## GENETICS

# nucMACC: An MNase-seq pipeline to identify structurally altered nucleosomes in the genome

Sara Wernig-Zorc[1]†, Fabian Kugler[1], Leo Schmutterer[2], Patrick Räß[2], Clemens Hausmann[2], Simon Holzinger[1], Gernot Längst[1]*, Uwe Schwartz[2]*

Micrococcal nuclease sequencing is the state-of-the-art method for determining chromatin structure and nucleosome positioning. Data analysis is complex due to the AT-dependent sequence bias of the endonuclease and the requirement for high sequencing depth. Here, we present the nucleosome-based MNase accessibility (nucMACC) pipeline unveiling the regulatory chromatin landscape by measuring nucleosome accessibility and stability. The nucMACC pipeline represents a systematic and genome-wide approach for detecting unstable ("fragile") nucleosomes. We have characterized the regulatory nucleosome landscape in *Drosophila melanogaster*, *Saccharomyces cerevisiae*, and mammals. Two functionally distinct sets of promoters were characterized, one associated with an unstable nucleosome and the other being nucleosome depleted. We show that unstable nucleosomes present intermediate states of nucleosome remodeling, preparing inducible genes for transcriptional activation in response to stimuli or stress. The presence of unstable nucleosomes correlates with RNA polymerase II proximal pausing. The nucMACC pipeline offers unparalleled precision and depth in nucleosome research and is a valuable tool for future nucleosome studies.

## INTRODUCTION

Nucleosomes are the basic unit of DNA packaging in eukaryotic cells. First electron microscopy images of chromatin spreads displayed regularly spaced nucleosome core particles, separated by a short linker DNA, appearing like "beads-on-a-string" (*1*). The nucleosome core particle consists of 147 bp of DNA wrapped around a histone octamer, composed of the histone proteins H2A, H2B, H3, and H4 (*2*, *3*). The stability of the nucleosome core is maintained by electrostatic interactions and more than 360 hydrogen bonds between the DNA and histone proteins (*4*). Linker DNA varies in length from 15 to 95 bp depending on the species, cell type, and even the chromatin state within the same nucleus (*5*–*7*). DNA in the linker is generally accessible to DNA binding factors, whereas access to DNA wrapped around the histone octamer is hindered. Nucleosome positioning, structure, and modification play a critical role in regulating gene expression, determining the accessibility of DNA to DNA binding factors, thereby defining cell type-specific gene activity and repression (*5*, *8*).

In vitro experiments have shown that nucleosome positioning depends on the bendability and structure of the underlying DNA sequence and the effect of steric hindrance between neighboring nucleosomes (*9*, *10*). On the other hand, in vivo nucleosome positioning is dynamically modified and influenced by additional factors, such as chromatin remodelers, histone variants, transcription factors (TFs), histone posttranslational modifications, or the transcriptional machinery (*11*). Regulatory regions, such as promoters, enhancers, and insulators, are characterized by well-positioned nucleosomes, which is a common feature of eukaryotic chromatin (*12*, *13*).

In addition to the positioning of nucleosomes, the structural property of the nucleosomes is central to DNA accessibility and chromatin function. Nucleosomes can vary in their histone composition and number. Histone variants modify the structural and functional properties of nucleosomes, playing specific roles in chromatin regulation, and being incorporated into nucleosomes independent of the cell cycle (*14*). For example, the histone H3 variant CENP-A is incorporated into nucleosomes at the centromeres. DNA-histone interactions are weakened in CENP-A–containing nucleosomes, partially unwinding DNA from the histone surface (*15*). In addition, the human testis–specific histone H3 variants, H3T and H3.5, form nucleosomes with reduced stability, as compared to nucleosomes with the canonical H3, as does the C-terminal half of H2A.Z1 (*16*–*18*). Hexasomes lacking an H2A-H2B dimer can be formed by a variety of mechanisms including transcription, histone assembly, and chromatin remodeling (*19*–*23*). The preferential presence of hexasomes near transcription start sites (TSSs) suggests a potential role in gene regulation (*24*). However, it is still unclear whether these altered/sub-nucleosomal structures are stable or short-lived nucleosome intermediates in vivo. Furthermore, there are several studies revealing structurally altered nucleosomes, like lexosome, alterosome, prenucleosome, tetrasome, and reversosome, that could play specific roles in maintaining an active genome (*25*–*27*).

Micrococcal nuclease sequencing (MNase-seq) has been widely used to study chromatin structure and nucleosome positioning. It involves the use of micrococcal nuclease (MNase), an enzyme that selectively cleaves linker DNA, leaving the nucleosome core particles intact. The resulting DNA fragments are sequenced and analyzed to determine the location and distribution of nucleosomes along the genome (*28*). MNase-seq provides a high-resolution map of the nucleosomal landscape and has contributed to our understanding of the dynamic and regulated nature of chromatin structure (*5*, *8*, *29*). Initially, high-MNase digestion conditions were used to completely hydrolyze chromatin to the mono-nucleosomal level. Later, we and others combined different MNase digestion conditions to study chromatin accessibility, nucleosome stability, and chromatin domain organization (*30*–*36*). Because of the complexity of data analysis in MNase titration experiments, metrics have been

[1]Regensburg Center for Biochemistry (RCB), University of Regensburg, Regensburg, Germany. [2]NGS Analysis Center Biology and Pre-clinical Medicine, University of Regensburg, Regensburg, Germany.
*Corresponding author. Email: uwe.schwartz@ur.de (U.S.); gernot.laengst@ur.de (G.L.)
†Present address: St. Anna Children's Cancer Research Institute (CCRI), Vienna, Austria.

introduced to provide a stable framework for the accurate quantification of chromatin accessibility and nucleosome occupancy (*31, 32*). However, up to six different MNase concentrations, from very low to high, were used per sample and additional spike-ins were suggested to be necessary, resulting in a high workload, high costs, and requiring large amounts of sample material. Furthermore, the lack of an established analytical pipeline makes it difficult to adapt these kinds of experiments to customized sample formats.

Here, we present a comprehensive and user-friendly nucleosome-based MNase accessibility (nucMACC) analysis pipeline. The nucMACC workflow starts directly with raw sequencing data, assesses the critical quality metrics relevant to MNase-seq experiments, and summarizes the results into user-friendly reports (fig. S1). The workflow reveals the genome-wide nucleosome positions and their corresponding nucMACC scores. The pipeline identifies hyper- and hypo-accessible nucleosome positions with exceptionally high or low nucMACC scores. As an additional feature of the nucMACC pipeline, nucleosome stability scores are inferred from sub-nucleosomal fragments identifying unstable or noncanonical nucleosomes. We show that the pipeline achieves robust results using only two MNase conditions per sample. In summary, the nucMACC pipeline is written in nextflow, available on Zenodo (DOI: 10.5281/zenodo.10777489, or GitHub repository uschwartz/nucMACC), and integrates Docker software container, making it easy to use, portable, reproducible, and scalable.

## RESULTS

### The nucMACC pipeline scores nucleosome accessibility and stability

MNase, which preferentially hydrolyzes accessible linker DNA between nucleosomes, releases nucleosome core particles from the chromatin. Depending on the MNase digestion conditions, nucleosomes are released at different rates (Fig. 1A). At mild MNase conditions, nucleosomes from accessible and active genomic sites are preferentially released, referred to as hyper-accessible nucleosomes (Fig. 1A, green nucleosome) (*30, 31*). Increasing the MNase concentration or digestion time increasingly releases nucleosome core particles from chromatin, until the nucleosomal chain is fully hydrolyzed. Furthermore, MNase is a processive enzyme. Once the nucleosome is released from chromatin, MNase progressively trims the linker DNA and then continues—at a lower rate—to hydrolyze the histone-bound DNA, resulting in sub-nucleosomal fragments. Canonical nucleosome core particles are highly resistant to intranucleosomal MNase cleavage, and the nucleosome core DNA remains mainly complete, even at high-MNase digestion conditions. However, noncanonical nucleosomes, often referred to as unstable or fragile nucleosomes, exhibit an increased MNase sensitivity and are degraded to sub-nucleosome-sized fragments already at low MNase concentrations (Fig. 1A, blue nucleosome) (*37–39*). We refer to these kinds of nucleosomes as unstable nucleosomes in the remainder of this manuscript. At high-MNase conditions, unstable nucleosomes are fully degraded and are therefore lost in this kind of MNase dataset. Hence, sub-nucleosomal fragments can be used to measure nucleosome stability in MNase titration experiments, as reflected by resistance to internal MNase cleavage. In conclusion, MNase titration experiments provide information about nucleosome positioning, accessibility, and stability. To exploit the full range of features for chromatin characterization, we have

developed the nucMACC pipeline (see details in Supplementary Text and fig. S1).

We used the MNase chromatin digestions coupled with H4 chromatin immunoprecipitation (ChIP) from the original MNase accessibility (MACC) publication to establish and validate the nucMACC pipeline (data S1) (*31*). The dataset collection consists of four different MNase concentration conditions (1.5, 6.25, 25, and 100 U), performed in two replicates each. As replicates are pooled later in the nucMACC workflow to achieve high nucleosome coverage, we first carefully checked the consistency of the replicates (fig. S2). The fragment size distribution is an important quality control parameter, as it clearly shows the degree of chromatin hydrolysis in each sample (Fig. 1B and fig. S2A). With increasing MNase concentrations, the di-nucleosome fraction is disappearing, and the levels of mono-nucleosomes are increasing. The dominant mono-nucleosomal peak is in the range of 140 to 200 bp, while the peak summit shifts with higher MNase concentrations closer to 147 bp, the protected DNA length of a nucleosome core particle. Fragment sizes shorter than 140 bp indicate sub-nucleosomal DNA fragments, which mainly arise from DNA cleavage within the nucleosomal core (fig. S3A). Sub-nucleosomal DNA fragments are barely present at mild MNase conditions but increase with higher MNase concentrations. On the basis of the shape of the fragment size distribution curve, we selected the fragments in the range of 140 to 200 bp in length for mono-nucleosome analysis and fragments shorter than 140 bp for sub-nucleosome analysis, respectively (Fig. 1B).

Plotting the independently normalized fragment frequencies of mono- and sub-nucleosomes at the TSS shows that the selected categories behave qualitatively differently, suggesting different structural information or functional behavior (Fig. 1, C and E). Analyzing the mono-nucleosomes of the two high-MNase conditions (25 U and 100 U) reveals the well-known nucleosome positioning pattern at the TSS: a nucleosome-depleted region (NDR) followed downstream by a phased nucleosome array, which starts with a well-positioned +1-nucleosome and spreads into the gene. The low-MNase conditions (1.5 U and 6.25 U) reveal the same mono-nucleosomal positions; however, specific positions, such as the +1-nucleosome, are markedly enriched compared to the high-MNase conditions. This effect clearly indicates that nucleosomes, depending on their genomic location, exhibit different release rates from chromatin. The nucMACC score provides a quantitative assessment of chromatin accessibility, revealing the change in mono-nucleosome fragment frequencies with increasing MNase concentration (Supplementary Text, Fig. 1D, and fig. S1).

Comparing the mono-nucleosome and the sub-nucleosome TSS profile at high-MNase conditions reveals a very similar pattern, suggesting that once the nucleosomes were released from chromatin, MNase trims the nucleosomal DNA, as observed by the change of fragment length distribution of the +1 nucleosome at different MNase concentrations (fig. S3A). However, in the sub-nucleosomal fraction, an additional dominant peak at the denominated NDR appears at low MNase conditions. As the analyzed dataset is based on an H4-ChIP, these observations suggest that the NDR is not free of histones but occupied by MNase-sensitive nucleosomes (Fig. 1E). Note that the population of sub-nucleosomal fragments under low-MNase conditions is rather small (Fig. 1B, fig. S2B, and data S1) and would not stand out when analyzed together with the mono-nucleosomes (fig. S3A). However, by looking at their relative abundance
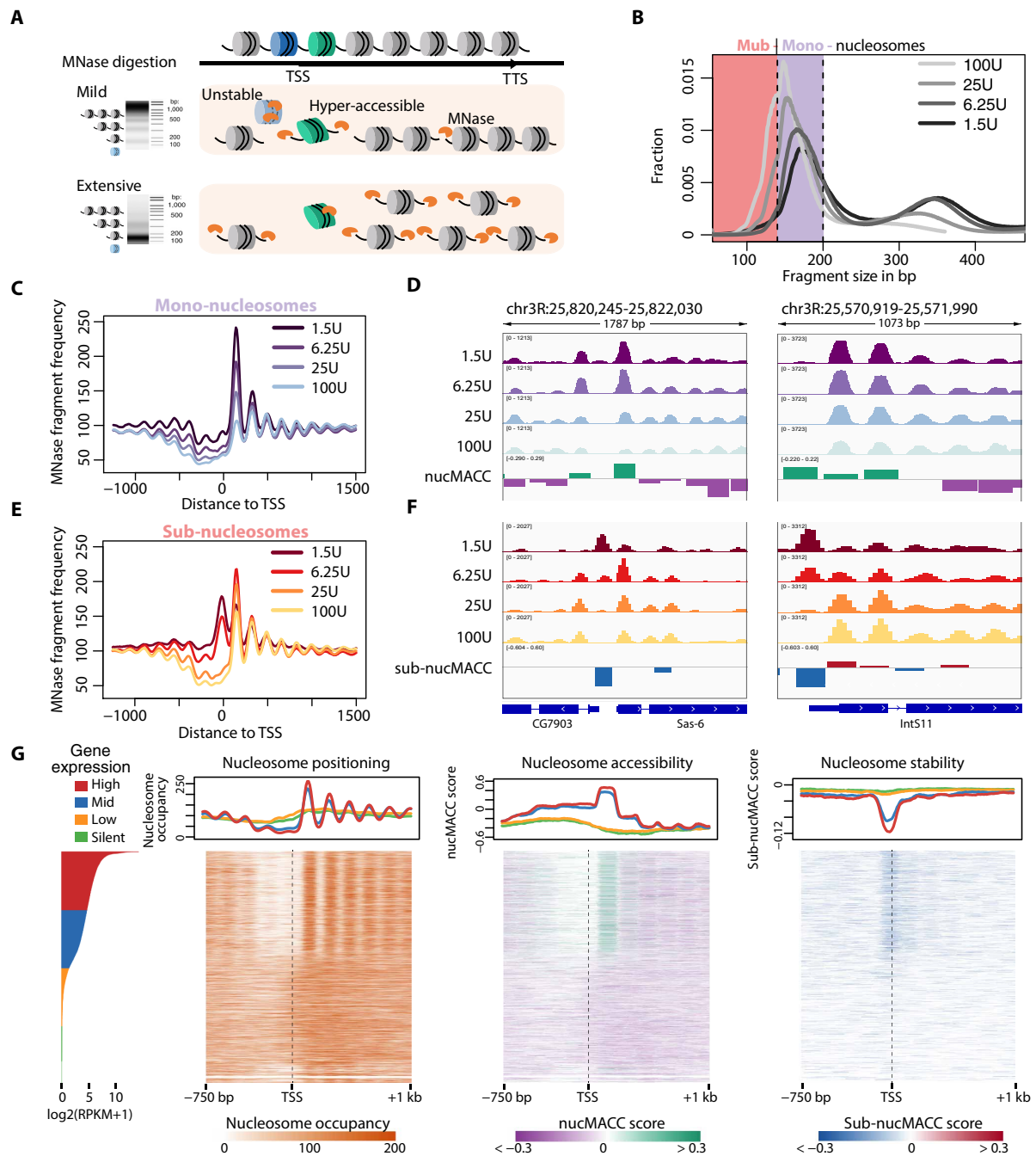
**Fig. 1. The nucMACC pipeline scores nucleosome accessibility and stability.** (**A**) Schematic representation of differential MNase chromatin digestion. Under mild MNase conditions, chromatin is partially digested and nucleosomes (green, hyper-accessible) from accessible regions are preferentially released (top). At extensive MNase conditions, chromatin is completely digested to the nucleosome core particles and MNase-sensitive nucleosomes (blue, unstable) are lost (bottom). (**B**) Fragment size distribution of MNase titrations. Fragments are selected based on the fragment size and grouped into sub-nucleosomes (<140 bp, light red section) or mono-nucleosomes (140 to 200 bp, purple section). (**C** and **E**) Average fragment frequencies at the TSS. (**D** and **F**) Genome browser snapshot showing the output of the nucMACC pipeline. nucMACC scores are deduced from mono-nucleosomal fragments (D) and sub-nucMACC scores from sub-nucleosomal fragments, respectively (F). (**G**) Heatmaps sorted by gene expression showing nucleosome positioning (left), accessibility (middle, nucMACC scores), and stability (right, sub-nucMACC scores) at TSS.

in the sub-nucleosomal fraction, it is clear that they accumulate at distinct sites that do not correlate with the abundance of mono-nucleosomes (Fig. 1, C to F). The sub-nucMACC score allows the quantification of the nucleosome stabilities at these sites, as it monitors the changes in sub-nucleosome fragment frequencies over MNase concentrations (Supplementary Text, Fig. 1F, and fig. S1).

Although the nucMACC and sub-nucMACC scores are calculated in a similar manner, their scores do not correlate at overlapping positions, confirming that both metrics reveal distinct nucleosomal features (fig. S3B). The correlation of the nucMACC scores with the original MACC scores (31) was modest [correlation coefficient ($R$) = 0.48], which can be explained by the fact that the scores in our analysis are centered on the actual nucleosome positions revealing functional details with high resolution. The sub-nucMACC scores, which is an additional category implemented in our pipeline, showed no correlation ($R = -0.24$), indicating additional features of the dynamic chromatin structure (fig. S3, C and D).

By centering the nucMACC scores on the TSS, we show that the +1 nucleosomes of actively transcribed genes are in general hyper-accessible, whereas nucleosomes further downstream in the gene body did not substantially change nucleosome accessibility (Fig. 1G). Remarkably, in genic regions, we observed no difference in accessibility between silent and highly expressed genes (fig. S3E), confirming that chromatin accessibility is not changed at a global scale but modulated only locally, changing the accessibility of individual nucleosomes at regulatory sites, such as the promoter (30).

The nucleosome stability profile at the TSS of active genes was different compared to the nucleosome accessibility score. The lowest nucleosome stability scores were obtained upstream of the +1 nucleosome directly at the TSS of active genes. The sub-nucleosome analysis uncovered unstable nucleosomes directly at the NDR, showing an overall low nucleosome occupancy in mono-nucleosome analysis (Fig. 1G). In contrast, the transcription end sites (TESs) are nucleosome depleted in *Drosophila*, as no unstable nucleosomes can be detected (fig. S3F), in contrast to studies performed in yeast and mouse (39, 40).

Overall, the nucleosome features provided by the nucMACC pipeline are robust and reproducible, as shown by the comparison to other MNase titration experiments that were combined with a H3-ChIP (fig. S4, A to G) (31). We observed the highest variation between the H3- and the H4-MNase-ChIP experiments in the number and location of the sub-nucleosome positions (fig. S4C). This can be explained by the substantially lower number of sub-nucleosomal fragments in the H3 low-MNase samples compared to the H4 dataset, giving rise to fewer called sub-nucMACC positions (data S1, dependency of sequencing depth on position calling addressed in more detail in Fig. 6, C and D). Nevertheless, the sub-nucMACC scores at overlapping positions revealed a high correlation (fig. S4D, $R = 0.83$).

In summary, splitting up MNase titration data by fragment sizes into mono-nucleosomes and sub-nucleosomes enables the nucMACC pipeline to reveal nucleosome positions, nucleosome accessibility, and nucleosome stability scores within the same experiment.

Next, we tested whether the additional histone ChIP step after chromatin digestion is required to obtain reliable nucleosome features. Without ChIP, mono- and sub-nucleosome positioning changed and overlapping positions showed only a modest correlation with whole chromatin extractions (fig. S4, H to K). Changes in the

fragment profile were clearly visible upstream of the TSS, and most pronounced at the −1 nucleosome position affecting nucMACC and sub-nucMACC scores (fig. S4, L to N). The −1 nucleosome position exhibited the highest accessibility throughout all conditions, whereas the sub-nucleosomal fraction exhibited a broad and pronounced enrichment at the promoter of expressed genes in the low MNase concentrations. This result is reminiscent to other methods measuring chromatin accessibility, such as DNase- or assay for transposase-accessible chromatin using sequencing (41, 42). The differences between whole chromatin analysis and the combined MNase-ChIP datasets can be explained by nonhistone proteins bound to DNA (31). Our analysis shows that an additional histone ChIP step is required, to obtain accessibility and stability measurements specifically relating to nucleosomes. Nevertheless, the nucMACC pipeline can be run using whole chromatin extractions and measures the overall accessibility and stability of chromatin across all molecules bound to DNA.

## nucMACC scores identify hyper- and hypo-accessible nucleosomes

To identify functionally important nucleosomes, the nucMACC pipeline extracts nucleosomes with exceptionally high or low accessibility, termed hyper- or hypo-accessible nucleosomes. For this, the nucMACC score of each nucleosome is plotted against its rank. Ranks are divided by the total number of nucleosomes and the nucMACC score, by the total range of scores. In that way, both the x and y axes have a value range of 1 (Fig. 2A and data S2). To geometrically identify the positions, where the signal decreases/increases rapidly, we determined the positions at the curve where the slope is the first/last time below one. Hypo-accessible nucleosomes are defined as the nucleosome positions before the slope of the curve decreases below 1 (Fig. 2A, purple dots), and hyper-accessible nucleosomes are defined as the nucleosome positions after the slope increases again above 1 (Fig. 2A, green dots). A similar strategy was previously applied to define super-enhancers (43).

In total, we identified 6763 (2% of all tested nucleosome positions; $n = 333.792$) hypo-accessible and 16,342 (4.9%) hyper-accessible nucleosomes in the *Drosophila* S2 H4-MNase-ChIP dataset. Hyper-accessible nucleosomes are strongly enriched at DNase I hypersensitive sites, whereas hypo-accessible nucleosomes are depleted, confirming the nucMACC score as a measure for general DNA accessibility (Fig. 2B). Genomic feature characterization of hyper- and hypo-accessible nucleosomes revealed a substantial enrichment at promoter and intergenic regions, respectively (Fig. 2C). Hyper-accessible nucleosomes are preferentially associated with H3K4me1, H3K4me3, and H3K27ac histone modifications, marking active TSS and enhancer, whereas hypo-accessible nucleosomes are associated with H3K9me3 and H3K27me3 modifications, indicating heterochromatic regions (Fig. 2D). In agreement with the histone modification analysis, hyper-accessible nucleosomes occupy mainly expressed genes, while hypo-accessible nucleosomes are present on silent genes (Fig. 2E). Nevertheless, we identified a substantial fraction of hypo-accessible nucleosomes at exons of expressed genes.

The nucMACC pipeline identifies functional nucleosome positions at regulatory sites of the genome, showing that the regulation of genome activity is associated with differential nucleosome accessibility.
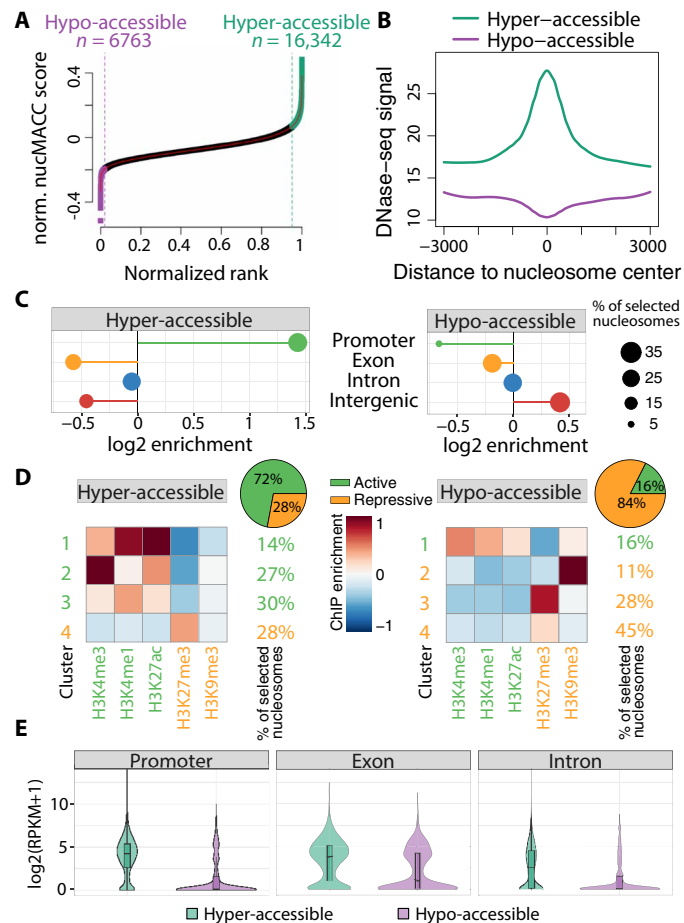
Fig. 2. Identification of hyper- and hypo-accessible nucleosomes. (A) Selection of hyper-accessible (green) and hypo-accessible (purple) nucleosomes based on the nucMACC scores. Nucleosomes were ranked by the nucMACC score. *x* and *y* axes were normalized to a range of one. Cutoffs were deduced from the slope of the LOESS curve fit. Nucleosomes before the point, where the slope goes the first time below 1 (purple dashed line), were classified as hypo-accessible. Nucleosomes after the point where the slope of the curve exceeds one again (green dashed line), were classified as hyper-accessible nucleosomes. (B) Correlation of DNase-seq sites with hyper-/hypo-accessible nucleosomes. (C) Enrichment of hyper-/hypo-accessible nucleosomes at genomic features. (D) K-means clustering ($n = 4$) of histone modifications at hyper-/hypo-accessible sites. Centroid values of each cluster are indicated by the color code. The relative number of nucleosomes represented in each cluster is indicated on the right side. Repressive histone modifications are highlighted in orange and active in green, respectively. The pie chart depicts the fraction of nucleosomes assigned to either active or repressive histone clusters. (E) Gene expression levels of genes exhibiting hyper-/hypo-accessible nucleosomes at distinct features. Promoter was defined as 500 bp upstream of the TSS to 300 bp downstream.

## Sub-nucMACC scores identify the presence of noncanonical and unstable nucleosomes

To measure nucleosome stability, sub-nucleosome positions were derived from the sub-nucleosome fragments in the lowest MNase concentration condition. Hyper-accessible nucleosomes are expected to have a higher MNase degradation rate as they are less protected once they are released from chromatin. Here, sub-nucleosome positions are selected, which either contained at least fourfold more normalized sub-nucleosome fragment counts than the mononucleosome fraction or did not overlap with any mono-nucleosome position (fig. S1).

Like the selection of hyper- and hypo-accessible nucleosomes, nucleosomes exhibiting extreme stability scores (sub-nucMACC) were selected (Fig. 3A and data S2). This selection strategy resulted in 4233 (6.7%) nucleosomes showing very low stability scores and therefore termed unstable nucleosomes. Additionally, a rather small subset of nucleosomes ($n = 355$, 0.6%) were enriched in the sub-nucleosome fraction of the lowest-MNase condition and remained stable at higher MNase concentrations. This group represents most likely structurally altered nucleosomes with additional MNase cleavage sites in the realm of the nucleosome core or nonoctamer histone compositions, such as hexa- or hemisomes (*15*, *24*, *44*). Nucleosomes enriched in the low sub-nucleosomal fractions and showing extraordinarily high stability scores are referred to here as noncanonical nucleosomes.

In contrast to noncanonical nucleosomes, unstable nucleosome positions are characterized by a high chromatin accessibility (Fig. 3B), being mainly located at gene promoters (53.8%), whereas noncanonical nucleosomes are enriched in intergenic and exonic regions (Fig. 3C). Preferentially, active histone marks are associated with unstable nucleosomes (Fig. 3D), and accordingly, unstable nucleosomes are present at expressed genes (Fig. 3E). We identify a subgroup of unstable nucleosomes associated with active H3K4me1 and repressive H3K27me3 marking bivalent enhancers. Conversely, noncanonical nucleosomes showed enrichment in repressive H3K27me3 and H3K4me1 (Fig. 3D). Nevertheless, 30% of the noncanonical nucleosomes are associated with active histone modifications, and therefore noncanonical nucleosomes are located at both expressed and silent genes (Fig. 3E).

## Unstable nucleosomes occupy promoters with high RNA polymerase pausing rate

As we observed that unstable nucleosomes are enriched upstream of the TSS (Fig. 1G and fig. S5A) of expressed genes, we addressed their role in gene regulation. Therefore, we divided all expressed genes into two groups: (i) those that contained an unstable nucleosome 200 bp upstream of the TSS ($n = 1588$), and (ii) those that exhibited an NDR in the promoter ($n = 5363$) (Fig. 4A, fig. S5B, and data S2). To discriminate between NDR or unstable nucleosome-containing promoters, it is obligatory to use MNase treatments coupled with histone immunoprecipitations, as DNA fragments protected by nonhistone proteins would dominate the promoter signal in whole chromatin MNase digestion, resulting in sub-nucleosomal fragments at all expressed promoters (fig. S5C). Overall, the steady-state transcript abundance of the corresponding genes was similar in both groups and only slightly elevated for genes carrying an unstable nucleosome (Fig. 4B, *P* value = $2 \times 10^{-15}$). Gene ontology (GO) enrichment analysis revealed that genes containing unstable nucleosomes in their promoter are stimulus or developmentally regulated (Fig. 4C). Furthermore, unstable nucleosome promoters were enriched for *Drosophila* promoter motifs Unknown1 to 3 and M1BP motif, but depleted in Initiator, TATA-box, and Unkown4 motif (Fig. 4D). As M1BP is a known factor to regulate RNA polymerase II (Pol II) pausing depending on nucleosome positioning (*45*), it is suggested that Pol II pausing provides an efficient way to induce gene transcription rapidly or synchronously, in response to developmental and environmental signals. We tested whether Pol II pausing differed between
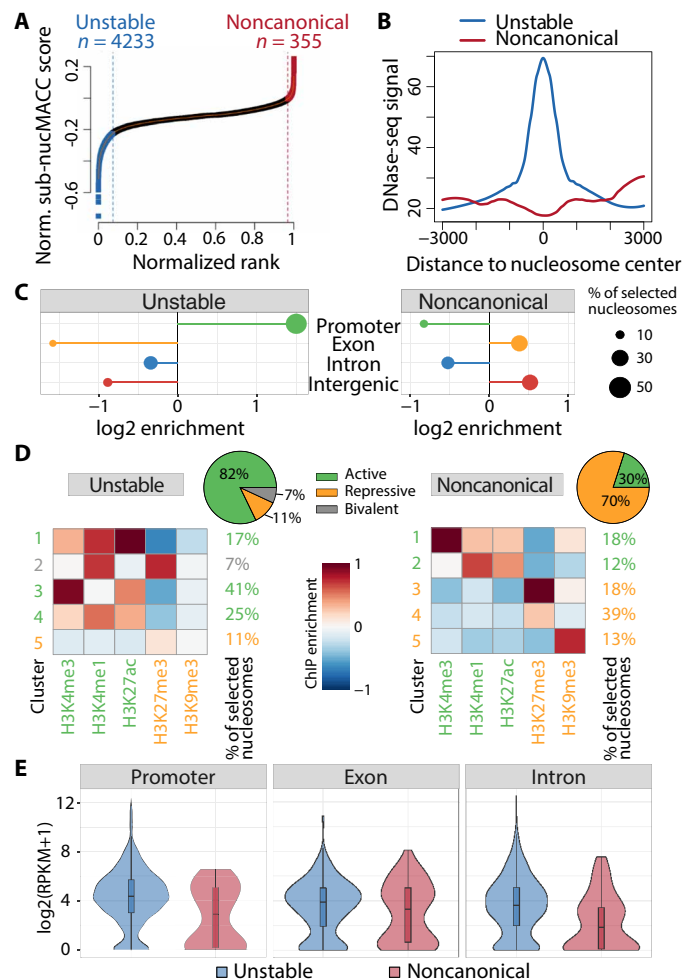
**Fig. 3. Identification of noncanonical and unstable nucleosomes.** (**A**) Selection of unstable (blue) and noncanonical (red) nucleosomes based on nucMACC scores. Nucleosomes were ranked by the sub-nucMACC score. *x* and *y* axes were normalized to a range of one. Cutoffs were deduced from the slope of the LOESS curve fit. Nucleosomes before the point, where the slope goes the first time below 1 (blue dashed line) were classified as unstable. Nucleosomes after the point where the slope of the curve exceeds one again (red dashed line) were classified as noncanonical nucleosomes. (**B**) Correlation of DNase-seq sites with unstable and noncanonical nucleosomes. (**C**) Enrichment of unstable or noncanonical nucleosomes at genomic features. (**D**) K-means clustering (*n* = 5) of histone modifications at unstable or noncanonical sites. Centroid values of each cluster are indicated by the color code. The relative number of nucleosomes represented in each cluster is indicated on the right side. Repressive histone modifications are highlighted in orange, and active histone modifications are highlighted in green. The bivalent cluster enriched for both repressive and active histone modifications is indicated in gray. The pie chart depicts the fraction of nucleosomes assigned to either active, repressive, or bivalent histone clusters. (**E**) Gene expression levels of genes exhibiting unstable or noncanonical nucleosomes at distinct features. Promoter was defined as 500 bp upstream of the TSS to 300 bp downstream.

NDR or unstable nucleosome promoter using nascent RNA sequencing data (*46*). In agreement with M1BP motif enrichment, unstable nucleosome promoters show a pronounced association with paused Pol II directly downstream of the TSS and higher pausing index rates compared to the NDR promoters, enabling a fast transition into productive transcription at responsive and developmental genes (Fig. 4, E and F).

## M1BP binds to DNA occupied by an unstable nucleosome

We next assessed the relationship between the binding of the transcription pausing factor M1BP and nucleosome properties. Here, we investigated nucleosome positioning, accessibility, and stability at annotated M1BP binding sites (*47*). We observed that M1BP binding sites are surrounded by hyper-accessible nucleosomes (Fig. 5, A to C). Directly in the binding center of M1BP, we identified a clear enrichment of unstable nucleosomes (Fig. 5, D to F). This observation suggests that M1BP does not bind to nucleosome-free DNA but rather to DNA, still co-occupied with histones like a pioneer TF (*48*). Another possibility is that M1BP competes with nucleosomes at the same position and the signal is a mixture of two different populations of cells that have either a nucleosome or M1BP bound at that position. Regardless, the enrichment of unstable nucleosomes clearly shows that histone DNA interactions are apparently weakened, making them sensitive to internal MNase digestions, which could be the result of TF binding to the nucleosome core (*49*) or active remodeling (*50*).

## Evaluating the robustness of the nucMACC pipeline

Previously, four or more MNase digestion conditions have been used to determine chromatin accessibility scores (*31*, *32*). We sought to determine the minimum number of titration points per sample to obtain reliable and robust results with our pipeline. We compared the nucMACC and sub-nucMACC scores using two or four titration points of the same experimental dataset. As a result, we show that only two different MNase conditions are required, being highly correlated with the four titration conditions. This correlation holds true, if the difference in the compared MNase concentration conditions is at least 10-fold (Fig. 6, A and B, and fig. S6, A and B). In addition, we only observed consistent results, when using the lowest available MNase concentration to detect unstable nucleosomes within the sub-nucleosomal fragments. At higher concentrations, fragments of unstable nucleosomes are often lost due to their fragile nature.

Next, we addressed the importance of sequencing depth for the robustness of the nucMACC pipeline, as MNase-seq analysis usually requires a high sequence coverage of the genome. To obtain a genome-independent quantification, we normalized the sequencing depth to the number of fragments per nucleosome (FPN). To estimate the total number of nucleosomes, we divided the size of the mappable genome by the expected nucleosome repeat length (here, we used 170 bp as an approximation):

$$\mathrm{Fragments\,per\,nucleosome\,(FPN)} = \frac{\mathrm{number\,of\,sequenced\,fragments}}{\frac{\mathrm{effective\,genome\,size}}{\mathrm{nucleosome\,repeat\,length}}}$$

To demonstrate the dependence on the experimental sequencing depth, we calculated FPN(total) using the total number of sequenced fragments (Fig. 6C). In addition, to assess mono-nucleosome and sub-nucleosome coverage independently, we used the respective size-selected fragments of all MNase concentrations for mono-nucleosomes [FPN(mono)] and of the lowest concentration for sub-nucleosomes [FPN(sub)] (Fig. 6D). We found that using half of the fragments, with a coverage of 160 FPN(total) or 53 FPN(mono), preserves 73% of the mono-nucleosome positions, and the nucMACC score correlation is reasonably high (*R* = 0.88) (Fig. 6, C and D, and fig. S6C), whereas further lowering the sequencing depth to 80
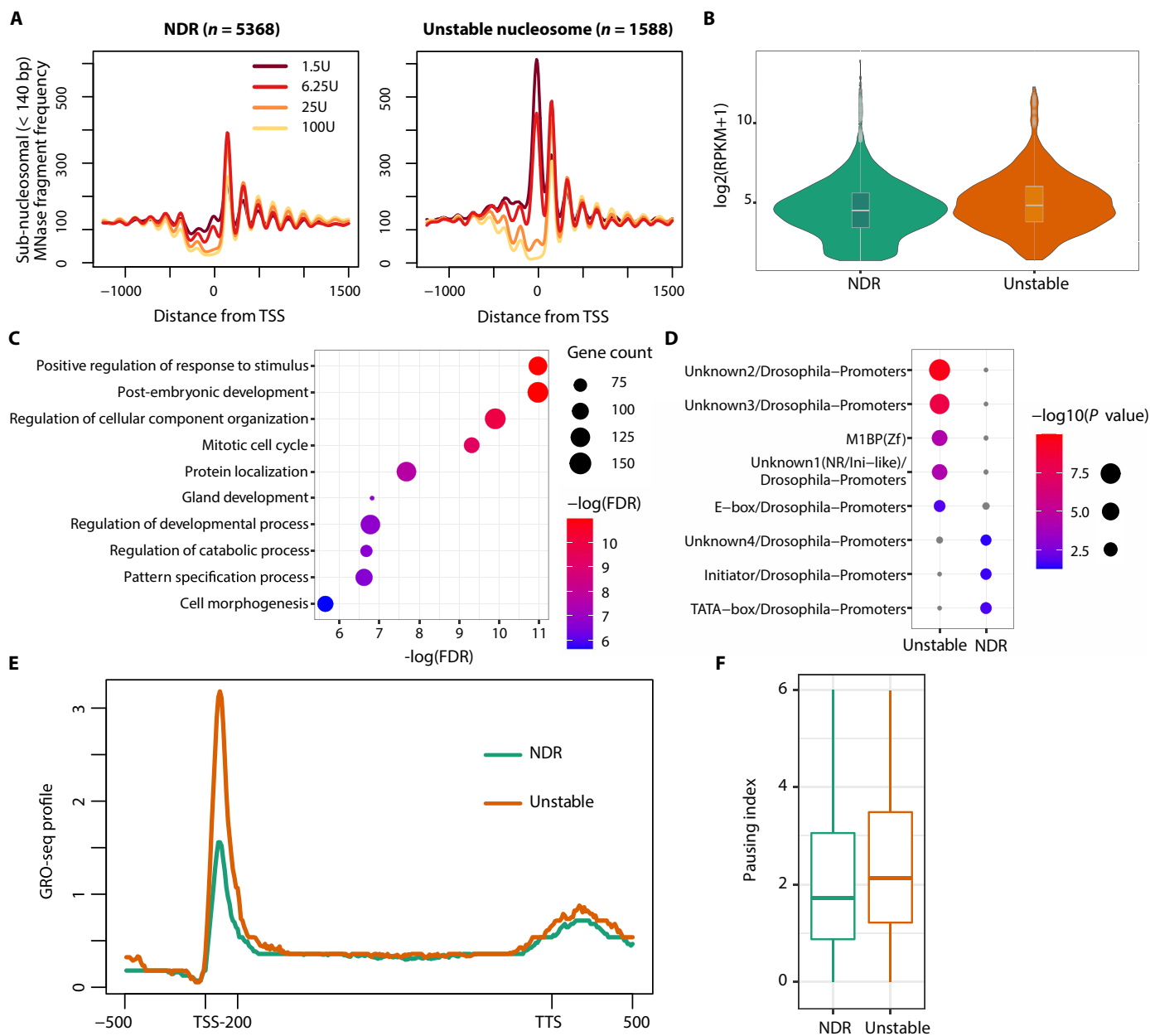
**Fig. 4. Unstable nucleosomes upstream of the TSS are associated with long RNA pol II pausing.** (**A**) Average fragment frequencies at the TSS of expressed genes either containing an unstable nucleosome (n = 1588, right) or exhibiting a nucleosome-depleted region (NDR) (n = 5368, left) directly upstream of the TSS. (**B**) Violin plot showing expression levels of genes either containing an unstable nucleosome (n = 1588, right) or exhibiting an NDR (n = 5368, left) directly upstream of the TSS. (**C**) Enrichment of biological processes associated with genes characterized by an unstable nucleosome promoter. (**D**) Motif enrichment analysis in the unstable nucleosome (left) or NDR (right) promoter. (**E**) Median GRO-seq profile showing nascent RNA abundance. Regions upstream of TSS, downstream of transcription termination site (TTS), and the first 200 bp downstream of TSS are unscaled. The remaining gene body is scaled to have the same length for each gene. (**F**) Boxplots showing the pausing index based on the GRO-seq signal. The pausing index was calculated as the fold change of the GRO-seq signal in the first 200 bp downstream of the TSS versus the remaining gene body.

FPN(total) or 27 FPN(mono) results in a considerable loss of called mono-nucleosome positions to 25%, albeit the nucMACC score correlation remains stable at overlapping nucleosome positions, with a correlation of 0.80. For sub-nucleosomes, a value of 320 FPN(total) or 1.8 FPN(sub) already results at a low coverage, because the sub-nucleosomal fragments are underrepresented in low MNase digestion conditions. Reducing the coverage to 240 FPN(total) or 1.3 FPN(sub) decreases the called sub-nucleosomes to 49% (Fig. 6, C and D).

Nonetheless, the correlation between sub-nucMACC scores remains very high at overlapping positions (fig. S6D). Below 240 FPN(total) or 1.3 FPN(sub) coverage, the sub-nucleosome analysis produces an insufficient amount of called sub-nucleosome positions, not allowing a meaningful analysis (Fig. 6, C and D). In summary, we recommend a minimum sequencing depth of 240 FPN(total) or higher to be used when analyzing nucleosome stability. In addition, specific enrichment of sub-nucleosomal–sized fragments before library
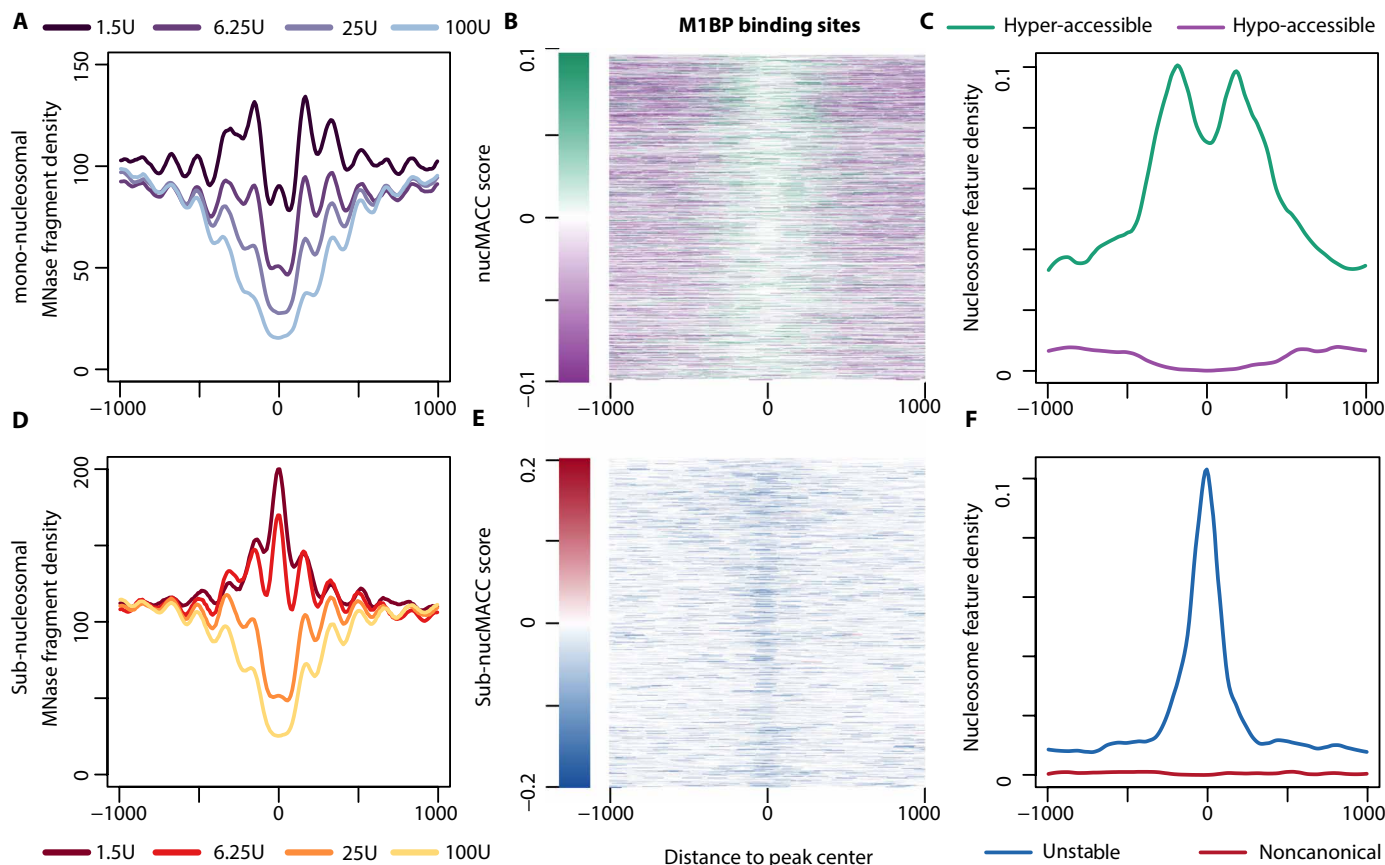
**Fig. 5. Nucleosome features at M1BP binding sites.** (**A** and **D**) Average fragment frequencies of mono- (A) and sub-nucleosomal (D) fragments at M1BP-bound sites. (**B** and **E**) Heatmap showing nucMACC (B) or sub-nucMACC (E) scores at M1BP-bound sites. (**C** and **F**) Distribution of hyper-/hypo-accessible (C) or unstable and noncanonical (F) nucleosomes at M1BP-bound sites.

preparation or setting a higher sequencing depth for the lowest MNase concentration could overcome the high sequence depth limitations of sub-nucleosome analysis.

Ultimately, we assessed the reliability and robustness of nucMACC and sub-nucMACC scores, testing if the differential number of nucleosomes released at various MNase titration conditions would bias the scores. Therefore, we ran our pipeline on a published dataset with spiked-in nucleosomes and compared the results with or without spike-in information. Remarkably, we find almost no difference in nucMACC scores when comparing the pipeline with or without spike-in information for mono-nucleosomes (Fig. 6E) and sub-nucleosomes (Fig. 6F).

In conclusion, the nucMACC pipeline does not require spike-ins and is robust and reproducible with only two MNase titrations.

## Unstable nucleosomes are associated with nucleosome remodeling

As the nucMACC pipeline produces consistent results with only two MNase titrations, we used appropriate published datasets to gain further insights into chromatin biology. A recent MNase-seq study in *Saccharomyces cerevisiae* questioned the existence of MNase-sensitive nucleosomes at promoter regions (*51*). We used our nucMACC pipeline to re-analyze the H4-MNase-ChIP-seq dataset starting from their raw sequencing files.

Investigating the results of the mono-nucleosome analysis, we observed that both hyper- and hypo-accessible nucleosomes are enriched at the promoter and predominantly at the +1 nucleosome position (fig. S7A and data S3). We selected the genes associated with either a hyper- or a hypo-accessible +1 nucleosome and compared the initiation frequency and pausing rate using native elongating transcript sequencing (NET-seq) data (*52*). In contrast to the results in *Drosophila melanogaster*, genes associated with a hypo-accessible +1 nucleosome exhibited higher expression rates (fig. S7B). This can be explained by the rather homogenous chromatin accessibility in yeast, due to the high gene density and overall high transcriptional activity (*53*, *54*). Here, the local accessibility to MNase is mainly restricted by the local crowdedness of chromatin and additional nonhistone proteins. Hypo-accessible +1 nucleosomes showed a higher occupancy of initiating RNA Pol II and factors of the initiation complex, indicating increased DNA protection from MNase digestion by the RNA Pol II preinitiation complex (fig. S7C). This is in agreement with the higher expression rates compared to hyper-accessible +1 nucleosomes. At normal conditions, we did not observe a characteristic enrichment of other factors at hyper-accessible promoters but detected a pronounced enrichment of the heat shock factor 1 (Hsf1) and Spt7. Spt7 is a subunit of the SAGA complex, located upstream of the TSS upon heat shock, suggesting that these promoters are kept accessible for rapid inducible activation (fig. S7D).
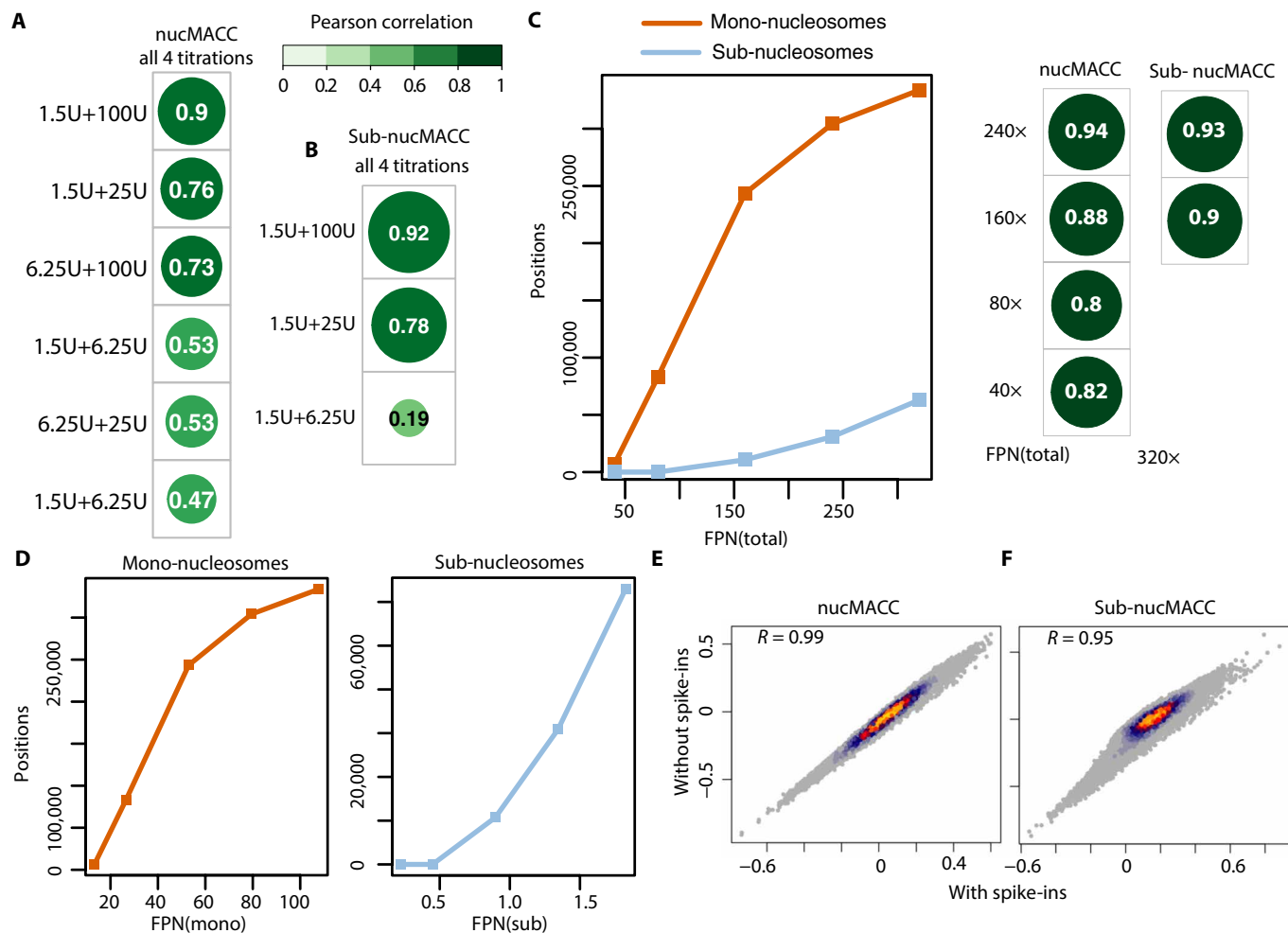
**Fig. 6. Robustness of the nucMACC results.** (**A** and **B**) Correlation of nucMACC (A) or sub-nucMACC (B) scores between samples with two or four MNase titrations. The Pearson correlation coefficient is indicated. (**C** and **D**) The number of called nucleosomes based on (C) total sequencing depth FPN(total) or (D) number of size-selected fragments for mono-nucleosomes FPN(mono) or sub-nucleosomes FPN(sub) (FPN, fragments per nucleosome). (C) Correlation of nucMACC or sub-nucMACC scores between samples with different sequencing depths is illustrated on the right. (**E** and **F**) Scatter density plot showing the correlation of nucMACC (E) or sub-nucMACC (F) scores with or without spike-in information. The Pearson correlation coefficient $R$ is indicated.

Next, we investigated the sub-nucleosomal fraction. A recent study suggested that noncanonical nucleosomes predominate in yeast cells (*55*); however, we detected only few stable noncanonical sub-nucleosomes (*n* = 105) using the nucMACC pipeline (data S3). We detected unstable nucleosomes not only at the TES (fig. S8A), as already described (*38*, *40*, *51*), but also directly upstream of the TSS, like in *D. melanogaster* (Fig. 7A and data S3). Next, we subdivided promoters of expressed genes into promoters with NDR and promoters harboring an unstable nucleosome. In total, 2942 (53%) promoters are occupied by an unstable nucleosome, illustrating the abundance of this promoter type (Fig. 7A and fig. S8B). Our classification allowed us to detect sub-nucleosomal structures in independent H2B- and H3-MNase-ChIP datasets, although higher MNase digestion conditions were used (fig. S8C) (*56*). Those sub-nucleosomal patterns were absent using MNase-ChIP directed against histone modifications, such as H3K4me3, indicating that they are not a result of cross-linking artifacts. Histone variant H2A.Z is enriched at unstable nucleosome promoters, agreeing with

recent studies showing sub-nucleosomal structures in *S. cerevisiae* (fig. S8C) (*40*, *50*). It was suggested that unstable nucleosomes are preferentially found at wide NDRs (>150 bp distance between +1 and −1 nucleosome) (*37*). Accordingly, we checked whether sub-nucleosomal fragments are preferentially enriched at wide NDRs but could not observe a distinct difference (fig. S8D). As the main difference between the work of Kubik *et al.* (*37*) and ours was that we used MNase-ChIP-seq data instead of sole MNase digestions, we reasoned that nonhistone proteins protecting DNA accessibility might have biased their analysis. The mono-nucleosome–sized fragments at wide NDRs accumulated an additional peak upstream of the TSS, which was not present in the H4-MNase-ChIP seq data (fig. S8E). Furthermore, the sub-nucleosome fraction revealed a high abundance of nonhistone proteins upstream of the TSS, likely resulting in nucleosome-sized fragments at wide NDRs under low-MNase conditions.

The transcription initiation rate at NDR and unstable nucleosome promoters did not significantly differ (Fig. 7B). Nevertheless,
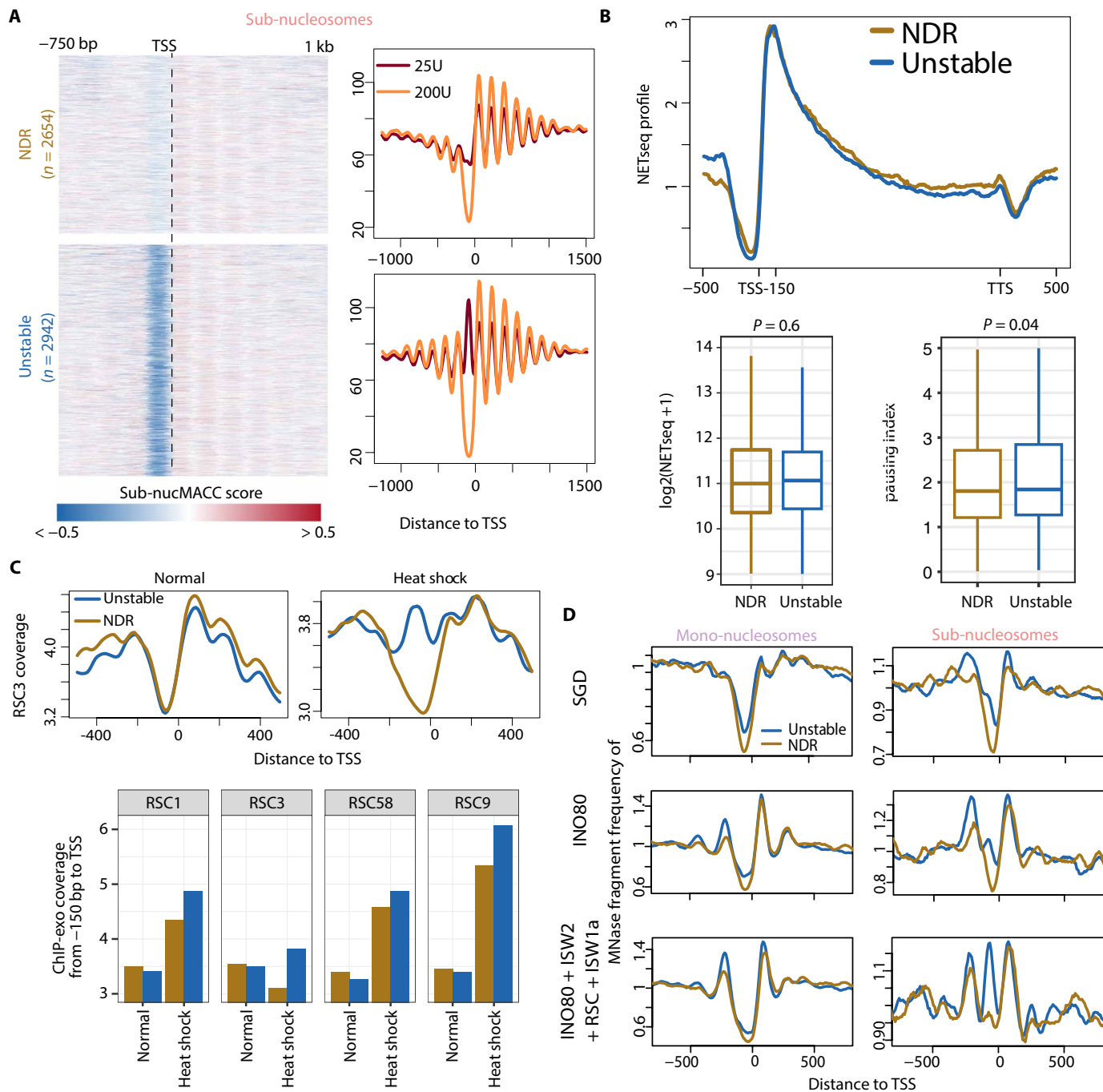
**Fig. 7. Unstable nucleosomes occupy promoters in yeast and are formed by chromatin remodeling.** (**A**) Grouping of promoter types into promoter harboring an unstable nucleosome (unstable; *n* = 2942) or a nucleosome-depleted region (NDR; *n* = 2654) in *S. cerevisiae* (*51*). Left: Heatmap showing the distribution of sub-nuccMACC scores. Right: Average plot showing the sub-nucleosomal fragment frequencies of the different MNase conditions. (**B**) Gene expression analysis of the promoter subgroups (unstable in blue and NDR in gold) using NETseq data. Top: Median NETseq profile showing nascent RNA abundance. Regions upstream of TSS, downstream of TTS, and the first 150 bp downstream of TSS are unscaled. The remaining gene body is scaled to have the same length for each gene. Bottom: Boxplots showing the RNA Pol II initiation rate as calculated from the median NETseq signal over the +1 nucleosome (first 150 bp downstream of TSS) is shown on the left. The pausing index is shown on the right. The pausing index was calculated as the fold change of the NETseq signal in the first 150 bp downstream of the TSS versus the remaining gene body. (**C**) Occupancy analysis of RSC components at unstable nucleosome or NDR promoter types in the context of heat shock using ChIP-exo data (*56*). Top panel shows the median normalized ChIP-exo coverage of the subunit RSC3 under normal and heat shock conditions. Bottom panel shows the median normalized ChIP-exo signal at the first 150 bp upstream of the TSS for RSC subunits RSC1, RSC3, RSC58, and RSC9. (**D**) MNase fragment density of in vitro reconstituted nucleosome arrays at unstable nucleosomes or NDR promoter. MNase fragments were size-selected into mono-nucleosomal (left) and sub-nucleosomal (right) fragments. Nucleosomes were assembled onto DNA by SGD and purified remodeler were added as indicated (*57*).

the pausing index was increased at the promoters harboring an unstable nucleosome as already seen in *D. melanogaster*.

Because it was suggested that RSC (remodeling the structure of chromatin) nucleosome remodeling leads to MNase-sensitive structures through partially unwrapped nucleosome intermediates (*50*), we tested whether subunits of the RSC complex are overrepresented at the sites of unstable nucleosome promoters compared to NDR promoters using the ChIP-exo datasets (*56*). We did not detect differences at normal conditions; however, upon heat shock–induced stress, we identified several RSC subunits, being enriched at the sites of the unstable nucleosomes (Fig. 7C).

Next, we addressed whether chromatin remodelers are sufficient to generate unstable nucleosomes. Hence, we used MNase-seq data of genome-wide in vitro reconstitutions, where the activity of purified remodelers was tested (*57*).

The nucleosomes, assembled by salt gradient dialysis (SGD) onto genomic DNA, exhibited a very low level of sub-nucleosomal structures at the unstable nucleosome promoters (Fig. 7D). The number of sub-nucleosomal fragments increased upon the addition of INO80, creating wider NDRs by positioning the +1 and −1 nucleosomes. The addition of ISW2, RSC, and ISW1 further increased the amount of sub-nucleosomal fragments, specifically at unstable nucleosome promoters. Such a distinction of sub-nucleosome fragments into two groups could not be achieved using the NDR width classification after (*37*) (fig. S8F). Nevertheless, INO80 alone seems to be sufficient to adjust the NDR width into either narrow or wide NDRs.

In summary, the nucMACC pipeline uncovered unstable nucleosomes upstream of the TSS of specific genes in *S. cerevisiae*. Unstable nucleosomes are formed by the activity of chromatin remodeling enzymes in reconstituted nucleosome arrays, suggesting that they generate partially unwrapped nucleosomal intermediates, or sub-nucleosomal structures, like hexasomes.

### Application of the nucMACC pipeline for the characterization of mammalian chromatin structures

To demonstrate the general applicability of the nucMACC pipeline across different species, we re-analyzed whole chromatin MNase-seq experiments performed in mouse and human cell lines (data S1) (*31*, *58*). The TF SOX2 plays a pivotal role in the regulation of pluripotency in embryonic stem cells (ESCs) and multipotency in neural progenitor cells (NPCs) (*59*). SOX2 mostly occupies distinct genomic regions in ESCs and NPCs, and only a small subset of binding sites are common (Fig. 8, A and B) (*59*). As a pioneer factor, SOX2 can bind to nucleosome-occupied motifs and initiate the process of chromatin opening (*48*, *49*). The computed nucMACC scores reveal nucleosome hyper-accessibility, specifically at sites of SOX2 binding, revealing a cell type–specific pattern (Fig. 8C). In addition, upon SOX2 binding to the nucleosome, we detect increased levels of sub-nucleosomal fragments in the low-MNase conditions, as demonstrated by the negative sub-nucMACC score (Fig. 8D). These results can be explained by a recent structural study showing that nucleosome binding of SOX2 introduces DNA distortions that destabilize the nucleosome (*48*). Accordingly, we show that the sub-nucMACC scores enable the detection of destabilized nucleosomes in mammalian genomes.

The required sequencing depth for MNase-seq analysis determines the main experimental cost. This is especially true for larger genomes, such as the mammalian genomes. To obtain high coverage

differential MNase-seq data in human cell lines, we re-analyzed a published dataset using the MNase-Transcription Start Site Sequence Capture method (mTSS-seq) to study the depletion of the H2A histone variant H2A.Z at the TSS (*58*, *60*) (data S1). Upon H2A.Z depletion (shH2A.Z), we observe the most abundant changes directly upstream of the TSS in the sub-nucleosomal fraction (Fig. 8E). Consistently, the nucMACC pipeline retrieved a notable increase of unstable protein-DNA interactions in the promoter of several genes (*n* = 680) upon shH2A.Z (Fig. 8, F and G, and data S4). With the nucMACC pipeline, it is now possible to obtain an in-depth characterization of the chromatin structure of the affected gene promoters (Fig. 8G to I). The unstable protein-DNA interactions detected in the sub-nucleosomal fraction are likely to occur due to TF binding, as chromatin hydrolysis was applied without the additional step of histone immunoprecipitations (Fig. 8G). Consistently, shH2A.Z results in increased chromatin accessibility at the respective gene promoters (Fig. 8H), enhancing the binding of additional factors to these sites. It would not have been possible to detect these chromatin features in a standard MNase-seq experiment, as no change in nucleosome positioning is detected (Fig. 8I). Changes in the chromatin structure can be attributed to the absence of H2A.Z, as the H2A.Z histone variant is abundant at the affected promoters under normal conditions (Fig. 8I). The genes characterized by unstable chromatin structures upon shH2A.Z treatment showed notable expression levels under standard conditions (Fig. 8J) and were preferentially down-regulated in shH2A.Z (Fig. 8K).

In conclusion, we show that the nucMACC pipeline produces consistent results in mammalian cells and can be used to compare chromatin structure under different conditions, thereby simplifying chromatin analysis in complex organisms.

### DISCUSSION

We describe a complete MNase-seq analysis pipeline called nucMACC, which is designed to analyze genome-wide nucleosome and sub-nucleosome positions, nucleosome accessibility, and stability using two or more MNase titration conditions. It provides a systematic approach to select nucleosome positions with exceptional properties that play regulatory roles in the genome. The systematic detection of unstable nucleosomes unveils their regulatory role in RNA Pol II promoter-proximal pausing and TF binding.

MNase-seq is an established method, and studies to analyze chromatin accessibility have previously been performed (*31*, *32*). However, the use of these pipelines remains challenging to implement, as they use specific annotation files, which are only available for a few organisms, or require preinstalled software packages, some of which are not maintained anymore or require subscriptions. The nucMACC pipeline is publicly available and utilizes the nextflow workflow management system and Docker software containerization guaranteeing a seamless start on different computer systems and reproducible results (*61*). In addition, the nucMACC pipeline uncovers additional properties of structurally altered nucleosomes and sub-nucleosomes, resulting in insights into chromatin function.

The nucMACC pipeline relates to the original MACC concept, which gives a qualitative and quantitative description of chromatin accessibility (*31*). We made two major modifications to obtain a higher resolution in mono-nucleosome analysis: (i) We selected only mono-nucleosomal fragments for the analysis by narrowing the fragment length selection to 140 to 200 bp. (ii) Instead of subdividing the
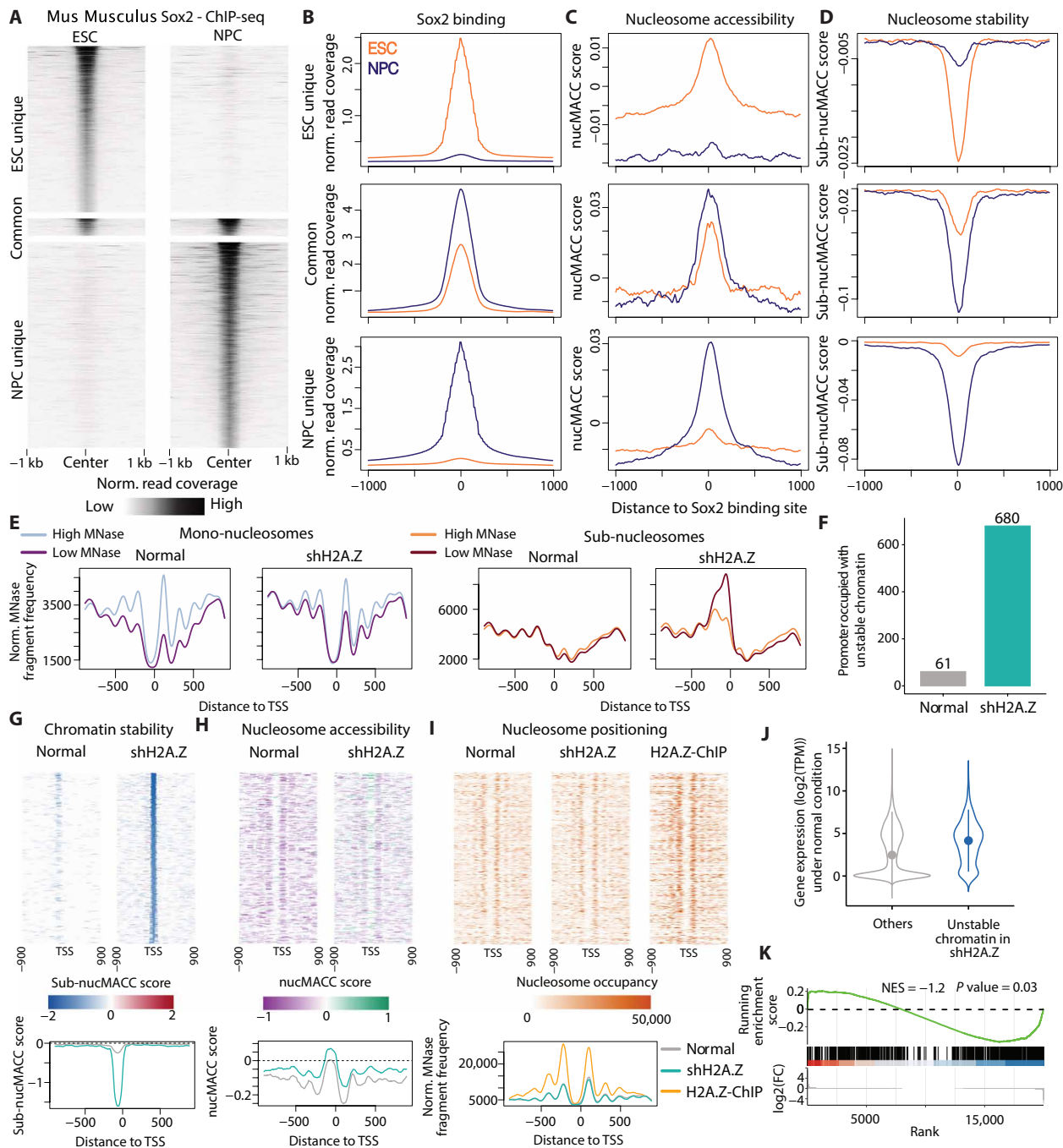
**Fig. 8. Mammalian chromatin analysis using the nucMACC pipeline.** (**A**) Heatmaps showing Sox2 enrichment in mouse ESCs and NPCs. Sox2 binding sites were grouped into ESC unique ($n = 12,399$), NPC unique ($n = 15,376$), and common ($n = 1271$) sites (*59*). (**B** to **D**) Meta-plots showing the average (B) Sox2 ChIP-seq read coverage, (C) nucMACC score as a measure for nucleosome accessibility, or (D) sub-nucMACC score as a measure for nucleosome stability for each Sox2 category. (**E**) Meta-TSS plots showing high and low MNase fragment frequencies in human MCF10A cells under normal conditions and upon H2A.Z depletion (shH2A.Z). Fragments were split by fragment sizes into mono-nucleosomes ($\geq 140$ bp and $<200$ bp; left side) and sub-nucleosomes ($<140$ bp; right side). (**F**) Barplot showing the number of gene promoters associated with unstable chromatin particles directly upstream of the TSS ($-150$ to $50$ bp distance). Unstable chromatin particles were defined using the sub-nucMACC analysis. (**G** to **I**) Heatmaps showing (G) sub-nucMACC scores as a measure of chromatin particle stability, (H) nucMACC scores as a measure of chromatin accessibility, and (I) mono-nucleosome fragment coverages as a measure of nucleosome positioning. Below, the corresponding meta-plot is shown. Only the promoter of genes exhibiting an unstable chromatin particle in shH2A.Z ($n = 680$) is shown. (I) In addition, MNase combined with H2A.Z ChIP data is shown to highlight H2A.Z positions. (**J**) Violin plots showing RNA abundance in normal condition. Protein-coding genes were spilt into genes, which exhibited an unstable chromatin particle in the promoter in shH2A.Z ($n = 680$) or without (others). Read counts per gene were normalized to TPMs (transcripts per million) and log-transformed. (**K**) GSEA of genes with unstable chromatin particle in shH2A.Z. All protein-coding genes were ranked based on log2 fold change in RNA abundance between normal and shH2A.Z condition. The analysis revealed an enrichment at down-regulated genes in shH2A.Z [normalized enrichment score (NES) = $-1.2$] with a *P* value of 0.03.

whole genome into arbitrary bins, we used the called nucleosome positions to measure chromatin accessibility. Consequently, every nucleosome is further characterized by its linker DNA accessibility. Nevertheless, the main findings of the original MACC study about chromatin accessibility could be confirmed with our pipeline.

A recent study used six different MNase titrations and the addition of spike-ins to accurately measure the kinetics of nucleosome release from chromatin (q-MNase-seq) (*32*). The high number of MNase experiments and high sequencing depth requirements make the data generation extremely laborious and costly, especially for organisms with large genomes. Here, we established the minimal parameters to reveal chromatin accessibility and could show that the nucMACC pipeline works robustly with only two MNase titration conditions that should differ in about one order of MNase units applied. The use of spike-ins did not improve the analysis and is therefore not required as the pipeline measures the relative change between the conditions. Nevertheless, to quantitatively measure absolute nucleosome occupancy between various conditions, spike-ins are compulsory. We could not compare the results of our nucMACC analysis directly with the q-MNase analysis (*32*), as analyzed data tracks are not available, and we were not able to reproduce the results shown in the publication based on the provided method description. However, our analysis confirmed the results of the q-MNase-seq analysis, that hetero- and euchromatic regions exhibit similar accessibility rates in *D. melanogaster* (*32*), which has been initially observed in human cells (*30*). As a unique feature, nucleosomes with extraordinary chromatin accessibility are automatically detected in the nucMACC pipeline. Hyper-accessible nucleosomes are enriched at the promoter and correlate with DNase-seq assays, transcriptional activity, and active histone marks, whereas hypo-accessible nucleosomes show opposite characteristics. However, with only 4.9% hyper- and 2% hypo-accessible nucleosomes in the entire *D. melanogaster* genome, most of the nucleosomes have similar accessibility. Therefore, our results are not consistent with the classical view of chromatin domain organization as a rigid two-state model consisting of either condensed and inaccessible heterochromatin or open and active euchromatin. Our results support a more recent model of chromatin organization in which chromatin forms condensed domains with liquid-like properties ensuring a general chromatin accessibility, while the active processes take place at the hyper-accessible surface of the condensed domain (*36*). Hypo-accessible sites can be established by local nucleosome arrangements forming dense nucleosome "clutches," as previously observed in super-resolution nanoscopy (STORM) or native chromatin cryo-electron tomography (*62*, *63*).

So far, MNase-sensitive nucleosomes have been identified by direct comparison of high- and low-MNase conditions at a limited subset of selected genomic regions, such as TSS, enhancers, or CTCF binding sites (*30*, *37*, *39*, *40*, *50*, *64*–*66*). The nucMACC pipeline provides a systematic, unbiased, and genome-wide approach to detect unstable nucleosomes. In agreement with previous reports, we find them preferentially upstream of the TSS of active genes. However, in contrast, we do not observe a correlation between transcription strength or NDR width and the presence of unstable nucleosomes (*37*, *39*, *50*, *66*, *67*). As the positions of the denominated MNase-sensitive nucleosomes overlap with binding sites of nonhistone factors at TSS, enhancer, or TF binding sites, we clearly show that it is necessary to use MNase combined with histone-ChIP datasets. Otherwise, DNA protected by nonhistone

proteins will result in nucleosome-sized fragments, preferentially at low-MNase conditions, and bias the results. Here, we show that promoters in *D. melanogaster* or *S. cerevisiae* can be classified into two distinct promoter nucleosome configurations, promoters occupied by an unstable nucleosome upstream of the TSS, and nucleosome-depleted promoters. Although their associated genes have similar transcript abundance, these promoter types are functionally different. We observed that genes associated with unstable nucleosomes had higher RNA Pol II pausing rates in both *D. melanogaster* and *S. cerevisiae*. In line with other studies, we find unstable nucleosomes being associated with specific genes that are transcriptionally activated in response to stimuli (*40*, *65*). Unstable nucleosomes are found in the promoters of genes before the induction of environmental changes, suggesting that these genes are poised for rapid up-regulation. Therefore, we hypothesize that nucleosome stability is encoded at gene promoters, contributing to the high transcriptional plasticity observed at these sites. We could confirm our hypothesis using the in vitro reconstituted nucleosome arrays on the yeast genome (*57*). In these datasets, we clearly observe that the addition of chromatin remodeling enzymes results in the formation of unstable nucleosomes at the same subset of promoters as in vivo. These findings highlight the different nucleosome configuration capacities of promoter sequences in recruiting remodeling activities, which shape gene function.

We can only speculate about the exact structure and composition of unstable nucleosomes. In vertebrates, it was shown that histone variants like H3.3 or H2A.Z contribute to nucleosome instability (*68*). We observed a partial enrichment of H2A.Z in yeast unstable nucleosomes, but this correlation is not sufficient to account for the stability of most unstable nucleosomes. Furthermore, we used cross-linked nucleosome datasets, which would overturn the weakened histone DNA interactions observed for histone variants (*40*).

We show that unstable nucleosomes are enriched in sub-nucleosomal fragments and are not limited to wide NDRs, but can also be found at narrow NDRs, where the space is too short to assemble full nucleosomes. These findings indicate the existence of smaller, noncanonical sub-nucleosomal structures, such as hexasomes or hemisomes (*69*). However, here we face one of the main limitations in MNase-seq data—looking at population wide averages. We cannot exclude the possibility that different nucleosome array configurations exist in individual cells. Here, it would be interesting to investigate unstable nucleosome promoters further, using single-molecule DNA sequencing approaches to measure nucleosome positions on individual chromatin fibers (*70*).

Another possibility might be that unstable nucleosomes represent sub-nucleosomal intermediate states of the adenosine triphosphate–dependent nucleosome remodeling process. It was reported that histones accumulate upstream of the TSS at unstable nucleosome promoters, upon mutation of the FACT subunit SPT16, indicating that those sites undergo active remodeling to promote RNA Pol II transcription through the +1 nucleosome (*67*, *71*). Furthermore, partially unwrapped nucleosome structures were proposed during INO80 or RSC remodeling, resulting in sub-nucleosomal structures (*23*, *50*). In line with these studies, we observed preferential enrichment of RSC components at unstable nucleosome positions upon heat shock and an increase in the number of unstable nucleosomes, by adding nucleosome remodeler to reconstituted nucleosome arrays.

In summary, we provide a publicly available pipeline for analyzing nucleosome positions and nucleosomal properties using MNase titrations, which provides a systematic framework to identify unstable nucleosomes.

## MATERIALS AND METHODS

### The nucMACC pipeline

All the steps from fastq files to annotate (sub-)nucMACC scores were integrated and automated within the nucMACC pipeline, which is available on Zenodo (DOI: 10.5281/zenodo.10777489) or on GitHub at uschwartz/nucMACC. The version of the nucMACC pipeline used in this study was v.1.2. The nucMACC pipeline was executed using nextflow, an open-source workflow management system, and runs the software within stable Docker containers, ensuring a reproducible analysis workflow (61).

Raw sequencing data in the form of fastq files were first subjected to quality control checks using FastQC (72). Reads were mapped against the reference genome (dm3/sacCer3) using Bowtie2 with the following parameters: --very-sensitive-local --no-discordant (73). Alignments with mapping quality below 30 were removed using samtools (74). Next, the alignments were filtered based on their size to categorize them into mono-nucleosomes (140 to 200 bp) or sub-nucleosomes (<140 bp) using the alignmentSieve function from the deepTools suite (75). Additionally, any blacklisted regions were removed during this step to ensure that only reliable regions were considered for further analysis. Nucleosome positions were identified using DANPOS2 (76). For this, pooled mono-nucleosome samples were used to call the positions of mono-nucleosomes, while sub-nucleosome positions were called from samples with the lowest MNase concentration. Here, library size was normalized to the effective genome size, and paired-end reads were centered to their midpoint and extended to 70 bp to obtain nucleosome-sized positions. The GC content at the identified nucleosome positions was measured using the bedtools genomecov function (77). To quantify the fragments associated with each nucleosome position, the featureCounts tool was used (78). Nucleosome positions with low fragment counts were filtered out to ensure robustness in the downstream analysis. Specifically, mono-nucleosome positions with fragment count less than 30 and sub-nucleosome positions with counts less than 5 were removed.

### nucMACC score calculation

First, the fragment counts were normalized based on the library size to obtain counts per million (CPM). A pseudocount was determined as the median of the sample median count. A linear regression analysis was performed between log2-transformed MNase concentration and the log2-transformed normalized counts with the added pseudocount. The raw nucMACC scores were estimated from the slope obtained from the regression analysis multiplied by −1. Next, the raw nucMACC scores were corrected for the underlying GC content. Nucleosome positions exhibiting rarely occurring extreme GC content were filtered out. A locally weighted scatterplot smoothing (LOESS) fit was calculated for the remaining GC content and corresponding raw nucMACC scores using a smoothing parameter α of 0.1. Raw nucMACC scores were subtracted by the deviation of the LOESS fit from the median raw nucMACC score at the corresponding GC content. The sub-nucMACC scores were determined in the same manner, except that the slope values were not multiplied by −1 to represent the stability scores.

### Identification of hypo-/hyper-accessible nucleosomes

To identify hypo- and hyper-accessible nucleosomes, the nucMACC scores were normalized to a deviation from the highest to lowest value of 1 and plotted against the ranks of the nucMACC scores, which were normalized by the total number of ranks. A LOESS smoothing was applied and the first derivate of the LOESS fit was calculated to deduce the slope of the curve. Hypo-accessible nucleosomes are determined as the lowest nucMACC scores until the slope of the curve falls below 1 and hyper-accessible nucleosomes are the highest nucMACC scores after the slope of the curve exceeds again a slope of 1.

### Identification of unstable and noncanonical nucleosomes

Unstable and noncanonical nucleosomes were similarly determined as hyper-/hypo-accessible nucleosomes, except that sub-nucleosome positions are selected, which either are unique or exhibit at least fourfold more normalized counts in the lowest-MNase condition compared to the mono-nucleosomes. Unstable nucleosomes are determined as the lowest sub-nucMACC scores until the slope of the curve falls below 1 and noncanonical nucleosomes are the highest sub-nucMACC scores after the slope of the curve exceeds again a slope of 1.

### Spike-in analysis

We compared the results of the nucMACC pipeline ignoring the spike-in information with a slightly modified version using the spike-ins for normalization. Therefore, reads were aligned to the sacCer3 *S. cerevisiae* genome and treated analogously to the nucMACC pipeline. The number of fragments per mono- and sub-nucleosome fraction was counted, and normalization factors for the nucMACC analysis in *D. melanogaster* were calculated by dividing the number of *S. cerevisiae* derived fragments by 1 million.

### Downstream analysis

Density correlation scatterplots were generated using the LSD R package (79). Heatmaps and metaplots were calculated using the computeMatrix, plotHeatmap, and plotProfile functions of the deepTools suite (v3.5.4) (75) and further modified in R using basic plot functions or the pheatmap package (80). Genomic features were annotated using the Bioconductor package ChIPseeker (81). Motif enrichment analysis was carried out using the findMotifs.pl script from the HOMER software suite (82). Motifs from the integrated fly database were tested for enrichment at the TSS ± 100 bp of genes associated either with an unstable nucleosome or an NDR. The DNA sequence around the TSS of all expressed genes served as background.

Gene set over-representation analysis was carried out using the Metascape online tool (83). Enrichment was performed using the GO database for biological processes (BP) and restricted to terms with at least 20 genes overlapping to the tested gene set. Gene set enrichment analysis was carried out using the R package clusterProfiler (84).

### Bulk RNA-/GRO-/NET-seq analysis

Bulk RNA-, GRO-, and NET-seq data have been analyzed using the in-house nextflow transcriptomics pipeline available on Zenodo (DOI: 10.5281/zenodo.10853116) or GitHub (uschwartz/RNAseq).

Initially, quality control of the raw sequence reads was conducted using FastQC (v0.11.8) (72). Subsequently, reads were mapped to the reference genome and corresponding gene annotation using the Spliced Transcripts Alignment to a Reference (STAR) software (v2.7.8a) (85). The following options were used to optimize the alignment process:

--outFilterType BySJout, --outFilterMultimapNmax 20, --alignSJoverhangMin 8, --alignSJDBoverhangMin 1, --outFilterMismatchNmax 999, --alignIntronMin 10, --alignIntronMax 1000000, --outFilterMismatchNoverReadLmax 0.04, --runThreadN 12, --outSAMtype BAM SortedByCoordinate, --outSAMmultNmax 1, and --outMultimapperOrder Random.

Postmapping quality control was performed using the rnaseq analysis mode of Qualimap (v2.2.1) (*86*). The level of PCR duplication was assessed using Picard MarkDuplicates (v2.21.8) (*87*) and dupRadar (v1.15.0) (*88*). Gene expression quantification was carried out using featureCounts (v1.6.3) (*78*).

### ChIP-exo data processing

ChIP-exo sequencing data in *S. cerevisiae* was downloaded and trimmed with TrimGalore (v0.6.7) (github.com/FelixKrueger/TrimGalore). Trimmed reads were aligned to the sacCer3 reference genome using bowtie2 v2.4.2 (*73*) with the following options: "--very-sensitive -I 0 -X 500 --fr --no-mixed --no-discordant–dovetail." Duplicate reads were identified and removed using picard v2.18.7 (*87*). Bigwigs mimicking standard ChIP were generated using the bamCoverage function from deepTools (v3.5.1) (*75*) with the following options: "-bs 10 --normalizeUsing CPM --extendReads --centerReads --minMappingQuality 1."

### Public datasets used in this study
#### *Drosophila melanogaster*

Whole chromatin MNase-seq, MNase-ChIP-H3-seq, MNase-ChIP-H4-seq and RNA-seq (GSE78984), M1BP-ChIP-seq (GSE101554), GRO-seq (GSE58955), and MNase-seq with spike-ins in S2 (GSE12-8689) were downloaded from the European nucleotide archive (ENA) database. Histone modification ChIP-seq data were retrieved through modENCODE, with the following accession numbers: H3K27ac (296), H3K27me3 (298), H3K4me1 (304), H3K9me3 (313), and H3K4me3 (3761). DNase-seq (GSE172753) data were retrieved from the Gene Expression Omnibus database.

#### *Saccharomyces cerevisiae*

Whole chromatin MNase-seq, MNase-ChIP-H4-seq (GSE83123), NET-seq (GSE25107), in vitro MNase-seq on SGD assembled nucleosomes (GSE72106), and ChIP-exo data of Rsc3, Rsc1, Rsc58, Rsc9, Spt7, Hsf1, Bdf1, CTDSer5, Ssl1, Ssl2, Taf2, Taf3, Tfa1, Tfb1, Tfb2, Tfb3, Tfb4, and Tfg1 (GSE147927) were downloaded from the ENA database.

#### *Mus musculus*

Sox2 ChIP-seq (GSE35496) and whole chromatin MNase-seq (GSE78984) in ESC and NPC were downloaded from the ENA database.

#### *Homo sapiens*

Whole chromatin and H2A.Z ChIP MNase-seq data (GSE134297) and RNA-seq data (GSE134298) in MCF10A cells were downloaded from the ENA database.

### Supplementary Materials

**This PDF file includes:**
Supplementary Text
Figs. S1 to S8
Legends for data S1 to S4

**Other Supplementary Material for this manuscript includes the following:**
Data S1 to S4

## REFERENCES AND NOTES

1. A. L. Olins, R. Douglas Carlson, D. E. Olins, Visualization of chromatin substructure: Upsilon bodies. *J. Cell Biol.* **64**, 528–537 (1975).
2. K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, T. J. Richmond, Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260 (1997).
3. R. D. Kornberg, Chromatin structure: A repeating unit of histones and DNA. *Science*, 868–871 (1974).
4. C. A. Davey, D. F. Sargent, K. Luger, A. W. Maeder, T. J. Richmond, Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *J. Mol. Biol.* **319**, 1097–1113 (2002).
5. A. Valouev, S. M. Johnson, S. D. Boyd, C. L. Smith, A. Z. Fire, A. Sidow, Determinants of nucleosome organization in primary human cells. *Nature* **474**, 516–520 (2011).
6. J. L. Compton, M. Bellard, P. Chambon, Biochemical evidence of variability in the DNA repeat length in the chromatin of higher eukaryotes. *Proc. Natl. Acad. Sci. U.S.A.* **73**, 4382–4386 (1976).
7. A. Prunell, R. D. Kornberg, Variable center to center distance of nucleosomes in chromatin. *J. Mol. Biol.* **154**, 515–523 (1982).
8. V. B. Teif, Y. Vainshtein, M. Caudron-herger, J. Mallm, C. Marth, T. Höfer, K. Rippe, Genome-wide nucleosome positioning during embryonic stem cell development. *Nat. Struct. Mol. Biol.* **19**, 1185–1192 (2012).
9. K. Struhl, E. Segal, Determinants of nucleosome positioning. *Nat. Struct. Mol. Biol.* **20**, 267–273 (2013).
10. N. Kaplan, I. K. Moore, Y. Fondufe-Mittendorf, A. J. Gossett, D. Tillo, Y. Field, E. M. LeProust, T. R. Hughes, J. D. Lieb, J. Widom, E. Segal, The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* **458**, 362–366 (2009).
11. R. V. Chereji, D. J. Clark, Major determinants of nucleosome positioning. *Biophys. J.* **114**, 2279–2289 (2018).
12. G. Hu, D. E. Schones, K. Cui, R. Ybarra, D. Northrup, Q. Tang, L. Gattinoni, N. P. Restifo, S. Huang, K. Zhao, Regulation of nucleosome landscape and transcription factor targeting at tissue-specific enhancers by BRG1. *Genome Res.* **21**, 1650–1658 (2011).
13. D. E. Schones, K. Cui, S. Cuddapah, T. Y. Roh, A. Barski, Z. Wang, G. Wei, K. Zhao, Dynamic regulation of nucleosome positioning in the human genome. *Cell* **132**, 887–898 (2008).
14. P. B. Talbert, S. Henikoff, Histone variants on the move: Substrates for chromatin dynamics. *Nat. Rev. Mol. Cell Biol.* **18**, 115–126 (2017).
15. D. Hasson, T. Panchenko, K. J. Salimian, M. U. Salman, N. Sekulic, A. Alonso, P. E. Warburton, B. E. Black, The octamer is the major form of CENP-A nucleosomes at human centromeres. *Nat. Struct. Mol. Biol.* **20**, 687–695 (2013).
16. S. Sato, N. Tanaka, Y. Arimura, T. Kujirai, H. Kurumizaka, The N-terminal and C-terminal halves of histone H2A.Z independently function in nucleosome positioning and stability. *Genes Cells* **25**, 538–546 (2020).
17. T. Urahama, A. Harada, K. Maehara, N. Horikoshi, K. Sato, Y. Sato, K. Shiraishi, N. Sugino, A. Osakabe, H. Tachiwana, M. Kagawa, Y. Ohkawa, H. Kurumizaka, Histone H3.5 forms an unstable nucleosome and accumulates around transcription start sites in human testis. *Epigenetics Chromatin* **9**, 2 (2016).
18. H. Tachiwana, W. Kagawa, A. Osakabe, K. Kawaguchi, T. Shiga, Y. Hayashi-Takanaka, H. Kimura, H. Kurumizaka, Structural basis of instability of the nucleosome containing a testis-specific histone variant, human H3T. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 10454–10459 (2010).
19. M. L. Kireeva, W. Walter, V. Tchernajenko, V. Bondarenko, M. Kashlev, V. M. Studitsky, Nucleosome remodeling induced by RNA polymerase II. *Mol. Cell* **9**, 541–552 (2002).
20. R. Prasad, S. D'Arcy, A. Hada, K. Luger, B. Bartholomew, Coordinated action of Nap1 and RSC in disassembly of tandem nucleosomes. *Mol. Cell. Biol.* **36**, 2262–2271 (2016).
21. Y. Liu, K. Zhou, N. Zhang, H. Wei, Y. Z. Tan, Z. Zhang, B. Carragher, C. S. Potter, S. D'Arcy, K. Luger, FACT caught in the act of manipulating the nucleosome. *Nature* **577**, 426–431 (2019).
22. S. Ramachandran, K. Ahmad, S. Henikoff, Transcription and remodeling produce asymmetrically unwrapped nucleosomal intermediates. *Mol. Cell* **68**, 1038–1053.e4 (2017).
23. L. J. Hsieh, M. A. Gourdet, C. M. Moore, E. N. Muñoz, N. Gamarra, V. Ramani, G. J. Narlikar, A hexasome is the preferred substrate for the INO80 chromatin remodeling complex, allowing versatility of function. *Mol. Cell* **82**, 2098–2112.e4 (2022).
24. H. S. Rhee, A. R. Bataille, L. Zhang, B. F. Pugh, Subnucleosomal structures and nucleosome asymmetry across a genome. *Cell* **159**, 1377–1388 (2014).
25. C. Lavelle, A. Prunell, Chromatin polymorphism and the nucleosome superfamily: A genealogy. *Cell Cycle* **6**, 2113–2119 (2007).
26. J. Fei, S. E. Torigoe, C. R. Brown, M. T. Khuong, G. A. Kassavetis, H. Boeger, J. T. Kadonaga, The prenucleosome, a stable conformational isomer of the nucleosome. *Genes Dev.* **29**, 2563–2575 (2015).
27. S. Ramachandran, S. Henikoff, Nucleosome dynamics during chromatin remodeling in vivo. *Nucleus* **7**, 20–26 (2016).

28. S. Baldi, P. Korber, P. B. Becker, Beads on a string—Nucleosome array arrangements and folding of the chromatin fiber. *Nat. Struct. Mol. Biol.* **27**, 109–118 (2020).

29. S. Diermeier, P. Kolovos, L. Heizinger, U. Schwartz, T. Georgomanolis, A. Zirkel, G. Wedemann, F. Grosveld, T. A. Knoch, R. Merkl, P. R. Cook, G. Längst, A. Papantonis, TNFα signalling primes chromatin for NF-κB binding and induces rapid and widespread nucleosome repositioning. *Genome Biol.* **15**, 536 (2014).

30. U. Schwartz, A. Németh, S. Diermeier, J. H. Exler, S. Hansch, R. Maldonado, L. Heizinger, R. Merkl, G. Längst, Characterizing the nuclease accessibility of DNA in human cells to map higher order structures of chromatin. *Nucleic Acids Res.* **47**, 1239–1254 (2018).

31. J. Mieczkowski, A. Cook, S. K. Bowman, B. Mueller, B. H. Alver, S. Kundu, A. M. Deaton, J. A. Urban, E. Larschan, P. J. Park, R. E. Kingston, M. Y. Tolstorukov, MNase titration reveals differences between nucleosome occupancy and chromatin accessibility. *Nat. Commun.* **7**, 11485 (2016).

32. R. V. Chereji, T. D. Bryson, S. Henikoff, Quantitative MNase-seq accurately maps nucleosome occupancy levels. *Genome Biol.* **20**, 198 (2019).

33. U. Schwartz, T. Komatsu, C. Huber, F. Lagadec, C. Baumgartl, E. Silberhorn, M. Nuetzel, F. Rayne, E. Basyuk, E. Bertrand, M. Rehli, H. Wodrich, G. Laengst, Changes in adenoviral chromatin organization precede early gene activation upon infection. *EMBO J.* **42**, e114162 (2023).

34. J. G. Henikoff, J. A. Belsky, K. Krassovsky, D. M. MacAlpine, S. Henikoff, Epigenome characterization at single base-pair resolution. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 18318–18323 (2011).

35. Ř. V. Chereji, T. W. Kan, M. K. Grudniewska, A. V. Romashchenko, E. Berezikov, I. F. Zhimulev, V. Guryev, A. V. Morozov, Y. M. Moshkin, Genome-wide profiling of nucleosome sensitivity and chromatin accessibility in *Drosophila melanogaster*. *Nucleic Acids Res.* **44**, 1036–1051 (2016).

36. K. Maeshima, S. Iida, M. A. Shimazoe, S. Tamura, S. Ide, Is euchromatin really open in the cell? *Trends Cell Biol.* **34**, 7–17 (2024).

37. S. Kubik, M. J. Bruzzone, P. Jacquet, J. L. Falcone, J. Rougemont, D. Shore, Nucleosome stability distinguishes two different promoter types at all protein-coding genes in yeast. *Mol. Cell* **60**, 422–434 (2015).

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

39. L. N. Voong, L. Xi, A. C. Sebeson, B. Xiong, J.-P. Wang, X. Wang, Insights into nucleosome organization in mouse embryonic stem cells through chemical mapping. *Cell* **167**, 1555–1570.e15 (2016).

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

41. J. D. Buenrostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, W. J. Greenleaf, Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).

42. J. Vierstra, H. Wang, S. John, R. Sandstrom, J. A. Stamatoyannopoulos, Coupling transcription factor occupancy to nucleosome architecture with DNase-FLASH. *Nat. Methods* **11**, 66–72 (2014).

43. W. A. Whyte, D. A. Orlando, D. Hnisz, B. J. Abraham, C. Y. Lin, M. H. Kagey, P. B. Rahl, T. I. Lee, R. A. Young, Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).

44. Y. Dalal, T. Furuyama, D. Vermaak, S. Henikoff, Structure, dynamics, and evolution of centromeric nucleosomes. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 15974–15981 (2007).

45. J. Li, D. S. Gilmour, Distinct mechanisms of transcriptional pausing orchestrated by GAGA factor and M1BP, a novel transcription factor. *EMBO J.* **32**, 1829–1841 (2013).

46. N. J. Fuda, M. J. Guertin, S. Sharma, C. G. Danko, A. L. Martins, A. Siepel, J. T. Lis, GAGA factor maintains nucleosome-free regions and has a role in RNA polymerase II recruitment to promoters. *PLOS Genet.* **11**, e1005108 (2015).

47. A. Zouaz, A. Auradkar, M. C. Delfini, M. Macchi, M. Barthez, S. E. Akoa, L. Bastianelli, G. Xie, W.-M. Deng, S. S. Levine, Y. Graba, A. J. Saurin, The Hox proteins Ubx and AbdA collaborate with the transcription pausing factor M1BP to regulate gene transcription. *EMBO J.* **36**, 2887–2906 (2017).

48. A. K. Michael, R. S. Grand, L. Isbel, S. Cavadini, Z. Kozicka, G. Kempf, R. D. Bunker, A. D. Schenk, A. Graff-Meyer, G. R. Pathare, J. Weiss, S. Matsumoto, L. Burger, D. Schübeler, N. H. Thomä, Mechanisms of OCT4-SOX2 motif readout on nucleosomes. *Science* **368**, 1460–1465 (2020).

49. M. Iwafuchi-Doi, K. S. Zaret, Pioneer transcription factors in cell reprogramming. *Genes Dev.* **28**, 2679–2692 (2014).

50. S. Brahma, S. Henikoff, RSC-associated subnucleosomes define MNase-sensitive promoters in yeast. *Mol. Cell* **73**, 238–249.e3 (2019).

51. R. V. Chereji, J. Ocampo, D. J. Clark, MNase-sensitive complexes in yeast: Nucleosomes and non-histone barriers. *Mol. Cell* **65**, 565–577.e3 (2017).

52. L. S. Churchman, J. S. Weissman, Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* **469**, 368–373 (2011).

53. L. T. Marr, J. Ocampo, D. J. Clark, J. J. Hayes, Global histone protein surface accessibility in yeast indicates a uniformly loosely packed genome with canonical nucleosomes. *Epigenetics Chromatin* **14**, 5 (2021).

54. C. Chen, H. H. Lim, J. Shi, S. Tamura, K. Maeshima, U. Surana, L. Gan, Budding yeast chromatin is dispersed in a crowded nucleoplasm in vivo. *Mol. Biol. Cell* **27**, 3357–3368 (2016).

55. Z. Y. Tan, S. Cai, A. J. Noble, J. K. Chen, J. Shi, L. Gan, Heterogeneous non-canonical nucleosomes predominate in yeast cells in situ. *eLife* **12**, RP87672 (2023).

56. M. J. Rossi, P. K. Kuntala, W. K. M. Lai, N. Yamada, N. Badjatia, C. Mittal, G. Kuzu, K. Bocklund, N. P. Farrell, T. R. Blanda, J. D. Mairose, A. V. Basting, K. S. Mistretta, D. J. Rocco, E. S. Perkinson, G. D. Kellogg, S. Mahony, B. F. Pugh, A high-resolution protein architecture of the budding yeast genome. *Nature* **592**, 309–314 (2021).

57. N. Krietenstein, M. Wal, S. Watanabe, B. Park, C. L. Peterson, B. F. Pugh, P. Korber, Genomic nucleosome organization reconstituted with pure proteins. *Cell* **167**, 709–721.e12 (2016).

58. L. Cole, S. Kurscheid, M. Nekrasov, R. Domaschenz, D. L. Vera, J. H. Dennis, D. J. Tremethick, Multiple roles of H2A.Z in regulating promoter chromatin architecture in human cells. *Nat. Commun.* **12**, 2524 (2021).

59. M. A. Lodato, C. W. Ng, J. A. Wamstad, A. W. Cheng, K. K. Thai, E. Fraenkel, R. Jaenisch, L. A. Boyer, SOX2 co-occupies distal enhancer elements with distinct POU factors in ESCs and NPCs to specify cell state. *PLOS Genet.* **9**, e1003288 (2013).

60. B. R. Druliner, D. Vera, R. Johnson, X. Ruan, L. M. Apone, E. T. Dimalanta, F. J. Stewart, L. Boardman, J. H. Dennis, Comprehensive nucleosome mapping of the human genome in cancer progression. *Oncotarget* **7**, 13429–13445 (2016).

61. P. D. I. Tommaso, M. Chatzou, E. W. Floden, P. P. Barja, E. Palumbo, C. Notredame, Nextflow enables reproducible computational workflows. *Nat. Biotechnol.* **35**, 316–319 (2017).

62. Z. Hou, F. Nightingale, Y. Zhu, C. MacGregor-Chatwin, P. Zhang, Structure of native chromatin fibres revealed by Cryo-ET in situ. *Nat. Commun.* **14**, 6324 (2023).

63. M. A. Ricci, C. Manzo, M. F. García-Parajo, M. Lakadamyali, M. P. Cosma, Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo. *Cell* **160**, 1145–1158 (2015).

64. M. Iwafuchi-Doi, G. Donahue, A. Kakumanu, J. A. Watts, S. Mahony, B. F. Pugh, D. Lee, K. H. Kaestner, K. S. Zaret, The pioneer transcription factor FoxA maintains an accessible nucleosome configuration at enhancers for tissue-specific gene activation. *Mol. Cell* **62**, 79–91 (2016).

65. T. E. Jeffers, J. D. Lieb, Nucleosome fragility is associated with future transcriptional response to developmental cues and stress in *C. elegans*. *Genome. Res.* **27**, 75–86 (2017).

66. H. Ishii, J. T. Kadonaga, B. Ren, MPE-seq, a new method for the genome-wide analysis of chromatin structure. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E3457–E3465 (2015).

67. P. Michl-Holzinger, S. Obermeyer, H. Markusch, A. Pfab, A. Ettner, A. Bruckmann, S. Babl, G. Längst, U. Schwartz, A. Tvardovskiy, O. N. Jensen, A. Osakabe, F. Berger, K. D. Grasser, Phosphorylation of the FACT histone chaperone subunit SPT16 affects chromatin at RNA polymerase II transcriptional start sites in Arabidopsis. *Nucleic Acids Res.* **50**, 5014–5028 (2022).

68. C. Jin, G. Felsenfeld, Nucleosome stability mediated by histone variants H3.3 and H2A.Z. *Genes Dev.* **21**, 1519–1529 (2007).

69. C. Y. Y. Chong, L. Gan, Are extraordinary nucleosome structures more ordinary than we thought? *Chromosoma* **132**, 139–152 (2023).

70. N. J. Abdulhay, C. P. McNally, L. J. Hsieh, S. Kasinathan, A. Keith, L. S. Estes, M. Karimzadeh, J. G. Underwood, H. Goodarzi, G. J. Narlikar, V. Ramani, Massively multiplex single-molecule oligonucleosome footprinting. *eLife* **9**, e59404 (2020).

71. S. Obermeyer, L. Schrettenbrunner, R. Stöckl, U. Schwartz, K. D. Grasser, Different elongation factors distinctly modulate RNA polymerase II transcription in Arabidopsis. *Nucleic Acids Res.* **51**, 11518–11533 (2023).

72. S. Andrews, FASTQC. A quality control tool for high throughput sequence data (2010); https://bioinformatics.babraham.ac.uk/projects/fastqc/.

73. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

74. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

75. F. Ramírez, D. P. Ryan, B. Grüning, V. Bhardwaj, F. Kilpert, A. S. Richter, S. Heyne, F. Dündar, T. Manke, deepTools2: A next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).

76. K. Chen, Y. Xi, X. Pan, Z. Li, K. Kaestner, J. Tyler, S. Dent, X. He, W. Li, DANPOS: Dynamic analysis of nucleosome position and occupancy by sequencing. *Genome Res.* **23**, 341–351 (2013).

77. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

78. Y. Liao, G. K. Smyth, W. Shi, FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

79. B. Schwalb, A. Tresch, P. Torkler, S. Duemcke, C. Demel, B. Ripley, B. Venables, LSD: Lots of superior depictions (2020); https://CRAN.R-project.org/package=LSD.

80. R. Kolde, pheatmap: Pretty heatmaps (2019); https://CRAN.R-project.org/package=pheatmap.

81. G. Yu, L. G. Wang, Q. Y. He, ChIPseeker: An R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382–2383 (2015).

82. S. Heinz, C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh, C. K. Glass, Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).

83. Y. Zhou, B. Zhou, L. Pache, M. Chang, A. H. Khodabakhshi, O. Tanaseichuk, C. Benner, S. K. Chanda, Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **10**, 1523 (2019).

84. T. Wu, E. Hu, S. Xu, M. Chen, P. Guo, Z. Dai, T. Feng, L. Zhou, W. Tang, L. Zhan, X. Fu, S. Liu, X. Bo, G. Yu, clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation* **2**, 100141 (2021).

85. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

86. F. García-Alcalde, K. Okonechnikov, J. Carbonell, L. M. Cruz, S. Götz, S. Tarazona, J. Dopazo, T. F. Meyer, A. Conesa, Qualimap: Evaluating next-generation sequencing alignment data. *Bioinformatics* **28**, 2678–2679 (2012).

87. Broad Institute, Picard toolkit (2019); https://Broadinstitute.Github.Io/picard/.

88. S. Sayols, D. Scherzinger, H. Klein, dupRadar: A Bioconductor package for the assessment of PCR artifacts in RNA-Seq data. *BMC Bioinformatics* **17**, 428 (2016).