

# Long-range single-molecule mapping of chromatin accessibility in eukaryotes

ZOHAR SHIPONY<sup>1,\*</sup>, GEORGI K. MARINOV<sup>1,\*</sup>, MATTHEW P. SWAFFER<sup>5</sup>, NICHOLAS A. SINNOTT-ARMSTRONG<sup>1</sup>, JAN M. SKOTHEIM<sup>5</sup>, ANSHUL KUNDAJE<sup>1,2</sup>, AND WILLIAM J. GREENLEAF<sup>1,3,4,#</sup>

<sup>1</sup>*Department of Genetics, Stanford University, Stanford, CA 94305, USA*

<sup>2</sup>*Department of Computer Science, Stanford University, Stanford, CA 94305, USA*

<sup>3</sup>*Department of Applied Physics, Stanford University, Stanford, CA 94305, USA*

<sup>4</sup>*Chan Zuckerberg Biohub, San Francisco, California, USA*

<sup>5</sup>*Department of Biology, Stanford University, Stanford, CA 94305, USA*

\**These authors contributed equally to this work*

#*Corresponding author*

## Abstract

Mapping open chromatin regions has emerged as a widely used tool for identifying active regulatory elements in eukaryotes. However, existing approaches, limited by reliance on DNA fragmentation and short-read sequencing, cannot provide information about large-scale chromatin states or reveal coordination between the states of distal regulatory elements. We have developed a method for profiling the accessibility of individual chromatin fibers (SMAC-seq, or Single-Molecule long-read Accessible Chromatin mapping sequencing assay), enabling the simultaneous, high-resolution, single-molecule assessment of chromatin states at multi-kilobase length scales. Our strategy is based on combining the preferential methylation of open chromatin regions by DNA methyltransferases with low sequence specificity, in this case EcoGII, an N<sup>6</sup>-methyladenosine (m<sup>6</sup>A) methyltransferase, and the ability of nanopore sequencing to directly read DNA modifications. We demonstrate that aggregate SMAC-seq signals match bulk-level accessibility measurements, observe single-molecule nucleosome and transcription factor protection footprints, and quantify the correlation between chromatin states of distal genomic elements.

In eukaryotes, open chromatin regions are associated with regulatory elements (REs), such as enhancers, promoters, and insulators. This property is highly useful for identifying candidate REs (cREs) and for understanding the functional organization of genomes. That REs exhibit greatly increased sensitivity to nuclease cleavage was already noted four decades ago<sup>1-3</sup>. Subsequent advances in microarray<sup>4,5</sup> and DNA sequencing technologies<sup>6,7</sup> enabled DNase hypersensitivity-based mapping of cREs genome-wide. Similarly, digestion of DNA is inhibited by nucleosomes, and MNase-seq is widely used to map nucleosome positioning<sup>9</sup>. More recently, the Tn5 transposase was adapted as a probe for chromatin accessibility<sup>8</sup>. However, while short read-based assays can map cREs and positioned nucleosomes, they provide little insight into the long-range physical organization of individual chromatin fibers as they remove linkage between distal segments.

We developed SMAC-seq, a single-molecule method that directly assays both open chromatin regions and nucleosome positioning within a single chromatin fiber at mul-

tilkilobase scales. We use SMAC-seq to study chromatin architecture and co-accessibility states in the yeast *Saccharomyces cerevisiae*. We assess the degree of coordination between positions of nearby nucleosome particles, enumerate mutually exclusive regulatory states along individual loci, and observe coordinated changes in nucleosome positioning and chromatin accessibility upon transcriptional activation. SMAC-seq allows for footprinting of transcription factor (TF) occupancy, and provides strand-specific information about the exposure of DNA occupied by nucleosomes. We expect future applications of, improvements on, and extensions of the SMAC-seq approach to enable novel insights into the dynamics of chromatin states in a wide variety of experimental systems.

## Results

**SMAC-seq maps chromatin accessibility and nucleosome positioning at the multikilobase scale**

SMAC-seq is built on the conceptual foundations of NOMe-seq/dSMF<sup>10-12</sup>. These methods rely on preferential modification of accessible DNA with M.CviPI and/or M.SssI (GpC/CpG-specific 5mC methyltransferases), followed by bisulfite conversion and Illumina-based sequencing readout (both enzymes can be used in the absence of endogenous 5mC methylation). We use the m<sup>6</sup>A methyltransferase EcoGII<sup>13</sup> as an alternative/addition to CpG/GpC, and use nanopore sequencing to generate single-molecule readouts of accessibility states over many kilobases (Figure 1a). Nanopore sequencing allows direct detection of these modifications<sup>14,15</sup>, enabling the generation of methylation maps for individual DNA molecules, which can then be interpreted in terms of chromatin accessibility.

The addition of m<sup>6</sup>A signal associated with accessible chromatin substantially improves both the resolution and applicability of SMAC-seq. Many genomes are endogenously methylated at 5mC positions<sup>16,17</sup>, usually in CpG contexts, but not always<sup>18</sup>, confounding CpG/GpC-based accessibility measurements. More importantly, CpG/GpC dinucleotides are rare. The average resolution achieved by combining the two methyltransferases is >10 bp in *Drosophila* and ~15 bp in yeast. It is ~25bp for GpC alone in mammals, and these are averages; in practice many individual regions either completely lack or contain too few informative positions. Using m<sup>6</sup>A increases SMAC-seq’s resolution down to a theoretical limit of ~3 bp in all model organisms, and ensures proper coverage over all individual loci (Supplementary Figures 1–3 and 4–12).

We initially developed the method in *S. cerevisiae* as it has no endogenous DNA methylation and has a small genome (~12 Mbp), enabling very high sequencing coverage. To verify the specificity and efficiency of enzymatic treatments, we carried out both dSMF experiments on chromatin and M.CviPI + M.SssI + EcoGII reactions on naked DNA (gDNA), followed by bisulfite sequencing. We observe ≥95% CpG/GpC methylation for gDNA, ≤10% for chromatin, and ~0% on untreated gDNA (Supplementary Figure 13). Comparing dSMF to DNase-seq and MNase-seq profiles around transcription start sites (TSSs) and positioned nucleosomes<sup>19</sup> revealed the expected nucleosome depletion/occupancy patterns (Figure 1b-d).

EcoGII’s methylation efficiency is more difficult to estimate as fully methylated templates are known to be difficult to sequence on the Oxford Nanopore platform. Using yeast gDNA or λ DNA treated with a high dose of EcoGII (Supplementary Table 1), the limited number of observed reads exhibited ~50% methylation levels (Supplementary Figure 14 and 15). We hypothesize these rates are underestimates, as biochemical reports suggest ≥50% methylation of gDNA after 5 minutes, increasing to ≥85% after an hour<sup>13</sup>.

We next applied SMAC-seq to unsynchronized *S. cerevisiae* cells. We obtained reads with a median length ~1.5 kbp from this initial experiment, allowing the capture of multiple promoter regions for much of the yeast genome (Supplementary Figures 16-17). We applied the Tombo<sup>20</sup>

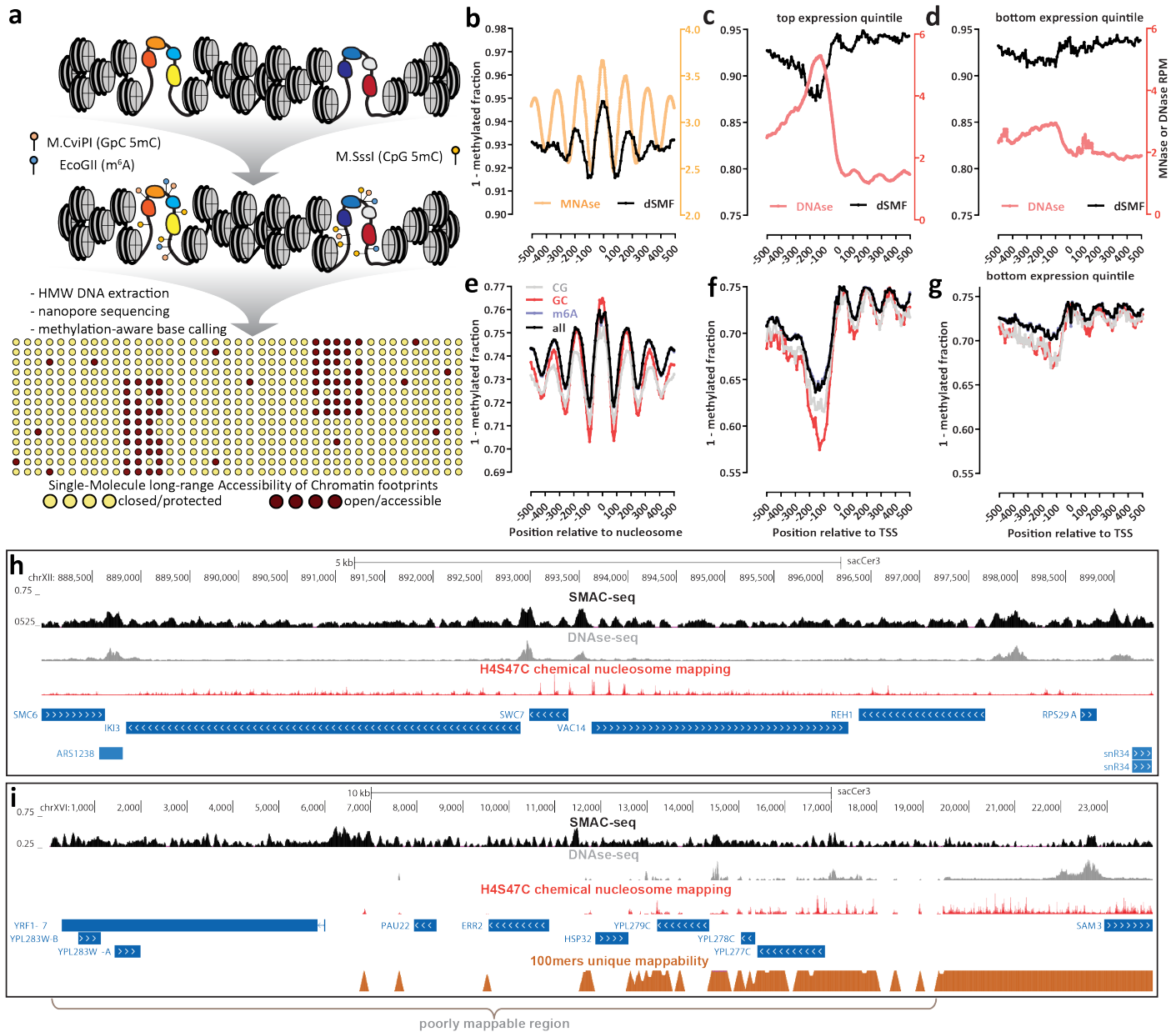
algorithm (running on top of the Minimap aligner<sup>21</sup>) for “resquigglng” of raw nanopore signal and general methylated base calling. We also analyzed our initial dataset with Nanopolish<sup>14</sup>, an alternative algorithm for identifying 5mC events in CpG/GpC context.

Unlike Illumina-based bisulfite sequencing, nanopore-based measurement of DNA methylation provides methylation probabilities. While per-base methylation probabilities are skewed towards 0 or 1, a substantial fraction lie in between those extremes (Supplementary Figure 18). We examined multiple approaches for binarizing methylation calls within single molecules to identify the optimal strategy in terms of signal-to-noise ratio (Supplementary Figures 19–20). We compared average SMAC-seq profiles to dSMF, MNase-seq, and DNase-seq, as well as ChIP-seq for RNA Polymerase (Pol2) and transcription initiation factors around known chromatin features (Figure 1e-g, Supplementary Figures 21-23). SMAC-seq faithfully reproduces nucleosomal positioning throughout the genome and nucleosome depletion around promoters. We also observe positive correlation between average SMAC-seq methylation levels and DNase/ATAC-seq coverage over promoter regions (Supplementary Figure 24). Strikingly, SMAC-seq has a larger observed dynamic range than dSMF data (possibly due to higher long-read mapping efficiency). Based on dSMF and untreated gDNA data, we estimate the false positive rate of methylation base calling to be ~20% for Tombo and 10-15% for Nanopolish (Supplementary Figures 14 and 25). We also examined potential sequence biases inherent to the combination of methylation enzymes and base calling algorithm. We find only modest differences in methylation levels for different *k*-mers (≤2-fold for *k* = 6; Supplementary Figures 26-27).

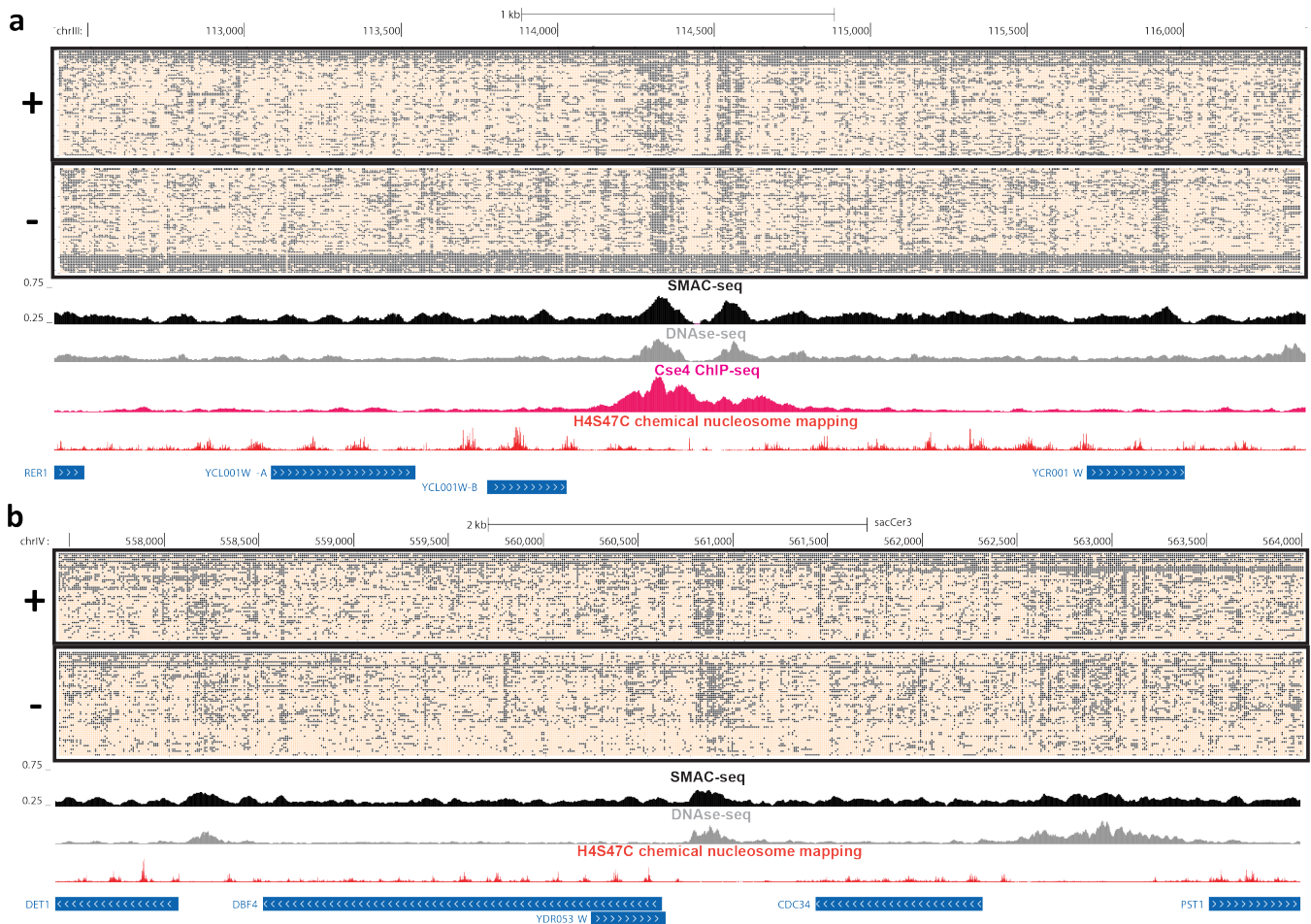
In practice, the biologically relevant scale of chromatin accessibility is larger than an individual base. We thus reasoned that sharing methylation information between adjacent bases should improve the reliability of accessibility measurements, and developed a Bayesian procedure to aggregate methylation probabilities and derive single-molecule accessibility calls over windows of arbitrary size (thereafter referred to as “aggregate” signal).

We observed a relatively small subpopulation of reads highly methylated over large segments (Supplementary Figure 28 and Figure 2a), which we interpret as originating from naked DNA molecules likely from dead cells. As such reads can confound many analyses, we filter these out (Supplementary Figure 29). However, at certain loci chromatin is indeed largely nucleosome-free in vivo; for such unique loci, we analyze all reads.

We then compared average SMAC-seq profiles against positioned nucleosomes, DNase-seq, and transcriptional activity maps at the level of individual genomic loci (Figure 1h). Qualitatively, we observe that large SMAC-seq peaks match very closely with DNase-seq peaks, while smaller SMAC-seq “bumps” inversely correlate with positioned nucleosomes, consistent with labeling of linker DNA. Thus



**Figure 1: The SMAC-seq assay for profiling chromatin accessibility and nucleosome positioning at the multikilobase scale.** (a) Outline of the SMAC-seq assay. Intact chromatin is treated with m<sup>6</sup>A and CpG and GpC 5mC methyltransferases, which preferentially methylate DNA bases in open chromatin regions. HMW DNA is then isolated, subjected to nanopore sequencing and methylated bases are used to reconstruct the open chromatin state within individual molecules. (b-h) SMAC-seq faithfully captures chromatin accessibility around promoters and positioned nucleosomes in *S. cerevisiae*; (b) MNase-seq and dSMF profiles around chemically mapped positioned nucleosome dyads; (c) DNase-seq and dSMF profiles around the top 20% highly expressed genes in *S. cerevisiae*; (d) DNase-seq and dSMF profiles around the bottom 20% expressed genes in *S. cerevisiae*; (e) Average SMAC-seq profile around chemically mapped positioned nucleosomes dyads (shown is the “diamide 0 min rep2” sample); (f) Average SMAC-seq profile around the top 20% highly expressed genes in *S. cerevisiae*; (g) Average SMAC-seq profile around the bottom 20% expressed genes in *S. cerevisiae*; (h) SMAC-seq correlates closely with both DNase-seq and nucleosome occupancy profiling at the level of individual loci, and provides a combined readout of accessibility and nucleosome positioning. Shown is aggregate SMAC-seq signal along the genome (aggregated over 50-bp windows sliding every 5 bp; see Methods for details) together with DNase-seq, nucleosome chemical mapping data, and transcriptional activity (measured by PRO-seq and PRO-cap). Large aggregate SMAC-seq signal enrichments match closely with DNase accessibility peaks, while smaller aggregate SMAC-seq peaks are inversely correlated with positioned nucleosomes; (i) SMAC-seq profiles chromatin accessibility in repetitive regions of the genomes that are “invisible” to short reads. Shown is the telomeric region of chrXVI



**Figure 2: SMAC-seq provides a single-molecule linked-read view of the chromatin landscape.** (a) Unfiltered nanopore reads fully spanning the 4-kilobase neighborhood of the centromere of *S. cerevisiae* chrIII (“aggregate” signal “Sample 1”). (b) Unfiltered nanopore reads fully spanning a 6.6-kilobase neighborhood encompassing several genes on chrIV (“aggregate” signal from “Sample 1”). In both cases, accessibility is shown at 10-bp resolution (see Methods section for details) for the single-molecule display, and aggregated over sliding (every 5 bases) 50-bp windows for the average SMAC-seq track.

SMAC-seq simultaneously identifies both open chromatin regions and positioned nucleosomes.

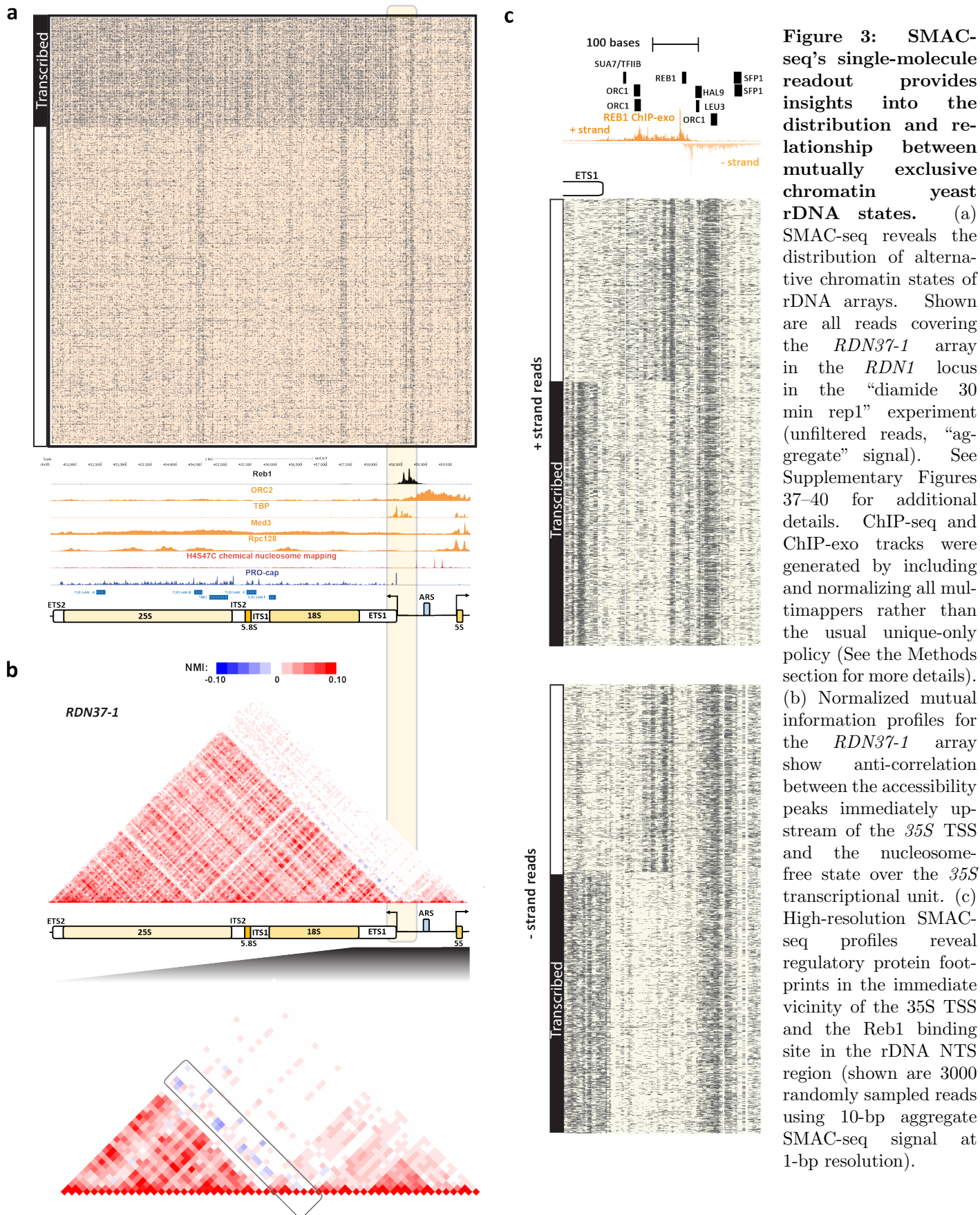
The long nanopore reads allow SMAC-seq to map accessibility for the whole yeast genome (Supplementary Figure 30). For example, SMAC-seq maps chromatin and nucleosomes in the repetitive telomere of chrXVI (Figure 1i), which contains several active promoters and numerous well-positioned nucleosomes. SMAC-seq also revealed open chromatin peaks around the promoters of multiple transposable elements (Supplementary Figure 31).

### SMAC-seq provides single-molecule accessibility profiles on individual chromatin fibers

To demonstrate SMAC-seq’s ability to map open chromatin within individual long molecules we investigated all reads spanning the 4-kb neighborhood around the chrIII centromere (Figure 2a). Yeast centromeres are specified by

precisely defined sequence elements and are occupied by a single nucleosome containing the H3 histone variant Cse4, thought to be nearly perfectly positioned<sup>22,23</sup>. We indeed observe strong nucleosomal positioning using SMAC-seq, with nearly all individual reads exhibiting the expected nucleosomal pattern. We also observe hints of substructure in the form of accessibility inside the protected centromeric region and potential protection footprints in its immediate vicinity. We find similarly strong positioning for many other centromeric nucleosomes (Supplementary Figures 32-35). We also examined a ~6.6-kb span of chrIX containing five genes and three open chromatin regions, one of them fairly large and diffuse. In contrast to the more localized accessibility observed elsewhere, this region exhibits considerable accessibility heterogeneity suggesting a complex protein occupancy landscape.

We next asked if SMAC-seq could reveal binary chro-



matin accessibility states by investigating ribosomal DNA (rDNA) loci. In yeast, rDNA is organized into multicopy ( $\sim 150$ ) arrays, each  $\sim 9.1$  kb unit of which contains a copy of the 35S precursor pre-rRNA, transcribed by Pol I, a 5S RNA, transcribed by Pol III, and a replication origin ARS element, located in non-transcribed (NTS) regions of the array. The number of units can vary between cells, and the **sacCer3** genome assembly only contains a single locus with two array copies. The rDNA chromatin structure adopts two distinct conformations<sup>25–27</sup> – an inactive nucleosomal state and a state largely devoid of nucleosomes due to extremely high transcription activity<sup>24,26,28</sup>. The two states are estimated to exist in roughly equal proportions in normally growing cells<sup>29</sup>. However, other studies have alternatively suggested that nucleosomes are present over actively transcribed rDNA arrays<sup>29</sup>. Of note, rDNA indeed appears to be extremely accessible in short-read assays – around half of reads in a typical yeast ATAC-seq dataset originate from rDNA arrays (Supplementary Figure 36). Single-molecule SMAC-seq **maps reveal** a striking picture of the two alternative, mutually exclusive rDNA states (Figure 3a). About a quarter of full-length molecules exhibit near-full accessibility over the 35S transcript, but not in the NTS; the rest show a typical nucleosomal state. A broadly similar picture is observed in all samples (Supplementary Figures 37–40). While it is possible that the two states are differentially represented due to biases against fully methylated long reads, we observe the fully accessible fraction is present in approximately 50% of shorter windows around the 35S promoter (Figure 3c). We also observe a region of localized accessibility just upstream of the 35S transcriptional unit present only in the nucleosomal subpopulation, suggesting the possibility of a regulatory switch at this location. Finally, we also observe at least two previously unreported regions inside 35S exhibiting strong accessibility in the nucleosome-protected fraction (Figure 3a).

To quantify (anti-)correlation between chromatin states, we developed a modified normalized mutual information (NMI) metric for assessing the degree of accessibility correlation between genomic regions. NMI analysis of rDNA confirmed the inverse correlation between the active 35S state and accessibility of this upstream element (Figure 3b).

What factors might be driving this observed chromatin state switch? Silencing of yeast rDNA is thought to be mediated by the Sir2-containing RENT complex<sup>30</sup>, and a NTS1 Reb1 binding site has been suggested to recruit corepressors<sup>29</sup>. We took a higher-resolution view of NTS1 by integrating SMAC-seq, Reb1 ChIP-exo data and transcription factor motif maps. We find a clear pattern of protection around the Reb1 motif, concordant with ChIP-exo (Figure 3c), and we also observe patterns consistent with footprinting for several other TFs. However, the anti-correlated accessibility profile seems to not be exclusively associated with Reb1 binding but rather with the region closer to the 35S TSS. Thus it appears that other proteins may be responsible for establishing this state.

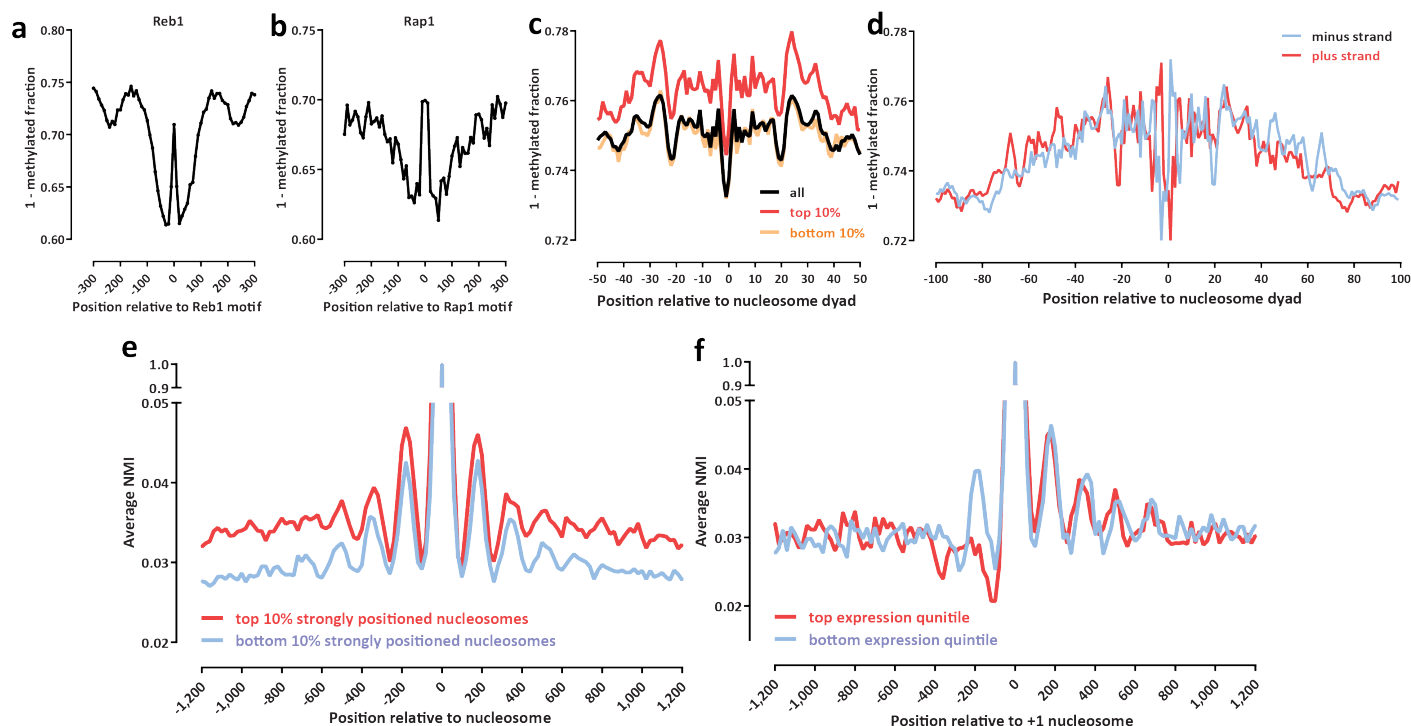
## SMAC-seq provides a high-resolution strand-specific view of protein occupancy on DNA

We next asked if SMAC-seq can generally identify TF footprints (Supplementary Figure 41). Averaging genome-wide SMAC-seq profiles over occupied motifs revealed strong protection footprints for several factors, such as Reb1, Rap1 and ORC1 (Figure 4a-b, Supplementary Figure 42). We observed high concordance between footprint profiles observed using either DNase-seq, ATAC-seq, and SMAC-seq with footprints identified previously by high-resolution DNase-seq<sup>7</sup> in aggregate and at some individual sites (Supplementary Figure 43–50). However, we did not observe strong footprinting for all TFs (e.g. Abf1 and Cbf1; Supplementary Figure 42) even though some do exhibit DNase-seq footprints<sup>7</sup>, perhaps because different enzymes differ in their ability to access DNA in the context of protein occupancy.

To further explore SMAC-seq’s resolution limits, we studied methylation patterns around positioned nucleosomes in more detail (Figure 4c-d, Supplementary Figure 23). We observe marked increase in accessibility signal at the dyad point, in contrast to the points of contact with DNA two helical turns away. The same pattern was observed for all nucleosomes irrespective of positioning strength. We did not observe similar patterns in deep DNase-seq data (Supplementary Figure 23e). We next quantified strand-specific DNA accessibility within nucleosomes, and observe a strand-asymmetric DNA accessibility patterns around the nucleosome particle (Figure 4d), especially within the dyad and at the points two helical turns away. The magnitude of these differences is similar to that observed between nucleosomes and flanking linker regions. This heterogeneity in methylation potential within the nucleosome may inform the manner by which TFs might interact with nucleosome-associated DNA *in vivo*<sup>31</sup>. We also note that these patterns are most clearly observed using m<sup>6</sup>A (Supplementary Figure 51).

## SMAC-seq reveals chromatin co-accessibility patterns

We next examined co-accessibility patterns in the yeast genome by assessing nucleosome positioning correlations. Average NMI profiles centered on positioned nucleosomes reveal detectable correlation between nucleosome positions up to three to four nucleosomes away (Figure 4e; Supplementary Figure 52), with strongly positioned nucleosomes exhibiting stronger overall correlation. These observations are consistent with nucleosomes imposing restrictions on one another, resulting in short-range correlation between protection footprints that dephases over longer ranges. We next measured co-accessibility in the vicinity of promoters (Figure 4f). Active yeast TSSs are characterized by an upstream nucleosome-depleted/free region (NFR) and a well-positioned +1 nucleosome. NMI profiles centered on the latter show significant differences between expressed and



**Figure 4: SMAC-seq provides a high-resolution strand-specific view of genomic occupancy by DNA-binding proteins and complexes.** (a-b) SMAC-seq allows for footprinting of transcription factor binding events. Shown is aggregate genome-wide SMAC-seq signal around occupied (as measured by ChIP-exo) Reb1 (a), and Rap1 (b) sequence recognition motifs. (c) SMAC-seq profiles around positioned nucleosome dyads reveal increased accessibility in the dyad and increased protection at the points of contact with the nucleosome (see Supplementary Figure 23 for additional details.) (d) SMAC-seq provides a strand-specific view of nucleosome occupancy and reveals differential accessibility between the two DNA strands depending on their position on the nucleosomal particle. (e-f) Coordination between the positions of individual nucleosomes at the level of single chromatin fibers. (e) Shown is the average normalized mutual information between each strongly or poorly positioned nucleosome in the yeast genome and its immediate genomic neighborhood (measured for windows of 10 bp length tiling at every genomic position centered on the nucleosome dyad). (f) Shown is the average normalized mutual information between each +1 nucleosome and its immediate genomic neighborhood in highly expressed and in mostly silent genes (measured for windows of 10-bp length tiling at every genomic position centered on the +1 nucleosome dyad).

inactive genes. While correlation decays downstream of the TSS similarly for both groups, active genes exhibit an inverse correlation pattern upstream of the TSS.

Active yeast genes often exist in a looped conformation, with promoter and termination regions in physical proximity, an arrangement thought to help enforce transcriptional directionality<sup>32,33</sup>. We wondered if accessibility would be correlated between the two gene ends. SMAC-seq reveals low levels of correlation between the NFR and 3' gene ends, and stronger correlation between positioned nucleosomes in these locations (Supplementary Figure 53). The correlation between the NFR and 3' ends increases for active genes and decreases for silent genes, suggesting that transcriptional activity and/or looping may help more strongly position nucleosomes at the two gene ends.

We next assessed coordinated accessibility between yeast TSSs. To this end, we devised an explicit test of co-

ordinated coaccessibility based on splitting reads into separate pieces, randomly reassembling them, then deriving an empirical coaccessibility distribution. We identified 1,115 TSS pairs as significantly correlated (out of 19,578 pairs covered with  $\geq 100$  reads; Supplementary Figure 54). Of these, 560 were located  $\geq 1$ kb from each other (for example see Supplementary Figure 55). One possible mechanism for this correlated accessibility signal is increased frequency of physical association in 3D space. Analysis of Micro-C data<sup>34</sup> shows that significantly coaccessible promoters interact more frequently than non-coaccessible ones at a similar distance (Supplementary Figure 56).

#### SMAC-seq charts coordinated accessibility changes during the yeast stress response

Finally, we carried out SMAC-seq during a time course of diamide treatment, to monitor chromatin states during a



dynamic response. Diamide oxidizes thiols in proteins, leading to activation of the stress response pathway and changes in the expression of hundreds of genes<sup>35</sup>. Yeast stress response is largely mediated by the Hsf1 and Msn2/4 TFs<sup>36</sup>.

We performed SMAC-seq at 0, 30 and 60 minutes after diamide treatment, as well as RNA-seq, ATAC-seq, and ChIP-seq for Pol2, the elongating Pol2pS2 version, and HSF1 (RNA and ATAC data were also collected at 15 and 45 minutes; Figure 5a). We observe several hundred genes exhibiting strong expression changes (Supplementary Figure 57), and strong Hsf1 occupancy induction at hundreds of sites (Supplementary Figure 58). SMAC-seq at 30 minutes shows strong footprinting over Hsf1 motifs within induced Hsf1 binding sites.

We illustrate the dynamic accessibility patterns we observe upon diamide treatment using the *TMA10* and *HSP26* genes in Figure 5d-e, and multiple others in Supplementary Figures 59–67. *TMA10* and *HSP26* are strongly upregulated at 15 minutes; *TMA10* expression subsequently declines somewhat and stabilizes (Figure 5c) while *HSP26* only declines at 60 minutes. SMAC-seq reveals a relatively modest level of upstream accessibility before diamide treatment. However, at 30 minutes and upon Hsf1 binding, dramatic changes are evident. Nearby nucleosomes are evicted in many cells, and nucleosome depletion increases within gene bodies, where RNA Pol2 ChIP-seq shows highly active transcription. At 60 minutes, this response dampens for *TMA10*, with the fraction of accessible reads decreasing; the effect is less pronounced for *HSP26* whose expression remains relatively higher. NMI co-accessibility maps (Supplementary Figures 59–68) frequently show loss of correlation between positioned nucleosomes within and upstream of activated gene bodies as a result of diamide response, consistent with increased nucleosome movement due to the activity of polymerases and chromatin remodelers. Aggregate SMAC-seq observations were largely corroborated by matched ATAC-seq on the same samples (Supplementary Figures 69–70). We observed a decrease in the transcribed rDNA array fraction (Figure 5f), consistent with decreased rDNA transcription upon activation of the stress response program.

## Discussion

SMAC-seq is a novel single-molecule method for profiling chromatin accessibility within individual chromatin fibers at a multikilobase resolution using nanopore sequencing. SMAC-seq generates accessibility signals similar to widely used short-read methods while enabling the simultaneous profiling of nucleosome positioning and accessible chromatin on a truly genome-wide scale, the measurement of the underlying distribution of accessibility states, and the identification of loci exhibiting significant co-accessibility.

Extending SMAC-seq to larger genomes will require significantly increased sequencing throughput, or selective enrichment of individual loci. Fortunately, nanopore through-

put is increasing rapidly, while selective enrichment methods are also becoming available<sup>37</sup>. Increases in read length will also be useful, especially for assaying co-accessibility of distal regulatory elements, which can often be tens of kilobases apart in mammalian genomes.

Base calling is another area of future improvement. One hurdle for the creation of more accurate base callers is the lack of ground truth controls for training base calling algorithms (e.g. pools of DNA templates with individual modifications in well-defined yet highly diverse sequence contexts). Alternatively, the use of tags bulkier than a methyl group<sup>38</sup> may provide much stronger modulation of the current passing through the nanopore, enabling more reliable signal identification. We also anticipate a diversity of DNA modifying enzymes available to carry out SMAC-seq variations.

Endogenous methylation in mammalian genomes also represents potentially confounding signal. To evaluate the sale of this concern, we generated low-coverage SMAC-seq data for human GM12878 cells using only EcoGII, and examined aggregate “m<sup>6</sup>A-SMAC” profiles around CTCF sites, open chromatin regions and TSSs. We recovered the expected features of chromatin accessibility (strong nucleosome positioning around CTCF<sup>62</sup>, accessibility peaks around all three features), and observed no significant difference between SMAC-seq profiles generated by filtering out A positions nearby CpGs, demonstrating that interference from endogenous methylation is likely modest (Supplementary Figures 71–73).

However, there are also species where m<sup>6</sup>A occurs endogenously and is strongly correlated with chromatin accessibility and nucleosome positioning<sup>39–41</sup>. Modifications such as 4mC<sup>42</sup> (**N<sup>4</sup>-methylcytosine**), cytidine deamination<sup>43</sup>, or 5-hydroxymethyluracil<sup>44</sup> are among the potential future alternatives in such cases. Finally, we believe that the integration of SMAC-seq with other measurements of the physical genome and the epigenome into single-molecule multiomic assays represents a potentially fruitful direction. We envision the possibility of simultaneous single-molecule, multikilobase-scale measurements of accessibility, nucleosomal positioning, endogenous DNA methylation, protein occupancy, chromatin interactions, and/or DNA replication. In principle, similar approaches may also be applicable to individual RNA molecules. We expect long-read single-molecule approaches to provide an important new class of tools for the study of the functional and physical organization of genomes.

## Materials and Methods

Except for when explicitly stated otherwise, all analyses were carried out using custom-written Python or R scripts (**available on request**).

## Cell lines and cell culture

The BY4741 *S. cerevisiae* strain (a kind gift from Ji-Ping Wang and Xiaozhong Wang) was used for all experiments except for Hsf1 ChIP-seq experiments where MS143 (H4S47C.Hsf1-V5::HphMX6, this study) was used. MS143 was generated by PCR-based C-terminal tagging of Hsf1 with the V5 epitope. Hsf1-V5 tagging was confirmed by colony PCR and western blotting. For all experiments, except the initial one (“Sample 1”), cells were grown in YPD media (30 °C) to OD~0.8 before collection.

## Yeast SMAC-seq experiments

### *Enzymatic treatment of chromatin*

We developed and optimized SMAC-seq using the equivalent of  $1 \times 10^6$  human cells, which in the case of *S. cerevisiae* translates in to  $2.5 \times 10^8$  (the size of the haploid human genome is  $\sim 3 \times 10^9$ bp while that of *S. cerevisiae* is  $1.2 \times 10^6$ bp). As yeast cells have a cell wall, we adapted the spheroplasting protocol previously used for carrying out ATAC-seq in yeast cells<sup>45</sup> for our SMAC-seq experiments.

Yeast cells in log phase ( $OD_{660} \leq 1.0$ ) were first centrifuged at 13,000 rpm for 1 minute, then washed with 100  $\mu$ L Sorbitol Buffer (1.4 M Sorbitol, 40 mM HEPES-KOH pH 7.5, 0.5 mM MgCl<sub>2</sub>), and centrifuged again at 13,000 rpm for 1 minute. Cells were then spheroplasted by resuspending in 200  $\mu$ L Sorbitol Buffer with DTT added at a final concentration of 10 mM and 0.5 mg/mL 100T Zymolase, followed by incubating for 5 minutes at 30 °C at 300 rpm in a Thermomixer. The pellet was centrifuged for 2 minutes at 5,000 rpm, washed in 100  $\mu$ L Sorbitol Buffer, and centrifuged again at 5,000 rpm for 2 minutes.

Cells were then resuspended in 100  $\mu$ L ice-cold Nuclei Lysis Buffer (10 mM Tris pH 7.4, 10 mM NaCl, 3 mM MgCl<sub>2</sub>, 0.1 mM EDTA, 0.5% NP-40) and incubated on ice for 10 minutes. Nuclei were then centrifuged at 5000 rpm for 5 min at 4 °C, resuspended in 100  $\mu$ L cold Nuclei Wash Buffer (10 mM Tris pH 7.4, 10 mM NaCl, 3 mM MgCl<sub>2</sub>, 0.1 mM EDTA), and centrifuged again at 5,000 rpm for 5 min at 4 °C. Finally, nuclei were resuspended in 100  $\mu$ L M.CviPI Reaction Buffer (50 mM Tris-HCl pH 8.5, 50 mM NaCl, 10 mM DTT).

Nuclei were then first treated with M.CviPI + EcoGII by adding 200 U of M.CviPI (NEB) and 200 units of EcoGII (NEB), SAM at 0.6 mM and sucrose at 300 mM, and incubating at 30 °C for 7.5 min. After this incubation, 128 pmol SAM and another 100 U of enzymes were added, and a further incubation at 30 °C for 7.5 min was carried out. Immediately after, M.SssI treatment followed, by adding 60 U of M.SssI (NEB), 128 pmol SAM, MgCl<sub>2</sub> at 10 mM and incubation at 30 °C for 7.5 min.

The reaction was stopped by adding an equal volume of Stop Buffer (20 mM Tris-HCl pH 8.5, 600 mM NaCl, 1% SDS, 10 mM EDTA).

### *High-molecular weight DNA isolation*

HMW DNA was isolated using the MagAttract HMW DNA Kit (Qiagen; cat # 67563) following the manufacturer’s instructions.

### *Enzymatic treatment of naked DNA*

Naked DNA was treated under exactly the same conditions as chromatin except that the reaction volume and enzyme amounts were reduced in half. HMW DNA was purified as described above

## GM12878 cell culture

The GM12878 human lymphoblastoid cell lines were grown in media containing RPMI1640-GlutaMAX (Life Technologies, Carlsbad, CA) supplemented with 10% fetal bovine serum.

## GM12878 SMAC-seq experiments

Briefly  $1 \times 10^6$  human GM12878 cells were washed with  $1 \times$  PBS then resuspended in 200  $\mu$ L ice-cold Nuclei Lysis Buffer (10 mM Tris pH 7.4, 10 mM NaCl, 3 mM MgCl<sub>2</sub>, 0.1 mM EDTA, 0.5% NP-40) and incubated on ice for 10 minutes. Nuclei were then centrifuged at 500 *g* for 5 min at 4 °C, resuspended in 200  $\mu$ L cold Nuclei Wash Buffer (10 mM Tris pH 7.4, 10 mM NaCl, 3 mM MgCl<sub>2</sub>, 0.1 mM EDTA), and centrifuged again at 500 *g* for 5 min at 4 °C. Finally, nuclei were resuspended in 200  $\mu$ L reaction buffer ( $1 \times$  NEB CutSmart buffer, 0.3 M sucrose). Nuclei were then treated with EcoGII by adding 200 units of EcoGII (NEB) and SAM at 0.6 mM, and incubating at 37 °C for 10 min. The reaction was stopped by adding 0.2% SDS, and HMW DNA was immediately isolated as previously described.

## SMAC-seq analysis

### *Nanopore sequencing*

HMW DNA was converted into libraries using the Ligation Sequencing Kit 1D (Oxford Nanopore Technologies, SQK-LSK108) following the manufacturer’s instructions. Nanopore sequencing was carried out on R9.4 MinION flow-cells (Oxford Nanopore Technologies) for up to 48 hours.

### *Nanopore base calling*

Nanopore events were converted to DNA sequence using Albacore (V2.3.3) using default settings. Reads were resquigged using Tombo<sup>20</sup>, version 1.3, using the `sacCer3` reference genome. Methylated bases were identified using Tombo in the “de novo model” mode.

### Aggregation of accessibility information over multibasepair windows

Even with the addition of m<sup>6</sup>A methylation, the resolution of SMAC-seq still does not cover every nucleotide in the genome, and it varies substantially between different locations depending on local sequence content differences. In addition to that, nanopore base calling is still far from being a fully resolved problem, and even more so in methylation-aware mode. For these reasons, for many of the analyses described in this study we aimed at assigning aggregate accessibility scores over windows, taking the totality of the available evidence into account, thus obtaining more reliable, if coarser-grained, views of accessibility patterns along the genome. We used a Bayesian approach to carry out aggregation, as follows.

For a given window of width  $w$  in the genome, specified by coordinates  $c, i, i + w$  (where  $c$  denotes the chromosome, and  $i$  the leftmost coordinate of the window), and for all reads  $r \in R_{c,i,i+w}$  fully spanning the window, we obtain all Tombo probabilities  $p_{r,(c,j)}$  such that  $j \in [i, i + w]$  for sequence contexts CpG, GpC and A on the corresponding genomic strand. We use a Beta prior  $B(\alpha, \beta)$ , with  $\alpha = \beta = 10$ , which we then updated based on each probability  $p_{r,(c,j)}$  for all  $j \in [i, i + w]$ . The final binary accessibility score  $p_{r,(c,i,i+w)}$  for read  $r$  and window  $c, i, i + w$  is determined by the final state of the prior.

### Read filtering

As discussed above, we sometimes observe a population of reads that are fully methylated across their whole length or over large segments of it. These reads most likely derive from dead cells, as our initial experiment, which was carried out on a very dense yeast population containing a substantial number of dead cells, exhibited much higher proportion of such reads compared to subsequent experiments using early log-phase cells. In order to remove such potentially artifactual reads, “filtered” sets of reads were obtained by removing all reads containing a  $\geq 1$ -kbp stretch that is  $\geq 75\%$  methylated (while also filtering out reads shorter than 1 kb).

### Read clustering

For most analyses presented in this manuscript, the `tblkmeans` package was used to cluster SMAC-seq reads (implemented in R, <https://bitbucket.org/tanaylab/tblkmeans>). In addition, the hierarchical clustering implementation in `scipy` was also used in certain cases.

### Co-accessibility assessment using Normalized Mutual Information

To evaluate co-accessibility patterns along the genome, we applied a Normalized Mutual Information as follows. Each chromosome in the genome  $c$  was split into windows of size  $w$ . For each such window  $(c, i, i + w)$ , we identified the maximum range to the right of it,  $(c, j, j + w)$  such that the span  $(c, i, j + w)$  is covered by  $\geq M$  reads. All reads spanning  $(c, i, j + w)$  were then extracted and subsampled down to  $M$  reads (usually  $M = 100$ , unless specified otherwise). Accessibility scores were then aggregated and binarized as described above for all windows located in the span  $(c, i, j + w)$ , and for all  $M$  reads fully spanning it, resulting in a local co-accessibility matrix  $LCM$  of size  $M \times (j + w - i)/w$ . We then calculated Normalized Mutual Information scores for each pair of columns  $LCM_k$  and  $LCM_l$  as follows:

$$\begin{aligned}
 MI(LCM_k, LCM_l) = & p(0, 0) \log_2 \left( \frac{p(0, 0)}{p_k(0) p_l(0)} \right) \\
 & + p(1, 1) \log_2 \left( \frac{p(1, 1)}{p_k(1) p_l(1)} \right) \\
 & + p(0, 1) \log_2 \left( \frac{p(0, 1)}{p_k(0) p_l(1)} \right) \\
 & + p(1, 0) \log_2 \left( \frac{p(1, 0)}{p_k(1) p_l(0)} \right)
 \end{aligned} \tag{1}$$

MI scores were then normalized and rescaled in the interval  $(-1, 1)$ :

$$NMI(LCM_k, LCM_l) = \begin{cases} \frac{MI(LCM_k, LCM_l)}{\sqrt{H(LCM_k)H(LCM_l)}} & \text{for } p(0, 0) + p(1, 1) \geq 0.5 \\ -\frac{MI(LCM_k, LCM_l)}{\sqrt{H(LCM_k)H(LCM_l)}} & \text{for } p(0, 0) + p(1, 1) < 0.5 \end{cases} \tag{2}$$

Where  $H$  refers to the entropy of each individual distribution.

For computational efficiency, local NMI matrices were calculated for even-sized (50kb) evenly spaced (every 10kb) tiles of the genome. The entries of the general genome-wide

NMI matrix were then calculated as the average of all local NMI matrices containing each entry.

### Testing for coordinated accessibility

Coordinated accessibility was evaluated as follows. For each pair of locations  $(c, i_1, i_1 + r_1)$  and  $(c, i_2, i_2 + r_2)$  (usually  $r_1 = r_2$ ), a minimum number of reads  $N$  was required that fully spans the  $(c, i_1, i_2 + r)$  interval. All such reads were then obtained for each pair, and then subsampled multiple times down to  $N$  reads (in order not to introduce bias in coordinated accessibility tests arising due to differential read coverage between locations closer/further apart). For each subsampling, the fraction of accessible regions  $p_1$  and  $p_2$  was estimated for each of the two locations using the Bayesian procedure described above, as well as the distribution of joint accessibilities over the four states  $(0, 0)$ ,  $(1, 0)$ ,  $(0, 1)$ , and  $(1, 1)$ . The two halves of the reads were then virtually split in half and recombined for a total of  $10^3$  random combinations. The empirical distribution  $\mathcal{N}(\mu, \sigma)$  of the four states was then estimated from these random combinations, where  $\mu = N * (p_{(0,0)} + p_{(1,1)})$  if  $p_{(0,0)} + p_{(1,1)} > 0.5$  and  $\mu = N * (p_{(1,0)} + p_{(0,1)})$  if  $p_{(0,0)} + p_{(1,1)} \leq 0.5$ . Empirical coordinated accessibility  $p$ -values were then estimated based on the observed counts  $|(0, 0)| + |(1, 1)|$  if  $|(0, 0)| + |(1, 1)| > 0.5 * N$  or  $|(0, 1)| + |(1, 0)|$  if  $|(0, 0)| + |(1, 1)| \leq 0.5 * N$ . Bonferroni correction was applied to account for multiple hypothesis testing.

### dSMF and Bisulfite sequencing

Illumina measurements of CpG and GpC methylation levels were carried out using the PBAT protocol<sup>46</sup> with modifications. HMW DNA ( $\sim 500$  ng) was bisulfite converted using the EZ DNA Methylation-Lightning Kit (Zymo, Cat # D5030) by mixing 20  $\mu$ L of purified DNA ( $\sim 500$  ng) with 130  $\mu$ L DNA Methylation Lightning Conversion reagent and incubating at 98 °C for 8 minutes and then at 64 °C for 60 minutes. Bisulfite converted DNA was then cleaned up using the EZ DNA Methylation-Lightning Kit following manufacturer's instructions.

First strand synthesis was carried out by mixing 20  $\mu$ L bisulfite converted DNA, 19.75  $\mu$ L H<sub>2</sub>O, 5  $\mu$ L 10 $\times$  Blue Buffer (ThermoFisher), 1.25  $\mu$ L 10 mM dNTP (NEB), and 4  $\mu$ L custom-designed biotinylated adapter. Samples were then incubated at 94 °C for 5 minutes, and at 4 °C for 5 minutes, after which 1.5  $\mu$ L Klenow (3'  $\rightarrow$  5' exo minus; MCLab) were added, and the reaction was incubated at 4 °C for 15 minutes, at 37 °C for 90 minutes, and at 70 °C for 5 minutes. First-strand reaction cleanup was carried out using 50  $\mu$ L AMPure XP beads (Beckman Coulter); DNA was eluted 50  $\mu$ L EB buffer.

Biotinylated DNA was captured on streptavidin beads. A total of 20  $\mu$ L streptavidin Dybaneads M-280 (ThermoFisher) per sample were added to a PCR tube, separated on a magnet and then resuspended in 50  $\mu$ L 2 $\times$  BW(Li) buffer (6.3 g LiCl, 0.5 mL Tris-HCL pH 8.0, and 0.1 mL 500 mM EDTA for 50 mL total volume), to which the 50  $\mu$ L of eluted first-strand reaction DNA was added. Beads were then incubated at room temperature for 30 minutes, washed

with 180  $\mu$ L 2 $\times$  BW(Li) buffer, twice with 0.1 N NaOH (by resuspending well and incubating at room temperature for 2 minutes), washed again with 180  $\mu$ L 2 $\times$  BW(Li) buffer, then with 180  $\mu$ L 10 mM Tris-HCL pH 7.5.

Second-strand synthesis was carried out by resuspending streptavidin beads in the following reaction mix: 5  $\mu$ L 10 $\times$  Blue Buffer, 1.25  $\mu$ L 10 mM dNTPs, 39.75  $\mu$ L H<sub>2</sub>O, 4  $\mu$ L custom-designed second-strand adapter. Samples were then incubated at 94 °C for 5 minutes, and at 4 °C for 5 minutes, after which 1.5  $\mu$ L Klenow (3'  $\rightarrow$  5' exo minus) were added, followed by further incubation at 4 °C for 15 minutes, at 37 °C for 30 minutes, and at 70 °C for 5 minutes.

Beads were separated on magnet and the chase reaction was carried out by resuspending in a mix of 5  $\mu$ L 10 $\times$  Thermo Pol Buffer, 1.25  $\mu$ L 10 mM dNTPs, 43.5  $\mu$ L H<sub>2</sub>O, and 1  $\mu$ L *Bst* DNA Polymerase Large Fragment (NEB). Samples were incubated at 65 °C for 30 minutes, then again separated on magnet.

PCR was performed on beads in 50  $\mu$ L reactions composed of 25  $\mu$ L 2 $\times$  NEB Next PCR Master Mix, 20  $\mu$ L H<sub>2</sub>O, 2.5  $\mu$ L i7 and 2.5  $\mu$ L i5 primers (both custom-designed), with initial extension at 72 °C for 3 min, denaturation at 98 °C for 30 sec, 15 cycles of 98 °C for 10 sec, 63 °C for 30 sec, and 72 °C for 30 sec, and final extension at 72 °C for 5 min. PCR reactions were cleaned up and size-selected using AMPure XP beads.

Libraries were sequenced on Illumina NextSeq or MiSeq instruments, as 2 $\times$ 75mers or 2 $\times$ 300mers, respectively.

### dSMF data processing

Bisulfite reads were trimmed using cutadapt (version 0.16) and Trim Galore (version 0.4.4), using the following settings (taking into account that the bisulfite sequencing libraries are generated with the PBAT protocol): `--clip_R1 9 --clip_R2 9 --three_prime_clip_r1 6 --three_prime_clip_r2 6 --paired`. Trimmed reads were mapped to the *sacCer3* version of the yeast genome using Bismark<sup>47</sup> (version 0.19.0) with the following settings: `--bowtie2 --pbat`. Methylation calls were extract using the `bismark_methylation_extractor` program within Bismark and the following settings: `-s --no_overlap --comprehensive --merge_non_CpG --cytosine_report --CX`.

### ATAC-seq

ATAC-seq was carried out on the same nuclei isolated for SMAC-seq as described above (before resuspension in M.CviPI Reaction Buffer), by resuspending nuclei with 25  $\mu$ L 2 $\times$  TD buffer (20 mM Tris-HCl pH 7.6, 10 mM MgCl<sub>2</sub>, 20% Dimethyl Formamide), 2.5  $\mu$ L transposase (custom produced) and 22.5  $\mu$ L nuclease-free H<sub>2</sub>O, and incubating at 37 °C for 30 min in a Thermomixer at 1000 RPM. Transposed DNA was isolated using the DNA Clean & Concentrator Kit (Zymo, cat # D4014) and PCR amplified as

described before<sup>48</sup>. Libraries were then sequenced on a Illumina NextSeq instrument as 2×36mers or as 2×75mers.

### ATAC-seq data processing

Demultiplexed fastq files were mapped to the `sacCer3` assembly of the *S. cerevisiae* genome as 2×36mers using Bowtie<sup>49</sup> with the following settings: `-v 2 -k 2 -m 1 --best --strata`. Duplicate reads were removed using `picard-tools` (version 1.99).

### ChIP-seq experiments

Cell lysis and ChIP reactions were performed as previously described<sup>50</sup> with minor modifications. Cells were fixed with 1% formaldehyde for 20 minutes (Rpb1-CTD and Rbp1-CTD-S2P ChIP) or 30 minutes (Hsf1-V5 ChIP) and quenched with 0.125 M glycine for 5 minutes. A total of ~50 ODs of cells were used per Rpb1-CTD or Rpb1-CTD-S2P ChIP and ~300 ODs per Hsf1-V5 ChIP. Fixed cell were washed 2× in cold 1× PBS, pelleted and stored at -80 °C. Pellets were lysed in 300 μL FA lysis buffer (50 mM HEPES-KOH pH 8.0, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate, 1 mM PMSF, Roche protease inhibitor) with ~1 mL ceramic beads on a Fastprep-24 (MP Biomedicals). The entire lysate was then collected and adjusted to 1 mL with FA lysis buffer before sonication with a 1/8' microtip on a Q500 sonicator (Qsonica) for 14 minutes (10 seconds on, 20 seconds off). The sample tube was held in a -20 °C 80% ethanol bath throughout sonication to prevent sample heating. After sonication, cell debris was pelleted and the supernatant was retained for ChIP. For each ChIP reaction, 30 μL Protein G Dynabeads (Invitrogen) were blocked (PBS + 0.5% BSA), prebound with 5-10 μL antibody (8wG16 Rpb1-CTD, Abcam cat # ab817); 3E10 Rpb1-CTD-S2P, Millipore cat # 04-1571-1) or SV5-Pk1 (anti-V5, BioRad cat # MCA1360G)) and washed 1× with PBS before incubation with supernatant (4 °C, overnight). Dynabeads were then washed (5 minutes per wash) 3× in FA lysis buffer, 3× in high-salt FA lysis buffer (50 mM Hepes-KOH pH 8.0, 500 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate, 1 mM PMSF), 1× in ChIP wash buffer (10 mM Tris-HCl pH 7.5, 0.25 M LiCl, 0.5% NP-40, 0.5% sodium deoxycholate, 1 mM EDTA, 1 mM PMSF) and 1× in TE wash buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 50 mM NaCl). DNA was eluted from the beads in ChIP elution buffer (50 mM Tris-HCl pH 7.5; 10 mM EDTA; 1% SDS) at 65 °C for 20 min. Eluted DNA was incubated at 65 °C overnight to reverse crosslinks, before treatment with RNase A (37 °C, 1 hour) and then Proteinase K (65 °C, 2 hours). DNA was purified using the ChIP DNA Clean & Concentrator kit (Zymo Research). Sequencing libraries were generated using the NEBNext Ultra II DNA Library Prep kit (NEB Cat # E7645) and sequenced on a Illumina NextSeq instrument as 2×36mers or as 2×75mers.

### ChIP-seq data processing

Demultiplexed fastq files were mapped to the `sacCer3` assembly of the *S. cerevisiae* genome as 2×36mers using Bowtie<sup>49</sup> with the following settings: `-v 2 -k 2 -m 1 --best --strata`. Duplicate reads were removed using `picard-tools` (version 1.99). Hsf1 peaks were called using MACS2<sup>51</sup> (version 2.1.0) with the following settings: `-g 12000000 -f BAMPE`.

### Multiread-preserving alignment and normalization

Multiread-preserving alignment and track generation was carried out by mapping reads to the `sacCer3` assembly of the *S. cerevisiae* genome using Bowtie<sup>49</sup> with the following settings: `-v 2 -a --best --strata`. Each alignment was then given a weight inversely proportional to the number of locations that the read maps to i.e. each position's score was normalized to RPMs as follows:

$$S_{c,i} = \frac{\sum_{R \in R_{c,i}} \frac{1}{NH_R}}{|R|} \cdot 10^6 \quad (3)$$

Where  $NH_R$  is the number of locations in the genome a read maps to.

### RNA-seq experiments

Cells (1 mL) were pelleted and flash frozen in liquid N<sub>2</sub>. Pellets were resuspended in 300 μL TRIzol and lysed with ~1 mL ceramic beads on a Fastprep-24 (MP Biomedicals). Cell debris were pelleted and RNA was extracted from the supernatant using the Direct-Zol RNA Microprep Kit (Zymo Research). RNA-seq libraries were generated using the NEBNext Ultra Directional RNA Library Prep Kit (NEB Cat # E7420)

### RNA-seq processing and gene expression quantification

RNA-seq reads were mapped to the yeast genome as 1×50mers (external datasets) or 2×75mers (diamide experiments) using TopHat version 2.0.8<sup>52</sup>. Gene-level quantifications in FPKMs (Fragments Per Kilobase per Million mapped reads) were generated using Cufflinks version 2.0.2<sup>52</sup>. The mean from all replicates was taken as the expression level for each gene for subsequent analyses.

### External sequencing datasets

A number of previously published *S. cerevisiae* genomics datasets were used in this study. ChIP-exo reads and called peaks for Abf1, Cbf1, Rap1 and Reb1 were downloaded from GEO accessions GSE93662 and GSE72106. ChIP-seq data for centromeric proteins was downloaded

from GEO accessions GSE31466 and GSE51949. PRO-seq and PRO-CAP data was obtained from GEO accession GSE76142. ORC ChIP-seq data was downloaded from GEO accession GSE16926. DNase-seq was downloaded from GEO accession GSE69651 while DGF data was downloaded from SRA accession SRP000620. MNase-seq data was obtained from GEO accessions GSE26493 and GSE29292, TBP ChIP-seq from GSE44200, Rpb1 ChIP-seq from GSE93190, Rpb3 ChIP-seq from GSE74787, RPC128 ChIP-seq from GSE39566, and Mediator subunits ChIP-seq from GSE95051. RNA-seq data from accession GSE85590 was also used. Except where otherwise stated, raw reads were aligned using Bowtie<sup>53</sup> with the following settings: `-v 2 --k 2 -m 1 --best --strata`, with the addition of `-X 1000` for paired-end reads. Paired-end reads were aligned as 2×25mers), while single-end reads were aligned as 1×36mers. PRO-seq and PRO-CAP data was aligned as 1×16mers.

### Micro-C data and processing

Micro-C data was downloaded from GEO accession GSE68016 and processed as described in the original publication<sup>34</sup>.

### Transcription factor motif mapping

Transcription factor motif recognition sequences were mapped genome-wide using FIMO<sup>54</sup> (version 4.11.2 of the MEME-Suite<sup>55</sup> using the CIS-BP database<sup>56</sup> as a reference set of position weight matrices.

### Gene annotation update

Publicly available gene models for *S. cerevisiae* do not contain TSS and TTS information for a major fraction of genes in the genome, only including the coding (“CDS”) portions instead. As the omission of UTRs presents a problem for TSS- and TTS-centered analyses, we updated the existing gene models following the approach described previously<sup>45</sup> and the *S. cerevisiae* TIF-seq dataset from GEO accession GSE39128<sup>57</sup>. New TSS and TTS positions were assigned to each gene for which such information was available based on the median UTR length as measured by TIF-seq.

### Nucleosome positioning information

H4S47C<sup>19,58</sup> chemical mapping data was downloaded from GEO accessions GSE59523 and GSE36063. H3Q85C<sup>59</sup> chemical mapping data was downloaded from GEO accession GSE97290. We used the nucleosome positioning calls obtained from the original 2012 Brogaard et al. study for our analyses, after transforming them from coordinates in the `sacCer2` version of the *S. cerevisiae* genome assembly to `sacCer3` using the `liftOver` function in the UCSC Genome Browser utilities toolkit.

### Mappability tracks generation

To evaluate unique read mappability, the whole genome was tiled with reads of given length at every position. The reads were then mapped back to the genome using the same settings used to map single-end ChIP-seq reads. For every position coverage by mapped reads was calculated, and mappability was scored as the ratio between read coverage and the read length used to tile the genome.

### Data availability

Short read datasets associated with this study are available through GEO accession GSE128290. Nanopore data is available through SRA BioProject PRJNA594057.

### Author contributions

Z.S, G.K.M and N.A.S.A. conceived and designed the study. Z.S, G.K.M and N.A.S.A. performed initial experiments. Z.S., M.P.S. and G.K.M. performed diamide time course experiments. G.K.M. and Z.S. analyzed data. W.J.G., A.K. and J.M.S. supervised the study. G.K.M., Z.S., W.J.G. and A.K. wrote the manuscript with input from all authors.

### Acknowledgments

This work was supported by NIH grants (P50HG007735, RO1 HG008140, U19AI057266 and UM1HG009442 to W.J.G., 1UM1HG009436 to W.J.G. and A.K., 1DP2OD022870-01 and 1U01HG009431 to A.K.), the Rita Allen Foundation (to W.J.G.), the Baxter Foundation Faculty Scholar Grant, and the Human Frontiers Science Program grant RGY006S (to W.J.G). W.J.G is a Chan Zuckerberg Biohub investigator and acknowledges grants 2017-174468 and 2018-182817 from the Chan Zuckerberg Initiative. Z.S. is supported by EMBO Long-Term Fellowship EMBO ALTF 1119-2016 and by Human Frontier Science Program Long-Term Fellowship HFSP LT 000835/2017-L. G.K.M. is supported by the Stanford School of Medicine Dean’s Fellowship. N.A.S.A. is funded by the Department of Defense through a National Defense Science and Engineering Grant and by a Stanford Graduate Fellowship. The authors would also like to thank members of the Greenleaf and Kundaje labs for their helpful suggestions and discussions on the subject over the course of the study.

### Competing Interest Statement

The authors declare no competing interests.

### References

1. Wu C. 1980. The 5' ends of *Drosophila* heat shock genes in chromatin are hypersensitive to DNase I. *Nature* **286**(5776):854–860.

2. Keene MA, Corces V, Lowenhaupt K, Elgin SC. 1981. DNase I hypersensitive sites in Drosophila chromatin occur at the 5' ends of regions of transcription. *Proc Natl Acad Sci U S A* **78**(1):143–146.
3. McGhee JD, Wood WI, Dolan M, Engel JD, Felsenfeld G. 1981. A 200 base pair region at the 5' end of the chicken adult  $\beta$ -globin gene is accessible to nuclease digestion. *Cell* **27**(1 Pt 2):45–55.
4. Dorschner MO, Hawrylycz M, Humbert R, Wallace JC, Shafer A, Kawamoto J, Mack J, Hall R, Goldy J, Sabo PJ, Kohli A, Li Q, McArthur M, Stamatoyannopoulos JA. 2004. High-throughput localization of functional elements by quantitative chromatin profiling. *Nat Methods* **1**(3):219–225.
5. Sabo PJ, Kuehn MS, Thurman R, Johnson BE, Johnson EM, Cao H, Yu M, Rosenzweig E, Goldy J, Haydock A, Weaver M, Shafer A, Lee K, Neri F, Humbert R, Singer MA, Richmond TA, Dorschner MO, McArthur M, Hawrylycz M, Green RD, Navas PA, Noble WS, Stamatoyannopoulos JA. 2006. Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays. *Nat Methods* **3**(7):511–518.
6. Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, Furey TS, Crawford GE. 2008. High-resolution mapping and characterization of open chromatin across the genome. *Cell* **132**(2):311–322.
7. Hesselberth JR, Chen X, Zhang Z, Sabo PJ, Sandstrom R, Reynolds AP, Thurman RE, Neph S, Kuehn MS, Noble WS, Fields S, Stamatoyannopoulos JA. 2009. Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. *Nat Methods* **6**(4):283–289.
8. 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**(12):1213–1218.
9. Schones DE, Cui K, Cuddapah S, Roh TY, Barski A, Wang Z, Wei G, Zhao K. 2008. Dynamic regulation of nucleosome positioning in the human genome. *Cell* **132**(5):887–898.
10. Kelly TK, Liu Y, Lay FD, Liang G, Berman BP, Jones PA. 2012. Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome Res* **22**(12):2497–2506.
11. Nabils NH, Deleyrolle LP, Darst RP, Riva A, Reynolds BA, Klädde MP. 2014. Multiplex mapping of chromatin accessibility and DNA methylation within targeted single molecules identifies epigenetic heterogeneity in neural stem cells and glioblastoma. *Genome Res* **24**(2):329–339.
12. Krebs AR, Imanci D, Hoerner L, Gaidatzis D, Burger L, Schübeler D. 2017. Genome-wide Single-Molecule Footprinting Reveals High RNA Polymerase II Turnover at Paused Promoters. *Mol Cell* **67**(3):411–422.e4.
13. Murray IA, Morgan RD, Luyten Y, Fomenkov A, Corrêa IR Jr, Dai N, Allaw MB, Zhang X, Cheng X, Roberts RJ. 2018. The non-specific adenine DNA methyltransferase M.EcoGII. *Nucleic Acids Res* **46**(2):840–848.
14. Simpson JT, Workman RE, Zuzarte PC, David M, Dursi LJ, Timp W. 2017. Detecting DNA cytosine methylation using nanopore sequencing. *Nat Methods* **14**(4):407–410.
15. Rand AC, Jain M, Eizenga JM, Musselman-Brown A, Olsen HE, Akeson M, Paten B. 2017. Mapping DNA methylation with high-throughput nanopore sequencing. *Nat Methods* **14**(4):411–413.
16. Feng S, Cokus SJ, Zhang X, Chen PY, Bostick M, Goll MG, Hetzel J, Jain J, Strauss SH, Halpern ME, Ukomadu C, Sadler KC, Pradhan S, Pellegrini M, Jacobsen SE. 2010. Conservation and divergence of methylation patterning in plants and animals. *Proc Natl Acad Sci U S A* **107**(19):8689–8694.
17. Zemach A, McDaniel IE, Silva P, Zilberman D. 2010. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* **328**(5980):916–919.
18. Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR. 2008. Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* **133**(3):523–536.
19. Brogaard K, Xi L, Wang JP, Widom J. 2012. A map of nucleosome positions in yeast at base-pair resolution. *Nature* **486**(7404):496–501.
20. Stoiber MH, Quick J, Egan R, Lee JE, Celniker SE, Neely R, Loman N, Pennacchio L, Brown JB. 2017. De novo Identification of DNA Modifications Enabled by Genome-Guided Nanopore Signal Processing. *bioRxiv* 094672
21. Li H. 2016. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics* **32**(14):2103–2110.
22. Cole HA, Howard BH, Clark DJ. 2011. The centromeric nucleosome of budding yeast is perfectly positioned and covers the entire centromere. *Proc Natl Acad Sci U S A* **108**(31):12687–12692.
23. Henikoff S, Ramachandran S, Krassovsky K, Bryson TD, Codomo CA, Brogaard K, Widom J, Wang JP, Henikoff JG. 2014. The budding yeast Centromere DNA Element II wraps a stable Cse4 hemisome in either orientation in vivo. *Elife* **3**:e01861.
24. Merz K, Hondele M, Goetze H, Gmelch K, Stoeckl U, Griesenbeck J. 2008. Actively transcribed rRNA genes in *S. cerevisiae* are organized in a specialized chromatin associated with the high-mobility group protein Hmo1 and are largely devoid of histone molecules. *Genes Dev* **22**(9):1190–1204.
25. Conconi A, Widmer RM, Koller T, Sogo JM. 1989. Two different chromatin structures coexist in ribosomal RNA genes throughout the cell cycle. *Cell* **57**(5):753–761.

26. French SL, Osheim YN, Cioci F, Nomura M, Beyer AL. 2003. In exponentially growing *Saccharomyces cerevisiae* cells, rRNA synthesis is determined by the summed RNA polymerase I loading rate rather than by the number of active genes. *Mol Cell Biol* **23**(5):1558–1568.
27. Goetze H, Wittner M, Hamperl S, Hondele M, Merz K, Stoeckl U, Griesenbeck J. 2010. Alternative chromatin structures of the 35S rRNA genes in *Saccharomyces cerevisiae* provide a molecular basis for the selective recruitment of RNA polymerases I and II. *Mol Cell Biol* **30**(8):2028–2045.
28. Panday A, Grove A. 2016. Yeast HMO1: Linker Histone Reinvented. *Microbiol Mol Biol Rev* **81**(1). pii: e00037–16.
29. Jones HS, Kawauchi J, Braglia P, Alen CM, Kent NA, Proudfoot NJ. 2007. RNA polymerase I in yeast transcribes dynamic nucleosomal rDNA. *Nat Struct Mol Biol* **14**(2):123–130.
30. Huang J, Moazed D. 2003. Association of the RENT complex with nontranscribed and coding regions of rDNA and a regional requirement for the replication fork block protein Fob1 in rDNA silencing. *Genes Dev* **17**(17):2162–2176.
31. Zhu F, Farnung L, Kaasinen E, Sahu B, Yin Y, Wei B, Dodonova SO, Nitta KR, Morgunova E, Taipale M, Cramer P, Taipale J. 2018. The interaction landscape between transcription factors and the nucleosome. *Nature* **562**(7725):76–81.
32. O’Sullivan JM, Tan-Wong SM, Morillon A, Lee B, Coles J, Mellor J, Proudfoot NJ. 2004. Gene loops juxtapose promoters and terminators in yeast. *Nat Genet* **36**(9):1014–1018.
33. Tan-Wong SM, Zaugg JB, Camblong J, Xu Z, Zhang DW, Mischo HE, Ansari AZ, Luscombe NM, Steinmetz LM, Proudfoot NJ. 2012. Gene loops enhance transcriptional directionality. *Science* **338**(6107):671–675.
34. Hsieh TH, Weiner A, Lajoie B, Dekker J, Friedman N, Rando OJ. 2015. Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. *Cell* **162**(1):108–119.
35. Weiner A, Hsieh TH, Appleboim A, Chen HV, Rahat A, Amit I, Rando OJ, Friedman N. 2015. High-resolution chromatin dynamics during a yeast stress response. *Mol Cell* **58**(2):371–386.
36. Morano KA, Grant CM, Moye-Rowley WS. 2012. The response to heat shock and oxidative stress in *Saccharomyces cerevisiae*. *Genetics* **190**(4):1157–1195.
37. Gabrieli T, Sharim H, Fridman D, Arbib N, Michaeli Y, Ebenstein Y. 2018. Selective nanopore sequencing of human BRCA1 by Cas9-assisted targeting of chromosome segments (CATCH). *Nucleic Acids Res* **46**(14):e87.
38. Staševskij Z, Gibas P, Gordevičius J, Kriukienė E, Klimišauskas S. 2017. Tethered Oligonucleotide-Primed Sequencing, TOP-Seq: A High-Resolution Economical Approach for DNA Epigenome Profiling. *Mol Cell* **65**(3):554–564.e6.
39. Fu Y, Luo GZ, Chen K, Deng X, Yu M, Han D, Hao Z, Liu J, Lu X, Doré LC, Weng X, Ji Q, Mets L, He C. 2015. N<sup>6</sup>-methyldeoxyadenosine marks active transcription start sites in *Chlamydomonas*. *Cell* **161**(4):879–892.
40. Wang Y, Chen X, Sheng Y, Liu Y, Gao S. 2017. N<sup>6</sup>-adenine DNA methylation is associated with the linker DNA of H2A.Z-containing well-positioned nucleosomes in Pol II-transcribed genes in *Tetrahymena*. *Nucleic Acids Res* **45**(20):11594–11606.
41. Luo GZ, Hao Z, Luo L, Shen M, Sparvoli D, Zheng Y, Zhang Z, Weng X, Chen K, Cui Q, Turkewitz AP, He C. 2018. N<sup>6</sup>-methyldeoxyadenosine directs nucleosome positioning in *Tetrahymena* DNA. *Genome Biol* **19**(1):200.
42. Timinskas A, Butkus V, Janulaitis A. 1995. Sequence motifs characteristic for DNA [cytosine-N4] and DNA [adenine-N6] methyltransferases. Classification of all DNA methyltransferases. *Gene* **157**(1–2):3–11.
43. Salter JD, Bennett RP, Smith HC. 2016. The APOBEC Protein Family: United by Structure, Divergent in Function. *Trends Biochem Sci* **41**(7):578–594.
44. Kawasaki F, Beraldi D, Hardisty RE, McInroy GR, van Delft P, Balasubramanian S. 2017. Genome-wide mapping of 5-hydroxymethyluracil in the eukaryote parasite *Leishmania*. *Genome Biol* **18**(1):23.
45. Schep AN, Buenrostro JD, Denny SK, Schwartz K, Sherlock G, Greenleaf WJ. 2015. Structured nucleosome fingerprints enable high-resolution mapping of chromatin architecture within regulatory regions. *Genome Res* **25**(11):1757–1770.
46. Miura F, Enomoto Y, Dairiki R, Ito T. 2012. Amplification-free whole-genome bisulfite sequencing by post-bisulfite adaptor tagging. *Nucleic Acids Res* **40**(17):e136.
47. Krueger F, Andrews SR. 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**(11):1571–1572.
48. Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S, Satpathy AT, Rubin AJ, Montine KS, Wu B, Kathiria A, Cho SW, Mumbach MR, Carter AC, Kasowski M, Orloff LA, Risca VI, Kundaje A, Khavari PA, Montine TJ, Greenleaf WJ, Chang HY. 2017. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods* **14**(10):959–962.
49. Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**(3):R25.
50. Hu B, Petela N, Kurze A, Chan KL, Chapard C, Nasmyth K. 2015. Biological chromodynamics: a general

- method for measuring protein occupancy across the genome by calibrating ChIP-seq. *Nucleic Acids Res* **43**(20):e132.
51. Feng J, Liu T, Qin B, Zhang Y, Liu XS. 2012. Identifying ChIP-seq enrichment using MACS. *Nat Protoc* **7**(9):1728–1740.
  52. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7**(3):562–578.
  53. Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**(3):R25.
  54. Grant CE, Bailey TL, Noble WS. 2011. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**(7):1017–1018.
  55. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* **37**(Web Server issue):W202–208.
  56. Weirauch MT, Yang A, Albu M, Cote AG, Montenegro-Montero A, Drewe P, Najafabadi HS, Lambert SA, Mann I, Cook K, Zheng H, Goity A, van Bakel H, Lozano JC, Galli M, Lewsey MG, Huang E, Mukherjee T, Chen X, Reece-Hoyes JS, Govindarajan S, Shaulsky G, Walhout AJ, Bouget FY, Ratsch G, Larrondo LF, Ecker JR, Hughes TR. 2014. Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **158**(6):1431–1443.
  57. Pelechano V, Wei W, Steinmetz LM. 2013. Extensive transcriptional heterogeneity revealed by isoform profiling. *Nature* **497**(7447):127–131.
  58. Ramachandran S, Zentner GE, Henikoff S. 2015. Asymmetric nucleosomes flank promoters in the budding yeast genome. *Genome Res* **25**(3):381–390.
  59. Chereji RV, Ramachandran S, Bryson TD, Henikoff S. 2018. Precise genome-wide mapping of single nucleosomes and linkers in vivo. *Genome Biol* **19**(1):19.
  60. Liang Z, Shen L, Cui X, Bao S, Geng Y, Yu G, Liang F, Xie S, Lu T, Gu X, Yu H. 2018. DNA N<sup>6</sup>-Adenine Methylation in *Arabidopsis thaliana*. *Dev Cell* **45**(3):406–416.e3.
  61. Greer EL, Blanco MA, Gu L, Sendinc E, Liu J, Ariztizábal-Corrales D, Hsu CH, Aravind L, He C, Shi Y. 2015. DNA Methylation on N6-Adenine in *C. elegans*. *Cell* **161**(4):868–878.
  62. Fu Y, Sinha M, Peterson CL, Weng Z. 2008. The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome. *PLoS Genet* **4**(7):e1000138.