

Article

Superstructure Detection in Nucleosome Distribution Shows Common Pattern within a Chromosome and within the Genome

Sujeet Kumar Mishra ^{1,2}, Kunhe Li ¹, Simon Brauburger ¹, Arnab Bhattacharjee ², Nestor Norio Oiwa ^{1,3} and Dieter W. Heermann ^{1,*}

¹ Institute for Theoretical Physics, Heidelberg University, D-69120 Heidelberg, Germany; sujeetsankrityan@gmail.com (S.K.M.); li@thphys.uni-heidelberg.de (K.L.); brauburger@stud.uni-heidelberg.de (S.B.); oiwa@thphys.uni-heidelberg.de (N.N.O.)

² School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi 110067, India; arnab@jnu.ac.in

³ Department of Basic Science, Universidade Federal Fluminense, Nova Friburgo 28625-650, Brazil

* Correspondence: heermann@thphys.uni-heidelberg.de

Abstract: Nucleosome positioning plays an important role in crucial biological processes such as replication, transcription, and gene regulation. It has been widely used to predict the genome's function and chromatin organisation. So far, the studies of patterns in nucleosome positioning have been limited to transcription start sites, CTCFs binding sites, and some promoter and loci regions. The genome-wide organisational pattern remains unknown. We have developed a theoretical model to coarse-grain nucleosome positioning data in order to obtain patterns in their distribution. Using hierarchical clustering on the auto-correlation function of this coarse-grained nucleosome positioning data, a genome-wide clustering is obtained for *Candida albicans*. The clustering shows the existence beyond hetero- and eu-chromatin inside the chromosomes. These non-trivial clusterings correspond to different nucleosome distributions and gene densities governing differential gene expression patterns. Moreover, these distribution patterns inside the chromosome appeared to be conserved throughout the genome and within species. The pipeline of the coarse grain nucleosome positioning sequence to identify underlying genomic organisation used in our study is novel, and the classifications obtained are unique and consistent.

Keywords: chromatin; nucleosome positioning; nucleosome distribution; heterochromatin; euchromatin; structure classification



Citation: Mishra, S.K.; Li, K.; Brauburger, S.; Bhattacharjee, A.; Oiwa, N.N.; Heermann, D.W. Superstructure Detection in Nucleosome Distribution Shows Common Pattern within a Chromosome and within the Genome. *Life* **2022**, *12*, 541. <https://doi.org/10.3390/life12040541>

Academic Editor: Eva Bartova

Received: 26 January 2022

Accepted: 23 March 2022

Published: 6 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The genomes of all higher eukaryotes are organised in different structures on multi-length scales [1,2]. Of these organisational structures, the chromosome is the biggest one, being observable under a normal light microscope. The smallest organisational structure, one level above the double helix DNA, is the nucleosome where 147 base pairs (bp) of DNA are wrapped 1.65 times around a histone octamer [3–5]. The arrays of nucleosomes organise to form the chromatin fibre, which folds into two mutually excluded structural domains, namely “heterochromatin” and “euchromatin”. The “heterochromatin” regions are enriched with inactive/repressive genes and are usually positioned closer to the periphery of the nucleus. The “euchromatin” regions contain transcriptionally active chromatin [3,6,7], which are genes located in the interior of the nucleus. The hierarchical packaging of chromatin renders the genome a very compact conformation that provides controlled accessibility of the regulatory DNA sequences (genes) by other DNA-binding proteins (DBPs) [8,9]. Thus, the chromatin organisation is tightly linked to gene regulation and warrants detailed investigation. Various experimental techniques have been developed to probe the hierarchical chromatin organisation at different length scales. For instance,

the “chromatin conformation capture” experiment (e.g., 3C and HiC) [2,10,11] captures the organisation of chromatin in a kbp to Mbp length scale, revealing the formation of topologically associated domains (TADs) [12] and chromatin loops [13,14]. Further characterisation of the chromatin fibre at the length scale of genes (\sim kbp) is achieved by the Micro-C technique that captures the intra-chromatin interactions at a resolution of \sim 100 bp within an organisation module called chromosomal interaction domains (CIDs) [15,16]. CIDs are much smaller but still similar to TADs. These structural organisations are strongly regulated by the nucleosome positions, length of linker regions, and presence of nucleosome-depleted regions (NDR) across the chromosome [17].

The term “nucleosome positioning” refers to the location of nucleosomes along the sequence of genomic DNA. Nucleosome positioning is determined by several factors, including DNA sequence [18,19], DNA-binding proteins [20,21], nucleosome remodelers [22–24], RNA polymerases [25], and more. Although nucleosome positioning is a dynamic process, the sequence-based mapping approach identifies its position only in a cell- and time-averaged manner. The technology of micrococcal nuclease (MNase) digestion combined with high-throughput sequencing (MNase-seq) [26] is a powerful method to map the genome-wide distribution of nucleosome positioning and its occupancy. The resulting occupancy maps are ensemble averages of heterogeneous cell populations and may also be influenced by titration [27]. However, it is necessary to retrieve the cell-specific features from the population average to reveal the mechanism of nucleosome organisation and its translocation along the genome. Zhang et al. has developed an algorithm called “Nucleosome Positioning from Sequencing” (NPS) to predict accurate nucleosome positioning from the MNase-seq data, which was later improved to iNPS (improved NPS) [28]. The nucleosome positioning here is considered as an average static picture where they implicitly consider the nucleosome dynamics in the form of snapshots at different time- and cell-averages. This nucleosome positioning provides the frequency of its occurrence from which peaks are annotated to obtain possible nucleosome location along the sequence. In short, the nucleosome positioning data from iNPS are simply the most probable nucleosome position along the chromosome. Furthermore, extensive studies have been performed to recognise nucleosome positioning patterns around CTCFs, transcription start sites (TSSs), exons and introns, promoter and loci regions locally. For instance, a typical nucleosome distribution around TSSs indicates nucleosome depletion, resulting in a nucleosome-free region (NFR), whereas the nucleosomes downstream of TSS are equally spaced [29]. A similar observation around CTCF is obtained: an array of well-positioned nucleosomes flank the sites occupied by the insulator binding protein CTCF across the human genome [30]. Despite the efforts, the global picture of nucleosome positioning remains elusive until a recent study that has reported three types of nucleosomal arrangement by analyzing the nucleosome spacing and phasing in a genome [31]. The evenly spaced nucleosomes in the array are termed as a regular array and irregular otherwise. At a given genomic location in the cell population, nucleosomes may also assume similar positions and are referred to as phased arrays. The phased-regular nucleosome arrays, being most prominent, are the hallmark of chromatin and found to be conserved from yeast to mammals. These phased-regular nucleosome arrays are mostly found near the promoter regions of transcribed genes in the yeast genome and near the binding sites of high-affinity DBPs in higher eukaryotes. However, the findings have limited applicability only at local regions of the chromatin fibre and provide absolutely no information about the nucleosome organisation along a complete chromosome or genome.

We used a theoretical approach to obtain a novel classification of segments across the chromosome based on the similarity in nucleosome patterns. The nucleosome positioning data are used as inputs that are systematically coarse-grained to analyze their auto-correlation function to search for any pattern. The results are processed using hierarchical clustering techniques to investigate if there exists any unique pattern of nucleosome. Our results suggest that the positions and occupancy of nucleosomes in a chromosome are not random; rather, they reveal distinct patterns of distribution within a chromosome. Interestingly, the patterns appear to be conserved within the genome as well and are in

agreement with the previous study that has reported three distinct nucleosome organisations across the genome. Furthermore, at the chromosome level, our approach could capture a few unique patterns in the range of the ~ 50 kbp length scale, which repeatedly occur throughout the chromosomes, indicating they might play a crucial role in regulating gene networks at a more local scale. The study underpins the nucleosome positioning architecture inside a genome that can provide insights into the genome organisation (c.f. Figure 1) not known before.

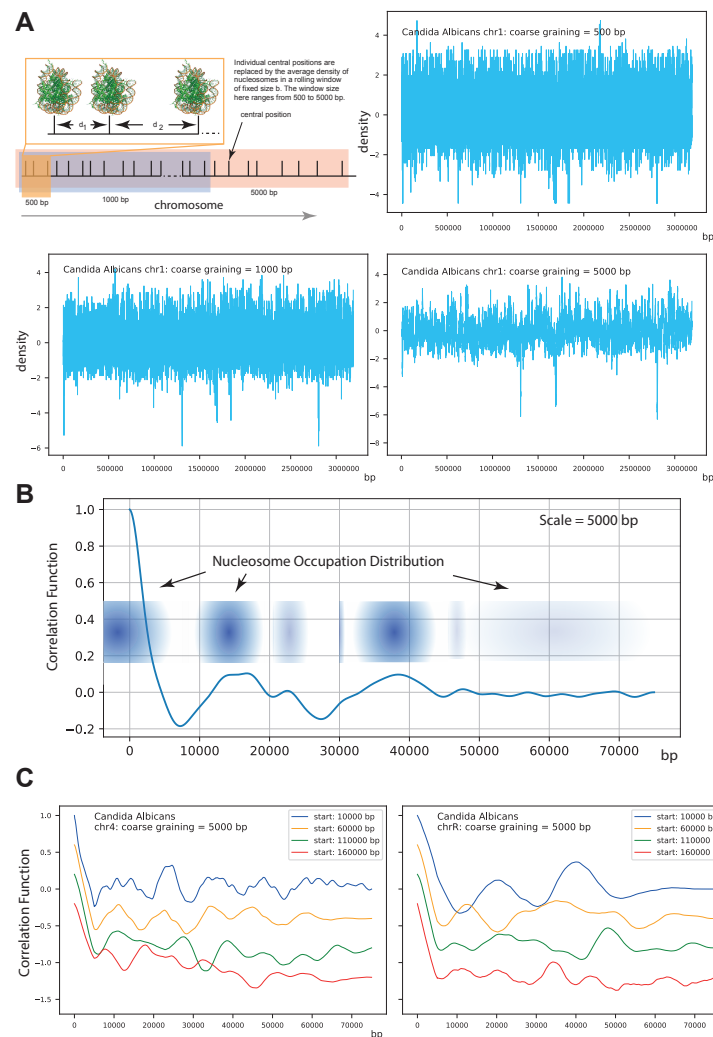


Figure 1. (A) shows the performed coarse-graining procedure and results for coarse-graining lengths L of 500 bp, 1000 bp, and 5000 bp. More structure is visible as b is increased. Going up even further washes out the structure. This is typical for systems with an intrinsic length scale. (B) shows the correlation among the coarse-grained super nucleosomes. The structure is that of a system exhibiting short range-order that is liquid-like with first and second nearest neighbor peaks. If there is no order or correlation, then the correlation function would be constant. On the other hand, if one would see strong regular peaks, this would indicate a regular ordering with the peak distances giving the preferred distance between the coarse-grained nucleosomes. The oscillatory characteristic with a larger first peak and smaller second peak indicates that two coarse-grained nucleosomes are on average located within a distance from the origin to the first peak and a second coarse-grained nucleosome at the distance indicated by the second peak. Since the peaks are decreasing, this ordering diminishes, much like the local ordering in a liquid. On larger scales larger than 50,000 bp, there is no order, i.e., there is no correlation. (C) shows for two chromosomes how the structure differs within as well as among chromosomes. The parameter start indicates from where in the chromosomes the structure was computed. One can see that the structure varies within a chromosome; nevertheless, common structures are found.

1.1. Data

The technology of micrococcal nuclease (MNase) digestion combined with high-throughput sequencing (MNase-seq) [26] is used to map the distribution of nucleosome occupancy genome-wide. In order to map the MNase-seq data to nucleosome positioning data, several programs were developed, such as NPS [32], nucleR [33], and DANPOS [34]. A nucleosome sequencing profile is generated to depict nucleosome distribution in wave-form where nucleosome peaks are detected. The improved nucleosome-positioning algorithm (iNPS) can be applied to identify peaks and correctly detect nucleosome positions [28]. One possible output of the iNPS algorithm is in the binary format, with 1s representing a nucleosome being present and 0s for the nucleosome-free regions or linker regions.

The genome-wide study of the species is a challenging task due to its large sequence size, which needs theoretical expertise and computational power. For our study, we have chosen *Candida albicans* as a simple completely sequenced organism [35] that is small enough to be computationally viable. Furthermore, *C. albicans* allows for similar mechanisms that are found in eucaryotes. Indeed, epigenetic mechanisms across animals, plants, and fungi include DNA methylation as a common epigenetic signalling mechanism, and it is present in *C. albicans*. A putative histone H1 has been identified [36]. Whereas these are technical decisions, we also wanted to select a species that should have a clinical prevalence. It consists of eight sets of chromosome pairs whose complete genome sequence is available. The raw data of the MNase-seq are available from the Gene Expression Omnibus (GSM1542419) and were measured by Puri et al. [37]. We also accessed the processed iNPS data in the NucMap database by Zhao et al. [38].

1.2. Methods

To obtain a consistent classification of the nucleosomal positioning data in genome-wide classes, we perform the following steps (explained in more detail below the list):

1. Each chromosome is divided into segments of 75 kbp of length.
2. For every chromosome, the positioning data are coarse-grained.
3. The coarse-grained nucleosome positioning data are used to calculate auto-correlation functions over the different sections.
4. A distance matrix is calculated over all the auto-correlation function data.
5. These segments are clustered. Various distance matrix and clustering algorithms are used to generalize the results.

1.2.1. Genome Section Classification

In order to extract the global pattern for areas in a genome, the whole genome is separated into sections with equal length. The section length L is an important scale parameter and needs to be properly set. L should not be too large to avoid all features from different areas bounded together. At the same time, L also should not be too small; otherwise, the global structure is flooded by the subtle differences and becomes a pattern for only a single nucleosome. The single nucleosome wrapping length L_n can be used as a lower bound for the choice of L . However, to obtain a relevant structure, we require that $L \gg L_n$. Considering the nucleosome length L_n is about 147 bp [3,4], L is chosen to be 50 kbp. Additionally, to avoid boundary effects, for each section, a 12.5 kbp intersection on both sides with its neighbor is added. Hence, the total section length L is 75 kbp. This binning is applied to each chromosome. Chr. 2 for example, with a length of 2,231,883 bp, is separated into 44 sections.

1.2.2. Coarse Graining

The idea of coarse graining is an established ansatz and tool in physics to describe complex systems on a scale that allows identifying structure. Typically, the structure appears as a collective phenomenon among smaller entities. The idea is to eliminate degrees of freedom, i.e., find a representation of the system on a larger time or space scale, iteratively moving to larger scales without changing the system. Over the last few years,

coarse graining has emerged as a way to model large complex systems and has successfully been applied to other biomolecules such as proteins [39].

After the whole genome is separated into sections, coarse graining is applied for each section. The method we implemented for coarse graining is the rolling mean method [40]. This method takes a window with a certain size (e.g., $b = 5$ kbp), computes the averaged value of the nucleosome positioning inside the window, and moves the window to the following location. After this value is computed for each location, coarse-grained data on the scale of the window size are returned. Here, Python pandas.DataFrame.rolling [41] is used to obtain the coarse-graining. To exclude the effect of telomeres, discrete ends of the sections and incorporation of the window size and offset was chosen to be at least

$$\text{offset} \geq \text{window size}/2 \quad (1)$$

1.2.3. Auto-Correlation Function Calculation

An auto-correlation function is a well-known approach in physics and pattern recognition, capturing the inner interaction pattern inside the data [40]. Particularly for structures that are liquid-like, the auto-correlation function, or in this context the radial distribution function, identifies typical length scales and patterns.

For each section j , it is applied on all the coarse-grained data ρ_j . The normalized auto-correlation function $C^j(\tau)$ with respect to distance τ for section j is:

$$C^{\alpha,j}(\tau) = \frac{E[(\rho_i^{\alpha,j} - \mu^{\alpha,j})(\rho_{i+\tau}^{\alpha,j} - \mu^{\alpha,j})]}{(\sigma^{\alpha,j})^2} \quad (2)$$

where $\rho_i^{\alpha,j}$ is the data at position i within the section j of chromosome α . $E(\dots)$ is the mean of everything in the parentheses over all indices i . μ^j is the mean of ρ and σ^j is the variance for the section j . Thus, associated with each section j is the function $C^{\alpha,j}(\tau)$ of chromosome α ; hence, at the end, we will have N functions $C^{\alpha,j}(\tau)$ where N is the section number for the particular chromosome.

1.2.4. Distance Matrix Calculation

To classify the functions, a similarity measure is applied, and a resulting distance matrix is computed. The distance matrix is a square matrix containing the pairwise distances between all the elements available in the dataset, measuring the proximity between the correlation functions. Interpreting the functions as high-dimensional vectors, we use the p -norm to define the distance d_p between two functions:

$$d_p(a, b) = \|a - b\|_p = \left(\sum_{i=1}^d |a_i - b_i|^p \right)^{1/p} \quad (3)$$

where a and b are the functions in the form of vectors. For $p = 2$, the p -norm corresponds to the Euclidean distance.

1.2.5. Clustering

To identify the unique nucleosome organisation or distribution function, there is a need to cluster the sections together on the basis of similarity among them. We used a clustering approach, i.e., hierarchical clustering [42]. This is an unsupervised algorithm that groups similar objects into groups called clusters. It uses a distance matrix to identify the two closest clusters first and then merge the two most similar clusters. This iterative process continues until the clusters are merged to get distinct clusters in a hierarchical manner.

Hierarchical clustering builds a hierarchy of clusters using two methods: agglomerative and divisive algorithms. We used the former, i.e., the Ward method [43], where each observation starts in its own cluster and pairs of clusters are merged, moving up the hierarchy.

1.2.6. Statistical Distributions Fitting

Fitting of the distributions was performed using the `scipy stats` package [44] under Python.

2. Results

The first indication of non-trivial ordering is given by the distribution of the nucleosome positioning data. The binary nucleosome positioning data for all chromosomes of *Candida albicans* (NucMap database [38]) are subjected to the described coarse graining and then analyzed (see the histogram of densities in the Supplementary Information Figures S1–S3, and Tables S1 and S2). The genome-wide normalised nucleosome density shows a non-Gaussian behaviour with a slight negative skew. Overall, a log-logistic distribution gives the best consistent fit for all chromosomes compared to a normal distribution on the same bin size and rolling average for all chromosomes.

Recall that each chromosome is divided into chunks of 75 kbp with 25 kbp overlapping on each side. The auto-correlation of each chunk is obtained on the coarse-grained nucleosome positioning data. The respective correlation function of each section for all chromosomes are shown in Figure 2 and in detail in Supplementary Information (Figures S9–S16). Shown are the correlation functions on the coarse-grained scale as well as a further smoothing to make the features that are common among a class more apparent (see below). The colour bar indicates the class. Even though there are variations within a class, certain common features are seen. These features are the first and second peak structure, the height of the peaks, and how long a structure persists. Recall that the zero line indicates that there is no correlation; i.e., there, the structure is that of a gas or an unordered behaviour. The first peak indicates an increased probability to find a coarse-grained nucleosome at the distance of the peak position, and the same applies to the second and additional peaks. If these peaks are of similar height, then there is a stronger long-range ordering. A particular example showing similar heights up to a third peak is in section 12 of chromosome no. 3 (see Supplementary Information Figure S11), while section 6 shows a drop in the peak heights. Nevertheless, due to the overall similarity, these fall into the same class.

With diminishing height, the likelihood of the ordering and the strictness of ordering vanishes. Notice that for some of the sections (within one class), many sub-peaks or side-peaks exist, indicating possible sub-orderings. An example on the more extreme side is chromosome 3 and sections such as 3, 5, 16, etc. Overall, the short-range order is much less pronounced. The orange smoothed line indicates that in this class, the salient feature is a smoothly decreasing function indicating a different kind of order than for the class with sections 0, 8 and 12, etc.

Even looking at the correlation functions without the indicated class mapping shows that there are universal features beyond fluctuations. Within a class, a more or less pronounced ordering feature is visible. Comparing the different correlation data between the chromosomes, these become apparent.

These observations can be proven more rigorously by applying similarity measures between the correlation functions. Figure 2 shows the resulting distance matrix between all chromosomes and all sections (the individual results are shown in the Supplementary Information Figures S4–S6). Shown is the distance matrix after reordering on the basis of similarity between sections. The colour indicates the similarity between the correlation functions. Notice the patterns that emerge from the sorting of the data into classes.

These classes, represented by different colours, are shown in the dendrogram. These classes were obtained by hierarchical clustering. In the lower part of the figure on the left are the typical correlation functions representing the corresponding class with its colour code. The orange-coloured class shows a fairly regular pattern and closely spaced ordering on a short scale, such as tightly packed heterochromatin, whereas the light blue class has lost the regularity and shows a less stringent regular but still pronounced pattern on a slightly larger scale. The blue-coloured class shows a rather very irregular pattern compared to the other two classes and corresponds more to euchromatin.

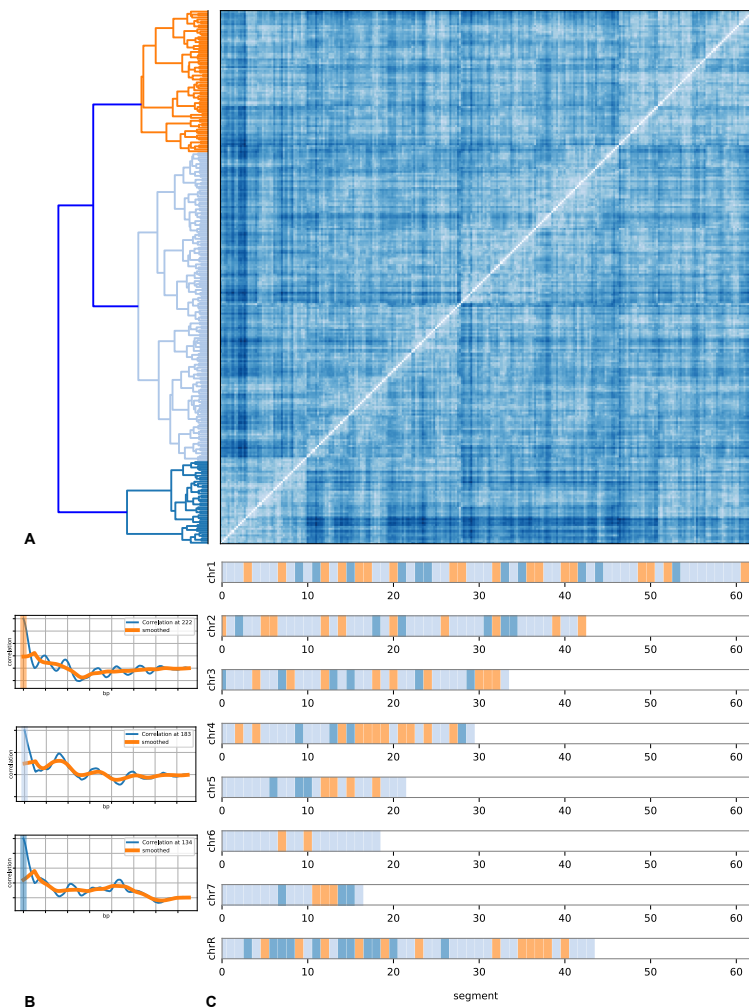


Figure 2. (A) shows the genome-wide distance matrix between the correlation functions between segments of size 75 kbp. Hierarchical clustering was applied to identify common patterns. The matrix was sorted according to the patterns. The left side shows the clustering. (B) shows the coarse-grained nucleosomal density correlation functions of *Candida albicans* at 5 kb coarse graining. (C) shows the genome-wide distribution of segments with colours corresponding to the classification. White space is due to not all chromosomes having the same length. The pattern classification was done genome-wide to yield three main patterns. These three patterns were assigned colours, and the segments of each chromosome corresponding to one of the three patterns are marked. The orange-coloured pattern is characterised by a closely and fairly regularly spaced ordering similar to the tightly packed heterochromatin. The dark and light-coloured blue patterns have lost the regularity and the longer range of the order and thus correspond more to euchromatin. However, note that both these two classes have a huge variety of subclasses. This is not surprising in the sense that one would expect a larger variety of not so ordered patterns in one dimension than for ordered patterns in one dimension.

These observations are consistent with the typical classification from microscopy data into hetero- and euchromatin. The data show that the orange and light blue classes can be mapped on heterochromatin. Thus, the blue-coloured class is euchromatin. The data also show that still, within any of these classes, the features have many sub-features that we salvaged for the larger patterns to allow a “coarse-grained” view on the ordering of the nucleosomes. These sub-features compose elaborated chromatin states such as solenoid [45], zig-zag ribbon [46], or other structures [47], which demand a cross-correlation analysis with CTCF binding sites [48], CpG island position [49], and other data.

Notice that this partitioning into classes is genome-wide. A consistent classification can be established. This is shown in the mapping of the positions of the section to the chromosomes. Notice that, as expected, not a random mixture of the three colours emerges but rather a clear pattern. The larger chromosomes appear to have more internal structuring compared to the smaller chromosomes that are more homogeneous in their internal structure. The partitioning into a clear pattern, genome-wide is not limited to species *Candida albicans*, but the pipeline is generalised and can be used for any species in which the whole genome has been sequenced.

3. Discussion

The structural organisation of the genome depends on the patterns of nucleosome positioning and their distribution in the genome. At a higher scale, the nucleosome positioning distribution varies across the chromosomes, which appear to be conserved along the entire genome. The classification of the chromosomes into segments of the distinct nucleosomal distribution shown here is in line with earlier studies. Although two major classifications of the chromosomal region as heterochromatin and euchromatin are suggested, we find that their organisations can be further subdivided. Nucleosomes can be well-positioned to form phased and unphased arrays consisting of regularly spaced nucleosomes or can be fuzzy to form irregular arrays of nucleosomes. The three distinct nucleosome distribution patterns along the genome obtained in our result are in agreement with this study. Moreover, further classification of nucleosomal distribution is obtained along each chromosome. Around five to seven different nucleosome distribution patterns are observed for all chromosomes. However, for the entire genome, three patterns are found to be conserved.

We have analysed the effect for different $p = 2, 7$ in the p -norm on the outcome of the clustering of similar correlation functions, and the outcome comes to be similar for all p . For high p values, some of the clusters split into further clusters. In addition, the cosine similarity norm was tested for further verification, yielding similar clustering (see Supplementary Information Video S1). This rules out that the clustering is an artifact of the model and its architecture.

Around five patterns of chromosomal organisation are obtained for each chromosome by analysing the nucleosome positioning data distribution. These patterns obtained are generally coincident with gene densities and lead to the distinct spatial organisation of genomic DNA. The genome's hierarchical structure–function relationship [12] is governed by chromatin domains and their higher-order folding. The formation of chromatin boundaries and associated TADs are controlled by the nucleosome distribution patterns. Recent studies by Wiese et al. [16] suggested that domain formation and genome organisation can be predicted with nucleosome positioning only. Pulivarty et al. [50] primarily focused on nucleosome studies, which are limited to a very local individual promoter and enhancer but can be a more general mechanism by which cells can regulate the accessibility of the genome during development at different scales. After an extensive analysis of nucleosome positioning data, the way of organisation of nucleosomal distribution patterns is found to be different at different scales and for different chromosomes. The distinct patterns obtained from our calculation correspond to different ways of nucleosome positioning and may control domain formation and genome organisation in the cell. However, the three distinct patterns of nucleosome organisation that appeared to be conserved in the genome show the global consistency of distribution patterns inside the genome. The consistency in different kinds of distinct patterns observed in the genome corresponds to identical gene densities and similar expression regions for specific locations inside the cell.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/life12040541/s1>, Figures S1–S27: Nucleosome Correlation Data; Tables S1 and S2: Nucleosome Correlation Data; Video S1: Coarse Graining.

Author Contributions: All authors were involved in the conception, processing of the data, analysis and drafting of the manuscript. D.W.H. and N.N.O. supervised the work. All authors have read and agreed to the published version of the manuscript.

Funding: This work is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC 2181/1-390900948 (the Heidelberg STRUC-TURES Excellence Cluster).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors gratefully acknowledge the data storage service SDS@hd supported by the Ministry of Science, Research and the Arts Baden-Württemberg (MWK) and the German Research Foundation (DFG) through grant INST 35/1314-1 FUGG and INST 35/1503-1 FUGG. Kunhe Li would like to acknowledge funding by the Chinese Scholarship Council (CSC). Sujeet Kumar Mishra would like to acknowledge funding by the India government Ministry of Science and Technology, Department of Biotechnology (DBT)-Interdisciplinary Research Center for Scientific Computing (IWR) PhD program.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Oluwadare, O.; Highsmith, M.; Cheng, J. An Overview of Methods for Reconstructing 3-D Chromosome and Genome Structures from Hi-C Data. *Biol. Proced. Online* **2019**, *21*, 7. [[CrossRef](#)]
- Jerkovic, I.; Cavalli, G. Understanding 3D genome organization by multidisciplinary methods. *Nat. Rev. Mol. Cell Biol.* **2021**, *22*, 511–528. [[CrossRef](#)]
- Routh, A.; Sandin, S.; Rhodes, D. Nucleosome repeat length and linker histone stoichiometry determine chromatin fiber structure. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 8872–8877. [[CrossRef](#)]
- Bohr, J.; Olsen, K. The size of the nucleosome. *arXiv* **2012**, arXiv:1102.0761.
- Staneva, D.; Georgieva, M.; Miloshev, G. *Kluyveromyces lactis* genome harbours a functional linker histone encoding gene. *FEMS Yeast Res.* **2016**, *16*, fow034. [[CrossRef](#)] [[PubMed](#)]
- Bohn, M.; Diesinger, P.; Kaufmann, R.; Weiland, Y.; Müller, P.; Gunkel, M.; Ketteler, A.; Lemmer, P.; Hausmann, M.; Heermann, D.; et al. Localization Microscopy Reveals Expression-Dependent Parameters of Chromatin Nanostructure. *Biophys. J.* **2010**, *99*, 1358–1367. [[CrossRef](#)]
- Tchasovnikarova, I.A.; Kingston, R.E. Beyond the Histone Code: A Physical Map of Chromatin States. *Mol. Cell* **2018**, *69*, 5–7. [[CrossRef](#)] [[PubMed](#)]
- Struhl, K.; Segal, E. Determinants of nucleosome positioning. *Nat. Struct. Mol. Biol.* **2013**, *20*, 267–273. [[CrossRef](#)] [[PubMed](#)]
- Hilbert, L.; Sato, Y.; Kuznetsova, K.; Bianucci, T.; Kimura, H.; Jülicher, F.; Honigmann, A.; Ziburdaev, V.; Vastenhouw, N.L. Transcription organizes euchromatin via microphase separation. *Nat. Commun.* **2021**, *12*, 1360. [[CrossRef](#)] [[PubMed](#)]
- Dekker, J.; Rippe, K.; Dekker, M.; Kleckner, N. Capturing Chromosome Conformation. *Science* **2002**, *295*, 1306–1311. [[CrossRef](#)] [[PubMed](#)]
- van Berkum, N.L.; Lieberman-Aiden, E.; Williams, L.; Imakaev, M.; Gnirke, A.; Mirny, L.A.; Dekker, J.; Lander, E.S. Hi-C: A method to study the three-dimensional architecture of genomes. *J. Vis. Exp.* **2010**, *6*, 1869. [[CrossRef](#)]
- Beagan, J.A.; Phillips-Cremins, J.E. On the existence and functionality of topologically associating domains. *Nat. Genet.* **2020**, *52*, 8–16. [[CrossRef](#)] [[PubMed](#)]
- Ghavi-Helm, Y.; Jankowski, A.; Meiers, S.; Viales, R.R.; Korb, J.O.; Furlong, E.E.M. Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. *Nat. Genet.* **2019**, *51*, 1272–1282. [[CrossRef](#)] [[PubMed](#)]
- Nora, E.P.; Lajoie, B.R.; Schulz, E.G.; Giorgetti, L.; Okamoto, I.; Servant, N.; Piolot, T.; van Berkum, N.L.; Meisig, J.; Sedat, J.; et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **2012**, *485*, 381–385. [[CrossRef](#)]
- Quentin, S.; Frédéric, B.; Giacomo, C. Principles of genome folding into topologically associating domains. *Sci. Adv.* **2022**, *5*, eaaw1668. [[CrossRef](#)]
- Wiese, O.; Marenduzzo, D.; Brackley, C.A. Nucleosome positions alone can be used to predict domains in yeast chromosomes. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 17307. [[CrossRef](#)] [[PubMed](#)]
- Kharerin, H.; Bai, L. Thermodynamic modeling of genome-wide nucleosome depleted regions in yeast. *PLoS Comput. Biol.* **2021**, *17*, e1008560. [[CrossRef](#)] [[PubMed](#)]
- Drew, H.R.; Travers, A.A. DNA bending and its relation to nucleosome positioning. *J. Mol. Biol.* **1985**, *186*, 773–790. [[CrossRef](#)]
- Chung, H.R.; Vingron, M. Sequence-dependent Nucleosome Positioning. *J. Mol. Biol.* **2009**, *386*, 1411–1422. [[CrossRef](#)]
- Parmar, J.J.; Marko, J.F.; Padinhateeri, R. Nucleosome positioning and kinetics near transcription-start-site barriers are controlled by interplay between active remodeling and DNA sequence. *Nucleic Acids Res.* **2014**, *42*, 128–136. [[CrossRef](#)] [[PubMed](#)]

21. Angermayr, M.; Oechsner, U.; Bandlow, W. Reb1p-dependent DNA bending effects nucleosome positioning and constitutive transcription at the yeast profilin promoter. *J. Biol. Chem.* **2003**, *278*, 17918–17926. [[CrossRef](#)]
22. Rippe, K.; Schrader, A.; Riede, P.; Strohner, R.; Lehmann, E.; Längst, G. DNA sequence- and conformation-directed positioning of nucleosomes by chromatin-remodeling complexes. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 15635–15640. [[CrossRef](#)] [[PubMed](#)]
23. Wippo, C.J.; Israel, L.; Watanabe, S.; Hochheimer, A.; Peterson, C.L.; Korber, P. The RSC chromatin remodelling enzyme has a unique role in directing the accurate positioning of nucleosomes. *EMBO J.* **2011**, *30*, 1277–1288. [[CrossRef](#)]
24. Shim, Y.S.; Choi, Y.; Kang, K.; Cho, K.; Oh, S.; Lee, J.; Grewal, S.I.; Lee, D. Hrp3 controls nucleosome positioning to suppress non-coding transcription in eu- and heterochromatin. *EMBO J.* **2012**, *31*, 4375–4387. [[CrossRef](#)] [[PubMed](#)]
25. Helbo, A.S.; Lay, F.D.; Jones, P.A.; Liang, G.; Grønbaek, K. Nucleosome Positioning and NDR Structure at RNA Polymerase III Promoters. *Sci. Rep.* **2017**, *7*, 41947. [[CrossRef](#)]
26. Klein, D.C.; Hainer, S.J. Genomic methods in profiling DNA accessibility and factor localization. *Chromosome Res.* **2020**, *28*, 69–85. [[CrossRef](#)] [[PubMed](#)]
27. Mieczkowski, J.; Cook, A.; Bowman, S.K.; Mueller, B.; Alver, B.H.; Kundu, S.; Deaton, A.M.; Urban, J.A.; Larschan, E.; Park, P.J.; et al. MNase titration reveals differences between nucleosome occupancy and chromatin accessibility. *Nat. Commun.* **2016**, *7*, 11485. [[CrossRef](#)]
28. Chen, W.; Liu, Y.; Zhu, S.; Green, C.D.; Wei, G.; Han, J.D.J. Improved nucleosome-positioning algorithm iNPS for accurate nucleosome positioning from sequencing data. *Nat. Commun.* **2014**, *5*, 4909. [[CrossRef](#)]
29. Georgakilas, G.K.; Perdikopanis, N.; Hatzigeorgiou, A. Solving the transcription start site identification problem with ADAPT-CAGE: A Machine Learning algorithm for the analysis of CAGE data. *Sci. Rep.* **2020**, *10*, 877. [[CrossRef](#)] [[PubMed](#)]
30. Oiwa, N.N.; Cordeiro, C.E.; Heermann, D.W. The Electronic Behavior of Zinc-Finger Protein Binding Sites in the Context of the DNA Extended Ladder Model. *Front. Phys.* **2016**, *4*, 13. [[CrossRef](#)]
31. Singh, A.K.; Mueller-Planitz, F. Nucleosome positioning and spacing: From mechanism to function. *J. Mol. Biol.* **2021**, *433*, 166847. [[CrossRef](#)] [[PubMed](#)]
32. Schöpflin, R.; Teif, V.B.; Müller, O.; Weinberg, C.; Rippe, K.; Wedemann, G. Modeling nucleosome position distributions from experimental nucleosome positioning maps. *Bioinformatics* **2013**, *29*, 2380–2386. [[CrossRef](#)]
33. Flores, O.; Orozco, M. nucleR: A package for non-parametric nucleosome positioning. *Bioinformatics* **2011**, *27*, 2149–2150. [[CrossRef](#)] [[PubMed](#)]
34. Chen, K.; Xi, Y.; Pan, X.; Li, Z.; Kaestner, K.; Tyler, J.; Dent, S.; He, X.; Li, W. DANPOS: Dynamic analysis of nucleosome position and occupancy by sequencing. *Genome Res.* **2013**, *23*, 341–351. [[CrossRef](#)]
35. Price, R.J.; Weindling, E.; Berman, J.; Buscaino, A. Chromatin Profiling of the Repetitive and Nonrepetitive Genomes of the Human Fungal Pathogen *Candida albicans*. *mBio* **2019**, *10*, e01376-19. [[CrossRef](#)] [[PubMed](#)]
36. Skrzypek, M.S.; Binkley, J.; Binkley, G.; Miyasato, S.R.; Simison, M.; Sherlock, G. The *Candida* Genome Database (CGD): Incorporation of Assembly 22, systematic identifiers and visualization of high throughput sequencing data. *Nucleic Acids Res.* **2017**, *45*, gkw924. [[CrossRef](#)]
37. Puri, S.; Lai, W.K.M.; Rizzo, J.M.; Buck, M.J.; Edgerton, M. Iron-responsive chromatin remodelling and MAPK signalling enhance adhesion in *Candida albicans*. *Mol. Microbiol.* **2014**, *93*, 291–305. [[CrossRef](#)] [[PubMed](#)]
38. Zhao, Y.; Wang, J.; Liang, F.; Liu, Y.; Wang, Q.; Zhang, H.; Jiang, M.; Zhang, Z.; Zhao, W.; Bao, Y.; et al. NucMap: A database of genome-wide nucleosome positioning map across species. *Nucleic Acids Res.* **2019**, *47*, D163–D169. [[CrossRef](#)] [[PubMed](#)]
39. Kmiecik, S.; Gront, D.; Kolinski, M.; Wieteska, L.; Dawid, A.E.; Kolinski, A. Coarse-Grained Protein Models and Their Applications. *Chem. Rev.* **2016**, *116*, 7898–7936. [[CrossRef](#)]
40. Reichl, L.E. *A Modern Course in Statistical Physics*, 4th ed.; Wiley: Hoboken, NJ, USA, 2016.
41. Pandas Development Team T. Pandas-Dev/Pandas: Pandas. 2020. Available online: <https://doi.org/10.5281/zenodo.3509134> (accessed on 12 February 2022).
42. Maimon, O.; Rokach, L. *Data Mining and Knowledge Discovery Handbook*; Springer: Berlin/Heidelberg, Germany, 2005.
43. Everitt, B.S.; Landau, S.; Leese, M. *Cluster Analysis*, 4th ed.; Oxford University Press: Oxford, UK, 2001.
44. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17*, 261–272. [[CrossRef](#)]
45. Finch, J.T.; Klug, A. Solenoidal model for superstructure in chromatin. *Proc. Natl. Acad. Sci. USA* **1976**, *73*, 1897–1901. [[CrossRef](#)] [[PubMed](#)]
46. Diesinger, P.M.; Kunkel, S.; Langowski, J.; Heermann, D.W. Histone depletion facilitates chromatin loops on the kilobasepair scale. *Biophys. J.* **2010**, *99*, 2995–3001. [[CrossRef](#)] [[PubMed](#)]
47. Williams, S.P.; Athey, B.D.; Muglia, L.J.; Schappe, R.S.; Gough, A.H.; Langmore, J.P. Chromatin fibers are left-handed double helices with diameter and mass per unit length that depend on linker length. *Biophys. J.* **1986**, *49*, 233–248. [[CrossRef](#)]
48. Norio Oiwa, N.; Li, K.; Cordeiro, C.E.; Heermann, D.W. Prediction and Comparative Analysis of CTCF Binding Sites based on a First Principle Approach. *arXiv* **2021**, arXiv:2110.10508.

49. Gardiner-Garden, M.; Frommer, M. CpG Islands in vertebrate genomes. *J. Mol. Biol.* **1987**, *196*, 261–282. [[CrossRef](#)]
50. Pulivarthy, S.R.; Lion, M.; Kuzu, G.; Matthews, A.G.W.; Borowsky, M.L.; Morris, J.; Kingston, R.E.; Dennis, J.H.; Tolstorukov, M.Y.; Oettinger, M.A. Regulated large-scale nucleosome density patterns and precise nucleosome positioning correlate with V(D)J recombination. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 201605543. [[CrossRef](#)] [[PubMed](#)]