



An integrated machine-learning model to predict nucleosome architecture

Alba Sala ^{1,†}, Mireia Labrador ^{1,†}, Diana Buitrago ¹, Pau De Jorge ¹, Federica Battistini ^{1,2}, Isabelle Brun Heath ¹ and Modesto Orozco ^{1,2,*}

¹Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Barcelona, Spain

²Departament de Bioquímica i Biomedicina, Universitat de Barcelona, Barcelona, Spain

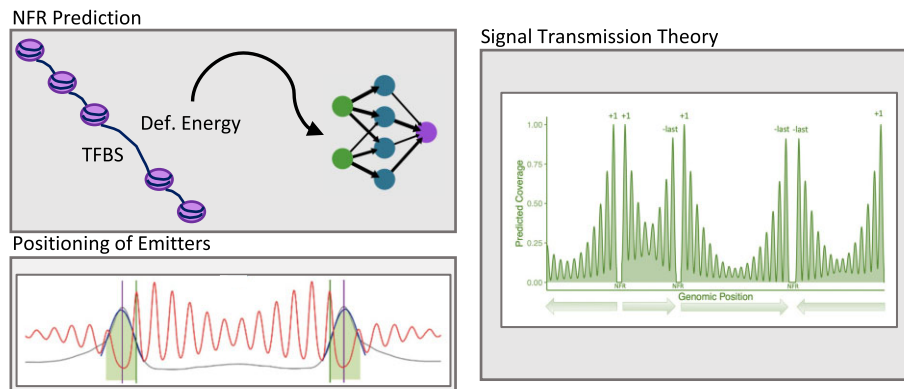
*To whom correspondence should be addressed. Tel: +34 93 40 37156; Fax: +35 93 403 7157; Email: modesto.orozco@irbbarcelona.org

†The first two authors should be regarded as Joint First Authors.

Abstract

We demonstrate that nucleosomes placed in the gene body can be accurately located from signal decay theory assuming two emitters located at the beginning and at the end of genes. These generated wave signals can be in phase (leading to well defined nucleosome arrays) or in antiphase (leading to fuzzy nucleosome architectures). We found that the first (+1) and the last (-last) nucleosomes are contiguous to regions signaled by transcription factor binding sites and unusual DNA physical properties that hinder nucleosome wrapping. Based on these analyses, we developed a method that combines Machine Learning and signal transmission theory able to predict the basal locations of the nucleosomes with an accuracy similar to that of experimental MNase-seq based methods.

Graphical abstract



Introduction

Nucleosomes (the basic units of eukaryotic chromatin) are formed by 147 bp of duplex DNA wrapped around an octamer of histones (1), followed by a linker DNA where, in complex eukaryotic organisms, an additional histone (H1) can be bound (2). Nucleosomes are not randomly placed but maintain a defined architecture along the genome, with certain positions occupied by well-positioned nucleosomes while others are nucleosome-free (3–9). Most significant nucleosome free regions (NFRs) are associated with the promoter regions of genes (upstream of the Transcription Start Sites, TSSs), the replication origins (ORIs) and the Transcription Termination Sites (TTSs) (10,11). The general consensus is that NFRs at TSSs are preferentially recognized by effector proteins in-

involved in the regulation of gene activity, and the widths of these regions correlate with gene expression (12). Furthermore, perturbation in nucleosome architectures associated to stress, changes in cell cycle phases, source of nutrients, or the cell metabolic cycle (6,11,13,14) proved the connection between nucleosome architecture and gene activity. The causality in this relationship is however unclear.

Over the last two decades, many efforts have been made to discover the main determinants of nucleosome positioning (15–20). Several studies have suggested that DNA physical properties are crucial for defining nucleosome positioning, with NFRs characterized by sequences where the mechanical cost of wrapping DNA around nucleosomes is very high (10,13,21). On the contrary, others have

Received: December 1, 2023. Revised: July 17, 2024. Editorial Decision: July 21, 2024. Accepted: July 29, 2024

© The Author(s) 2024. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License

(<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

Table 1. Genes modified with the 81-nt sequence in each strain

Gene	Strain	Strand	Start	End	Chromosome	Insert Position	Phasing
UBX5	1	+	1 127 872	1 129 374	chrIV	1 128 586	Phased
CKB2	2	+	405 768	406 544	chrXV	406 248	Phased
PPT1	3	-	736662	738 203	chrVII	737 615	Phased
TRP4	4	+	1 184 747	1 185 889	chrIV	1 185 196	Phased
BSP1	1	+	883 828	885 558	chrXVI	884 578	Control (not-phased)
DGK1	2	-	899056	899 928	chrXV	899 667	Control (not-phased)
SLM3	3	-	392659	393 912	chrIV	393 462	Control (not-phased)
PAN5	4	-	224030	225 169	chrVIII	224 581	Control (not-phased)

suggested that nucleosome positioning is dictated exclusively by cellular machinery involving a complex interplay between chromatin remodelers, transcription factors and RNA polymerase activity (22–25). Chromatin reconstitution experiments (13,22,26,27) demonstrated that NFRs are well reproduced *in vitro*, but their boundaries are not precise in the absence of cellular effectors. These findings suggest that while physical principles can signal NFRs, cellular machinery is required for the correct definition of their boundaries (18,25,28). Nonetheless, it is unclear how these intrinsic and extrinsic signals are combined to define nucleosome architecture.

We explore here whether the basal nucleosome architecture can be determined by easily available DNA descriptors such as sequence-dependent physical properties (29,30) and sequence annotations of transcription factor binding sites (TFBSs) (31). Additionally, we investigated whether changes in nucleosome architecture are a reason for, or a consequence of gene expression. With this aim, we developed a method that combines Machine Learning (ML) and signal transmission theory (STT) able to predict the most probable nucleosome architectures in yeast with accuracy comparable to experimental techniques. Furthermore, synthetic biology experiments demonstrate that the structural fingerprint of active genes (characterized by wide NFRs and phased nucleosome arrays) is a consequence, rather than a reason for their gene expression level (14,32).

Materials and methods

Yeast strains and growth conditions

The *Saccharomyces cerevisiae* PPY1 strain (*MATa his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 bar1::leu2*) was transformed with the appropriate DNA fragments to generate all the mutant strains used in this work. The PPY1 strain was obtained from Oscar Aparicio's lab at the University of Southern California, USA. For the selection of the mutant strains, we used YPD with or without 5-FOA (5-fluoroorotic acid) and SD (synthetic defined) with the required amino acids.

Mutant strains generation

We generated 4 mutant strains, with the 81-nt DNA sequence (5'-GCGTGTGTTGTGTTTTCTCCGAGGAGAAACATTCA AATCTTGTGCTATGGCTTTGCCTACCGTCTGCGCCA TCCATCTTTTCGC-3') inserted in the coding sequence of 2 selected genes per strain (Table 1). We selected four non-essential genes which showed phased nucleosomes (UBX5, CKB2, PPT1, TRP4; see phase definition below) together with four non-essential control genes which were unphased (BSP1, DGK1, SLM3, PAN5). The 81-nt sequence was designed not

to match any existing yeast sequence and not to favor nor disfavor nucleosome formation or affect the reading frame (see Results). The strains were produced using the *Delitto Perfetto* strategy described in (33).

RNA extraction and RT-qPCR

Three independent colonies from each yeast strain were grown until exponential phase and then arrested at late G1 by alpha-factor. RNA was obtained from 10 ml yeast cultures (OD₆₀₀ 0.8) using the hot-phenol method. cDNA synthesis was done with the First Strand cDNA Synthesis Kit (Roche) using oligo dT and following the provider instructions. Gene expression levels were determined by quantitative PCR using the Light-Cycler 480 sybr green I master (Roche). The oligonucleotides used for the qPCR are listed in [Supplementary Table S1](#).

Transcription inhibition

In order to determine the correct incubation time to inhibit transcription without killing the cells, we selected 2 genes with low RNA stability (RPA135 and NMD3) and 2 genes with high RNA stability (ACT1 and DGK1) to serve as controls (34). We then measured their mRNA level by qPCR after incubation with 10-phenanthroline at 100 μg/ml at 30°C during 0, 5, 15, 30 and 45 minutes. Using this approach, we observed that the amount of RPA135 and NMD3 mRNA started to decrease after 30 min. This incubation time was selected to perform the MNase-seq experiments on cells with inhibited transcription.

MNase digestion

The Micrococcal nuclease (MNase) digestion was performed on semi-intact yeast cells prepared as described elsewhere (35). We optimized the MNase digestion conditions for each sample to obtain about 80% of mononucleosomes. The integrity and size distribution of digested fragments were determined using the microfluidics-based platform Bioanalyzer (Agilent) prior to sample preparations and sequencing. The sample preparation was done using the Illumina TruSeq DNA sample preparation kit for whole genome sequencing, following the Illumina standard protocol. The libraries (paired-end) were sequenced paired-end on a HiSeq2000, v4, 2 × 75 bp, with approximately 10 M PE reads/sample.

Nucleosome calling

MNase-seq paired-end reads were mapped to customized versions of the yeast genome (SacCer3, UCSC), containing the inserted sequences in the modified genes, using the Bowtie (36) aligner, allowing up to two mismatches. Output files were imported in to R where reads were trimmed to 50 bp

maintaining the original center and transformed to reads per million bp. Peak calling was performed, after noise filtering, with the nucleR package implemented in the Nucleosome Dynamics platform (37,38) using the standard parameters for yeast: peak width of 147 bp, peak detection threshold of 35% and maximum overlap of 80 bp. Nucleosome calls were considered well-positioned when nucleR's peak width score and height score were higher than 0.6 and 0.4 (39) respectively, and fuzzy otherwise.

Nucleosome periodicity and phasing

Periodicity in nucleosome positioning was determined for each gene by computing the autocorrelation coefficient, as seen in (40):

$$R(T) = \int_{X_1}^{X_2} I(x) \cdot I(x - T) dx \quad (1)$$

where X_1 and X_2 stand for the limits of a sampling window (e.g. the position of TSS and TTS), I is the function representing nucleosome coverage for all genes and T is the period. This thus reflects the continuity of a nucleosome repeat length and will have a maximum when all the nucleosome peaks are T units apart as $R(T)$ will be the highest. In other words, autocorrelation is defined by the correlation between a profile and shifted versions of itself. This method can uncover hidden patterns in the signal that wouldn't be clear by just examining the strength of the signal itself. Autocorrelation coefficients for different periods were normalized as shown in:

$$\hat{R}(T) = \frac{R(T)}{R(0)} \quad (2)$$

Nucleosome period is defined as the value of T that optimizes $\hat{R}(T)$, and periodic genes are those showing large autocorrelation coefficient values (eq. 1). Phased genes are defined as those where the +1 to -last distance (L) is a multiple of the period (T). Anti-phased genes are those where the distance from integer (DFI) score, defined as the modulus of the ratio length/periodicity, is close to $T/2$. Unphased genes refer to intermediate values.

$$DFI = L - T \cdot \text{round}\left(\frac{L}{T}\right) \quad (3)$$

Signal transmission theory for nucleosome positions

Having observed experimentally two clear NFRs at the beginning and end of genes, we propose a simple signal decay model, where the coverage at a given position is given by the addition of two independent positioning signals emitted from the two ends of a gene, one starting from the +1 (i.e. the nucleosome right after the NFR at promoter region), and another one from the -last nucleosome (i.e. right before the TTS). The strength of the signal emitted from the two ends has been chosen to better reproduce the nucleosome coverage pattern. For this reason, the emitted value is strongest at the position of the +1 ($Cov^{+1}(X)$) compared to the emitter from the -last nucleosome ($Cov^{-last}(X')$), as can be seen in:

$$Cov^{+1}(X) = \left(1 + \alpha + \sin\left(\frac{\pi}{2} + 2 \cdot \frac{\pi}{T} X\right)\right) \sigma^{\left(\frac{|X|}{T}\right)} \quad (4)$$

$$Cov^{-last}(X') = \left(1 + \sin\left(\frac{\pi}{2} + 2 \cdot \frac{\pi}{T} X'\right)\right) \sigma^{\left(\frac{|X'|}{T}\right)} \quad (5)$$

where X is the distance from the +1 nucleosome and X' is the distance from the -last nucleosome, $X' = L - X$. The shifting factor α corrects for the higher density of reads at the +1 nucleosome and the decay factor σ accounts for the reduced coverage as we move away from the NFR. We evaluated different values for α and σ and selected those that maximized the correlation between the observed experimental coverage and the predicted (α was set to 0.2 and σ was set to 0.7). The total coverage is then normalized to guarantee an effective decay of the signal:

$$Cov(X) = \frac{Cov^{+1}(X) + Cov^{-last}(X')}{Cov^{+1}(0) + Cov^{-last}(0)} \quad (6)$$

where, $Cov^{+1}(0)$, $Cov^{-last}(0)$ are the values of the two emitting signals at the +1 nucleosome dyad, which are used as denominator to normalize the $Cov(X)$ to 1 at this position. The difference in strengths will be determined by the factor α which affects the normalization of the total coverage.

Deformation energy

The elastic energy associated to the DNA deformation from the naked to the nucleosome DNA was calculated in the harmonic regime using:

$$Energy = \frac{\sum_{j=1}^{146} E_j}{146} \text{ with } E_j = \frac{1}{2} \sum_{s=1}^6 \sum_{t=1}^6 k_{st}^j \Delta X_s^j \Delta X_t^j \quad (7)$$

where j stands for each of the 146 bp steps of the DNA stretches. E_j is the elastic energy required at each base pair step determined using the stiffness matrix (K), and ΔX_s^j and ΔX_t^j are the differences between the nucleosome and equilibrium values for the 6 bp step helical parameters (roll, twist, tilt, slide, rise or shift). The equilibrium values and stiffness constants for each individual base pair step were taken from MD simulations that cover all the unique base pair steps in all the possible tetranucleotide environments from microsecond-long parmbc1 simulations (29,30).

Machine learning

A Neural Network (NN) was developed to predict NFRs (and non-NFRs, i.e. regions occupied by nucleosomes) based exclusively on sequence information. The selected NN was defined by three layers: an *input layer* which depended on the selected centered window length, a 30-neuron *hidden layer* and a 2-neuron *output layer* to obtain the respective probabilities for the two studied classes. We used a *ReLU* activation function for the *hidden layer* and a *sigmoid* activation function for the output layer as defined in Keras (<https://keras.io>). The *binary crossentropy* function was minimized using a stochastic gradient descent optimizer with parameters 0.001 for the learning rate and 0.003 for the momentum. The NN was built using the Python library scikit-learn (v 0.20.3) (41), the software library TensorFlow (42) and its API Keras.

Data on nucleosome positions were obtained from MNase-seq experiments of yeast cells synchronized at the G1 phase and processed with nucleR as implemented in our Nucleosome Dynamics package (37,38). NFRs defined by nucleR proximal to TSS or TTS were used for training. As non-NFRs are more present in the data than NFRs, we randomly removed points from non-NFRs to obtain a balanced data set corresponding to the NFRs/one and non-NFRs/zero classes. A train/test ratio of 80/20 was used to reduce overtraining

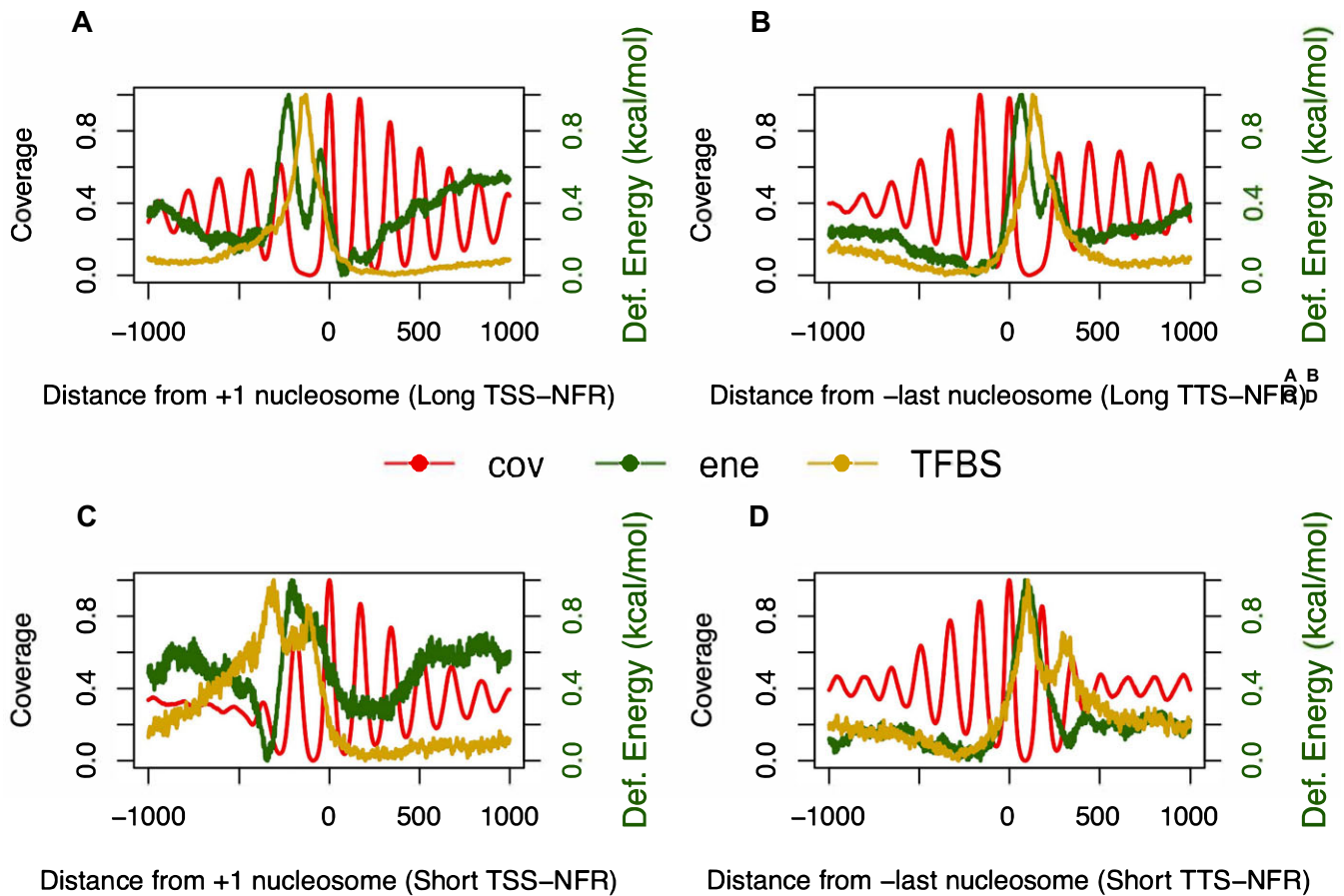


Figure 1. Average nucleosome coverage (red), TFBS density (yellow) and deformation energy (green) around well positioned +1 and -last nucleosomes for open NFRs (wider than 215 bp): Long TSSs-NFR (panel **A**, 2749 genes), Long TTSs-NFR (panel **B**, 1134 genes); and closed NFRs (shorter than 215 bp): Short TSSs-NFR (panel **C**, 644 genes) and Short TTSs-NFR (panel **D**, 942 genes).

artifacts and to choose the best ML algorithm. To explore the generality of the model some additional tests were done by training the model only with data from one chromosome (chrIV) and validating the model on the whole genome. Additionally, to avoid any potential bias training and testing on the same set of data, we re-trained our model with data from DANPOS (43) and tested on nucleR defined maps.

Results

NFRs are characterized by unique DNA physical properties and by high density of protein-recognition sequences.

Our analysis of the entire yeast genome reveals the placement of NFRs at the TSS and TTS of genes; the latter being present in different classes of gene, tandem and convergent, showing that the NFRs at the TTS is not a duplication of a neighboring TSS (see [Supplementary Figure S1](#)). Interestingly, both NFRs correspond to regions where the harmonic deformation energy (see [Methods](#)) required to wrap the DNA around the histone core is unusually high and where there is a large density of potential TFBS (Figure 1). Note that this behavior was found for both open (Figure 1A, B) and closed (Figure 1C, D) NFRs, suggesting the existence of a sequence-coded fingerprint characterizing all NFRs (at least at functionally relevant gene positions).

The differential characteristics of NFRs allowed us to train a Neural Network (NN) classifier to predict NFRs using as predictive features the deformation energy and experimental TFBS density profiles through the entire genome (see [Materials and methods](#)). These features are taken as stacked vector windows of size N around a center point, which define the first input layer of our neural network consisting of $N \times 2$ neurons (see [Supplementary Figure S2](#)). The resulting method has a good NFR prediction power as shown by the Area under the ROC Curve (AUC) of 96% and accuracy of 92% (Figure 2A, B; see also some examples of NFR predictions along the genome shown in [Supplementary Figure S3](#)). As described in [Methods](#), in order to demonstrate the consistency of our model we trained our predictor using only one large chromosome (chrIV) and tested it on all remaining chromosomes with good results (accuracy of 90% on the remaining chromosomes).

The NN-prediction peaks were then fitted to a Gaussian curve to define the 95% range (2 standard deviations of the mean; see Figure 2C-F) at which we place the +1 (TSS) and -last (TTS) nucleosomes. In order to test the precision of this procedure, we then considered the overall averaged experimental maps and located (44) the first peak before and after the NFR to which we assigned the dyad of the -last and +1 nucleosomes, respectively. After the positioning we calculated the absolute difference between the experimental positions and our predicted ones considering the Gaussian fitting. This

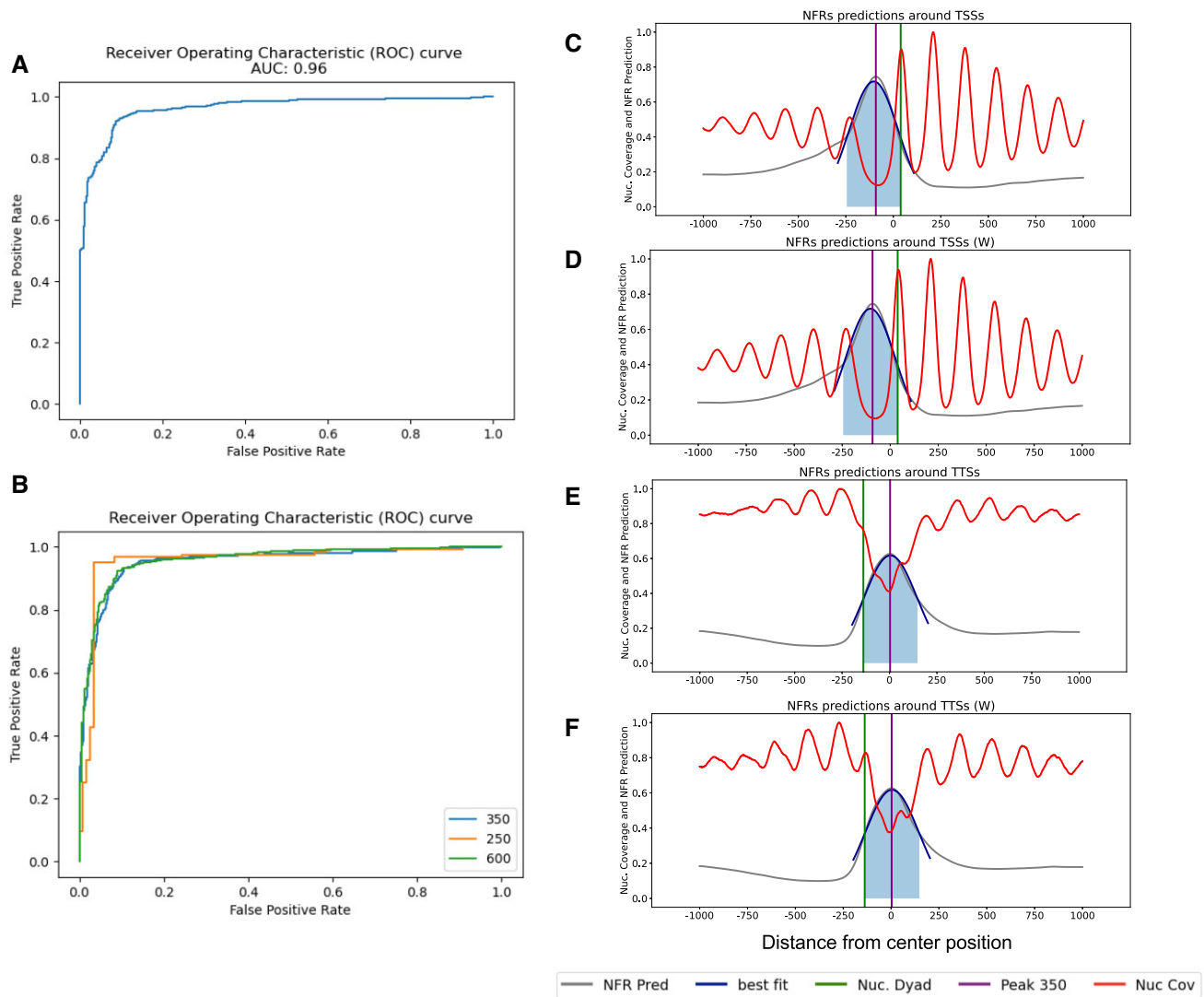


Figure 2. (A, B) panels showing the receiver operating characteristic (ROC) curve results from our Neural Network for (A) a 350 bp window and (B) different window sizes ROC curves (350 bp in blue, 250 bp in orange and 600 bp in green). NFR prediction (grey) against nucleosome experimental coverage (red) for (C) all TSSs (5676 genes), (D) well defined TSSs (3393 genes), (E) all TTSSs (5676 genes) and (F) well defined TTSSs (2076 genes). Green lines denote the average prediction of the +1 (in C and D) and -last (in E and F) nucleosomes, 2 stds from a fitted Gaussian distribution (dark blue). Purple lines mark the peak of the NFR probability prediction. Around 19% of the genes were excluded from the analysis given that they were missing the +1 and/or -last nucleosome experimental calls (see [Supplementary Table S2](#)).

simple method allowed us to position the +1 and -last nucleosomes with striking accuracy: only 4 bp (+1) and 17 bp (-last) away from the detected peaks (44) (Figure 2C, D for TSSs and 2E, F for TTSSs). The results from increasing and decreasing our predetermined standard deviation from our Gaussian fitting allowed us to observe a decrease in accuracy as we move away from two standard deviations. This was also the case when considering separately and optimizing for different classes of genes (tandem or convergent). Additionally, training our model and performing the same analysis for larger and shorter windows (600 and 250 bp), we observed similar or worse results (data not shown).

We repeated the same procedure but training only using a single chromosome (chrIV) and obtained similar results (see [Supplementary Figure S4A–D](#)). Quite interestingly, our method works for both open (the distance between the two dyads is wider than 215 bp) and closed NFRs (the distance between the two dyads is shorter than 215 bp) (39), especially when predicting the crucial +1 nucleosome (see

[Supplementary Figure S4E–H](#)). In order to further validate our model and discard possible biases given by training and testing the algorithm on the same dataset, we trained our algorithm with nucleosome positioning maps obtained from a widely used calling tool DANPOS (43) and tested it with our experimental data. The results obtained showed that our model trained with DANPOS nucleosome positions was still able to position the +1 and -last nucleosomes, respectively 9 and 11 bp away from the experimental average peak (see [Supplementary Figure S5](#)). This is similar to the distances obtained when the algorithm was trained with maps from nucleR.

Nucleosome positioning along gene body is determined by distance-decayed periodic signals.

We showed above how nucleosome positioning at the beginning and end of genes can be defined by a combination of intrinsic and extrinsic properties coded in the DNA sequence.

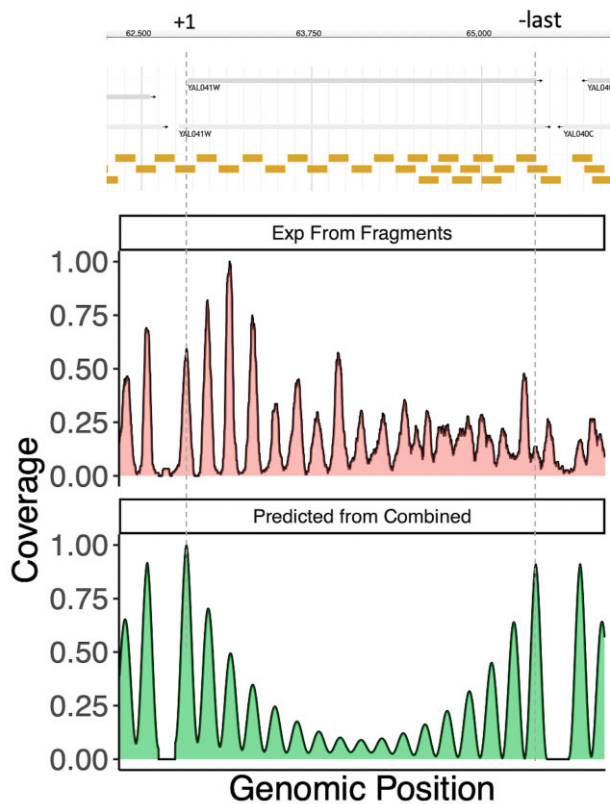


Figure 3. Scheme of resulting calls using nucleR (top golden boxes) with the coverage coming from the experimental mapping (red profile) against the predicted coverage from our signal theory combined prediction; see Materials and methods (green profile). Note the cell variability detected as multiple peak callings detected by nucleR in the top plot.

Our next step was to predict the placement of nucleosomes in the gene body and to understand why some of these nucleosomes appear well-positioned giving clear signal in the experimental MNase-seq maps, while others appear quite delocalized leading to fuzzy signals. Firstly, we determined nucleosome periodicity by computing the autocorrelation coefficient (see Methods for details) from the nucleosome coverage from one of our MNase-seq experiments for different given periods (T ; see Eqs. 1–2), finding a clear peak at 165 bp (the average nucleosome repeat length in yeast (45); see Supplementary Figure S6). This establishes reasonably well the distance between the +1 and the –last nucleosomes (DFI, see Methods) as a multiple of T , but as expected not the distance between the TSS and TTS which we find to be more uniform (Supplementary Figure S7).

Taking the experimental positions of the +1 and –last nucleosomes as emitting sites of a distance decaying signal (see Materials and methods, Eqs. 4–6), we can predict the positions of the intragenic nucleosomes (this is defined throughout the paper as our *combined prediction*) with good accuracy (Figure 3). As expected, the model performs best when predicting those nucleosomes which are well positioned close to the TSS, while the largest deviations are found in central regions, where nucleosomes show more cell-dependent variability (see the green error bars in Supplementary Figure S8A), leading to fuzzier signals. Similarly, at the TTS our prediction improves when the experimental nucleosomes are well

positioned, and we perform worse when we observe fuzzier architectures.

Globally, we could reproduce well the nucleosome architecture within the gene (green box in Supplementary Figure S8B) with 85% of the experimental nucleosomes correctly predicted, and an average distance to the experimental peak (as determined by nucleR) of 19bp. Note that, matching the prediction from our model, experimental MNase-seq maps (Figure 4A) show the presence of phased genes, where the +1 and –last signals add up to define clear and periodic nucleosome patterns, and unphased genes, where signals can partially cancel out in the middle of the gene, leading to diffuse nucleosome patterns (see examples in Figure 4B, and profiles in Figure 4C), illustrating cell variability in these regions. Additionally, it is also worth noting the stronger intensity detected experimentally of the +1 nucleosome emitter compared to the –last one, something that is considered and well reproduced by our signal-decay model (see Materials and methods).

Predicting nucleosome positioning along the entire genome

We can now join the NFR predictor and the periodic signals from STT to reconstruct the nucleosome architecture at the genome level (*full prediction* throughout the paper). This method allowed us to reproduce 78% of the nucleosome profile with an average distance of 32 ± 22 bp from the experimental peaks determined by nucleR (blue box in Supplementary Figure S8B), without any experimental information on the position of the +1 and –last nucleosomes. The average distance to nucleR peaks compares well to the average experimental distance of 37 bp found between the centers of each individual read and the corresponding peak obtained from the coverage of all the reads (37) (red box Supplementary Figure S8B), indicating that our nucleosome position estimate is within the intrinsic experimental noise derived from cellular heterogeneity. In our *full prediction* we detected a milder increase in noise as we displaced to the center of the gene than that found with our *combined prediction* (i.e. STT prediction based on experimental +1 and –last nucleosomes (blue bars in Supplementary Figure S8A)). Finally, our *full prediction* model is able to distinguish well between phase and unphased genes as detected experimentally (Supplementary Figure S9).

Next, we explored the robustness of the predicted nucleosome arrays to changes in gene expression (10). To this end, we checked the ability of our G1-trained model to reproduce nucleosome arrays within cells collected at different cell cycle phases or determined from completely different experiments (11). As seen in Supplementary Figure S10 the model performs very well in reproducing not only G1, but also M and S data. This suggests that there is a basal nucleosome architecture, which can be reasonably well reproduced by our simple predictive model, despite of specific changes related to the action of the cellular machinery.

Nucleosome architecture and gene expression are coupled in a complex way

Results above strongly suggest that nucleosomes along the genes are located based on periodicity rules from signals derived from the presence of NFRs, which position the +1 and –last nucleosomes, organizing the rest of the nucleosome string, which can be well ordered in the case of phased genes,

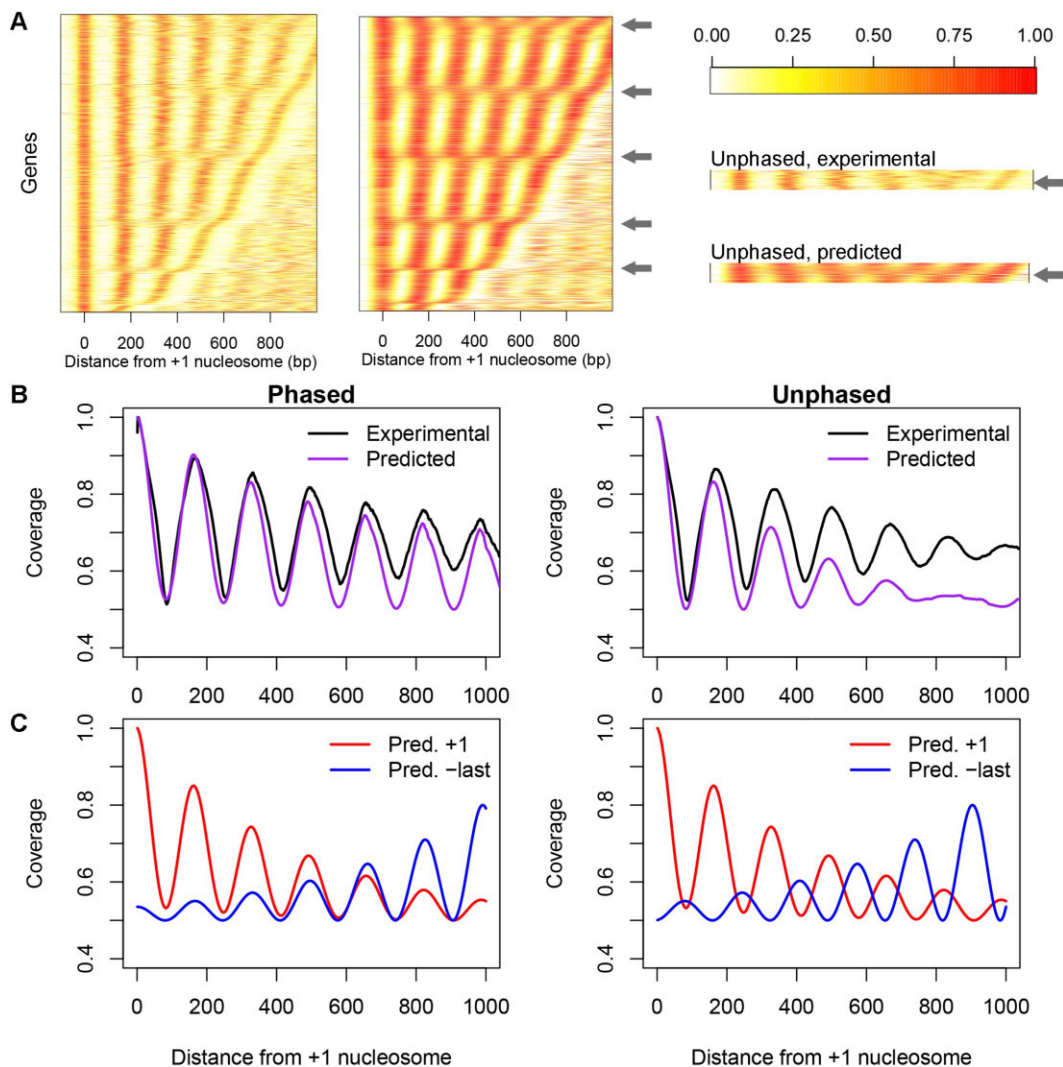


Figure 4. Signal decay model of nucleosome positioning. (A) Experimental (left panel) and predicted (right panel) nucleosome coverage for each gene, with respect to the +1 nucleosome. Genes are sorted by the distance between the +1 and the -last nucleosomes. Colour scale corresponds to normalized nucleosome coverage, from 1 (red) to 0 (white). (B) Nucleosome coverage, experimental (black) and predicted (purple, see Materials and methods Eq. 6) from the +1 nucleosome, averaged across all genes. Genes are split into phased or unphased based on $DFI < 10$ and $DFI > 40$, respectively. (C) Signals from the +1 (red) and the -last nucleosomes (blue) to predict the experimental coverage (see Materials and methods Eqs. 4–5).

or fuzzier in the case of the non-phased ones. Analysis of MNase-seq and RNA-seq datasets from (46) show that, as expected (14,32), transcriptionally active genes are associated with wider NFRs. Interestingly, while nucleosome positions are not altered dramatically, the coverage is more periodic along the gene bodies in transcribed regions in comparison to the inactive counterparts (Figure 5A and C) as was previously observed (40). This finding, which is also clear when looking at predicted nucleosome coverage (Figure 5B and D), reveals a correlation between expression and periodicity, a result that agrees with the ‘crystal-like’ behavior of nucleosome in active genes suggested by Vaillant *et al.* (7).

While the correlation between the architecture of the nucleosome array and gene activity is clear, the causality link is not so obvious. To clarify this point, we inserted an innocuous 81-nt sequence in a linker region approximately at the middle of the coding sequence of 8 non-essential genes (4 genes with phased nucleosomes and 4 with unphased nucleosomes). The insert was placed in a linker region to avoid direct interference with specific nucleosomes (see Figure 6A and

Supplementary Figure S11). Additionally, we tested the effect on nucleosome positioning and gene expression. For technical reasons, we only modified 2 genes per strain so we built 4 strains in total, with one phased and one unphased gene modified per strain (see Figure 6, Table 1 and Materials and methods). In principle, we could expect three different scenarios: (i) a displacement of the +1 and -last nucleosomes to recover the original phasing; (ii) a coordinated small displacement of all nucleosomes to recover phasing; and (iii) an increase in the fuzziness of the nucleosome string. Results in Table 2, Figure 6B and C, Supplementary Figures S12–S14 demonstrate that the introduction of the DNA segment in the phase genes leads to the generation of fuzzier nucleosome arrays, but neither to a significant change in the placement of the +1 and -last nucleosomes, nor to a coordinate sliding of nucleosomes. This suggests that irrespectively of other nucleosomes, the +1 and -last are placed in well-defined regions marked by sequence-dependent intrinsic and extrinsic factors, while the rest of the nucleosomes are placed based on periodicity considerations as described above.

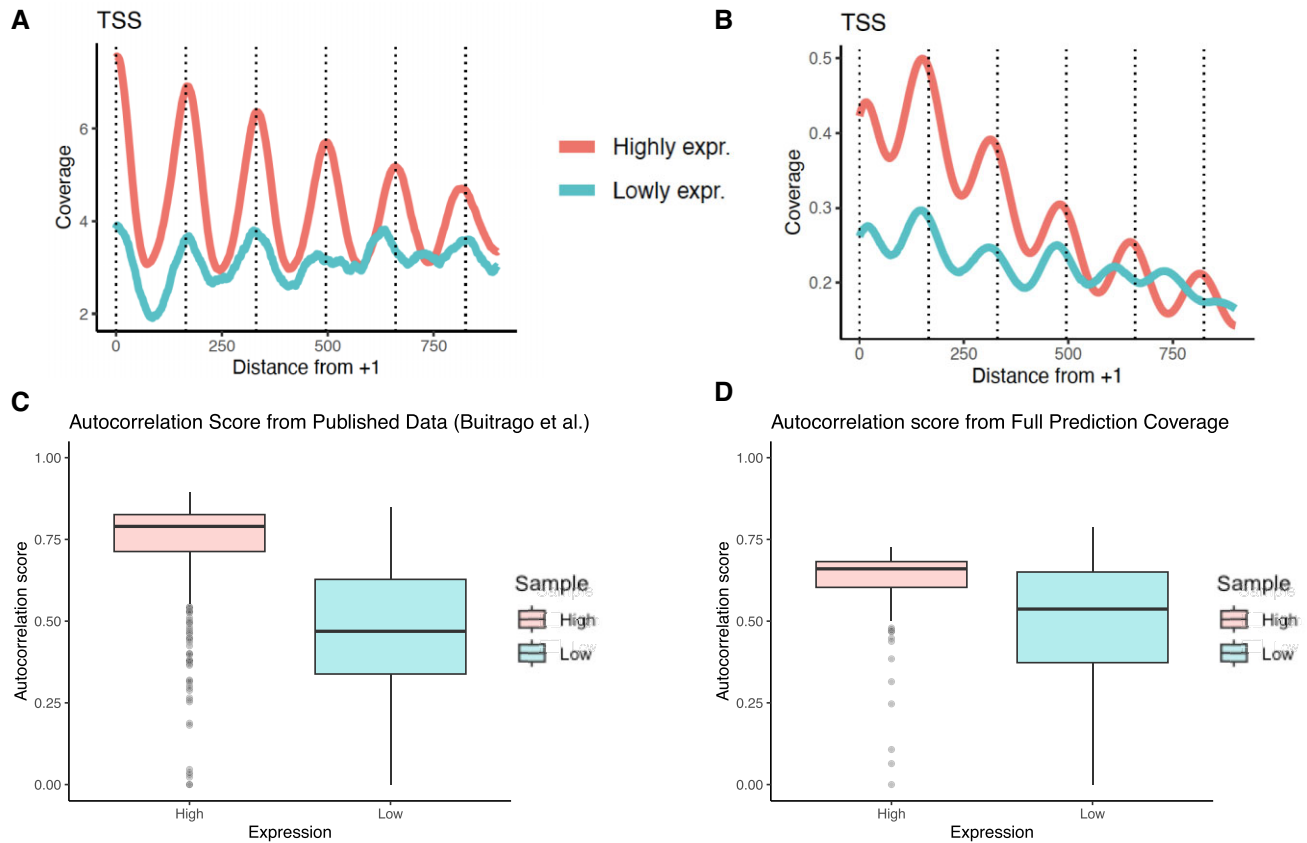


Figure 5. Coverage for highly (713 genes) and lowly (669 genes) expressed genes against a periodic nucleosome repeat length for (A) our experimental method and (B) combined predicted coverage. Autocorrelation scores for highly (red) and lowly (blue) expressed genes derived from (C) MNase-seq experimental data and (D) our coverage from our full prediction (see Materials and methods).

Interestingly, analysis of mRNA levels for the eight modified genes with or without the 81-nt segment only showed significant changes for two of the eight genes (a 2.2 \log_2 fold change for PPT1 transcript, and a mild decrease for CKB2 transcript) (Supplementary Figure S15) suggesting that changes in nucleosome architecture do not necessarily lead to changes in expression.

In order to investigate this further, the treatment with 1,10-pt was performed in conditions that lead to a decrease in RNAPII signals in ChIP-seq experiments (47). Inhibition of transcription by treatment with 1,10-phenanthroline (1,10-pt) led to an increased fuzziness of the nucleosome array and a slight displacement of the +1 and (specially) the -last nucleosomes (Figure 7A, B, Table 3 and Supplementary Figure S16), with a significant decrease of the autocorrelation score observed in the four strains tested (Figure 7C and Supplementary Table S3). 1,10-Phenanthroline is a metal chelating agent with high affinity for divalent cations. However, its mechanism of action to inhibit transcription has not been clearly established. Therefore, to discard an artefact caused by the use of 1,10-pt, we repeated the analysis using data from a previously published work on transcription and nucleosome positioning (12). We observed that inactivation of RNAPII using the temperature-sensitive (ts) allele *rpb1-1* causes an increase in nucleosome fuzziness and leads to a wider NFR at the TSS due mostly to a shift of the -1 nucleosome, in agreement with our results obtained with 1,10-pt (see Table 3 and Supplementary Figure S17).

In conclusion, we can outline a preferential (most likely not unique) causality arrow: expression activity \rightarrow changes in nucleosome architecture, with a direct role of polymerase or the elongation complex in reinforcing positioning signals within the gene body.

Finally, we benchmarked our predictor to the state-of-the-art NuPoP (48) method, taking as reference our experimental MNase nucleosome maps (grey profiles in Supplementary Figure 18). The profile calculated using our full predictor (blue profile in Supplementary Figure 18) could position with high accuracy the +1 and -last nucleosomes and the periodicity in the middle of the genes, the largest differences being in the definition of the intensity of the peaks. NuPoP (green profile in Supplementary Figure 18) while being quite accurate, has more problems in the localization of the +1 nucleosome and generates a very fuzzy and irregular profile, overpopulating some regions of nucleosomes, while fully depleting others, and leading to flat peaks that do not correspond to the read distributions found experimentally.

While our model reproduces with high accuracy the nucleosome architecture in yeast, other elements come in to play when understanding more complex organisms such as mammalian genome architectures. We performed a similar analysis to predict the probability of having a NFR in human data using annotated TFBS from the UCSC Genome Browser (49) and we calculated the deformation energy. The model was trained and tested on data from a single chromosome (chr1)

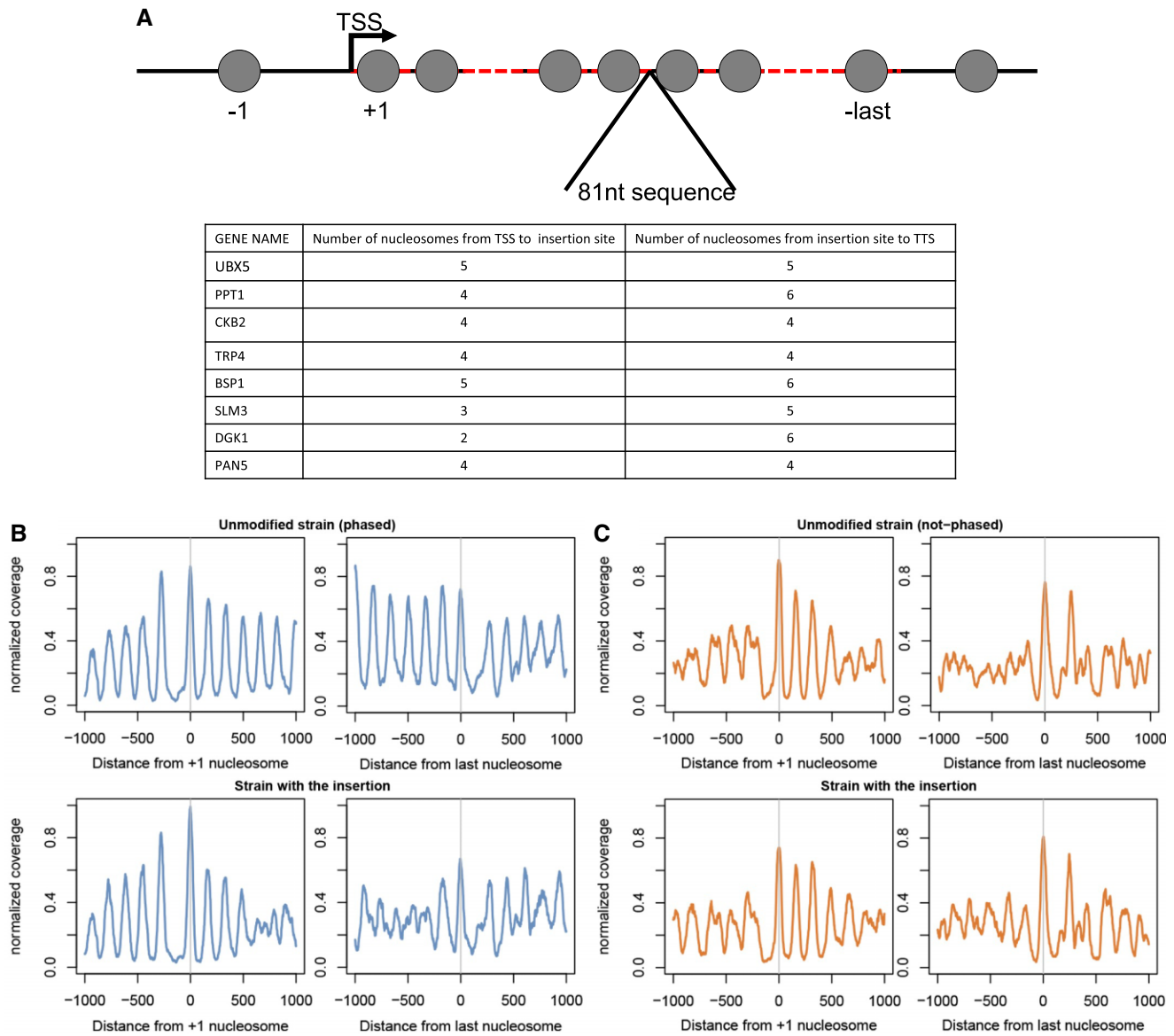


Figure 6. (A) Schematic representation of the experimental design. The exact location of the 81-nt insertion for each of the eight genes is indicated in Table 1 and represented in Supplementary Figures S13 and S14B and C. Nucleosome coverage for the selected genes in the unmodified strain (top panels) and the strain with the 81-nt insertion (bottom panels). Average of (B) the four genes phased in the unmodified strain (UBX5, CKB2, PPT1 and TRP4) (blue line) and (C) the four genes unphased in the unmodified strain (BSP1, DGK1, SLM3 and PAN5) (orange line).

Table 2. Phase score (DFI) and autocorrelation (R) in the unmodified strain and in the strain with the 81-nt insertion in the selected genes. Genes UBX5, CKB2, PPT1 and TRP4 are phased, and genes BSP1, DGK1, SLM3 and PAN5 are unphased in the unmodified strain

Gene	Unmodified strain		Strain with the 81-nt insert	
	DFI	R	DFI	R
UBX5	7	0.79 424	79	0.76 687
CKB2	1	0.81 862	79	0.68 771
PPT1	13	0.88 553	71	0.82 500
TRP4	6	0.80 702	80	0.73 771
BSP1	77	0.73 841	7	0.77 164
DGK1	46	0.69 007	31	0.81 921
SLM3	37	0.61 947	42	0.60 019
PAN5	44	0.73 341	46	0.81 708

obtaining an AUC of almost 70% (see Supplementary Figure S19) outperforming what would be expected from random.

Discussion

Nucleosome positioning in the gene body can be predicted with good accuracy by signal transmission theory (STT), assuming the existence of two well-positioned nucleosomes at the +1 and -last positions, which emit periodic signals whose intensities decay with distance. Phased genes (i.e. those whose distances between the +1 and -last is a multiple of 165) have periodic nucleosome signals, while unphased genes (and at a lower extend non-phased genes) tend to have fuzzy nucleosomes in the middle of the gene body. Change in distance between the +1 and -last leads to changes in nucleosome periodicity fully predictable by the theory and in the fuzzi-

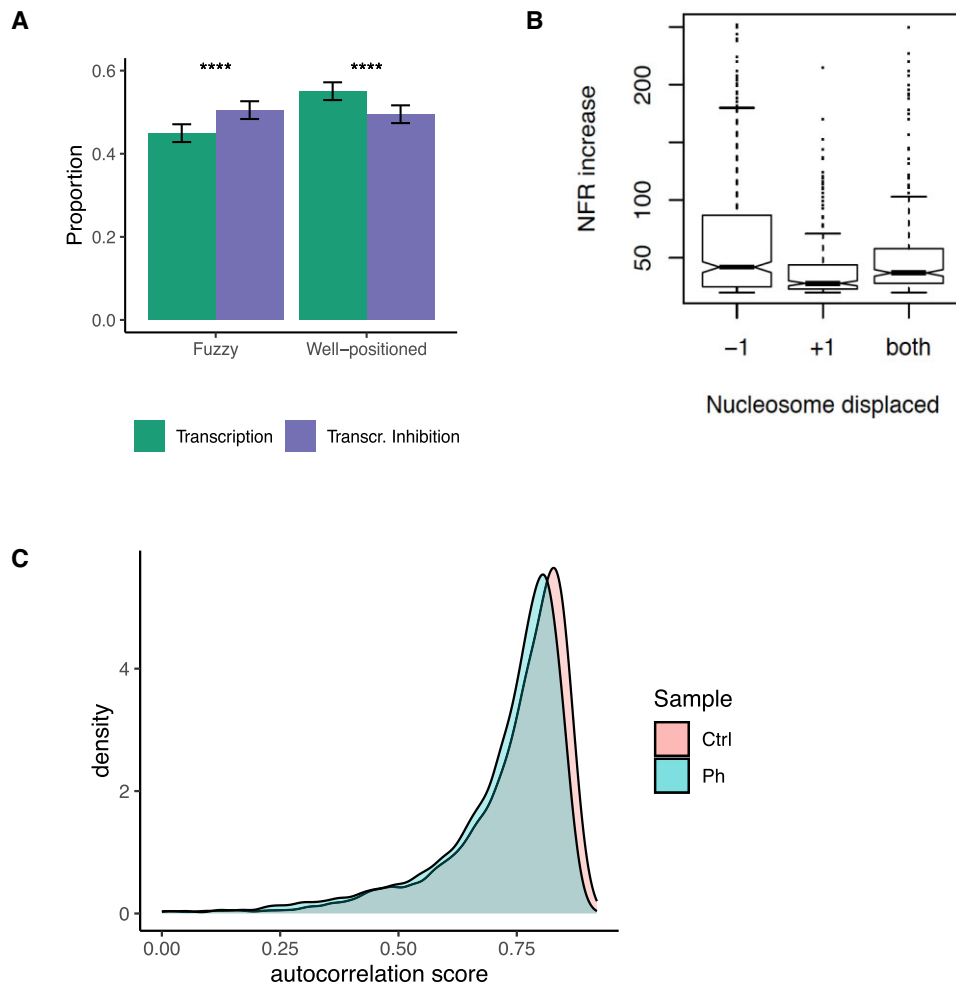


Figure 7. Effect of transcription on nucleosome positioning. **(A)** Change in the proportion of Fuzzy and Well-positioned nucleosomes upon transcription inhibition, with bars indicating relative standard error. **(B)** Change in NFRs' width at the TSS (–1 to +1 nucleosome distance) upon transcription inhibition in the presence of 1,10-phenanthroline (only cases with significant displacements (>20 bp) are considered in the box plots). **(C)** Mean autocorrelation scores of control (Ctrl) and phenanthroline (Ph) samples for the 4 strains previously described (see Figure 6A and Table 1) and for all the genes having a well-defined +1 and a -last nucleosome.

Table 3. NFRs' width increase at the TSS for genes displacing –1, +1 or both nucleosomes upon transcription inhibition from our 1,10-phenanthroline inhibition and published work on the inactivation of RNAPII by a ts-allele (12)

NFR width mean with std deviation (bp)	1,10-Phenanthroline inhibition	ts-allele inhibition
Transcription (1)	242.5 ± 3.9	196.5 ± 65
Transcription inhibition (2)	251.1 ± 1.2	214.5 ± 69
Difference (2, 1)	8.6 ± 4.6	18.0 ± 73

ness of nucleosomes in the middle of the gene. Very interestingly, the placement of the +1 and –last nucleosomes can be defined by the vicinity of NFRs, i.e. segments of DNA depleted of nucleosomes, which in turn, can be predicted by a simple neural network considering physical descriptors of DNA and TFBS densities. This is the case for both tandem and convergent genes, where the overall effects of neighboring genes are still captured by the model even though we consider each gene independently. Future models can investigate if a more specific gene level model that takes into account neigh-

boring effects/cooperativity would benefit from a substantial increase in accuracy or contrary, not compensate for the increase in complexity. The combination of our machine learning algorithm to localize the NFR, and consequently the +1 and –last nucleosomes, with the emission of two periodic signals on opposite direction allows us to predict the ground state positioning of the nucleosomes through the gene body with an accuracy similar to the experimental noise associated to MNase-seq. This suggests that nucleosome positions are quite well defined in the absence of complex mechanisms involving chromatin remodelers. Obviously, this 'ground state' nucleosome architecture can be modified to satisfy cellular needs by a myriad of factors, including among others, epigenetic signals, effector proteins or remodelers. Our signal transmission theory using two emitters assembled a pattern that could accurately reproduced the state positions of nucleosome and the dependency of coverage on transcription activity. To precisely reproduce the experimental coverage, new attributes should be included in the model to match the intensity. However, the predictive power of the G1-trained model is maintained for nucleosome architectures detected in phase S and M, suggesting the existence of a basal nucleosome con-

figuration, which can be modified to adapt to the cellular needs.

When we applied our model to the human genome, we obtained more accurate results than what would be expected from a random model. Considering the additional layers to deconvolute when studying more complex organisms and that the current model and architecture was optimized for yeast, our current approach with a fine tuning methodology and additional experimental data, shows the potential to be used to study any nucleosome positioning array.

Results presented here show the existence of a clear connection between expression level and the organization of nucleosome arrays, but while changes in nucleosome phasing do not lead to alteration in gene activity, transcription inhibition results in loss of order in the nucleosome string. Thus, our results suggest a causal order expression level \rightarrow nucleosome architecture, with a role of RNA polymerase, and/or the transcription elongation complex, in refining nucleosome strings that goes beyond unfolding nucleosomes (50), supporting the mechanism of pausing and histone transfer suggested by recent cryoEM studies of RNA polymerase II and the histone chaperone FACT (Facilitates Chromatin Transcription) (51–53).

Data availability

All relevant data supporting the key findings of this study are available within the article and the Supplementary Information. The datasets generated and/or analyzed during the current study are available in the ArrayExpress repository under the following accession number EMTAB-13613 and in GEO under GSE255857. For a detailed description of the generated data see [Supplementary Table S4](#). The code for the subsequent analysis is available in the GitHub repository (<https://github.com/Jalbiti/NucleosomePeriodicity>) and on Zenodo (<https://doi.org/10.5281/zenodo.12793703>).

Supplementary data

[Supplementary Data](#) are available at NAR Online.

Acknowledgements

This work was supported by the Center of Excellence for HPC H2020 European Commission; ‘BioExcel-2 – Centre of Excellence for Computational Biomolecular Research’ [823830]; BioExcel-3: Centre of Excellence for Computational Biomolecular Research [European Union: 101093290; Ministerio de Ciencia e Innovación: PCI2022-134976-2]; Spanish Ministry of Science [PID2021-122478NB-I00]; Instituto de Salud Carlos III – Instituto Nacional de Bioinformática, Fondo Europeo de Desarrollo Regional [ISCIII PT 17/0009/0007]; European Regional Development Fund, ERFD Operative Programme for Catalunya, the Catalan Government AGAUR [SGR2021 00863]. The IRB Barcelona is the recipient of a Severo Ochoa Award of Excellence from the MINECO.

Funding

Ministerio de Ciencia e Innovación [PCI2022-134976-2, PID2021-122478NB-I00]; European Regional Development Fund, ERFD Operative Programme for Catalunya,

the Catalan Government AGAUR [SGR2021 00863]; Center of Excellence for HPC H2020 European Commission [101093290, 823830]; Instituto de Salud Carlos III–Instituto Nacional de Bioinformática, Fondo Europeo de Desarrollo Regional [ISCIII PT 17/0009/0007]. Funding for open access charge: Center of Excellence for HPC H2020 European Commission; ‘BioExcel-2 – Centre of Excellence for Computational Biomolecular Research’ [823830].

Conflict of interest statement

None declared.

References

- Richmond, T.J. and Davey, C.A. (2003) The structure of DNA in the nucleosome core. *Nature*, **423**, 145–150.
- Izzo, A., Kamieniarz, K. and Schneider, R. (2008) The histone H1 family: specific members, specific functions? *Bchm*, **389**, 333–343.
- Yuan, G.-C., Liu, Y.-J., Dion, M.F., Slack, M.D., Wu, L.F., Altschuler, S.J. and Rando, O.J. (2005) Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science*, **309**, 626–630.
- Mavrich, T.N., Ioshikhes, I.P., Venters, B.J., Jiang, C., Tomsho, L.P., Qi, J., Schuster, S.C., Albert, I. and Pugh, B.F. (2008) A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Res.*, **18**, 1073–1083.
- Mavrich, T.N., Jiang, C., Ioshikhes, I.P., Li, X., Venters, B.J., Zanton, S.J., Tomsho, L.P., Qi, J., Glaser, R.L., Schuster, S.C., *et al.* (2008) Nucleosome organization in the *Drosophila* genome. *Nature*, **453**, 358–362.
- Shivaswamy, S., Bhinge, A., Zhao, Y., Jones, S., Hirst, M. and Iyer, V.R. (2008) Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation. *PLoS Biol.*, **6**, e65.
- Vaillant, C., Palmeira, L., Chevereau, G., Audit, B., D’Aubenton-Carafa, Y., Thermes, C. and Arneodo, A. (2010) A novel strategy of transcription regulation by intragenic nucleosome ordering. *Genome Res.*, **20**, 59–67.
- Valouev, A., Johnson, S.M., Boyd, S.D., Smith, C.L., Fire, A.Z. and Sidow, A. (2011) Determinants of nucleosome organization in primary human cells. *Nature*, **474**, 516–520.
- Baldi, S., Krebs, S., Blum, H. and Becker, P.B. (2018) Genome-wide measurement of local nucleosome array regularity and spacing by nanopore sequencing. *Nat. Struct. Mol. Biol.*, **25**, 894–901.
- Deniz, Ö., Flores, O., Battistini, F., Pérez, A., Soler-López, M. and Orozco, M. (2011) Physical properties of naked DNA influence nucleosome positioning and correlate with transcription start and termination sites in yeast. *Bmc Genomics [Electronic Resource]*, **12**, 489.
- Deniz, Ö., Flores, O., Aldea, M., Soler-López, M. and Orozco, M. (2016) Nucleosome architecture throughout the cell cycle. *Sci. Rep.*, **6**, 19729.
- Weiner, A., Hughes, A., Yassour, M., Rando, O.J. and Friedman, N. (2010) High-resolution nucleosome mapping reveals transcription-dependent promoter packaging. *Genome Res.*, **20**, 90–100.
- Kaplan, N., Moore, I.K., Fondufe-Mittendorf, Y., Gossett, A.J., Tillo, D., Field, Y., LeProust, E.M., Hughes, T.R., Lieb, J.D., Widom, J., *et al.* (2009) The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature*, **458**, 362–366.
- Nocetti, N. and Whitehouse, I. (2016) Nucleosome repositioning underlies dynamic gene expression. *Genes Dev.*, **30**, 660–672.
- Jiang, C. and Pugh, B.F. (2009) Nucleosome positioning and gene regulation: advances through genomics. *Nat. Rev. Genet.*, **10**, 161–172.
- Segal, E. and Widom, J. (2009) What controls nucleosome positions? *Trends Genet.*, **25**, 335–343.

17. Clark,D.J. (2010) Nucleosome positioning, Nucleosome spacing and the Nucleosome code. *J. Biomol. Struct. Dyn.*, **27**, 781–793.
18. Struhl,K. and Segal,E. (2013) Determinants of nucleosome positioning. *Nat. Struct. Mol. Biol.*, **20**, 267–273.
19. Lieleg,C., Krietenstein,N., Walker,M. and Korber,P. (2015) Nucleosome positioning in yeasts: methods, maps, and mechanisms. *Chromosoma*, **124**, 131–151.
20. Chereji,R.V. and Clark,D.J. (2018) Major determinants of nucleosome positioning. *Biophys. J.*, **114**, 2279–2289.
21. Suter,B. (2000) Poly(dAmiddle dotdT) sequences exist as rigid DNA structures in nucleosome-free yeast promoters in vivo. *Nucleic Acids Res.*, **28**, 4083–4089.
22. Zhang,Y., Moqtaderi,Z., Rattner,B.P., Euskirchen,G., Snyder,M., Kadonaga,J.T., Liu,X.S. and Struhl,K. (2009) Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo. *Nat. Struct. Mol. Biol.*, **16**, 847–852.
23. Hughes,A.L., Jin,Y., Rando,O.J. and Struhl,K. (2012) A functional evolutionary approach to identify determinants of nucleosome positioning: a unifying model for establishing the genome-wide pattern. *Mol. Cell*, **48**, 5–15.
24. Lorch,Y., Maier-Davis,B. and Kornberg,R.D. (2014) Role of DNA sequence in chromatin remodeling and the formation of nucleosome-free regions. *Genes Dev.*, **28**, 2492–2497.
25. Kubik,S., Bruzzone,M.J., Challal,D., Dreos,R., Mattarocci,S., Bucher,P., Libri,D. and Shore,D. (2019) Opposing chromatin remodelers control transcription initiation frequency and start site selection. *Nat. Struct. Mol. Biol.*, **26**, 744–754.
26. Krietenstein,N., Wal,M., Watanabe,S., Park,B., Peterson,C.L., Pugh,B.F. and Korber,P. (2016) Genomic nucleosome organization reconstituted with pure proteins. *Cell*, **167**, 709–721.
27. Zhang,Z., Wippo,C.J., Wal,M., Ward,E., Korber,P. and Pugh,B.F. (2011) A packing mechanism for nucleosome organization reconstituted across a eukaryotic genome. *Science*, **332**, 977–980.
28. Zhang,Z. and Pugh,B.F. (2011) High-resolution genome-wide mapping of the primary structure of chromatin. *Cell*, **144**, 175–186.
29. Dans,P.D., Balaceanu,A., Pasi,M., Patelli,A.S., Petkevičiūtė,D., Walther,J., Hospital,A., Bayarri,G., Lavery,R., Maddocks,J.H., et al. (2019) The static and dynamic structural heterogeneities of B-DNA: extending Calladine–Dickerson rules. *Nucleic Acids Res.*, **47**, 11090–11102.
30. Walther,J., Dans,P.D., Balaceanu,A., Hospital,A., Bayarri,G. and Orozco,M. (2020) A multi-modal coarse grained model of DNA flexibility mappable to the atomistic level. *Nucleic Acids Res.*, **48**, e29.
31. Pachkov,M., Balwierz,P.J., Arnold,P., Ozonov,E. and van Nimwegen,E. (2012) SwissRegulon, a database of genome-wide annotations of regulatory sites: recent updates. *Nucleic Acids Res.*, **41**, D214–D220.
32. Bai,L. and Morozov,A.V. (2010) Gene regulation by nucleosome positioning. *Trends Genet.*, **26**, 476–483.
33. Storici,F. and Resnick,M.A. (2006) The Delitto perfetto approach to In vivo site-directed mutagenesis and chromosome rearrangements with synthetic oligonucleotides in yeast. *Methods Enzymol.*, **409**, 329–345.
34. Grigull,J., Mnaimneh,S., Pootoolal,J., Robinson,M.D. and Hughes,T.R. (2004) Genome-wide analysis of mRNA stability using transcription inhibitors and microarrays reveals posttranscriptional control of ribosome biogenesis factors. *Mol. Cell Biol.*, **24**, 5534–5547.
35. Schlenstedt,G., Hurt,E., Doye,V. and Silver,P.A. (1993) Reconstitution of nuclear protein transport with semi-intact yeast cells. *J. Cell Biol.*, **123**, 785–798.
36. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
37. Flores,O. and Orozco,M. (2011) nucleR: a package for non-parametric nucleosome positioning. *Bioinformatics*, **27**, 2149–2150.
38. Buitrago,D., Codó,L., Illa,R., de Jorge,P., Battistini,F., Flores,O., Bayarri,G., Royo,R., Del Pino,M., Heath,S., et al. (2019) Nucleosome Dynamics: a new tool for the dynamic analysis of nucleosome positioning. *Nucleic Acids Res.*, **47**, 9511–9523.
39. Flores,O., Deniz,Ö., Soler-López,M. and Orozco,M. (2014) Fuzziness and noise in nucleosomal architecture. *Nucleic Acids Res.*, **42**, 4934–4946.
40. Wan,J., Lin,J., Zack,D.J. and Qian,J. (2009) Relating periodicity of nucleosome organization and gene regulation. *Bioinformatics*, **25**, 1782–1788.
41. Pedregosa,F., Varoquaux,G., Gramfort,A., Michel,V., Thirion,B., Grisel,O., Blondel,M., Müller,A., Nothman,J., Louppe,G., et al. (2012) Scikit-learn: machine learning in Python. arXiv doi: <https://arxiv.org/abs/1201.0490>, 05 January 2018, preprint: not peer reviewed.
42. Abadi,M., Barham,P., Chen,J., Chen,Z., Davis,A., Dean,J., Devin,M., Ghemawat,S., Irving,G., Isard,M., et al. (2016) TensorFlow: a system for large-scale machine learning. arXiv doi: <https://arxiv.org/abs/1605.08695>, 31 May 2016, preprint: not peer reviewed.
43. Chen,K., Xi,Y., Pan,X., Li,Z., Kaestner,K., Tyler,J., Dent,S., He,X. and Li,W. (2013) DANPOS: dynamic analysis of nucleosome position and occupancy by sequencing. *Genome Res.*, **23**, 341–351.
44. Virtanen,P., Gommers,R., Oliphant,T.E., Haberland,M., Reddy,T., Cournapeau,D., Burovski,E., Peterson,P., Weckesser,W., Bright,J., et al. (2020) SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods*, **17**, 261–272.
45. Ocampo,J., Chereji,R.V., Eriksson,P.R. and Clark,D.J. (2016) The ISW1 and CHD1 ATP-dependent chromatin remodelers compete to set nucleosome spacing in vivo. *Nucleic Acids Res.*, **44**, 4625–4635.
46. Buitrago,D., Labrador,M., Arcon,J.P., Lema,R., Flores,O., Esteve-Codina,A., Blanc,J., Villegas,N., Bellido,D., Gut,M., et al. (2021) Impact of DNA methylation on 3D genome structure. *Nat. Commun.*, **12**, 3243.
47. Martin,B.J.E., Brind’Amour,J., Kuzmin,A., Jensen,K.N., Liu,Z.C., Lorincz,M. and Howe,L.J. (2021) Transcription shapes genome-wide histone acetylation patterns. *Nat. Commun.*, **12**, 210.
48. Wang,J.-P., Fondufe-Mittendorf,Y., Xi,L., Tsai,G.-F., Segal,E. and Widom,J. (2008) Preferentially quantized linker DNA lengths in *Saccharomyces cerevisiae*. *PLoS Comput. Biol.*, **4**, e1000175.
49. Nassar,L.R., Barber,G.P., Benet-Pagès,A., Casper,J., Clawson,H., Diekhans,M., Fischer,C., Gonzalez,J.N., Hinrichs,A.S., Lee,B.T., et al. (2023) The UCSC Genome Browser database: 2023 update. *Nucleic Acids Res.*, **51**, D1188–D1195.
50. Schwabish,M.A. and Struhl,K. (2004) Evidence for eviction and rapid deposition of histones upon transcriptional elongation by RNA polymerase II. *Mol. Cell Biol.*, **24**, 10111–10117.
51. Žumer,K., Maier,K.C., Farnung,L., Jaeger,M.G., Rus,P., Winter,G. and Cramer,P. (2021) Two distinct mechanisms of RNA polymerase II elongation stimulation in vivo. *Mol. Cell*, **81**, 3096–3109.
52. Kujirai,T., Ehara,H., Fujino,Y., Shirouzu,M., Sekine,S. and Kurumizaka,H. (2018) Structural basis of the nucleosome transition during RNA polymerase II passage. *Science*, **362**, 595–598.
53. Kujirai,T., Ehara,H., Sekine,S. and Kurumizaka,H. (2023) Structural transition of the nucleosome during transcription elongation. *Cells*, **12**, 1388.

Received: December 1, 2023. Revised: July 17, 2024. Editorial Decision: July 21, 2024. Accepted: July 29, 2024

© The Author(s) 2024. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.