nucMACC: An optimized MNase-seq pipeline measures genome-wide nucleosome accessibility and stability

AUTHORS:

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

[1] Biochemistry III, University of Regensburg, Regensburg, Germany
[2] NGS Analysis Center Biology and Pre-clinical Medicine, University of Regensburg, Regensburg, Germany
*Co-corresponding authors

ABSTRACT:

Micrococcal nuclease sequencing (MNase-seq) is the state-of-the-art method for determining the 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 high sequencing depth requirement. Therefore, we developed an optimized analysis pipeline written in nextflow and publicly available on GitHub. The nucMACC pipeline provides high-resolution nucleosome positions and associated properties such as nucleosome accessibility and stability. Hyper-/hypo-accessible nucleosomes or non-canonical and unstable nucleosome positions in the genome are called automatically. We provide recommendations for the minimum sequencing depth and show robust results using two differential MNase conditions.

Using the nucMACC pipeline, we characterized the nucleosomal properties in D. melanogaster and explored the correlation between these properties and gene expression. Intriguingly, we found a distinct group of unstable nucleosomes enriched in promoters marked with the M1BP motif. The expression rate of the genes harboring an unstable nucleosome in the promoter was not altered compared to genes with a nucleosome-depleted promoter region, whereas the RNA polymerase pausing rate was affected.

INTRODUCTION:

Nucleosomes are the basic unit of DNA packaging in eukaryotic cells. Electron microscopy of chromatin spreads showed regularly spaced nucleosome core particles separated by linker DNA, giving the appearance of a “beads-on-string” structure (Olins, Douglas Carlson and Olins, 1975). The nucleosome core particle consists of 147 bp of DNA wrapped around a histone octamer, composed of histone proteins H2A, H2B, H3, and H4 (Kornberg, 1974; Luger *et al.*, 1997). The stability of the nucleosome core is maintained by over 360 hydrogen bonds between the DNA and histone proteins (Davey *et al.*, 2002). 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 (Valouev *et al.*, 2011; Silberhorn *et al.*, 2016; Baldi *et al.*, 2018). DNA in the linker is generally accessible to DNA-binding factors, whereas access to DNA in the nucleosome core is hindered. Nucleosome positioning, structure, and modification play crucial roles in regulating gene expression and determining the accessibility of DNA for DNA-templated processes, thereby defining cell type-specific gene activity and repression.

In vitro experiments have revealed that nucleosome positioning depends on the bendability and structure of the underlying DNA sequence and the effect of steric hindrance between neighboring nucleosomes (Kaplan *et al.*, 2009; Struhl and Segal, 2013). On the other hand, *in vivo* nucleosome positioning is more complex and is affected by additional factors, such as chromatin remodelers, histone variants, transcription factors (TF), histone posttranslational modifications, or transcription machinery (Chereji and Clark, 2018). Regulatory regions, such as promoters, enhancers, and insulators, are marked by well-positioned nucleosomes, which is a common feature of eukaryotic chromatin (Schones *et al.*, 2008; Hu *et al.*, 2011).

Besides the positioning of nucleosomes, studying their structural characteristics is central to revealing the functions of chromatin. Nucleosomes vary in their histone composition and number. Histone variants have distinct structural and functional characteristics, serve specific roles in chromatin regulation, and are produced independently of the cell cycle. The histone H3 variant CENP-A is incorporated into nucleosomes at the centromeres. DNA histone interactions are weakened in CENP-A-containing nucleosomes, partially unwrapping DNA from the histone surface (Hasson *et al.*, 2013). Hexasomes, which miss one H2A-H2B dimer, can form through various mechanisms such as transcription, histone chaperones, and chromatin remodeling complexes (Kireeva *et al.*, 2002; Prasad *et al.*, 2016; Ramachandran, Ahmad and Henikoff, 2017; Liu *et al.*, 2019). The presence of hexasomes near transcription start sites was suggested in a recent study (Rhee *et al.*, 2014), highlighting their potential role

66 in gene regulation. However, it is still unclear whether these sub-nucleosomal structures are
67 stable or short-lived nucleosome intermediates *in vivo*.
68 MNase-seq has been widely used to study chromatin structure and nucleosome positioning.
69 It involves the use of micrococcal nuclease (MNase), an enzyme that selectively cleaves linker
70 DNA and leaves the nucleosome core particles intact. The resulting DNA fragments were
71 sequenced and analyzed to determine the location and distribution of nucleosomes along the
72 genome (Baldi, Korber and Becker, 2020). MNase-seq provides a high-resolution map of the
73 nucleosome landscape and has contributed to our understanding of the dynamic and
74 regulated nature of chromatin structure (Valouev *et al.*, 2011; Teif *et al.*, 2012; Diermeier *et*
75 *al.*, 2014). Initially, high MNase digestions were applied, hydrolyzing chromatin completely to
76 the mono-nucleosomal level. Later, we and others combined different degrees of MNase
77 digestion to study chromatin accessibility and nucleosome stability (Henikoff *et al.*, 2011;
78 Chereji *et al.*, 2015; Mieczkowski *et al.*, 2016; Schwartz *et al.*, 2018; Chereji, Bryson and
79 Henikoff, 2019). As the data analysis of MNase titration experiments is complex, new metrics
80 have been introduced to provide a stable framework for accurate quantification of chromatin
81 accessibility and nucleosome occupancy (Mieczkowski *et al.*, 2016; Chereji, Bryson and
82 Henikoff, 2019). However, at least six titrations of MNase per sample and additional spike-ins
83 were required, resulting in a high workload and the required amount of material.
84 Furthermore, application to new experiments is challenging because complete analysis
85 pipelines are missing.
86 Here, we introduce a comprehensive and user-friendly nucleosome-based MNase
87 accessibility (nucMACC) analysis pipeline. The nucMACC workflow starts directly with raw
88 sequencing data, assesses the critical quality metrics relevant to MNase-seq experiments, and
89 summarizes the results into user-friendly reports. The workflow reveals the genome-wide
90 nucleosome positions and their corresponding nucleosome accessibility (nucMACC) scores.
91 Hyper- and hypo-accessible nucleosome positions with exceptionally high or low nucMACC
92 scores are determined. Additionally, as a unique feature of the nucMACC pipeline,
93 nucleosome stability scores are inferred from sub-nucleosomal fragments uncovering un-
94 stable or non-canonical nucleosomes (Figure S1). We show that the pipeline obtains robust
95 results using only two MNase conditions per sample. The nucMACC pipeline is written in
96 nextflow, available at GitHub (https://github.com/uschwartz/nucMACC), and integrates the
97 Docker software container, making it easy to use, portable, reproducible, and scalable.

RESULTS:

nucMACC pipeline overview

The nucMACC pipeline requires at least two MNase-seq titrations as input, preferentially in combination with a histone immunoprecipitation step to exclude non-histone protein-mediated DNA-binding events. 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 (Figure 1A). Under mild MNase conditions, nucleosomes from accessible and active genomic sites are preferentially released, referred to as hyper-accessible nucleosomes (Figure 1A, green nucleosome) (Mieczkowski *et al.*, 2016; Schwartz *et al.*, 2018). Increasing the MNase concentration or digestion duration continuously releases more nucleosome core particles from the chromatin until the chromatin is fully hydrolyzed. Additionally, MNase is a processive enzyme. Once the nucleosome is released from chromatin, MNase progressively trims the linker DNA and hydrolyzes it within the realm of the nucleosomal DNA, resulting in sub-nucleosomal fragments. Canonical nucleosome core particles are resistant to internal MNase cleavage and remain mainly complete even under extensive chromatin digestion conditions. However, it has been suggested that a special group of nucleosomes, often referred to as un-stable or fragile nucleosomes, exists, which are MNase-sensitive and start to be degraded already under mild MNase conditions to sub-nucleosome-sized fragments (Figure 1A, blue nucleosome) (Weiner *et al.*, 2010; Kubik *et al.*, 2015; Voong *et al.*, 2016). We refer to such nucleosomes as un-stable nucleosomes in the remainder of this manuscript. Under extensive MNase conditions, un-stable nucleosomes are completely degraded and not any more detectable. Therefore, sub-nucleosomal fragments in MNase titration experiments can be used to measure nucleosome stability, as reflected by resistance against internal MNase cleavage.

Starting from raw sequencing files in the FASTQ format, the nucMACC pipeline first assesses the sequencing quality. Next, the reads are aligned to the provided reference genome and subsequently mapped reads are filtered based on mapping quality (Figure S2). Optionally

131 ambiguous genomic elements, such as blacklisted regions or mitochondrial chromosomes,
132 can be removed from the analysis. To obtain high-resolution nucleosome positions, fragments
133 between 140 and 200 bp in length, the typical size of mono-nucleosomal DNA, were selected
134 (Figure 1B). Size-selected fragments of all MNase titrations are pooled and used to derive a
135 comprehensive map of the genome-wide nucleosome positions.
136 Recently, a metric for measuring chromatin accessibility, termed MACC (MNase accessibility),
137 was introduced (Mieczkowski et al., 2016). As the MACC scores provide an elegant way to
138 measure chromatin accessibility based on MNase-seq data, we adapted the MACC scoring
139 system to increase resolution and accuracy. Instead of counting the MNase reads in arbitrary
140 genomic bins as in the original MACC version, we quantify the MNase accessibility directly at
141 the defined nucleosome positions, which results in a specific accessibility score for each
142 nucleosome (nucMACC). Next, linear regression is conducted on the normalized fragment
143 frequencies for each nucleosome (Figure S2, see Methods section for details). The slope of
144 the regression line multiplied by minus one is used as the nucMACC score. Since MNase
145 exhibits an increasing AT preference with higher concentrations, raw nucMACC scores are
146 normalized for the underlying GC content using a LOESS regression.
147 As an additional feature of the nucMACC pipeline fragments shorter than 140 bp, which
148 represent sub-nucleosomal DNA fragments, are used to measure nucleosome stability.
149 Here, we use only the lowest MNase concentration to call nucleosome positions in the
150 following called sub-nucleosomes, as in the higher MNase titrations un-stable nucleosomes
151 are completely hydrolyzed. Analogous to the nucMACC scores, sub-nucleosomal fragment
152 frequencies for each MNase titration are quantified at the sub-nucleosome positions. The
153 slope of the regression is used to measure nucleosome stability and normalized to the
154 underlying GC content. The normalized score is referred to as the sub-nucleosome MNase
155 accessibility score (sub-nucMACC).
156
157 **The nucMACC pipeline scores nucleosome accessibility and stability**
158
159 To establish the nucMACC pipeline, we used mainly the MNase chromatin digestions coupled
160 with H4 immunoprecipitation from the original MACC publication (Mieczkowski *et al.*, 2016).
161 The dataset collection contains four different MNase concentrations (1.5U, 6.25U, 25U, and
162 100U) done in two replicates each. The fragment size distribution is an important quality

163 control step as it clearly shows the degree of chromatin digestion in each sample (Figure 1B).
164 With increasing MNase concentrations the di-nucleosome fraction is disappearing, and the
165 levels of mono-nucleosomes are increasing. The main mono-nucleosomal peak is in the range
166 of 140–200 bp, while the peak summit is shifted with higher MNase concentrations closer to
167 147 bp, the length of DNA in a nucleosome core particle. Fragment sizes shorter than 140 bp
168 indicate sub-nucleosomal DNA fragments, which mainly arise from DNA cleavage within the
169 core nucleosome. Sub-nucleosomal DNA fragments are barely present in mild MNase
170 conditions but increase in extensive MNase samples. Based on the curve shape of the
171 fragment size distribution, we selected the fragments in the range of 140 – 200 bp in length
172 for mono-nucleosome analysis and shorter than 140 bp for sub-nucleosome analysis,
173 respectively (Figure 1B).
174 Plotting the fragment frequencies of mono- and sub-nucleosomes at the TSS shows that the
175 selected categories contain different information (Figure 1C and 1E). Looking at the mono-
176 nucleosomes of the two extensive MNase conditions (25U and 100U) reveals a well-known
177 nucleosome positioning pattern at the TSS: a nucleosome depleted region (NDR) followed
178 downstream by a phased nucleosome array, which starts with a well-positioned +1-
179 nucleosome and propagates into the gene. The mild MNase samples (1.5U and 6.25U) reveal
180 the same mono-nucleosome positions, however, distinct positions, such as the +1-
181 nucleosome, are markedly enriched compared to the extensive MNase samples indicating
182 that nucleosomes exhibit different release rates from chromatin. The nucMACC score
183 quantifies chromatin accessibility as mono-nucleosome fragment frequencies change with
184 MNase titrations (Figure 1D and S2).
185 Comparing the mono-nucleosome and the sub-nucleosome TSS profile the extensive MNase
186 samples are very similar showing that once the nucleosomes are released from chromatin,
187 MNase trims the DNA within the nucleosome core. Nevertheless, an additional dominant
188 peak is appearing at the denominated (NDR) in the sub-nucleosomal fraction of the mild
189 MNase condition, indicating that this position is not completely free of histones, but occupied
190 by MNase sensitive nucleosomes, which are completely degraded in extensive MNase
191 digestions (Figure 1E). The sub-nucMACC score quantifies nucleosome stability as changes in
192 sub-nucleosome fragment frequencies over MNase titrations (Figure 1F and S2).
193 Even though the way how nucMACC and sub-nucMACC scores are calculated is very similar,
194 their scores are not correlated at overlapping positions, confirming that both metrics show

195 distinct features of nucleosomes (Figure S3A). The correlation of the nucMACC scores with
196 the original MACC scores (Mieczkowski *et al.*, 2016) was modest (R=0.48), whereas the sub-
197 nucMACC scores showed no correlation (R=-0.24) indicating that our pipeline uncovers new
198 features in MNase titration experiments (Figure S3B-C).
199 Centering the nucMACC scores on the TSS shows that the +1 nucleosomes of actively
200 transcribed genes are in general hyper-accessible, whereas nucleosomes further downstream
201 in the gene body did not significantly change nucleosome accessibility (Figure 1G).
202 Remarkably, in genic regions, we observed no difference in accessibility between silent and
203 highly expressed genes, except at the +1 nucleosome (Figure S3D), confirming that chromatin
204 accessibility is not changed at a global scale but modulated at defined local sites, such as the
205 promoter, to regulate gene transcription (Schwartz *et al.*, 2018).
206 The nucleosome stability profile at the TSS of active genes was different compared to the
207 nucleosome accessibility score. The lowest nucleosome stability scores were obtained
208 upstream of the +1 nucleosome directly at the TSS of active genes. The sub-nucleosome
209 analysis uncovered un-stable nucleosomes directly at the denominated NDRs showing overall
210 low nucleosome occupancy in mono-nucleosome analysis (Figure 1G). In contrast, the
211 transcription end sites (TES) are nucleosome depleted in Drosophila, since we do not observe
212 un-stable nucleosomes there as recently proposed in yeast and mouse (Figure S3E) (Chereji,
213 Ocampo and Clark, 2016; Voong *et al.*, 2016).
214 Overall, the nucleosome features provided by the nucMACC pipeline are robust and
215 reproducible, as shown by a comparison with MNase titrations combined with H3-ChIP data
216 (Figure S4A-G). We observed the highest variation between H3 and H4 experiments in the
217 number and location of the sub-nucleosome positions (Figure S4C). This can be explained, as
218 the number of sub-nucleosomal fragments in the H3 mild MNase samples was significantly
219 lower compared to the H4 dataset revealing fewer sub-nucMACC positions (addressed in
220 more detail later). Nevertheless, the sub-nucMACC scores at overlapping positions showed a
221 high correlation (Figure S4C, R=0.8).
222 In summary, splitting up MNase titration data by fragment sizes into mono-nucleosomes and
223 sub-nucleosomes enables the nucMACC pipeline to reveal nucleosome positions, nucleosome
224 accessibility, and nucleosome stability scores within the same experiment.
225 Furthermore, we tested whether the additional histone immunoprecipitation step after
226 chromatin digestion is required to obtain reliable nucleosome features. Mono- and sub-

227 nucleosome positioning changed and overlapping positions showed only a modest correlation
228 with whole chromatin extractions (Fig S4H-K). Changes in the fragment profile were clearly
229 visible upstream of the TSS and most pronounced at the -1 nucleosome position affecting
230 nucMACC and sub-nucMACC scores (Figure S4L-N). The -1 nucleosome position exhibited the
231 highest accessibility in whole chromatin extractions and the sub-nucleosomal fraction in the
232 mild MNase sample showed a broad and pronounced enrichment at the promoter of
233 expressed genes reminiscent of other methods to measure chromatin accessibility such as
234 DNase- or ATAC-seq (Buenrostro *et al.*, 2013; Vierstra *et al.*, 2014). The differences between
235 whole chromatin and histone immunoprecipitated MNase datasets can be addressed by non-
236 histone proteins bound to DNA (Mieczkowski *et al.*, 2016). To obtain accessibility and stability
237 measurements specific to nucleosomes, an additional histone immunoprecipitation step is
238 necessary. Nevertheless, the nucMACC pipeline can be run at whole chromatin extractions
239 and measures overall chromatin accessibility and stability over all molecules bound to DNA.
240
241 **Characterization of nucleosome accessibility scores**
242 To identify functionally relevant nucleosomes, the nucMACC pipeline extracts nucleosomes
243 with exceptionally high or low accessibility, termed hyper- or hypo-accessible nucleosomes.
244 Therefore, the nucMACC scores of each nucleosome are plotted against their ranks. Ranks are
245 normalized by the total number of nucleosomes and the MACC score by the total range of
246 scores. In that way, both the x and y axes have a value range of 1 (Figure 2A). To geometrically
247 identify the positions, where the signal decreases/increases rapidly, we determined the
248 positions at the curve where the slope is the first/last time below one. Hypo-accessible
249 nucleosomes are defined as the nucleosome positions before the slope of the curve decreases
250 below 1 (Figure 2A, purple dots), and hyper-accessible nucleosomes after the slope increases
251 again above 1, respectively (Figure 2A, green dots). A similar strategy is applied to define
252 super-enhancer (Whyte *et al.*, 2013).
253 In total, we identified 6,763 (2 %) hypo-accessible and 16,342 (4,9 %) hyper-accessible
254 nucleosomes in the Drosophila S2 H4-ChIP data set. Hyper-accessible nucleosomes are
255 strongly enriched at DNase hypersensitive sites, whereas hypo-accessible nucleosomes are
256 depleted, confirming the nucMACC score as a measure for accessibility (Figure 2B). Genomic
257 feature characterization of hyper- and hypo-accessible nucleosomes revealed an enrichment
258 at the promoter and intergenic regions, respectively (Figure 2C). Hyper-accessible

259 nucleosomes are preferentially associated with H3K4me1, H3K4me3, and H3K27ac histone
260 modifications marking active TSS and enhancer, whereas hypo-accessible nucleosomes are
261 associated with H3K9me3 and H3K27me3 modifications indicative for heterochromatic
262 regions (Figure 2D). In agreement with the histone modification analysis, hyper-accessible
263 nucleosomes occupy mainly expressed genes, while hypo-accessible nucleosomes are present
264 on silent genes (Figure 2E). Nevertheless, we found a substantial fraction of hypo-accessible
265 nucleosomes at the exons of expressed genes potentially regulating differential exon usage.
266 The nucMACC pipeline identifies functional nucleosome positions in the genome regulating
267 DNA accessibility and therefore genome activity.
268
269 **Characterization of nucleosome stability scores**
270 To measure nucleosome stability, sub-nucleosome positions were derived from the sub-
271 nucleosome fragments in the lowest MNase condition. Hyper-accessible nucleosomes are
272 expected to have a higher MNase degradation rate as they are less protected once they are
273 released from chromatin. Therefore, sub-nucleosome positions are selected, which either
274 contained at least four-fold more normalized sub-nucleosome fragment counts than the
275 mono-nucleosome fraction or did not overlap with any mono-nucleosome position.
276 Similar to the selection of hyper- and hypo-accessible nucleosomes, nucleosomes exhibiting
277 extreme stability scores (sub-nucMACC) were determined (Figure 3A). This selection strategy
278 resulted in 4,233 (6.7 %) nucleosomes showing very low stability scores and therefore termed
279 un-stable nucleosomes. Additionally, a rather small subset of nucleosomes (n=355, 0.6 %)
280 were enriched in the sub-nucleosome fraction of the lowest MNase condition and remained
281 stable at higher MNase concentrations. This group represents most likely structurally altered
282 nucleosomes with additional MNase cleavage sites in the realm of the nucleosome core or
283 non-octamer histone compositions, such as hexa- or hemisomes (Dalal *et al.*, 2007; Hasson *et*
284 *al.*, 2013; Rhee *et al.*, 2014). Nucleosomes enriched in the low sub-nucleosomal fractions and
285 showing extraordinarily high stability scores are referred here to as non-canonical
286 nucleosomes.
287 In contrast to non-canonical nucleosomes, un-stable nucleosomes positions are characterized
288 by high chromatin accessibility (Figure 3B) and are mainly located at gene promoter (53,8 %),
289 whereas non-canonical nucleosomes are enriched in intergenic and exonic regions (Figure
290 3C). Preferentially, active histone marks are associated with un-stable nucleosomes (Figure

291 3D) and accordingly, un-stable nucleosomes are present at expressed genes (Figure 3E).
292 Interestingly, we identify a subgroup of un-stable nucleosomes associated with active
293 H3K4me1 and repressive H3K27me3 marking bivalent enhancers. Conversely, non-canonical
294 nucleosomes showed enrichment in repressive H3K27me3 and H3K4me1 (Figure 3D).
295 Nevertheless, 30% of the non-canonical nucleosomes are associated with active histone
296 modifications, and therefore non-canonical nucleosomes are located at both expressed and
297 silent genes (Figure 3E).
298
299 **Un-stable nucleosomes at the promoter and their function in the regulation of gene**
300 **expression**
301
302 As we observed that un-stable nucleosomes are often occurring directly upstream of the TSS
303 (Figure 1G and Figure 3C) of expressed genes, we wanted to investigate their role in gene
304 regulation. Thus, we divided all expressed genes into two groups: 1) those which contained
305 an un-stable nucleosome 200 bp upstream of the TSS (n=1,588) and 2) those which had
306 instead a nucleosome depleted region (NDR) in the promoter (n=5,363) (Figure 4A). To define
307 NDR or un-stable nucleosome promoter it is absolutely required to use MNase chromatin
308 digestions coupled with histone immunoprecipitations, as non-histone proteins will dominate
309 the signal at the promoter in whole chromatin extractions (Figure S5). Overall, the expression
310 rate of the corresponding genes was similar in both groups and only slightly elevated for genes
311 carrying an un-stable nucleosome (Figure 4B, p-value = $2x10^{-15}$). GO enrichment analysis
312 revealed that genes containing un-stable nucleosome in their promoter are stimulus or
313 developmentally regulated (Figure 4C). Furthermore, unstable nucleosome promoter were
314 enriched for Drosophila promoter motifs Unknown1-3 and M1BP motif, but depleted in
315 Initiator, TATA-box, and Unkown4 motif (Figure 4D). As M1BP is a known factor to regulate
316 RNA Pol II pausing depending on nucleosome positioning (Li and Gilmour, 2013), which
317 provides an efficient way to induce gene transcription rapidly or synchronously in response
318 to developmental and environmental signals, we tested whether Pol II pausing differed
319 between NDR or unstable nucleosome promoter using nascent RNA sequencing data (Fuda *et*
320 *al.*, 2015). In agreement with M1BP motif enrichment, unstable nucleosome promoter show
321 a pronounced association with paused Pol II directly downstream of the TSS and higher

322 pausing index rates compared to NDR promoter enabling a fast transition into productive
323 transcription at responsive and developmental genes. (Figure 4E-D).
324
325 **M1BP binds to un-stable nucleosomes**
326 We next assessed the relationship between the binding of the transcription pausing factor
327 M1BP and nucleosome properties. Here, we investigated nucleosome positioning,
328 accessibility, and stability at annotated M1BP binding sites (Zouaz *et al.*, 2017). We found that
329 M1BP binding sites are surrounded by hyper-accessible nucleosomes (Figure 5A-C). Directly
330 in the binding center of M1BP, we observed a clear enrichment of un-stable nucleosomes
331 (Figure 5D-F) suggesting that M1BP does not bind to nucleosome-free DNA but rather to DNA
332 still co-occupied with histones. Nevertheless, histone DNA interactions seemed to be
333 weakened upon M1BP binding making them sensitive to internal MNase digestions.
334 **Evaluating the robustness of the nucMACC pipeline**
335 Previously four or more MNase digestion conditions have been used to determine chromatin
336 accessibility scores (Mieczkowski *et al.*, 2016; Chereji, Bryson and Henikoff, 2019). We sought
337 to determine the minimum number of titration points per sample to obtain reliable and
338 robust results with our pipeline. We compared the nucMACC and sub-nucMACC scores of two
339 or four titration points of the same experimental dataset. We observed that the results of
340 only two different MNase conditions are highly correlated with 4 titration points, as long as
341 the difference in the MNase concentration pairs is at least 10-fold (Figure 6A and B; Figure
342 S6A an B). In addition, we observed only consistent results using the lowest available MNase
343 concentration to detect un-stable nucleosomes using the sub-nucleosomal fragments. At
344 higher concentrations, fragments of unstable nucleosomes are often lost due to their fragile
345 nature.
346 Next, we addressed the importance of sequencing depth on the results of the nucMACC
347 pipeline, as MNase-seq analysis requires a high coverage of the genome. To obtain a genome
348 independent quantification we normalized the sequencing depth to the number of fragments
349 per kilobase of the mappable genome (FPKG):

350 $$Fragments\ per\ kilobase\ of genome\ (FPKG) = \frac{number\ of\ sequenced\ fragments}{\frac{effective\ genome\ size}{1{,}000}}$$

351 We found that using half of the fragments, with coverage of 122 FPKG preserves 73% of the
352 mono-nucleosome positions, and the nucMACC score correlation is reasonably high (R=0.88)
353 (Figure 6C and Figure S6C). Whereas further lowering the sequencing depth to 61 FPKG,
354 results in a considerable loss of called mono-nucleosome positions (25%), albeit the nucMACC
355 score correlation remains stable at overlapping nucleosome positions, with a correlation of
356 0.80. For sub-nucleosomes, starting with 244 FPKG results already at low coverage, since the
357 sub-nucleosomal fragments are underrepresented in mild MNase digestions. Reducing the
358 coverage to 183 FPKG (75% of sequencing fragments) amounts to a decrease in called sub-
359 nucleosomes to 49% (Figure 6D and Figure S6D). Nonetheless, the correlation between sub-
360 nucMACC scores remains very high at overlapping positions. Below 183 FPKG coverage, the
361 sub-nucleosome analysis produces an insufficient amount of called sub-nucleosome positions
362 (Figure 6D).
363 In summary, we recommend a minimum sequencing depth of 180 FPKG or higher to be used
364 when analyzing nucleosome stability.
365 Ultimately, we tested the reliability and robustness of nucMACC and sub-nucMACC scores
366 and tested if the differential amount of nucleosomes released at various MNase titrations
367 would bias the scores. Therefore, we ran our pipeline on a published dataset with spiked-in
368 nucleosomes and compared the results with or without spike-in information. Remarkably, we
369 find almost no difference in nucMACC scores when comparing the pipeline with or without
370 spike-in information for mono- (Figure 6E) and sub-nucleosomes (Figure 6F).
371 In conclusion, the nucMACC pipeline doesn't require spike-ins and is robust and reproducible
372 with only two MNase titrations.
373

374 # Discussion:

375
376 In the current study, we set out to create an optimized MNase-seq pipeline and characterize
377 nucleosomes based on their accessibility and stability. We provide a unique way to select
378 nucleosome positions showing extraordinary accessibility and stability properties. Hyper-
379 accessible nucleosomes were enriched in promoter regions, whereas hypo-accessible
380 nucleosomes showed enrichment in intergenic regions. We conclude most nucleosomes have
381 similar levels of accessibility, with local effects of hyper- and hypo- accessibility, which has
382 been also previously observed in human cell lines (Schwartz *et al.*, 2018).

383 The nucMACC pipeline provides the first systematic and genome wide approach to detect un-
384 stable nucleosomes. We find un-stable nucleosomes enriched on active promoter of
385 expressed genes, while non-canonical nucleosomes are enriched at intergenic and exonic
386 regions of both silent and expressed genes. Furthermore, we show that some non-canonical
387 nucleosomes exhibit shorter DNA length and are resistant to high MNase digestions.
388 However, we cannot discriminate between nucleosome structures, such as hemisomes,
389 hexasomes or lexasomes. Intriguingly we find a special group of un-stable nucleosomes
390 enriched at TSS of promoters marked with M1BP TF-specific motif (Li and Gilmour, 2013).
391 Undeniably, we see a high enrichment of un-stable nucleosomes at M1BP binding sites. This
392 made us speculate that un-stable nucleosomes are involved in RNA pol-II pausing, which we
393 confirmed with GRO-seq data and we detected a clear difference in RNA pol-II pausing rate
394 compared to nucleosome depleted promoter. We hypothesize that un-stable nucleosomes
395 may stimulate binding properties of specific TF, despite nucleosomes normally occluding
396 binding. This may enable a rapid response to stimuli and developmentally regulated gene
397 expression. The un-stable nucleosome represents a gene regulation facet, adding to the
398 complexity of eukaryotic gene regulation.
399 Moreover, we provide a robust pipeline and guidelines for the experimental setup of the
400 MNase-seq protocol by analyzing different datasets with two or four MNase titrations per
401 sample, with or without spike-in information, different sequencing depths, and the
402 correlation between replicates, and correlation between different anti-histone antibody
403 pulldowns. In short, we show that only two MNase titrations are required, albeit at a high
404 sequencing depth. A minimum sequencing depth of 122 FPKG per sample for mono- and 244
405 FPKG for sub-nucleosomes is recommended to attain high-quality results. Of course, this
406 depends on the experimental design. For example, if mono- and sub-nucleosome fractions
407 are specifically enriched, such as through gel excision, the sequencing depth required per
408 sample will be lower.
409 For nucleosome analysis, applying an anti-histone ChIP after MNase treatment is crucial to
410 differentiate between histone and non-histone proteins bound to DNA, which cannot be
411 guaranteed just by fragment size selection.
412 We also analyzed the nucMACC scores with or without spike-in information. Spike-ins are
413 typically used to compare antibody pulldowns with different efficiencies or to make an
414 absolute quantification of RNA expression or, in our case, nucleosome occupancy. We see no

415 improvement or difference in nucMACC scores when using spike-in information. In
416 conclusion, for determining absolute nucleosome occupancy, spike-ins are still compulsory
417 and should be included in the experimental design, whereas using relative differences in the
418 MNase fragment frequencies is sufficient to calculate the nucleosome accessibility and
419 stability scores. Sequencing depth seems to have a more detrimental effect on the accuracy
420 of nucMACC scores than the absence of spike-in information. In a recent publication, authors
421 suggest spike-ins are necessary to accurately estimate the nucleosome digestion rate and
422 subsequently accessibility scores (Chereji, Bryson and Henikoff, 2019). Unfortunately, we
423 could not directly compare their scores and ours in the same dataset since they did not
424 provide analyzed data tracks. Furthermore, we could not reproduce their results with the
425 available information about the workflow they used, and an automated pipeline was not
426 provided.
427 In summary, we provide a publicly available Nextflow pipeline for analyzing nucleosome
428 positions and nucleosomal properties using MNase titrations. We provide guidelines for the
429 experimental design of MNase-seq experiments. Lastly, using our pipeline on public MNase-
430 seq datasets in D. melanogaster we find a novel mechanism of gene regulation in eukaryotes,
431 via nucleosome stability modulation and RNA pol-II pausing rate.
432

433 # METHODS:

434

435 **Public datasets used in this study**

436 Whole chromatin MNase-seq, MNase-ChIP-H3-seq, MNase-ChIP-H4-seq and RNA-seq
437 (GSE78984), MNase-ChIP-H2B-seq, MNase-ChIP-H3-seq (GSE69336), M1BP-ChIP-seq
438 (SRX3011239), GRO-seq (GSM3274626), and MNase-seq with spike-ins in S2 and Kc167 cell
439 lines (PRJNA528497) were downloaded from the European nucleotide archive (ENA)
440 database. Histone modification and TF ChIP-seq or ChIP-seq data were retrieved through
441 modENCODE, with the following accession numbers: H3K27ac (296), H3K27me3 (298),
442 H3K4me1 (304), H3K9me3 (313), RNA pol-II ChIP-seq (329), H3K36me3 (3189), H3K4me3
443 (3761), CBP TF ChIP-seq (858), GAF TF ChIP-seq (858). DNAse-seq (GSM5258764) data were
444 retrieved from the GEO database.
445

446 **Processing of MNase-seq data**

447 Raw sequencing reads were mapped to the UCSC Drosophila melanogaster dm3 genome,
448 using Bowtie2 with the following parameters --very-sensitive-local --no-discordant. Aligned
449 reads were filtered for their mapping quality (MAPQ > 30), fragment size (length 140 – 200
450 bp for mono-nucleosomes and length < 140 bp for sub-nucleosomes), and ambiguous
451 chromosomes and blacklisted regions were removed using DeepTools2 alignmentSieve. Often
452 there is a shift in the fragment sizes due to variations in MNase enzyme activity; thus, each
453 dataset should be assessed individually. To achieve higher sequencing depth, mono-
454 nucleosome positions were called on pooled MNase titrations, using DANPOS, with dpos -m
455 1 --extend 70 -c 162367812 -u 0 -z 1 -a 1 -e 1, parameters. Sub-nucleosome positions were
456 called using only the lowest digestion samples, with the following parameters: dpos -m 1 - -
457 extend 70 -z 70 -c 162367812 -u 0 -e 1. Here library size was normalized to effective dm3
458 genome size (162367812), paired-end reads centered to their midpoint and extended to 70
459 bp to obtain nucleosome-sized positions. Sub-nucleosome positions were further filtered by
460 two parameters: a non-overlapping position with a mono-nucleosome or an overlapping
461 position with an enriched signal over a mono-nucleosome signal. Here a cut-off of > 4x higher
462 signal in sub-nucleosomal data as opposed to mono-nucleosomal data was chosen.
463
464 **NucMACC score calculation**
465 Reads per sample per nucleosome position were counted using featureCounts, using a
466 custom SAF file with nucleosome positions and the following parameters -SAF -p -B -C -
467 largestOverlap. Information on the GC content of each nucleosome was added to the read
468 count file using BEDtools nuc. Nucleosome positions were obtained by converting the
469 DANPOS annotation file to SAF format using AWK. NucMACC scores were calculated using R
470 and DESeq2 R package. Firstly, raw fragments were filtered for sequencing coverage, with cut-
471 off > 30 fragments per mono-nucleosome and > 5 fragments for sub-nucleosomes. Then,
472 fragments were normalized based on the library size (counts per million, CPM) or to the
473 calculated library size based on normalization factors for spike-in analysis. Linear regression
474 was calculated using normalized fragment counts and log2 MNase concentrations. Finally,
475 nucMACC scores were normalized for the inherent GC bias using the locally weighted
476 scatterplot smoothing (LOESS) algorithm. Henceforth, the loess function implemented in R
477 was applied to estimate the influence of GC content on MNase concentration titrations, using
478 the slope of the linear regression line. The function predicts a local score based on the GC

479 content and trained values. The predicted scores were used as a normalization factor for the
480 final nucMACC scores. The nucMACC score can be positive or negative. For mono-
481 nucleosomes, the nucMACC score was multiplied by -1 for a more intuitive interpretation.
482 Here, a negative score represents hypo-accessible DNA, whereas a positive score represents
483 hyper-accessible DNA. For sub-nucleosomes, a negative score represents unstable
484 nucleosomes, whereas a positive score characterizes nucleosomes with non-canonical
485 structures.
486
487 **NucMACC score cut-off**
488 A universal cut-off of the nucMACC score was obtained by ranking nucleosomes based on the
489 score and the first derivative and followed by a local regression, which predicts the slope
490 based on the scores. The cut-off for differentially accessible nucleosomes is when the slope =
491 1; this happens twice for positive and negative scores.
492
493 **Spike-in analysis**
494 Yeast reads were aligned to the sacCer3 Saccharomyces cerevisiae genome and treated
495 analogously to the nucMACC pipeline. The number of reads per mono- and sub-nucleosome
496 fraction was counted, and normalization factors were calculated by dividing the number of
497 reads by 10e6.
498
499 **Nucleosome characterization**
500 Characterization of nucleosomes and figures used in the manuscript were done in R.
501 Nucleosomes were assigned to genes using bitr R package and
502 TxDb.Dmelanogaster.UCSC.dm3.ensGene annotation R package. Pathway enrichment of
503 genes was performed with clusterProfiler R package. For expression correlation with
504 nucleosome positioning, nucleosomes on promoter and gene body (exon and intron) regions
505 were used. For heatmaps, LSD R package was used. For figures ChIPpeakAnno, ChIPseeker
506 and ggplot2 R packages were used. For TF enrichment, HOMER findMotifsGenome was used,
507 with a custom BED file containing the positions of TSS-un-stable nucleosomes and the *D.*
508 *melanogaster* TF database.
509
510 **Pausing index**

511 Pausing index is defined as the ratio of the GRO-seq read density at TSS to gene body density.
512 It was calculated as read density +/-100 nt from TSS divided by read density +150 nt from TSS
513 to the TES.
514
515 **Minimum sequencing coverage analysis**
516 Samples were sub-samples using SAMtools view -bs 42.X, where X represents the percentage
517 to which the data was reduced to, and 42 as the seed number to obtain reproducible results.
518
519 **Processing of RNA-seq data**
520 Sequencing reads were aligned to the UCSC Drosophila melanogaster dm3 genome using
521 STAR v2.7.2a with the following parameters --outFilterType BySJout --
522 outFilterMultimapNmax 20 --alignSJoverhangMin 8 --alignSJDBoverhangMin 1 --
523 outFilterMismatchNmax 999 --outFilterMismatchNoverReadLmax 0.04 --alignIntronMin 20 --
524 alignIntronMax 1000000 – alignMatesGapMax. Aligned reads were filtered for their mapping
525 quality (MAPQ > 30), using SAMtools v1.9. Reads per exon were counted using featureCounts
526 v1.6.2 and parameters -C -p -B -t exon -g gene_id.
527
528 {Bibliography}
529 Baldi, S. *et al.* (2018) 'array regularity and spacing by nanopore sequencing', *Nature*
530 *Structural & Molecular Biology*, 25(September). Available at:
531 https://doi.org/10.1038/s41594-018-0110-0.
532 Baldi, S., Korber, P. and Becker, P.B. (2020) 'Beads on a string—nucleosome array
533 arrangements and folding of the chromatin fiber', *Nature Structural and Molecular Biology*.
534 Nature Research, pp. 109–118. Available at: https://doi.org/10.1038/s41594-019-0368-x.
535 Buenrostro, J.D. *et al.* (2013) 'Transposition of native chromatin for fast and sensitive
536 epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position',
537 *Nature Methods*, 10(12), pp. 1213–1218. Available at: https://doi.org/10.1038/nmeth.2688.
538 Chereji, R. V., Bryson, T.D. and Henikoff, S. (2019) 'Quantitative MNase-seq accurately maps
539 nucleosome occupancy levels', *Genome biology*, 20(1), p. 198. Available at:
540 https://doi.org/10.1186/s13059-019-1815-z.
541 Chereji, R. V. and Clark, D.J. (2018) 'Major Determinants of Nucleosome Positioning',
542 *Biophysical Journal*, 114(10), pp. 2279–2289. Available at:
543 https://doi.org/10.1016/J.BPJ.2018.03.015.
544 Chereji, R. V, Ocampo, J. and Clark, D.J. (2016) 'MNase-sensitive Complexes in Yeast :
545 Nucleosomes and Non-histone Barriers', (December), pp. 565–577. Available at:
546 https://doi.org/10.1016/j.molcel.2016.12.009.
547 Chereji, V. *et al.* (2015) 'Genome-wide profiling of nucleosome sensitivity and chromatin
548 accessibility in Drosophila melanogaster', pp. 1–16. Available at:
549 https://doi.org/10.1093/nar/gkv978.

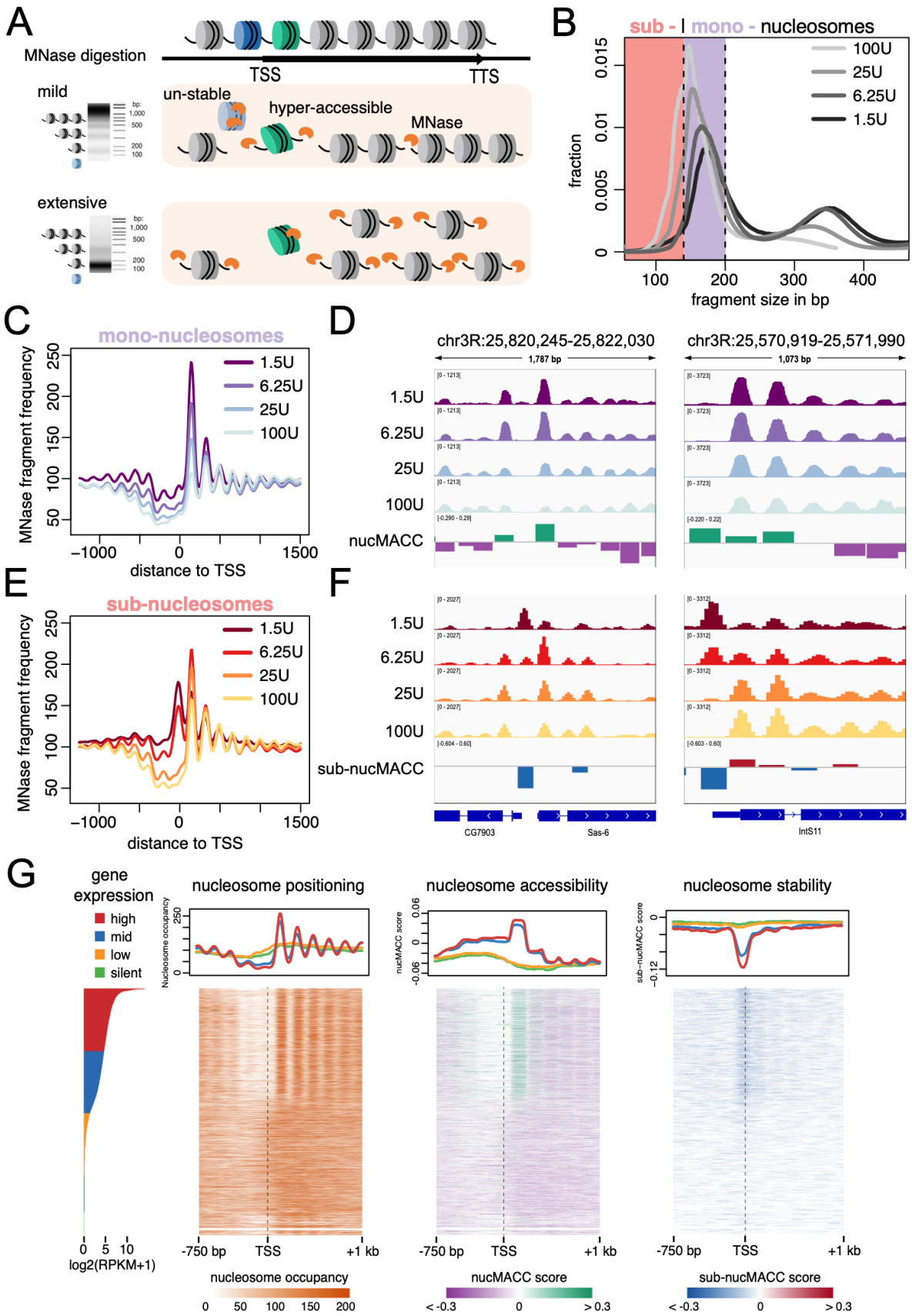
550 Dalal, Y. *et al.* (2007) 'Structure, dynamics, and evolution of centromeric nucleosomes',
551 *Proceedings of the National Academy of Sciences of the United States of America*, 104(41),
552 pp. 15974–15981. Available at: https://doi.org/10.1073/PNAS.0707648104.
553 Davey, C.A. *et al.* (2002) 'Solvent mediated interactions in the structure of the nucleosome
554 core particle at 1.9 a resolution', *Journal of molecular biology*, 319(5), pp. 1097–1113.
555 Available at: https://doi.org/10.1016/S0022-2836(02)00386-8.
556 Diermeier, S. *et al.* (2014) 'TNFα signalling primes chromatin for NF-κB binding and induces
557 rapid and widespread nucleosome repositioning.', *Genome biology*, 15(12), p. 536. Available
558 at: https://doi.org/10.1186/s13059-014-0536-6.
559 Fuda, N.J. *et al.* (2015) 'GAGA Factor Maintains Nucleosome-Free Regions and Has a Role in
560 RNA Polymerase II Recruitment to Promoters', *PLOS Genetics*, 11(3), p. e1005108. Available
561 at: https://doi.org/10.1371/JOURNAL.PGEN.1005108.
562 Hasson, D. *et al.* (2013) 'The octamer is the major form of CENP-A nucleosomes at human
563 centromeres', *Nature Structural & Molecular Biology 2013 20:6*, 20(6), pp. 687–695.
564 Available at: https://doi.org/10.1038/nsmb.2562.
565 Henikoff, J.G. *et al.* (2011) 'Epigenome characterization at single base-pair resolution'.
566 Available at: https://doi.org/10.1073/pnas.1110731108/-
567 /DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1110731108.
568 Hu, G. *et al.* (2011) 'Regulation of nucleosome landscape and transcription factor targeting
569 at tissue-specific enhancers by BRG1', *Genome research*, 21(10), pp. 1650–1658. Available
570 at: https://doi.org/10.1101/GR.121145.111.
571 Kaplan, N. *et al.* (2009) 'The DNA-encoded nucleosome organization of a eukaryotic
572 genome', *Nature*, 458(7236), pp. 362–366. Available at:
573 https://doi.org/10.1038/nature07667.
574 Kireeva, M.L. *et al.* (2002) 'Nucleosome remodeling induced by RNA polymerase II: loss of
575 the H2A/H2B dimer during transcription', *Molecular cell*, 9(3), pp. 541–552. Available at:
576 https://doi.org/10.1016/S1097-2765(02)00472-0.
577 Kornberg, R.D. (1974) 'Chromatin structure: A repeating unit of histones and DNA', *Science*,
578 184(4139), pp. 868–871. Available at: https://doi.org/10.1126/science.184.4139.868.
579 Kubik, S. *et al.* (2015) 'Nucleosome Stability Distinguishes Two Different Promoter Types at
580 All Protein-Coding Genes in Yeast', *Molecular Cell*, 60(3), pp. 422–434. Available at:
581 https://doi.org/10.1016/j.molcel.2015.10.002.
582 Li, J. and Gilmour, D.S. (2013) 'Distinct mechanisms of transcriptional pausing orchestrated
583 by GAGA factor and M1BP, a novel transcription factor', *EMBO Journal*, 32(13), pp. 1829–
584 1841. Available at: https://doi.org/10.1038/EMBOJ.2013.111.
585 Liu, Y. *et al.* (2019) 'FACT caught in the act of manipulating the nucleosome', *Nature 2019*
586 *577:7790*, 577(7790), pp. 426–431. Available at: https://doi.org/10.1038/s41586-019-1820-
587 0.
588 Luger, K. *et al.* (1997) 'Crystal structure of the nucleosome core particle at 2.8 Å resolution',
589 *Nature*, 389(6648), pp. 251–260. Available at: https://doi.org/10.1038/38444.
590 Mieczkowski, J. *et al.* (2016) 'MNase titration reveals differences between nucleosome
591 occupancy and chromatin accessibility.', *Nature communications*, 7(May), p. 11485.
592 Available at: https://doi.org/10.1038/ncomms11485.
593 Olins, A.L., Douglas Carlson, R. and Olins, D.E. (1975) 'Visualization of chromatin
594 substructure: ν bodies', *Journal of Cell Biology*, 64(3), pp. 528–537. Available at:
595 https://doi.org/10.1083/jcb.64.3.528.

596 Prasad, R. *et al.* (2016) 'Coordinated Action of Nap1 and RSC in Disassembly of Tandem
597 Nucleosomes', *Molecular and cellular biology*, 36(17), pp. 2262–2271. Available at:
598 https://doi.org/10.1128/MCB.00195-16.
599 Ramachandran, S., Ahmad, K. and Henikoff, S. (2017) 'Transcription and Remodeling
600 Produce Asymmetrically Unwrapped Nucleosomal Intermediates', *Molecular Cell*, 68(6), pp.
601 1038-1053.e4. Available at: https://doi.org/10.1016/j.molcel.2017.11.015.
602 Rhee, H.S. *et al.* (2014) 'Subnucleosomal structures and nucleosome asymmetry across a
603 genome', *Cell*, 159(6), pp. 1377–1388. Available at:
604 https://doi.org/10.1016/j.cell.2014.10.054.
605 Schones, D.E. *et al.* (2008) 'Dynamic Regulation of Nucleosome Positioning in the Human
606 Genome', *Cell*, 132(5), pp. 887–898. Available at: https://doi.org/10.1016/j.cell.2008.02.022.
607 Schwartz, U. *et al.* (2018) 'Characterizing the nuclease accessibility of DNA in human cells to
608 map higher order structures of chromatin', *Nucleic Acids Research*, 47(3), pp. 1239–1254.
609 Available at: https://doi.org/10.1093/nar/gky1203.
610 Silberhorn, E. *et al.* (2016) 'Plasmodium falciparum Nucleosomes Exhibit Reduced Stability
611 and Lost Sequence Dependent Nucleosome Positioning', *PLOS Pathogens*. Edited by K.W.
612 Deitsch, 12(12), p. e1006080. Available at: https://doi.org/10.1371/journal.ppat.1006080.
613 Struhl, K. and Segal, E. (2013) 'Determinants of nucleosome positioning', *Nature Publishing*
614 *Group*, 20(3), pp. 267–273. Available at: https://doi.org/10.1038/nsmb.2506.
615 Teif, V.B. *et al.* (2012) 'Genome-wide nucleosome positioning during embryonic stem cell
616 development', *Nature Structural & Molecular Biology*, 19(11), pp. 1185–1192. Available at:
617 https://doi.org/10.1038/nsmb.2419.
618 Valouev, A. *et al.* (2011) 'Determinants of nucleosome organization in primary human cells',
619 *Nature*, 474(7352), pp. 516–520. Available at: https://doi.org/10.1038/nature10002.
620 Vierstra, J. *et al.* (2014) 'Coupling transcription factor occupancy to nucleosome architecture
621 with DNase-FLASH', 11(1). Available at: https://doi.org/10.1038/NMETH.2713.
622 Voong, L.N. *et al.* (2016) 'Insights into Nucleosome Organization in Mouse Embryonic Stem
623 Cells through Chemical Mapping', *Cell*, 167(6), pp. 1555-1570.e15. Available at:
624 https://doi.org/10.1016/j.cell.2016.10.049.
625 Weiner, A. *et al.* (2010) 'High-resolution nucleosome mapping reveals transcription-
626 dependent promoter packaging', *Genome research*, 20(1), pp. 90–100. Available at:
627 https://doi.org/10.1101/GR.098509.109.
628 Whyte, W.A. *et al.* (2013) 'Master transcription factors and mediator establish super-
629 enhancers at key cell identity genes', *Cell*, 153(2), pp. 307–319. Available at:
630 https://doi.org/10.1016/J.CELL.2013.03.035/ATTACHMENT/7BB94B8B-BF47-473D-9788-
631 A9CE552FB4BF/MMC9.XLSX.
632 Zouaz, A. *et al.* (2017) 'The Hox proteins Ubx and AbdA collaborate with the transcription
633 pausing factor M1BP to regulate gene transcription', *The EMBO Journal*, 36(19), pp. 2887–
634 2906. Available at: https://doi.org/10.15252/EMBJ.201695751.
635

636

637

Figures and Legends:



641

642 **Figure 1**

643 (A) Schematic representation of differential MNase chromatin digestion. Under mild MNase
644 conditions, chromatin is partially digested and nucleosomes (green, hyper-accessible) from
645 accessible regions are preferentially released (upper panel). At extensive MNase conditions,
646 chromatin is completely digested to nucleosome core particles and MNase-sensitive
647 nucleosomes (blue, un-stable) are lost. (B) Fragment size distribution of MNase titrations.
648 Fragments are selected based on the fragment size and grouped into sub- (< 140 bp, light red
649 section) or mono-nucleosomes (140-200 bp, purple section). (C) and (E) average fragment
650 frequencies at the TSS. (D) and (J) Genome browser snapshot showing the output of the
651 nucMACC pipeline. nucMACC scores are deduced from mono-nucleosomal fragments (D) and
652 sub-nucMACC scores from sub-nucleosomal fragments, respectively (J). (E) Heatmaps sorted
653 by gene expression showing nucleosome positioning (left), accessibility (middle, nucMACC
654 scores), and stability (right, sub-nucMACC scores) at TSS.

655

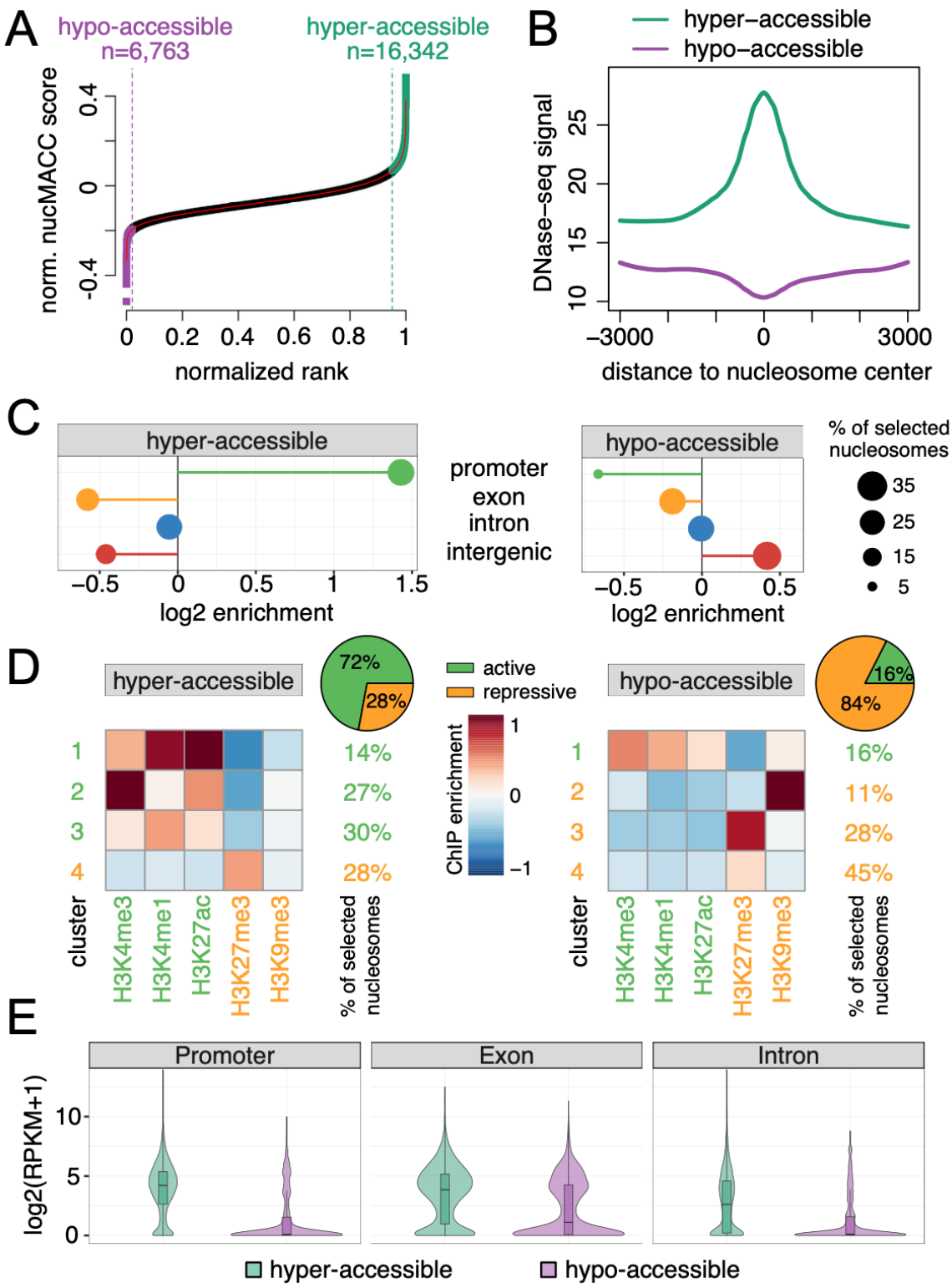
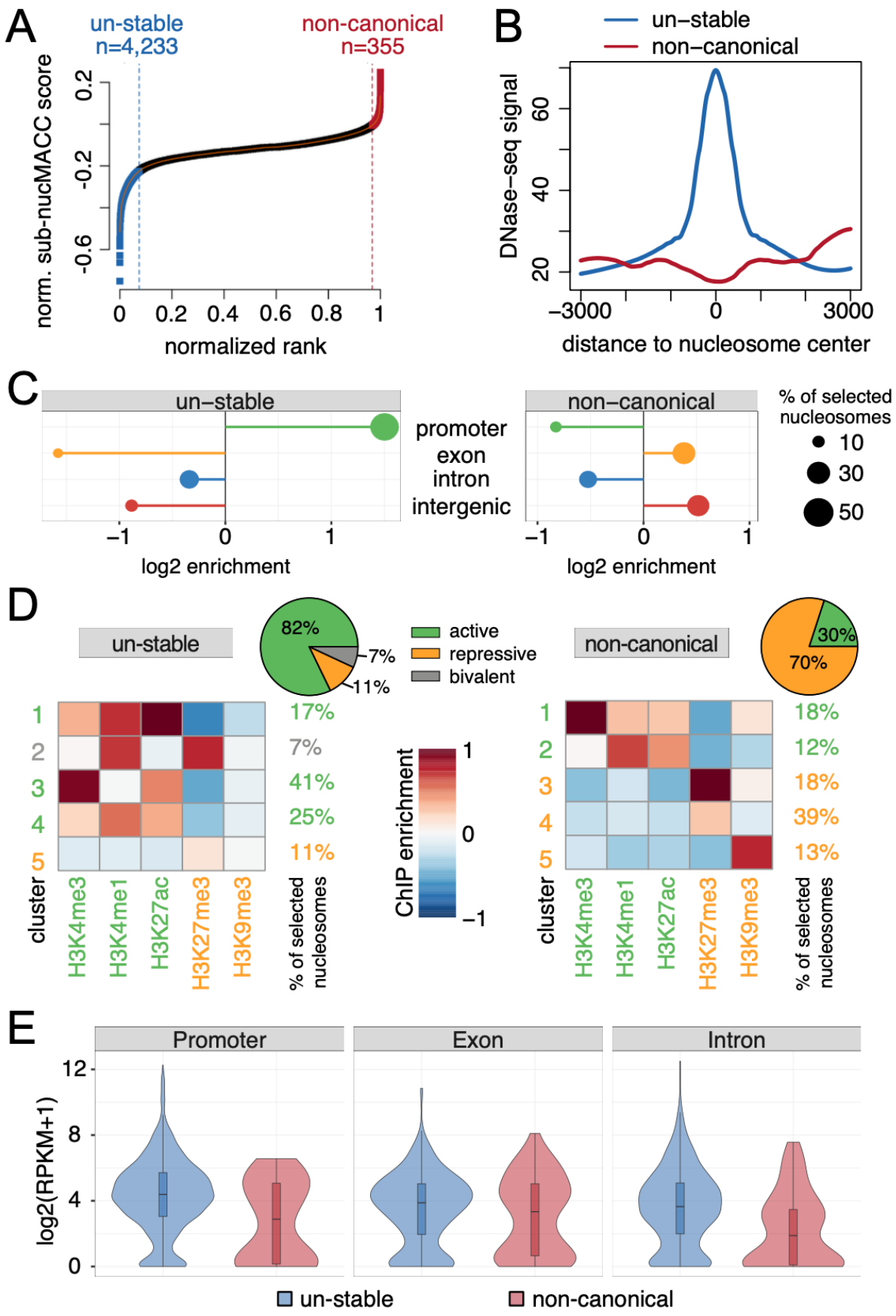


Figure 2

(A) Selection of hyper- (green) and hypo-accessible (purple) nucleosomes based on nucMACC scores. Nucleosomes were ranked by nucMACC score. X- and y-axis were normalized to a range of one. Cut-offs were deduced from the slope of the LOESS curve fit. Nucleosomes

662 before the point, where the slope goes the first time below 1 (purple dashed line) were
663 classified as hypo-accessible. Nucleosomes after the point where the slope of the curve
664 exceeds one again (green dashed line), were classified as hyper-accessible nucleosomes. (B)
665 Correlation of DNase-seq sites with hyper-/hypo-accessible nucleosomes. (C) Enrichment of
666 hyper-/hypo-accessible nucleosomes at genomic features. (D) K-means clustering (n=4) of
667 histone modifications at hyper-/hypo-accessible sites. Centroid values of each cluster are
668 indicated by the color code. The relative number of nucleosomes represented in each cluster
669 is indicated on the right side. Repressive histone modifications are highlighted in orange and
670 active in green, respectively. The pie chart depicts the fraction of nucleosomes assigned to
671 either active or repressive histone clusters. (E) Gene expression levels of genes exhibiting
672 hyper-/hypo-accessible nucleosomes at distinct features. Promoter was defined as 500 bp
673 upstream of the TSS to 300 bp downstream.
674
675


A
un-stable
n=4,233
non-canonical
n=355
norm. sub-nucMACC score
0.2
-0.2
-0.6
0
0.2
0.4
0.6
0.8
1
normalized rank
B
un−stable
non−canonical
DNase−seq signal
60
40
20
−3000
0
3000
distance to nucleosome center
C
un−stable
non−canonical
promoter
exon
intron
intergenic
−1
0
1
log2 enrichment
% of selected nucleosomes
10
30
50
D
un-stable
82%
7%
11%
active
repressive
bivalent
non-canonical
30%
70%
1
2
3
4
5
17%
7%
41%
25%
11%
18%
12%
18%
39%
13%
cluster
H3K4me3
H3K4me1
H3K27ac
H3K27me3
H3K9me3
% of selected nucleosomes
ChIP enrichment
1
0
−1
E
Promoter
Exon
Intron
log2(RPKM+1)
12
8
4
0
un-stable
non-canonical

676

677

678 **Figure 3**

679 (A) Selection of un-stable (blue) and non-canonical (red) nucleosomes based on nucMACC
680 scores. Nucleosomes were ranked by sub-nucMACC score. X- and y-axis were normalized to a
681 range of one. Cut-offs were deduced from the slope of the LOESS curve fit. Nucleosomes
682 before the point, where the slope goes the first time below 1 (blue dashed line) were classified
683 as un-stable. Nucleosomes after the point where the slope of the curve exceeds one again
684 (red dashed line), were classified as non-canonical nucleosomes. (B) Correlation of DNase-seq
685 sites with un-stable and non-canonical nucleosomes. (C) Enrichment of un-stable or non-
686 canonical nucleosomes at genomic features. (D) K-means clustering (n=5) of histone
687 modifications at un-stable or non-canonical sites. Centroid values of each cluster are indicated
688 by the color code. The relative number of nucleosomes represented in each cluster is
689 indicated on the right side. Repressive histone modifications are highlighted in orange, active
690 in green, respectively. The bivalent cluster enriched for both repressive and active histone
691 modifications is indicated in grey. The pie chart depicts the fraction of nucleosomes assigned
692 to either active, repressive, or bivalent histone clusters. (E) Gene expression levels of genes
693 exhibiting un-stable or non-canonical nucleosomes at distinct features. Promoter was defined
694 as 500 bp upstream of the TSS to 300 bp downstream.

695

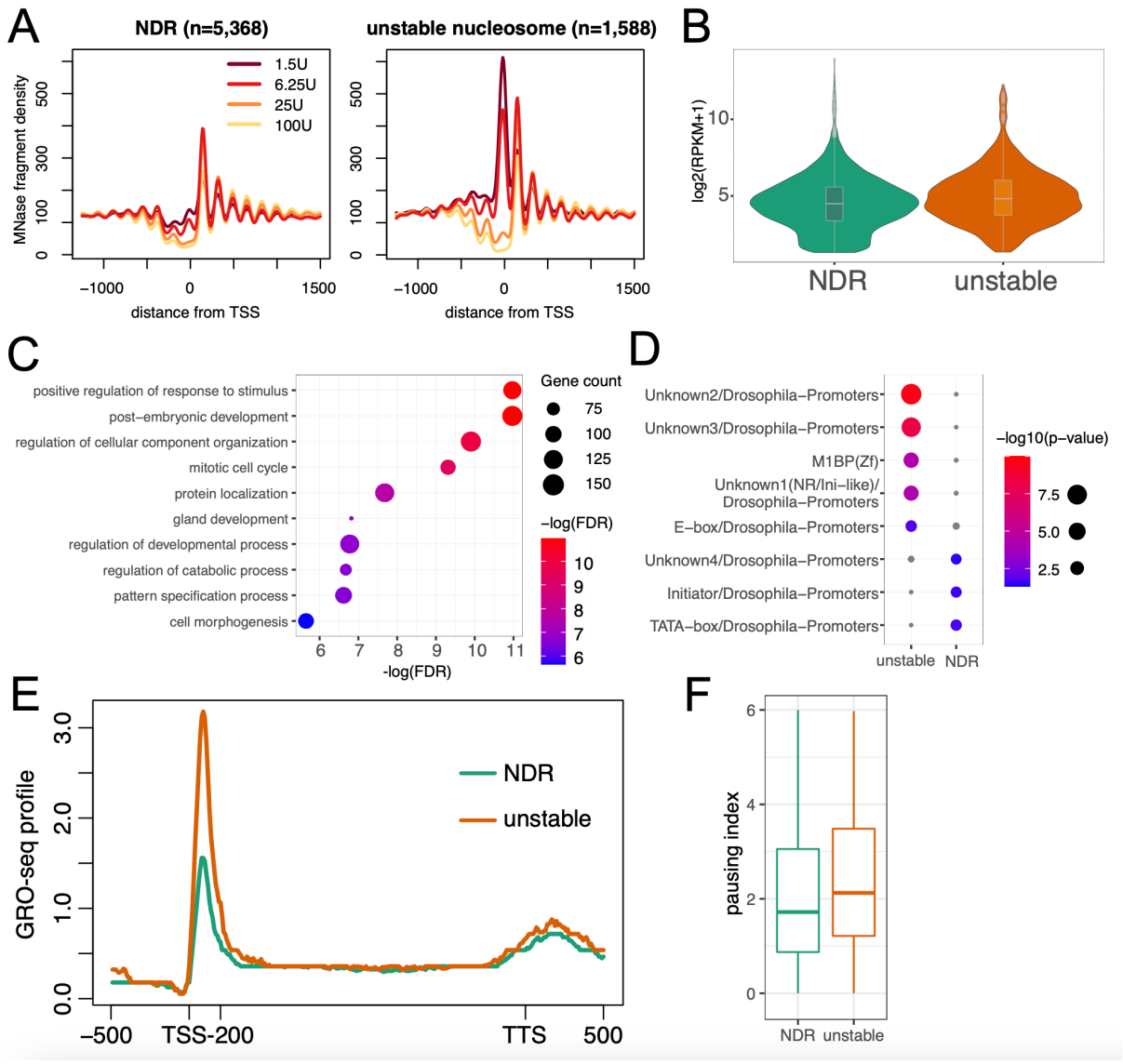
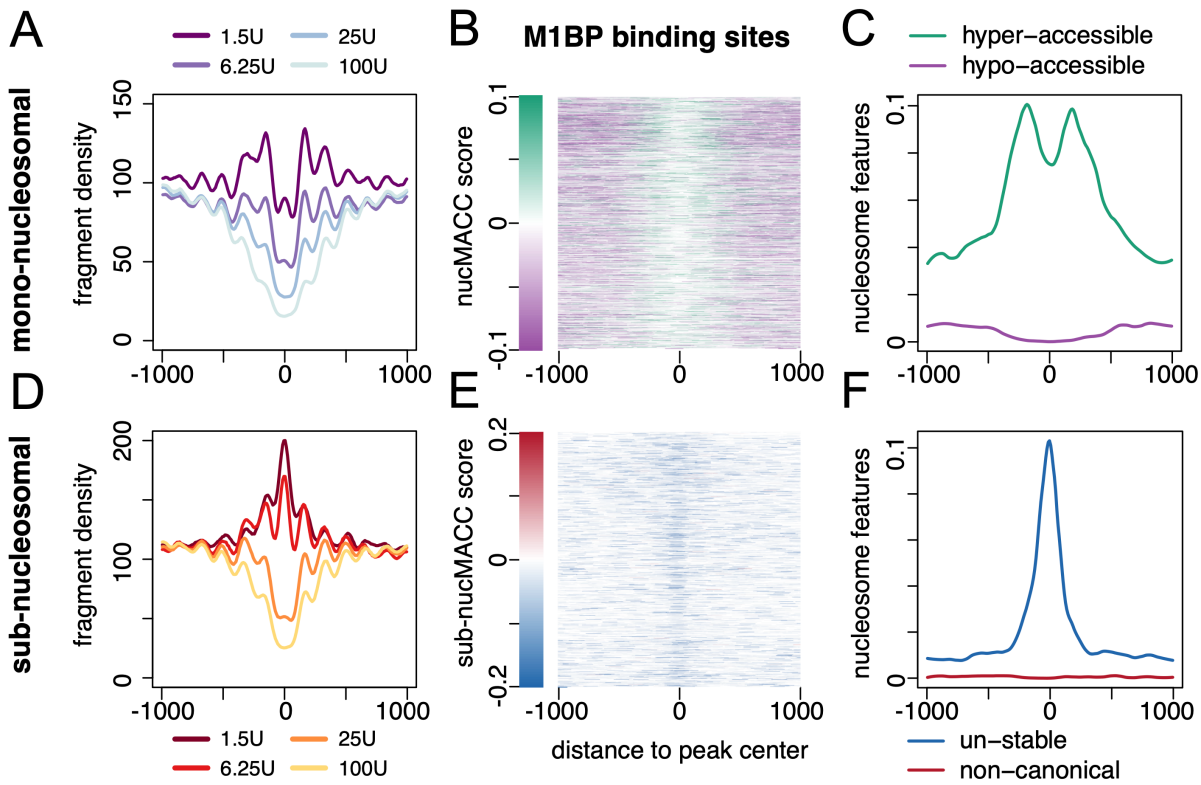


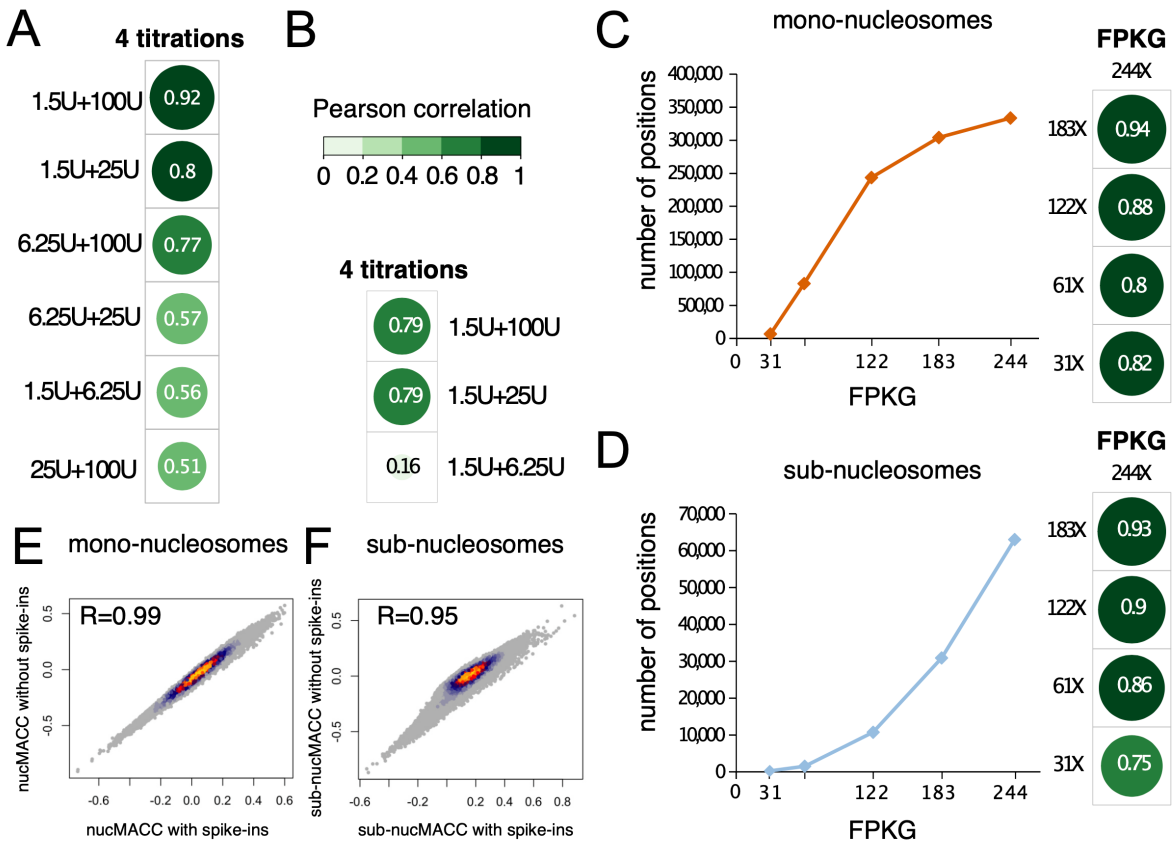
Figure 4

(A) Average fragment frequencies at the TSS of expressed genes either containing an un-stable nucleosome (n=1,588, right) or exhibiting a nucleosome depleted region (NDR) (n=5,368, left) directly upstream of the TSS. (B) Violin plot showing expression levels of genes either containing an un-stable nucleosome (n=1,588, right) or exhibiting a nucleosome depleted region (NDR) (n=5,368, left) directly upstream of the TSS. (C) Enrichment of biological processes associated with genes characterized by an un-stable nucleosome promoter. (D) Motif enrichment analysis in the un-stable nucleosome (left) or NDR (right) promoter. (E) Median GRO-seq profile showing nascent RNA abundance. Regions upstream of TSS, downstream of 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

709 the pausing index based on the GRO-seq signal. Pausing index was calculated as the fold
710 change of the GRO-seq signal in the first 200 bp downstream of the TSS versus the remaining
711 gene body.
712
713



714
715 **Figure 5**
716 (A) and (D) average fragment frequencies of mono- (A) and sub-nucleosomal (D) fragments at
717 M1BP bound sites. (B) and (E) Heatmap showing nucMACC (B) or sub-nucMACC (E) scores at
718 M1BP bound sites. (C) and (F) Distribution of hyper-/hypo-accessible (C) or un-stable and non-
719 canonical (F) nucleosomes at M1BP bound sites.
720



721

722

723 **Figure 6**

724 (A-B) Correlation of nucMACC (A) or sub-nucMACC (B) scores between samples with two or
725 four MNase titrations. The pearson correlation coefficient is indicated (C-D) The number of
726 called nucleosomes based on sequencing depth (FPKG – fragments per kilobase of genome).
727 Correlation of nucMACC (C) or sub-nucMACC (D) scores between samples with different
728 sequencing depths is illustrated at the right. (E-F) Scatterdensityplot showing correlation of
729 nucMACC (E) or sub-nucMACC (F) scores with or without spike-in information. The Pearson
730 correlation coefficient is indicated.