ChromoSkein: Untangling Three-Dimensional Chromatin Fiber With a Web-Based Visualization Framework

Matúš Talčík, Filip Opálený, Tereza Clarence, Katarína Furmanová, Jan Byška, Barbora Kozlíková, and David Kouřil

Abstract—We present ChromoSkein, a web-based tool for visualizing three-dimensional chromatin models. The spatial organization of chromatin is essential to its function. Experimental methods, namely Hi-C, reveal the spatial conformation but output a 2D matrix representation. Biologists leverage simulation to bring this information back to 3D, assembling a 3D chromatin shape prediction using the 2D matrices as constraints. Our overview of existing chromatin visualization software shows that the available tools limit the utility of 3D through ineffective shading and a lack of advanced interactions. We designed ChromoSkein to encourage analytical work directly with the 3D representation. Our tool features a 3D view that supports understanding the shape of the highly tangled chromatin fiber and the spatial relationships of its parts. Users can explore and filter the 3D model using two interactions. First, they can manage occlusion both by toggling the visibility of semantic parts and by adding cutting planes. Second, they can segment the model through the creation of custom selections. To complement the 3D view, we link the spatial representation with non-spatial genomic data, such as 2D Hi-C maps and 1D genomic signals. We demonstrate the utility of ChromoSkein in two exemplary use cases that examine functional genomic loci in the spatial context of chromosomes and the whole genome.

Index Terms—Biological visualization, chromatin, 3D, genomic data, interaction.

1 INTRODUCTION

Organization of a genome in three-dimensional (3D) space significantly impacts its function. In eukaryotic cells, the DNA is packed into a micron-sized nucleus with the help of proteins, which together build a so-called chromatin fiber. Chromatin represents a unique combination of two data characteristics. First, the underlying DNA molecule is a linear nucleotide sequence. Second, the fiber folds into a densely packed 3D shape.

The specific spatial configuration of chromatin in a nucleus is not yet fully determined. Thus, it is an object of intense study in biology. Over the years, several methods for revealing the organization of chromatin have been developed. Most prominent is the chromosome conformation capture (3C) suite of experiments [15]. These methods cross-link together segments of the DNA that are in physical proximity. Aggregating the resulting interactions across many cells gives a measure of how likely two genomic parts are mutually interacting. The Hi-C method [22], a variant of 3C, resolves all-to-all interactions between same-sized DNA segments. The output of this method is usually in the form of a contact matrix, see Figure 1. The matrix values can then be used to assemble a prediction of chromatin’s conformation in 3D.

While a number of tools feature a 3D view, the 3D aspect is underutilized in genomic visualization. We hypothesize that the lack of tools supporting 3D interaction is caused mainly by two reasons. First, the Hi-C experiment outputs 2D contact maps. Therefore, most tools naturally work with the 2D representation. Second, 3D visualization inherently presents issues like occlusion that make it harder to work with 3D data on a two-dimensional display. Nusrat et al. [28] further remark that visualizing 3D spatial chromatin data suffers from the fact that it only shows a prediction and omits information about the resolved structure’s ambiguity.

Despite these claims, certain theories can only be examined in the 3D context. For example, positioning of a gene in a chromosome and its spatial distance to functional landmarks, such as promoters and enhancers, has a direct impact on gene regulation [10]. Certain data characteristics—e.g., density or mutual spatial orientation—are more intuitively obtained from the 3D representation. The examined structure, i.e., chromatin fiber, is by definition a spatial object and, according to recent biological research, it makes sense to examine it in its natural context [4]. We found out that many tools use 3D chromatin models merely for illustration while interaction with the 3D data is severely limited.

To address this gap in available chromatin visualization tools, we set out to design and develop a novel tool that places the 3D representation at the forefront. We focus primarily on two facets: a) performant and high-fidelity visual representations that highlight the spatiality of the underlying data; and b) interaction allowing direct exploration of the 3D chromatin model. We deliver our visualization solution for the web environment to promote easier adoption and collaboration.
In summary, we present the following key contributions:

- An extensive overview of existing genomic tools featuring 3D visualization, reviewing available visual representations, interactions, and analytical features.
- Design and prototypical implementation of a 3D-centric visualization tool for chromatin data, called ChromoSkein1.
- Discussion and demonstration of ways of linking the 3D representation with conventional genomic 1D and 2D visualizations.
- Two exemplary use case scenarios demonstrating the utility of 3D representation in domain-specific analytical tasks.

2 BACKGROUND & MOTIVATION

Our work is motivated by a specific biological application domain. We, therefore, start with a brief introduction to chromatin conformation research and outline the high-level motivation for developing a novel tool.

2.1 Background

While methods for resolving the linear DNA sequence are well established, the acquisition of DNA’s three-dimensional organization is still challenging. It is known that the double helix is further organized in nucleosomes, which in turn form a chromatin fiber packaged into chromosomes in the cell nucleus.

The organization of chromatin in a cell is the subject of several experimental methods, referred to as chromosome conformation capture (3C) techniques [15]. These methods can indicate the spatial organization of chromatin. In general, 3C techniques are based on measurements of the frequency of interactions between fragments of DNA. Earlier methods could yield results only for a few of these fragments. However, high-throughput variations of 3C methods, such as Hi-C [2], [22], enable biologists to infer the number of interactions between all equally-sized regions of the whole chromatin fiber. The trade-off is a very limited precision. Depending on the experiment’s parameters, the size of these regions is in the order of thousands (kilobases, kb) to millions of nucleotides (megabases, Mb). One such group of nucleotides is then referred to as a bin.

The output of a Hi-C assay is a 2D contact frequencies matrix, sometimes also called interaction frequencies matrix, that assigns each pair of bins a number proportional to the number of interactions observed between these two bins. Further analysis of Hi-C matrix data reveals multi-scale organization of genomes. The organization manifests in the Hi-C map as specific patterns, e.g., a chessboard-like pattern signifying A/B compartments [22], or point peaks in distant sequential regions indicating topologically associating domains (TADs), important in gene expression [7].

The contact frequency maps coming from the Hi-C experiment imply a genome’s spatial configuration. Computational biologists thus infer the 3D structure using probabilistic algorithms, polymer simulations, or various other methods [29]. The Hi-C matrix values serve as constraints for the computation, as illustrated in Figure 1. These methods generate a three-dimensional model prediction, where every bin is assigned a position in 3D space. Examples of tools for genome structure prediction are LorDG [41] or 3DMax [30].

2.2 Biological Motivation

The three-dimensional chromatin structure can be used to analyze and confirm hypotheses. For example, Stevens et al. [37], who first reconstructed whole-genome models using single-cell Hi-C from haploid mouse cells, examined the 3D model to confirm known spatial features. They found that chromosomes occupy specific areas—territories—in the nucleus. They did so by isolating individual chromosomes and de-emphasizing the rest to indicate the chromosomes’ nuclear position.

Similarly, Tan et al. [37] reconstructed whole genome chromatin models but from diploid human cells. They also look at chromosome territories. The shape and mutual arrangement of chromosomes are evident from the 3D model, such as how the chromosomes intermingle. Tan et al. also show how gene-rich chromosomes occupy space closer to the nuclear center while gene-poor chromosomes lie on its periphery. With the 20kb resolution, the 3D model exhibits a hierarchical, fractal organization of chromatin, with regions of high spatial clustering and segregation. Thanks to Tan et al.’s novel method for capturing conformation of diploid cells, the resulting structures showcase differences in the shape of active and inactive X chromosomes. While this can be quantified by, for example, a radius of gyration, this feature can be intuitively identified during the visual exploration of the 3D shape.

The development of ChromoSkein is motivated by these recent investigations of chromatin conformation. Spatial representations clearly play an important role in several stages of the analytical process. We see an opportunity for intuitive visualization and interaction methods to expedite the analysis of 3D chromatin data.

2.3 Why Focus on 3D?

Bioinformaticians trained in Hi-C analysis can imply insights from the contact matrix representation. Neverthe-
less, they are still limited by their cognitive abilities. Hi-C maps are an abstract representation that captures many relationships between genomic regions. 3D structure prediction allows generating a spatial model that satisfies all the constraints and gives one possible conformation in space. This relieves scientists of having to reconstruct a mental image of the spatial conformation.

Additionally, certain characteristics of a spatial structure cannot be inferred purely from the 2D representation. For example, biologists might be interested in density of functional landmarks (e.g., genes) in a certain nucleus compartment. This metric can only be determined from the modeled 3D structures, either visually or by radius of gyration. Similarly, mutual orientation of elements, e.g., chromosomes, is a relationship indirectly captured in the Hi-C map but observable only once the structure is visualized in 3D.

Finally, Hi-C maps are often quite extensive, occupying gigabytes in storage. At a common 15kb resolution, where one bin corresponds to a sequence of 15 thousand basepairs, a human genome results in approx. 207 thousand bins. Storing the information about observed interactions for all possible pairs of bins thus results in \(207 \times 10^6\) numbers. As the Hi-C map is symmetric, some viewers only show half of it, i.e., triangular Hi-C map (see Figure 2B), resulting in \(N^2/2\) numbers to store. However, the asymptotic complexity \(O(n^2)\) remains the same. The 3D representation, on the other hand, requires only a list of bin positions, reducing the space complexity from \(O(n^2)\) to \(O(n)\). The 3D model thus serves as a more efficient form for storage and transfer.

Naturally, the 3D representation has its disadvantages. As Nusrat et al. [28] note, 3D chromatin models represent only a prediction that satisfies the constraints defined by the Hi-C map. This uncertainty is, however, understood and accounted for by biologists investigating 3D chromatin. 3D visualization also inherently presents issues related to showing 3D phenomena on a 2D display, such as occlusion. Occlusion management, however, has been a topic of research both in general computer graphics and in visualization specifically and we now have at our disposal a number of techniques to deal with this issue. Overall, we do not argue that a 3D representation should completely replace the 2D matrix representation. We do, however, believe that it can serve as a valuable addition and that existing tools underutilize this representation.

### 3 RELATED WORK

While building on general molecular visualization, tools visualizing genomic data have developed in their own strand of research. Here we review visualization’s role in the investigation of genomes’ spatial organization and survey existing visualization tools. Table 1 summarizes the reviewed tools and highlights available and missing functionality.

Marti-Renom and Mirny [25] discuss the challenges and approaches in resolving the biological structure spanning several magnitudes of scales with available experimental and imaging technology. Goodstadt and Marti-Renom [12] further survey means of visualizing existing genomic data types and list available software tools. Waldispühl et al. [44] provide additional insight into the technical challenges related to visualizing 3D genomes.

Initially, chromatin structure was examined purely using Hi-C matrices, i.e., in 2D representation. Juicebox [9] first enabled interactive zooming in the large-scale space of contact frequencies. Its authors leverage a tiling approach inspired by web map services, such as Google Maps. HiGlass [19] further expanded on the idea of zoomable Hi-C maps by implementing additional capabilities, e.g., several linked views and juxtaposing genomic (mostly 1D) data to the Hi-C map. Other visualization types complementing the Hi-C matrices are also common, such as the arc-plot variations used by 3DIV [46].

Most publications on visualizing genomes come from bioinformatics research and therefore largely focus on the applications. Nusrat et al. [28] provide a comprehensive survey of the topic from the visualization research perspective. Recently, L’Yi et al. [23] introduced Gosling, a framework for genomic data visualization that mostly focuses on 2D plots. They leverage a grammar-based approach popularized in the visualization community by Vega-Lite [33].

Working with three-dimensional genomic data presents several additional challenges. Goodstadt and Marti-Renom [13] debate the challenges of visualizing 3D genomes directly. Many domain experts employ existing molecular graphics tools, e.g., PyMOL [34] or ChimeraX [31], to perform the analysis of 3D chromatin. These tools often have a decades-long history and over time have developed into colossus suites tailored mostly for the analysis of smaller molecules, such as proteins. Their applicability to 3D genomic data is limited due to a high learning curve and sometimes technical limitations on larger datasets. Furthermore, including other views (e.g., Hi-C matrix) requires external software or extensive scripting.

One of the first tools tailored to the exploration of the genome in its spatial context is Genome3D [1]. The tool allows switching between three levels of scale that correspond to the inherent hierarchy of the genome. GMOL [27] expanded on Genome3D and increased the number of explorable levels to six. Interaction is mostly done through command-line input. Both of these tools were created as desktop programs, limiting wider adoption in research.

Software published as a web application, on the other hand, is instantly available. In genomic research, tools like the UCSC Genome Browser [18] benefited from the decision to target the web platform. Consequently, many tools for exploring 3D genomic data were developed for the web to lower the adoption threshold. 3DGB [3] devises a solution for storage, querying, and mesh-based rendering of 3D genomic data. Li et al. [21] use Disease to analyze spatial chromatin rearrangement in cancer and developmental diseases. They employ a plotting library to visualize small genomic sub-parts: TADs. TADs can be more closely examined in TADkit [11]. TADkit is a WebGL-based viewer for the analysis of TADs utilizing TADbit [35] library developed by the same team. CSynth [40] combines 3D structure modeling from contact frequencies with an interactive visualization of the result, allowing human-in-the-loop workflow. Delta [38] features a view able to show 3D conformation of small part of the whole genome.

Many of the mentioned tools begin with a form requiring to specify parts of the dataset, e.g., genomic coordinates range, to visualize and therefore inhibit holistic analysis of
Comparison of 3D visualization genomic tools for three-dimensional datasets.

| Tool                  | Representations | Annotations | Selections | View linking | Occlusion management | Depth cueing shader | Web-based |
|-----------------------|-----------------|-------------|------------|--------------|----------------------|---------------------|-----------|
| Genome3D [1]          | Tube, Spheres, Glyphs | Coloring | Selection Gizmo | —            | —                    | Yes                  | No        |
| 3DGB [3]              | Tube            | Coloring   | Cube, Linear | 1D ⇔ 3D     | —                    | No                   | Yes*      |
| GMOL [27]             | Ball and Stick  | Coloring, Text labels, Interaction lines | —         | —            | —                    | No                   | No        |
| 3Disease [21]         | Tube, 2D Heatmap | Coloring   | —          | ?            | ?                    | No                   | Yes*      |
| TADkit [11]           | Tube, 2D Heatmap | Coloring   | —          | 1D, 2D ⇔ 3D | —                    | No                   | Yes       |
| HIC-3DViewer [8]      | Tube            | Coloring, Interaction lines | —         | 1D, 2D ⇔ 3D | —                    | No                   | Yes*      |
| Delta [38]            | Tube, Spheres   | Coloring, Labels, Interaction lines | Point, Linear, Loop | 1D ⇔ 3D | —                    | No                   | Yes*      |
| GenomeFlow [42]       | Tube            | Coloring, Labels, Line thickness | Drag rectangle | —       | Hide regions         | No                   | No        |
| CSynth [40]           | Tube            | Coloring, Labels | Point, Gene | 2D ⇔ 3D | —                    | No                   | Yes       |
| Nucleome Browser [47] | Line, Tube, Sphere, Billboard crosses | Coloring | Point, Chromosome | 1D, 2D, 3D ⇔ 3D | Hide regions | No                   | Yes*      |
| SpaceWalk [16]        | Line, Tube, Sphere, Point cloud | Coloring, Labels, Glyphs | Point | 1D ⇔ 3D | —                    | No                   | Yes       |
| WashU Epigenome Browser [20] | Line, Tube, Sphere, Billboard crosses | Coloring, Labels, Glyps, Arrows | Point | 1D, 2D ⇔ 3D | Hide uniparental chromosomes, Line opacity | No | Yes* |
| Chromoskein           | Line, Tube, Sphere | Coloring, Glyphs | Point, Linear, Sphere | 2D, 3D ⇔ 3D | Cutting planes, Hide regions | Yes | Yes |

* Needs a server for computations or data management.
† Restricted to two views.
? Unknown due to unavailability of the tool and/or missing information in the accompanying publication.

Comparison of 3D visualization genomic tools for three-dimensional datasets.

both local parts as well as global context. Furthermore, while the above-listed tools feature 3D visualization in some form, this aspect is often underutilized and used only for illustration, while the actual analysis is done in other views, e.g., 2D Hi-C contact maps, or 1D feature tracks. We identified four tools where 3D views play a larger role within the analytical workflow and allow performing tasks directly in the 3D representation.

GenomeFlow [42], a continuation of GMOL, offers functionality for modeling and analysis of 3D genomic data. Users can examine predicted 3D structures and augment them by loading additional data, e.g., a list of chromatin loops or gene annotations, which are then overlaid over the 3D model. Interaction with the 3D model itself is, however, limited. GenomeFlow does not allow visibility management and bins cannot be selected from the 3D view. Trieu et al. [42] also do not mention the ability of linked views and how the interaction between the 2D Hi-C maps and 3D predicted structure is coordinated. The tool is implemented as a desktop application and has not been maintained recently, in our opinion, limits its utility for recent datasets.

Similarly, HiC-3DViewer [8] combines modeling and visualization albeit on web using client-server architecture. To work with the 3D structure, it offers mapping of genomic signals onto the 3D model. The 3D view is complemented with a pop-up Hi-C map and a 1D track panel. Selection in the Hi-C map is reflected in the 3D view, but selection in the opposite direction is not possible. The rendering uses flat colors without shading, limiting the perception of the chromatin’s overall shape. The tool does not include cutaways or filtering, leading to high visual complexity.

Nucleome Browser [47] developed by the 4D Nucleosome consortium is the third relevant tool we identified. It combines linked views offering all possible modalities: 1D for genomic signal tracks, 2D for Hi-C contact maps, and 3D for predicted 3D chromatin structure. The overall functionality and interaction inside the 3D view are rudimentary. The browser allows switching between global and local views but only at three granularities. Arbitrary selections are only unidirectional: users can select genomic regions in the 1D or 2D views which are reflected in the 3D, but not the other way around. Only a whole chromosome or a single bin can be selected from the 3D view.

WashU Epigenome Browser [20] is a genomic browser combining and linking different views. Its 3D view implements bidirectional linking in a limited way, allowing only single-point selections. This tool, however, offers a large number of options for annotating the 3D structure. Besides the prevalent coloring by numerical values, it can label the 3D structure with text and glyphs.

Currently, most web-based 3D genome visualizations use WebGL or libraries built on it, e.g., three.js. WebGL lacks in capabilities compared to graphics APIs for native desktops. This limits both the data sizes web viewers can render interactively and the visual fidelity. Furthermore, visualizing large chromatin datasets opens up an issue with
occlusion, which inhibits a proper exploration of the 3D dataset. None of the reviewed existing tools include any techniques for occlusion management to allow peeking into a dense 3D dataset, apart from a few tools that implement hiding selected regions. Finally, a lack of interaction options for exploring and manipulating the 3D genome model is prominent across all the surveyed tools. For the most part, the linking functionality is implemented in the 1D-to-3D or 2D-to-3D direction but the other way around is often limited. Users can usually select only singular bins, and have to turn back to 1D and 2D views for advanced selections. This hinders the exploration and prevents reasoning about the observed chromatin structure.

4 Requirements

We base our requirements on discussions with a domain scientist during the initial phases of our collaboration. Furthermore, we consider our investigation of related tools and the exposed feature set limits detailed in the previous section. As a result, we define the following requirements for a novel chromatin visualization tool:

- **Efficient 3D rendering:** The tool should be able to handle large datasets with sizes in the magnitude of hundreds of thousands of bins and render them in real-time with interactive framerates (30+ FPS).
- **Support shape comprehension:** The visual representations should highlight the spatiality of data and help users to gain an understanding of the shape and size of the model.
- **Visibility management:** Chromatin models tend to be dense and highly occluded. Biologists need be able to strategically highlight salient parts while removing the occlusion that prevents localizing them. Only the least amount of information should be removed to preserve context.
- **Bi-directional linking:** Users should be able to select genomic loci in the 3D view or the other (1D and 2D) representations. These selections should be mirrored in the remaining views as they are all better suited to different tasks. After localizing a significant part in the 3D view, it may be desired to look at multiple correlating data in several tracks of 1D views.

The requirements served as constraints and guiding principles for the design of our novel tool, which we describe in the following section.

5 3D-Centric Design of ChromoSkein

In this section, we describe our design choices and detail the implementation of a 3D-centric chromatin visualization tool, shown in Figure 2, which we call ChromoSkein. We describe how we turn the 3D data into visual representations and how we render and shade the visual marks. Afterward, we discuss interactions with the 3D visualization.

5.1 Visual Representations

3D chromatin modeling methods output a set of discrete 3D positions associated with nucleotide bins. The set is ordered, i.e., the order of bins corresponds to the order of individual segments in the DNA sequence. We reviewed representations used in the existing 3D chromatin visualization tools, see Table 1. In the majority of cases, tools include spherical and tubular representations, sometimes combining both in what is known in molecular visualization as balls-and-stick representation. Some tools implement more unconventional marks, such as crosses in Nucleome Browser [47]. Working with large 3D models, visual clutter is of primary concern. For that reason, simple and clear visual marks are preferred. We decided to include three established spatial representations—a *spherical* and two *continuous tubular* representations, see Figure 3.
When rendering the tubular representations, we estimate the optimal tube thickness based on the proximity of bins. To ensure a reasonable thickness, we limit it to half the distance between two bins so that two bins do not overlap and visibly merge into one. Depending on the used model reconstruction algorithm, the spacing between consecutive bins is not always uniform. Therefore, we use a statistical rule based on the interquartile range (IQR), to bound it between: 
\[ \frac{Q_1}{Q_3} \pm 1.5 * IQR, \]
where \( IQR = Q_3 - Q_1 \), \( Q_1 = \) lower quartile, \( Q_3 = \) upper quartile. The default thickness is set in the middle of this range and users can adjust the value within the range.

When it comes to shading, virtually all the existing chromatin visualization tools we reviewed employ a simple Phong shading model. Some tools even skip all shading and only show the 3D model with flat colors [8]. Phong shading model considers only local illumination omitting shading by neighboring structures. For larger chromatin models, it leads to an ambiguous view where it is impossible to discriminate global features such as holes and crevices. Thus, the comprehension of an overall shape is hindered.

To enhance perception of such spatial features, we shade the scene with real-time screen space ambient occlusion (SSAO) [39]. It can be difficult to configure the SSAO radius parameter for large datasets, as a large radius accentuates only bins deeply inside the structure. In contrast, a small radius highlights only differences between close objects. We rectify this by stacking two results of SSAO computations with small (for near objects) and large (for deeply buried objects) radii. Compared to the Phong shading model, in ChromoSkein we are clearly able to comprehend shape features, such as a hole in the center of a genome occupied in cells by nucleolus. A comparison can be seen in Figure 4.

5.3 Interactions With 3D Chromatin

We aim to give bioinformaticians the tools necessary to go beyond static visualization and to support them in analyzing the 3D model directly in its spatial context. Next, we describe two interconnected interactions used to filter, explore, and examine the 3D data.

5.3.1 Selections

One of the key interactions with 3D chromatin required by biologists is the ability to select genomic regions. The selection task is a prominent interaction across all visualization tools [45]. Available genomic tools, for the most part, implemented selections in 1D and 2D representations and highlighted the selected parts in the 3D model. Very few of them allow selecting objects directly in the 3D model. A bioinformatician might wish to select bins based on some spatial feature, e.g., bins close to a surface or, the opposite, deeply nested in the nucleus. Such task is impossible to do in the matrix representation. Therefore, in ChromoSkein, we designed methods for selecting bins in the 3D view.

In the 3D view, the atomic element is a single bin. Thus, we choose to represent selections as a subset of the model bins. We allow users to create an arbitrary number of selections. In ChromoSkein, we implemented three selection tools for the 3D view that together support all possible interactions we identified as critical for the users:

- Rule-based selection: users can apply a rule-based selection tool to highlight bins that meet specific criteria.
- Manual selection: users can manually select bins by clicking on them.
- Region selection: users can select a region of interest by drawing a selection box.

These selection tools allow users to efficiently explore and analyze the chromatin data in 3D.

5.2 3D Rendering & Shading

We currently deal with chromatin spatial models containing over tens of thousands of bins: our largest model of Mouse genome comprises of 25 thousand bins. In the future we expect even bigger models, as experimental methods improve in resolution. Such data sizes present a challenge for rendering in real-time, especially in the web environment. To achieve high performance, we avoid polygonal representation and instead use billboards with screen-space ray-traced parametric objects. The advantage of this approach is twofold: We lower the number of vertices that need to be updated every frame and we obtain a pixel-perfect objects representation. For the straight tubular representation, we use efficient and visually correct rounded cones described by Groß and Gumhold [14].

To compute the smooth tubular representation precisely, we would need to fit a cubic spline through the data bins. However, to achieve fast rendering, we only approximate the cubic spline with quadratic Bezier curves, as proposed by Truong et al. [43]. To render the tube itself, we use the method by Reshetov and Luebke [32].

The spherical representation is the most straightforward way of displaying the raw modeling data. Each bin—representing a certain number of nucleotides—is mapped to a sphere. This visual encoding indicates that the positional information is available only at this granularity for this whole genomic segment, i.e., it is impossible to say where each nucleotide lies in 3D space. The lost connectivity information is an undeniable limitation of the spherical representation. Therefore, we also provide two continuous representations. In the first, straight tubular representation, the consecutive bin positions are connected with straight tubes. This representation communicates the connectivity, while the original data points can still be inferred from the positions of the tube joints. However, the results can look unnatural, particularly for low-resolution data, where long straight segments and sharp angles at joints occur. Therefore, the last, smooth tubular representation, uses the bin positions to define a spline that forms the centerline of a tube. This technique produces more visually pleasing and organic-looking results, but it comes at the cost of no longer providing precise information about original bin positions.

We chose these three options mainly because of their simplicity. We specifically decided not to include a ball-and-stick representation: In this visual mapping there are twice as many marks on the screen, which leads to visual clutter without any added benefit.

Fig. 3. Three types of visual representations available in ChromoSkein. From left: spherical, straight tubular, and smooth tubular representation.
Point selection, Continuous sequence selection, and Sphere selection, all demonstrated in Figure 5.

The simplest way of selecting bins is selecting one bin at a time. Point selection serves for the definition of small selections of visually prominent bins or fine selection refinements. It is performed by clicking on the given bin in the 3D view. One of the most important features found in 3D chromatin is which bins or genomic loci are located in a neighborhood to a selected location. Therefore, Sphere selection allows selecting bins within a certain spatial distance from a given point or bin. Users can again pick a single bin by clicking on the 3D representation, but in this case, all bins within a spherical radius are selected. The radius is adjustable and the bins within the radius are highlighted upon hover. Finally, biologists are sometimes interested in chromatin loops, where the loop end points are distant in the linear sequence but spatially close. To allow selecting loops in 3D, we use Continuous sequence selection. When two bins are picked in the 3D view, the linear sequence of bins between them is selected. We display the selections in 3D by coloring the visual marks representing bins. The colors are randomly generated for each track but can be adjusted by the user.

5.3.2 Occlusion Management

Chromatin fiber is tightly packed in a nucleus, resulting in a highly intertwined structure with large parts hidden due to occlusion. Biologists are interested to know, for example, how deeply nested a genomic locus is—the position can affect the function or regulation of genes located in the locus. Therefore, providing bioinformaticians tools to manage occlusion is essential. We include two techniques to filter the model by adjusting visibility of its parts: one semantic and one based on spatial position.

ChromoSkein allows users to segment the chromatin model by adding bins to selections tracks. The first tool for filtering the 3D model is by toggling visibility of each such segmentation track. Users can hide any track which leads to hiding of associated bins from the 3D representation. Segmentation tracks can also be loaded from an external file. This allows, for example, loading chromosome segmentation and exploring the model by toggling visibility of individual chromosomes.

Sometimes the focus of the analysis cannot be defined semantically but rather spatially. To support these cases, ChromoSkein provides cuttings planes. Users can add one or more cutting planes either along one of the major axes or an arbitrarily oriented plane spanning from the camera’s point of view. The model is then cut by the plane: The primitives building the 3D bin representations are clipped at the intersection and resulting surface holes are filled. Clipping planes work in tandem with our rendering style supporting depth cues, highlighting both local and global structural features, such as holes nested inside the model. Users can also choose to keep some selections visible at all times which can be useful for studying immersed parts while keeping parts of their surroundings. All these features are presented in Figure 6.

6 Linking 3D With Other Data

Due to the complexity of investigating organization and function of cell nucleus—brought by the dynamicity and scale range—it is essential that biologists integrate several available modalities, each focusing at different aspects of chromatin fiber. Other genomic data and visualizations have been proposed, as reviewed by Nusrat et al. [28].

In this section, we dive into the design of linking the 3D representation described above with non-spatial data types and their conventional visual representations.
The distance map is generated from the 3D model and in ChromoSkein ical values denoting frequencies of contact between bins. In 6.1.2 Track View: Distance Map
tions are drawn as horizontally stacked bar charts. display signal tracks as simple line charts while segmenta-
come from an external file or result from user selections. We in BED format. Segmentations can also be relevant: e.g., A/B compartments based on chromosome segmentation, as seen in several 3D recon-
We display the distance map as a triangular shape. The horizontal axis shows bins and each triangular field carries the distance information mapped to a color value. Users can zoom and pan around the 2D map. As an optimization, the distance map is calculated on-demand based on which part of the map is zoomed into.

We employ a simple level-of-detail scheme. If the size of a distance map is such that it cannot allocate enough pixels for all bins, a level change is applied. We merge multiple consecutive bins by averaging their positions and calculate the distance between those larger bins. Distance maps can be also generated from custom selections. This can be useful to observe distance relationships of a subset of chromatin.

6.1 Linked Views
Visualization systems typically connect different data by using multiple linked views [25]. Not all tasks are best performed in 3D view. Additionally, many biologists are already trained on and used to 1D and 2D visualizations. Therefore, we include two conventional genomic visualizations in ChromoSkein. First, we implemented a genomic browser to display 1D genomic signals. Second, we included a distance map to serve as a proxy for any 2D maps, e.g., Hi-C matrix viewer. In ChromoSkein, we couple these two linear data visualizations into a track view, shown in Figure 7.

Both the distance map and the genomic browser are prototypical implementations to complete the feature set required by the bioinformaticians analytical workflow. There are tools that focus more on each modality, e.g., HiGlass for Hi-C maps, but do not include ChromoSkein’s capabilities for 3D data.

6.1.1 Track View: Genomic Browser
Genomic browsers are the most common way of working with genomic data. They display data laid out along one axis—usually the horizontal x-axis—annotated by genomic coordinates. Two typical examples of data shown in genomic browsers are genomic signals coming from biological experiments, e.g., ChIP-seq which analyzes protein-DNA interactions [17], and gene annotations marking positions of genes on the DNA sequence.

In ChromoSkein, we differentiate two types of 1D data: segmentations and signals. Segmentations assign genomic regions to a segmentation track while signals assign numerical values to each bin. The signals can be loaded into ChromoSkein in BED format. Segmentation tracks can either come from an external file or result from user selections. We display signal tracks as simple line charts while segmentations are drawn as horizontally stacked bar charts.

6.1.2 Track View: Distance Map
Hi-C experiments typically output 2D matrices with numerical values denoting frequencies of contact between bins. In ChromoSkein, a distance map stands in for a full Hi-C map. The distance map is generated from the 3D model and in principle should contain the same information. 3D reconstruction algorithms are rated according to the difference between the generated distance map and the original Hi-C map. If the reconstruction is good, the two 2D maps should be more or less equal.

We display the distance map as a triangular shape. The horizontal axis shows bins and each triangular field carries the distance information mapped to a color value. Users can zoom and pan around the 2D map. As an optimization, the distance map is calculated on-demand based on which part of the map is zoomed into.

We employ a simple level-of-detail scheme. If the size of a distance map is such that it cannot allocate enough pixels for all bins, a level change is applied. We merge multiple consecutive bins by averaging their positions and calculate the distance between those larger bins. Distance maps can be also generated from custom selections. This can be useful to observe distance relationships of a subset of chromatin.

6.1.3 Interactive Linking
Nusrat et al. [28] mention that views can be weakly, medium, or strongly linked. That simple categorization applies for 1D or 2D views that share reference frame of genomic coordinates. The 3D view steps out of this conformity, which extends the design space of possible linking.

We allow users to create many panels of the two types, i.e., 3D view or track view. Users can synchronize cameras between 3D views. That way they can observe one 3D model with different adjustments, e.g., hidden specific chromosomes. The 2D distance map and 1D genomic tracks in a track view are linked by default: Zooming and panning will be reflected across all tracks in one track view.

The linking between a 3D view and a track view is accomplished through custom selections. In the 3D view biologists select parts that are interesting because of their spatial features. On the other hand, they might also want to let the selections be guided by either the linear data, e.g., signal peaks, or patterns in the 2D map. For that reason, we implemented several selection tools also for the track view.

6.2 Annotating the 3D Structure
While linked views are based on the concepts of juxtaposition, superimposing data in the 3D is also possible. This allows biologists to contextualize 1D genomic data in space. Biologists call this annotating the 3D model. There are several ways that the color channel can be used to augment the 3D structure with other information. All types of annotation are demonstrated in Figure 8.

The first type of annotation is coloring the 3D representation based on chromosome segmentation, as seen in several existing 3D chromatin visualization tools. However, other segmentations can also be relevant: e.g., A/B compartments or TADs. In ChromoSkein, we therefore generalize this type of annotation. We allow users to segment the model by selecting its parts using several selection tools or loading the segmentation from an external file. We then color the 3D representation as we described in Section 5.3.1.

However, annotating the 3D model based on segmentation is not the only way to use the color channel. We can also map genomic signals from other biological experiments and superimpose the 1D data onto 3D.

Fig. 7. Example of a track view containing 4 tracks (from top to bottom): distance map, segmentation into chromosomes (loaded as selections from a file), custom user-made selection in 3D viewport, and 1D track (in this case solvent accessible surface area).
These types of data typically have more fine-grained resolution than 3D structures. Therefore, we allow users to choose how multiple values per bin are mapped, i.e., a minimum, maximum, average, or median. In order to color bins, we normalize the data and then apply one of the color maps recommended by the scientific color guide [5].

In some cases, the values might not come from an external file but instead can be computed on-the-fly. Bioinformatics tools often require sophisticated computation and powerful hardware but some algorithms can be implemented and run directly on a client device. This removes the need for running an external tool and importing the results into a visualization system. Users can thus iterate faster and often re-define the model subset meant for computation. As an example, we include computation of Solvent Accessible Surface Area (SASA) in ChromoSkein.

Finally, biologists are often interested in highlighting short segments of the DNA that carry functional elements such as genes. These short segments, called loci (singular locus), generally fall into a range less than a single bin. Therefore, semantically, we can consider them as short segments and implement them by selections. However, due to convention, we chose to highlight such loci using markers. Since markers most often span just a single bin, we display them as spheres and make them salient by using a greater radius and a different color, adjustable by the user.

7 Implementation

ChromoSkein is implemented as a client-only web application, which makes it usable without the need to install desktop software. The omission of a remote server component ensures that potentially confidential data need not be uploaded to a third-party server.

ChromoSkein itself is divided into two main parts: the application itself, written using the popular React front-end framework, and a narrowly focused visualization library written using the WebGPU API [24]. We decided to use WebGPU as it offers low-level access to GPU and addresses many of the issues of WebGL. Note that WebGPU is currently in development and requires experimental versions of modern browsers, e.g., Google Chrome Canary. The project is open-sourced at https://github.com/chromoskein/chromoskein, along with instructions on how to set up the browser to run our tool.

8 Exemplary Use Cases

We present two use cases prepared in collaboration with a biologist to demonstrate the capabilities of ChromoSkein. The biologist first performed analysis purely from 2D matrices, pointing out its deficiencies. Afterward, they utilized 3D visualization to gain additional insight beyond what is possible from only 1D and 2D genomic data.

8.1 Case study I

In the first case study, we look at chromosomal dynamics from Di Stefano et al. [6], focusing on the 3D model of a region surrounding mouse gene Sox2 and the gene itself. Without the utilization of 3D visualization, one has very limited options for exploring the dynamical behavior of Sox2 gene from pure XYZ coordinates; some of the potential analysis approaches are shown in Figure 9a-b.

Firstly, the number of pairwise contacts to Sox2 gene was explored as a function of time and number of interactions up to certain distance. In most of the selected distance cutoffs, the number of contacts increases in the first 10 000 steps of dynamical simulation while steadily lowering below starting point until the end of the simulation (step 60 000). This suggests a tight packing of the chromatin fiber around Sox2 gene at the beginning of the simulation where the number of the interactions is higher. As the simulation progresses, the fiber around Sox2 gene will likely unfold.

Secondly, one can monitor the spatial trajectory of Sox2 gene to understand its dynamics and mobility. If the monitored position fluctuates in all three axes uniformly, its mobility is considered isotropic. Constraints in one of the axes hint at the presence of anisotropy. This can be related to the stability of specific conformation, which may play an important role in the regulation of gene transcription.

However, without further visualization and exploration of 3D dynamics, it is not possible to infer further details. Visualization of Sox2 (Figure 9b) trajectory helps to understand these processes better and clarify the hypothesis based on previous analysis (Figure 9a-b). Indeed, it confirms the unfolding of chromatin locus in later stages of simulation (>10 000 s) and the rearrangement of chromatin folding around Sox2. Moreover, Chromoskein can calculate solvent accessible surface area (SASA), which allows further interpretation of un/folding events. Although Sox2 region converges to conformation with fewer interactions (Figure 9b), SASA values of most bins are lower towards the end of the simulation (brighter blue in Figure 9a) than in the initial conformation, making it less accessible to potential transcription factors (specific proteins regulating gene activity). This suggests the tendency for more permanent gene expression of Sox2 gene with less flexible regulation.
8.2 Case study II

In the second case study, we inspected the same Sox2 gene in the context of whole genome organization, as published by Stevens et al. [36], along with the addition of Lmo7, Neurod6, Rergl, Tet2 genes.

Exploration of Hi-C matrices or single coordinates without 3D visualization will inform us about probabilistic interactions among all the chromosomes but not about their mutual spatial arrangements/colocalization. Distance distribution estimates of a selected gene towards the rest of the chromosome can indirectly inform us about the folding of a region surrounding the gene (Figure 10a-b).

3D visualization of whole genome structure allowed us to uncover that Tet2 and Rergl genes are located on the ‘surface’ of the genome. Their corresponding chromosomes are likely interacting with so-called nuclear lamina (Figure 10c1-c2). This may suggest genes’ participation in specific processes, such as lamina regulation or nuclear transport.

There are many non-genome arrangements that cannot be captured via Hi-C experiment. However, they can manifest themselves within the 3D genome model as cavities or holes; therefore, depth cues and usage of cutting planes can provide us with information about them.

By sliding a cutting plane through the whole genome 3D structure (Figure 10c3), we can uncover the rest of genes, located on chromosomes buried deeply inside the chromatin core. We used the ability to keep chromosomes visible even under the cutting plane to keep the context. We find that Neurod gene is positioned in close proximity to a spatial caveat inside the genome, which is most likely occupied by the nucleolus, suggesting an important regulatory role or potential interaction with rRNA or rDNA.

9 Conclusion

We are currently receiving progressively better predictions of how DNA interacts in the 3D space. Formulating hypotheses about overall chromatin conformation intrinsically requires appropriate visual support, which is lacking in the currently available tools. In our work, we aim to bring the 3D spatial representation of chromatin to the forefront and design a novel tool that would assist experts in several stages of their research. We clearly identified a gap in the available software, where the semantics of the domain problem elicit new methods, making existing molecular visualization tools insufficient.

We believe that our newly proposed ChromoSkein tool will contribute to understanding of the importance of chromatin spatial organization. We foresee that the description of our design choices, the architecture of the tool, along with the availability of the source code, will serve the communities of chromatin experts, software developers, and visualization experts in adopting our solutions.

The described visualization framework gives us a solid basis for further research. Several topics already came up.
in preliminary discussions with domain experts. First, like other biological phenomena, chromatin is a dynamic structure that is constantly changing. Studying these movements is highly important since chromatin adjusts to the cell cycle and drastically changes its conformation. From the visualization perspective, time series data for 3D models of chromatin present intriguing challenges for which novel solutions will be required. Second, studying the evolutionary changes in chromatin structure and spatial organization among and across species can be naturally supported by comparative visualization. Finally, chromatin research is intrinsically based on integration of experimental methods, data types, and data sources. The same approach is being applied to tooling, now also enabled thanks to modern web technologies. In the future, we foresee a high potential in integrating our tool with applications such as HiGlass, that are targeting to solve one specific problem.

ACKNOWLEDGMENTS
The presented work has been supported by the Czech Ministry of Education project no. LIT20035. Authors wish to thank Marc Marti-Renom for consultations and Hanka Pokojá for creating ChromoSkein’s logo.

REFERENCES
[1] T. M. Ashbury, M. Mitman, J. Tang, and W. J. Zheng. "Genome 3D: A viewer-modeling framework for integrating and visualizing multi-scale genome information within a three-dimensional genome." BMC bioinformatics 11.1 (2010): 444.
[2] J. M. Belton, R. P. McCord, J. H. Gibeaut, N. Naumova, Y. Zhan, and J. Dekker. "Hi-C: A comprehensive technique to capture the conformation of genomes." Methods 58.3 (2012): 282-292.
[3] A. Butaeva, P. Mavlyutov, M. Blanchette, P. Cudré-Mauroux, and M. N. Goodstadt. "Challenges for visualizing three-dimensional in vivo protein-DNA interactions." Methods 58.3 (2012): 323-334.
[4] Y. Chi, J. Shi, D. Xing, and L. Tan. "Every gene everywhere all at once: High-precision measurement of 3D chromosome architecture with single-cell Hi-C." Frontiers in Molecular Biosciences 9 (2022): 10.3389/fmolb.2022.959688.
[5] J. Nowotny, A. Wells, O. Oluwadare, L. Xu, R. Cao, T. Trieu, J. Waldispühl, "A low-latency, big database system and browser for storage, querying and visualization of 3D genomic data," Nucleic Acids Research 49.5 (2021): 3941-3954.
[6] D. Li, S. Hsu, D. Purushotham, R. L. Sears, and T. Wang. "WashU Epigenome Browser update 2019." Nucleic Acids Research 47.13 (2019): W115-W126.
[7] P. Kerpedjiev, N. Abdennur, F. Lekschas, C. McCallum, K. Dinkla, H. Strobel, J. M. Luiter, S. B. Quellet, A. Aziz, N. Kumar, J. Hwang, S. Lee, B. H. Alver, H. Pfister, L. A. Mirny, F. Park, and N. Gehrke. "HiGlass: web-based visual exploration and analysis of genome interaction data." Genome Biology 19.9 (2018): 1-18.
[8] T. Graf, and M. A. Marti-Renom. "Transcriptional activation during cell reprogramming correlates with the formation of 3D open chromatin hubs." Nature Communications 11.1 (2020): 1-11.
[9] N. D.版本, S. Selvaraj, R. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, and B. Ken. "Topological domains in mammalian genomes identified by analysis of chromatin interactions." Nature 485.7398 (2012): 380-383.
[10] M. N. Goodstadt, M. Wang, M. Q. Zhang, and J. Gao. "HiC 3D View: a new tool to visualize Hi-C data in 3D space." Quantitative Biology 7.2 (2020): 183-190.
[11] N. C. Durand, J. T. Robinson, M. S. Shamim, I. Machol, P. Mesirov, E. S. Lander, and E. L. Aiden. "Jukebox provides a visualization system for Hi-C contact maps with unlimited zoom." Cell Systems 3.1 (2015): 99-101.
[12] J. Gibeaut and J. Dekker. "The Hierarchy of the 3D Genome." Molecular Cell 49.5 (2013): 775-782.
[13] M. N. Goodstadt, D. Castillo, and M. A. Marti-Renom. "TADkit v.1,01." 2015.
