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

AUTHORS:

Wernig-Zorc Sara[1, 2], Kugler Fabian[1], Schmutterer Leo[3], Räß Patrick[3], Hausmann Clemens[3], Holzinger Simon[1], Längst Gernot[1]*, Schwartz Uwe[3]*

[1] Regensburg Center for Biochemistry (RCB), University of Regensburg, Regensburg, Germany
[2] Present address: St. Anna Children's Cancer Research Institute (CCRI), Vienna, Austria
[3] 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 chromatin structure and nucleosome positioning. Data analysis is complex due to the AT-dependent sequence bias of the endonuclease, and the requirement for high sequencing depth. Here, we present the newly developed nucleosome-based MNase accessibility (nucMACC) pipeline unveiling the regulatory chromatin landscape by measuring nucleosome accessibility and stability. nucMACC represents the first systematic, and genome-wide approach for detecting unstable ("fragile") nucleosomes.

We characterized the regulatory nucleosomal landscape in *D. melanogaster* and *S. cerevisiae*. Two functionally distinct sets of promoters were identified, one associated with an unstable nucleosome and the other being nucleosome depleted. Chromatin structure analysis shows that unstable nucleosomes present intermediate states of nucleosome remodeling, preparing inducible genes for transcriptional activation in response to stimuli or stress. The presence of unstable nucleosomes correlates with RNA polymerase II proximal pausing. The nucMACC pipeline offers unparalleled precision and depth in nucleosome research and is a valuable tool for future nucleosome studies.

Teaser:

The nucMACC pipeline quantifies the local and global functional alterations of chromatin structure.

INTRODUCTION:

Nucleosomes are the basic unit of DNA packaging in eukaryotic cells. First electron microscopy images of chromatin spreads displayed regularly spaced nucleosome core particles, separated by a short linker DNA, appearing like “beads-on-a-string” (Olins *et al*, 1975). The nucleosome core particle consists of 147 bp of DNA wrapped around a histone octamer, composed of the histone proteins H2A, H2B, H3, and H4 (Luger *et al*, 1997; Kornberg, 1974). 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; Compton *et al*, 1976; Prunell & Kornberg, 1982). DNA in the linker is generally accessible to DNA-binding factors, whereas access to DNA wrapped around the histone octamer is hindered. Nucleosome positioning, structure, and modification play a critical role in regulating gene expression, determining the accessibility of DNA to DNA-binding factors, thereby defining cell type-specific gene activity and repression (Valouev *et al*, 2011; Teif *et al*, 2012).

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

In addition to the positioning of nucleosomes, the structural property of the nucleosomes is central to DNA accessibility and chromatin function. Nucleosomes can vary in their histone composition and number. Histone variants modify the structural and functional properties of nucleosomes, playing specific roles in chromatin regulation, and being incorporated into nucleosomes independent of the cell cycle (Talbert & Henikoff, 2017). For example, the histone H3 variant CENP-A is incorporated into nucleosomes at the centromeres. DNA-histone interactions are weakened in CENP-A-containing nucleosomes, partially unwinding DNA from the histone surface (Hasson *et al*, 2013). In addition, the human testis-specific histone H3

68 variants, H3T and H3.5, form nucleosomes with reduced stability, as compared to
69 nucleosomes with the canonical H3, as does the C-terminal half of H2A.Z1 (Sato *et al*, 2020;
70 Urahama *et al*, 2016; Tachiwana *et al*, 2010). Hexasomes lacking an H2A-H2B dimer, can be
71 formed by a variety of mechanisms including transcription, histone assembly, and chromatin
72 remodeling (Kireeva *et al*, 2002; Prasad *et al*, 2016; Liu *et al*, 2019; Ramachandran *et al*, 2017;
73 Hsieh *et al*, 2022). The preferential presence of hexasomes near transcription start sites
74 suggests a potential role in gene regulation (Rhee et al, 2014). However, it is still unclear
75 whether these altered/sub-nucleosomal structures are stable or short-lived nucleosome
76 intermediates *in vivo*. Furthermore, there are several studies revealing structurally altered
77 nucleosomes, like lexosome, alterosome, pre-nucleosome, tetrasome and reversosome, that
78 could play specific roles in maintaining an active genome (Lavelle & Prunell, 2007; Fei *et al*,
79 2015; Ramachandran & Henikoff, 2016).
80 MNase-seq has been widely used to study chromatin structure and nucleosome positioning.
81 It involves the use of micrococcal nuclease (MNase), an enzyme that selectively cleaves linker
82 DNA, leaving the nucleosome core particles intact. The resulting DNA fragments are
83 sequenced and analyzed to determine the location and distribution of nucleosomes along the
84 genome (Baldi *et al*, 2020). MNase-seq provides a high-resolution map of the nucleosomal
85 landscape and has contributed to our understanding of the dynamic and regulated nature of
86 chromatin structure (Valouev *et al*, 2011; Teif *et al*, 2012; Diermeier *et al*, 2014). Initially, high
87 MNase digestion conditions were used to completely hydrolyze chromatin to the mono-
88 nucleosomal level. Later, we and others combined different MNase digestion conditions to
89 study chromatin accessibility and nucleosome stability (Schwartz *et al*, 2018; Henikoff *et al*,
90 2011; Mieczkowski *et al*, 2016; Chereji *et al*, 2019, 2015; Schwartz *et al*, 2023). Due to the
91 complexity of data analysis in MNase titration experiments, new metrics have been
92 introduced to provide a stable framework for accurate quantification of chromatin
93 accessibility and nucleosome occupancy (Mieczkowski *et al*, 2016; Chereji *et al*, 2019).
94 However, up to six different MNase concentrations, from very low to high, were used per
95 sample and additional spike-ins were suggested to be necessary, resulting in a high workload,
96 high costs, and requiring large amounts of sample material. Furthermore, the lack of an
97 established analytical pipeline makes it difficult to adapt these kinds of experiments to new
98 sample formats.

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

RESULTS:

The nucMACC pipeline requires a minimum of two MNase-seq titration conditions 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). At mild MNase conditions, nucleosomes from accessible and active genomic sites are preferentially released, referred to as hyper-accessible nucleosomes (Figure 1A, green nucleosome) (Schwartz *et al*, 2018; Mieczkowski *et al*, 2016). Increasing the MNase concentration or digestion time, increasingly releases nucleosome core particles from chromatin, until the nucleosomal chain is fully hydrolyzed. Furthermore, MNase is a processive enzyme. Once the nucleosome is released from chromatin, MNase progressively trims the linker DNA and then continues - at a lower rate - to hydrolyze the histone bound DNA, resulting in sub-nucleosomal fragments. Canonical nucleosome core particles are highly resistant to intra-nucleosomal MNase cleavage, and the nucleosome core DNA remains mainly complete, even at high MNase digestion conditions. However, non-canonical

131 nucleosomes, often referred to as unstable or fragile nucleosomes, exhibit an increased
132 MNase sensitivity and are degraded to sub-nucleosome-sized fragments already at low
133 MNase concentrations (Figure 1A, blue nucleosome) (Kubik *et al*, 2015; Weiner *et al*, 2010;
134 Voong *et al*, 2016). We refer to these kinds of nucleosomes as unstable nucleosomes in the
135 remainder of this manuscript. At high MNase conditions, unstable nucleosomes are fully
136 degraded and are therefore lost in this kind of MNase dataset. Hence, sub-nucleosomal
137 fragments can be used to measure nucleosome stability in MNase titration experiments, as
138 reflected by resistance to internal MNase cleavage.
139
140 Starting from raw sequencing files in the FASTQ format, the nucMACC pipeline first assesses
141 the sequencing quality. Next, the reads are aligned to the provided reference genome and
142 subsequently mapped reads are filtered based on mapping quality (Figure S2). Optionally
143 ambiguous genomic elements, such as blacklisted regions or mitochondrial chromosomes,
144 can be removed from the analysis. To obtain high-resolution nucleosome positions, fragments
145 between 140 and 200 bp in length, the typical size of mono-nucleosomal DNA, are selected
146 (Figure 1B). Size-selected fragments of all MNase titration conditions are pooled and used to
147 derive a comprehensive map of the genome-wide nucleosome positions.
148 Recently, a metric to measure chromatin accessibility, termed MACC (MNase accessibility),
149 was introduced (Mieczkowski et al., 2016). As the MACC scores provide an elegant way to
150 measure chromatin accessibility based on MNase-seq data, we adapted the MACC scoring
151 system but changed the principle of calculation to obtain a higher annotation resolution and
152 accuracy. Instead of counting the MNase reads in arbitrary genomic bins as in the original
153 MACC version, we quantify the MNase accessibility directly at the defined nucleosome
154 positions, which results in a specific accessibility score for each individual nucleosome
155 (nucMACC). Next, linear regression is conducted on the normalized fragment frequencies for
156 each nucleosome (Figure S2, see Methods section for details). The slope of the regression line
157 multiplied by minus one is used as the nucMACC score. Since MNase exhibits an increasing
158 AT preference at higher concentrations, raw nucMACC scores are normalized for the
159 underlying GC content using a LOESS regression (Schwartz *et al*, 2018).
160 As a new feature of the nucMACC pipeline fragments shorter than 140 bp, representing sub-
161 nucleosomal DNA fragments, are used to measure nucleosome stability. We use only the
162 lowest MNase concentration to call these sub-nucleosomal positions (called sub-

163 nucleosomes in the remaining text), as at higher MNase titration conditions, these unstable
164 nucleosomes are fully hydrolyzed (Figure 1A, blue nucleosome). Analogous to the nucMACC
165 scores, sub-nucleosomal fragment frequencies for each MNase titration condition are
166 quantified at the defined sub-nucleosome positions. The slope of the regression is used to
167 measure nucleosome stability and normalized to the underlying GC content. The normalized
168 score is referred to as the sub-nucleosome MNase accessibility score (sub-nucMACC).
169
170 **The nucMACC pipeline scores nucleosome accessibility and stability**
171 To establish and validate the nucMACC pipeline, we used the MNase chromatin digestions
172 coupled with H4 immunoprecipitation from the original MACC publication (Mieczkowski *et al*,
173 2016). The dataset collection consists of four different MNase concentration conditions (1.5U,
174 6.25U, 25U, and 100U), performed in two replicates each. The fragment size distribution is an
175 important quality control parameter, as it clearly shows the degree of chromatin hydrolysis
176 in each sample (Figure 1B). With increasing MNase concentrations the di-nucleosome fraction
177 is disappearing, and the levels of mono-nucleosomes are increasing. The dominant mono-
178 nucleosomal peak is in the range of 140–200 bp, while the peak summit shifts with higher
179 MNase concentrations closer to 147 bp, the protected DNA-length of a nucleosome core
180 particle. Fragment sizes shorter than 140 bp indicate sub-nucleosomal DNA fragments, which
181 mainly arise from DNA cleavage within the nucleosomal core. Sub-nucleosomal DNA
182 fragments are barely present at mild MNase conditions but increase with higher MNase
183 concentrations. Based on the shape of the fragment size distribution curve, we selected the
184 fragments in the range of 140 – 200 bp in length for mono-nucleosome analysis and fragments
185 shorter than 140 bp for sub-nucleosome analysis, respectively (Figure 1B).
186 Plotting the fragment frequencies of mono- and sub-nucleosomes at the TSS shows that the
187 selected categories behave differently, suggesting different structural information or
188 functional behavior (Figure 1C and 1E). Analyzing the mono-nucleosomes of the two high
189 MNase conditions (25U and 100U) reveals the well-known nucleosome positioning pattern at
190 the TSS: a nucleosome depleted region (NDR) followed downstream by a phased nucleosome
191 array, which starts with a well-positioned +1-nucleosome and spreads into the gene. The low
192 MNase conditions (1.5U and 6.25U) reveal the same mono-nucleosomal positions, however,
193 specific positions, such as the +1-nucleosome, are markedly enriched compared to the high
194 MNase conditions. This effect clearly indicates that nucleosomes, depending on their genomic

195 location, exhibit different release rates from chromatin. The nucMACC score provides a
196 quantitative assessment of chromatin accessibility, revealing the change in mono-
197 nucleosome fragment frequencies with increasing MNase concentration (Figure 1D and S2).
198 Comparing the mono-nucleosome and the sub-nucleosome TSS profile at high MNase
199 conditions, reveals a very similar pattern, suggesting that once the nucleosomes were
200 released from chromatin, MNase trims the nucleosomal DNA. However, in the sub-
201 nucleosomal fraction, an additional dominant peak at the denominated NDR appears at low
202 MNase conditions. As the analyzed dataset is based on a H4-ChIP, these observations suggest
203 that the NDR is not free of histones but occupied by MNase sensitive nucleosomes (Figure
204 1E). The sub-nucMACC score allows the quantification of the nucleosome stabilities at these
205 sites, as it monitors the changes in sub-nucleosome fragment frequencies over MNase
206 concentrations (Figure 1F and S2).
207 Even though the nucMACC and sub-nucMACC scores are calculated in a similar manner, their
208 scores do not correlate at overlapping positions, confirming that both metrics reveal distinct
209 nucleosomal features (Figure S3A). The correlation of the nucMACC scores with the original
210 MACC scores (Mieczkowski *et al*, 2016) was modest (R=0.48), which can be explained by the
211 fact that the scores in our analysis are centered on the actual nucleosome positions revealing
212 functional details with high resolution. The sub-nucMACC scores, which is a novel category
213 implemented in our pipeline, showed no correlation (R=-0.24) indicating new features of the
214 dynamic chromatin structure (Figure S3B-C).
215 By centering the nucMACC scores on the TSS, we show that the +1 nucleosomes of actively
216 transcribed genes are in general hyper-accessible, whereas nucleosomes further downstream
217 in the gene body did not significantly change nucleosome accessibility (Figure 1G).
218 Remarkably, in genic regions, we observed no difference in accessibility between silent and
219 highly expressed genes (Figure S3D), confirming that chromatin accessibility is not changed at
220 a global scale but modulated only locally, changing the accessibility of individual nucleosomes
221 at regulatory sites, such as the promoter (Schwartz *et al*, 2018).
222 The nucleosome stability profile at the TSS of active genes was different compared to the
223 nucleosome accessibility score. The lowest nucleosome stability scores were obtained
224 upstream of the +1 nucleosome directly at the TSS of active genes. The sub-nucleosome
225 analysis uncovered unstable nucleosomes directly at the NDR, showing an overall low
226 nucleosome occupancy in mono-nucleosome analysis (Figure 1G). In contrast, the

227 transcription end sites (TES) are nucleosome depleted in *Drosophila*, as no unstable
228 nucleosomes can be detected (Figure S3E), in contrast to studies performed in yeast and
229 mouse (Voong *et al*, 2016; Xi *et al*, 2011).
230 Overall, the nucleosome features provided by the nucMACC pipeline are robust and
231 reproducible, as shown by the comparison to other MNase titration experiments that were
232 combined with a H3-ChIP (Figure S4A-G) (Mieczkowski *et al*, 2016). We observed the highest
233 variation between the H3- and the H4-MNase-ChIP experiments in the number and location
234 of the sub-nucleosome positions (Figure S4C). This can be explained, by the significantly lower
235 number of sub-nucleosomal fragments in the H3 low MNase samples compared to the H4
236 dataset, giving rise to fewer called sub-nucMACC positions (addressed in more detail below).
237 Nevertheless, the sub-nucMACC scores at overlapping positions revealed a high correlation
238 (Figure S4C, R=0.8).
239 In summary, splitting up MNase titration data by fragment sizes into mono-nucleosomes and
240 sub-nucleosomes enables the nucMACC pipeline to reveal nucleosome positions, nucleosome
241 accessibility, and nucleosome stability scores within the same experiment.
242 Next, we tested whether the additional histone immunoprecipitation step after chromatin
243 digestion is required to obtain reliable nucleosome features. Without ChIP, mono- and sub-
244 nucleosome positioning changed and overlapping positions showed only a modest correlation
245 with whole chromatin extractions (Fig S4H-K). Changes in the fragment profile were clearly
246 visible upstream of the TSS, and most pronounced at the -1 nucleosome position affecting
247 nucMACC and sub-nucMACC scores (Figure S4L-N). The -1 nucleosome position exhibited the
248 highest accessibility throughout all conditions, whereas the sub-nucleosomal fraction
249 exhibited a broad and pronounced enrichment at the promoter of expressed genes in the low
250 MNase concentrations. This result is reminiscent to other methods measuring chromatin
251 accessibility, such as DNase- or ATAC-seq (Vierstra *et al*, 2014; Buenrostro *et al*, 2013). The
252 differences between whole chromatin analysis and the combined MNase-ChIP datasets can
253 be explained by non-histone proteins bound to DNA (Mieczkowski *et al*, 2016). Our analysis
254 shows that an additional histone immunoprecipitation step is required, to obtain accessibility
255 and stability measurements specifically relating to nucleosomes. Nevertheless, the nucMACC
256 pipeline can be run using whole chromatin extractions and measures overall chromatin
257 accessibility and stability over all molecules bound to DNA.
258

259 **nucMACC scores identify hyper- and hypo-accessible nucleosomes**

260 To identify functionally important nucleosomes, the nucMACC pipeline extracts nucleosomes
261 with exceptionally high or low accessibility, termed hyper- or hypo-accessible nucleosomes.
262 For this, the nucMACC score of each nucleosome is plotted against its rank. Ranks are divided
263 by the total number of nucleosomes and the nucMACC score, by the total range of scores. In
264 that way, both the x and y axes have a value range of 1 (Figure 2A, Supplemental Dataset 1).
265 To geometrically identify the positions, where the signal decreases/increases rapidly, we
266 determined the positions at the curve where the slope is the first/last time below one. Hypo-
267 accessible nucleosomes are defined as the nucleosome positions before the slope of the curve
268 decreases below 1 (Figure 2A, purple dots), and hyper-accessible nucleosomes after the slope
269 increases again above 1, respectively (Figure 2A, green dots). A similar strategy was previously
270 applied to define super-enhancers (Whyte *et al*, 2013).

271 In total, we identified 6,763 (2 %) hypo-accessible and 16,342 (4,9 %) hyper-accessible
272 nucleosomes in the *Drosophila* S2 H4-MNase-ChIP data set. Hyper-accessible nucleosomes
273 are strongly enriched at DNase I hypersensitive sites, whereas hypo-accessible nucleosomes
274 are depleted, confirming the nucMACC score as a measure for general DNA accessibility
275 (Figure 2B). Genomic feature characterization of hyper- and hypo-accessible nucleosomes
276 revealed a significant enrichment at promoter and intergenic regions, respectively (Figure 2C).
277 Hyper-accessible nucleosomes are preferentially associated with H3K4me1, H3K4me3, and
278 H3K27ac histone modifications, marking active TSS and enhancer, whereas hypo-accessible
279 nucleosomes are associated with H3K9me3 and H3K27me3 modifications, indicating
280 heterochromatic regions (Figure 2D). In agreement with the histone modification analysis,
281 hyper-accessible nucleosomes occupy mainly expressed genes, while hypo-accessible
282 nucleosomes are present on silent genes (Figure 2E). Nevertheless, we identified a substantial
283 fraction of hypo-accessible nucleosomes at exons of expressed genes, potentially regulating
284 differential exon usage.

285 The nucMACC pipeline identifies functional nucleosome positions at regulatory sites of the
286 genome, showing that the regulation of genome activity is associated with differential
287 nucleosome accessibility.

288

289 **sub-nucMACC scores identify the presence of non-canonical and unstable nucleosomes**

290 To measure nucleosome stability, sub-nucleosome positions were derived from the sub-
291 nucleosome fragments in the lowest MNase concentration condition. Hyper-accessible
292 nucleosomes are expected to have a higher MNase degradation rate as they are less
293 protected once they are released from chromatin. Here, sub-nucleosome positions are
294 selected, which either contained at least four-fold more normalized sub-nucleosome
295 fragment counts than the mono-nucleosome fraction or did not overlap with any mono-
296 nucleosome position (Figure S1).
297 Like the selection of hyper- and hypo-accessible nucleosomes, nucleosomes exhibiting
298 extreme stability scores (sub-nucMACC) were selected (Figure 3A, Supplemental Dataset 1).
299 This selection strategy resulted in 4,233 (6.7 %) nucleosomes showing very low stability scores
300 and therefore termed unstable nucleosomes. Additionally, a rather small subset of
301 nucleosomes (n=355, 0.6 %) were enriched in the sub-nucleosome fraction of the lowest
302 MNase condition and remained stable at higher MNase concentrations. This group represents
303 most likely structurally altered nucleosomes with additional MNase cleavage sites in the
304 realm of the nucleosome core or non-octamer histone compositions, such as hexa- or
305 hemisomes (Hasson *et al*, 2013; Rhee *et al*, 2014; Dalal *et al*, 2007). Nucleosomes enriched in
306 the low sub-nucleosomal fractions and showing extraordinarily high stability scores are
307 referred here to as non-canonical nucleosomes.
308 In contrast to non-canonical nucleosomes, unstable nucleosome positions are characterized
309 by a high chromatin accessibility (Figure 3B), being mainly located at gene promoters (53,8
310 %), whereas non-canonical nucleosomes are enriched in intergenic and exonic regions (Figure
311 3C). Preferentially, active histone marks are associated with unstable nucleosomes (Figure
312 3D), and accordingly, unstable nucleosomes are present at expressed genes (Figure 3E).
313 Interestingly, we identify a subgroup of unstable nucleosomes associated with active
314 H3K4me1 and repressive H3K27me3 marking bivalent enhancers. Conversely, non-canonical
315 nucleosomes showed enrichment in repressive H3K27me3 and H3K4me1 (Figure 3D).
316 Nevertheless, 30% of the non-canonical nucleosomes are associated with active histone
317 modifications, and therefore non-canonical nucleosomes are located at both expressed and
318 silent genes (Figure 3E).
319
320 **Unstable nucleosomes occupy promoters with high RNA polymerase pausing rate**

321 As we observed that unstable nucleosomes are enriched upstream of the TSS (Figure 1G and
322 Figure 3C) of expressed genes, we addressed their role in gene regulation. Therefore, we
323 divided all expressed genes into two groups: 1) those that contained an unstable nucleosome
324 200 bp upstream of the TSS (n=1,588), and 2) those that exhibited a nucleosome-depleted
325 region (NDR) in the promoter (n=5,363) (Figure 4A, Supplemental Dataset 1). To discriminate
326 between NDR or unstable nucleosome containing promoters, it is obligatory to use MNase
327 treatments coupled with histone immunoprecipitations, as DNA fragments protected by non-
328 histone proteins would dominate the promoter signal in whole chromatin MNase digestion,
329 resulting in sub-nucleosomal fragments at all expressed promoters (Figure S5). Overall, the
330 steady-state transcript abundance of the corresponding genes was similar in both groups and
331 only slightly elevated for genes carrying an unstable nucleosome (Figure 4B, p-value = $2x10^{-}$
332 ^{15}). GO enrichment analysis revealed that genes containing unstable nucleosomes in their
333 promoter are stimulus or developmentally regulated (Figure 4C). Furthermore, unstable
334 nucleosome promoters were enriched for *Drosophila* promoter motifs Unknown1-3 and
335 M1BP motif, but depleted in Initiator, TATA-box, and Unkown4 motif (Figure 4D). As M1BP is
336 a known factor to regulate RNA Pol II pausing depending on nucleosome positioning (Li &
337 Gilmour, 2013), it is suggested that RNA Pol pausing provides an efficient way to induce gene
338 transcription rapidly or synchronously, in response to developmental and environmental
339 signals. We tested whether Pol II pausing differed between NDR or unstable nucleosome
340 promoter using nascent RNA sequencing data (Fuda *et al*, 2015). In agreement with M1BP
341 motif enrichment, unstable nucleosome promoters show a pronounced association with
342 paused Pol II directly downstream of the TSS and higher pausing index rates compared to the
343 NDR promoters, enabling a fast transition into productive transcription at responsive and
344 developmental genes (Figure 4E-D).
345
346 **M1BP binds to DNA occupied by an unstable nucleosome**
347 We next assessed the relationship between the binding of the transcription pausing factor
348 M1BP and nucleosome properties. Here, we investigated nucleosome positioning,
349 accessibility, and stability at annotated M1BP binding sites (Zouaz *et al*, 2017). We observed
350 that M1BP binding sites are surrounded by hyper-accessible nucleosomes (Figure 5A-C).
351 Directly in the binding center of M1BP, we identified a clear enrichment of unstable
352 nucleosomes (Figure 5D-F). This observation suggests that M1BP does not bind to

353 nucleosome-free DNA but rather to DNA, still co-occupied with histones like a pioneer
354 transcription factor (Michael *et al*, 2020). Another possibility is that M1BP competes with
355 nucleosomes at the same position and the signal is a mixture of two different populations of
356 cells that have either a nucleosome or M1BP bound at that position. Regardless, the
357 enrichment of unstable nucleosomes clearly shows that histone DNA interactions are
358 apparently weakened making them sensitive to internal MNase digestions, which could be
359 the result of transcription factor binding to the nucleosome core (Iwafuchi-doi & Zaret, 2014)
360 or active remodeling (Brahma & Henikoff, 2019).
361
362 **Evaluating the robustness of the nucMACC pipeline**
363 Previously four or more MNase digestion conditions have been used to determine chromatin
364 accessibility scores (Mieczkowski *et al*, 2016; Chereji *et al*, 2019). We sought to determine the
365 minimum number of titration points per sample to obtain reliable and robust results with our
366 pipeline. We compared the nucMACC and sub-nucMACC scores using two or four titration
367 points of the same experimental dataset. As a result, we show that only two different MNase
368 conditions are required, being highly correlated with the four titration conditions. This
369 correlation holds true, if the difference in the compared MNase concentration conditions is
370 at least 10-fold (Figure 6A and B; Figure S6A an B). In addition, we only observed consistent
371 results, when using the lowest available MNase concentration to detect unstable
372 nucleosomes within the sub-nucleosomal fragments. At higher concentrations, fragments of
373 unstable nucleosomes are often lost due to their fragile nature.
374 Next, we addressed the importance of sequencing depth for the robustness of the nucMACC
375 pipeline, as MNase-seq analysis usually requires a high sequence coverage of the genome. To
376 obtain a genome independent quantification we normalized the sequencing depth to the
377 number of fragments per nucleosome (FPN). To estimate the total number of nucleosomes
378 we divided the size of the mappable genome by the expected nucleosome repeat length (here
379 we used 170 bp as an approximation):

380 $$Fragments\ per\ nucleosome\ (FPN) = \frac{number\ of\ sequenced\ fragments}{\frac{effective\ genome\ size}{nucleosome\ repeat\ length}}$$

381 We found that using half of the fragments, with coverage of 160 FPN preserves 73% of the
382 mono-nucleosome positions, and the nucMACC score correlation is reasonably high (R=0.88)

383 (Figure 6C and Figure S6C). Whereas further lowering the sequencing depth to 80 FPN, results
384 in a considerable loss of called mono-nucleosome positions to 25%, albeit the nucMACC score
385 correlation remains stable at overlapping nucleosome positions, with a correlation of 0.80.
386 For sub-nucleosomes, a value of 320 FPN results already at low coverage, since the sub-
387 nucleosomal fragments are underrepresented in low MNase digestion conditions. Reducing
388 the coverage to 240 FPN decreases the called sub-nucleosomes to 49% (Figure 6D).
389 Nonetheless, the correlation between sub-nucMACC scores remains very high at overlapping
390 positions (Figure S6D). Below 240 FPN coverage, the sub-nucleosome analysis produces an
391 insufficient amount of called sub-nucleosome positions, not allowing meaningful analysis
392 (Figure 6D). In summary, we recommend a minimum sequencing depth of 240 FPN or higher
393 to be used when analyzing nucleosome stability.
394 Ultimately, we assessed the reliability and robustness of nucMACC and sub-nucMACC scores,
395 testing if the differential amount of nucleosomes released at various MNase titration
396 conditions would bias the scores. Therefore, we ran our pipeline on a published dataset with
397 spiked-in nucleosomes and compared the results with or without spike-in information.
398 Remarkably, we find almost no difference in nucMACC scores when comparing the pipeline
399 with or without spike-in information for mono- (Figure 6E) and sub-nucleosomes (Figure 6F).
400 In conclusion, the nucMACC pipeline does not require spike-ins and is robust and reproducible
401 with only two MNase titrations.
402
403 **Unstable nucleosomes are associated with nucleosome remodeling**
404 As the nucMACC pipeline produces consistent results with only two MNase titrations, we used
405 appropiate published datasets to gain further insights into chromatin biology. A recent
406 MNase-seq study in *Saccharomyces cerevisiae* questioned the existence of MNase-sensitive
407 nucleosomes at promoter regions (Chereji *et al*, 2017). We used our nucMACC pipeline to re-
408 analyze the H4-MNase-ChIP-seq data set starting from their raw sequencing files.
409 Investigating the results of the mono-nucleosome analysis, we observed that both hyper- and
410 hypo-accessible nucleosomes are enriched at the promoter and predominantly at the +1
411 nucleosome position (Figure S7A, Supplemental Dataset 2). We selected the genes either
412 associated with a hyper- or hypo-accessible +1 nucleosome and compared the initiation
413 frequency and pausing rate using NET-seq data (Churchman & Weissman, 2011). In contrast
414 to the results in *D. melanogaster*, genes associated with a hypo-accessible +1 nucleosome

415 exhibited higher expression rates (Figure S7B). This can be explained by the rather
416 homogenous chromatin accessibility in yeast, due to the high gene density and overall high
417 transcriptional activity (Marr *et al*, 2021; Chen *et al*, 2016). Here, the local accessibility to
418 MNase is mainly restricted by the local crowdedness of chromatin and additional non-histone
419 proteins. Hypo-accessible +1 nucleosomes showed a higher occupancy of initiating RNA Pol II
420 and factors of the initiation complex indicating increased DNA protection from MNase
421 digestion by the RNA Pol II pre-initiation complex (Figure S7C). This is in agreement with the
422 higher expression rates compared to hyper-accessible +1 nucleosomes. At normal conditions
423 we did not observe a characteristic enrichment of other factors at hyper-accessible promoters
424 but detected a pronounced enrichment of the heat shock factor 1 (Hsf1) and Spt7. Spt7 is a
425 subunit of the SAGA complex, located upstream of the TSS upon heat shock, suggesting these
426 promoters are kept accessible for rapid inducible activation (Figure S7D).
427 Next, we investigated the sub-nucleosomal fraction. We detected unstable nucleosomes not
428 only at the TTS (Figure S8A), as already described (Chereji *et al*, 2017; Xi *et al*, 2011; Weiner
429 *et al*, 2010), but also directly upstream of the TSS, like in *D. melanogaster* (Figure 7A,
430 Supplemental Dataset 2). Next, we subdivided promoters of expressed genes into promoters
431 with nucleosome-depleted regions (NDR) and promoters harboring an unstable nucleosome.
432 In total 2,942 (53%) promoters are occupied by an unstable nucleosome, illustrating the
433 abundance of this promoter type (Figure 7A). Our classification allowed us to detect sub-
434 nucleosomal structures in independent H2B- and H3-MNase-ChIP data sets, even though
435 higher MNase digestion conditions were used (Figure S8B) (Rossi *et al*, 2021). Those sub-
436 nucleosomal patterns were absent using MNase-ChIP directed against histone modifications,
437 such as H3K4me3, confirming that they are not a result of cross-linking artifacts. Remarkably,
438 histone variant H2A.Z is enriched at unstable nucleosome promoters, agreeing with recent
439 studies showing sub-nucleosomal structures in *S. cerevisiae* (Figure S8B) (Brahma & Henikoff,
440 2019; Xi *et al*, 2011). It was suggested that unstable nucleosomes are preferentially found at
441 wide NDRs (> 150 bp distance between +1 and -1 nucleosome) (Kubik *et al*, 2015). Accordingly,
442 we checked whether sub-nucleosomal fragments are preferentially enriched at wide NDRs
443 but could not observe a significant difference (Figure S8C). As the main difference between
444 the study of Kubik et al. and ours was that we used MNase-ChIP-seq data instead of sole
445 MNase digestions, we reasoned that non-histone proteins protecting DNA accessibility might
446 have biased their analysis. Indeed, the mono-nucleosome-sized fragments at wide NDRs

447 accumulated an additional peak upstream of the TSS, which was not present in the H4-
448 MNase-ChIP seq data (Figure S8D). Furthermore, the sub-nucleosome fraction revealed a high
449 abundance of non-histone proteins upstream of the TSS, likely resulting in nucleosome-sized
450 fragments at wide NDRs under low MNase conditions.
451 The transcription initiation rate at NDR and unstable nucleosome promoters did not
452 significantly differ (Figure 7B). Nevertheless, the pausing index was increased at the
453 promoters harboring an unstable nucleosome as already seen in *D. melanogaster*.
454 Since it was suggested that RSC (remodeling the structure of chromatin) nucleosome
455 remodeling leads to MNase-sensitive structures through partially unwrapped nucleosome
456 intermediates (Brahma & Henikoff, 2019), we tested whether subunits of the RSC complex
457 are overrepresented at the sites of unstable nucleosome promoters compared to NDR
458 promoters using the ChIP-exo datasets (Rossi *et al*, 2021). We did not detect differences at
459 normal conditions, however, upon heat-shock induced stress we identified several RSC
460 subunits, being enriched at the sites of the unstable nucleosomes (Figure 7C).
461 Next, we addressed, whether chromatin remodelers are sufficient to generate unstable
462 nucleosomes. Hence, we used MNase-seq data of genome-wide *in vitro* reconstitutions,
463 where the activity of purified remodelers was tested (Krietenstein *et al*, 2016).
464 The nucleosomes, assembled by salt gradient dialysis (SGD) onto genomic DNA, exhibited a
465 very low level of sub-nucleosomal structures at the unstable nucleosome promoters (Figure
466 7D). The number of sub-nucleosomal fragments increased upon the addition of INO80,
467 creating wider NDRs by positioning the +1 and -1 nucleosomes. The addition of ISW2, RSC and
468 ISW1 further increased the amount of sub-nucleosomal fragments, specifically at unstable
469 nucleosome promoters. Such a distinction of sub-nucleosome fragments into two groups
470 could not be achieved using the NDR width classification after Kubik et al. (Figure S8E).
471 Nevertheless, INO80 alone seems to be sufficient to adjust the NDR width into either narrow
472 or wide NDRs.
473 In summary, the nucMACC pipeline uncovered unstable nucleosomes upstream of the TSS of
474 specific genes in *S. cerevisiae*. Unstable nucleosomes are formed by the activity of chromatin
475 remodeling enzymes in reconstituted nucleosome arrays, suggesting that they generate
476 partially unwrapped nucleosomal intermediates, or sub-nucleosomal structures, like
477 hexasomes.
478

Discussion:

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

MNase-seq is not a new method, and studies to analyze chromatin accessibility have previously been performed (Mieczkowski *et al*, 2016; Chereji *et al*, 2019). However, the use of these pipelines remains challenging to implement, as they use specific annotation files, which are only available for a few organisms, or require preinstalled software packages, some of which are not maintained anymore or require subscriptions. The nucMACC pipeline is publicly available and utilizes the nextflow workflow management system and Docker software containerization guaranteeing a seamless start on different computer systems and reproducible results (DI Tommaso *et al*, 2017). In addition, the nucMACC pipeline uncovers additional properties of structurally altered nucleosomes and sub-nucleosomes, resulting in new insights into chromatin function.

The nucMACC pipeline relates to the original MACC concept, which gives a qualitative and quantitative description of chromatin accessibility (Mieczkowski *et al*, 2016). We made two major modifications to obtain a higher resolution in mono-nucleosome analysis: 1) We selected only mono-nucleosomal fragments for the analysis by narrowing the fragment length selection to 140 – 200 bp. 2) Instead of subdividing the whole genome into arbitrary bins, we used the called nucleosome positions to measure chromatin accessibility. Consequently, every nucleosome is further characterized by its linker DNA accessibility. Nevertheless, the main findings of the original MACC study about chromatin accessibility could be confirmed with our pipeline.

A recent study used six different MNase titrations and the addition of spike-ins to accurately measure the kinetics of nucleosome release from chromatin (q-MNase-seq) (Chereji *et al*, 2019). The high number of MNase experiments and high sequencing depth requirements, make the data generation extremely laborious and costly, especially for organisms with large

511 genomes. Here we established the minimal parameters to reveal chromatin accessibility and
512 could show that the nucMACC pipeline works robustly with only two MNase titration
513 conditions that should differ in about one order of MNase units applied. The use of spike-ins
514 did not significantly improve the analysis and is therefore not required as the pipeline
515 measures the relative change between the conditions. Nevertheless, to quantitively measure
516 absolute nucleosome occupancy between various conditions spike-ins are compulsory. We
517 could not compare the results of our nucMACC analysis directly with the q-MNase analysis
518 (Chereji *et al*, 2019), as analyzed data tracks are not available, and we were not able to
519 reproduce the results shown in the publication based on the provided method description.
520 However, our analysis confirmed the results of the q-MNase analysis, that hetero- and
521 euchromatic regions exhibit similar accessibility rates in *D. melanogaster* (Chereji *et al*, 2019),
522 which has been initially observed in human cells (Schwartz *et al*, 2018). As a unique feature,
523 nucleosomes with extraordinary chromatin accessibility are automatically detected in the
524 nucMACC pipeline. Hyper-accessible nucleosomes are enriched at the promoter and correlate
525 with DNase-seq assays, transcriptional activity, and active histone marks, whereas hypo-
526 accessible nucleosomes show opposite characteristics.
527 So far MNase sensitive nucleosomes have been identified by direct comparison of high and
528 low MNase conditions at a limited subset of selected genomic regions, such as TSS, Enhancers
529 or CTCF binding sites (Ishii *et al*, 2015; Schwartz *et al*, 2018; Brahma & Henikoff, 2019; Kubik
530 *et al*, 2015; Voong *et al*, 2016; Iwafuchi-Doi *et al*, 2016; Jeffers & Lieb, 2017; Xi *et al*, 2011).
531 The nucMACC pipeline provides the first systematic, unbiased and genome wide approach to
532 detect unstable nucleosomes. In agreement with previous reports, we find them
533 preferentially upstream of the TSS of active genes. However, in contrast, we do not observe
534 a correlation between transcription strength or NDR width and the presence of unstable
535 nucleosomes (Kubik *et al*, 2015; Brahma & Henikoff, 2019; Voong *et al*, 2016; Ishii *et al*, 2015;
536 Michl-Holzinger *et al*, 2022). As the positions of the denominated MNase-sensitive
537 nucleosomes overlap with binding sites of non-histone factors at TSS, enhancer or
538 transcription factor binding sites, we clearly show that it is necessary to use MNase combined
539 with histone-ChIP datasets. Otherwise, DNA protected by non-histone proteins will result in
540 nucleosome-sized fragments, preferentially at low MNase conditions and bias the results.
541 Here we show that promoters in D. melanogaster or S. cerevisiae can be classified into two
542 distinct promoter nucleosome configurations, promoters occupied by an unstable

543 nucleosome upstream of the TSS, and nucleosome depleted promoters. Although their
544 associated genes have similar transcript abundance, these promoter types are functionally
545 different. We observed that genes associated with unstable nucleosomes had higher RNA pol
546 II pausing rates in both *D. melanogaster* and *S. cerevisiae*. In line with other studies, we find
547 unstable nucleosomes being associated with specific genes that are transcriptionally
548 activated in response to stimuli (Jeffers & Lieb, 2017; Xi *et al*, 2011). Importantly, unstable
549 nucleosomes are found in the promoters of genes before the induction of environmental
550 changes, suggesting that these genes are poised for rapid upregulation. Therefore, we
551 hypothesize that nucleosome stability is encoded at gene promoters, contributing to the high
552 transcriptional plasticity observed at these sites. We could confirm our hypothesis using the
553 *in vitro* reconstituted nucleosome arrays on the yeast genome (Krietenstein *et al*, 2016). In
554 these datasets, we clearly observe that the addition of chromatin remodeling enzymes,
555 results in the formation of unstable nucleosomes at the same subset of promoters as *in vivo.*
556 These findings highlight the different nucleosome configuration capacities of promoter
557 sequences in recruiting remodeling activities, which shape gene function.
558 About the exact structure and composition of unstable nucleosomes we can only speculate.
559 In vertebrates, it was shown that histone variants like H3.3 or H2A.Z contribute to nucleosome
560 instability (Jin & Felsenfeld, 2007). Indeed, we observed a partial enrichment of H2A.Z in yeast
561 unstable nucleosomes, but this correlation is not sufficient to account for the stability of most
562 unstable nucleosomes. Furthermore, we used cross-linked nucleosome data datasets, which
563 would overturn the weakened histone DNA interactions observed for histone variants (Xi *et*
564 *al*, 2011).
565 We show that unstable nucleosomes are enriched in sub-nucleosomal fragments and are not
566 limited to wide NDRs, but can also be found at narrow NDRs, where the space is too short to
567 assemble full nucleosomes. These findings indicate the existence of smaller, non-canonical
568 sub-nucleosomal structures, like hexasomes or hemisomes (Chong & Gan, 2023). However,
569 here we face one of the main limitations in MNase-seq data - looking at population wide
570 averages. We cannot exclude the possibility that different nucleosome array configurations
571 exist in individual cells. Here it would be interesting to investigate unstable nucleosome
572 promoters further, using single-molecule DNA sequencing approaches to measure
573 nucleosome positions on individual chromatin fibres (Abdulhay *et al*, 2020).

574 Another possibility might be, that unstable nucleosomes represent sub-nucleosomal
575 intermediate states of the ATP dependent nucleosome remodeling process. It was reported
576 that histones accumulate upstream of the TSS at unstable nucleosome promoters, upon
577 mutation of the FACT subunit SPT16, indicating that those sites undergo active remodeling to
578 promote RNA Pol II transcription through the +1 nucleosome (Obermeyer *et al*, 2023; Michl-
579 Holzinger *et al*, 2022). Furthermore, partially unwrapped nucleosome structures were
580 proposed during INO80 or RSC remodeling resulting in sub-nucleosomal structures (Brahma
581 & Henikoff, 2019; Hsieh *et al*, 2022). In line with these studies, we observed preferential
582 enrichment of RSC components at unstable nucleosome positions upon heat shock and an
583 increase in the number of unstable nucleosomes, by adding nucleosome remodeler to
584 reconstituted nucleosome arrays.
585 In summary, we provide a publicly available pipeline for analyzing nucleosome positions and
586 nucleosomal properties using MNase titrations, which for the first time provides a systematic
587 framework to identify unstable nucleosomes.
588
589
590

591 # METHODS:
592

593 **Public datasets used in this study**
594 *Drosophila melanogaster*:
595 Whole chromatin MNase-seq, MNase-ChIP-H3-seq, MNase-ChIP-H4-seq and RNA-seq
596 (GSE78984), M1BP-ChIP-seq (GSE101554), GRO-seq (GSE58955), and MNase-seq with spike-
597 ins in S2 (GSE128689) were downloaded from the European nucleotide archive (ENA)
598 database. Histone modification ChIP-seq data were retrieved through modENCODE, with the
599 following accession numbers: H3K27ac (296), H3K27me3 (298), H3K4me1 (304), H3K9me3
600 (313), H3K4me3 (3761). DNase-seq (GSE172753) data were retrieved from the GEO database.
601 *Saccharomyces cerevisiae*:
602 Whole chromatin MNase-seq, MNase-ChIP-H4-seq (GSE83123), NET-seq (GSE25107), *in vitro*
603 MNase-seq on SGD assembled nucleosomes (GSE72106), ChIP-exo data of Rsc3, Rsc1, Rsc58,
604 Rsc9, Spt7, Hsf1, Bdf1, CTDSer5, Ssl1, Ssl2, Taf2, Taf3, Tfa1, Tfb1, Tfb2, Tfb3, Tfb4, Tfg1
605 (GSE147927) were downloaded from the European nucleotide archive (ENA) database.

606

607

608 **The nucMACC pipeline**

609 All the steps from fastq files to annotate (sub-)nucMACC scores were integrated and
610 automated within the nucMACC pipeline, which is available on GitHub at
611 uschwartz/nucMACC. The version of the nucMACC pipeline used in this study was v.1.2. The
612 nucMACC pipeline was executed using nextflow, an open-source workflow management
613 system, and runs the software within stable Docker containers, ensuring a reproducible
614 analysis workflow (DI Tommaso *et al*, 2017).

615 Raw sequencing data in the form of fastq files were first subjected to quality control checks
616 using FastQC (Andrews, 2010). Reads were mapped against the reference genome (dm3 /
617 sacCer3) using Bowtie2 with the following parameters --very-sensitive-local --no-discordant
618 (Langmead & Salzberg, 2012). Alignments with mapping quality (MAPQ) below 30 were
619 removed using samtools (Li *et al*, 2009). Next, the alignments were filtered based on their size
620 to categorize them into mono-nucleosomes (140-200 bp) or sub-nucleosomes (< 140 bp)
621 using the alignmentSieve function from the deepTools suite (Ramírez *et al*, 2016).
622 Additionally, any blacklisted regions were removed during this step to ensure that only
623 reliable regions were considered for further analysis. Nucleosome positions were identified
624 using DANPOS2 (Chen *et al*, 2013). For this, pooled mono-nucleosome samples were used to
625 call the positions of mono-nucleosomes, while sub-nucleosome positions were called from
626 samples with the lowest MNase concentration. Here library size was normalized to the
627 effective genome size, paired-end reads centered to their midpoint and extended to 70 bp to
628 obtain nucleosome-sized positions. The GC content at the identified nucleosome positions
629 was measured using bedtools genomecov function (Quinlan & Hall, 2010). To quantify the
630 fragments associated with each nucleosome position, the featureCounts tool was employed
631 (Liao *et al*, 2014). Nucleosome positions with low fragment counts were filtered out to ensure
632 robustness in the downstream analysis. Specifically, mono-nucleosome positions with
633 fragment count less than 30 and sub-nucleosome positions with counts less than 5 were
634 removed.

635

636 **nucMACC score calculation**

637 First the fragment counts were normalized based on the library size to obtain counts per
638 million (CPM). A pseudocount was determined as the median of the sample median count. A
639 linear regression analysis was performed between log2-transformed MNase concentration
640 and the log2-transformed normalized counts with the added pseudocount. The raw nucMACC
641 scores were estimated from the slope obtained from the regression analysis multiplied by
642 minus one. Next, the raw nucMACC scores were corrected for the underlying GC content.
643 Nucleosome positions exhibiting rarely occurring extreme GC content were filtered out. A
644 locally weighted scatterplot smoothing (LOESS) fit was calculated for the filtered GC content
645 and corresponding raw nucMACC scores using a smoothing parameter α of 0.1. Raw
646 nucMACC scores were subtracted by the deviation of the loess fit from the median raw
647 nucMACC score at the corresponding GC content. The sub-nucMACC scores were determined
648 in the same way, except that the slope values were not multiplied by minus one to represent
649 the stability scores.
650
651 **Identification of hypo-/hyper-accessible nucleosomes**
652 To identify hypo- and hyper-accessible nucleosomes, the nucMACC scores were normalized
653 to a deviation from the highest to lowest value of 1 and plotted against the ranks of the
654 nucMACC scores, which were normalized by the total number of ranks. A LOESS smoothing
655 was applied and the first derivate of the LOESS fit was calculated to deduce the slope of the
656 curve. Hypo-accessible nucleosomes are determined as the lowest nucMACC scores until the
657 slope of the curve falls below 1 and hyper-accessible nucleosomes are the highest nucMACC
658 scores after the slope of the curve exceeds again a slope of 1.
659
660 **Identification of unstable and non-canonical nucleosomes**
661 Unstable and non-canonical nucleosomes were similarly determined as hyper-/hypo-
662 accessible nucleosomes, except that sub-nucleosome positions are selected, which either are
663 unique or exhibit at least four-fold more normalized counts in the lowest MNase condition
664 compared to the mono-nucleosomes. Unstable nucleosomes are determined as the lowest
665 sub-nucMACC scores until the slope of the curve falls below 1 and non-canonical nucleosomes
666 are the highest sub-nucMACC scores after the slope of the curve exceeds again a slope of 1.
667
668 **Spike-in analysis**

669 We compared the results of the nucMACC pipeline ignoring the spike-in information with a
670 slightly modified version using the spike-ins for normalization. Therefore, reads were aligned
671 to the sacCer3 *Saccharomyces cerevisiae* genome and treated analogously to the nucMACC
672 pipeline. The number of fragments per mono- and sub-nucleosome fraction was counted, and
673 normalization factors for the nucMACC analysis in *D.melanogaster* were calculated by dividing
674 the number of *S. cerevisiae* derived fragments by 1 Mio.
675
676 **Downstream analysis**
677 The code used to generate the figures in this manuscript can be found in the project GitHub
678 repository uschwartz/nucMACC_Paper. Density correlation scatterplots were generated
679 using the LSD R package (Schwalb *et al*, 2020). Heatmaps and metaplots were calculated using
680 the computeMatrix, plotHeatmap, plotProfile functions of the deeptools suit (Ramírez *et al*,
681 2016) and further modified in R using basic plot functions or the pheatmap package (Kolde,
682 2019). Genomic features were annotated using the Bioconductor package ChIPseeker (Yu *et*
683 *al*, 2015). Motif enrichment analysis was carried out using the findMotifs.pl script from the
684 HOMER software suite (Heinz *et al*, 2010). Motifs from the integrated fly database were
685 tested for enrichment at the TSS +/- 100 bp of genes associated either with an unstable
686 nucleosome or a nucleosome depleted region. The DNA sequence around the TSS of all
687 expressed genes served as background.
688 Gene set over-representation analysis was carried out using the Metascape online tool (Zhou
689 *et al*, 2019). Enrichment was performed using the Gene Ontology (GO) database for biological
690 processes (BP) and restricted to terms with at least 20 genes overlapping to the tested gene
691 set.
692
693 **GRO-seq / NET-seq analysis**
694 GRO- and NET-seq data have been analyzed using the in-house nextflow transcriptomics
695 pipeline available at GitHub (uschwartz/RNAseq).
696 Initially, quality control of the raw sequence reads was conducted using FastQC (v0.11.8)
697 (Andrews, 2010). Subsequently, reads were mapped to the reference genome and
698 corresponding gene annotation using the Spliced Transcripts Alignment to a Reference (STAR)
699 software (version v2.7.8a)(Dobin *et al*, 2013). The following options were used to optimize
700 the alignment process: --outFilterType BySJout, --outFilterMultimapNmax 20, --

701 alignSJoverhangMin 8, --alignSJDBoverhangMin 1, --outFilterMismatchNmax 999, --
702 alignIntronMin 10, --alignIntronMax 1000000, --outFilterMismatchNoverReadLmax 0.04, --
703 runThreadN 12, --outSAMtype BAM SortedByCoordinate, --outSAMmultNmax 1, and --
704 outMultimapperOrder Random.
705 Post-mapping quality control was performed using the rnaseq analysis mode of Qualimap
706 (v2.2.1)(García-Alcalde *et al*, 2012). The level of PCR duplication was assessed using Picard
707 MarkDuplicates (v2.21.8) (Broad Institute, 2019), and dupRadar (version v1.15.0) (Sayols *et*
708 *al*, 2016). Gene expression quantification was carried out using featureCounts (v1.6.3) (Liao
709 *et al*, 2014).
710
711
712 **ChIP-exo data processing**
713 ChIP-exo sequencing data in *S. cerevisiae* was downloaded and trimmed with TrimGalore
714 v0.6.7 (github.com/FelixKrueger/TrimGalore). Trimmed reads were aligned to the sacCer3
715 reference genome using bowtie2 v2.4.2 (Langmead & Salzberg, 2012) with the following
716 options "--very-sensitive -I 0 -X 500 --fr --no-mixed --no-discordant –dovetail". Duplicate reads
717 were identified and removed using picard v2.18.7 (Broad Institute, 2019). Bigwigs mimicking
718 standard ChIP were generated using the bamCoverage function from deeptools v3.5.1
719 (Ramírez *et al*, 2016) with the options: "-bs 10 --normalizeUsing CPM --extendReads --
720 centerReads --minMappingQuality 1".
721

722 # ACKNOWLEDGEMENTS:

723
724 The study was financed by the DFG SFB960 (to GL and US). We thank Philipp Korber for
725 fruitful discussions and Joachim Griesenbeck for helpful comments on this manuscript.
726

727 # AUTHOR CONTRIBUTIONS:

728
729 ***Wernig-Zorc Sara:*** Methodology, Validation, Formal analysis, Investigation, Visualization,
730 Writing - Review & Editing ***Kugler Fabian:*** Methodology, Software, Validation, Formal
731 analysis, Investigation, Visualization ***Schmutterer Leo:*** Software, Formal analysis,
732 Investigation ***Räß Patrick:*** Software ***Hausmann Clemens:*** Software ***Holzinger Simon:***
733 Methodology, Formal analysis, Writing - Review & Editing ***Längst Gernot:*** Conceptualization,

Funding Acquisition, Resources, Project administration, Supervision, Writing - Review & Editing ***Schwartz Uwe:*** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing - Original Draft, Supervision, Visualization

DISCLOSURE AND COMPETING INTERESTS STATEMENT

The authors declare no competing interests.

REFERENCES

{Bibliography}

Abdulhay NJ, McNally CP, Hsieh LJ, Kasinathan S, Keith A, Estes LS, Karimzadeh M, Underwood JG, Goodarzi H, Narlikar GJ, *et al* (2020) Massively multiplex single-molecule oligonucleosome footprinting. *Elife* 9: 1–23

Andrews S (2010) FASTQC. A quality control tool for high throughput sequence data. (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) [PREPRINT]

Baldi S, Korber P & Becker PB (2020) Beads on a string—nucleosome array arrangements and folding of the chromatin fiber. *Nat Struct Mol Biol* 27: 109–118 doi:10.1038/s41594-019-0368-x [PREPRINT]

Brahma S & Henikoff S (2019) RSC-Associated Subnucleosomes Define MNase- Sensitive Promoters in Yeast Article RSC-Associated Subnucleosomes Define MNase-Sensitive Promoters in Yeast. *Mol Cell*: 1–12

Broad Institute (2019) Picard toolkit.

Buenrostro JD, Giresi PG, Zaba LC, Chang HY & Greenleaf WJ (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* 10: 1213–1218

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

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

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

Chereji R V. & Clark DJ (2018) Major Determinants of Nucleosome Positioning. *Biophys J* 114: 2279–2289

Chereji R V., Ocampo J & Clark DJ (2017) MNase-Sensitive Complexes in Yeast: Nucleosomes and Non-histone Barriers. *Mol Cell* 65: 565-577.e3

Chereji V, Kan T, Grudniewska MK, Romashchenko V, Berezikov E, Zhimulev IF, Guryev V, Morozov V & Moshkin YM (2015) Genome-wide profiling of nucleosome sensitivity and chromatin accessibility in Drosophila melanogaster. 1–16

Chong CYY & Gan L (2023) Are extraordinary nucleosome structures more ordinary than we thought? *Chromosoma* 132

Churchman LS & Weissman JS (2011) Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature 2011 469:7330* 469: 368–373

Compton JL, Bellard M & Chambon P (1976) Biochemical evidence of variability in the DNA repeat length in the chromatin of higher eukaryotes. *Proceedings of the National Academy of Sciences* 73: 4382–4386

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

Davey CA, Sargent DF, Luger K, Maeder AW & Richmond TJ (2002) Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution. *J Mol Biol* 319: 1097–1113

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

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

Fei J, Torigoe SE, Brown CR, Khuong MT, Kassavetis GA, Boeger H & Kadonaga JT (2015) The prenucleosome, a stable conformational isomer of the nucleosome. *Genes Dev* 29: 2563–2575

Fuda NJ, Guertin MJ, Sharma S, Danko CG, Martins AL, Siepel A & Lis JT (2015) GAGA Factor Maintains Nucleosome-Free Regions and Has a Role in RNA Polymerase II Recruitment to Promoters. *PLoS Genet* 11: e1005108

García-Alcalde F, Okonechnikov K, Carbonell J, Cruz LM, Götz S, Tarazona S, Dopazo J, Meyer TF & Conesa A (2012) Qualimap: evaluating next-generation sequencing alignment data. *Bioinformatics* 28: 2678–2679

Hasson D, Panchenko T, Salimian KJ, Salman MU, Sekulic N, Alonso A, Warburton PE & Black BE (2013) The octamer is the major form of CENP-A nucleosomes at human centromeres. *Nature Structural & Molecular Biology 2013 20:6* 20: 687–695

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H & Glass CK (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 38: 576–89

Henikoff JG, Belsky JA, Krassovsky K, Macalpine DM & Henikoff S (2011) Epigenome characterization at single base-pair resolution.

Hsieh LJ, Gourdet MA, Moore CM, Muñoz EN, Gamarra N, Ramani V & Narlikar GJ (2022) A hexasome is the preferred substrate for the INO80 chromatin remodeling complex, allowing versatility of function. *Mol Cell* 82: 2098-2112.e4

Hu G, Schones DE, Cui K, Ybarra R, Northrup D, Tang Q, Gattinoni L, Restifo NP, Huang S & Zhao K (2011) Regulation of nucleosome landscape and transcription factor targeting at tissue-specific enhancers by BRG1. *Genome Res* 21: 1650–1658

Ishii H, Kadonaga JT & Ren B (2015) MPE-seq , a new method for the genome-wide analysis of chromatin structure. 1–9

Iwafuchi-Doi M, Donahue G, Kakumanu A, Watts JA, Mahony S, Pugh BF, Lee D, Kaestner KH & Zaret KS (2016) The Pioneer Transcription Factor FoxA Maintains an Accessible Nucleosome Configuration at Enhancers for Tissue-Specific Gene Activation. *Mol Cell* 62: 79–91

Iwafuchi-doi M & Zaret KS (2014) Pioneer transcription factors in cell reprogramming. 2679–2692

824 Jeffers TE & Lieb JD (2017) Nucleosome fragility is associated with future transcriptional
825 response to developmental cues and stress in C. Elegans. *Genome Res* 27: 75–86
826 Jin C & Felsenfeld G (2007) Nucleosome stability mediated by histone variants H3.3 and
827 H2A.Z. *Genes Dev* 21: 1519–1529
828 Kaplan N, Moore IK, Fondufe-Mittendorf Y, Gossett AJ, Tillo D, Field Y, LeProust EM, Hughes
829 TR, Lieb JD, Widom J, *et al* (2009) The DNA-encoded nucleosome organization of a
830 eukaryotic genome. *Nature* 458: 362–366
831 Kireeva ML, Walter W, Tchernajenko V, Bondarenko V, Kashlev M & Studitsky VM (2002)
832 Nucleosome remodeling induced by RNA polymerase II: loss of the H2A/H2B dimer
833 during transcription. *Mol Cell* 9: 541–552
834 Kolde R (2019) pheatmap: Pretty Heatmaps. (https://CRAN.R-
835 project.org/package=pheatmap) [PREPRINT]
836 Kornberg RD (1974) Chromatin structure: A repeating unit of histones and DNA. *Science*
837 *(1979)* 184: 868–871
838 Krietenstein N, Wal M, Watanabe S, Park B, Peterson CL, Pugh BF & Korber P (2016)
839 Genomic Nucleosome Organization Reconstituted with Pure Proteins. *Cell* 167: 709-
840 721.e12
841 Kubik S, Bruzzone MJ, Jacquet P, Falcone JL, Rougemont J & Shore D (2015) Nucleosome
842 Stability Distinguishes Two Different Promoter Types at All Protein-Coding Genes in
843 Yeast. *Mol Cell* 60: 422–434
844 Langmead B & Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods*
845 9: 357–9
846 Lavelle C & Prunell A (2007) Chromatin polymorphism and the nucleosome superfamily: a
847 genealogy. *Cell Cycle* 6: 2113–2119
848 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G & Durbin R
849 (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078–9
850 Li J & Gilmour DS (2013) Distinct mechanisms of transcriptional pausing orchestrated by
851 GAGA factor and M1BP, a novel transcription factor. *EMBO Journal* 32: 1829–1841
852 Liao Y, Smyth GK & Shi W (2014) FeatureCounts: An efficient general purpose program for
853 assigning sequence reads to genomic features. *Bioinformatics* 30: 923–930
854 Liu Y, Zhou K, Zhang N, Wei H, Tan YZ, Zhang Z, Carragher B, Potter CS, D'Arcy S & Luger K
855 (2019) FACT caught in the act of manipulating the nucleosome. *Nature 2019 577:7790*
856 577: 426–431
857 Luger K, Mäder AW, Richmond RK, Sargent DF & Richmond TJ (1997) Crystal structure of the
858 nucleosome core particle at 2.8 Å resolution. *Nature* 389: 251–260
859 Marr LT, Ocampo J, Clark DJ & Hayes JJ (2021) Global histone protein surface accessibility in
860 yeast indicates a uniformly loosely packed genome with canonical nucleosomes.
861 *Epigenetics Chromatin* 14
862 Michael AK, Grand RS, Isbel L, Cavadini S, Kozicka Z, Kempf G, Bunker RD, Schenk AD, Graff-
863 Meyer A, Pathare GR, *et al* (2020) Mechanisms of OCT4-SOX2 motif readout on
864 nucleosomes. *Science (1979)* 368
865 Michl-Holzinger P, Obermeyer S, Markusch H, Pfab A, Ettner A, Bruckmann A, Babl S, Längst
866 G, Schwartz U, Tvardovskiy A, *et al* (2022) Phosphorylation of the FACT histone
867 chaperone subunit SPT16 affects chromatin at RNA polymerase II transcriptional start
868 sites in Arabidopsis. *Nucleic Acids Res* 50: 5014–5028

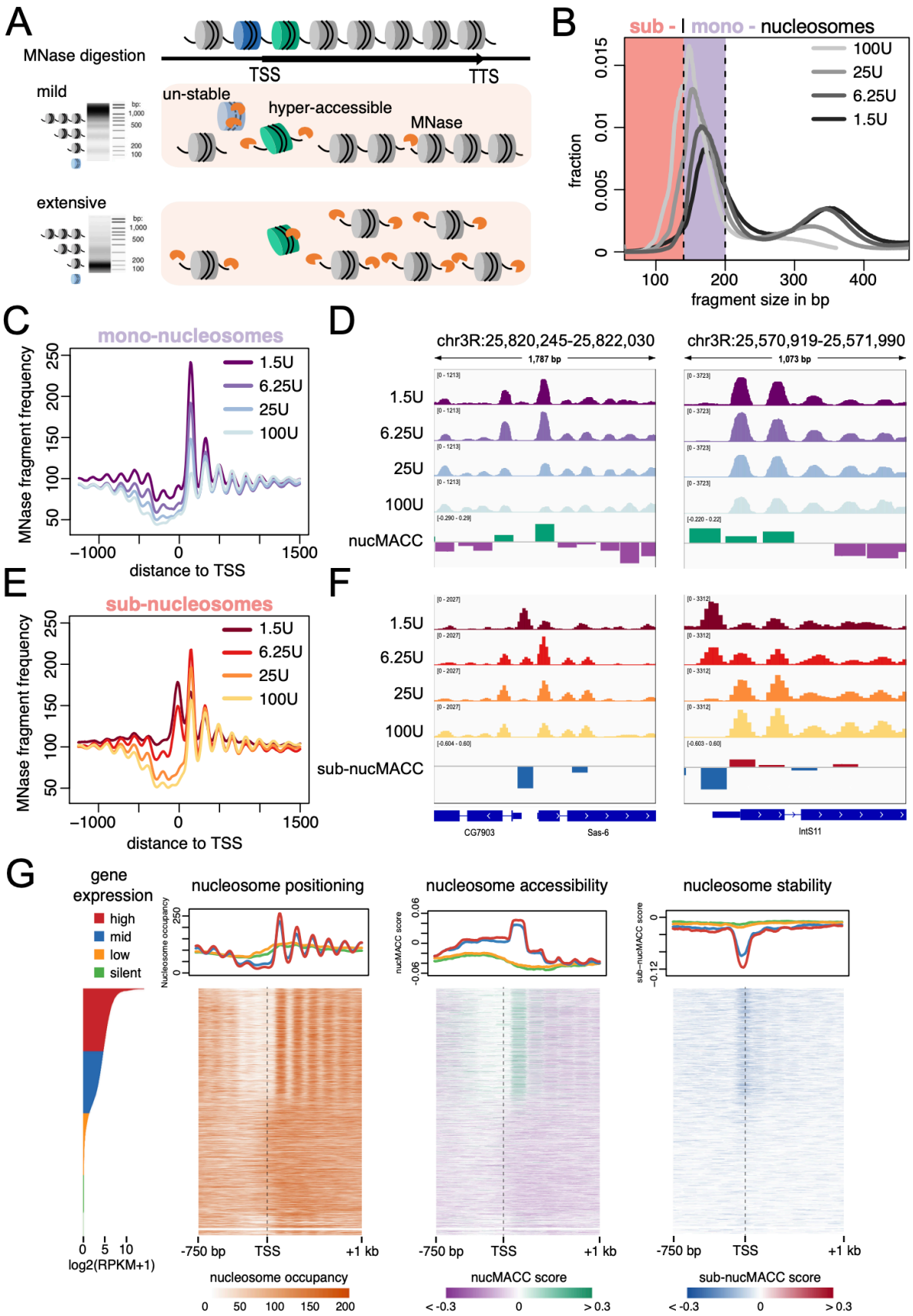
869 Mieczkowski J, Cook A, Bowman SK, Mueller B, Alver BH, Kundu S, Deaton AM, Urban JA,
870 Larschan E, Park PJ, *et al* (2016) MNase titration reveals differences between
871 nucleosome occupancy and chromatin accessibility. *Nat Commun* 7: 11485
872 Obermeyer S, Schrettenbrunner L, Stöckl R, Schwartz U & Grasser KD (2023) Different
873 elongation factors distinctly modulate RNA polymerase II transcription in Arabidopsis.
874 *Nucleic Acids Res* 1: 13–14
875 Olins AL, Douglas Carlson R & Olins DE (1975) Visualization of chromatin substructure: ν
876 bodies. *Journal of Cell Biology* 64: 528–537
877 Prasad R, D'Arcy S, Hada A, Luger K & Bartholomew B (2016) Coordinated Action of Nap1
878 and RSC in Disassembly of Tandem Nucleosomes. *Mol Cell Biol* 36: 2262–2271
879 Prunell A & Kornberg RD (1982) Variable center to center distance of nucleosomes in
880 chromatin. *J Mol Biol* 154: 515–523
881 Quinlan AR & Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic
882 features. *Bioinformatics* 26: 841–842
883 Ramachandran S, Ahmad K & Henikoff S (2017) Transcription and Remodeling Produce
884 Asymmetrically Unwrapped Nucleosomal Intermediates. *Mol Cell* 68: 1038-1053.e4
885 Ramachandran S & Henikoff S (2016) Nucleosome dynamics during chromatin remodeling in
886 vivo. *Nucleus* 7: 20–26
887 Ramírez F, Ryan DP, Bj¨ B, Grüning B, Grüning G, Bhardwaj V, Kilpert F, Richter AS, Heyne S,
888 Undar FD¨, *et al* (2016) deepTools2: a next generation web server for deep-sequencing
889 data analysis. *Web Server issue Published online* 44
890 Rhee HS, Bataille AR, Zhang L & Pugh BF (2014) Subnucleosomal structures and nucleosome
891 asymmetry across a genome. *Cell* 159: 1377–1388
892 Rossi MJ, Kuntala PK, Lai WKM, Yamada N, Badjatia N, Mittal C, Kuzu G, Bocklund K, Farrell
893 NP, Blanda TR, *et al* (2021) A high-resolution protein architecture of the budding yeast
894 genome. *Nature 2021 592:7853* 592: 309–314
895 Sato S, Tanaka N, Arimura Y, Kujirai T & Kurumizaka H (2020) The N-terminal and C-terminal
896 halves of histone H2A.Z independently function in nucleosome positioning and
897 stability. *Genes Cells* 25: 538–546
898 Sayols S, Scherzinger D & Klein H (2016) dupRadar: A Bioconductor package for the
899 assessment of PCR artifacts in RNA-Seq data. *BMC Bioinformatics* 17: 1–5
900 Schones DE, Cui K, Cuddapah S, Roh TY, Barski A, Wang Z, Wei G & Zhao K (2008) Dynamic
901 Regulation of Nucleosome Positioning in the Human Genome. *Cell* 132: 887–898
902 Schwalb B, Tresch A, Torkler P, Duemcke S, Demel C, Ripley B & Venables B (2020) LSD: Lots
903 of Superior Depictions. (https://CRAN.R-project.org/package=LSD) [PREPRINT]
904 Schwartz U, Komatsu T, Huber C, Lagadec F, Baumgartl C, Silberhorn E, Nuetzel M, Rayne F,
905 Basyuk E, Bertrand E, *et al* (2023) Changes in adenoviral chromatin organization
906 precede early gene activation upon infection. *EMBO J* 42: e114162
907 Schwartz U, Németh A, Diermeier S, Exler JH, Hansch S, Maldonado R, Heizinger L, Merkl R &
908 Längst G (2018) Characterizing the nuclease accessibility of DNA in human cells to map
909 higher order structures of chromatin. *Nucleic Acids Res* 47: 1239–1254
910 Struhl K & Segal E (2013) Determinants of nucleosome positioning. *Nature Publishing Group*
911 20: 267–273
912 Tachiwana H, Kagawa W, Osakabe A, Kawaguchi K, Shiga T, Hayashi-Takanaka Y, Kimura H &
913 Kurumizaka H (2010) Structural basis of instability of the nucleosome containing a
914 testis-specific histone variant, human H3T. *Proc Natl Acad Sci U S A* 107: 10454–10459

915 Talbert PB & Henikoff S (2017) Histone variants on the move: substrates for chromatin
916 dynamics. *Nat Rev Mol Cell Biol* 18: 115–126
917 Teif VB, Vainshtein Y, Caudron-herger M, Mallm J, Marth C, Höfer T & Rippe K (2012)
918 Genome-wide nucleosome positioning during embryonic stem cell development. *Nat*
919 *Struct Mol Biol* 19: 1185–1192
920 DI Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E & Notredame C (2017) Nextflow
921 enables reproducible computational workflows. *Nature Biotechnology 2017 35:4* 35:
922 316–319
923 Urahama T, Harada A, Maehara K, Horikoshi N, Sato K, Sato Y, Shiraishi K, Sugino N, Osakabe
924 A, Tachiwana H, *et al* (2016) Histone H3.5 forms an unstable nucleosome and
925 accumulates around transcription start sites in human testis. *Epigenetics Chromatin* 9
926 Valouev A, Johnson SM, Boyd SD, Smith CL, Fire AZ & Sidow A (2011) Determinants of
927 nucleosome organization in primary human cells. *Nature* 474: 516–520
928 Vierstra J, Wang H, John S, Sandstrom R & Stamatoyannopoulos JA (2014) Coupling
929 transcription factor occupancy to nucleosome architecture with DNase-FLASH. 11
930 Voong LN, Xi L, Sebeson AC, Xiong B, Wang J-P & Wang X (2016) Insights into Nucleosome
931 Organization in Mouse Embryonic Stem Cells through Chemical Mapping. *Cell* 167:
932 1555-1570.e15
933 Weiner A, Hughes A, Yassour M, Rando OJ & Friedman N (2010) High-resolution nucleosome
934 mapping reveals transcription-dependent promoter packaging. *Genome Res* 20: 90–100
935 Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, Rahl PB, Lee TI & Young RA
936 (2013) Master transcription factors and mediator establish super-enhancers at key cell
937 identity genes. *Cell* 153: 307–319
938 Xi Y, Yao J, Chen R, Li W & He X (2011) Nucleosome fragility reveals novel functional states
939 of chromatin and poises genes for activation. *Genome Res* 21: 718–724
940 Yu G, Wang LG & He QY (2015) ChIP seeker: An R/Bioconductor package for ChIP peak
941 annotation, comparison and visualization. *Bioinformatics* 31
942 Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C & Chanda SK
943 (2019) Metascape provides a biologist-oriented resource for the analysis of systems-
944 level datasets. *Nat Commun* 10
945 Zouaz A, Auradkar A, Delfini MC, Macchi M, Barthez M, Akoa SE, Bastianelli L, Xie G, Deng
946 W-M, Levine SS, *et al* (2017) The Hox proteins Ubx and AbdA collaborate with the
947 transcription pausing factor M1BP to regulate gene transcription. *EMBO J* 36: 2887–
948 2906
949

950

951

Figures and Legends:



955

956 **Figure 1**

957 (A) Schematic representation of differential MNase chromatin digestion. Under mild MNase
958 conditions, chromatin is partially digested and nucleosomes (green, hyper-accessible) from
959 accessible regions are preferentially released (upper panel). At extensive MNase conditions,
960 chromatin is completely digested to nucleosome core particles and MNase-sensitive
961 nucleosomes (blue, unstable) are lost. (B) Fragment size distribution of MNase titrations.
962 Fragments are selected based on the fragment size and grouped into sub- (< 140 bp, light red
963 section) or mono-nucleosomes (140-200 bp, purple section). (C) and (E) average fragment
964 frequencies at the TSS. (D) and (J) Genome browser snapshot showing the output of the
965 nucMACC pipeline. nucMACC scores are deduced from mono-nucleosomal fragments (D) and
966 sub-nucMACC scores from sub-nucleosomal fragments, respectively (J). (E) Heatmaps sorted
967 by gene expression showing nucleosome positioning (left), accessibility (middle, nucMACC
968 scores), and stability (right, sub-nucMACC scores) at TSS.

969

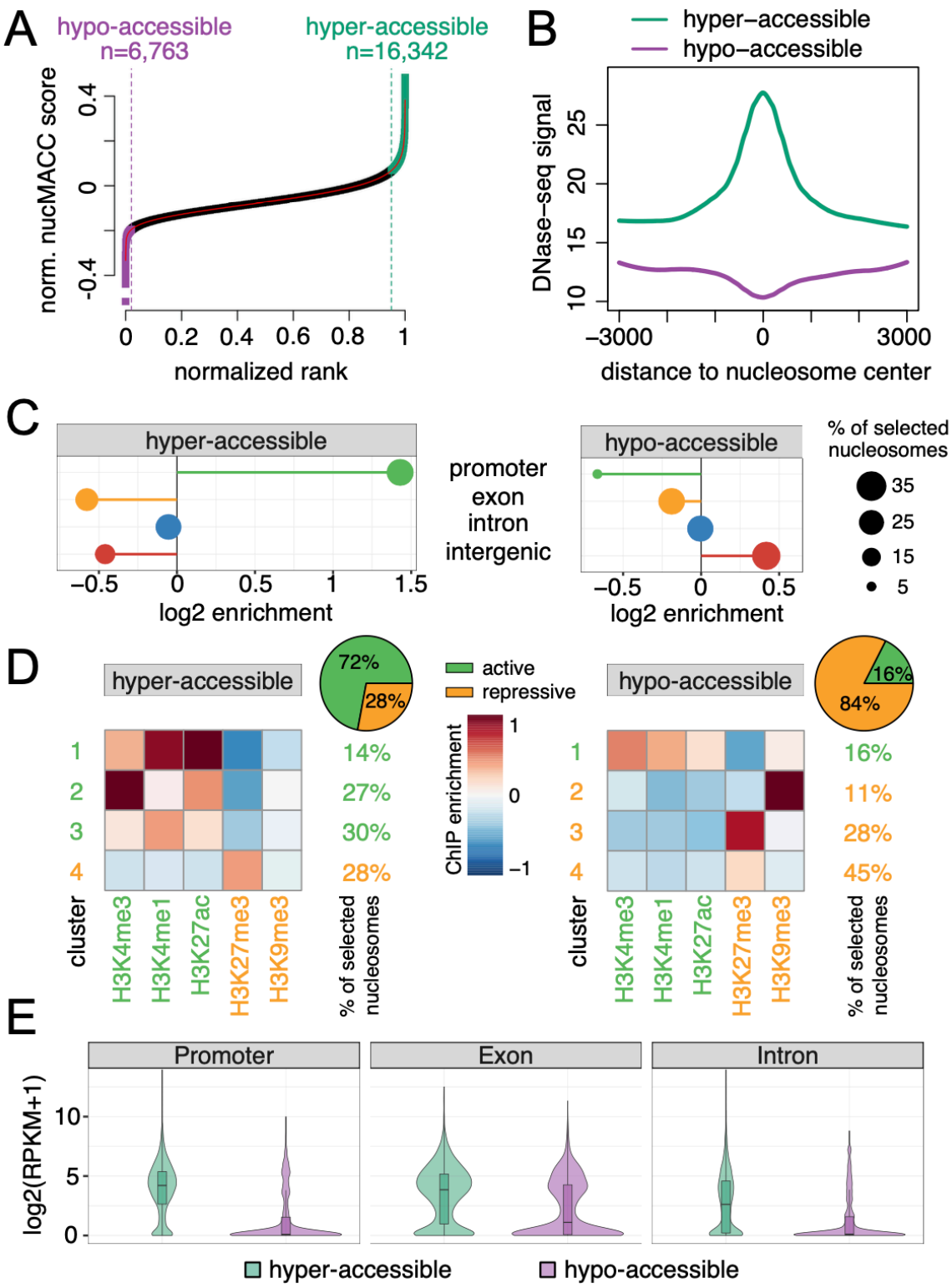
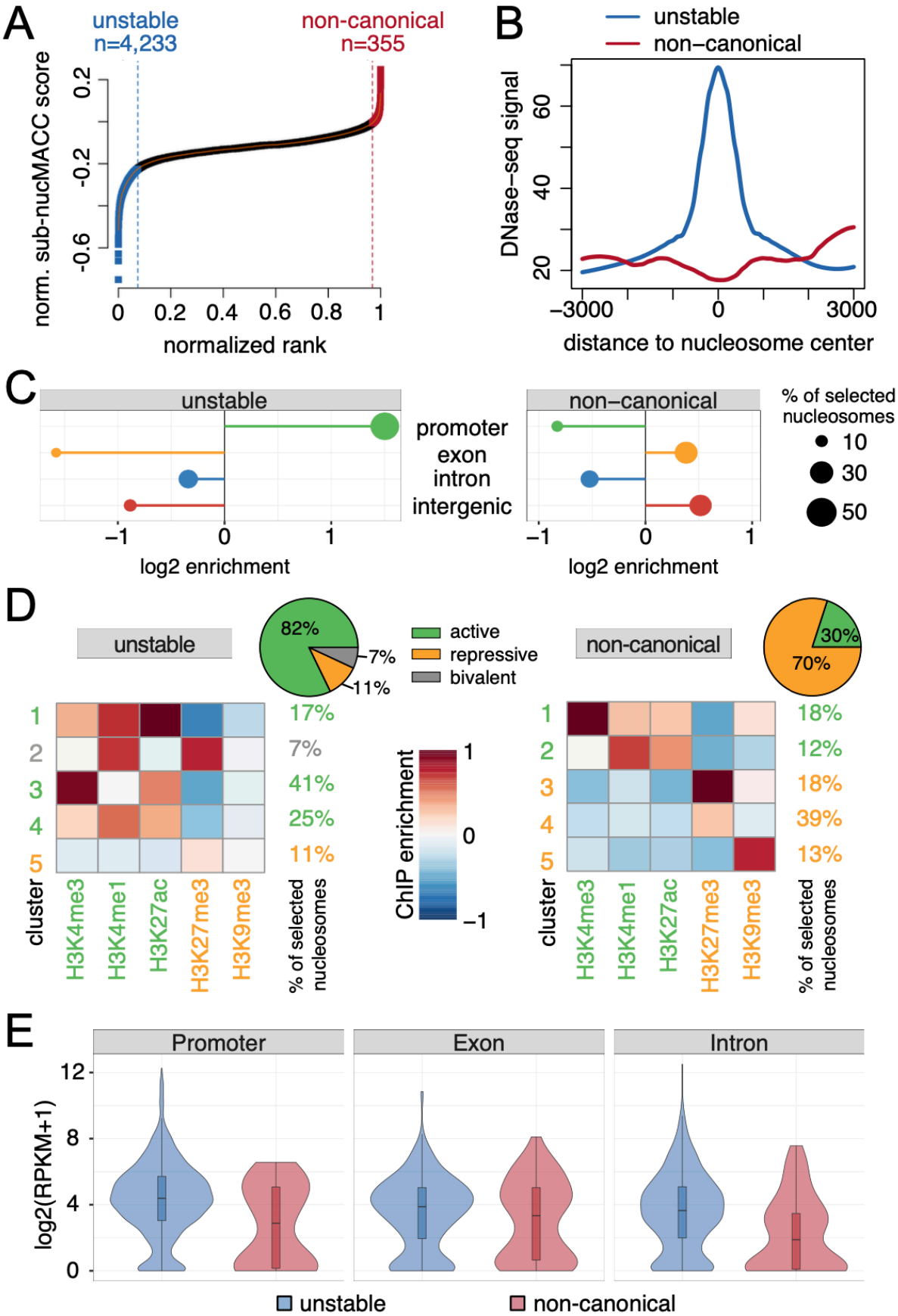


Figure 2

(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

976 before the point, where the slope goes the first time below 1 (purple dashed line) were
977 classified as hypo-accessible. Nucleosomes after the point where the slope of the curve
978 exceeds one again (green dashed line), were classified as hyper-accessible nucleosomes. (B)
979 Correlation of DNase-seq sites with hyper-/hypo-accessible nucleosomes. (C) Enrichment of
980 hyper-/hypo-accessible nucleosomes at genomic features. (D) K-means clustering (n=4) of
981 histone modifications at hyper-/hypo-accessible sites. Centroid values of each cluster are
982 indicated by the color code. The relative number of nucleosomes represented in each cluster
983 is indicated on the right side. Repressive histone modifications are highlighted in orange and
984 active in green, respectively. The pie chart depicts the fraction of nucleosomes assigned to
985 either active or repressive histone clusters. (E) Gene expression levels of genes exhibiting
986 hyper-/hypo-accessible nucleosomes at distinct features. Promoter was defined as 500 bp
987 upstream of the TSS to 300 bp downstream.
988
989


A
unstable
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
unstable
non–canonical
DNase–seq signal
60
40
20
–3000
0
3000
distance to nucleosome center
C
unstable
non–canonical
promoter
exon
intron
intergenic
–1
0
1
log2 enrichment
% of selected nucleosomes
10
30
50
D
unstable
82%
7%
11%
active
repressive
bivalent
non-canonical
30%
70%
1
2
3
4
5
cluster
H3K4me3
H3K4me1
H3K27ac
H3K27me3
H3K9me3
% of selected nucleosomes
17%
7%
41%
25%
11%
ChIP enrichment
1
0
–1
18%
12%
18%
39%
13%
E
Promoter
Exon
Intron
log2(RPKM+1)
12
8
4
0
unstable
non-canonical

990

991

992 **Figure 3**

993 (A) Selection of unstable (blue) and non-canonical (red) nucleosomes based on nucMACC
994 scores. Nucleosomes were ranked by sub-nucMACC score. X- and y-axis were normalized to a
995 range of one. Cut-offs were deduced from the slope of the LOESS curve fit. Nucleosomes
996 before the point, where the slope goes the first time below 1 (blue dashed line) were classified
997 as unstable. Nucleosomes after the point where the slope of the curve exceeds one again (red
998 dashed line), were classified as non-canonical nucleosomes. (B) Correlation of DNase-seq sites
999 with unstable and non-canonical nucleosomes. (C) Enrichment of unstable or non-canonical
1000 nucleosomes at genomic features. (D) K-means clustering (n=5) of histone modifications at
1001 unstable or non-canonical sites. Centroid values of each cluster are indicated by the color
1002 code. The relative number of nucleosomes represented in each cluster is indicated on the
1003 right side. Repressive histone modifications are highlighted in orange, active in green,
1004 respectively. The bivalent cluster enriched for both repressive and active histone
1005 modifications is indicated in grey. The pie chart depicts the fraction of nucleosomes assigned
1006 to either active, repressive, or bivalent histone clusters. (E) Gene expression levels of genes
1007 exhibiting unstable or non-canonical nucleosomes at distinct features. Promoter was defined
1008 as 500 bp upstream of the TSS to 300 bp downstream.

1009

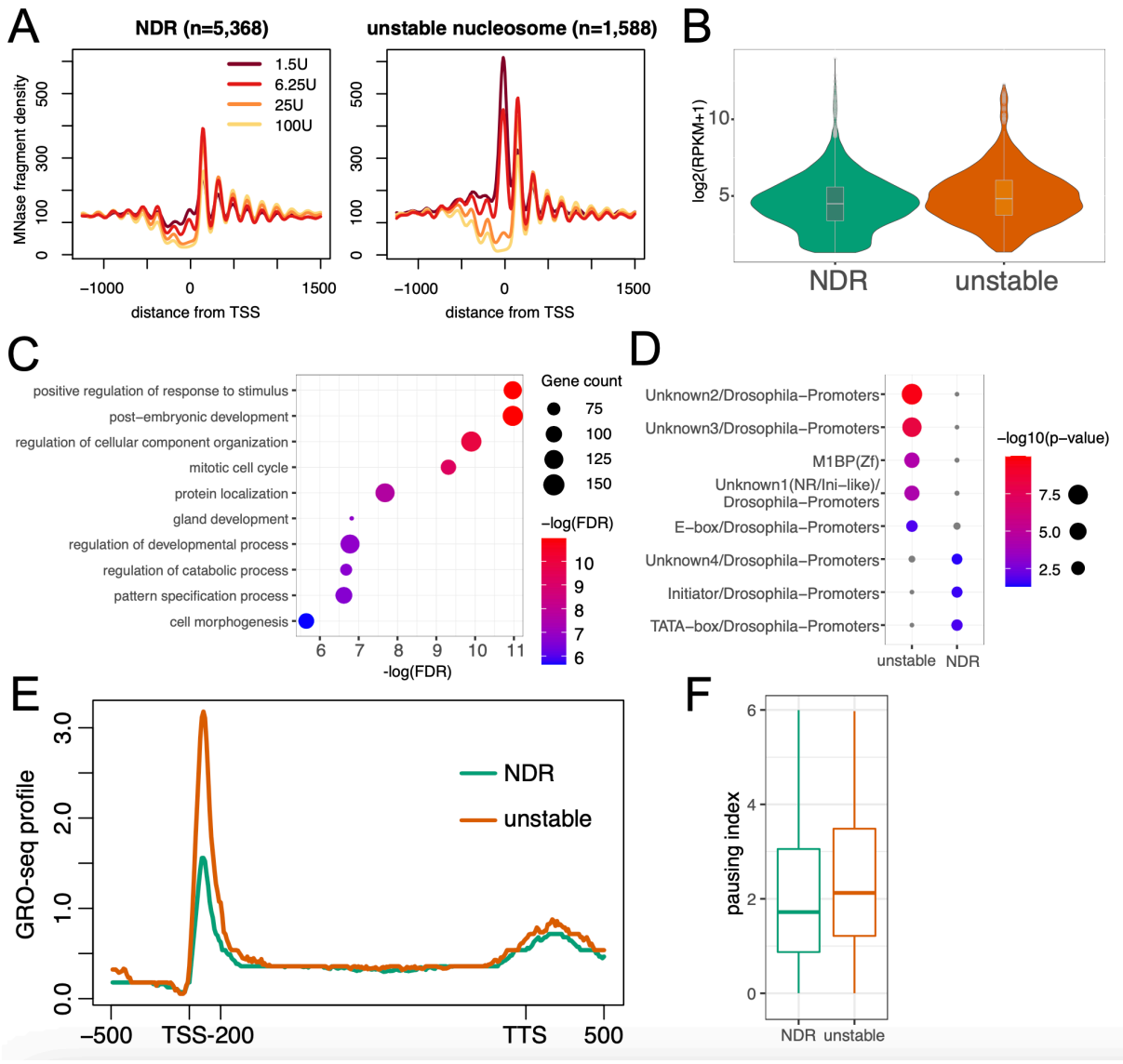
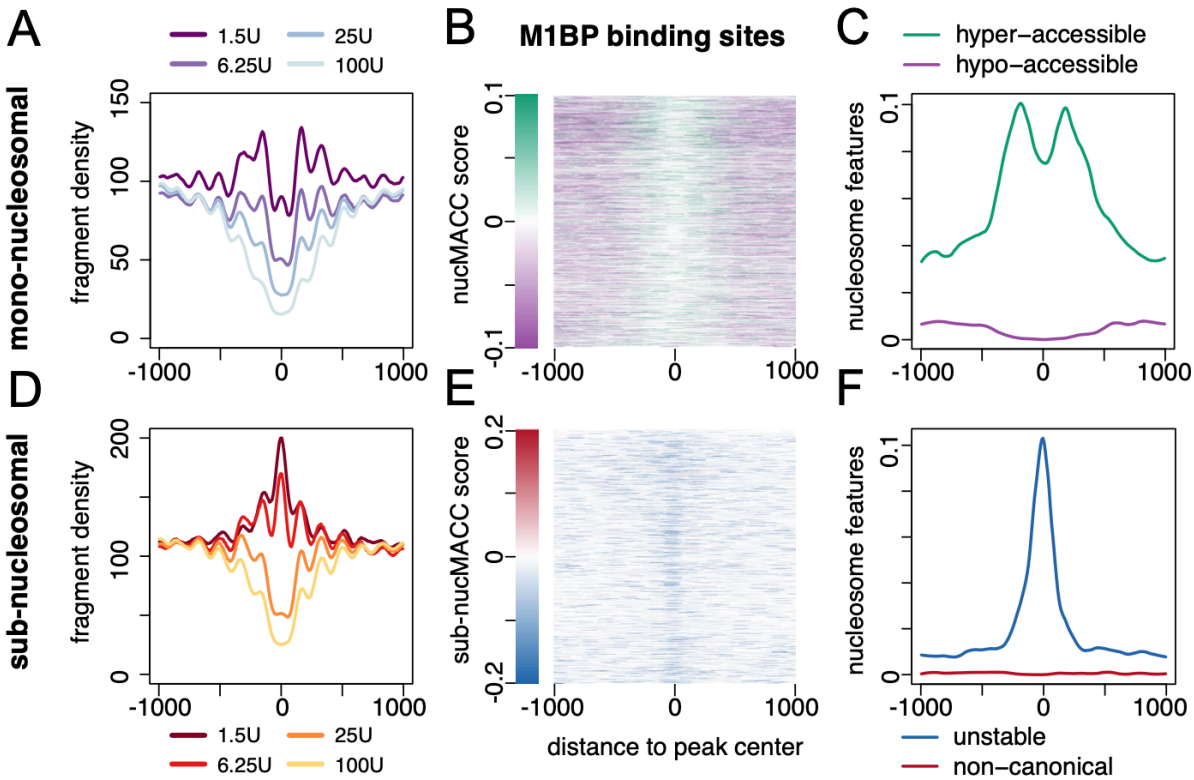


Figure 4

(A) Average fragment frequencies at the TSS of expressed genes either containing an unstable 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 unstable 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 unstable nucleosome promoter. (D) Motif enrichment analysis in the unstable nucleosome (left) or NDR (right) promoter. (E) Median GRO-seq profile showing nascent RNA abundance. Regions upstream of TSS, downstream of TTS and the first 200 bp downstream of TSS are unscaled. The remaining gene body is scaled to have the same length for each gene. (F) Boxplots showing the pausing index

1023 based on the GRO-seq signal. Pausing index was calculated as the fold change of the GRO-seq
1024 signal in the first 200 bp downstream of the TSS versus the remaining gene body.
1025
1026



1027
1028
1029 **Figure 5**
1030 (A) and (D) average fragment frequencies of mono- (A) and sub-nucleosomal (D) fragments at
1031 M1BP bound sites. (B) and (E) Heatmap showing nucMACC (B) or sub-nucMACC (E) scores at
1032 M1BP bound sites. (C) and (F) Distribution of hyper-/hypo-accessible (C) or unstable and non-
1033 canonical (F) nucleosomes at M1BP bound sites.
1034

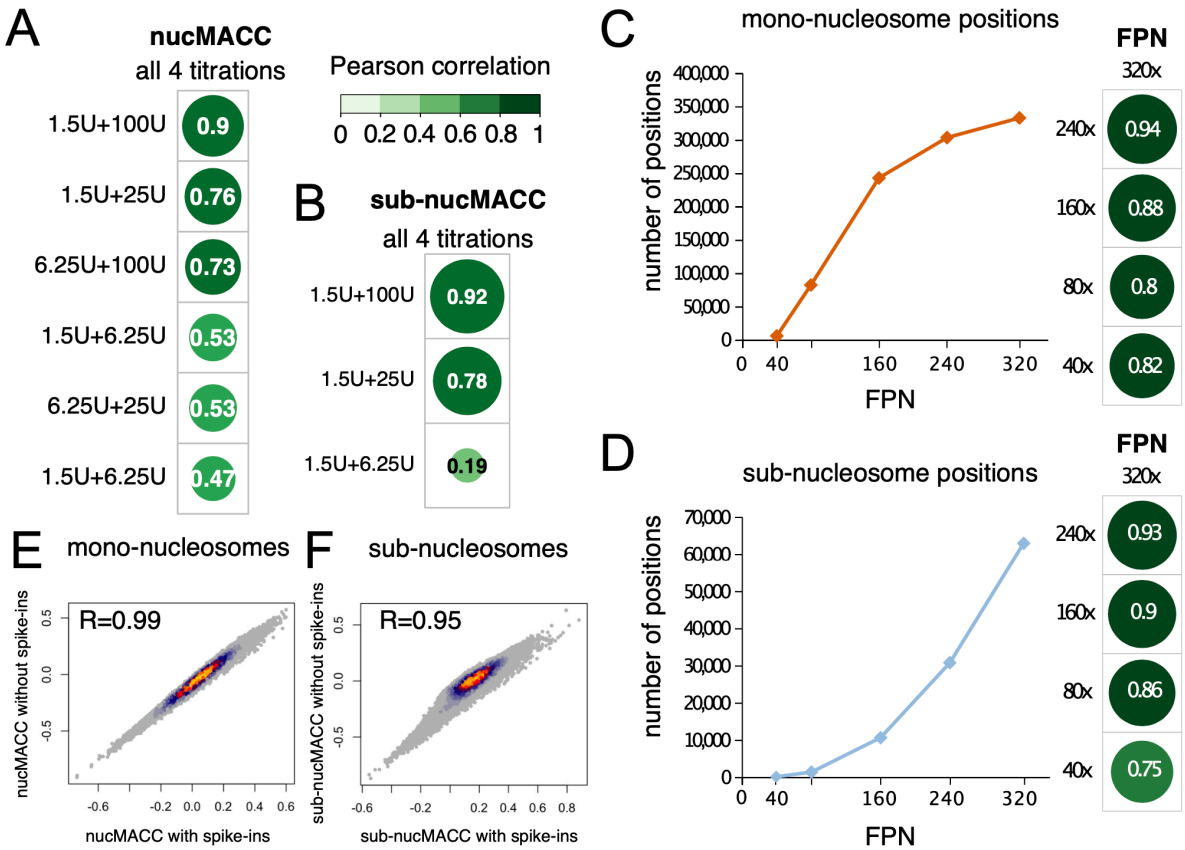


Figure 6

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

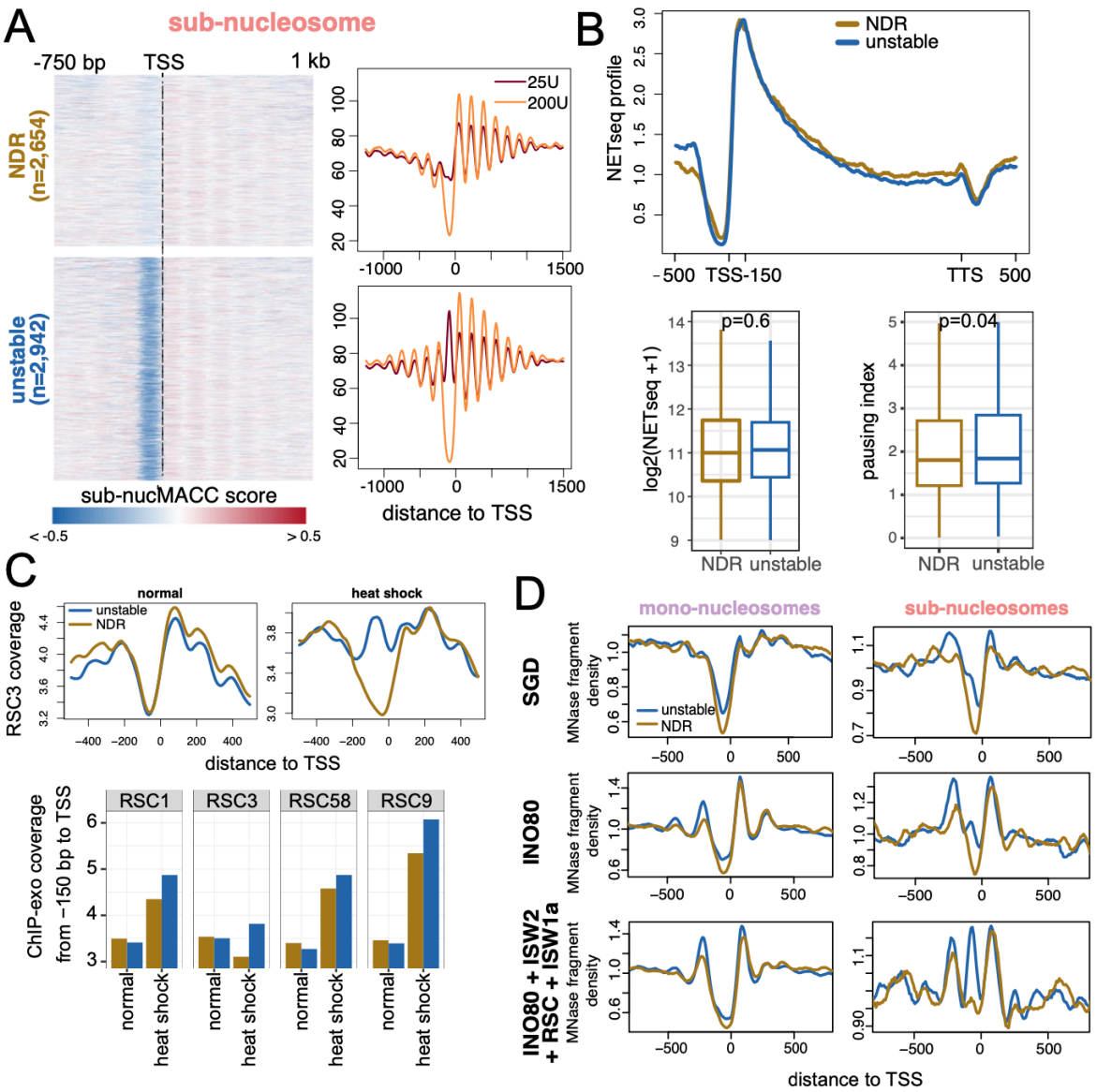


Figure 7

(A) Grouping of promoter types into promoter harboring an unstable nucleosome (unstable; n=2,942) or a nucleosome depleted region (NDR; n=2,654) in *S. cerevisiae*. Left panel: Heatmap showing the distribution of sub-nuccMACC scores. Right panel: Average plot showing the sub-nucleosomal fragment frequencies of the different MNase conditions. (B) Gene expression analysis of the promoter subgroups (unstable in blue and NDR in gold) using NETseq data. Top panel: Median NETseq profile showing nascent RNA abundance. Regions upstream of TSS, downstream of TTS and the first 150 bp downstream of TSS are unscaled. The remaining gene body is scaled to have the same length for each gene. Bottom panel: Boxplots showing the RNA Pol II initiation rate as calculated from the median NETseq signal

1058 over the +1 nucleosome (first 150 bp downstream of TSS) is shown on the left. The pausing
1059 index is shown on the right. Pausing index was calculated as the fold change of the NETseq
1060 signal in the first 150 bp downstream of the TSS versus the remaining gene body. (C)
1061 Occupancy analysis of RSC components at unstable nucleosome or NDR promoter types in
1062 the context of heat shock using ChIP-exo data (Rossi *et al*, 2021). Top panel shows the median
1063 normalized ChIP-exo coverage of the subunit RSC3 under normal and heat shock conditions.
1064 Bottom panel shows the median normalized ChIP-exo signal at the first 150 bp downstream
1065 of the TSS for RSC subunits RSC1, RSC3, RSC58, RSC9. (D) MNase fragment density of *in vitro*
1066 reconstituted nucleosome arrays at unstable nucleosomes or NDR promoter. MNase
1067 fragments were size selected into mono-nucleosomal (140 bp – 200 bp; left panel) and sub-
1068 nucleosomal fragments (< 140 bp; right panel). Nucleosomes were assembled onto DNA by
1069 salt gradient dialysis (SGD) and purified remodeler were added as indicated (Krietenstein *et*
1070 *al*, 2016).
1071