MNase than stable nucleosomes [52,88]. Such experiment in *C. elegans* showed that fragile nucleosomes are associated with high AT content of the underlying DNA sequence, and low expression plus high transcriptional plasticity when they are localized at the promoter of genes [52]. In *S. cerevisiae*, it has been shown that genes can be classified according to their nucleosomal organization [55,80,90]. Some genes have a “crystal” nucleosomal organization, with n nucleosomes on the body of the genes and a precise, constant NRL. Others have a “bistable” nucleosomal organization, with the possibility to put n or $n + 1$ nucleosomes on the body of the gene, the $n + 1$ organization being associated with a higher expression level. These two classes of nucleosomal organization are, like in maize, associated with different transcription plasticity. Indeed, growth genes are associated with “crystal” organization, where stress genes exhibit a “bistable” organization [55,80,90]. Finally, in human, about 70% of promoters are associated to CpG islands (GC rich regions with a CpG dinucleotide content higher than elsewhere on the genome) [91]. These CpG islands have been described to be accessible without the need for ATP-dependent remodeling [92]. This could be due to their DNA sequence inhibiting nucleosome formation, although the well described nucleosome-free region surrounding the TSS of eukarotic genes could also be implicated [93]. All the examples mentioned here show that in a range of organisms, sequence-encoded nucleosomal organization at genes is strongly linked to expression pattern.

Selection of nucleosomal positioning at genes has also been linked to the complexity of organisms (Figure 2) [29,90]. In yeast, the majority of promoter exhibit a NDR, both in vivo and in vitro, indicating that this nucleosomal conformation is encoded directly in the DNA sequence. In contrast, if NDRs can be found in human at the promoter of expressed genes in vivo, it has a rare occurrence in vitro, and sequence-encoded NDRs are typically absent from promoters. In fact, in human, prediction of nucleosomal positioning from sequence showed that promoters are generally occupied by nucleosome attracting regions (NAR), that are the opposite of NDR. One explanation of this difference could lie in the fact that yeasts are unicellular organisms when humans are complex multicellular ones. Most of yeast genes are supposed to be used almost constantly, unlike human genes that are mostly tissue-specific. Following this hypothesis, it could be advantageous for yeast to have a default organization of “open and ready to transcribe” chromatin at their promoter, and to actively close the promoters of the genes that need to be expressed in specific conditions only. In contrast, it could be advantageous in human to adopt the opposite default organization of “closed and repressed” chromatin at promoters and to open specifically the few genes needed in each cell. The comparison of sequence-predicted chromatin conformation at promoters of several species confirmed this hypothesis [29,90]. The nucleosomal organization at promoters follows a gradient, from “mostly NDR” to “mostly NAR”, that corresponds to the complexity of the organisms (identified as the number of different tissues composing the organism) [29]. In other words, yeast, a simple unicellular organisms, exhibited the most sequence-encoded open chromatin at their promoters. Interestingly, the same rule applies in archaea possessing nucleosome-like structures, where the histone core is tetrameric instead of octameric in eukaryotes, leading to the wrapping of only about 80 bp of DNA in archeal nucleosomes instead of 147 bp in eukaryotic nucleosomes [94]. Inversely, vertebrates like zebrafish and mammals, which are

multicellular complex organisms, exhibited the most sequence-encoded closed chromatin at their promoters. Between them a range of intermediate signals was found, but with a clear progression from full NDR model for unicellular, to hybrid NDR-NAR and full NAR model in multicellular organisms, according to the increase in organism complexity. This result seems to confirm the hypothesis mentioned earlier about the two models of chromatin at promoter. However, following this hypothesis, genes that are expressed in all cell types of complex multicellular organisms should exhibit a NDR at their promoter, because the “open and ready to transcribe” model would then be advantageous for these genes. Interestingly, this is not the case, and the promoters of these gene are even stronger NAR than cell-type specific genes. To explain this result, it has been proposed that the presence of NAR at promoters could also be linked to a retention of nucleosomes at promoters in cells generally depleted in nucleosomes such as sperm cells, to ensure transmission of epigenetic informations [29]. Regardless of the real biological meaning of these different sequence-encoded nucleosome organizations at promoters, this example shows that it has been modified during the evolution, and that these changes are mainly the result of sequence modifications, with NDR in yeast and NAR in mammals.

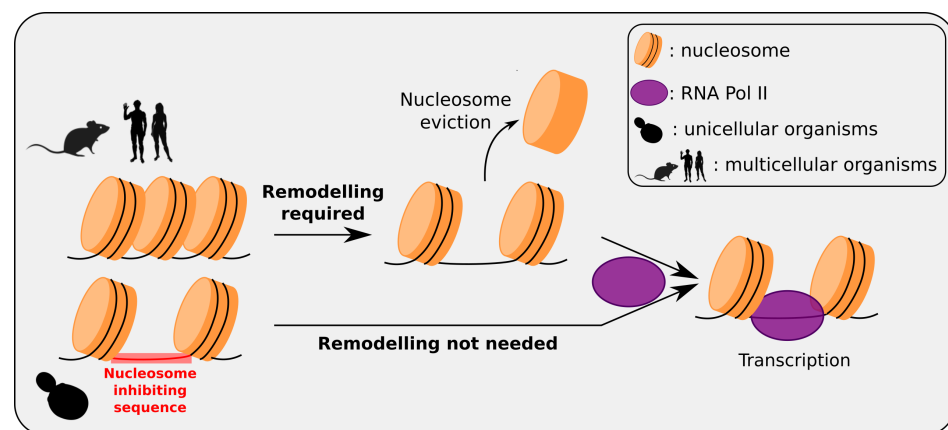


Figure 2. Multicellular/unicellular strategy for intrinsic nucleosomal organization at promoters. In human and most multicellular organisms, typical promoters are intrinsically occupied by nucleosomes, with no inhibiting sequences at these loci; genes are “repressed-by-default” and activated only when needed by the active removal of a nucleosome at the promoter making it accessible to the transcription machinery. In unicellular Eukaryotes such as yeast, inhibiting sequences have been selected at typical gene promoters, to avoid nucleosome formation at these loci; genes are “activated-by-default” since promoters are directly accessible to the transcription machinery, avoiding a remodeling step. These different strategies can be understood as in multicellular organisms, most genes (except housekeeping genes) present tissue-specificity, whereas in a unicellular Eukaryotes most genes are susceptible to be used in every cell. (RNA Pol II is not drawn to scale).

3.3. Is Chromatin Organization Selected Genome-Wide?

Examples of selection on specific nucleosomal organization at genes through selection of DNA sequences were described in Section 3.2. As nucleosome organization has a direct impact on the expression of genes, either driving expression level or transcriptional plasticity, it has a direct consequence on the fitness of individuals. Hence, selection of the corresponding sequence motifs in the course of evolution makes sense. However, genes represent only a very small fraction of the genome of most multicellular organisms. At numerous loci, nucleosomes are positioned by the intrinsic properties of the DNA sequence on which they are formed (Section 2). For example, nucleosomes are encoded in the DNA sequence over about 37% of the human genome through the statistical positioning at the border of NIEBs [28]. This genome-wide encoding of nucleosomes through nucleosomal barriers seems universal among vertebrates, as predicted in human but also in mouse, cow, pig, chicken and zebrafish [95]. This raises the question of the selection of this nucleosome positioning mechanism. In other words, are nucleosome positions also selected

at the genome-wide level? One NIEB feature that is common across vertebrates is the oscillating GC-content profile at NIEB borders, with very low GC at the internal border of NIEBs, then high GC on the ~140 bp adjacent to the barrier (corresponding to the first stacked nucleosome position), then again low GC over ~10 bp (first linker), then high GC over the second nucleosome location, low GC on the second linker, and so on. The oscillating pattern becomes less and less pronounced as we move away from the NIEBs, with barely no oscillation detectable after the third nucleosome. However, in the vicinity of NIEBs (~500 bp of each border), the oscillations are very clear and observed across vertebrates species. As low GC is associated with inhibition of nucleosome formation, and higher GC content in general is associated with nucleosome positioning, the nucleosome organization at the border of NIEBs should also be conserved across these species, through the conservation of GC content. It was indeed observed that there is a link between a higher GC content at the location of nucleosome dyads compared to linker regions and sequence evolution [27,28]. By comparing the interspecies mutations between human and chimpanzee to intraspecies mutations obtained from the 1000 Genomes project [96] in human, several types of selection reinforcing the oscillation of GC content at the border of NIEBs have been observed [28]. First, signature of positive selection for mutations towards A and T nucleotides were described at the internal border of NIEBs and at the linker loci. Inversely, signatures of purifying selection (counterselection) were observed against these mutations at the positions corresponding to nucleosomal DNA. This confirmed an earlier observation of C-to-T mutations favored in linkers and disfavored in nucleosomes [27]. Second, mutations towards G and C nucleotides followed the exact opposite pattern, with purifying selection in NIEBs and in linkers, and positive selection in nucleosomal DNA. Finally, mutations disrupting TTT or AAA sequences (tTt-to-tAt or aAa-to-aTa mutations) were highly counter-selected in NIEBs and linkers, and favored in nucleosomal DNA. As these sequences strongly impair nucleosome formation, this suggests that natural selection is acting on NIEBs to maintain the nucleosomal organization at their borders. In a nutshell, evolution at human NIEBs loci favored mutations towards A and T in non-nucleosomal DNA, and mutations toward C and G in nucleosomal DNA, leading to the oscillating GC content also observed in each vertebrate analyzed, and reinforcing the positioning of two to three nucleosomes at these loci.

3.4. Are Transposable Elements Involved in Chromatin Organization?

For now, most studies about the interplay between sequence evolution and nucleosome positioning focused on single nucleotide variations (SNVs), analyzing their position relative to the nucleosomes. However, little is known about other types of mutations such as insertions or deletions in this context. The insertions of transposable elements (TEs) could in fact be important to fully capture the coupling between sequence-mediated nucleosome organization and genome evolution. Indeed, TEs are able to integrate and spread within genomes through a mechanism called transposition [97,98]. They are major components of Eukaryotic genomes, representing for example at least 45% of the human genome [99], although there is a high diversity in terms of TE composition in vertebrate genomes [100]. There are many families of TEs, according to their transposition mechanism, size, DNA base composition, etc. [101]. Some of these elements have been associated to a biological function. For example, a TE insertion can be at the origin of the formation of a new gene, an event called TE domestication (reviewed in [102]). Some TE copies were found to be implicated in various biological processes, for example in the sexual development and function in various animal species [103]. In contrast, some TE insertions have been found to have deleterious effects, with TEs being associated with various diseases [104]. Thus, TEs are major components of the evolution of genomic sequences, their transposition bringing DNA fragments to new locations, inserting from a few tens to several thousands of base pairs of DNA at the insertion site. If the sequence effects of these insertions have been largely investigated such as TFBS transport or coding sequence disruption, the effect of the insertion on the nucleosomal organization remains largely unknown. The insertion

of TEs, by disrupting the sequence at the insertion site, could either disrupt or reinforce the sequence-encoded nucleosomal organization, according to the nucleosome-associated properties of the inserted sequence. Thus, apart from being drivers of sequence evolution, TEs could also be drivers of the evolution of nucleosomal organization. Some results already point into this direction such as the presence in human of Alu transposable elements at the border of about half of the NIEBs mentioned in Section 2.3 [95]. The family of Alu TEs is specific to primate genomes [105]. They are short retrotransposons of about 300 bp, with a DNA sequence compatible with the positioning of two nucleosomes [106]. One hypothesis to explain the distribution of Alu TE at the border of human NIEBs is that NIEBs being NDRs and thus accessible to external factors, they could represent preferential target sites for the insertion of Alu TEs. Another hypothesis is that Alu TEs could be at the origin of new NIEBs formation, i.e., nucleosome organization would be a consequence of Alu insertion. Note that these hypothesis are not mutually exclusive, and the link that was observed between NIEBs and Alu TEs in human could result from the interplay between several mechanisms and selection. Moreover, strongly positioned nucleosomes were observed on newly inserted TEs, possibly participating to their regulation [107]. The presence of these nucleosome could both decrease the accessibility to these TEs for transposition machinery, making new transpositions more difficult, and increase the mutation rates on them, because DNA repair is less efficient in nucleosomes than in naked DNA [107]. In a general fashion, there seems to be an interconnection between TEs and the evolution of nucleosomal positioning that still needs to be investigated to fully understand the coupling between sequence evolution and chromatin evolution.

4. Feedback of Nucleosomal Positioning on Mutational Patterns

As we saw in Section 3.3, signatures of selection have been clearly identified at the borders of NIEBs. However, another phenomenon could participate to the observed reinforcement of the local GC content. In fact, profiles of mutational rates at NIEB borders were calculated, for both inter- and intra-specific human mutations [28]. This showed for example that interspecies mutation rates towards A and T were higher in non-nucleosomal DNA than in nucleosomal DNA. As discussed above, positive selection would favor these mutations in non-nucleosomal DNA while counterselection would act in nucleosomal DNA. In addition, some oscillations of mutation rates were also observed for intraspecies mutations, for which selection had way less time to influence the mutational pattern. Thus, it seems that even in the presence of weak to no selection, the mutations are not randomly distributed at the borders of the NIEBs. This suggests that nucleosome occupancy has a direct influence on the mutational patterns. The presence of a well-positioned nucleosome, meaning that it almost always covers the same DNA fragment, could then create a mutational bias on this DNA fragment, favoring some mutations type in the nucleosomal DNA with respect to the linker DNA. Nucleosome could bias mutations towards some specific nucleotides on the nucleosomal DNA, by its interaction with DNA damage mechanisms, or the DNA repair machinery. Next generation sequencing progress now permits to establish cartographies of specific DNA damage mechanisms on the genome, and to quantify the efficiency of DNA repair machinery. This made it possible to explore the direct influence of nucleosomes on mutational processes.

Early in the 2000s, it was shown that the excision repair mechanisms of DNA such as base excision repair (BER) or nucleotide excision repair (NER) are hampered by the presence of nucleosomes [108]. It was confirmed a decade later that DNA damages are more persistent in nucleosomal DNA [109]. As DNA damages can lead to mutations, notably during replication, the inhibition of BER and NER has a direct influence on mutational patterns. Nucleosomes also directly modulate the formation rate of certain type of DNA lesions [110]. These properties can be related to the stability of the DNA double helix in the nucleosomal context, as illustrated by the lower degradation rate after cell death of nucleosomal DNA compared to linker DNA in ancient DNA samples [111,112]. Nonetheless, it is crucial to decipher the interplay between nucleosomes and DNA lesion formation

and repair mechanisms to understand the influence of nucleosomes on the mutational pattern, and take it into account in evolutionary approaches. Two types of mutational biases have been described in relation to nucleosome positioning [31], associated to nucleosomal occupancy and the rotational positioning of nucleosomal DNA in regards to the histone core, respectively.

Concerning nucleosome occupancy, it has been shown that C-to-T mutations were depleted in nucleosomal DNA relative to linker DNA [113]. As discussed in Section 3.3, natural selection is implicated in the mutational biases [27,28], but a mutational mechanism itself could also be implicated. Indeed, C-to-T mutations usually results from spontaneous deamination of cytosines and 5mCs [31]. This mechanism is more efficient when there is a local opening, called “breathing”, of the DNA double helix. Such breathing of DNA is inhibited in nucleosomal DNA, due to strong structural constraints imposed in the wrapping of DNA around histones, but remains possible in linker DNA, which is free from these constraints [114]. This wrapping is a hindrance to mutations leading to a depletion of the main C-to-T mutations in nucleosomal DNA as compared to linker DNA [113]. Similarly, experiments to map oxidatively induced DNA damages such as 8-oxoguanine (8-oxoG) in *S. cerevisiae* showed that they are modulated by nucleosome occupancy [115]. However, as 8-oxoG persistence depends on the equilibrium between DNA susceptibility to oxidation damage and efficiency of BER, it is still unclear whether the cause of the modulation by nucleosome occupancy is the influence on damage formation or on the efficiency of the repair mechanism [115]. Both hypotheses are not mutually exclusives. Further studies in yeast BER-deficient strains should provide insights about this question.

The effect of nucleosome occupancy on the mutational patterns has also been investigated in cancers where whole genome sequencing of tumors allows to examine the interplay between nucleosomes and mutational signatures [116–118]. These signatures correspond to unique combinations of mutation types, generated by specific mutational processes, in one or several types of cancers [119]. For example, mutational signature 1 found in all cancer types results from spontaneous deamination of 5-methylcytosine, and the type of mutation is mainly C-to-T mutation, with preferences for ACG, CCG, GCG and TCG contexts [119]. Mutations from signatures 17 and 18 are mainly T-to-G and C-to-A mutations, respectively, for which the mutational processes involved are unknown. In breast tumors, these two mutational signatures have been found to be more frequent in nucleosomes than it was expected from the sequence composition of the associated DNA fragments [116]. At transcription factor binding sites (TFBSs) flanked by regularly ordered nucleosomes following the model of statistical positioning by anchors (Section 2.3), melanoma mutations (principally induced by UV light) exhibit a periodic distribution associated to nucleosome positioning with a maximal density at nucleosome dyads, which differs from the expected pattern based on sequence composition [120]. More generally, a pan-cancer analysis revealed that for many cancer mutational processes, there are differences in mutation rates between nucleosomal DNA and linker DNA [121]. It also brought new observations, like tobacco-linked mutations occurring more frequently in linker than in nucleosomal DNA. The inhibition of both BER and NER repair systems is hypothesized to be a major player of UV-induced mutational biases. For tobacco-induced mutational bias, the mutational process (bulky DNA adducts at guanines (BPDE-dG)) is known to be inhibited in nucleosomes, leading to the “linker preference” for this type of mutations. The different examples mentioned here show that nucleosome dyad position (the so-called translational positioning of nucleosomes) has an influence on mutational patterns, through the modulation of the efficiency of either the DNA damage processes, or the repair mechanisms, or both, altogether leading to differences in mutation rates and biases between nucleosomal DNA and linker DNA.

Mutations are also modulated at a higher resolution than the nucleosome-linker dichotomy. Indeed, depending on which of the minor or the major groove of a DNA base pair faces the histones (the so-called rotational positioning of DNA within the nucleosome), mutation rates can be variable and, because DNA histone contact points are specifically with

the minor groove, each nucleosome translational positioning imposes a specific rotational positioning. A comparison between different *D. melanogaster* populations and between this species and a closely related species showed that C-to-T substitutions were more frequent where the minor groove of DNA faces the nucleosome (minor-in), than where the minor groove of DNA faces away from histones (minor-out) [122]. As at minor-in loci, the DNA is structurally constrained by chemical groups of histones H3 and H4, A/T (or WW) di-nucleotides could be favored for their higher flexibility and low steric hindrance [4]. The periodic occurrence of C-to-T mutations in nucleosomal DNA has been interpreted as a sign of selection on more favorable DNA fragments for nucleosome. However, an alternative hypothesis is that the interaction ability between DNA and mutagenic agents or repair machinery are different at minor-in and minor-out stretches of DNA, resulting in different mutation rates between these loci. This hypothesis is supported by the demonstrated decreased activity of BER at minor-in loci, resulting in lower repair efficiency of methylated guanines, the corollary of this being a higher mutation rate [123]. Experiments with DNase I showed that the accessibility to DNA could be a reason for the decreased activity of BER [124].

Another example of modulation of mutational processes along nucleosomes is for the UV-induced formation of cyclobutane pyrimidine dimers (CPDs) and (6-4) photoproducts (6,4-Pps) in DNA. Both DNA lesions are formed on TT, TC, CT and CC di-nucleotides. In nucleosomal DNA, a ~10 bp periodicity has been observed in CPDs formation [125]. In fact, this periodic pattern correlates with the rotational positioning of nucleosomes, with preferential CPD formation at minor-out loci [125,126]. The 10 bp periodic pattern and the correlation have been observed genome-wide in yeast and human thanks to a NGS-based damage mapping method named CPD-seq [123,127,128]. The UV-irradiation of the same naked DNA fragment (without nucleosomes) resulted in an opposite CPD formation pattern, with CPDs occurring at positions corresponding to minor-in loci, probably because of the increase of TT dinucleotides at these regions (Section 2) [127]. This means that the underlying sequence is not the cause of the periodic formation pattern of CPDs in nucleosomal DNA, in fact the sequence would even favor the opposite pattern. The presence of a nucleosome, and the structural constraints associated with its formation, override the sequence preferences of CPDs to promote UV-damage at minor-out regions, where the DNA is more accessible. So, nucleosomes have a strong influence on this DNA damage process.

Distribution patterns favoring the minor-out stretches of DNA such as the CPD distribution described above are also found in some types of cancer. In melanoma, the vast majority of somatic mutations are C-to-T transitions at dimers of pyrimidine, characteristic of the UV mutational signature [129]. Analysis of melanoma mutations showed the same ~10 bp periodicity in well-positioned nucleosomes as the one described for CPD mutations above [121,128]. The same pattern has been retrieved in lung cancer mutations, with high density at minor-out and low density at minor-in [121]. A high resolution genome-wide mapping of DNA damage process implicated in lung cancer mutations would help to understand if it is inhibited at minor-in stretches like CPD formation, or if the DNA damage distribution is constant, which would suggest an implication of the DNA repair mechanisms. On the other hand, other cancers exhibit an opposite mutational pattern, with high mutation densities at minor-in stretches and low mutation densities at minor-out stretches [121]. It has been proposed that a lower efficiency of BER mechanism at minor-in stretches could explain this periodicity. DNA damage at these loci would then be more persistent than at minor-out loci, leading to an increase in mutation rate [121]. Moreover, the presence of nucleosomes impair the recognition of single-strand breaks localized on the non-template DNA strand (NT-SSBs) by poly(adenosine diphosphate-ribose) polymerase 1 (PARP1), and in turn their repair through BER [130]. Undetected nucleosomal NT-SSBs can be repaired during transcription through transcription-coupled nucleotide excision repair (TC-NER). The efficiency of this repair mechanism is higher when the NT-SSBs is localized on the side of nucleosomal DNA facing the histone octamer than when it

is localized at more accessible loci in nucleosomal DNA. This could be because these accessible loci are more prone to be repaired by BER [130]. Although in the case of NT-SSBs, the interactions between nucleosome and DNA damage and repair mechanisms do not seem to lead to modulated mutation rates, it is another good example of the importance of nucleosome in DNA damage and repair processes.

Hence, the nucleosome has a strong influence on mutational patterns (Figure 3). It affects the effectiveness of excision repair mechanisms like BER and NER [131–134]. A lot of DNA damages are repaired through BER or NER, like UV-induced, tobacco-induced and oxidative DNA damage. The resulting mutational densities are generally increased in nucleosomal DNA compared to naked linker DNA. However, the activity of BER and NER are also modulated within the nucleosome, with a higher efficiency at minor-out loci, where DNA is more accessible than at minor-in loci, where structural constraints are stronger. The accessibility to DNA inside the nucleosome has also an influence on DNA damage mechanisms, some of them having a preference for minor-out stretches of DNA, and other for minor-in stretches [121]. All these possible influences of nucleosome on mutational patterns need to be taken into account in further attempts to decipher the evolution of DNA sequence regarding nucleosome positioning, to avoid considering as the sign of selection pressure some mutation biases induced by the presence of a nucleosome and its interplay with mutational processes.

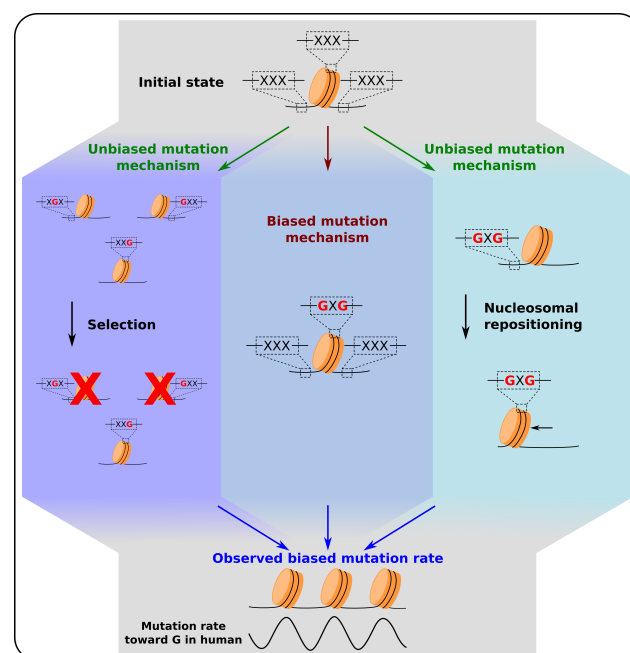


Figure 3. Source of biased mutation rates relative to nucleosomal positioning. Three mechanisms that can lead to biased mutation rates relative to nucleosome positioning as described in Sections 3 and 4. **(Left)** Mutations that facilitate the positioning of nucleosomes at specific loci are positively selected, those favoring alternative positions are purified. Such mechanism is for example observed at yeast promoters (Section 3.2). **(Center)** A biased mutation mechanism where the presence of a nucleosome drives mutations notably through interactions between nucleosomes and DNA damage and repair mechanisms (Section 4). **(Right)** A nucleosome repositioning model, in which mutations lead to the repositioning of nucleosomes that can also explain the observed biased mutation rates relative to nucleosome positioning when the latter is assumed to remain unchanged during evolution (Section 5). Since all three mechanisms have been observed at the genome scale, the global biased rate of mutations observed is likely to come from a combination of all three mechanisms. The cartoons illustrate possible evolutionary scenarios for 3 trinucleotides (XXX) located in the nucleosomal DNA, and the linker DNA upstream and downstream of a nucleosome. The figure only represents mutations toward G, but these three models are also valid for the other mutational biases (Sections 3–5).

5. Concluding Remarks

Nucleosome positions in genomes are at least partially encoded in the DNA sequence, through two main mechanisms (Section 2; Figure 1). The first one consists of an interplay between anti-positioning sequences (such as homopolymers like poly(dA:dT)) and high density of nucleosomes, leading to positioning by confinement between nucleosomal barriers [28,55,57,75]. The second mechanism consists in a fine-tuning of nucleosome positioning at the base-pair resolution, with preferences for A/T rich sequences where the DNA is making contact with histone proteins at minor-in positions, and G/C rich sequences where the DNA minor groove is facing away from histones [67]. In vivo and in vitro maps of nucleosomes present high similarities, indicating that the sequence properties are relevant even in the presence of other factors influencing the position of nucleosomes such as ATP-dependent remodelers. As nucleosomes have a functional importance, notably by modulating the accessibility to regulatory regions of genes, sequence evolution should be constrained by the effect of mutations on nucleosome positions (Figure 3). We reviewed several cases of sequence selection for their nucleosomal properties, in yeast, but also in multicellular organisms such as maize and human (Section 3). However, some caveats need to be taken into account when we try to decipher the evolution of sequence regarding its effect on nucleosomes. The first bias that must be considered is the feedback of nucleosome positioning on the mutational patterns. Nucleosomes influence both the mechanisms of DNA damage and DNA repair, leading to difference in the mutational patterns, either between nucleosomal DNA and linker DNA, and within the nucleosomes, between the minor-in and minor-out positions (Section 4; Figure 3). Now that these biases are described, one must be careful interpreting sequence changes in regards to nucleosomal positioning, and properly separate the contribution of selective pressure from the mutational biases induced by the presence of nucleosomes. A second caveat needs to be considered about the interpretation of some observations such as the signature of selective pressure on nucleosomal positioning, as initially raised in [30]. In many studies, mutational data obtained through the comparison of phylogenetically related species assume that the nucleosome organization was identical in the ancestor of the extant species. This “static” view of nucleosome positions in the course of evolution may not be a correct assumption [30], because nucleosomes are frequently repositioned following the evolution of sequences (Section 3). For example, in Eukaryotes, it was observed that A/T-to-G/C mutations are more frequent at the nucleosome dyad. It was interpreted as either a mutational bias caused by the nucleosome, or selection acting on these mutations to reinforce nucleosome positioning, assuming an evolutionary stable nucleosome organization. However, another scenario is compatible with the observations, where the A/T-to-G/C mutations would have repositioned the nucleosomes because of the preference of the dyad for GC-rich motifs [30], i.e., nucleosome positioning would follow the mutations towards G/C (Figure 3). To properly address this possibility, one needs to reconstruct the in vivo nucleosome organization at the time of the mutation. In regions where sequence-encoded nucleosome positioning is relevant in vivo, this can directly be done by applying the nucleosome position prediction tools available (Section 2.4) on the phylogenetically reconstructed ancestral sequences. Otherwise, one would need to compare experimental maps of nucleosome positioning in germline cells of all the species considered. However, for now, just about a handful of species have their nucleosomes mapped experimentally, mainly in somatic or cancer cell lines [54]. Thus, regions where the nucleosomal array appears not to be remodeled, such as NIEB loci in vertebrates, are the best candidates to distinguish between selection, repositioning, and the biased mutation events, and to estimate the relative importance of each mechanism to explain the mutational patterns at nucleosomes.

In this article, we reviewed here findings about the interplay between sequence-encoded nucleosome positioning and evolutionary constraints. Yet, the contribution of the collective properties and functions of the nucleosomal array depending on the position of nucleosomes relative to other nucleosomes have not been addressed. In genomes, the formation of nucleosomal arrays with regularly spaced nucleosomes is conserved across

Eukaryotic organisms [135]. These arrays are associated with various functions, such as chromatin condensation in higher order structure, but also with long-range contacts between enhancers and promoters, or inhibition of cryptic transcripts or protection of DNA from double-strand breaks [135]. The formation of nucleosomal arrays depends on various external factors, including remodeling, but also on DNA-binding factors creating nucleosomal barriers against which nucleosomes are stacked, following the model described in Section 2.3. Some sequence motifs such as NIEBs can also act as barriers. If one NIEB does not seem to be sufficient to position more than two to three nucleosomes at each of its borders, two close NIEBs can lead to a regularly spaced array of up to six nucleosomes between them [28]. In vertebrates, the relative position of NIEBs is constrained, with consecutive NIEBs being spaced by distances that are multiple of ~153 bp [28,95], which was interpreted as a constraint that an integer numbers of compact nucleosomes fits between two close NIEBs. In this way, consecutive NIEBs could form regularly spaced nucleosomal array of controlled NRL following the statistical positioning model between two barriers [55,73,75]. The constraint of NIEB positioning regarding other NIEBs could arise to favor the apparition of such arrays, to use their properties on chromatin condensation and long-range contacts as described above. For example, short NRLs would assure that the intrinsic nucleosome arrays are in an open state, permissive to epigenetic regulation, allowing cell type specific regulation [28]. The influence of DNA sequence on nucleosome positioning and its interplay with the evolution of both sequences and nucleosome positions must then be considered not only at the nucleosome scale, but also at the scale of the nucleosomal array, thus taking into account higher order chromatin structures.

Author Contributions: Conceptualization, J.B., C.V., J.-N.V., F.G.B. and B.A.; writing—original draft preparation, J.B. and B.A.; writing—review and editing, J.B., C.V., J.-N.V., F.G.B. and B.A.; supervision, J.-N.V., F.G.B. and B.A. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Agence Nationale de la Recherche (ANR-20-CE12-013), and the ENS de Lyon “Projets Emergents” program. J.B. acknowledges support from the PhD funding program of ENS de Lyon.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors thank the French CNRS network GDR “Architecture et Dynamique Nucléaire” (ADN) for stimulating workshops.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

BER	Base excision repair
bp	Base pair
CPD	Cyclobutane pyrimidine dimer
MNase	Micrococcal nuclease
NAR	Nucleosome attracting region
NER	Nucleotide excision repair
NDR	Nucleosome depleted region
NIEB	Nucleosome-inhibiting energy barrier
NRL	Nucleosome repeat length
PTM	Post-translational modification

SNV	Single nucleotide variation
TE	Transposable element
TFBS	Transcription factor binding site
TSS	Transcription start site

References

- Segal, E.; Widom, J. What controls nucleosome positions? *Trends Genet.* **2009**, *25*, 335–343. [[CrossRef](#)]
- Kornberg, R.D.; Lorch, Y. Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell* **1999**, *98*, 285–294. [[CrossRef](#)]
- Finch, J.T.; Lutter, L.C.; Rhodes, D.; Brown, R.S.; Rushton, B.; Levitt, M.; Klug, A. Structure of nucleosome core particles of chromatin. *Nature* **1977**, *269*, 29–36. [[CrossRef](#)] [[PubMed](#)]
- McGinty, R.K.; Tan, S. Nucleosome Structure and Function. *Chem. Rev.* **2015**, *115*, 2255–2273. [[CrossRef](#)] [[PubMed](#)]
- Lantermann, A.B.; Straub, T.; Strålfors, A.; Yuan, G.C.; Ekwall, K.; Korber, P. Schizosaccharomyces pombe genome-wide nucleosome mapping reveals positioning mechanisms distinct from those of Saccharomyces cerevisiae. *Nat. Struct. Mol. Biol.* **2010**, *17*, 251–257. [[CrossRef](#)] [[PubMed](#)]
- Athey, B.D.; Smith, M.F.; Rankert, D.A.; Williams, S.P.; Langmore, J.P. The diameters of frozen-hydrated chromatin fibers increase with DNA linker length: Evidence in support of variable diameter models for chromatin. *J. Cell Biol.* **1990**, *111*, 795–806. [[CrossRef](#)]
- Kurumizaka, H.; Kujirai, T.; Takizawa, Y. Contributions of Histone Variants in Nucleosome Structure and Function. *J. Mol. Biol.* **2021**, *433*, 166678. [[CrossRef](#)]
- Szenker, E.; Ray-Gallet, D.; Almouzni, G. The double face of the histone variant H3.3. *Cell Res.* **2011**, *21*, 421–434. [[CrossRef](#)]
- Correll, S.J.; Schubert, M.H.; Grigoryev, S.A. Short nucleosome repeats impose rotational modulations on chromatin fibre folding. *EMBO J.* **2012**, *31*, 2416–2426. [[CrossRef](#)]
- Valouev, A.; Johnson, S.M.; Boyd, S.D.; Smith, C.L.; Fire, A.Z.; Sidow, A. Determinants of nucleosome organization in primary human cells. *Nature* **2011**, *474*, 516–520. [[CrossRef](#)]
- Pisano, S.; Galati, A.; Cacchione, S. Telomeric nucleosomes: Forgotten players at chromosome ends. *Cell. Mol. Life Sci.* **2008**, *65*, 3553–3563. [[CrossRef](#)] [[PubMed](#)]
- Tommerup, H.; Dousmanis, A.; de Lange, T. Unusual chromatin in human telomeres. *Mol. Cell. Biol.* **1994**, *14*, 5777–5785. [[CrossRef](#)] [[PubMed](#)]
- Makarov, V.L.; Lejnine, S.; Bedoyan, J.; Langmore, J.P. Nucleosomal organization of telomere-specific chromatin in rat. *Cell* **1993**, *73*, 775–787. [[CrossRef](#)]
- Muyldermans, S.; De Jonge, J.; Wyns, L.; Travers, A.A. Differential association of linker histones H1 and H5 with telomeric nucleosomes in chicken erythrocytes. *Nucleic Acids Res.* **1994**, *22*, 5635–5639. [[CrossRef](#)]
- Smilenov, L.B.; Dhar, S.; Pandita, T.K. Altered telomere nuclear matrix interactions and nucleosomal periodicity in ataxia telangiectasia cells before and after ionizing radiation treatment. *Mol. Cell. Biol.* **1999**, *19*, 6963–6971. [[CrossRef](#)]
- Lejnine, S.; Makarov, V.L.; Langmore, J.P. Conserved nucleoprotein structure at the ends of vertebrate and invertebrate chromosomes. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 2393–2397. [[CrossRef](#)]
- Fajkus, J.; Kovarik, A.; Královics, R.; Bezděk, M. Organization of telomeric and subtelomeric chromatin in the higher plant *Nicotiana tabacum*. *Mol. Gen. Genet.* **1995**, *247*, 633–638. [[CrossRef](#)]
- Sýkorová, E.; Fajkus, J.; Ito, M.; Fukui, K. Transition between two forms of heterochromatin at plant subtelomeres. *Chromosome Res.* **2001**, *9*, 309–323. [[CrossRef](#)] [[PubMed](#)]
- Vershinin, A.V.; Heslop-Harrison, J.S. Comparative analysis of the nucleosomal structure of rye, wheat and their relatives. *Plant Mol. Biol.* **1998**, *36*, 149–161. [[CrossRef](#)] [[PubMed](#)]
- Rando, O.J.; Ahmad, K. Rules and regulation in the primary structure of chromatin. *Curr. Opin. Cell Biol.* **2007**, *19*, 250–256. [[CrossRef](#)]
- Struhl, K.; Segal, E. Determinants of nucleosome positioning. *Nat. Struct. Mol. Biol.* **2013**, *20*, 267–273. [[CrossRef](#)] [[PubMed](#)]
- Tillo, D.; Hughes, T.R. G+C content dominates intrinsic nucleosome occupancy. *BMC Bioinform.* **2009**, *10*, 442. [[CrossRef](#)] [[PubMed](#)]
- Field, Y.; Fondufe-Mittendorf, Y.; Moore, I.K.; Mieczkowski, P.; Kaplan, N.; Lubling, Y.; Lieb, J.D.; Widom, J.; Segal, E. Gene expression divergence in yeast is coupled to evolution of DNA-encoded nucleosome organization. *Nat. Genet.* **2009**, *41*, 438–445. [[CrossRef](#)] [[PubMed](#)]
- Tsankov, A.M.; Thompson, D.A.; Socha, A.; Regev, A.; Rando, O.J. The role of nucleosome positioning in the evolution of gene regulation. *PLoS Biol.* **2010**, *8*, e1000414. [[CrossRef](#)]
- Chen, J.; Li, E.; Lai, J. The coupled effect of nucleosome organization on gene transcription level and transcriptional plasticity. *Nucleus* **2017**, *8*, 605–612. [[CrossRef](#)] [[PubMed](#)]
- Chen, J.; Li, E.; Zhang, X.; Dong, X.; Lei, L.; Song, W.; Zhao, H.; Lai, J. Genome-wide Nucleosome Occupancy and Organization Modulates the Plasticity of Gene Transcriptional Status in Maize. *Mol. Plant* **2017**, *10*, 962–974. [[CrossRef](#)]
- Prendergast, J.G.D.; Semple, C.A.M. Widespread signatures of recent selection linked to nucleosome positioning in the human lineage. *Genome Res.* **2011**, *21*, 1777–1787. [[CrossRef](#)]

28. Drillon, G.; Audit, B.; Argoul, F.; Arneodo, A. Evidence of selection for an accessible nucleosomal array in human. *BMC Genom.* **2016**, *17*, 526. [[CrossRef](#)]
29. Tompitak, M.; Vaillant, C.; Schiessel, H. Genomes of Multicellular Organisms Have Evolved to Attract Nucleosomes to Promoter Regions. *Biophys. J.* **2017**, *112*, 505–511. [[CrossRef](#)]
30. Warnecke, T.; Becker, E.A.; Facciotti, M.T.; Nislow, C.; Lehner, B. Conserved substitution patterns around nucleosome footprints in eukaryotes and Archaea derive from frequent nucleosome repositioning through evolution. *PLoS Comput. Biol.* **2013**, *9*, e1003373. [[CrossRef](#)]
31. Gonzalez-Perez, A.; Sabarinathan, R.; Lopez-Bigas, N. Local Determinants of the Mutational Landscape of the Human Genome. *Cell* **2019**, *177*, 101–114. [[CrossRef](#)] [[PubMed](#)]
32. Widlund, H.R.; Cao, H.; Simonsson, S.; Magnusson, E.; Simonsson, T.; Nielsen, P.E.; Kahn, J.D.; Crothers, D.M.; Kubista, M. Identification and characterization of genomic nucleosome-positioning sequences. *J. Mol. Biol.* **1997**, *267*, 807–817. [[CrossRef](#)] [[PubMed](#)]
33. Lowary, P.T.; Widom, J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.* **1998**, *276*, 19–42. [[CrossRef](#)]
34. Thåström, A.; Bingham, L.M.; Widom, J. Nucleosomal locations of dominant DNA sequence motifs for histone-DNA interactions and nucleosome positioning. *J. Mol. Biol.* **2004**, *338*, 695–709. [[CrossRef](#)] [[PubMed](#)]
35. Gencheva, M.; Boa, S.; Fraser, R.; Simmen, M.W.; A Whitelaw, C.B.; Allan, J. In Vitro and in Vivo nucleosome positioning on the ovine beta-lactoglobulin gene are related. *J. Mol. Biol.* **2006**, *361*, 216–230. [[CrossRef](#)]
36. Parmar, J.J.; Marko, J.F.; Padinhateeri, R. Nucleosome positioning and kinetics near transcription-start-site barriers are controlled by interplay between active remodeling and DNA sequence. *Nucleic Acids Res.* **2014**, *42*, 128–136. [[CrossRef](#)] [[PubMed](#)]
37. Sexton, B.S.; Avey, D.; Druliner, B.R.; Fincher, J.A.; Vera, D.L.; Grau, D.J.; Borowsky, M.L.; Gupta, S.; Girmurugan, S.B.; Chicken, E.; et al. The spring-loaded genome: Nucleosome redistributions are widespread, transient, and DNA-directed. *Genome Res.* **2014**, *24*, 251–259. [[CrossRef](#)]
38. Vavouri, T.; Lehner, B. Chromatin organization in sperm may be the major functional consequence of base composition variation in the human genome. *PLoS Genet* **2011**, *7*, e1002036. [[CrossRef](#)]
39. Zhang, Z.; Wippo, C.J.; Wal, M.; Ward, E.; Korber, P.; Pugh, B.F. A packing mechanism for nucleosome organization reconstituted across a eukaryotic genome. *Science* **2011**, *332*, 977–980. [[CrossRef](#)]
40. Sulkowski, E.; Laskowski, M. Mechanism of action of micrococcal nuclease on deoxyribonucleic acid. *J. Biol. Chem.* **1962**, *237*, 2620–2625. [[CrossRef](#)]
41. Axel, R. Cleavage of DNA in nuclei and chromatin with staphylococcal nuclease. *Biochemistry* **1975**, *14*, 2921–2925. [[CrossRef](#)]
42. Pajoro, A.; Muiño, J.M.; Angenent, G.C.; Kaufmann, K. Profiling Nucleosome Occupancy by MNase-seq: Experimental Protocol and Computational Analysis. *Methods Mol. Biol.* **2018**, *1675*, 167–181. [[CrossRef](#)]
43. Yuan, G.C.; Liu, Y.J.; Dion, M.F.; Slack, M.D.; Wu, L.F.; Altschuler, S.J.; Rando, O.J. Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science* **2005**, *309*, 626–630. [[CrossRef](#)] [[PubMed](#)]
44. Schones, D.E.; Cui, K.; Cuddapah, S.; Roh, T.Y.; Barski, A.; Wang, Z.; Wei, G.; Zhao, K. Dynamic regulation of nucleosome positioning in the human genome. *Cell* **2008**, *132*, 887–898. [[CrossRef](#)]
45. Gaffney, D.J.; McVicker, G.; Pai, A.A.; Fondufe-Mittendorf, Y.N.; Lewellen, N.; Michelini, K.; Widom, J.; Gilad, Y.; Pritchard, J.K. Controls of nucleosome positioning in the human genome. *PLoS Genet.* **2012**, *8*, e1003036. [[CrossRef](#)] [[PubMed](#)]
46. Lai, B.; Gao, W.; Cui, K.; Xie, W.; Tang, Q.; Jin, W.; Hu, G.; Ni, B.; Zhao, K. Principles of nucleosome organization revealed by single-cell micrococcal nuclease sequencing. *Nature* **2018**, *562*, 281–285. [[CrossRef](#)]
47. Oberbeckmann, E.; Wolff, M.; Krietenstein, N.; Heron, M.; Ellins, J.L.; Schmid, A.; Krebs, S.; Blum, H.; Gerland, U.; Korber, P. Absolute nucleosome occupancy map for the *Saccharomyces cerevisiae* genome. *Genome Res.* **2019**, *29*, 1996–2009. [[CrossRef](#)] [[PubMed](#)]
48. Lee, W.; Tillo, D.; Bray, N.; Morse, R.H.; Davis, R.W.; Hughes, T.R.; Nislow, C. A high-resolution atlas of nucleosome occupancy in yeast. *Nat. Genet.* **2007**, *39*, 1235–1244. [[CrossRef](#)]
49. Mavrich, T.N.; Jiang, C.; Ioshikhes, I.P.; Li, X.; Venters, B.J.; Zanton, S.J.; Tomsho, L.P.; Qi, J.; Glaser, R.; Schuster, S.C.; et al. Nucleosome organization in the *Drosophila* genome. *Nature* **2008**, *453*, 358–362. [[CrossRef](#)] [[PubMed](#)]
50. Zhang, T.; Zhang, W.; Jiang, J. Genome-Wide Nucleosome Occupancy and Positioning and Their Impact on Gene Expression and Evolution in Plants. *Plant Physiol.* **2015**, *168*, 1406–1416. [[CrossRef](#)] [[PubMed](#)]
51. Teif, V.B.; Vainshtein, Y.; Caudron-Herger, M.; Mallm, J.P.; Marth, C.; Höfer, T.; Rippe, K. Genome-wide nucleosome positioning during embryonic stem cell development. *Nat. Struct. Mol. Biol.* **2012**, *19*, 1185–1192. [[CrossRef](#)]
52. Jeffers, T.E.; Lieb, J.D. Nucleosome fragility is associated with future transcriptional response to developmental cues and stress in *C. elegans*. *Genome Res.* **2017**, *27*, 75–86. [[CrossRef](#)]
53. Kaplan, N.; Moore, I.K.; Fondufe-Mittendorf, Y.; Gossett, A.J.; Tillo, D.; Field, Y.; LeProust, E.M.; Hughes, T.R.; Lieb, J.D.; Widom, J.; et al. The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* **2009**, *458*, 362–366. [[CrossRef](#)] [[PubMed](#)]
54. Teif, V.B. Nucleosome positioning: Resources and tools online. *Brief. Bioinform.* **2016**, *17*, 745–757. [[CrossRef](#)] [[PubMed](#)]
55. Chevereau, G.; Arnéodo, A.; Vaillant, C. Influence of the genomic sequence on the primary structure of chromatin. *Front. Life Sci.* **2011**, *5*, 29–68. [[CrossRef](#)]

56. Segal, E.; Fondufe-Mittendorf, Y.; Chen, L.; Thåström, A.; Field, Y.; Moore, I.K.; Wang, J.P.Z.; Widom, J. A genomic code for nucleosome positioning. *Nature* **2006**, *442*, 772–778. [[CrossRef](#)] [[PubMed](#)]
57. Vaillant, C.; Audit, B.; Arneodo, A. Experiments confirm the influence of genome long-range correlations on nucleosome positioning. *Phys. Rev. Lett.* **2007**, *99*, 218103. [[CrossRef](#)]
58. Field, Y.; Kaplan, N.; Fondufe-Mittendorf, Y.; Moore, I.K.; Sharon, E.; Lubling, Y.; Widom, J.; Segal, E. Distinct Modes of Regulation by Chromatin Encoded through Nucleosome Positioning Signals. *PLoS Comput. Biol.* **2008**, *4*. [[CrossRef](#)]
59. Miele, V.; Vaillant, C.; d'Aubenton Carafa, Y.; Thermes, C.; Grange, T. DNA physical properties determine nucleosome occupancy from yeast to fly. *Nucleic Acids Res.* **2008**, *36*, 3746–3756. [[CrossRef](#)]
60. Scipioni, A.; De Santis, P. Predicting nucleosome positioning in genomes: Physical and bioinformatic approaches. *Biophys. Chem.* **2011**, *155*, 53–64. [[CrossRef](#)]
61. Kaplan, N.; Hughes, T.R.; Lieb, J.D.; Widom, J.; Segal, E. Contribution of histone sequence preferences to nucleosome organization: Proposed definitions and methodology. *Genome Biol.* **2010**, *11*, 140. [[CrossRef](#)]
62. Trifonov, E.N.; Sussman, J.L. The pitch of chromatin DNA is reflected in its nucleotide sequence. *Proc. Natl. Acad. Sci. USA* **1980**, *77*, 3816–3820. [[CrossRef](#)] [[PubMed](#)]
63. Trifonov, E.N. Sequence-dependent deformational anisotropy of chromatin DNA. *Nucleic Acids Res.* **1980**, *8*, 4041–4053. [[CrossRef](#)]
64. Levene, S.D.; Crothers, D.M. A computer graphics study of sequence-directed bending in DNA. *J. Biomol. Struct. Dyn.* **1983**, *1*, 429–435. [[CrossRef](#)] [[PubMed](#)]
65. Rhodes, D. Nucleosome cores reconstituted from poly (dA-dT) and the octamer of histones. *Nucleic Acids Res.* **1979**, *6*, 1805–1816. [[CrossRef](#)]
66. Simpson, R.T.; Künzler, P. Chromatin and core particles formed from the inner histones and synthetic polydeoxyribonucleotides of defined sequence. *Nucleic Acids Res.* **1979**, *6*, 1387–1415. [[CrossRef](#)]
67. Drew, H.R.; Travers, A.A. DNA bending and its relation to nucleosome positioning. *J. Mol. Biol.* **1985**, *186*, 773–790. [[CrossRef](#)]
68. Satchwell, S.C.; Drew, H.R.; Travers, A.A. Sequence periodicities in chicken nucleosome core DNA. *J. Mol. Biol.* **1986**, *191*, 659–675. [[CrossRef](#)]
69. Johnson, S.M.; Tan, F.J.; McCullough, H.L.; Riordan, D.P.; Fire, A.Z. Flexibility and constraint in the nucleosome core landscape of *Caenorhabditis elegans* chromatin. *Genome Res.* **2006**, *16*, 1505–1516. [[CrossRef](#)]
70. Albert, I.; Mavrich, T.N.; Tomsho, L.P.; Qi, J.; Zanton, S.J.; Schuster, S.C.; Pugh, B.F. Translational and rotational settings of H2A.Z nucleosomes across the *Saccharomyces cerevisiae* genome. *Nature* **2007**, *446*, 572–576. [[CrossRef](#)] [[PubMed](#)]
71. Raveh-Sadka, T.; Levo, M.; Shabi, U.; Shany, B.; Keren, L.; Lotan-Pompan, M.; Zeevi, D.; Sharon, E.; Weinberger, A.; Segal, E. Manipulating nucleosome disfavoring sequences allows fine-tune regulation of gene expression in yeast. *Nat. Genet.* **2012**, *44*, 743–750. [[CrossRef](#)]
72. Fedor, M.J.; Lue, N.F.; Kornberg, R.D. Statistical positioning of nucleosomes by specific protein-binding to an upstream activating sequence in yeast. *J. Mol. Biol.* **1988**, *204*, 109–127. [[CrossRef](#)]
73. Kornberg, R.D.; Stryer, L. Statistical distributions of nucleosomes: Nonrandom locations by a stochastic mechanism. *Nucleic Acids Res.* **1988**, *16*, 6677–6690. [[CrossRef](#)]
74. Iyer, V.R. Nucleosome positioning: Bringing order to the eukaryotic genome. *Trends Cell Biol.* **2012**, *22*, 250–256. [[CrossRef](#)]
75. Milani, P.; Chevereau, G.; Vaillant, C.; Audit, B.; Haftek-Terreau, Z.; Marilley, M.; Bouvet, P.; Argoul, F.; Arneodo, A. Nucleosome positioning by genomic excluding-energy barriers. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 22257–22262. [[CrossRef](#)] [[PubMed](#)]
76. Drillon, G.; Audit, B.; Argoul, F.; Arneodo, A. Ubiquitous human ‘master’ origins of replication are encoded in the DNA sequence via a local enrichment in nucleosome excluding energy barriers. *J. Phys. Condens. Matter* **2015**, *27*, 064102. [[CrossRef](#)] [[PubMed](#)]
77. Tompitak, M.; Barkema, G.T.; Schiessel, H. Benchmarking and refining probability-based models for nucleosome-DNA interaction. *BMC Bioinform.* **2017**, *18*, 157. [[CrossRef](#)] [[PubMed](#)]
78. Nishida, H. Nucleosome Positioning. *ISRN Mol. Biol.* **2012**, *2012*, 245706. [[CrossRef](#)]
79. Chevereau, G.; Palmeira, L.; Thermes, C.; Arneodo, A.; Vaillant, C. Thermodynamics of intragenic nucleosome ordering. *Phys. Rev. Lett.* **2009**, *103*, 188103. [[CrossRef](#)]
80. Vaillant, C.; Palmeira, L.; Chevereau, G.; Audit, B.; d'Aubenton Carafa, Y.; Thermes, C.; Arneodo, A. A novel strategy of transcription regulation by intragenic nucleosome ordering. *Genome Res.* **2010**, *20*, 59–67. [[CrossRef](#)]
81. Brogaard, K.; Xi, L.; Wang, J.P.; Widom, J. A map of nucleosome positions in yeast at base-pair resolution. *Nature* **2012**, *486*, 496–501. [[CrossRef](#)]
82. Shivaswamy, S.; Bhinge, A.; Zhao, Y.; Jones, S.; Hirst, M.; Iyer, V.R. Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation. *PLoS Biol.* **2008**, *6*, e65. [[CrossRef](#)]
83. Bochkis, I.M.; Przybylski, D.; Chen, J.; Regev, A. Changes in nucleosome occupancy associated with metabolic alterations in aged mammalian liver. *Cell Rep.* **2014**, *9*, 996–1006. [[CrossRef](#)] [[PubMed](#)]
84. Hu, Z.; Chen, K.; Xia, Z.; Chavez, M.; Pal, S.; Seol, J.H.; Chen, C.C.; Li, W.; Tyler, J.K. Nucleosome loss leads to global transcriptional up-regulation and genomic instability during yeast aging. *Genes Dev.* **2014**, *28*, 396–408. [[CrossRef](#)]
85. Sexton, B.S.; Druliner, B.R.; Vera, D.L.; Avey, D.; Zhu, F.; Dennis, J.H. Hierarchical regulation of the genome: Global changes in nucleosome organization potentiate genome response. *Oncotarget* **2016**, *7*, 6460–6475. [[CrossRef](#)] [[PubMed](#)]
86. Weghorn, D.; Lässig, M. Fitness landscape for nucleosome positioning. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 10988–10993. [[CrossRef](#)] [[PubMed](#)]

87. Eslami-Mossallam, B.; Schram, R.D.; Tompitak, M.; van Noort, J.; Schiessel, H. Multiplexing Genetic and Nucleosome Positioning Codes: A Computational Approach. *PLoS ONE* **2016**, *11*, e0156905. [[CrossRef](#)] [[PubMed](#)]
88. 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. MNase titration reveals differences between nucleosome occupancy and chromatin accessibility. *Nat. Commun.* **2016**, *7*, 1–11. [[CrossRef](#)] [[PubMed](#)]
89. Chereji, R.V.; Bryson, T.D.; Henikoff, S. Quantitative MNase-seq accurately maps nucleosome occupancy levels. *Genome Biol.* **2019**, *20*, 198. [[CrossRef](#)]
90. Arneodo, A.; Drillon, G.; Argoul, F.; Audit, B. 2-The Role of Nucleosome Positioning in Genome Function and Evolution. In *Nuclear Architecture and Dynamics; Translational Epigenetics*; Lavelle, C., Victor, J.M., Eds.; Academic Press: Boston, MA, USA, 2018; Volume 2, pp. 41–79. [[CrossRef](#)]
91. Saxonov, S.; Berg, P.; Brutlag, D.L. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 1412–1417. [[CrossRef](#)]
92. Ramirez-Carrozzi, V.R.; Braas, D.; Bhatt, D.M.; Cheng, C.S.; Hong, C.; Doty, K.R.; Black, J.C.; Hoffmann, A.; Carey, M.; Smale, S.T. A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. *Cell* **2009**, *138*, 114–128. [[CrossRef](#)]
93. Deaton, A.M.; Bird, A. CpG islands and the regulation of transcription. *Genes Dev.* **2011**, *25*, 1010–1022. [[CrossRef](#)]
94. Sandman, K.; Reeve, J.N. Structure and functional relationships of archaeal and eukaryal histones and nucleosomes. *Arch. Microbiol.* **2000**, *173*, 165–169. [[CrossRef](#)]
95. Brunet, F.G.; Audit, B.; Drillon, G.; Argoul, F.; Volff, J.N.; Arneodo, A. Evidence for DNA Sequence Encoding of an Accessible Nucleosomal Array across Vertebrates. *Biophys. J.* **2018**. [[CrossRef](#)]
96. Durbin, R.M.; Altshuler, D.; Durbin, R.M.; Abecasis, G.R.; Bentley, D.R.; Chakravarti, A.; Clark, A.G.; Collins, F.S.; De La Vega, F.M.; Donnelly, P.; et al. A map of human genome variation from population-scale sequencing. *Nature* **2010**, *467*, 1061–1073. [[CrossRef](#)]
97. Biémont, C.; Vieira, C. Junk DNA as an evolutionary force. *Nature* **2006**, *443*, 521–524. [[CrossRef](#)]
98. Bourque, G.; Burns, K.H.; Gehring, M.; Gorbunova, V.; Seluanov, A.; Hammell, M.; Imbeault, M.; Izsvák, Z.; Levin, H.L.; Macfarlan, T.S.; et al. Ten things you should know about transposable elements. *Genome Biol.* **2018**, *19*, 199. [[CrossRef](#)]
99. Pace, J.K.; Feschotte, C. The evolutionary history of human DNA transposons: Evidence for intense activity in the primate lineage. *Genome Res.* **2007**, *17*, 422–432. [[CrossRef](#)] [[PubMed](#)]
100. Chalopin, D.; Naville, M.; Plard, F.; Galiana, D.; Volff, J.N. Comparative Analysis of Transposable Elements Highlights Mobilome Diversity and Evolution in Vertebrates. *Genome Biol. Evol.* **2015**, *7*, 567–580. [[CrossRef](#)] [[PubMed](#)]
101. Wicker, T.; Sabot, F.; Hua-Van, A.; Bennetzen, J.L.; Capy, P.; Chalhoub, B.; Flavell, A.; Leroy, P.; Morgante, M.; Panaud, O.; et al. A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **2007**, *8*, 973–982. [[CrossRef](#)] [[PubMed](#)]
102. Etchegaray, E.; Naville, M.; Volff, J.N.; Haftek-Terreau, Z. Transposable element-derived sequences in vertebrate development. *Mob. DNA* **2021**, *12*, 1. [[CrossRef](#)]
103. Dechaud, C.; Volff, J.N.; Schartl, M.; Naville, M. Sex and the TEs: Transposable elements in sexual development and function in animals. *Mob. DNA* **2019**, *10*. [[CrossRef](#)] [[PubMed](#)]
104. Payer, L.M.; Burns, K.H. Transposable elements in human genetic disease. *Nat. Rev. Genet.* **2019**, *20*, 760–772. [[CrossRef](#)] [[PubMed](#)]
105. Deininger, P. Alu elements: Know the SINES. *Genome Biol.* **2011**, *12*, 236. [[CrossRef](#)]
106. Tanaka, Y.; Yamashita, R.; Suzuki, Y.; Nakai, K. Effects of Alu elements on global nucleosome positioning in the human genome. *BMC Genom.* **2010**, *11*, 309. [[CrossRef](#)] [[PubMed](#)]
107. Li, C.; Luscombe, N.M. Nucleosome positioning stability is a modulator of germline mutation rate variation across the human genome. *Nat. Commun.* **2020**, *11*, 1363. [[CrossRef](#)] [[PubMed](#)]
108. Hara, R.; Mo, J.; Sancar, A. DNA damage in the nucleosome core is refractory to repair by human excision nuclease. *Mol. Cell. Biol.* **2000**, *20*, 9173–9181. [[CrossRef](#)] [[PubMed](#)]
109. Rodriguez, Y.; Smerdon, M.J. The structural location of DNA lesions in nucleosome core particles determines accessibility by base excision repair enzymes. *J. Biol. Chem.* **2013**, *288*, 13863–13875. [[CrossRef](#)]
110. Smerdon, M.J.; Conconi, A. Modulation of DNA damage and DNA repair in chromatin. *Prog. Nucleic Acid Res. Mol. Biol.* **1999**, *62*, 227–255. [[CrossRef](#)]
111. Pedersen, J.S.; Valen, E.; Velazquez, A.M.V.; Parker, B.J.; Rasmussen, M.; Lindgreen, S.; Lilje, B.; Tobin, D.J.; Kelly, T.K.; Vang, S.; et al. Genome-wide nucleosome map and cytosine methylation levels of an ancient human genome. *Genome Res.* **2014**, *24*, 454–466. [[CrossRef](#)]
112. Hanghøj, K.; Seguin-Orlando, A.; Schubert, M.; Madsen, T.; Pedersen, J.S.; Willerslev, E.; Orlando, L. Fast, Accurate and Automatic Ancient Nucleosome and Methylation Maps with epiPALEOMIX. *Mol. Biol. Evol.* **2016**, *33*, 3284–3298. [[CrossRef](#)]
113. Chen, X.; Chen, Z.; Chen, H.; Su, Z.; Yang, J.; Lin, F.; Shi, S.; He, X. Nucleosomes suppress spontaneous mutations base-specifically in eukaryotes. *Science* **2012**, *335*, 1235–1238. [[CrossRef](#)] [[PubMed](#)]
114. Makova, K.D.; Hardison, R.C. The effects of chromatin organization on variation in mutation rates in the genome. *Nat. Rev. Genet.* **2015**, *16*, 213–223. [[CrossRef](#)]

115. Wu, J.; McKeague, M.; Sturla, S.J. Nucleotide-Resolution Genome-Wide Mapping of Oxidative DNA Damage by Click-Code-Seq. *J. Am. Chem. Soc.* **2018**, *140*, 9783–9787. [[CrossRef](#)]
116. Morganella, S.; Alexandrov, L.B.; Glodzik, D.; Zou, X.; Davies, H.; Staaf, J.; Sieuwerts, A.M.; Brinkman, A.B.; Martin, S.; Ramakrishna, M.; et al. The topography of mutational processes in breast cancer genomes. *Nat. Commun.* **2016**, *7*, 11383. [[CrossRef](#)] [[PubMed](#)]
117. Nik-Zainal, S.; Morganella, S. Mutational Signatures in Breast Cancer: The Problem at the DNA Level. *Clin. Cancer Res.* **2017**, *23*, 2617–2629. [[CrossRef](#)]
118. ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. *Nature* **2020**, *578*, 82–93. [[CrossRef](#)] [[PubMed](#)]
119. Alexandrov, L.B.; Kim, J.; Haradhvala, N.J.; Huang, M.N.; Tian Ng, A.W.; Wu, Y.; Boot, A.; Covington, K.R.; Gordenin, D.A.; Bergstrom, E.N.; et al. The repertoire of mutational signatures in human cancer. *Nature* **2020**, *578*, 94–101. [[CrossRef](#)] [[PubMed](#)]
120. Sabarinathan, R.; Mularoni, L.; Deu-Pons, J.; Gonzalez-Perez, A.; López-Bigas, N. Nucleotide excision repair is impaired by binding of transcription factors to DNA. *Nature* **2016**, *532*, 264–267. [[CrossRef](#)]
121. Pich, O.; Muiños, F.; Sabarinathan, R.; Reyes-Salazar, I.; Gonzalez-Perez, A.; Lopez-Bigas, N. Somatic and Germline Mutation Periodicity Follow the Orientation of the DNA Minor Groove around Nucleosomes. *Cell* **2018**, *175*, 1074–1087.e18. [[CrossRef](#)]
122. Langley, S.A.; Karpen, G.H.; Langley, C.H. Nucleosomes shape DNA polymorphism and divergence. *PLoS Genet.* **2014**, *10*, e1004457. [[CrossRef](#)]
123. Mao, P.; Brown, A.J.; Malc, E.P.; Mieczkowski, P.A.; Smerdon, M.J.; Roberts, S.A.; Wyrick, J.J. Genome-wide maps of alkylation damage, repair, and mutagenesis in yeast reveal mechanisms of mutational heterogeneity. *Genome Res.* **2017**, *27*, 1674–1684. [[CrossRef](#)] [[PubMed](#)]
124. Mao, P.; Wyrick, J.J. Organization of DNA damage, excision repair, and mutagenesis in chromatin: A genomic perspective. *DNA Repair* **2019**, *81*, 102645. [[CrossRef](#)] [[PubMed](#)]
125. Gale, J.M.; Nissen, K.A.; Smerdon, M.J. UV-induced formation of pyrimidine dimers in nucleosome core DNA is strongly modulated with a period of 10.3 bases. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 6644–6648. [[CrossRef](#)] [[PubMed](#)]
126. Mao, P.; Wyrick, J.J.; Roberts, S.A.; Smerdon, M.J. UV-Induced DNA Damage and Mutagenesis in Chromatin. *Photochem. Photobiol.* **2017**, *93*, 216–228. [[CrossRef](#)] [[PubMed](#)]
127. Mao, P.; Smerdon, M.J.; Roberts, S.A.; Wyrick, J.J. Chromosomal landscape of UV damage formation and repair at single-nucleotide resolution. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 9057–9062. [[CrossRef](#)]
128. Brown, A.J.; Mao, P.; Smerdon, M.J.; Wyrick, J.J.; Roberts, S.A. Nucleosome positions establish an extended mutation signature in melanoma. *PLoS Genet.* **2018**, *14*. [[CrossRef](#)]
129. Hayward, N.K.; Wilmott, J.S.; Waddell, N.; Johansson, P.A.; Field, M.A.; Nones, K.; Patch, A.M.; Kakavand, H.; Alexandrov, L.B.; Burke, H.; et al. Whole-genome landscapes of major melanoma subtypes. *Nature* **2017**, *545*, 175–180. [[CrossRef](#)]
130. Pestov, N.A.; Gerasimova, N.S.; Kulaeva, O.I.; Studitsky, V.M. Structure of transcribed chromatin is a sensor of DNA damage. *Sci. Adv.* **2015**, *1*, e1500021. [[CrossRef](#)]
131. Adar, S.; Hu, J.; Lieb, J.D.; Sancar, A. Genome-wide kinetics of DNA excision repair in relation to chromatin state and mutagenesis. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E2124–E2133. [[CrossRef](#)]
132. Li, W.; Hu, J.; Adebali, O.; Adar, S.; Yang, Y.; Chiou, Y.Y.; Sancar, A. Human genome-wide repair map of DNA damage caused by the cigarette smoke carcinogen benzo[a]pyrene. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 6752–6757. [[CrossRef](#)]
133. Polak, P.; Karlić, R.; Koren, A.; Thurman, R.; Sandstrom, R.; Lawrence, M.; Reynolds, A.; Rynes, E.; Vlahoviček, K.; Stamatoyannopoulos, J.A.; et al. Cell-of-origin chromatin organization shapes the mutational landscape of cancer. *Nature* **2015**, *518*, 360–364. [[CrossRef](#)] [[PubMed](#)]
134. Hu, J.; Lieb, J.D.; Sancar, A.; Adar, S. Cisplatin DNA damage and repair maps of the human genome at single-nucleotide resolution. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 11507–11512. [[CrossRef](#)] [[PubMed](#)]
135. Singh, A.K.; Mueller-Planitz, F. Nucleosome Positioning and Spacing: From Mechanism to Function. *J. Mol. Biol.* **2021**, *433*, 166847. [[CrossRef](#)] [[PubMed](#)]