Optimized MNase-seq pipeline reveals a novel gene regulation via differential nucleosome stability

AUTHORS

Wernig-Zorc Sara[1,2], Schwartz Uwe[1,*], and Längst Gernot[1,*]

[1] University of Regensburg, Biochemistry III, Universitaetsstraße 31, Regensburg, Germany
[2] Present address: St. Anna Children's Cancer Research Institute, Zimmermannplatz 10, Vienna, Austria
* Co-corresponding authors

ABSTRACT

Micrococcal nuclease sequencing (MNase-seq) is the state-of-the-art method for determining chromatin structure and nucleosome positioning. Analysis of MNase-seq data is complex due to variable degrees of sample quality, sequence bias of the endonuclease, and a high sequencing depth requirement. Therefore, we developed an optimized analysis pipeline, publicly available via Nextflow. We show the nucMACC pipeline is consistent and robust and provide recommendations for minimum sequencing depth, MNase titrations, and spike-in use. Besides, the pipeline provides high-resolution nucleosome positions and an automated way of calling non-canonical nucleosomes, un-stable nucleosomes, hyper-accessible nucleosomes, hypo-accessible nucleosomes, and stable canonical nucleosomes. Using the new nucMACC pipeline, we characterized nucleosomal properties, such as stability and accessibility, and explore how these properties change to regulate gene expression. Intriguingly, we find a distinct group of un-stable nucleosomes enriched at the TSS of promoters marked with motif one, an M1BP TF-specific motif. We show that Eukaryotic organisms can modulate nucleosome stability and change the RNA polymerase pausing rate. Modulation of stability is a dynamic process regulating development and response to environmental stimuli and represents a novel mechanism of gene regulation in *D. melanogaster.*

INTRODUCTION

In Eukaryotes, DNA is wrapped around nucleosomes and compacted into chromatin. A nucleosome consists of 147 bp of DNA wrapped around a histone octamer, composed of one H3/H4 tetramer and two H2A/H2B histone dimers. DNA is bound to histones by over 360 direct and indirect hydrogen bonds, rendering the nucleosome a stable particle (1). Neighboring nucleosomes are connected by linker DNA, which varies in length from 15 to 95 bp depending on species and cell type (2). The underlying transcription machinery and regulatory factors require access to chromatin to extract genetic information. Nucleosomal DNA occludes most sequence-specific DNA binding factors, while linker DNA remains accessible. The structure and positioning of nucleosomes play an essential role in regulating

of DNA-dependent processes. Nucleosome positioning patterns and histone tail modifications determine cell-type-specific gene expression, defining genes' active and repressed state (3).

In vitro nucleosome positioning depends on the bendability and structure of the underlying DNA sequence and the effect of steric hindrance between neighboring nucleosomes (4, 5). On the other hand, *in vivo* nucleosome positioning is more complex and is determined by additional factors, such as ATP-dependent chromatin remodelers, histone variants, transcription factors (TF), histone posttranslational modifications, non-coding RNAs, components of the transcription machinery, and RNA polymerase II (Pol-II). Chromatin remodelers are evolutionarily conserved enzymes that can slide histone octamer along the DNA (nucleosome sliding), change the histone composition of nucleosomes (histone exchange), and disrupt (nucleosome eviction) or assemble nucleosomes *de novo* (6). Regulatory regions such as promoters, enhancers, terminators, and insulators are marked by well-positioned nucleosomes (7–9), a common feature of eukaryotic chromatin. In eukaryotes, nucleosomes are actively positioned by chromatin remodelers or locally constrained by barrier elements (10). Tandem repeat sequences and boundary elements like the DNA bound protein CTCF can serve as boundaries or local barriers to position neighboring nucleosomes by a passive mechanism termed statistical positioning (7). Hence, exact maps of nucleosome positioning and chromatin accessibility, in addition to its dynamic changes in cellular processes, are essential for understanding the epigenetic regulation of gene expression.

Micrococcal nuclease (MNase) digestion followed by high-throughput sequencing (MNase-seq) is the method of choice for determining genome-wide nucleosome positions and structure. MNase is a processive endo-exonuclease, preferentially hydrolyzing the accessible linker DNA. Coupling MNase-seq with anti-histone antibody immunoprecipitation (IP) ensures that only histone-bound regions of genomic DNA are sequenced (11, 12), allowing the exact determination of nucleosome boundaries. Initially, complete MNase digestion, hydrolyzing chromatin down to the mono-nucleosomal level, was used to study nucleosome positioning and occupancy. These studies revealed well-positioned nucleosomes flanking Transcription Start Sites (TSS) of expressed genes and nucleosome depleted regions (NDR) at TSS and Transcription Termination Sites (TTS) (13, 14). It was discovered that nucleosome phasing correlates with the level of RNA Pol-II at promoter regions, with more pronounced phasing linked to transcriptional activity (15). Differential sensitivity of MNase to core nucleosomes was observed before the invention of high throughput sequencing (16); however, it took several decades to be applied in genome-wide experiments. We and others recently suggested that different MNase concentration conditions should be combined to identify nucleosomes that are sensitive to MNase digestion and are commonly lost at conditions of high MNase concentration (17, 18). Treatment with a high concentration of MNase will cause hydrolysis of nucleosomal DNA, occupied by hyper-accessible and unstable nucleosomes or nucleosomes positioned at AT-rich regions of the genome (19). Nucleosome stability can be affected by several factors, for example, histone composition, altered nucleosomal structure (20), underlying DNA sequence (21), histone post-translational modifications (PTM), and other factors associated with chromatin (22, 23). Multiple reactions with different MNase concentrations ensure that nucleosomes are released from the genome and provide information on global chromatin structure and accessibility. Current MNase-seq pipelines include 4-6 titrations of

MNase per sample (17, 18). A recent pipeline additionally advises using spike-ins for accurate DNA accessibility scoring (18), resulting in a high workload, required amount of material, and complexity of data assessment.

A particular interest in recent years has been the fragile nucleosomes (FN) which are difficult to study *in vivo*. The term FN was first coined in the 70s and has remained a controversial topic ever since (24). Initially, FNs were proposed to take a split nucleosome form (also called lexosome) as an intermediate during RNA pol-II transcription. Evidence for the lexosome model was based on electron microscopy studies (24) and the accessibility of the usually buried cysteine in H3 (25). The term FNs has been loosely defined in the literature and may exemplify a hemisome, hexasome, lexosome, or octamer with non-canonical histone variants. The commonality between these is a sub-nucleosomal-sized particle in an unstable state. FNs directly upstream of the TSS have been shown by ChIP-exo in *S. cerevisiae* (26), MPE-ChIP-seq in mouse embryonic stem cells (ESC) (27), chemical mapping in mouse ESCs (28), MNase-seq titrations in *C. elegans* embryos (29), CUT&RUN-ChIP in *S. cerevisiae* (12), MNase-seq titrations in *D. melanogaster* S2 cell line (30), human HeLa cell line (11) and *S. cerevisiae* (31). In *C. elegans,* the FN particles have been associated with inducible genes prepped for the transcriptional response to developmental and environmental cues (29). In contrast, they were proposed to be associated with nucleosome eviction in budding yeast (12). Non-canonical nucleosome structures have been associated with FN. For example, hexasomes were present at the +1 position of expressed genes in *D. melanogaster* and promoted RNA Pol-II dependent gene transcription (32). Nucleosomes with the histone variant CENPA at human centromeres were shown to be octasomes with transiently unwrapped DNA (33). In yeast, results suggested both the hemisome (26) and hexasome models (12).

In recent years, several studies have shown that nucleosomes are far more dynamic than initially proposed and can take several transient states to promote transcription (34), heterochromatin formation (35), and form phase-separated condensates. Reshaping the octamer core by Swi6 chromatin remodeler increased nucleosome accessibility, increasing the opportunities for multivalent interactions between nucleosomes (35). The term nucleosome reshaping refers to changes in nucleosome properties resulting in modified nucleosomes stability.

In the present study, we designed a robust and reproducible MNase-seq pipeline to obtain high-resolution nucleosomal positions with their associated stability and accessibility. We are the first to provide a robust and universal way of classifying differentially accessible nucleosomes and nucleosomes with reduced stability herein referred to as unstable nucleosomes. We show that in *D. melanogaster* embryo development, approximately 4.9% of nucleosomes exhibit hyper-accessibility and are enriched in promoter regions of transcribed genes. At the same time, hypo-accessible nucleosomes represent about 2% of all nucleosomes and are enriched in exonic and intergenic regions, possibly having a role in gene expression regulation. We found a special group of nucleosomes based on their stability, the unstable nucleosomes, enriched at M1BP Transcription Factor (TF) binding motifs associated with differential RNA pol-II staling. A sub-group of unstable nucleosomes, representing 1418 nucleosomes, was enriched at the TSS of genes containing motif one, an M1BP-specific binding motif. Promoters bound by M1BP TF have been previously shown to control RNA pol-II pausing rate. Several

studies have suggested that nucleosomes are involved in the M1BP-dependent mechanism of differential RNA pol-II pausing (36). Nonetheless, it was not shown to date how nucleosomes contribute to this mechanism. Using GRO-seq data to quantify RNA pol-II pausing we observe a correlation between genes associated with an unstable nucleosome on the TSS and a reduced RNA pol-II pausing time. Interestingly, most genes with unstable nucleosome coincide with an increased RNA pol-II pausing time. Those are enriched in developmental pathways, which allows us to suggest that unstable nucleosome mark genes are important for spatiotemporal expression of developmental and stress response signals.

Moreover, separating mono- and sub- nucleosomes based on the insert length of sequencing fragments dramatically changes the resolution of MNase-seq results and provides new evidence for nucleosome-regulated gene transcription. Our robust pipeline requires only two digestion conditions of MNase and not requiring spike-ins, thus reducing the workload and costs.

RESULTS

Overview of the nucMACC pipeline.

The nucMACC pipeline requires raw data in .fastq format and a minimum of two MNase titrations per experimental condition. The pipeline works on both MNase-seq and MNase-ChIP-seq data; however, we strongly recommend using MNase and anti-histone immunoprecipitation data when analyzing sub-nucleosomal particles. It starts with read mapping to the reference genome and quality filtering. In the next step, fragments are divided into sub-nucleosomal- and mono-nucleosomal- sized-fragments which are processed separately (Figure 1).

Mono-nucleosomes

MNase titrations are pooled to obtain high sequencing depth and are used to call mono-nucleosome positions. A nucMACC score is calculated by counting fragments per each nucleosome position and each MNase titration condition. A linear regression slope is fitted through MNase titrations for each nucleosome. The slope of the regression line represents the raw score, which is normalized to the underlying GC% content in the following step. The normalized score is referred to as the nucleosome MNase accessibility score (nucMACC). In the final step, we select extreme nucleosomes groups, with nucMACC score considerably deviating from the mean.

Sub-nucleosomes

The lowest MNase titration point is used to call sub-nucleosomal positions and is consequently removed by the mono-nucleosomal positions to obtain proper sub-nucleosome positions. Similar to the mono-nucleosomes, a regression slope is fitted on the MNase titration fragment counts and is normalized to the underlying GC% content. The normalized score is referred to as the sub-nucleosome MNase accessibility score (sub-nucMACC). In the final step, we select extreme nucleosomes groups, with nucMACC score considerably deviating from the mean.

Output

The final output of the pipeline is the quality matrix, unique sub-nucleosome positions with sub-nucMACC scores, mono-nucleosome positions with nucMACC score, positions of hyper-accessible hypo-accessible nucleosomes, un-stable (fragile) nucleosomes, and non-canonical nucleosomes (Figure 1).

To distinguish different nucleosome populations, we use the following definitions:

1. **Fragile/Unstable nucleosomes:** Nucleosomes with reduced stability cause MNase to hydrolyze within the realm of the nucleosomal DNA at low MNase concentrations. They are fully hydrolyzed at high MNase digestion conditions, resulting in a decrease of occupancy with increasing MNase concentration. Present only in the sub-nucleosomal fraction at low MNase digestion.

2. **Non-canonical nucleosomes:** This is a diverse group of nucleosomes with either a non-canonical nucleosome structure (hemisome – a nucleosome particle containing one H3/H4 dimer and one H2A/H2B dimer (37); lexosome – a nucleosome particle containing one H3/H4 tetramer and two H2A/H2B dimer, split in half (24); hexasome - a nucleosome particle containing one H3/H4 tetramer and one H2A/H2B dimer (38) or histone variants. This group represents structurally altered nucleosomes with additional MNase cleavage sites within the realm of the nucleosome, generating a sub-nucleosomal fraction that also remains stable at higher MNase concentrations. Over-digested nucleosomes cannot be distinguished from nucleosomes with non-canonical structures or histone variants without additional experimental validation (ChIP-seq or MNase-ChIP-seq data on histone variants).

3. **Stable nucleosomes:** Canonical nucleosomes with average stability, released at intermediate digestion levels. Present in the mono-nucleosomal fraction in the low and intermediate digestions and might be over-digested (present in the sub-nucleosomal fraction) at higher digestions.

4. **Hyper-accessible nucleosomes:** Hyper accessible nucleosomes are released at lower MNase concentrations as full-sized nucleosomal DNA fragments. They exhibit a regular and stable nucleosomal structure but reside in regions of reduced chromatin compaction, revealing the high accessibility of their DNA linkers for the endonuclease. These nucleosomes are present at mono-nucleosomal fragment size and have a decreased occupancy with higher digestion levels.

5. **Hypo-accessible nucleosomes:** Nucleosomes residing in compacted chromatin regions, correlating with reduced DNA linker accessibility. These nucleosomes are released at higher MNase concentrations in the form of mono-nucleosomal DNA fragments and exhibit a higher occupancy with increasing MNase concentrations.

Characterization of sub-nucleosomal and mono-nucleosomal fragments.

Under mild MNase conditions, nucleosomes at accessible regions in the genome (37) or nucleosomes with reduced stability are preferentially released (Figure 2, A). The differential accessibility of cellular DNA to MNase, a 16.9 kDa endonuclease, is due to chromatin compaction or crowding, DNA sequence

composition, differential nucleosome structure, and DNA linker accessibility. MNase is a processive enzyme; therefore, the digestion kinetics will result in the progressive trimming of nucleosomal particles and even stable nucleosomes at high MNase concentrations.

For the establishment of the pipeline, we used the MNase-seq datasets (GSE78984) – with H3 or H4 immunoprecipitation after MNase digestion, followed by high throughput sequencing. All datasets contain four MNase concentration titrations (1.5U, 6.25U, 25U, and 100U) per sample and two replicates each.

Looking at the fragment sizes of MNase-seq data, we observe disappearance in di-nucleosomes and increasing levels of mono- and sub- nucleosomes with increased MNase digestion (Figure 1, B). Based on the shape of the curve of the fragment size distribution, we selected the size ranges corresponding to mono-nucleosomal DNA (140 – 200 base pairs, Figure 2, B – purple fraction and Figure S1, A) and sub-nucleosomal DNA (50 – 139 base pairs, Figure 2, B – pink fraction and Figure S1, A). When plotting the two fractions relative to the transcription start site (TSS), we observe striking differences in the location of the nucleosomal and sub-nucleosomal fraction on transcribed genes (Figure 2, C-D). The sub-nucleosomes show the highest coverage at the nucleosome depleted region (NDR) upstream of the TSS at very low MNase concentrations. The occupancy is increasingly lost with higher MNase concentrations (Figure 2, C). Since the data result from ChIP-seq after MNase fragmentation, it is evident that very unstable histone-DNA complexes cover the promoter regions of active genes. This is evident for the histone H4 ChIP (Figure 2, C) and the H3 ChIP experiments (Figure S1, B-C). The nucleosomal DNA fraction, in contrast, has a distinct MNase concentration dependent TSS profile over expressed genes. A precisely positioned mono nucleosome is located at the so-called +1 site, exhibiting the highest occupancy at low MNase concentrations (Figure 2, D). Not surprisingly, we find the +1 nucleosomes of expressed genes in a hyper-accessible configuration, meaning the fragment count decreases progressively with increasing MNase concentrations. In contrast, the fragment count does not change significantly for the nucleosomes +2, +3, +4, and subsequent nucleosomes (Figure 2, D). The binding of the initiation complex and ongoing transcription renders the +1 nucleosomes into an open conformation that allows preferential DNA linker cleavage by MNase. We can observe this feature on a per gene basis (Figure 2, E), where a hyper-accessible +1 nucleosome and a hypo-accessible +2, +3, +4 nucleosome can be observed.

We define the nucleosome accessibility and stability scores by comparing the differential MNase digestion conditions, considering the normalized fragment count of the mono- and sub- nucleosomal fractions, respectively. We separate nucleosomes based on gene expression into high, mid, low expression, and silent genes, confirming previously published observation that non-expressed and lowly expressed genes exhibit greater overall fuzziness in nucleosome positioning (Figure 12, F). Similarly, we also observe a difference in nucleosome accessibility and stability scores around the TSS, based on gene expression levels, with most high and mid expressed genes displaying a hyper-accessible +1 nucleosome and a sub-group of those with an un-stable nucleosome at the TSS (Figure 2, F).

Characterization of the nucleosome accessibility scores.

We next set out to find a universal nucMACC score cut-off, to identify nucleosomes exhibiting extraordinary accessibilities. Here we ranked the nucMACC scores, and only the nucleosomes where the score was exceptionally different from the main distribution were selected. The cut-off was defined as the point of the curve where the slope is greater than 1, a strategy that is also used to find super-enhancers (Figure 3, A). In our analysis, we could identify a set of nucleosomes with an extraordinary low nucMACC score, characterizing 2% (6763) of all nucleosomes, indicative of hypo-accessible nucleosomes. We also find hyper-accessible nucleosomes with high nucMACC scores, representing 4.9% (16342) of all nucleosomes (Figure 3, A). Characterization of hyper- and hypo-accessible nucleosomes revealed an enrichment in promoter and exon regions, respectively, suggesting nucleosomes with distinct accessibility preferentially occupy regulatory regions (Figure 3, C).

Interestingly, hyper-accessible nucleosomes were present mainly on expressed genes, while hypo-accessible nucleosomes were mainly present on silent genes (Figure 3, D). This suggests an association with local chromatin states and is also reflected in histone modifications. Hyper-accessible nucleosomes show enrichment in active enhancers and promoters (H3K4me1, H3K27ac, and H3K4me3) and hypo-accessible nucleosomes in heterochromatin (H3K9me3 and H3K27me3) (Figure 3, E). Additionally, we confirmed and validated the accuracy of nucleosome accessibility scores with DNase-seq and ATAC-seq data (Figure 3, F and G). Moreover, hyper-accessible nucleosomes are enriched in GAF and CBP TF occupancy. GAF and CBP TFs have been shown to be involved in nucleosome displacement and gene regulation, respectively (Figure 3, H and I).

Characterization of the nucleosome stability scores.

Next, nucleosomes with extraordinary stability or non-canonical structures were characterized. First, sub-nucleosomes were filtered if they overlapped with a mono-nucleosome position or had less than 4x higher signal than the mono-nucleosomal fraction. This resulted in 63014 positions. Second, in the same way as for mono-nucleosomes, a cut-off for sub-nucMACC scores was devised to define extreme nucleosome groups. Genes marked with a low sub-nucMACC score, i.e., un-stable nucleosomes showed significant enrichment in promoter regions (Figure 4, A) of expressed genes (Figure 4, D) and represented 6.7% (n = 4232) of all sub-nucleosomes (Figure 4, A). Not surprisingly, un-stable nucleosomes showed enrichment in H3K27ac and H3K4me3 histone marks, representing active promoters (Figure 4, F). We also detected sub-populations of un-stable nucleosomes with significant enrichment in H3K27me3 histone modification, marking repressive chromatin state and in H3K27ac and H3K4me1, marking poised enhancers (Figure 4, F). Nucleosomes with a high sub-nucMACC score, i.e., nucleosomes with non-canonical structures, represent 0.5% of all sub-nucleosomes (n = 355) and favor silent genes (Figure 4, D). Non-canonical nucleosomes are enrichment in intergenic and exonic regions (Figure 4, C). These showed enrichment in H3K27me3 and H3K9me3 histone modifications, marking heterochromatin (Figure 4, F).

Gene regulation via modulation of nucleosome stability.

We identified a specific group of un-stable nucleosomes, positioned at TSS of expressed genes, referred to as TSS-un-stable nucleosomes (n = 1418). When comparing the TTS-un-stable nucleosomes with

stable-TSS-nucleosomes of expressed genes (n = 6010), we do not see any difference in gene expression between the groups (Figure 5, A), however we could observe a clear difference in nucleosome occupancy at the TSS (Figure 1, C and Figure S2, A). Strikingly, we find a significant enrichment in M1BP TF motifs, within the promoters of genes harboring the un-stable nucleosomes (Figure 5, C). We asked next whether un-stable nucleosomes prevent TF recruitment and gene activation genes or whether the un-stable nucleosomes allow both to bind situationally. We overlapped the un-stable nucleosomes with ChIP-seq data for M1BP and found significant overlaps and enrichment over expressed genes without a TSS-un-stable nucleosome (Figure 5, D). M1BP TF has been shown to bind to specific promoters marked with motif “one”, having a higher RNA pol-II transcription rate than other promoters in *D. melanogaster* (38). Therefore, we analyzed GRO-seq data and looked for RNA pol-II staling rate for genes occupied with or without a TSS-un-stable nucleosome. We observed a clear difference between the two gene groups and their RNA pol-II occupancy and pausing index (Figure 5, E and G). Next, we tested if un-stable nucleosomes occupy genes with a common biological function. Monitoring for GO terms, we observed an enrichment in developmental, stimuli response, and morphological terms for TSS-un-stable nucleosomes associated genes (Figure 5, F and Figure S5). In comparison, expressed genes without TSS-un-stable nucleosomes are enriched in the GO terms endosomal transport, mitochondrial translation and expression, mitotic division, DNA repair, and RNA splicing (Figure 5, F and Figure S6). Results clearly reflect the developmental stage of the S2 cell line and suggests genes with TSS-un-stable nucleosomes have a regulatory and predictive role in response to stimuli and spatiotemporal expression during the development of the fruit fly.

Evaluating the robustness of the nucMACC pipeline.

Next, we wanted to test the robustness of our pipeline by comparing MNase-H3-ChIP-seq and MNase-H4-ChIP-seq data. We find that 90% of mono-nucleosomes overlap between MNase-H3-ChIP-seq and MNase-H4-ChIP-seq data and exhibit a correlation of 0.68 between the nucMACC scores (Figure 6, A). Sub-nucleosomal-sized fragments show a higher variation in sequencing depth than mono-nucleosomal-size fragments and thus, not surprisingly, show a lower overlap of 39.8% between experimental setups; nevertheless, the correlation between the nucMACC scores was high - corr = 0.83 (Figure 6, B).

Previously four or more MNase digestion conditions have been used to determine DNA accessibility scores (17, 18, 39). 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 scores of two or four titration points of the same experimental dataset. We found the combination of 1.5U + 25U, 1.5U + 100U titrations to have a high correlation for both mono- and sub-nucleosomes, with correlation scores of 0.8, 0.92 and 0.79, 0.79, respectively (Figure 6, C; Figure S7, A; Figure S8, A). The number of called mono-nucleosome positions also correlated with the nucMACC scores (Figure S7, B). In summary, the choice of one low and one high MNase concentration used to analyze mono-nucleosomes is not critical unless the difference in MNase concentration or digestion time is at least 15-fold. In contrast, to detect the sub-nucleosomal population, a very low MNase concentration is required, due to their fragile nature.

Next, we addressed the importance of sequencing depth on data quality, as MNase-seq is a rather expensive experiment. We devised a sequencing coverage matrix independent of the genome size (Formula 1), defined as the number of fragments per kilobase of genome size (FPKG). For mono-nucleosomes, the pooled data had a coverage of 634 FPKG. We found that using half of the fragments, with coverage of 312 FPKG preserves 73% of the mono-nucleosome positions, and the nucMACC score correlation is reasonably high, 0.88 (Figure 6, E and Figure S8, B). Whereas further lowering the sequencing depth to 156 FPKG, results in a considerable loss of called mono-nucleosome positions (25%), albeit the nucMACC score correlation remains stable, with a correlation of 0.80. For sub-nucleosomes, we start with 183 FPKG, already at low coverage. Reducing the coverage to 135 FPKG (75% of sequencing fragments) amounts to a decrease in called sub-nucleosomes to 49% (Figure 6, F and Figure S8, C). Nonetheless, the correlation between nucMACC scores remains very high – 0.90 (Figure 6, F and Figure S8, C). Below this coverage, the sub-nucleosome analysis produces very low amount of called nucleosomes (17%) with a high sub-nucMACC score correlation of 0.86 (Figure 6, F and Figure S8, C). Since nucleosomes with hypo-stability represent a smaller portion of the genome, it is necessary to sequence the genome at a higher coverage. In summary, we recommend a minimum sequencing depth of 180 FPKG or higher to be used when analyzing nucleosomes with differential stability.

Formula 1:

$$\text{FPKG (fragments per kilobase of genome size)} = \frac{\text{Number of fragments} \times 1\text{ kb}}{\text{Genome size (dm3: 162367812)}}$$

Ultimately, we tested the reliability and robustness of our accessibility and stability scores and determined if sequencing depth would bias the scores. This prompted us to use our pipeline on a published dataset with spike-ins and compare 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- (Figure 5, G) and sub- nucleosomes (Figure 5, H). In addition, we can observe a high overlap between extreme nucleosome groups (Figure S9, A-D).

In conclusion, nucMACC pipeline is a robust, reproducible, doesn't require spike-ins and provides clear data with only two MNase titrations. The pipeline is completely automated and publicly available via Nextflow. We are the first to provide an automated way for calling special nucleosome groups in addition to nucleosome positions.

DISCUSSION

In the current study, we set out to optimize and improve the MNase-seq pipeline and characterize nucleosomes based on their accessibility and stability. We observed local changes in nucleosome accessibility, which represent around 7% of all nucleosomes. The majority of these exhibits an increase in accessibility (4.9%), and 2% show a decreased accessibility. Hyper-accessible nucleosomes were enriched in promoter regions, whereas hypo-accessible nucleosomes showed enrichment in intergenic regions (Figure 2, C). We conclude most nucleosomes have similar levels of accessibility, with local

effects of hyper- and hypo- accessibility, which has been also previously observed in human cell lines (11).

Moreover, we show for the first time the global distribution of un-stable and non-canonical nucleosomes and provide an automated way for identifying their positions in the genome (Figure 4, A). We find un-stable nucleosomes enriched on active promoter regions (Figure 4, C and F) of mainly expressed genes (Figure 4, D), while non-canonical nucleosomes are enriched at intergenic and exonic regions (Figure 4, C) of mainly silent genes (Figure 4, D) at heterochromatic regions (Figure 4, F). Furthermore, we show by ChIP-MNase-seq that un-stable nucleosomes contain H4, H3, and H2B histones (Figure 1, D; Figure S1, B; Figure S2, A) and exhibit shorter DNA length. However, we cannot discriminate between nucleosome structures, such as hemisomes, hexasomes or lexasomes.

Intriguingly we find a special group of un-stable nucleosomes enriched at TSS of promoters marked with motif one (Figure 5, C), which has been shown to harbor an M1BP TF-specific motif (38). Undeniably, we see a high enrichment of M1BP ChIP-seq peaks over these un-stable nucleosomes (Figure 5, D). This made us speculate their role in RNA pol-II pausing, which we confirmed with GRO-seq data (Figure 5, E). We could detect a clear difference in RNA pol-II pausing between genes marked with a TSS-un-stable nucleosome (Figure 4, E and G). We hypothesize un-stable nucleosomes located directly on the TSS allow TF binding, despite nucleosomes normally occluding binding. This may enable a rapid response to stimuli and cell type-specific gene expression. The un-stable nucleosome represents a novel gene regulation machinery, adding to the complexity of eukaryotic gene regulation.

Additionally, we also find un-stable nucleosomes within local heterochromatin (Figure 4, F), which is consistent with a recent study by Sanulli et al., where they show nucleosomes in heterochromatic regions are hyper-accessible and promote multivalent interactions with other nucleosomes and histone tails, enabling chromatin compaction into liquid condensates (35). In this study, we separate nucleosome accessibility and stability and show that heterochromatic nucleosomes are unstable.

Moreover, we provide a robust pipeline and guidelines for the experimental setup of the MNase-seq protocol by analyzing different datasets with two or four MNase titrations per sample, with or without spike-in information, different sequencing depths, and the correlation between replicates, and correlation between different anti-histone antibody pulldowns. In short, we show that only two MNase titrations are required, albeit at a high sequencing depth. A minimum sequencing depth of 312 FPKG per sample for mono- and 183 FPKG for sub- nucleosomes is required to attain high-quality results (Figure 6, E and F). This will, of course, depends on the experimental design. For example, if mono- and sub- nucleosomes are excised from the gel, the sequencing depth required per sample will be lower. Mono-nucleosomes correlate well between H3 and H4 data (Figure 6, A). Likewise, the correlation between sub-nucMACC scores of H3 and H4 data is high (Figure 6, B). The differences observed between H3 and H4 data could be due to biological differences in histone composition, antibody efficiency, and the dynamic nature of sub-nucleosomal particles. For sub-nucleosomes, it is crucial to validate them by anti-histone ChIP after MNase treatment to differentiate them from TF bound to DNA

at these sites. Ideally, an anti-TF ChIP-seq would be used for comparison to see whether it leads to structural changes in the nucleosome.

We also analyzed the nucMACC scores with or without spike-in information. Spike-ins are typically used to compare antibody pulldowns with different efficiencies or to make an absolute quantification of RNA expression or, in our case, nucleosome occupancy. We see no improvement or difference in nucMACC scores when using spike-in information (Figure 6, G-H and Figure S9, A-D). In conclusion, for determining absolute nucleosome occupancy, spike-ins are still compulsory and should be included in the experimental design, whereas using relative differences in the occupancy is sufficient to calculate the nucleosome accessibility and stability scores. Sequencing depth seems to have a more detrimental effect on the accuracy of nucMACC scores than the absence of spike-in information. In a recent publication, authors suggest spike-ins are necessary to accurately estimate the nucleosome digestion rate and subsequently accessibility scores (40). Unfortunately, we could not directly compare their scores and ours in the same dataset due to the lack of IP-experiments after MNase digestion. However, we showed that the accessibility scores coincide with ATAC-seq and DNase-seq data (Figure 4, F and G).

In summary, we provide a publicly available Nextflow pipeline for analyzing nucleosome positions and nucleosomal properties using MNase titrations. We provide guidelines for the experimental design of MNase-seq experiments. Lastly, using our pipeline on public MNase-seq datasets in *D. melanogaster* we find a novel mechanism of gene regulation in eukaryotes, via nucleosome stability modulation and RNA pol-II pausing rate.

MATERIALS AND METHODS

Public datasets used in this study.

Whole chromatin MNase-seq, MNase-ChIP-H3-seq, MNase-ChIP-H4-seq and RNA-seq (GSE78984), MNase-ChIP-H2B-seq, MNase-ChIP-H3-seq (GSE69336), M1BP-ChIP-seq (SRX3011239), GRO-seq (GSM3274626), and MNase-seq with spike-ins in S2 and Kc167 cell lines (PRJNA528497) were downloaded from the European nucleotide archive (ENA) database. Histone modification and TF ChIP-seq or ChIP-seq data was retrieved through modENCODE, with the following accession numbers: H3K27ac (296), H3K27me3 (298), H3K4me1 (304), H3K9me3 (313), RNA pol-II ChIP-seq (329), H3K36me3 (3189), H3K4me3 (3761), CBP TF ChIP-seq (858), GAF TF ChIP-seq (858). DNAse-seq (GSM5258764) and ATAC-seq (GSM2756640) were retrieved from the GEO database.

Processing of MNase-seq data

Adapters were removed from raw sequencing reads with Trimmomatic and mapped to the UCSC Drosophila melanogaster dm3 genome, using the Bowtie2 with the following parameters --very-sensitive-local --no-discordant. Aligned reads were filtered for their mapping quality (MAPQ > 30), fragment size (length 140 – 200 nt for mono-nucleosomes and length 50 – 139 nt for sub-nucleosomes)

and ambiguous chromosomes and blacklisted regions were removed using DeepTools2 alignmentSieve. Often there is a shift in the fragment sizes due to variation in MNase enzyme activity; thus, each dataset should be assessed individually. In case of no fragment size selection after MNase digestion, a dominant peak around 150 nt can be observed and a second smaller peak around 340 nt, representing mono- and di- nucleosomes, respectively. Larger di-nucleosomal fragments only reduce the quality of nucleosome positions and should be excluded from the analysis. To achieve higher sequencing depth, mono-nucleosome positions were called on pooled MNase titrations, using DANPOS, with dpos -m 1 --extend 70 -c 162367812 -u 0 -z 1 -a 1 -e 1, parameters. Sub-nucleosome positions were called using only the mild digestion samples, with the following parameters: dpos -m 1 --extend 70 -z 70 -c 162367812 -u 0 -e 1. Here library size was normalized to effective dm3 genome size (162367812), paired-end reads centered to their midpoint and extended to 70bp to obtain nucleosome-sized positions. Sub-nucleosome positions were further filtered by two parameters: a non-overlapping position with a mono-nucleosome or an overlapping position with an enriched signal over a mono-nucleosome signal. Here a cut-off of > 4x higher signal in sub-nucleosomal data as opposed to mono-nucleosomal data was chosen.

NucMACC score calculation

Reads per sample per nucleosome position were counted using featureCounts, using a custom SAF file with nucleosome positions and the following parameters -SAF -p -B -C -largestOverlap. Information on the GC content of each nucleosome was added to the read count file using BEDtools nuc. Nucleosome positions were obtained by converting the DANPOS annotation file to SAF format using AWK. NucMACC scores were calculated using R and edgeR R package. Firstly, raw fragments were filtered for sequencing coverage, with cut-off > 30 fragments per mono-nucleosome and > 5 fragments for sub-nucleosomes. Then, fragments were normalized based on the library size (counts per million, CPM) or to the calculated library size based on normalization factors for spike-in analysis. Linear regression was calculated using normalized fragment counts and log2 MNase concentrations or time points to preserve the similar distance between samples. Finally, nucMACC scores were normalized for the inherent GC bias using the locally weighted scatterplot smoothing (LOESS) algorithm. Henceforth, the loess function implemented in R was applied to estimate the influence of GC content on MNase concentration/time titrations, using the slope of the linear regression line. The function predicts a local score based on the GC content and trained values. The predicted scores were used as a normalization factor for the final nucMACC scores. The nucMACC score can be positive or negative. For mono-nucleosomes, the nucMACC score was multiplied by -1 for a more intuitive interpretation. Here, a negative score represents hypo-accessible DNA, whereas a positive score represents hyper-accessible DNA. For sub-nucleosomes, a negative score represents unstable nucleosomes, whereas a positive score characterizes nucleosomes with non-canonical structures.

NucMACC score cut-off

A universal cut-off of the nucMACC score was obtained by ranking nucleosomes based on the score and the first derivative and followed by a local regression, which predicts the slope based on the scores.

The cut-off for differentially accessible nucleosomes is when the slope = 1; this happens twice for positive and negative scores.

Spike-in analysis

Yeast reads were aligned to the sacCer3 Saccharomyces cerevisiae genome using the Bowtie2 with the following parameters --very-sensitive-local --no-discordant. Aligned reads were filtered for their mapping quality (MAPQ > 30) and fragment size (length 140 – 200 nt for mono-nucleosomes and length 50 – 139 nt for sub-nucleosomes) using DeepTools alignmentSieve. The number of reads per mono- and sub- nucleosome fraction was counted, and normalization factors were calculated by dividing the number of reads by 10e6.

Nucleosome characterization

Characterization of nucleosomes and figures used in the manuscript were done in R. Nucleosomes were assigned to genes using bitr R package and TxDb.Dmelanogaster.UCSC.dm3.ensGene annotation R package. Pathway enrichment of genes was performed with clusterProfiler R package. For expression correlation with nucleosome positioning, nucleosomes on promoter and gene body (exon and intron) regions were used. For heatmaps, LSD R package was used. For figures ChIPpeakAnno, ChIPseeker and ggplot2 R packages were used. For TF enrichment, HOMER findMotifsGenome was used, with a custom BED file containing the positions of TSS-un-stable nucleosomes and the *D. melanogaster* TF database.

Pausing index

Pausing index is defined as the ratio of the GRO-seq read density at TSS to gene body density. It was calculated as read density +/- 100 nt from TSS divided by read density +150 nt from TSS to the TES.

Minimum sequencing coverage analysis

Samples were sub-samples using SAMtools view -bs 42.X, where X represents the percentage to which the data was reduced to, and 42 as the seed number to obtain reproducible results.

Processing of RNA-seq data

Sequencing reads were aligned to the UCSC Drosophila melanogaster dm3 genome using STAR v2.7.2a with the following parameters --outFilterType BySJout --outFilterMultimapNmax 20 --alignSJoverhangMin 8 --alignSJDBoverhangMin 1 --outFilterMismatchNmax 999 --outFilterMismatchNoverReadLmax 0.04 --alignIntronMin 20 --alignIntronMax 1000000 –alignMatesGapMax. Aligned reads were filtered for their mapping quality (MAPQ > 30), using SAMtools v1.9. Reads per exon were counted using featureCounts v1.6.2 and parameters -C -p -B -t exon -g gene_id. Differential gene expression was done in R v.3.6.1, using the edgeR and limma Bioconductor packages.

REFERENCES

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

2. Van Holde,K. (1985) Physical biochemistry Prentice-Hall, Englewood Cliffs, NJ.

3. Oruba,A., Saccani,S. and van Essen,D. (2020) Role of cell-type specific nucleosome positioning in inducible activation of mammalian promoters. *Nat Commun*, **11**, 1–18.

4. Lowary,P.T. and Widom,J. (1997) Nucleosome packaging and nucleosome positioning of genomic DNA. *Proc Natl Acad Sci U S A*, **94**, 1183–1188.

5. Denisov,D.A., Shpigelman,E.S. and Trifonov,E.N. (1997) Protective nucleosome centering at splice sites as suggested by sequence-directed mapping of the nucleosomes. In *Gene*. Gene, Vol. 205, pp. 145–149.

6. Längst,G. and Manelyte,L. (2015) Chromatin remodelers: From function to dysfunction. *Genes (Basel)*, **6**, 299–324.

7. 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.

8. Fu,Y., Sinha,M., Peterson,C.L. and Weng,Z. (2008) The Insulator Binding Protein CTCF Positions 20 Nucleosomes around Its Binding Sites across the Human Genome. *PLoS Genet*, **4**, e1000138.

9. Wiechens,N., Singh,V., Gkikopoulos,T., Schofield,P., Rocha,S. and Owen-Hughes,T. (2016) The Chromatin Remodelling Enzymes SNF2H and SNF2L Position Nucleosomes adjacent to CTCF and Other Transcription Factors. *PLoS Genet*, **12**, e1005940.

10. Baldi,S., Jain,D.S., Harpprecht,L., Zabel,A., Scheibe,M., Butter,F., Straub,T. and Becker,P.B. (2018) Genome-wide Rules of Nucleosome Phasing in Drosophila. *Mol Cell*, **72**, 661-672.e4.

11. Schwartz,U., Emeth,A.N.´, Diermeier,S., Exler,J.H., Hansch,S., Maldonado,R., Heizinger,L., Merkl,R. and Angst,G.L.¨ (2018) Characterizing the nuclease accessibility of DNA in human cells to map higher order structures of chromatin. *Nucleic Acids Res*, 10.1093/nar/gky1203.

12. Brahma,S. and Henikoff,S. (2019) RSC-Associated Subnucleosomes Define MNase-Sensitive Promoters in Yeast. *Mol Cell*, 10.1016/j.molcel.2018.10.046.

13. Schep,A.N., Buenrostro,J.D., Denny,S.K., Schwartz,K., Sherlock,G., Greenleaf,W.J. and Alicia N. Schep,1 Jason D. Buenrostro,1 Sarah K. Denny,2 Katja Schwartz,1 Gavin Sherlock,1 and William J. Greenleaf1 (2015) Structured nucleosome fingerprints enable high-resolution mapping of chromatin architecture within regulatory regions. *Genome Res*, **25**, 1757–1770.

14. Kaplan,N., Moore,I.K., Fondufe-Mittendorf,Y., Gossett,A.J., Tillo,D., Field,Y., LeProust,E.M., Hughes,T.R., Lieb,J.D., Widom,J., *et al.* (2009) The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature*, **458**, 362–366.

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

16. Längst,G., Schätz,T., Langowski,J. and Grummt,I. (1997) Structural analysis of mouse rDNA: coincidence between nuclease hypersensitive sites, DNA curvature and regulatory elements in the intergenic spacer Oxford University Press.

17. 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**, 1–11.

18. Razvan,V.C., Bryson D.,T. and Henikoff,S. (2019) Quantitative MNase-seq accurately maps nucleosome occupancy levels. *Genome Biol*.

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

20. Zhou,K., Gaullier,G. and Luger,K. (2019) Nucleosome structure and dynamics are coming of age. *Nat Struct Mol Biol*, **26**, 3–13.

21. Culkin,J., de Bruin,L., Tompitak,M., Phillips,R. and Schiessel,H. (2017) The role of DNA sequence in nucleosome breathing. 10.1140/epje/i2017-11596-2.

22. North,J.A., Shimko,J.C., Javaid,S., Mooney,A.M., Shoffner,M.A., Rose,S.D., Bundschuh,R., Fishel,R., Ottesen,J.J. and Poirier,M.G. (2012) Regulation of the nucleosome unwrapping rate controls DNA accessibility. *Nucleic Acids Res*, **40**, 10215–10227.

23. Anderson,J.D., Lowary,P.T. and Widom,J. (2001) Effects of histone acetylation on the equilibrium accessibility of nucleosomal DNA target sites. *J Mol Biol*, **307**, 977–985.

24. Tsanev, R., and Petrov,P. (1976) The substructure of chromatin and its variations as revealed by electron microscopy. *J. Microsc. Biol. Cell.*

25. Lee,M.S. and Garrard,W.T. (1991) Transcription-induced nucleosome ‘splitting’: an underlying structure for DNase I sensitive chromatin. *EMBO J*, **10**, 607–615.

26. Rhee,H.S., Bataille,A.R., Zhang,L. and Pugh,B.F. (2014) Subnucleosomal structures and nucleosome asymmetry across a genome. *Cell*, **159**, 1377–1388.

27. Ishii,H., Kadonaga,J.T. and Ren,B. (2015) MPE-seq, a new method for the genome-wide analysis of chromatin structure. *Proc Natl Acad Sci U S A*, **112**, E3457-65.

28. Voong,L.N., Xi,L., Sebeson,A.C., Xiong,B., Wang,J.P. and Wang,X. (2016) Insights into Nucleosome Organization in Mouse Embryonic Stem Cells through Chemical Mapping. *Cell*, **167**, 1555-1570.e15.

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

30. 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.

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

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

33. Hasson,D., Panchenko,T., Salimian,K.J., Salman,M.U., Sekulic,N., Alonso,A., Warburton,P.E. and Black,B.E. (2013) The octamer is the major form of CENP-A nucleosomes at human centromeres. *Nat Struct Mol Biol*, **20**, 687–695.

34. Gallego,L.D., Schneider,M., Mittal,C., Romanauska,A., Gudino Carrillo,R.M., Schubert,T., Pugh,B.F. and Köhler,A. (2020) Phase separation directs ubiquitination of gene-body nucleosomes. *Nature*, **579**, 592–597.

35. Sanulli,S., Trnka,M.J., Dharmarajan,V., Tibble,R.W., Pascal,B.D., Burlingame,A.L., Griffin,P.R., Gross,J.D. and Narlikar,G.J. (2019) HP1 reshapes nucleosome core to promote phase separation of heterochromatin. *Nature*, **575**, 390–394.

36. Fuda,N.J., Guertin,M.J., Sharma,S., Danko,C.G., Martins,A.L., Siepel,A. and Lis,J.T. (2015) GAGA Factor Maintains Nucleosome-Free Regions and Has a Role in RNA Polymerase II Recruitment to Promoters. *PLoS Genet*, **11**.

37. Schwartz,U., Németh,A., Diermeier,S., Exler,J.H., Hansch,S., Maldonado,R., Heizinger,L., Merkl,R. and Längst,G. (2019) Characterizing the nuclease accessibility of DNA in human cells to map higher order structures of chromatin. *Nucleic Acids Res*, **47**, 1239.

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

39. Mueller,B., Mieczkowski,J., Kundu,S., Wang,P., Sadreyev,R., Tolstorukov,M.Y. and Kingston,R.E. (2017) Widespread changes in nucleosome accessibility without changes in nucleosome occupancy during a rapid transcriptional induction. *Genes Dev*, **31**, 451–462.

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

MAIN FIGURES

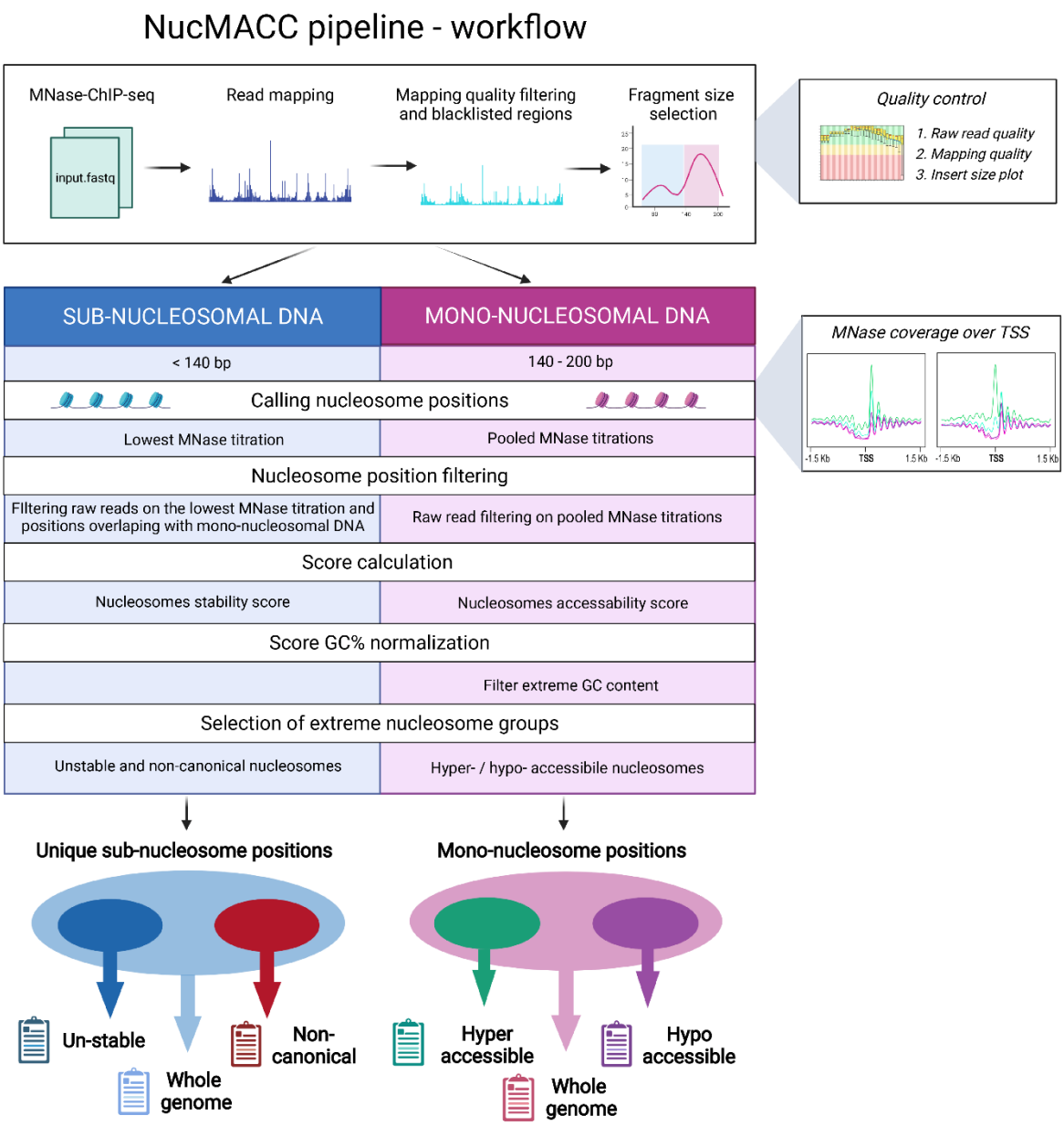

Figure 1. nucMACC pipeline workflow. The nucMACC pipeline requires raw data in .fastq format and a minimum of two MNase titrations per experimental condition. It starts with read mapping to the reference genome and quality filtering. In the next step, fragments are divided into sub-nucleosomal- and mono-nucleosomal- sized-fragments which are processed separately. Then a nucMACC score is calculated by counting fragments per each nucleosome position and each MNase titration. A linear regression slope is then fitted through MNase titrations for each nucleosome. The slope of the regression line represents the raw score, which is normalized to the underlying GC% content in the following step. The normalized score is referred to as the nucleosome MNase accessibility score (nucMACC). For nucleosome position calling all MNase titrations are pulled, whereas for sub-nucleosomes, only the lowest MNase titration is used to call sub-nucleosomal positions and is consequently filtered by previously called mono-nucleosomal positions to obtain true sub-nucleosome positions. In the final step, special nucleosome groups are obtained by analyzing where the nucMACC score considerably deviates from the mean. From the mono-nucleosome fraction we obtain stable, hypo-, and hyper- accessible nucleosomes. While the special nucleosomes from the sub-nucleosomal fraction represent un-stable and non-canonical nucleosomes.

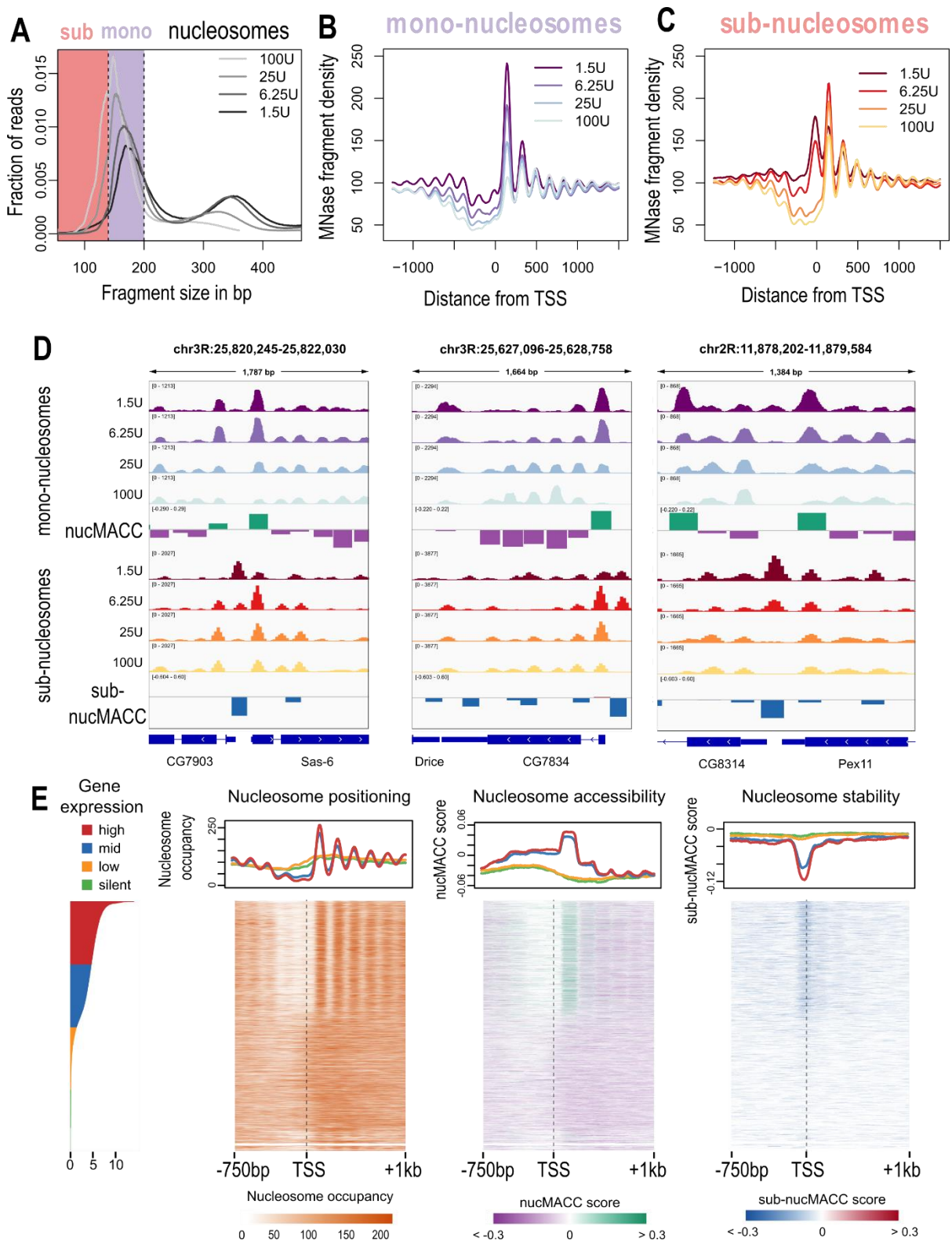

Figure 2. Characterization of the nucMACC score for ChIP-H4-MNase-seq data. (A) In silico fragment size selection based on the insert size distribution plot. In the blue square, the sub-nucleosomal fraction is indicated. In the pink square, the mono-nucleosomal fraction is indicated. (B-C) Average MNase signal over the TSS of expressed genes for mono- (B) and sub- (C) nucleosomes. (D) Examples of the output of the nucMACC pipeline. (E) Nucleosome positions (left), nucleosome accessibility (middle) and stability (right) scores over TSSs, sorted by gene expression.

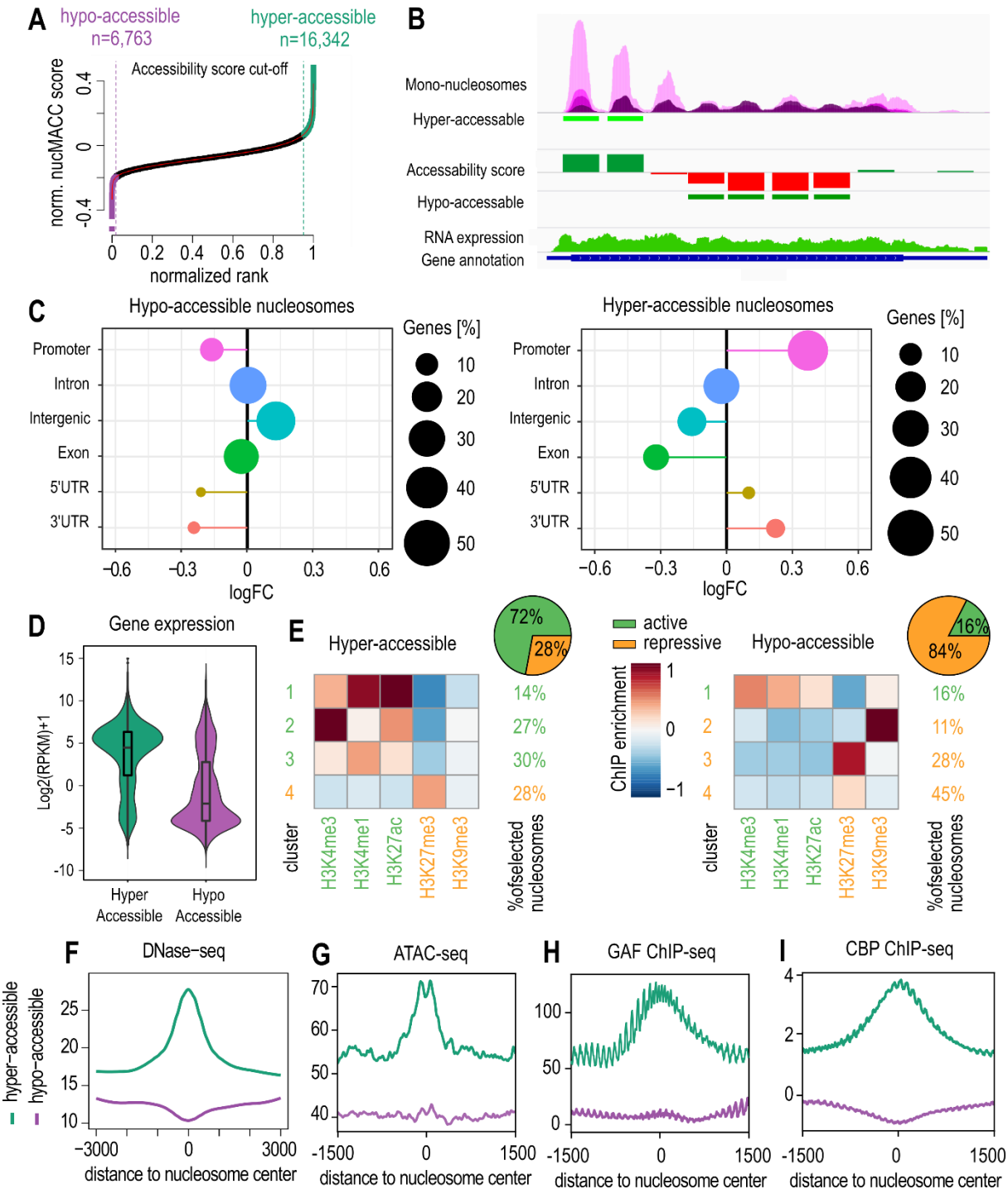

Figure 3. Characterization of the nucleosome accessibility score. (A) Nucleosome accessibility score cut-off. Hyper-accessible nucleosomes are indicated in green, whereas hypo-accessible nucleosomes are indicated in purple. (B) Example genomic region indicating a hyper-accessible +1 nucleosome and a hypo-accessible gene body. (C) Feature distribution of hypo- (left) and hyper- (right) accessible nucleosomes vs. all mono-nucleosomes. (D) Gene expression distribution for hyper- and hypo-accessible nucleosomes. (E) Enrichment of histone marks (H3K4me1, H3K27ac, H3K4me3, H3K9me3, and H3K27me3) in hyper- and hypo-accessible nucleosomes. (F) Correlation between DNase-seq data and hyper-, hypo- accessible nucleosomes. (G) Correlation between ATAC-seq data and hyper-, hypo- accessible nucleosomes. (H) Enrichment of GAF ChIP-seq data over hyper-, hypo-accessible nucleosomes. (I) Enrichment of CBP ChIP-seq data over hyper-, hypo- accessible nucleosomes.

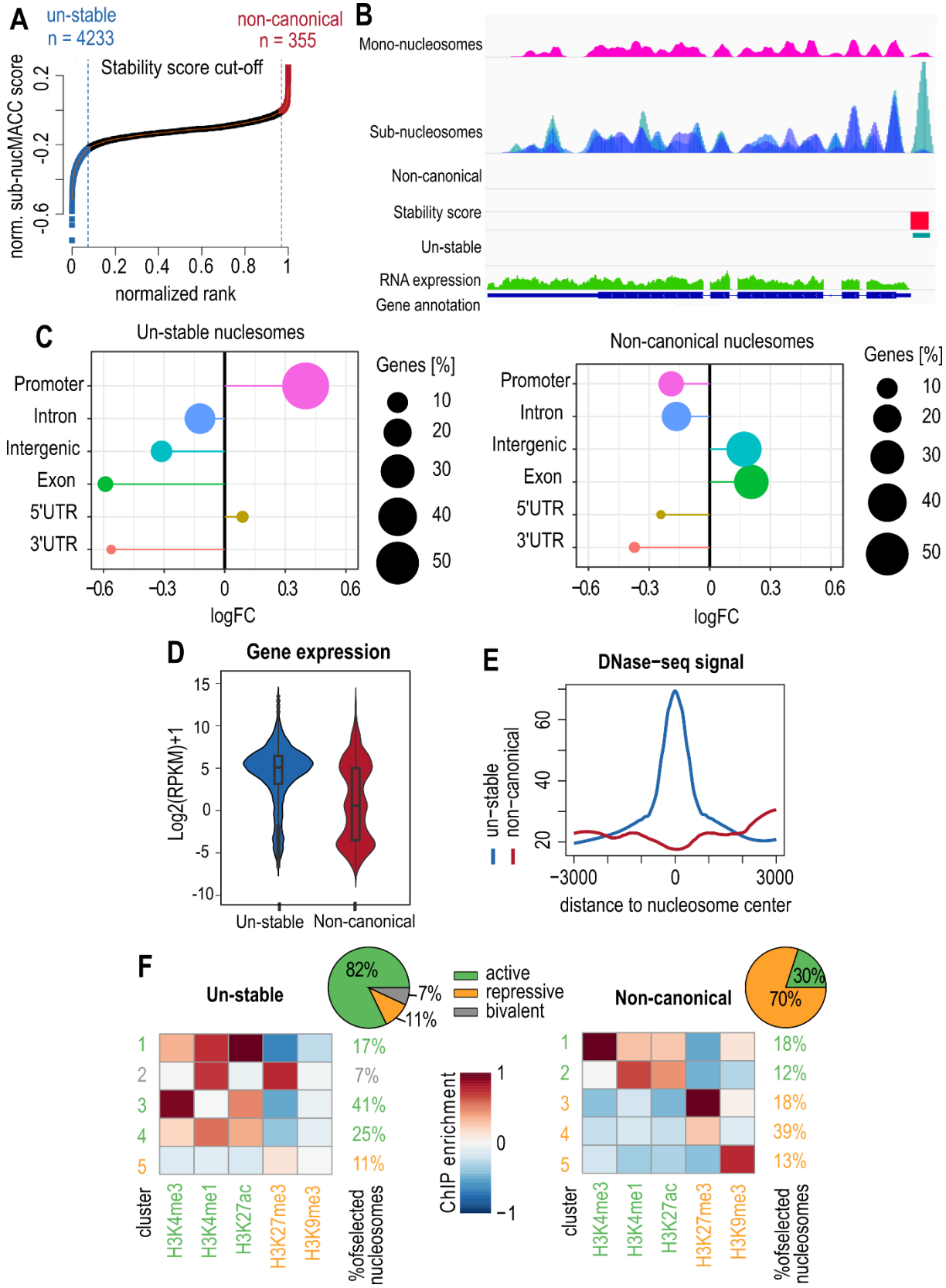

Figure 4. Characterization of the nucleosome stability score. (A) Nucleosome stability score cut-off. Un-stable nucleosomes are indicated in blue, whereas non-canonical nucleosomes are indicated in red. Stable canonical nucleosomes are indicated in black. (B) Example region indicating an un-stable nucleosome at TSS. (C) Feature distribution of un-stable and non-canonical nucleosomes over all sub-nucleosomes. (D) Gene expression distribution of un-stable and non-canonical nucleosomes. (E) Correlation between DNase-seq data and unstable, non-canonical nucleosomes. (F) Enrichment of histone modifications (H3K4me1, H3K27ac, H3K4me3, H3K9me3, and H3K27me3) in un-stable and non-canonical nucleosomes.

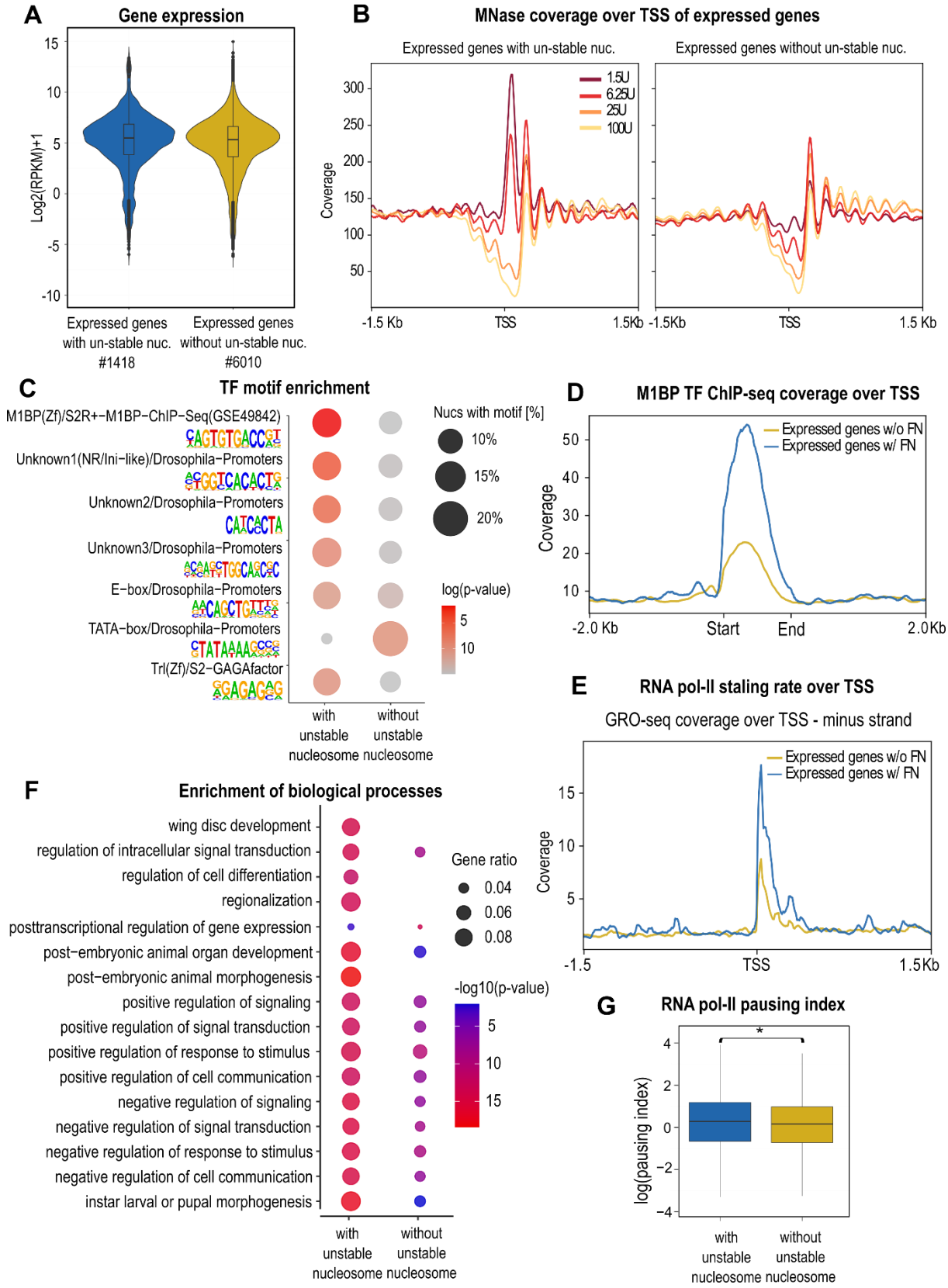

Figure 5. Characterization of un-stable nucleosomes. (A) Gene expression distribution of expressed genes with (blue) or without (gold) an un-stable nucleosome. (B) MNase coverage over TSS of expressed genes with (left) or without (right) an un-stable nucleosome. (C) TF motif enrichment of expressed genes with (left) or without (right) an un-stable nucleosome. (D) M1BP ChIP-seq seq coverage over TSS of genes with (blue) or without (gold) an un-stable nucleosome. (E) GRO-seq coverage over genes with (blue) or without (gold) an un-stable nucleosome. (F) Enrichment of biological

processes associated with genes with (left) or without (right) an un-stable nucleosome. (G) RNA pol-ii pausing index for genes with (blue) or without (gold) an un-stable nucleosome.

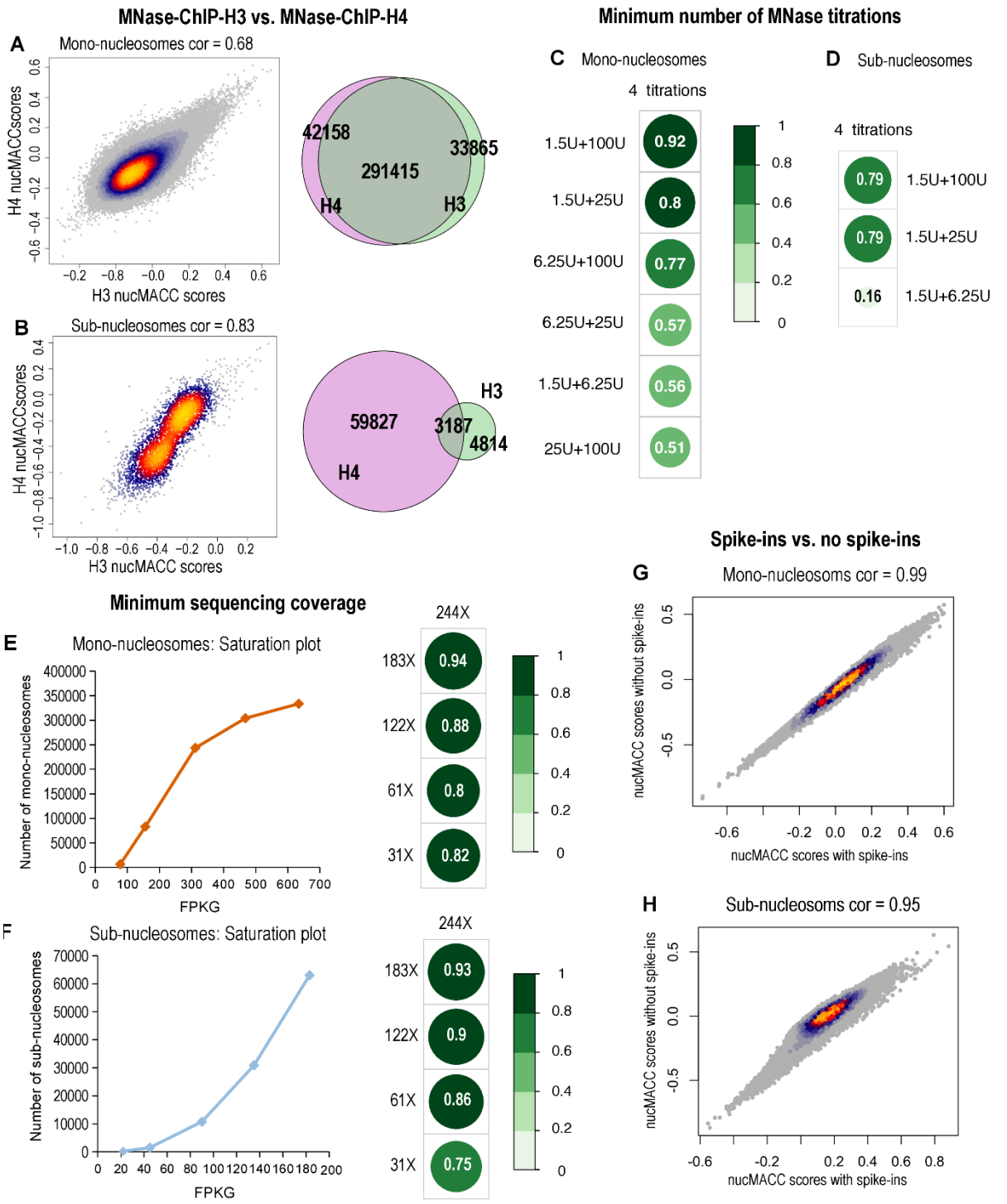

Figure 6. Evaluating the robustness of the nucMACC pipeline. (A-B) Correlation between nucMACC scores and the overlap between called nucleosome between H3 and H4 data in mono- (A) and sub- (B) nucleosomes. (C-D) Correlation between nucMACC scores for samples with two or four MNase titrations in mono- (C) and sub- (D) nucleosomes. (E-F) The number of called nucleosomes based on sequencing depth (FPKG – fragments per kilobase of the genome size) and the correlation between nucMACC scores of samples with different sequencing depths in mono- (E) and sub- (F) nucleosomes. (G-H) Correlation between nucMACC scores with or without spike-in information for mono- (G) and sub- (H) nucleosomes.