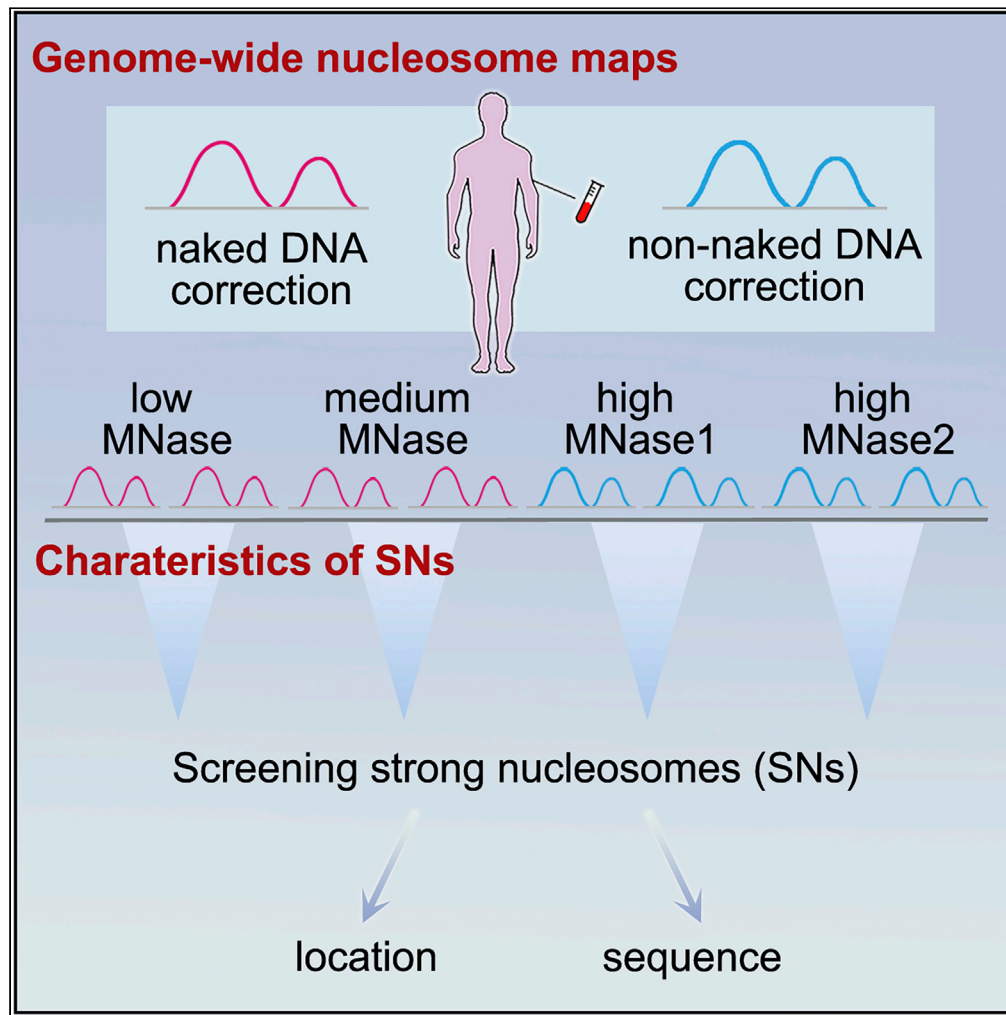


Article

A study of strong nucleosomes in the human genome



Lin Wang,  
Chunnan Dong,  
Chaolong Lu,  
Shujin Li, Lihong  
Fu, Bin Cong

hbydbincong@126.com

**Highlights**

Naked DNA correction improved the accuracy of nucleosome map in partial digestion

Level of MNase digestion has effects on nucleosome organization

A type of strong nucleosomes (SNs) exist across different nucleosome maps

Nucleosome stability may be related to its location and the DNA sequence

Wang et al., iScience 25, 104593  
July 15, 2022 © 2022 The Author(s).  
<https://doi.org/10.1016/j.isci.2022.104593>



## Article

## A study of strong nucleosomes in the human genome

Lin Wang,<sup>1,3</sup> Chunnan Dong,<sup>2,3</sup> Chaolong Lu,<sup>1</sup> Shujin Li,<sup>1</sup> Lihong Fu,<sup>1</sup> and Bin Cong<sup>1,4,\*</sup>

## SUMMARY

Micrococcal nuclease (MNase) is widely used to map nucleosomes. However, nucleosomes are highly dynamic and susceptible to experimental conditions, resulting in extreme variability across nucleosome maps, which complicates the generation of accurate nucleosome organization data. We mapped nucleosomes from different individuals using improved MNase-seq. The improvements included setting different digestion levels (low, medium, high) and naked DNA correction to remove the noise caused by experimental manipulation and comparing maps to obtain the accurate position and occupancy of strong nucleosomes (SNs) in the whole genome. In addition, the characteristics of SNs were further excavated. SNs were enriched in Alu elements and near the centromere of Chr12. SNs contain some specific sequences, and the GC content of SNs is different from that of dynamic nucleosomes. The findings suggest that nucleosome location in the genome and the DNA sequence may affect nucleosome stability.

## INTRODUCTION

Nucleosomes are the basic repeating structural unit of eukaryotic chromatin (Kornberg, 1974). Each nucleosome is composed of a histone octamer and 147 bp DNA wrapped around it (Luger et al., 1997). As nucleosomes spatially exclude other proteins that bind to DNA, their organization in the genome will directly affect all DNA-related processes in the nucleus, such as replication, repair, recombination, and transcription (Han and Grunstein, 1988; Jiang and Pugh, 2009).

The organization of nucleosomes is described by nucleosome occupancy and positioning. Nucleosome positioning can be divided into translational positioning and rotational positioning (Albert et al., 2007). The translational positioning refers to the position of the nucleosome dyad in the genome. Nucleosomes with the same translational position in each cell are called "well" or "strongly" positioned, whereas nucleosomes with highly variable positions are considered "weakly" or "fuzzily" positioned (Kaplan et al., 2010a). However, there is no unified definition of strong nucleosomes (SNs). The DNA sequence is the most important factor in determining nucleosomes *in vitro*, so nucleosomes with strongly periodic sequences are called SNs (Salih et al., 2015). *In vivo*, nucleosomes with consistent translational positions across different samples are defined as SNs (Gaffney et al., 2012; Cai and Luscombe, 2020); in this article, we adopt this definition.

The use of micrococcal nuclease (MNase) to cleave the accessible linker DNA between two nucleosomes followed by high-throughput sequencing (MNase-seq) is the most classic and frequent method for mapping the nucleosome position and occupancy *in vivo* (Mieczkowski et al., 2016; Chereji et al., 2016, 2019). However, despite great advances in the accuracy of nucleosome maps, there are serious differences in the precise position of individual nucleosomes *in vivo*. In different studies, and even with different MNase digestion levels in the same study, some well-positioned nucleosomes may be fuzzy or non-existent (Huebert et al., 2012; Bai and Morozov, 2010; Ozonov and Van Nimwegen, 2013). The differences among nucleosome maps hinder the accurate position of nucleosomes and affect the study of the characteristics of nucleosome positioning. The difference is due partly to the dynamic nature of the nucleosome itself (Lai and Pugh, 2017) and partly to the noise caused by experimental methods and data analysis methods (Flores et al., 2014).

To accurately locate the nucleosome in the human genome, we genome-wide screened a special class of SNs that have consistent positioning and occupancy across multiple cell types and different individuals. The research of SNs is an unexplored but promising field (Flores et al., 2014). The analysis of their

<sup>1</sup>College of Forensic Medicine, Hebei Medical University, Hebei Key Laboratory of Forensic Medicine, Shijiazhuang 050017, China

<sup>2</sup>Department of Pathogen Biology, Hebei Medical University, Shijiazhuang 050017, China

<sup>3</sup>These authors contributed equally

<sup>4</sup>Lead contact

\*Correspondence:

hbydbincong@126.com

<https://doi.org/10.1016/j.isci.2022.104593>



characteristics enables us to reach more reliable conclusions than previous studies on common nucleosomes, a more in-depth understanding of the laws and characteristics of nucleosome positioning, the relationship of nucleosomes and gene transcription and other processes and provides a theoretical basis for the application of nucleosomes in forensic analysis (Flores et al., 2014) and archaeology (Flores et al., 2014). In addition, the analysis of their characteristics enables us to have more valuable conclusions than previous studies on common nucleosomes. For example, Cai Li et al. (Cai and Luscombe, 2020) found that a nucleosome with higher translational stability scores also exhibited higher mutation rates and showed the mutational mechanisms linked to SNs. Thus, the research on SNs is important for understanding human diseases and genome evolution.

Here, we obtained accurate nucleosome maps of the human genome and analyzed the stability of nucleosome organization across individuals to investigate the characteristics of nucleosome organization in the population. To be as similar as possible to the human body environment and simplify the operation process, we generated high-resolution personal nucleosome maps on leukocytes isolated from the peripheral blood of four healthy donors. After removing the noise generated by sequencing as much as possible, the samples with the same degree of digestion were compared to find SNs. We found that the nucleosome organization of human leukocytes isolated from different donors was quite similar overall, but there were a variety of situations specific to the individual nucleosomes. Some of the nucleosomes changed remarkably, including the translational position and occupancy change, while some kept the same organization, namely SNs. Furthermore, the proportion of SNs to total nucleosomes was affected by the level of MNase digestion, and the highest proportion of SNs was about 10% under the high level of digestion. Additionally, nucleosome location and sequence characteristics may greatly influence the stability of nucleosomes. The findings have important implications for human genetics and our understanding of genome stability.

## RESULTS

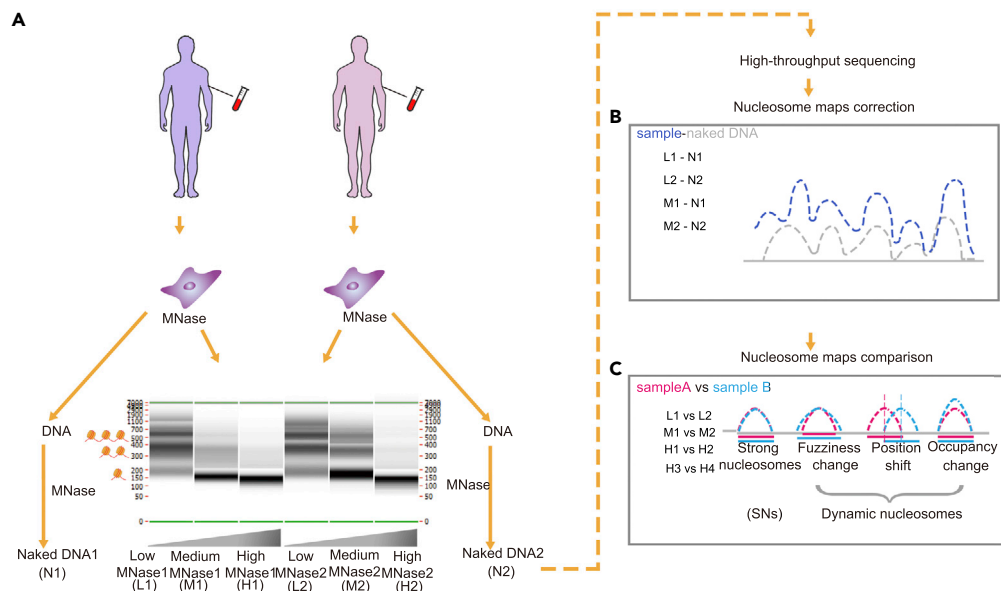
### Generation of nucleosome maps

We digested chromatin from the peripheral blood leukocytes of four healthy donors using three different concentrations of MNase (15 U, 30 U, 120 U) to obtain a more comprehensive map of nucleosomes; the other reaction conditions were the same. In addition, the naked DNA of donors 1 and 2 was prepared and digested to about 150–500 bp fragments by MNase. After digestion, the genome shows typical ladder-like bands on a gel, and the fragments represent mononucleosomes, bionucleosomes, and trinucleosomes from small to large. The digestion level was calculated according to the percentage of the amount of mononucleosome DNA in total DNA. The mononucleosome DNA was chosen for library construction and paired-end sequencing of all samples (Figure 1A). The sequencing depth of the sample is 31 x- 43 x, and the data volume is 91G-136G (Figure S1).

The nucleosome positioning of all samples was analyzed and compared by the software DANPOS. DANPOS (Chen et al., 2013) is a comprehensive bioinformatics pipeline explicitly designed for dynamic nucleosome analysis at single-nucleotide resolution. Owing to the sequence preference, the cutting rate of the AT sequence is 30 times faster than that of GC (Chereji et al., 2019). To eliminate this bias, Gutierrez et al. developed a novel method using the naked DNA signal as the background to correct the low digestion MNase-seq data at the genome-wide level in yeast. The results show that this method can improve the data quality and help to find the differences between samples hidden in the background noise, which can be applied to large-scale experiments (Gutiérrez et al., 2017). We applied it to the human genome, prepared naked DNA 1 and naked DNA 2 (N1 and N2) for sequencing, and corrected the data of partial digestion (L1, M1, L2, M2) (Figure 1B). To eliminate the influence of different digestion levels, we compared the maps with the same digestion level (Figure 1C). We wanted to see whether SNs have consistent translational positioning and occupancy across multiple cell types among different individuals.

### Improvement of nucleosome mapping by subtracting the noise in low and medium digestions

As shown in Figure 2A on IGV, the naked DNA signal has peaks and valleys, indicating that MNase does have a bias for DNA cleavage. We found several differences by comparing the data before and after correction on three digestion levels. Naked DNA correction significantly improved the accuracy of the nucleosome map. There were two different pieces of evidence. First, we could see a cleaner signal from the corrected nucleosome map as in the previous study. The peak became clearer, and the map with two different



**Figure 1. Summary of MNase experiments and nucleosome maps included in the study**

(A) The peripheral blood leukocytes from two healthy donors were digested using three different concentrations of MNase (15 U, 30 U, 120 U). Additionally, the naked DNA of donors 1 and 2 was prepared. The peripheral blood leukocytes from the other two healthy donors were digested with high MNase levels (H3, H4) and are not shown in figure a. The mononucleosomal DNA of digestion products (eight samples) was chosen for sequencing.

(B) Schematic illustration of nucleosome map correction. (Gray axis) DNA sequence. (Blue dashed curves) Nucleosome occupancies before correction. (Gray dashed curves) Nucleosome occupancies in naked DNA. Partial digestion data (L1, L2, M1, M2) were corrected by naked DNA data (N1, N2) by DANPOS.

(C) A cartoon to show a comparison of two maps and four types of nucleosomes. Red and sky-blue dashed curves represent nucleosome occupancy in samples A and B, respectively. Nucleosome maps from two donors with the same digestion level (L1 vs. L2, M1 vs. M2, H1 vs. H2, H3 vs. H4) were compared by DANPOS. Strong nucleosomes (SNs) and dynamic nucleosomes were selected and analyzed. SNs were selected with consistent translational positioning and occupancy between two nucleosome maps. See also [Figures S1](#), [Table S1](#) and [S2](#).

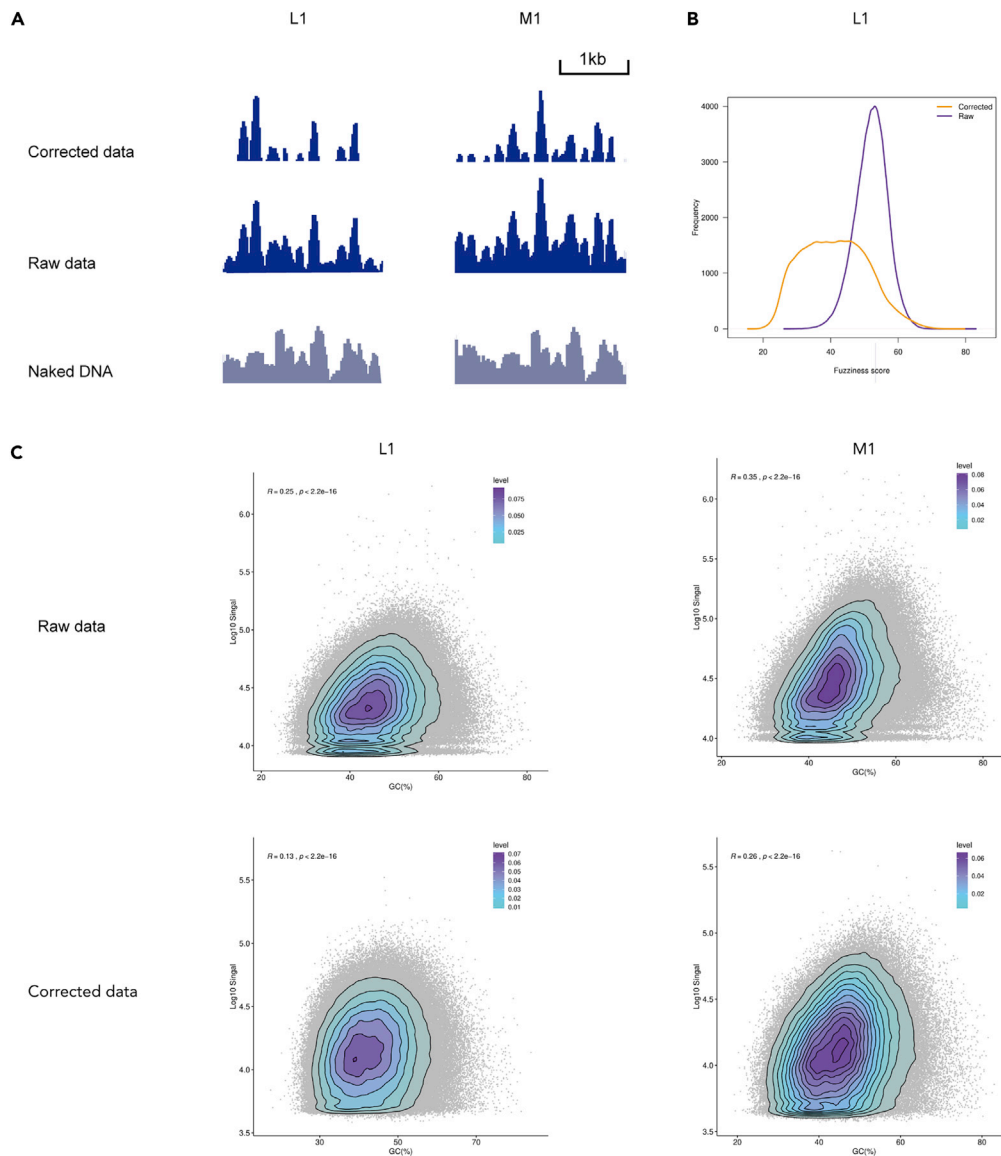
digestion levels had the same pattern ([Figures 2A](#) and [S2A](#)). Second, to further explore whether this phenomenon exists in the whole genome or only in particular locations, we investigated the distribution of the fuzziness score of each sample before and after correction ([Figures 2B](#) and [S2B](#)). We observed that the fuzziness score of the corrected data was significantly lower than that of the raw data, and their distribution was markedly different. This difference means that a substantial part of the apparent fuzziness of raw data was noise owing to the MNase sequence bias.

The correction attenuated the sequence bias of MNase cleavage. We correlated the GC content and nucleosome occupancy level of naked DNA from individual 1. We observed a weak correlation between GC content and occupancy ( $R = 0.23$ ,  $p < 2.2e-16$ ) ([Figure S2C](#)), indicating that the DNA sequence influences MNase cleavage in the human genome. Indeed, there is an AT sequence preference when cutting DNA with MNase. To test whether naked DNA correction could decrease the sequence preference, a correlation analysis was also performed on the raw and corrected data. We found that the correction coefficient at two partial digestion levels (low-MNase and medium-MNase) was reduced ([Figure 2C](#)), indicating that the correction weakens the sequence preference.

Altogether, these results indicated that the naked DNA correction significantly improved data quality and reduced MNase sequence preference. In addition, we demonstrated that naked DNA correction is also effective in complex genomes such as that of humans.

### Level of MNase digestion has effects on nucleosome organization

Because different digestion levels have effects on nucleosome organization, including nucleosome occupancy and precise translational positioning, it is necessary to prepare samples with various digestion levels.



**Figure 2. Naked DNA correction significantly improves the accuracy of nucleosome maps**

(A) The IGV browser snapshot shows the changes in the nucleosome map after correction of the locations randomly selected in the genome. The three snapshots in each column are at the same location on the genome.

(B) Taking one of the eight samples as an example (L1), the fuzziness score was significantly reduced after the naked DNA correction.

(C) The correlation of the log<sub>10</sub> signal intensity and the GC content of fragments. Pearson's correlation is shown ( $p < 2.2e-16$ ). After the naked DNA correction at three MNase digestion levels, the correlation was reduced. See also [Figures S2](#).

As previously observed, the fragment length of mononucleosome bands under different digestion levels varies greatly. Taking the samples of individual 1 as an example, the fragment length under low digestion (20%) was longer, mainly concentrated on 186 bp. Higher the level of digestion, the shorter the fragment length. The fragment length at medium digestion (58%) was 160 bp, and that of the high digestion (86%) was 149 bp ([Figure S3A](#) and [Table S1](#)). MNase first functions as an endonuclease, cutting the genome into 180 and 190 bp fragments. It then acts mainly as an exonuclease, modifying the end of the fragment until the histone octamer blocks it. These results indicated that the digestion level is a very important factor in mapping nucleosome organization, which should be clear when preparing samples. If the comparison is based on samples with different digestion levels, the accuracy of the results will be seriously affected.

Additionally, it could be seen from the base percentage of reads (Figure S3B) that most of the mononucleosomes at low-MNase concentration come from AT-rich regions in the genome. With the progress of digestion, mononucleosomes in GC-rich regions are released from chromatin, and the AT content of mononucleosomes tends to be close to GC content.

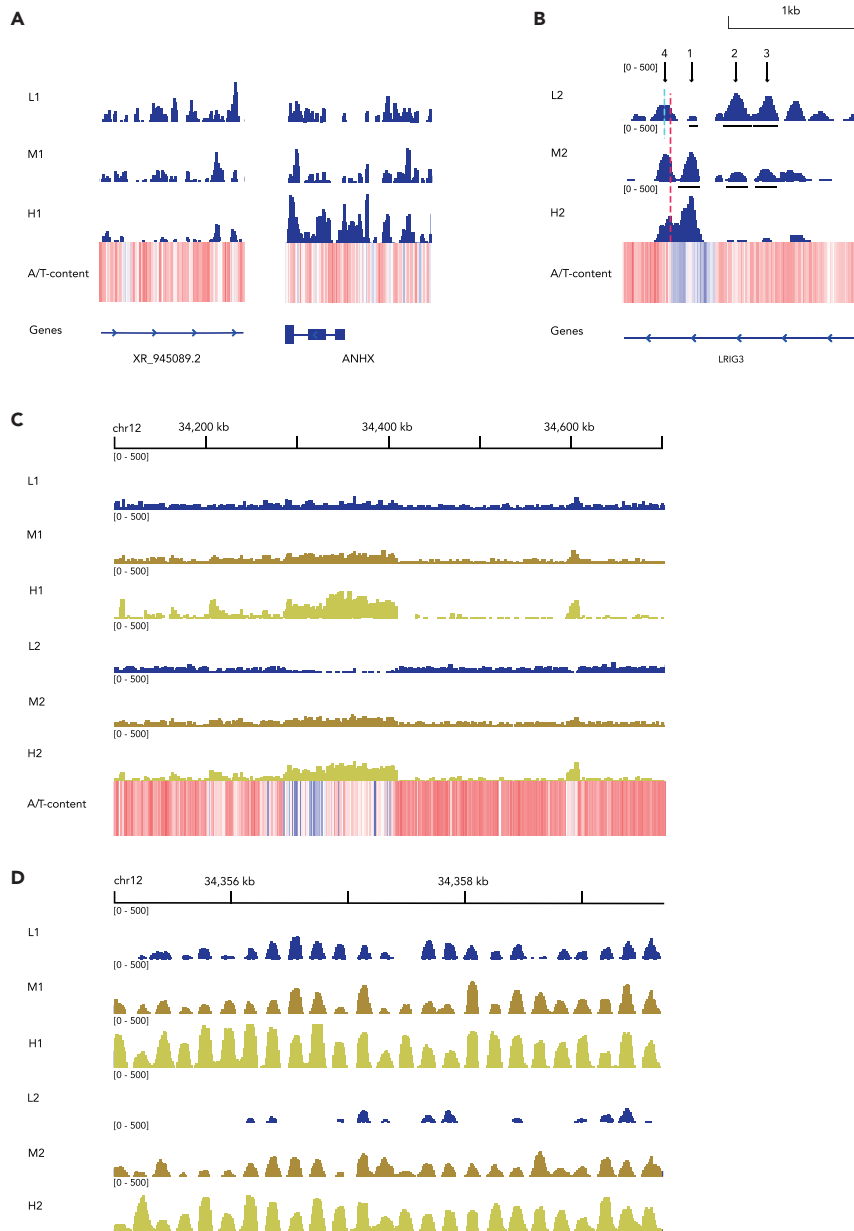
The level of digestion has an influence on nucleosome occupancy. Nucleosome patterns are quite similar in profiles at three MNase digestion conditions. However, there were significant differences among them on individual nucleosomes. The data volume and sequencing depth of each sample were different but similar (Table S1). Therefore, we did not make a quantitative comparison of nucleosome occupancy among different samples but only looked at the changing trend of nucleosome occupancy. In some regions, the height of nucleosome peaks under high-MNase digestion conditions was significantly lower than that at low-MNase digestion conditions, and some nucleosome peaks even disappeared at the high-MNase digestion conditions (Figure 3A). On the contrary, some nucleosome peaks were only found in some regions at high-MNase digestion conditions (Figure 3A). To further investigate whether the difference in nucleosome occupancy is related to the AT content of nucleosome DNA, we prepared a heat map of AT content. Taking the Chr.12 *LRIG3* gene as an example, Peak1 and Peak2 are adjacent to each other in the genome, but the nucleosome occupancy differs among samples at three MNase digestion conditions (Figure 3B). Peak1 had the highest peak value in the low-MNase data and disappeared in the high-MNase data, and the corresponding genome AT content was higher than the average level of the genome. In contrast, peak2 had the highest peak value in the high-MNase data, and the corresponding AT content of the genome was lower than the average level. These results suggest that AT-rich nucleosomes were the first to be released from chromatin and are more accessible to degradation with increasing MNase digestion conditions. At the same time, the GC-rich nucleosomes are released more slowly, and they are more resistant to degradation.

Most studies only focused on the effect of digestion level on nucleosome occupancy, but little attention was paid to nucleosome positioning (Chereji et al., 2019; Mieczkowski et al., 2016). Our results show that there may be subtle differences in nucleosome positioning in addition to the occupancy changes. For example, peak1, 2, and 3 have fuzziness change, and peak4 has position shift under different digestion levels (Figure 3B). We infer that owing to the difference in digestion levels, the number and degree of pruning of the ends of mononucleosome fragments among the three samples were different, leading to different length distributions and dyad and widths of nucleosome peaks. To maintain the consistency between samples (except for the digestion level), although we selected the peripheral blood leukocytes of the same healthy individual simultaneously, we still cannot rule out that these subtle changes may come from inter-replica variation and cell population heterogeneity. Although the difference is subtle, it will directly affect the comparison results of translational position between samples.

The nucleosome array region is more obvious at the high digestion level. It was found that many nucleosomes in the human genome are in regularly arranged arrays, among which the centromeric region of chromosome 12 is a particularly extreme example, containing more than 400 continuously arranged nucleosomes in a single array (Gaffney et al., 2012). This nucleosome array was also found at the same location in samples of ancient human DNA. (Pedersen et al., 2014; Hanghoj et al., 2016). To explore the behavior of this special nucleosome array at different MNase digestion levels, we observed and found a well-positioned nucleosome array region of 121 kb (Chr12.34286–34407 kb) near the centromeric region of Chr12, as in previous studies. Especially at a high digestion level, the occupancy of the nucleosome array region was significantly higher than that of the surrounding region (Figure 3D), and the AT content was significantly lower than that of the surrounding region. For example, the average height of nucleosomes in sample high-MNase1 was 348, higher than the average height of Chr.12, which was 236. This array in the low-MNase data is not as typical as that in high digestion. The centromeric region may be heterochromatin with low chromatin accessibility that is not digested at low enzyme concentrations.

### There are differences in nucleosome maps between different individuals, but strong nucleosomes are kept

We wondered how different the nucleosome profiles were among different individuals. We compared the corrected profiles with the same digestion level by DANPOS (Figure 1C). After minimizing the noise, we found that the nucleosome map of the two individuals was similar in general. However, focusing on individual nucleosomes, there will be many types of subtle changes. The types of dynamic nucleosomes can



**Figure 3. The degree of digestion has influence on the mapping of nucleosomes**

IGV browser shows nucleosome maps at different levels of digestion with MNase.

(A) With the increase in digestion, nucleosome occupancy decreased or even disappeared in some regions. At the same time, there are regions with different behaviors. Nucleosome occupancy increased with the increase of digestion. Heat map represents changes in A/T content along the genome. The white represents the average genome-wide A/T content (~0.58), the red represents the A/T-rich region, and the blue represents the G/C-rich region.

(B) Example of a locus showing profiles obtained from different MNase levels. There are different types of changes even in nearby areas.

(C) and (D) There is a nucleosome array region near the centromere region of Chr12.34286–34407 kb. See also [Figures S3](#).

be divided into three categories: occupancy change, fuzziness change, and position shift, the latter two types describing differences in translational position. Shift change represents the uniform change in the nucleosome dyad of each cell in a cell population, while changes in fuzziness represent the change in the distribution of nucleosome dyad, and some cells of the cell population are shifted. Coverage change means a change in the number of cells with the nucleosome at a given location.

We found that 5% of nucleosomes had changes in occupancy at a low digestion level, 28.35% of nucleosomes had shift changes (shift  $\geq 5$  bp), and 6.41% of nucleosomes had fuzziness changes. Interestingly, the proportion of shift change increased to 41.56% and 41.30% for high-MNase1 and high-MNase2, while the proportion of fuzziness and occupancy change decreased to only 0.05% and 0.05%, 2.14%, and 4.43% at high digestion level, respectively, indicating that the higher the degree of digestion, the more the differences among different individuals showed as shift change, and fuzziness and occupancy change were almost unchanged. Here, a strict screening threshold is used to differentiate between occupancy and fuzziness. If the threshold is lowered, more nucleosomes will be divided into the differential nucleosome. Our results show that the position and occupancy of some nucleosomes in different individuals vary with types and degrees.

A type of SNs has been identified, accounting for 8.7% of human nucleosomes from seven lymphocyte lines. Although this method is suitable for multiple samples, it does not consider clonal reads generated by sequencing and whether the sequencing depth of each sample is consistent. Thus, it is not suitable for our purpose to obtain accurate information on SNs (Gaffney et al., 2012). Here, we compared maps of two individuals by DANPOS. DANPOS can remove clonal reads, adjust nucleosome size to enhance the signal-to-noise ratio, and normalize samples with different sequencing depths to make them comparable. Each peak has three indices to evaluate its consistency. In this process, we used strict selection criteria. The standards are: (1) the translational position is consistent; namely, the position shift is zero (treat2control\_dis = 0), fuzziness score no difference (fuzziness\_diff\_FDR  $\geq 0.01$ , and |fuzziness\_log2FC|  $\leq 1$ ) occupancy no difference (smt\_val\_diff\_FDR  $\geq 0.01$ , and |smt\_val\_log2FC|  $\leq 1$ ) between two samples; (2) Nucleosome is well-positioned in a sample, which means the position and occupancy are consistent across multiple cell types (smt\_val  $\geq 60$ , fuzziness  $\leq 70$ ).

Using this method, we obtained four groups of SNs, namely low-MNase, medium-MNase, high-MNase1, and high-MNase2, for which the numbers of SNs were 195,241, 223,020, 674,975, and 378,049, respectively. Our results indicate that SNs exist across different individuals and multiple cell types. SNs accounted for 2.14%, 2.767%, 10.97%, and 9.47% of all nucleosomes at three MNase digestion levels. We found that SNs were the lowest at the low digestion level and the highest at the high digestion level. Only a few mononucleosomes were released and selected for sequencing at a low digestion level, and most of them still existed with large fragments; the number of these SNs present in low and high digestion samples was even smaller. Most mononucleosomes could be released from chromatin at a high digestion level, and most of these SNs exist in both samples. Therefore, we hypothesize that nucleosomes are stronger at a high digestion level in the population.

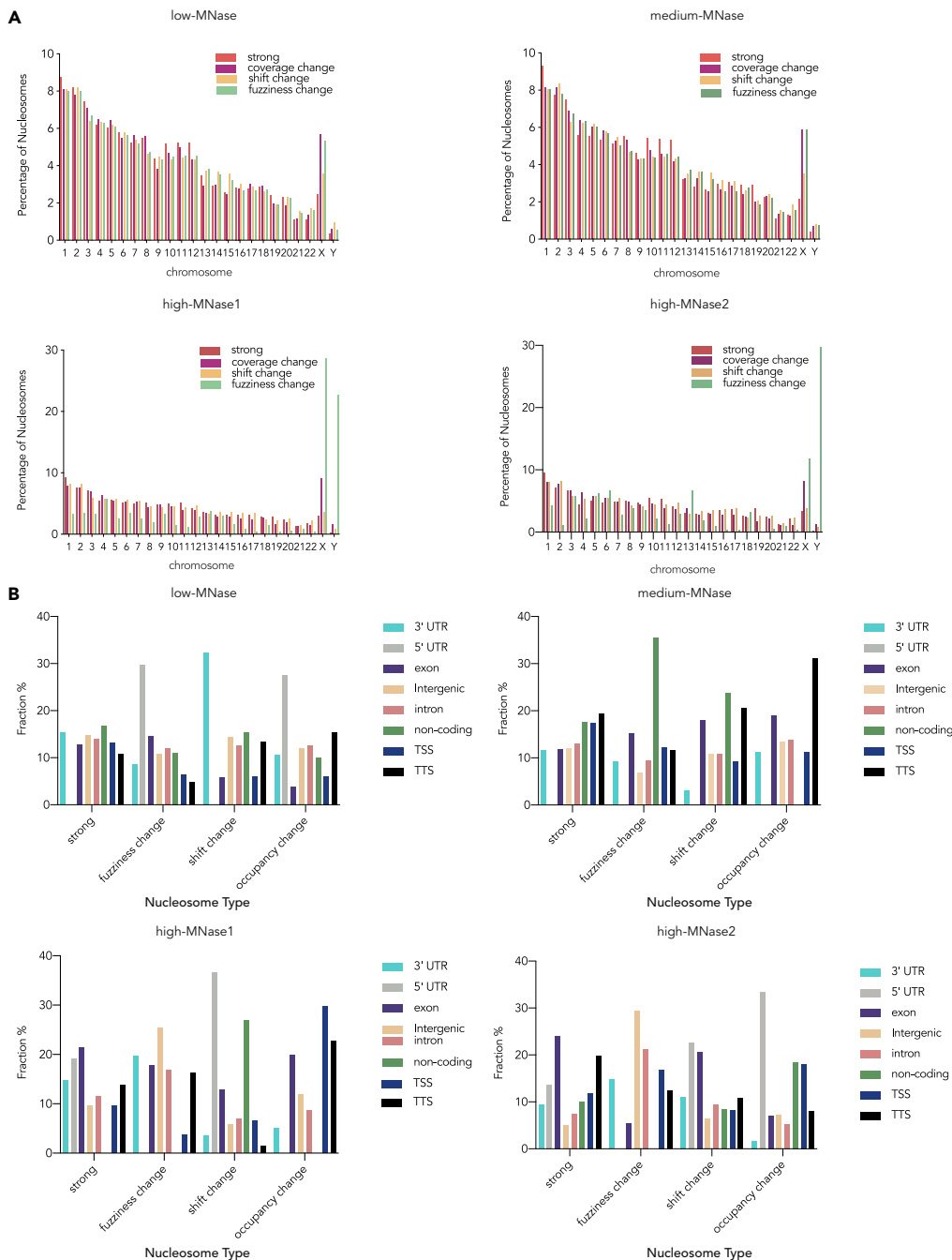
### Nucleosome stability may be related to its location in the genome

To study the distribution characteristics of SNs in the genome, we screened about 500 of the strongest nucleosomes from four groups of SNs for subsequent analysis. At the same time, about 500 nucleosomes with the largest difference in shift, fuzziness, and occupancy change were screened out (Table S2). During the screening, it was found that a large proportion of nucleosomes with occupancy change were also accompanied by fuzziness change to some degree. Therefore, to better explore the differences in occupancy and fuzziness change, we exclude those nucleosomes with both changes.

First, we focused on the distribution of four types of nucleosomes on each chromosome. Overall, SNs were evenly distributed on each chromosome in the genome, which was generally correlated with chromosome length. The longer the chromosome, the more SNs there were. The three types of dynamic nucleosomes also follow this pattern (Figure 4A). In addition, the proportions of four types of nucleosomes on each chromosome were very similar. However, the X chromosome is unique in that the number of three types of dynamic nucleosomes is significantly higher than in other chromosomes. The results showed that dynamic nucleosomes were enriched on the X chromosome. In the high-level digestion, the enrichment of the X chromosome was the most obvious, and that of the Y chromosome. The reason may be that the structure of the X and Y chromosome is more complex than that of the autosomes. We speculate that nucleosome stability may be related to the chromosome it is on and that nucleosomes are stronger on an autosome.

Second, we analyzed the distribution of SNs in different regions of the genome. The results showed that SNs had the largest number of intergenic regions and introns (Figure S4), which was consistent with the proportion of each region in the human genome. We further reanalyzed the number of nucleosomes of each





**Figure 4. Genomic distribution of four types of nucleosomes from four groups of comparison (low-MNase, medium-MNase, high-MNase1, and high-MNase2)**

(A) Distribution of four types of nucleosomes on each chromosome.

(B) The percentage of the number of nucleosomes per kilobase in each functional area of four types of nucleosomes. See also [Figures S4](#).

type per kilobase in each functional area. To allow direct comparison among each type, we have normalized the number of nucleosomes of each type per kilobase. This was conducted by calculating the percentage of the number of nucleosomes per kilobase in each functional area in the total number of the same sample ([Figure 4B](#)). The results showed no significant difference in the percentage of SNs in each region among the four groups, indicating that SNs were not enriched in a certain region but evenly distributed

in the whole genome. There are differences in the distribution of three types of dynamic nucleosomes. The common point of the dynamic nucleosomes is that they are enriched in the regulatory regions of 3' UTR, 5' UTR, and TTS, especially the shift and occupancy change nucleosomes. The differential nucleosomes are concentrated near TSS and TTS. The shift or occupancy change in nucleosomes near TSS and TTS affects the accessibility of DNA to proteins, including various chromatin regulators and transcription machinery. The accessibility of DNA to proteins regulates transcription by either promoting or inhibiting transcription initiation and termination. These results suggest that the stability of nucleosomes is related to the genomic region they are located.

Finally, the SNs in the nucleosome array region of Chr.12 at a high digestion level were analyzed. The entire length of Chr.12 is ~ 133,000 Kb, with a total of 28,930 SNs and an average of 22 nucleosomes per kb. The nucleosome array near the centromeric region of Chr.12 was 121 kb in length, and there were 56 SNs in total, with an average of 47 SNs/Kb, which was much higher than the mean for Chr.12. This implies that SNs are enriched in the array of well-positioned nucleosomes.

### Nucleosome stability may be related to the DNA sequence

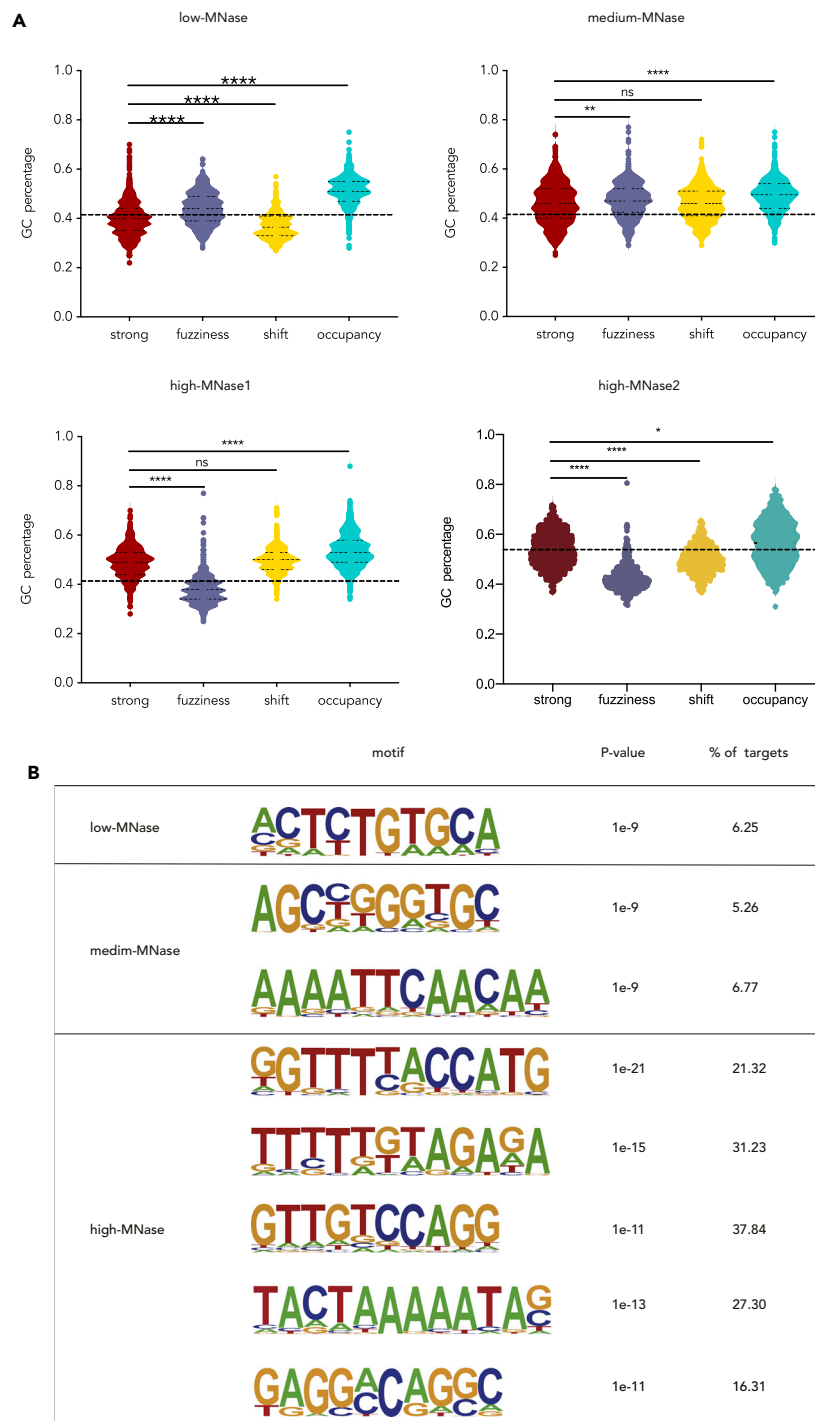
The role of DNA sequences in determining nucleosome organization has been controversial for many years (Kaplan et al., 2010b). *In vivo*, the contribution of DNA sequences is sometimes superseded by the actions of many regulatory factors, including chromatin remodelers and general regulators (Chereji and Clark, 2018; Struhl and Segal, 2013). Furthermore, the sequences of SNs are favorable to the stability of nucleosomes, while the sequences of dynamic nucleosomes are unfavorable to their stability. Therefore, only analyzing all nucleosomes without distinguishing their stability will affect the reliability of the results.

First, we thought about whether the GC content of DNA affected the stability of nucleosomes. The GC content of four types of nucleosomes was analyzed and compared using the Kruskal-Wallis test. We found that GC content has significant differences between SNs and occupancy change nucleosomes at three digestion levels, the same as SNs and fuzziness change nucleosomes. The results suggest that the consistency of nucleosomes on occupancy and fuzziness score might be related to the GC content of DNA sequences. The difference between SNs and Shift nucleosomes is significant at low digestion levels but not at medium and high digestion levels, indicating that the stability of nucleosome dyads is affected by GC content at low digestion levels, but not at the medium-high digestion level (Figure 5A). The reason may be that the shift change nucleosomes we used to compare with SNs were 500 nucleosomes with the largest shift difference ( $\geq 90$  bp). In this case, the nucleosome positioning in most cells had changed. Therefore, we speculate that the main cause of large shift change is GC content, but other factors cannot be excluded. The change in nucleosome occupancy and fuzzy score represents the change in the proportion of cells that contain a nucleosome at a given position in the cell population and the change of nucleosome position in some cells, respectively (Chereji et al., 2019; Chen et al., 2013), which are both changes of small degree compared with the nucleosome shift. These types of changes are related to the GC content of the sequence.

In addition, we found that the GC content of the fuzzy dynamic nucleosomes was significantly higher than that of SNs at low and medium digestion levels. In comparison, the GC content of the fuzzy differential nucleosome was significantly lower than that of SNs at a high digestion level. Thus, it may be that under high-MNase digestion, some nucleosomes with low GC content have begun to degrade, leading to the irregular edges of mononucleosome fragments and a differential fuzziness score.

In general, we speculate that, at low and medium digestion levels, nucleosomes with low GC content are stronger at nucleosome occupancy and positioning compared to the high digestion level condition, while at high digestion levels, the nucleosomes with low GC content are more stable at nucleosome occupancy and more unstable at the translational position.

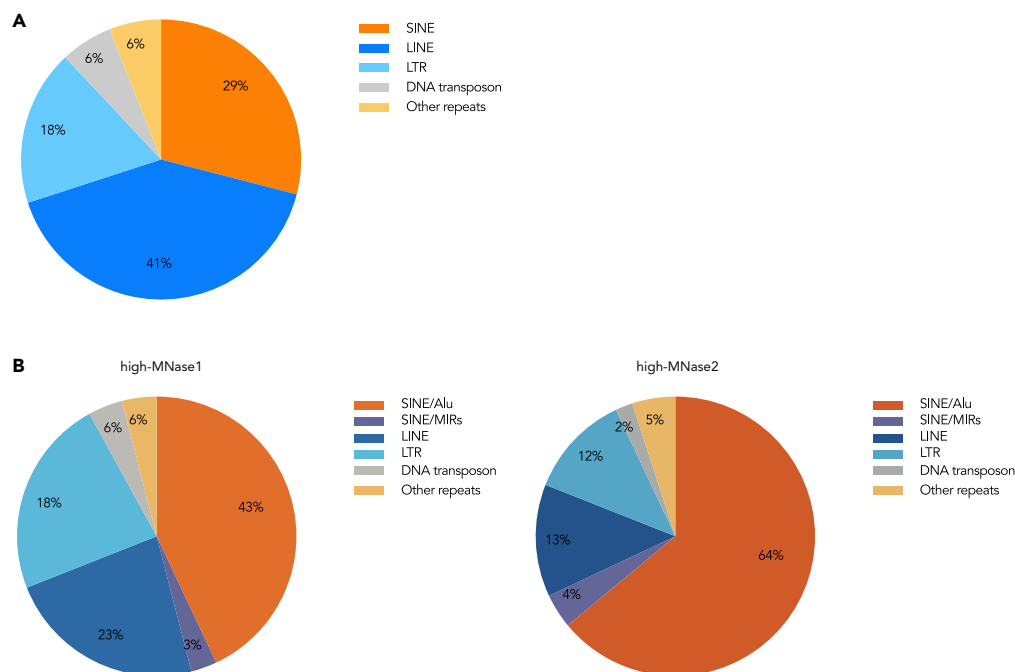
In addition to GC content, searching for nucleosome motifs has always received ample research attention. Is the DNA of SNs composed of specific sequences? If we think of SNs as a special class of nucleosomes, are there particular sequences that bind strongly on histone octamer? The Homer *de novo* algorithm was used to identify SNs motifs. We conducted a motif search in SN sequences at three levels of MNase digestion. Many potential nucleosome motifs were found, and most of them were



**Figure 5. GC content and motifs of SN sequences**

(A) GC content comparison of DNA sequences of the four types of nucleosomes. The horizontal line represents GC content in the human genome. \* indicates that the GC content of SN sequence is significantly different from that of dynamic nucleosomes with the Kruskal-Wallis test ( $P < 0.05$ ). \*\*, \*\*\*, \*\*\*\* indicate that  $P < 0.01$ ,  $P < 0.001$ ,  $P < 0.0001$ , respectively.

(B) Motifs in SN sequences identified by *de novo* motif analysis at three levels of MNase digestion. On the right side of the logo is the concentration of the motif and the proportion of its presence in the SN sequence.



**Figure 6. SNs and repeated elements**

(A) The proportion of various repeated elements in the human genome.

(B) The panel shows the proportion of repeated elements occupied by the SN sequences at high digestion levels. See also [Table S3](#) and [S4](#).

of 10 or 12-bp length. We identified the most significant motifs by combining statistical significance and the number of motifs in the nucleosome sequence ([Figure 5B](#)). GGTTTTACCATG, TTTTGTAGAGA, and GTTGTCAGG at a high digestion level had the highest enrichment, and these motifs were found in 21%, 31%, and 37.84% of SNs sequences, respectively. It is speculated that these sequences bind strongly to the histone octamers.

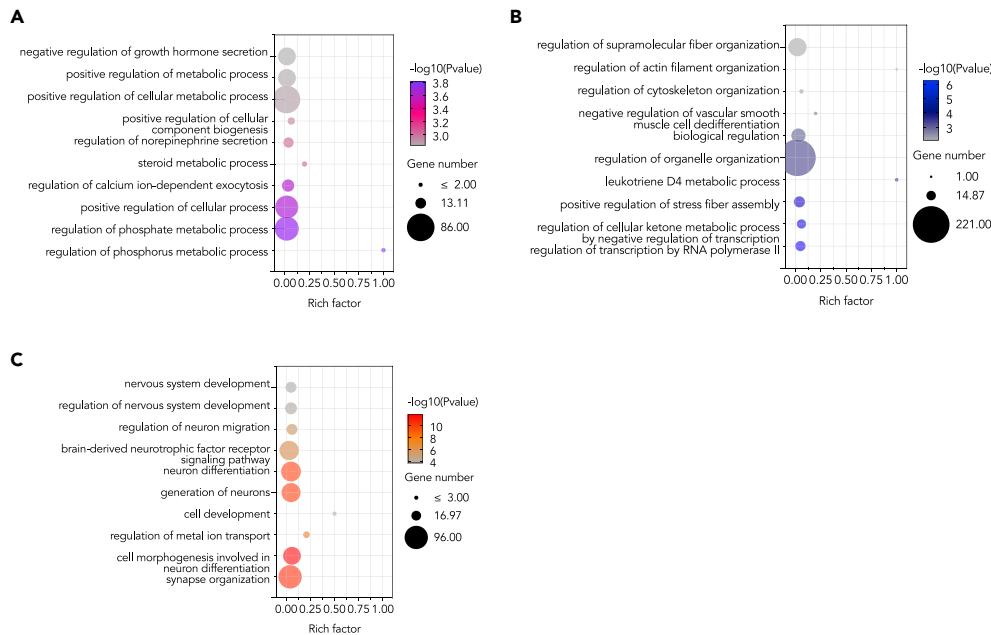
### Relationship between strong nucleosomes and repeated elements

To test whether SNs are related to repeated elements, we first detected four SNs with repeats. It was found that about 40%–50% of the SN sequence overlaps with repeats ([Treangen and Salzberg, 2012](#)), which is consistent with the proportion of repetitive elements in the human genome ([Table S3](#)).

We also focused on the proportion of repeats on the human genome and SNs. The proportion of various repetitive elements in SNs under low MNase was similar to that of the genome ([Figure 6A](#)), while SINE/Alu elements were enriched under medium MNase and high MNase (high-MNase1 and high-MNase2), accounting for 35%, 43%, and 64%, respectively. Especially in the SNs of high-mnase2, the proportion of SINE/Alu (64%) is much higher than that in the human genome (24%) ([Figure 6B](#) and [Table S4](#)). Studies have found that Alu sequences show lower deformation energy than the surrounding regions ([Cai and Luscombe, 2020](#); [Tolstorukov et al., 2008](#)), and lower deformation energy cost is more conducive to the stability of histone octamer binding to DNA. In addition, some studies have shown that nucleosomes may adopt optimal rotational positioning in human repeated elements (including SINE/Alu elements) ([Wright and Cui, 2019](#)). These factors may promote the enrichment of SNs in SINE/Alu elements.

### Strong nucleosomes function

To explore nucleosome function, GO analysis was performed on the genes of SNs, shift, and occupancy change nucleosomes. As fuzziness changes, the nucleosome is dynamic with small differences; no analysis was performed herein. However, the enrichment results for the same type of nucleosomes at different digestion levels were similar. Therefore, in the subsequent analysis, the nucleosome with different degradation levels was analyzed together. We found that shift and occupancy differential



**Figure 7. GO enrichment analysis**

- (A) Enrichment of biological processes in shift change nucleosomes.  
 (B) Enrichment of biological processes in occupancy changes nucleosomes.  
 (C) Enrichment of biological processes in SNs.

nucleosomes are associated with regulating multiple life processes. The results showed that the nucleosomes with shift differences were associated with the binding of protein kinases and the sequence-specific DNA binding of the RNA polymerase II regulatory region (Figure 7A). Changes in occupancy are associated with RNA polymerase II transcriptional regulation, regulation of cell migration, and leukotriene metabolism (Figure 7B). It is suggested that the instability of the nucleosome is owing to the influence of various life processes such as transcription and that the nucleosomes in the human genome are highly dynamic.

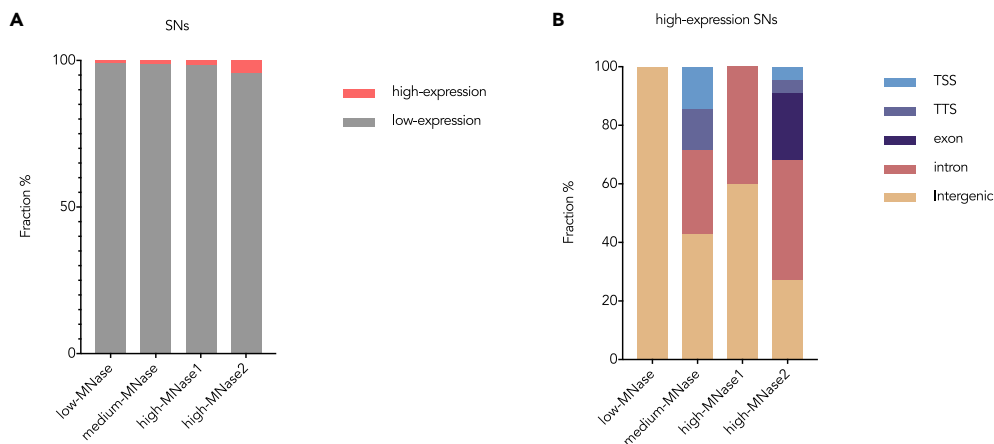
Interestingly, we found that the function of SNs is associated with the development of the nervous system (Figure 7C). This has not been found before, and the underlying mechanism remains to be explored in the future.

### Strong nucleosomes and gene expression

To explore the relationship between SNs and gene expression levels, we obtained the RPKM value of genes in which four groups of SNs were located and stratified genes based on the RPKM value. First, we analyzed the expression level of the genes in which SNs are located. We found that more than 95% of SNs were located on genes with low expression levels (Figure 8A). Second, we analyzed the distribution of SNs in different regions of the gene with high expression levels. We found that SNs are mainly distributed in an intergenic region and gene body, and there are only a few near the TSS (Figure 8B). One possible explanation is that the existence of SNs in the promoter of transcriptionally active genes is not conducive to the accessibility of DNA to the transcriptional machinery. Finally, we focused on the SNs located in the TSS of high expression levels. We found that only two SNs from four groups were located in the TSS of the genes (*SLAMF7* and *ARSA*) (Table S5). *SLAMF7* is associated with a variety of immune system tumors. The presence of SNs in the TSS may be one of the reasons for this abnormal phenomenon.

### DISCUSSION

Although there are large amounts of high-resolution data on nucleosome positioning, it remains challenging to accurately determine the nucleosome position owing to the remarkable variation between different nucleosome datasets. The reason for this may be two-fold. First, it could be owing to the



**Figure 8. SNs of four groups and gene expression levels**

(A) The percentage of SNs located in genes with high and low expression.

(B) Distribution of SNs in genes with high expression. See also [Table S5](#).

differences in MNase digestion level, sequencing method, data analysis method, and other experimental techniques. Second, nucleosome positioning is affected by multiple life processes owing to nucleosome dynamics. In this article, to overcome the limitations of MNase-seq and improve the accuracy of nucleosome position on the human genome, we improved the experimental method according to the source of noise as follows. First, the leucocyte of each donor was divided into three parts for MNase-seq with three levels of digestion (low, medium, and high). Then, the nucleosome maps at the same digestion level were compared to exclude the influence of digestion level. Second, the sequencing platform and method of each sample are the same. Third, all samples were analyzed using the same biological information software DANPOS to position nucleosomes. Finally, to correct the bias of MNase and improve the accuracy of nucleosome maps, naked DNA was prepared from each donor's leucocyte as the control. Bioinformatics software DANPOS was used to correct each sample at low-MNase and medium-MNase, after which comparisons between samples were made.

After removing the noise sources that may cause differences as much as possible, we used SN screening to remove the nucleosomes affected by various life processes and obtain accurate nucleosome positions in the population. We then explored the factors that may affect nucleosome stability. To our knowledge, our study is the first to analyze the characteristics of SNs at the genome-wide scale and explore the factors that may affect nucleosome stability. We found that nucleosomes are stronger on an autosome and SNs were mainly distributed on genes with low expression levels and non-regulatory regions. GGTTTTACCATG, TTTTGTAGAGA, and GTTGTCCAGG were highly enriched in SNs. The GC content differed from that of dynamic nucleosomes, suggesting that the sequence affected nucleosome stability. The enrichment of SNs in the Alu element and centromere region suggests that SNs may play an important role in maintaining genomic stability.

At present, there are different views on the research of naked DNA. On the one hand, studies have shown that there is a high correlation between the nucleosomal and naked MNase profiles, and the intrinsic physical properties of naked DNA determine major nucleosome positioning in yeast, especially those located in TSS and TTS, which affect the accessibility of DNA to regulatory proteins and ultimately impact gene regulation ([Deniz et al., 2011](#)). On the other hand, we have also consulted many references showing that MNase has sequence preference and preferential cutting of AT sequence leads to wrong enrichment of GC. One way to avoid MNase sequence preference is complete digestion (overdigestion), and the other is to use naked DNA correction for partially digested samples. More recently, the naked DNA signal could be removed at the whole-genome level to avoid the influence of MNase itself on nucleosome positioning ([Gutiérrez et al., 2017](#); [Flores et al., 2014](#); [Pazin et al., 2010](#)).

In this study, we used DANPOS to subtract the naked DNA signal from the chromatin data in partially digested MNase-seq experiments to avoid MNase sequence preference. The method of naked DNA correction is the same as that used by Gutierrez et al. ([Gutiérrez et al., 2017](#)), but with minor differences. First,

Gutierrez et al. studied the yeast genome, and we are the first to apply this method to the human genome. Second, Gutierrez et al. reported a stronger correlation between GC content and occupancy of naked DNA than our results. The reason may be that different species have different sequence characteristics and base compositions. Third, we found that the corrected samples still correlate with GC content and nucleosome occupancy. We think the reason might be that the GC-rich sequence is more flexible, which is more conducive to bending the DNA double helix and the winding of DNA around the histone octamers (Bettecken et al., 2011; Zuo and Li, 2011). We believe that naked DNA correction is helpful to improve the accuracy of nucleosome mapping and should be generalized in the future.

There has been no consensus on the definition of SNs. Here, we set a more rigorous definition of SNs that includes two conditions. First, the translational position of nucleosomes in the two samples was the same (shift  $\leq 5$  bp), and there was no difference in fuzziness scores and occupancy level. Second, the nucleosome is well-positioned (peak height  $\geq 60$  and fuzziness score  $\leq 70$ ). Based on the method of screening SNs reported in a previous study (Gaffney et al., 2012), we considered the consistency of the translational position of nucleosomes and nucleosome occupancy in the first condition. Because nucleosome positioning and occupancy are co-determinants of nucleosome organization (Lai and Pugh, 2017), the two cannot be completely separated. Nucleosome occupancy represents the number of cells with nucleosome dyad at a certain location in the cell population (Segal et al., 2006; Hughes and Rando, 2014); it is an important aspect of chromatin organization that also affects gene regulation and all other DNA-related processes in cells (Chereji et al., 2019). Therefore, we set the second condition to identify the nucleosomes representing most cell positions. The SNs we identified are independent of life processes, independent of individual differences, and independent of different cell types in leucocytes, representing nucleosomes localized in most cells. This finding provides a new idea for studying the organization and function of nucleosomes and provides reliable data for other studies. However, owing to the limitations of experimental technology and funding, we have only completed the study of SNs between two unrelated donors. In the future, a study involving more donor participants will be conducted to obtain additional accurate SN information on the human population.

Since the 1980s, scientists have tried many methods to find the preferred sequence of nucleosomes but have faced challenges. Crothers et al. attempted to search for sequences with strong histone affinity using recombinant nucleosomes *in vitro* and found that nucleosomes were reconstructed on repeated 20-mer TCGGTGTTAGCCTGTAAC 10-mer TGTTAGTCGTG sequences were more stable than those constructed on 5sRNA genes (Trifonov et al., 2015). With the rapid development of sequencing technology in the past decades, many nucleosome maps were obtained from experiments *in vitro*, and researchers screened SNs by searching for sequences with strong periodicity in the genome sequence (Salih et al., 2015). However, the real environment *in vivo* is much more complicated. Compared to those synthesized *in vitro*, 95% of nucleosomes in natural chromatin sequences are considered less stable (Lowary and Widom, 1997). In addition to the DNA sequence, nucleosome positioning *in vivo* is also affected by DNA replication, transcription, chromatin remodeling, and crown modification, among other processes. Therefore, the experiments *in vitro* cannot represent the real nucleosome position. For example, Perales et al. found that nucleosomes are not strongly formed on the 601 elements of the open reading frame (ORF) (GAL1-YLR454W) or intergenic region, which is considered a powerful nucleosome positioning element *in vitro* in the yeast genome (Perales et al., 2011). At the same time, the same problem can be found by directly searching for strongly periodic sequences in the genome. The position of some nucleosomes may be affected by many life activities, which may cause shift or deletion, resulting in an inaccurate map of the nucleosome and thus affecting the sequence features of the nucleosome. To overcome the above limitations, we classified nucleosomes according to their stability and then screened stable nucleosomes and excluded dynamic nucleosomes. Various biological activities may change the position of these dynamic nucleosomes. Using this method, we obtained more precise nucleosome positions. We identified a group of sequences with the highest enrichment degree in SNs, which indicates that the binding of these sequences to histones is the most stable. This finding is of great significance for revealing the sequence characteristics of nucleosome positioning.

Many studies, including this one, have found that nucleosomes are involved in gene regulation. However, the focus has largely been on chromatin functional regions, such as promoter regions, and little attention was paid to other genome regions. We mapped the nucleosome position of the whole genome and screened out SNs. It was found that SNs not only exist in the promoter and transcriptional termination

site but were enriched in the intron and intergenic regions, as well as non-gene sequences such as the Alu elements and the centromere region. In the human genome, the Alu is the transposable element (TE) with the largest copy number. TEs can affect local genome stability through insertion, deletion, recombination, transduction, and other mechanisms and even cause genome rearrangement, which affects the stability of the genome. Most TE inserts are neutral or harmful and may even lead to disease. It has been demonstrated in cancer research that genomic instability is a common feature of almost all cancers (Negrini et al., 2010). Enrichment for SNs in the Alu element may inhibit the gene's activity in which the Alu is located, thus reducing the transcription level, and may have the same effect on neighboring genes, which is conducive to maintaining the stability of the genome. It has been suggested that the specific localization of nucleosomes in the Alu element may be important in concealing the unwanted effects of Alu on adjacent genes (Tanaka et al., 2010). In addition, we found that SNs were also enriched on the nucleosome array in the centromeric region on Chr12, which was consistent with the results found by Widlund et al. in the mouse genome through FISH (Widlund et al., 1997). Furthermore, the nucleosome array was observed on the whole-genome sequencing data generated from hair shafts of a 4000-yr-old Saqqaq individual (Pedersen et al., 2014). The fragmentation of ancient DNA occurs post-mortem owing to a combination of enzymes and other environmental factors. This suggests that SNs in the nucleosome array region could resist not only the digestion of MNase, but also the destruction and degradation of various enzymes and the natural environment, and remain stable for a long time. Exploring genetic markers on the sequence of SNs could provide a new method for detecting highly degraded DNA in forensic medicine. These studies indicate that SNs may play an important role in maintaining genomic stability.

In summary, we provide a method for minimizing the noise of MNase-seq in human cells and obtaining accurate and comprehensive nucleosome maps, which may help guide the standardization of the MNase-seq assay. By comparing the maps of different individuals, we found a group of SNs, explored their characteristics, and excluded the nucleosomes with positioning and occupancy affected by chromatin remodeling, DNA transcription, and other processes. Therefore, a reliable preference sequence of nucleosomes was mined, and a new method for studying nucleosome sequence characteristics was presented. We found signatures of the location of SNs in the genome and discussed the critical role of SNs in maintaining genomic stability and the important role of dynamic nucleosomes in gene regulation, which enriches our understanding of nucleosomes. These findings are of great significance to many research fields. First, in forensic science, based on the protective effect of a nucleosome on DNA, the genetic markers of nucleosome DNA resistant to degradation can be obtained to improve the success rate of forensic analysis of highly degraded DNA (Dong et al., 2016; Freire-Aradas et al., 2012). Second, in the study of ancient human DNA, the combination of ancient nucleosome maps with strong nucleosome information can help explain ancient gene expression levels and functions (Hanghoj et al., 2016; Pedersen et al., 2014). Third, the study of SNs could be relevant for the CRISPR DNA editing field because nucleosomes may regulate the accessibility to CRISPR-Cas nucleases and influence DNA targeting in eukaryotic cells (Strohkendl et al., 2021). Finally, this study is important for understanding nuclear stability and dynamics, chromogenic stability, human genome evolution, and human diseases such as *de novo* mutations (Cai and Luscombe, 2020).

### Limitations of the study

In this study, we used DANPOS to compare two maps of different individuals, found four groups of SNs, and explored their characteristics. Owing to the limitations of the existing methods, the number of samples is not very large. In the future, it is hoped that there will be better methods that can directly compare more nucleosome maps. And more samples are expected to validate our findings on SNs. In addition, we did not conduct the experiment of female samples in the study. In the future, we will select female donors to explore the influence of sex or gender on nucleosome positioning.

### STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS



● **METHOD DETAILS**

- Differential MNase digestion and DNA isolation
- Sequencing data alignment and processing
- Nucleosome peak calling and comparison between samples
- DNA sequencing characterization analysis and functional enrichment analysis of SNs
- Expression analysis

● **QUANTIFICATION AND STATISTICAL ANALYSIS**

**SUPPLEMENTAL INFORMATION**

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2022.104593>.

**ACKNOWLEDGMENTS**

This work was supported by the National Natural Science Foundation of China Youth Science Fund Project (81701865), Hebei Province Natural Science Foundation Project (H2020206107), and Hebei Provincial Department of Education Funded Project for Cultivating Innovative Ability of Graduate Students (CXZZBS2020109). In addition, we thank Junyan Wang and Shaohan Yang for their help in imaging.

**AUTHOR CONTRIBUTIONS**

Bin Cong, Lin Wang, and Chunnan Dong conceived the study. Lin Wang and Chaolong Lu performed most of the bioinformatics analysis. Lin Wang and Chunnan Dong verified the underlying data. Chunnan Dong, Chaolong Lu, Shujin Li, and Lihong Fu were involved in the discussion of the project. All authors read and approved the final version of the article.

**DECLARATION OF INTERESTS**

The authors declare no conflict of interest relevant to the present article.

Received: September 9, 2021

Revised: March 3, 2022

Accepted: June 8, 2022

Published: July 15, 2022

**REFERENCES**

- Albert, I., Mavrich, T.N., Tomsho, L.P., Qi, J., Zanton, S.J., Schuster, S.C., and Pugh, B.F. (2007). Translational and rotational settings of H2A.Z nucleosomes across the *Saccharomyces cerevisiae* genome. *Nature* 446, 572–576. <https://doi.org/10.1038/nature05632>.
- Bai, L., and Morozov, A.V. (2010). Gene regulation by nucleosome positioning. *Trends Genet.* 26, 476–483. <https://doi.org/10.1016/j.tig.2010.08.003>.
- Bettecken, T., Frenkel, Z.M., and Trifonov, E.N. (2011). Human nucleosomes special role of CG dinucleotides and Alu nucleosomes. *BMC Genom.* 12, 273. <https://doi.org/10.1186/1471-2164-12-273>.
- Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
- Cai, L., and Luscombe, N.M. (2020). Nucleosome positioning stability is a modulator of germline mutation rate variation across the human genome. *Nat. Commun.* 11, 1363. <https://doi.org/10.1038/s41467-020-15185-0>.
- Chen, K.F., Xi, Y.X., Pan, X.W., Li, Z.Y., Kaestner, K., Tyler, J., Dent, S., He, X.W., and Li, W. (2013). DANPOS: dynamic analysis of nucleosome position and occupancy by sequencing. *Genome Res.* 23, 341–351. <https://doi.org/10.1101/gr.142067.112>.
- Chereji, R.V., and Clark, D.J. (2018). Major determinants of nucleosome positioning. *Biophys. J.* 114, 2279–2289. <https://doi.org/10.1016/j.bpj.2018.03.015>.
- Chereji, R.V., Bryson, T.D., and Henikoff, S. (2019). Quantitative MNase-seq accurately maps nucleosome occupancy levels. *Genome Biol.* 20, 198. <https://doi.org/10.1186/s13059-019-1815-z>.
- Chereji, R.V., Kan, T.-W., Grudniewska, M.K., Romashchenko, A.V., Berezikov, E., Zhimulev, I.F., Guryev, V., Morozov, A.V., and Moshkin, Y.M. (2016). Genome-wide profiling of nucleosome sensitivity and chromatin accessibility in *Drosophila melanogaster*. *Nucleic Acids Res.* 44, 1036–1051. <https://doi.org/10.1093/nar/gkv978>.
- Deniz, Ö., Flores, O., Battistini, F., Pérez, A., Soler-López, M., and Orozco, M. (2011). Physical properties of naked DNA influence nucleosome positioning and correlate with transcription start and termination sites in yeast. *BMC Genom.* 12, 489. <https://doi.org/10.1186/1471-2164-12-489>.
- Dong, C.-N., Yang, Y.-D., Li, S.-J., Yang, Y.-R., Zhang, X.-J., Fang, X.-D., Yan, J.-W., and Cong, B. (2016). Whole genome nucleosome sequencing identifies novel types of forensic markers in degraded DNA samples. *Sci. Rep.* 6, 26101. <https://doi.org/10.1038/srep26101>.
- Falcon, S., and Gentleman, R. (2006). Using GOstats to test gene lists for GO term association. *Bioinformatics* 23, 257–258. <https://doi.org/10.1093/bioinformatics/btl567>.
- Flores, O., Deniz, Ö., Soler-López, M., and Orozco, M. (2014). Fuzziness and noise in nucleosomal architecture. *Nucleic Acids Res.* 42, 4934–4946. <https://doi.org/10.1093/nar/gku165>.
- Freire-Aradas, A., Fondevila, M., Kriegel, A.K., Phillips, C., Gill, P., Prieto, L., Schneider, P.M., Carracedo, Á., and Lareu, M.V. (2012). A new SNP assay for identification of highly degraded human DNA. *Forensic Sci. Int. Genet.* 6, 341–349. <https://doi.org/10.1016/j.fsigen.2011.07.010>.
- Gaffney, D.J., Mcvicker, G., Pai, A.A., Fondufe-Mittendorf, Y.N., Lewellen, N., Michelini, K., Widom, J., Gilad, Y., and Pritchard, J.K. (2012). Controls of nucleosome positioning in the human genome. *PLoS Genet.* 8, e1003036. <https://doi.org/10.1371/journal.pgen.1003036>.
- Gutiérrez, G., Millán-Zambrano, G., Medina, D.A., Jordán-Pla, A., Pérez-Ortín, J.E., Peñate, X., and Chávez, S. (2017). Subtracting the sequence bias from partially digested MNase-seq data reveals a

- general contribution of TFIIIS to nucleosome positioning. *Epigenet. Chromatin* 10, 58. <https://doi.org/10.1186/s13072-017-0165-x>.
- Han, M., and Grunstein, M. (1988). Nucleosome loss activates yeast downstream promoters *in vivo*. *Cell* 55, 1137–1145. [https://doi.org/10.1016/0092-8674\(88\)90258-9](https://doi.org/10.1016/0092-8674(88)90258-9).
- Hanghøj, K., Seguin-Orlando, A., Schubert, M., Madsen, T., Pedersen, J.S., Willerslev, E., and Orlando, L. (2016). Fast, accurate and automatic ancient nucleosome and methylation maps with epiPALEOMIX. *Mol. Biol. Evol.* 33, 3284–3298.
- Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murte, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 38, 576–589. <https://doi.org/10.1016/j.molcel.2010.05.004>.
- Huebert, D.J., Kuan, P.F., Keles, S., and Gasch, A.P. (2012). Dynamic changes in nucleosome occupancy are not predictive of gene expression dynamics but are linked to transcription and chromatin regulators. *Mol. Cell Biol.* 32, 1645–1653. <https://doi.org/10.1128/MCB.06170-11>.
- Hughes, A.L., and Rando, O.J. (2014). Mechanisms underlying nucleosome positioning *in vivo*. *Annu. Rev. Biophys.* 43, 41–63. <https://doi.org/10.1146/annurev-biophys-051013-023114>.
- Strohkendl, I., Saifuddin, F.A., Gibson, B.A., Rosen, M.K., Russell, R., and Finkelstein, I.J. (2021). Inhibition of CRISPR-Cas12a DNA targeting by nucleosomes and chromatin. *Sci. Adv.* 7, eabd6030. <https://doi.org/10.1126/sciadv.abd6030>.
- Jiang, C., and Pugh, B.F. (2009). Nucleosome positioning and gene regulation: advances through genomics. *Nat. Rev. Genet.* 10, 161–172. <https://doi.org/10.1038/nrg2522>.
- Jurka, J., Kapitonov, V.V., Pavlicek, A., Klonowski, P., Kohany, O., and Walichiewicz, J. (2005). Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res.* 110(1-4), 462–467. <https://doi.org/10.1159/000084979>.
- Kaplan, N., Hughes, T.R., Lieb, J.D., Widom, J., and Segal, E. (2010a). Contribution of histone sequence preferences to nucleosome organization: proposed definitions and methodology. *Genome Biol.* 11, 140. <https://doi.org/10.1186/gb-2010-11-11-140>.
- Kaplan, N., Moore, I., Fondufe-Mittendorf, Y., Gossett, A.J., Tillo, D., Field, Y., Hughes, T.R., Lieb, J.D., Widom, J., and Segal, E. (2010b). Nucleosome sequence preferences influence *in vivo* nucleosome organization. *Nat. Struct. Mol. Biol.* 17, 918–920. <https://doi.org/10.1038/nsmb0810-918>.
- Khan, A., Fornes, O., Stigliani, A., Gheorghie, M., Castro-Mondragon, J.A., Van Der Lee, R., Bessy, A., Chèneby, J., Kulkarni, S.R., Tan, G., et al. (2018). JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* 46, D260–D266. <https://doi.org/10.1093/nar/gkx1126>.
- Kornberg, R.D. (1974). Chromatin structure: A repeating unit of histones and DNA. *SCIENCE* 184, 868–871. <https://doi.org/10.1126/science.184.4139.868>.
- Lai, W.K.M., and Pugh, B.F. (2017). Understanding nucleosome dynamics and their links to gene expression and DNA replication. *Nat. Rev. Mol. Cell Biol.* 18, 548–562. <https://doi.org/10.1038/nrm.2017.47>.
- Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* 26, 589–595. <https://doi.org/10.1093/bioinformatics/btp698>.
- Lowary, P.T., and Widom, J. (1997). Nucleosome packaging and nucleosome positioning of genomic DNA. *Proc. Natl. Acad. Sci. USA.* 94, 1183–1188. <https://doi.org/10.1073/pnas.94.4.1183>.
- Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389, 251–260. <https://doi.org/10.1038/38444>.
- Mieczkowski, J., Cook, A., Bowman, S.K., Mueller, B., Alver, B.H., Kundu, S., Deaton, A.M., Urban, J.A., Larschan, E., Park, P.J., et al. (2016). MNase titration reveals differences between nucleosome occupancy and chromatin accessibility. *Nat. Commun.* 7, 11485. <https://doi.org/10.1038/ncomms11485>.
- Negrini, S., Gorgoulis, V.G., and Halazonetis, T.D. (2010). Genomic instability—an evolving hallmark of cancer. *Nat. Rev. Mol. Cell Biol.* 11, 220–228. <https://doi.org/10.1038/nrm2858>.
- Ozgyin, L., Horvath, A., Hevessy, Z., and Balint, B.L. (2019). Extensive epigenetic and transcriptomic variability between genetically identical human B-lymphoblastoid cells with implications in pharmacogenomics research. *Sci. Rep.* 9, 4889. <https://doi.org/10.1038/s41598-019-40897-9>.
- Ozonov, E.A., and Van Nimwegen, E. (2013). Nucleosome free regions in yeast promoters result from competitive binding of transcription factors that interact with chromatin modifiers. *PLoS Comput. Biol.* 9, e1003181. <https://doi.org/10.1371/journal.pcbi.1003181>.
- Pazin, M.J., Chung, H.-R., Dunkel, I., Heise, F., Linke, C., Krobitch, S., Ehrenhofer-Murray, A.E., Sperling, S.R., and Vingron, M. (2010). The effect of micrococcal nuclease digestion on nucleosome positioning data. *PLoS One* 5, e15754. <https://doi.org/10.1371/journal.pone.0015754>.
- 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. (2014). Genome-wide nucleosome map and cytosine methylation levels of an ancient human genome. *Genome Res.* 24, 454–466. <https://doi.org/10.1101/gr.163592.113>.
- Perales, R., Zhang, L., and Bentley, D. (2011). Histone occupancy *in vivo* at the 601 nucleosome binding element is determined by transcriptional history. *Mol. Cell Biol.* 31, 3485–3496. <https://doi.org/10.1128/MCB.05599-11>.
- Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26. <https://doi.org/10.1038/nbt.1754>.
- Salih, B., Tripathi, V., and Trifonov, E.N. (2015). Visible periodicity of strong nucleosome DNA sequences. *J. Biomol. Struct. Dyn.* 33, 1–9. <https://doi.org/10.1080/07391102.2013.855143>.
- Segal, E., Fondufe-Mittendorf, Y., Chen, L., Thåström, A., Field, Y., Moore, I.K., Wang, J.P.Z., and Widom, J. (2006). A genomic code for nucleosome positioning. *Nature* 442, 772–778. <https://doi.org/10.1038/nature04979>.
- Struhl, K., and Segal, E. (2013). Determinants of nucleosome positioning. *Nat. Struct. Mol. Biol.* 20, 267–273. <https://doi.org/10.1038/nsmb.2506>.
- Tanaka, Y., Yamashita, R., Suzuki, Y., and Nakai, K. (2010). Effects of Alu elements on global nucleosome positioning in the human genome. *BMC Genom.* 11, 309. <https://doi.org/10.1186/1471-2164-11-309>.
- Tolstorukov, M.Y., Choudhary, V., Olson, W.K., Zhurkin, V.B., and Park, P.J. (2008). NuScore: a web-interface for nucleosome positioning predictions. *Bioinformatics* 24, 1456–1458. <https://doi.org/10.1093/bioinformatics/btn212>.
- Treangen, T.J., and Salzberg, S.L. (2012). Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat. Rev. Genet.* 13, 36–46. <https://doi.org/10.1038/nrg3117>.
- Trifonov, E.N., Nibhani, R., and Chaires, J. (2015). Review fifteen years of search for strong nucleosomes. *Biopolymers* 103, 432–437. <https://doi.org/10.1002/bip.22604>.
- Widlund, H.R., Cao, H., Simonsson, S., Magnusson, E., Simonsson, T., Nielsen, P.E., Kahn, J.D., Crothers, D.M., and Kubista, M. (1997). Identification and characterization of genomic nucleosome-positioning sequences. *J Mol Biol* 267(4), 807–817. <https://doi.org/10.1006/jmbi.1997.0916>.
- Wright, G.M., and Cui, F. (2019). The nucleosome position-encoding WW/SS sequence pattern is depleted in mammalian genes relative to other eukaryotes. *Nucleic Acids Res.* 47, 7942–7954. <https://doi.org/10.1093/nar/gkz544>.
- Zuo, Y.-C., and Li, Q.-Z. (2011). Identification of TATA and TATA-less promoters in plant genomes by integrating diversity measure, GC-Skew and DNA geometric flexibility. *Genomics* 97, 112–120. <https://doi.org/10.1016/j.ygeno.2010.11.002>.

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Biological samples</b>		
Human leukocytes of peripheral blood	Isolated from four healthy donors	N/A
<b>Chemicals, peptides, and recombinant proteins</b>		
MNase	thermoscientific	EN0181
EDTA	Solarbio	Cat#E1170
<b>Critical commercial assays</b>		
E.Z.N.A.TM Blood DNA MIDI Kit	OMEGA	D3494-03
Wizard SV Gel and PCR Clean-Up System	Promega	Cat#9281/2/5
DNA HiSens Reagent Kit	PerkinElmer	CLS760672
<b>Deposited data</b>		
High throughput sequencing data	This paper	GSA-Human: HRA002461
human reference genome (hg38)	Genome Reference Consortium	<a href="http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/">http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/</a>
<b>Software and algorithms</b>		
Trimmomatic version 0.36	<a href="#">Bolger et al., 2014</a>	<a href="http://www.usadellab.org/cms/index.php?page=trimmomatic">http://www.usadellab.org/cms/index.php?page=trimmomatic</a>
BWA-MEM version 0.7.12	<a href="#">Li and Durbin, 2010</a>	<a href="http://bio-bwa.sourceforge.net/bwa.shtml">http://bio-bwa.sourceforge.net/bwa.shtml</a>
DANPOS2 toolkit	<a href="#">Chen et al., 2013</a>	<a href="http://code.google.com/p/danpos/">http://code.google.com/p/danpos/</a>
IGV browser	<a href="#">Robinson et al., 2011</a>	<a href="https://igv.org/app">https://igv.org/app</a>
homer2 version 4.11	<a href="#">Heinz et al., 2010</a>	<a href="https://github.com/homer2">https://github.com/homer2</a>
JASPAR	<a href="#">Khan et al., 2018</a>	<a href="https://jaspar.genereg.net">https://jaspar.genereg.net</a>
RepeatMasker V4.1.1	<a href="#">Jurka et al. 2005</a>	<a href="http://www.repeatmasker.org">http://www.repeatmasker.org</a>

### RESOURCE AVAILABILITY

#### Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Bin Cong ([hbydbincong@126.com](mailto:hbydbincong@126.com)).

#### Materials availability

This study did not generate new unique reagents.

#### Data and code availability

The raw sequence data of MNase-seq reported in this paper have been deposited in the Genome Sequence Archive in National Genomics Data Center, China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences, under accession number GSA-Human: HRA002461 that are accessible at <https://ngdc.cncb.ac.cn/gsa-human>. The raw data are publicly available as of the date of publication. Accession number is listed in the [key resources table](#).

This paper does not report original code.

Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

### EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human peripheral blood leukocytes of four healthy male donors were collected from the Hebei Blood Center. The ages of the four donors are 25 (individual1), 24 (individual2), 24 (individual3), and

24 (individual4). In order to cover the data of all autosomal chromosomes and sex chromosomes (X and Y) and ensure comparability between sex chromosomes, we selected all male volunteers. This study was conducted with the approval of the Ethics Committee of School and the College of Forensic Medicine, Hebei Medical University. We confirmed that informed consent was obtained from all subjects.

## METHOD DETAILS

### Differential MNase digestion and DNA isolation

Blood samples from the four donors were collected from the Hebei Blood Center. Blood samples were processed immediately after donation to avoid quality changes resulting from sample storage. PBS was added to the blood for dilution, and the leukocyte precipitation was obtained by centrifugation. The erythrocyte lysate was added to lyse erythrocytes, and leukocyte precipitates were obtained by centrifugation at 4,500 rpm.

Leukocytes were counted under a microscope, A total of eight tubes (each with  $5 \times 10^6$  cells) were prepared from four individuals. The cells in each tube were centrifuged and resuspended in a 0.5% Triton solution, lysed on ice for 15 min, and washed with PBS three times. The MNase concentrations were 15, 30, and 120 U, and CaCl<sub>2</sub> (100 mM, 250  $\mu$ L) was added to each tube of leukocytes to prepare samples with different digestion levels of MNase. The cells were incubated at 37°C for 3.5 h, and finally, 150 microliter 500 mM EDTA was added to terminate the reaction. DNA was extracted and purified by E.Z.N.A.<sup>TM</sup> Blood DNA MIDI Kit (OMEGA). The digested DNA was analyzed by Labchip® GX Touch24 to obtain the fragment sizes of each sample and calculate the digestion level. The mononucleosome fragment (about 150 bp) of agarose gels was cut, and the DNA was excised and purified using Wizard SV Gel and PCR Clean-Up System (Promega).

Two naked DNA samples were prepared from two blood samples of individual1 and individual2, and the materials and methods were the same as those used for the previous eight samples. The difference was that naked DNA was extracted from leukocytes first, and then 0.45 U MNase was added to digest DNA.

### Sequencing data alignment and processing

MNase sequencing was performed for eight samples (six mononucleosome samples and two naked DNA samples). For each sample, a paired-end (2 × 150 bp) sequencing library was generated and then sequenced on Illumina Novaseq 6,000 platform according to the manufacturer's protocol. Raw reads performed quality control with Trimmomatic (version 0.36) (Bolger et al., 2014) and filtered reads were aligned to the human reference genome (hg38) using BWA-MEM (version 0.7.12) with default parameters (Li and Durbin, 2010).

### Nucleosome peak calling and comparison between samples

Nucleosome peaks called DANPOS2 toolkit (Chen et al., 2013) in dpos mode with default parameters for eight mononucleosome samples and two naked DNA samples. We compared nucleosome maps between two samples using the triple algorithm (L1 and L2, M1 and M2, H1 and H2, H3 and H4), where each sample was corrected by naked DNA sample (N1 or N2). To obtain accurate results of SNs and dynamic nucleosomes between samples, we set strict filter conditions. To visualize specific genome sites, IGVTools were used to create trails (TDF files) that can be viewed in the IGV browser (James T Robinson et al., 2011).

Nucleosome fuzziness refers to the degree that read positions in each nucleosome peak deviate from the most preferred nucleosome position, and the standard deviation of read positions in each peak was used as a fuzziness score. Nucleosome fuzziness change was determined by an F-test on the p-value of difference in fuzziness score of two nucleosome peaks. Nucleosome occupancy change was determined by a Poisson test on the difference between the maximum values of the two nucleosome peaks. Nucleosome position shift was estimated by calculating the relative distance between two nucleosome peaks.

### DNA sequencing characterization analysis and functional enrichment analysis of SNs

We obtained four groups of SNs at three digestion levels. To facilitate the analysis further, we selected the strongest nucleosomes, about 500, according to two indicators: the fuzzy score difference (| fuzziness\_log2FC |) and occupancy difference (| smt\_val\_log2FC |). The motif *de novo* prediction and annotation of nucleosome peaks was performed with homer2 (version 4.11) and JASPAR (Heinz et al., 2010; Khan

et al., 2018). The repeat elements were identified using RepeatMasker (V4.1.1) (Jurka et al., 2005). The main types of repeated elements were selected for analysis, including short interspersed nuclear elements (SINE) (Alu, MIR), long interspersed nuclear elements (LINE) (L1, L2, and L3/CR1), long terminal repeats (LTR) (ERV1, ERV2, ERVL, and ERVL-MALRS), DNA transposon (TCMar-Mariner, Hat-Charlie). The rest of the repeated elements fall into the "other" category, which includes simple repeats, satellites, small RNAs, and so on. Functional enrichment for Gene Ontology was performed with GOstats (R package) (Falcon and Gentleman, 2006).

### Expression analysis

Since there is no RNA-seq data for human peripheral blood leukocytes, we obtained RNA-seq data of B-lymphoblastoid cell lines (LCLs) from the database (Gene Expression Omnibus: GSE121926). We then selected and downloaded two RNA-seq data replicates from one healthy male individual, and then processed, mapped the raw reads to hg38, and calculated the transcriptional abundance (RPKM), using the same method as Anton valouev et al. (Ozgyin et al., 2019).

### QUANTIFICATION AND STATISTICAL ANALYSIS

For the correlation of the log<sub>10</sub> signal intensity and the GC content of fragments, Pearson's correlation was used ( $p < 0.001$ ) (Figures 2 and S2C). The GC content of four types of nucleosomes was analyzed and compared using the Kruskal-Wallis test (Figure 5A). Statistical significance was set at  $p < 0.05$ . Statistical analyses were conducted using GraphPad Prism 9 (GraphPad Software).