Computers & Graphics (2024)

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Virtual reality inspection of chromatin 3D and 2D data

Elena Molina[a,4], David Kouřil[b,1], Tobias Isenberg[c,2], Barbora Kozlíková[d,4], Pere-Pau Vázquez[1,4,*]

[a]Jordi Girona 3-5, Barcelona, 08034, Spain
[b]10 Shattuck Street, Suite 514, Boston, MA 02115, US
[c]Bât 660, Digiteo–Moulon, Orsay, 91405, France
[d]Botanická 554/68a, Brno, 602 00, Czech Republic

ARTICLE INFO

Article history:
Received 3 May 2024
Received in final form 7 August 2024
Accepted 20 August 2024

ABSTRACT

Understanding the packing of long DNA strands into chromatin is one of the ultimate challenges in genomic research. An intrinsic part of this complex problem is studying the chromatin's spatial structure. Biologists reconstruct 3D models of chromatin from experimental data, yet the exploration and analysis of such 3D structures is limited in existing genomic data visualization tools. To improve this situation, we investigated the current options of immersive methods and designed a prototypical VR visualization tool for 3D chromatin models that leverages virtual reality to deal with the spatial data. We showcase the tool in three primary use cases. First, we provide an overall 3D shape overview of the chromatin to facilitate the identification of regions of interest and the selection for further investigation. Second, we include the option to export the selected regions and elements in the BED format, which can be loaded into common analytical tools. Third, we integrate epigenetic modification data along the sequence that influence gene expression, either as in-world 2D charts or overlaid on the 3D structure itself. We developed our application in collaboration with two domain experts and gathered insights from two informal studies with five other experts.

1. Introduction

Gene expression and regulation are largely influenced by the packing of the DNA strand into the 3D structure of the chromatin in the cell nucleus [38]. Unfortunately, the captured data of chromosomal conformation, using techniques such as Hi-C [29], are difficult to interpret with traditionally used 2D visual representations. Most of the current analysis tools either do not provide a 3D inspection mode, or they support only a basic 3D projection on the desktop, with limited interaction options [8, 27, 49]. Immersive systems, however, have repeatedly proven their superior ability to help users understand complex spatial arrangements of 3D structures [4, 6], as well as to provide more elaborate and intuitive interactions [24, 36]. Especially for chromatin data, 3D analysis is highly relevant because the genome is compartmentalized. Different regions have different behaviors, and to understand those, knowledge about the 3D structure of the genome is crucial. For example, understanding the spatial arrangement of chromatin may increase the chance of finding options for treating a disease by applying genetic therapy techniques to the neighboring genes, instead of the genes themselves. Furthermore, they also help to reveal how local changes may affect other regions of the genome that are spatially close but may be distant in the sequence (i.e., along the chromatin fiber).

To explore the options that the fusion of virtual reality with visualizing the spatial arrangement of chromatin can offer, we de-

*Corresponding author: Tel.: +34-93-4137990; fax: +0-000-000-0000;
e-mail: pere.pau.vazquez@upc.edu (Pere-Pau Vázquez)
[1]Department of Biomedical Informatics, Harvard Medical School
[2]Université Paris-Saclay, CNRS, Inria, LISN
[3]Masaryk University
[4]ViRVIG Group, UPC

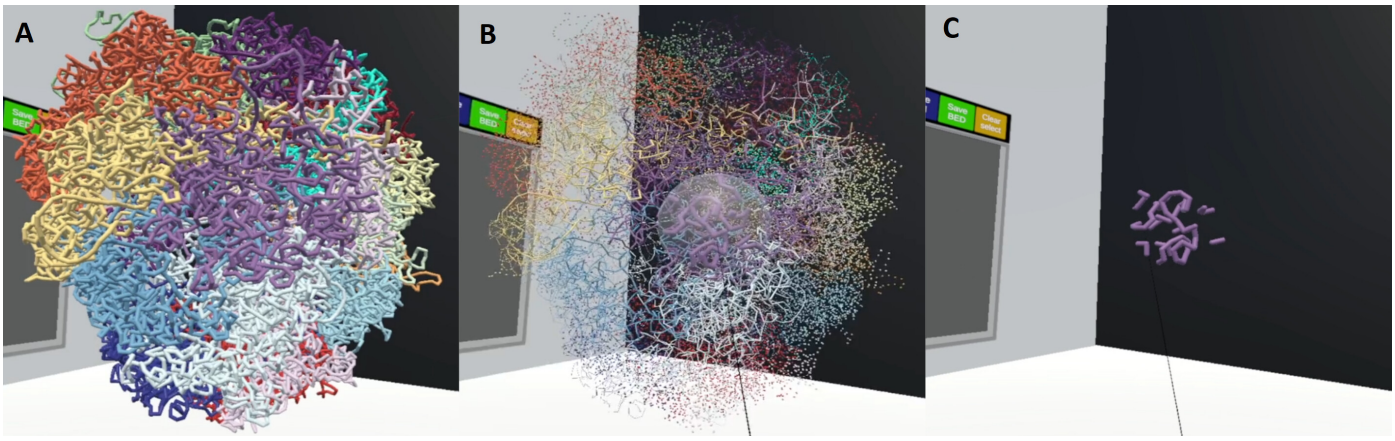

Fig. 1. Virtual Reality inspection of chromatin 3D and 2D data. (a) Visualization of the 3D chromatin model in the virtual environment. (b) Model exploration with the selection of a Region of Interest (ROI) encapsulated in a sphere. (c) The chromatin structure is hidden, except for the selected ROI.

signed a new Virtual Reality (VR) tool, see Figure 1, to showcase that it can complement the work pipeline of bioinformaticians and biologists in their analysis of chromatin data. We intensively discussed the design and functions with experts in genomic research. Our system can load 3D chromatin models, allows users to explore and select regions of interest, and provides a visualization of epigenetic modification data on the models for an initial analysis. Together with the experts and based on the results of our experiments, we identified four exemplary applications of the exploration for chromatin data, where VR capabilities can play a vital role: *(i)* the initial inspection of the 3D shape to enable users to gain a comprehensive understanding of the 3D shape and identify regions of interest that may otherwise not be detected, *(ii)* the selection of interesting portions of the model that are exported as BED files [37] for further inspection in the traditional software, *(iii)* an initial analysis of so-called tracks—files with epigenetic modification information of the chromatin—that can be rendered over the surfaces of the chromosomes to make sense on their relationship with the 3D space, and *(iv)* the assessment of hypotheses after 2D inspection has been carried out, i.e., to determine whether certain hypotheses are physically correct. The epigenetic modifications (tracks) can induce changes in chromatin structure, either locally or over long distances, and changes in gene expression. For instance, histone acetylation disrupts electrostatic interactions between histones and DNA, leading to a less compact state of the structure, thus enabling the access of other proteins [3].

Through interviews with two domain experts, we first defined initial requirements, based on which we created an initial version of the tool, capable of interactively exploring 3D chromatin models in VR. We then iteratively refined the tool with one of the experts. Next, we asked five additional and independent experts to examine the tool: the first of them as a pilot test, while the other four completed a full experiment. The qualitative feedback revealed that the experts are highly interested in using VR as a new instrument to complement their pipelines. The contribution of our work is thus not a new visualization paradigm, the use of immersive analytics [14] in general, or new interaction paradigms. Rather, we contribute a design study of the use of VR-based visualization in a practical use case and the demonstration that application domains, which traditionally rely solely on desktop tools, can take advantage of the advanced spatial comprehension facilitated by immersive technology [15, 23].

2. Previous Work

VR rapidly evolves into a mature technology with applications across many data visualization domains. Our case of 3D genomic data brings together advances in biochemical research and subsequent computational processing, with work on efficient data visualization and interaction in immersive environments.

2.1. Genomic Data Visualization

Visualization has long been a fundamental part of genomic data analysis. Nusrat et al. [35] give an overview of the diverse visual data representations employed in genomics. Historically, the UCSC Genome Browser [21] and the Ensembl genome database [31] defined the now conventional paradigm of aligning genomic data—often called *genomic signals*—along the genomic DNA sequence in a linear fashion. Such linear data offer only limited means for studying the genome's structural features, which often play a significant role in diseases [5].

In 2002, Dekker et al. [9] developed the Chromosome Conformation Capture (3C) technique, which gave researchers the necessary tool to investigate the spatial organization of genomic material in cells. This advance was followed by extensions, e.g., 4C [40], 5C [12]. Ultimately, Lieberman-Aiden et al. [29] came up with the Hi-C technique that provided genome-wide interactions between all equally-sized regions called *bins*. The results are 2D matrices encoding contact frequencies between the bins, i.e., what is the likelihood of the two regions being in close proximity in vivo. Several tools for visualizing these 2D matrices—often extensive and multiscale—have been developed (e.g., Juicebox [13] or HiGlass [22]). Over the years, biologists and bioinformaticians have become accustomed to examining the spatial structural features of DNA through the 2D matrix representation. Different 3D features are displayed as specific patterns in the 2D matrix, e.g., A/B compartments [29] or Topologically Associated Domains (TADs) [10].

In summary, researchers in the genomics field learned to combine one-dimensional datasets, i.e., genomic signals, and two-dimensional datasets, i.e., matrices encoding interaction frequencies, to investigate both functional and structural features of the genome. The DNA folding in the cell nucleus, however, by definition happens in three dimensions. Many scientists are thus interested in generating 3D DNA models [16, 17, 30]. While certain mechanisms of DNA folding have been identified, the whole packing process remains unexplained. We know, for instance, that the DNA polymer is wound around histones (proteins providing structural support), forming nucleosomes. The whole folded structure with other supporting proteins is often called chromatin—a term that we also use throughout the paper. Further, high-level folding is currently difficult to observe due to the scale at which it occurs. Modeling chromatin in silico and visualizing the generated 3D structures aims to address questions regarding spatial conformations. 3D models generated this way are the focus of our work presented in this paper.

2.2. 3D Chromatin Visualization

Multiple ways exist to acquire a 3D model of chromatin folding in the cell nucleus. First, one can generate such structure from the experimental data, i.e., Hi-C matrices. Another option involves polymer simulations, using known physical properties of the DNA fiber as constraints. Imakaev et al. [20] give more details about the various paths to a 3D model of a genome. In both cases, the methods generate a prediction of the chromatin fiber's spatial conformation, satisfying the input conditions.

The visualization of such 3D structure predictions has been proposed by several authors and implemented in several bioinformatics tools. For example, two general-purpose web-based genome browsers include a 3D component: Nucleome Browser [49] and WashU Epigenome Browser [27]. Both tools are actively developed and maintained and provide the baseline functionality for 3D chromatin visualization on desktops.

In the past, a handful of other tools proposed 3D visualization in genomics. Genome3D [2] was one of the earliest tools for visualizing genomes in 3D. GMOL [34] decided to separate the multiscale model into discrete scales and allowed users to switch between them. 3Disease Browser [28] focused on disease-specific conformation changes. HiC-3DViewer [11] facilitated 1D-to-2D-to-3D data mapping, although not the other way around. Delta [42] also supplemented the 3D model with other conventional genomic visualizations. GenomeFlow [45] proposed various ways to link 1D, 2D, and 3D genomic data in a desktop application, whereas CSynth [44] combines simulation with 2D matrix integration for the 3D view.

Despite this publication history, the full potential of 3D visualization as it is established in the visualization domain has not yet been used. In particular, VR can underline the spatial aspects of the data, which we aim to explore in our work.

2.3. VR in Genomic and Molecular Visualization

There are just a few examples of virtual reality applied to 3D chromatin data. CSynth [44] is a web application that renders 3D chromatin models in the browser. They provide an option to show the 3D chromatin models in virtual reality, mainly for education purposes. CSynth links the 3D structure with a 2D matrix by highlighting selected regions. Furthermore, users can map a single genomic track onto the color of the 3D chromatin model. Our work aims to support linking the chromatin model with multiple genomic tracks at the same time. Delta AR [43] also features immersive environment viewing in Augmented Reality through HoloLens glasses. However, they only show the 3D model and leave the interactions with other data, such as tracks, to 2D displays. Reiske et al. [39] propose an ideogram-inspired representation of a genome in VR, and compare three querying techniques in VR and desktop. However, they did not use 3D chromatin structures, but textual gene information. Moreover, the study only comprised the 3D interaction techniques and did not include domain experts. In a related field of general molecular data visualization, many more works have been published. Kuťák et al. [24] overview the landscape of VR in molecular visualization. Despite some attempts at using Virtual Reality for chromatin visualization, as we have seen, they primarily intend to serve as a showcase component, rather than forming an integral part of the workflow of domain experts. Our proposal fills this gap by creating a VR tool that can effectively solve problems that are not feasible (or are more difficult) to solve using standard software tools.

3. Chromatin Exploration in VR

Depending on the source organism and resolution of the data, a single genome might result in models containing thousands to tens of thousands of bins. This situation presents several issues for intuitive exploration, primarily due to occlusion from projecting the 3D model on a two-dimensional screen. Furthermore, these models are composed of features at different scales: e.g., chromosomes, TADs, and genes. We aim to provide tools that not only enable global exploration of the whole model but also support filtering these features, while helping with the comprehension of spatial features thanks to the use of VR.

3.1. Motivation and Requirements

Only 5% of human chromatin is composed of genes, approximately 20K, and only about 25% are typically active, being transcribed into proteins. Such a transcription event is adjusted by regulatory sequences of DNA that promote or silence gene expression. These regulations occur not only locally on nearby genes but also on distant elements that, when the DNA folds in space, come into spatial proximity. These regulatory processes are determined by modifications on the proteins packaged with the DNA, information that is covered in the track files we had mentioned. When scientists wish to modify the expression of a gene, for instance, due to its association with a disease, they can identify the associated regulator. With the aid of Hi-C maps, researchers are capable of detecting an interaction in a one-to-one analysis. However, if a gene therapy by editing the regulator is proposed, it is important to note that it could alter other genes, potentially causing unwanted side effects. Hence, it is necessary to investigate other interactions. The analysis of the 3D spatial environment facilitates inspection beyond one-to-one contacts.

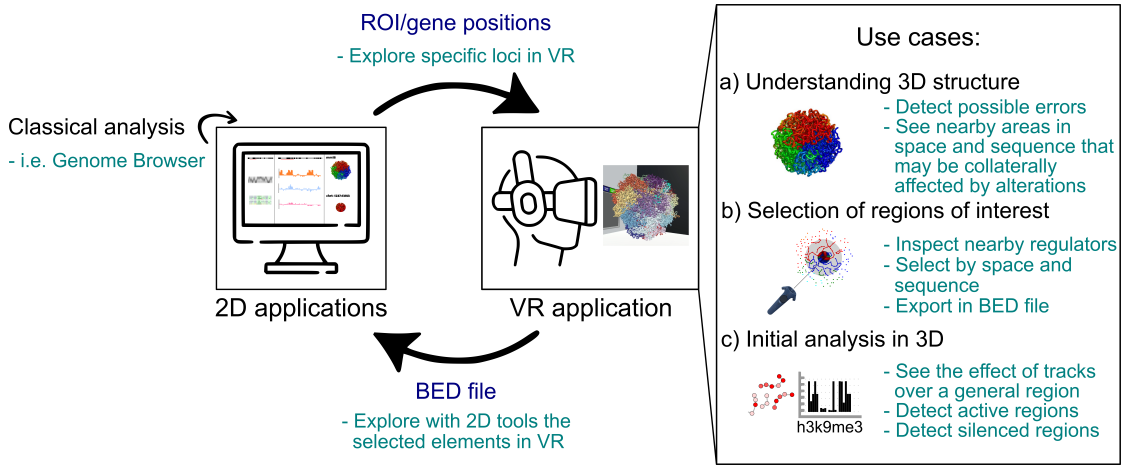

Fig. 2. Combination of classical desktop applications with our proposed VR tool. Our tool can be used before starting the work with standard applications, to understand the 3D structure (1), selection of regions of interest (2), or performing an initial analysis using the tracks of the model (3).

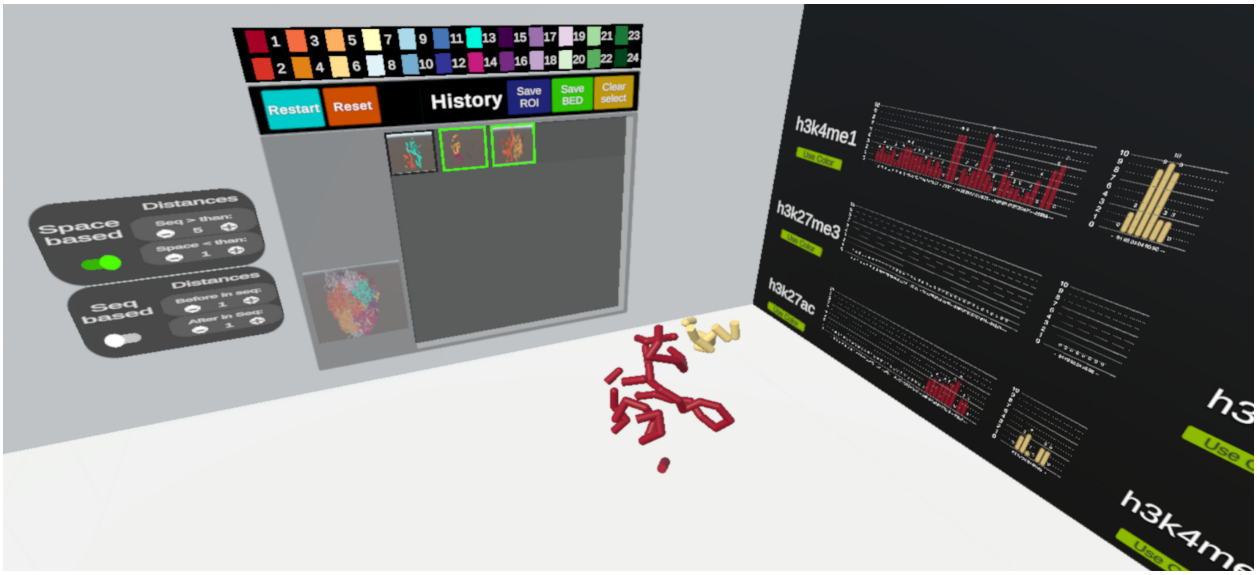

Fig. 3. Snapshot of our virtual environment: Control panel on the left wall, visualization wall on the right, and the 3D model (ROI) in the center.

However, understanding the 3D space using 2D projections is difficult. Thus, when discerning potential modifications is crucial, it is essential to employ a method that provides clear insight. For proteins, whose 3D structure is crucial for comprehending their function and designing novel drugs, VR has proven advantageous, enhancing spatial comprehension and facilitating collaborative environments. We therefore approached two domain experts to determine the requirements for a VR application for chromatin.

One of them, our reference domain expert, has more than 25 years of experience with genome data, and around 15 years working specifically with chromatin models. He directs several international consortia that study chromatin folding.

Chromatin study relies on extensive 2D data that inform about genes, expression levels of regulatory factors, and more. Our proposed methodology involves a combination of a VR tool and conventional desktop tools. The user would use VR for specific moments that require structural visualization, and desktop tools for accessing 2D data like gene databases. Thus, the VR tool should initially focus on enabling 3D exploration of chromatin and facilitating data transfer between 3D and 2D environments. The input for 2D environments typically comes from files indicating positions to analyze, such as BED files. The input for the VR environment could be either a specific position to be loaded onto the scene, or a free exploration of the model, allowing the user to select the region to explore.

Based on our findings, we developed a minimum set of features for our solution:

R1: Loading and displaying 3D chromatin data in real-time: This need is obvious, but rendering such complex models is not simple in VR.

R2: Identifying a concrete segment of the chromatin data to be displayed: When experts analyze a chromatin model, they typically have an initial idea about the segment they want to inspect. This access could be achieved through a spatial search using 3D exploration, or through a semantic search by entering the name or position of the gene.

R3: Brushing, filtering, and selection of regions of interest: For

a proper exploration, we need to provide several tools that facilitate the brushing and filtering of the data. Moreover, loading and saving of ROI selections is essential.

R4: Exporting segments to BED files [37], a standard format for the continued data exchange with the existing data analysis tools, forms a part of their normal workflow.

R5: Displaying epigenetic modifications charts for an initial analysis, whith which the domain experts spend a lot of time when searching for their objects of interest.

R6: Rendering of multiple information tracks simultaneously, shown over the chromatin model, to allow users to understand how those are represented along the chromosome. Upon selection of the proper chart, this information should be overlaid on the elements.

The resulting workflow should thus combine both desktop applications and the use of our VR tool, as illustrated in Fig. 2 and inspired by previous Immersive Visualization work [47, 48].

3.2. Application Scenarios

In collaboration with the two experts, we also identified the following three cases that would represent common scenarios to be addressed with our approach and tool.

Use case 1: Understanding 3D structure. Having access to a tool that enables experts to explore a 3D model in a manner that is comparable to that of the real world facilitates a more comfortable perception of the characteristics of that structure. According to our collaborators, this kind of visualization explains the territories of the chromosomes and helps them to identify which chromosome is next to another one. Even though the overall organization of the genome is generally consistent across genomes, there are notable exceptions, such as in the case of the eyes of mammals, where the silenced genes are located in the center and the rest in the periphery, which is in direct contrast to common genes. Hence, the capability to translate and rotate the model in a virtual environment facilitates its exploration and comprehension, owing to the stereoscopy.

Another utility of the 3D exploration, besides the general understanding of the model, is the detection of potential errors in the 3D coordinate reconstruction model, or positions that need further inspection. We illustrate this issue in Fig. 4.

Use case 2: Selection of regions of interest. The 3D exploration has the potential to provide researchers with valuable insights that would otherwise be difficult to discover, and to uncover intriguing regions that warrant further investigation. During one of our meetings with one of the two domain experts, for instance, we presented a model to him, and he promptly identified a three-dimensional region in the space where three chromosomes were in close proximity. This was unexpected, and he immediately stated that this region deserved further exploration. This exploration would require the ability to isolate the specified region and conduct further selections of its interesting elements. In addition, the information on those selected elements should be made available to continue the analysis in other tools, for instance by creating a BED file.

In addition to this selection function, it would be useful to include a logging option that allows users to replicate exploration steps (or skip them and proceed to the current state of a previous session). The log can be used to provide both history and the current state, the former being useful for retracing steps and the latter yielding a session state akin to the experts' desktop tools.

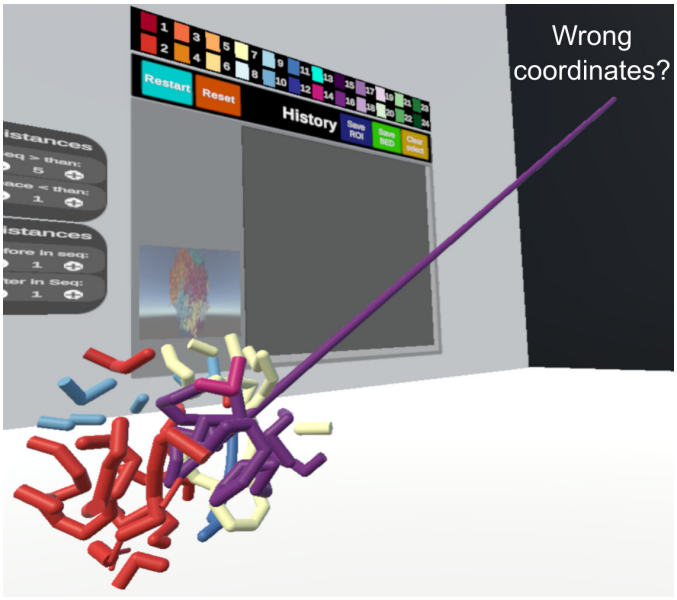

Fig. 4. A possible error (or position that needs further analysis) in the reconstructed coordinates detected thanks to the 3D view.

Use case 3: Initial analysis in 3D. While the goal of our tool is to serve as a companion to desktop software, the experts also found it interesting to add the possibility of inspecting the tracks together with the 3D model, since the tracks mark regions with certain behavior. For example, track `h3k4me3` encodes the trimethylation of lysine 4 on histone H3, which can indicate active genes, while track `h3k27me3` is a repressive mark associated with gene silencing. The domain experts thus suggested adding an option to overlay the tracks on the 3D model.

3.3. Design and Implementation

3.3.1. Design Objectives

Taking everything discussed in Sect. 3.1 into account, we designed our application as a virtual room, into which the 3D model of our chromatin can be loaded. The user can interact directly with the model or use the menus on the walls to perform various functions, as it is currently implemented in protein VR applications. We decided to use a virtual wall for the menus in order to avoid the need to manage them (make them appear and disappear, or reposition them) to the users. This reduces interaction complexity and makes users always aware of their position. We believe that this overcomes the disadvantage of them being partially occluded. We provide details about the virtual environment in Sect. 3.3.3. To address the experts' requirements, the available functions must include: selecting ROIs, saving the actual ROI, creating a BED file, maintaining a session history, loading tracks, and displaying track values on the model. Furthermore, we decided to incorporate additional functions such as the ability to restart the session to facilitate more precise selections within the ROIs of potentially interesting elements and to

modify the selection criteria. We provide a detailed explanation of all these functions in Sect. 3.3.4.

3.3.2. Technologies and Data

Based on these requirements, we developed a prototype application using Unity and based on the HTC Vive HMD (Head-Mounted Display) and its controllers. We created the charts using the Graph And Chart Unity asset [1]. We used a system with the following characteristics: Intel(R) Core(TM) i7-6700 CPU @ 3.40GHz 3.40 GHz, 32GB RAM, and an NVIDIA GeForce GTX 1070 with 24GB.

We derived our chromatin model from the work of Stevens et al. [41]: the chromatin of a mouse embryonic stem cell at a resolution of 100kb per bin (i.e., 100,000 DNA base pairs in each bin), reconstructed from Hi-C data. We represent it using spheres, each sphere corresponding to a bin at the reconstructed positions, and we use cylinders to show their connections, from here on referred to as bonds. This model has 25,709 spheres and 25,690 cylinders, 43.6M of triangles; without any enhancement it is thus unfeasible to achieve real-time rendering. As a result, we decided to test alternatives and chose to instantiate each object and apply the corresponding color through *MaterialPropertyBlockSetting* in Unity.

3.3.3. Overview of the Tool

To describe our tool and its functions, we begin by describing the virtual environment in which we work. The application creates a virtual scenario where we load the inspected object and render it in the center. We also create two virtual static walls, one for the control panel and other interaction elements, and another for the visualization of the charts, as shown in Fig. 3. The rationale behind this design is to facilitate the interaction. An alternative to a fixed region for the menus and other interaction widgets would be to make them appear on-demand after user input and let the user configure their virtual position in the scene. However, this would require the user to often reposition the elements to avoid overlap with the 3D model. In our case, both the control panel and the visualization wall require a lot of space. The control panel, in addition to the input widgets for controlling selection parameters, also provides the history log with snapshots of the regions of interest. The visualization component on the right also loads several charts at the same time. Some particular feature (definition of a subsegment of the model, given in chromosome number and positions, and toggling tracks) is currently implemented to be initiated from the program's console but could be moved to the UI if required. The main controls work as follows.

Control panel. Our control panel is always accessible from one of the walls of the virtual environment. This design allows for easy access to its features without requiring them to learn another system. Furthermore, it has been demonstrated that spatial menus are preferable to body-referenced menus, and they are less physically and mentally demanding than other designs [26]. In the future, we will examine shortcuts to common tasks to improve the user experience. With the control panel (Fig. 5), positioned on the left wall, we can restart the session or reposition the structure. We can save regions and maintain a history to reload them at a later time, delete selected elements, and change the method used to calculate distances for highlighting neighborhoods.

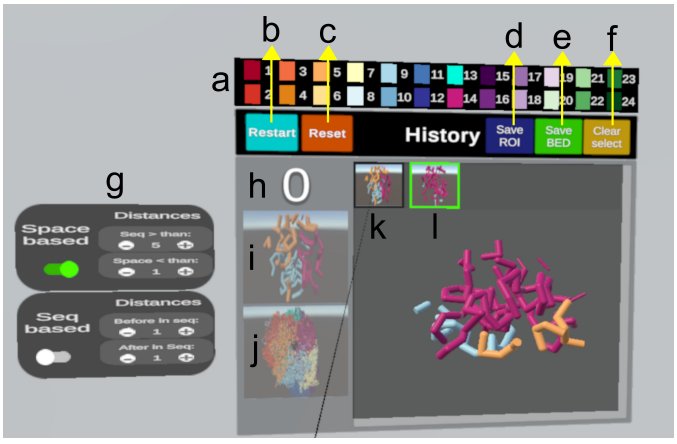

Fig. 5. Control panel parts. a) Legend of chromosome colors. b) "Restart": Reload the structure at the start position. c) "Reset": reset session, delete saved ROIs. d) "Save ROI": save ROI in history with a snapshot. e) "Save BED": save ROI and create a BED file. f) "Clear select": delete current selections. g) Distance method menu. h) ROI number displayed when its snapshot is hovered. i) Augmented snapshot when hovered. j) Snapshot of the whole structure; black sphere represents the hovered ROI's position. k) Snapshot of a saved ROI. l) Snapshot of a saved ROI plus BED file.

Visualization wall. This component hosts the tracks—represented as bar charts—to take advantage of the way experts normally visualize track data in tools such as Nucleome Browser. These tracks encode the DNA position on the x-axis and the signal intensity on the y-axis. The right wall consequently displays the tracks upon request. Users can also interactively modify the model colors based on the signal value or brush the bars to select the corresponding spheres, as we explain below.

3.3.4. Functions

We use two controllers for the interaction, the left one for repositioning the structure and interacting with the UI, and the right one for the filtering and ROI selection tasks as we show in Fig. 6. We designed this mapping since we expect the users to spend more time on detailed filtering and ROI selection operations than navigating through the menus.

Basic inspection. Users can explore the 3D scene by moving in the physical space or moving the structure using the *grip* button of the left controller. The UI also has buttons to restart

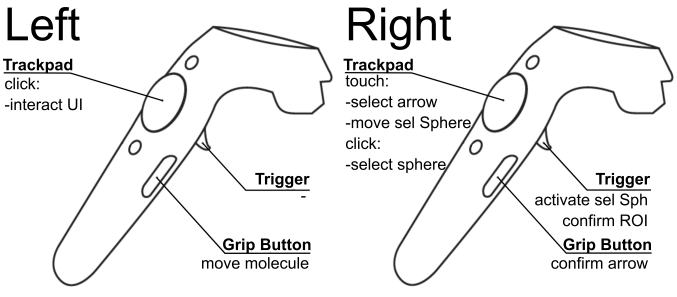

Fig. 6. Mapping from the controller's buttons to functions. Controller illustration from [46], used under CC BY 4.0 International license.

the initial state or to reset the session, deleting the saved ROIs (Fig. 5). We facilitate a detailed exploration of the structure via filtering and ROI selection procedures, as we explain next.

ROI selection. To facilitate the selection of regions of interest, we provide a cutaway tool using a selection sphere, similar to the filtering approach by Le Muzic *et al.* [25] for mesoscopic biological models. This conveys focus and context of our location or the surrounding environment. This sphere is activated with the trigger button on the right controller. In the scene, we display a semi-transparent sphere centered on the ray coming from the right controller. We can move it closer or farther away using the touchpad on the controller. If we touch the bottom region, the sphere moves along the ray towards the controller, whereas if we touch the upper region, it moves away from the controller.

To prevent occlusion and clutter, we clip elements that are located outside the sphere and between the sphere and the user. For the remaining elements, we display them according to two thresholds, namely the sphere radius and the distance limit. We display the elements within the sphere in their original size. Elements that exceed the established distance limit are depicted as spheres devoid of any bonds. This technique preserves the surrounding context when highlighting focus regions, which may prove crucial for some tasks [32]. In this case our context is the whole chromatin model. For the objects between those thresholds, we calculate the width of the sphere and the bond based on the distance, and we make both elements smaller as they get farther away from the viewer. We show an example in the center of Fig. 1 and in step 1 of Fig. 10.

Once we have identified a region of interest, we can confirm it with the same *trigger* button, and the region within the selection sphere is the only one displayed, as we show in Fig. 1. This aspect is one of the components that addresses **R3**.

Detailed selection. With a selected ROI, we can proceed to interact with the spheres (bins) in the model. We distinguish between hovered spheres (black spheres), potential interesting neighbors (red spheres), neighbors accessible for selection (red spheres with arrows), and selected spheres (yellow spheres), see Fig. 10. We explain these concepts later in this section. We implemented an exploration method to facilitate navigation between neighboring elements using the touchpad. The design of the colored spheres and the exploration method with arrows are based on the approach by Molina and Vázquez [33]. In their user study, they found that users responded well to this method of selecting small elements in scenes with occlusion when working with proteins. In their case, they employed two techniques based on arrows or colored rings. Here, we decided to combine them, as explained later, to take advantage of the benefits of both methods.

When we hover over a sphere with the right controller, we display it in black, as well as potential interesting neighbors as red spheres—those who meet the active distance criteria, which we explain in the following section.

If we want to fix these black and red spheres to further explore, we need to click on the correspondent hovered sphere. This way, they will remain even if we move the controller around.

Furthermore, we display colored arrows over certain red spheres, which indicate the neighbors that can be selected. The arrows provide the necessary visual feedback to choose which elements to select. The direction of the arrow corresponds to the direction on the controller, and the color of the arrow ensures that, even when we turn our heads or move the model, we can identify the desired direction. It is based on the following steps:

- The algorithm scans the list of neighbors to find the up to eight closest ones to the black sphere.
- To provide visual feedback, we project the candidates onto a virtual plane that is perpendicular to the axis and extends from the viewer to the center of the black sphere.
- We split the virtual plane into distinct sectors and assign the spheres to the sector in which they land. If multiple spheres project to a single sector, we keep the one closest to the camera.

This approach produces a set of up to eight candidate neighbors with visual cues that indicate how to move to them. Users can navigate through this set pushing the touchpad in the direction of the arrow. When a particular direction is indicated, the arrow turns white, indicating its selection. One can confirm the selection by pressing the *grip* button on the right controller at that moment, and we turn the corresponding sphere yellow. To select the center black sphere, it is possible to press the *grip* button without specifying a direction. In steps 3–7 of Fig. 10 we show the evolution of this process. Moreover, we equipped the controller with a guide that enables us to observe the arrows and the selected direction, as shown in steps 4–6 of Fig. 10.

Knowing spatially neighboring elements can help an expert to understand the effects that occur on gene expression. For instance, modifications in proteins can either inhibit or promote expression, as we discussed with tracks. In addition, specific DNA regions play a role in regulating nearby elements. For instance, promoters are DNA sequences initiating RNA synthesis, and enhancers activate promoters' activity. Conversely, other sequences such as silencers have the opposite effect.

Filtering. It is useful to allow domain experts to filter data both in 3D space, based on distance, and the sequence. This interaction completes the detailed selection required for **R3**. The first one is spatial-based, defining a sphere and radius to explore adjacent neighbors in space while discarding those close in the sequence. This allows for focusing on complex 3D spatial interactions between those elements, irrespective of the chromosome 2D organization. The second method enables the selection of elements based on the sequence. This facilitates the focus on elements from the same chromosome, thereby providing visual cues on how the sequence is organized in space. These approaches are depicted in Fig. 7.

Space-based filtering: With this method, on the one hand, we set a spatial limit, based on a sphere of the indicated radius, and hide elements that are farther than this distance. On the other hand, we have another limit in the sequence of the selected element: we hide elements that are closer in sequence than the indicated limit, both in front or behind.

Sequence-based filtering: In this case, we are only interested in elements from the same chromosome as the selected sphere. We can specify the number of bins preceding and following the selected bin in the sequence, and display this range.

Saving ROIs, creating BED files. Although this was not a requirement, we decided to incorporate the option to save

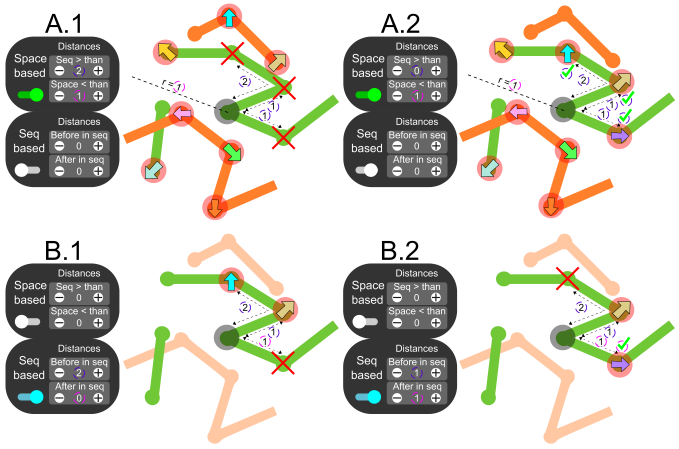

Fig. 7. Representation of the effects of changing filtering approach (A to B) and changing some parameters of each approach (1 to 2). A.1. Shows the filtering of the neighborhood based on the *space-based* method. Elements further away than 2 in the sequence and inside a sphere or radius 1 are considered. A.2. Shows the effect of changing some parameters with this method, in this case, we want to include all the bins in the same sequence. B.1. Uses the same position as before, and filters the model using the *sequence-based* method. It only considers elements in the same chromosome, in this case, the two previous elements in the sequence and none after. B.2. Shows the effect of changing some parameters with this method, in this case, we want to include only the bins right before and after the selected one. Appendix E provides an analogous representation of these variations as they would appear in the application.

analysis states to facilitate the experts' work by keeping a record of the desired steps. We provide users with a means to store the ROIs defined for a model and re-load them later, via the *Save ROI* button in the control panel. For each saving action, we also generate a representative image of the ROI in the history section of the menu (Fig. 5). We decided to include a snapshot of the ROI to facilitate subsequent identification when the user wishes to return to that point later in the session.

Domain experts were also interested in getting selected data transferred to their other desktop analysis tools using BED format [37], as described in **R4**. We implemented this method, which can be accessed through the *Save BED* option in the control panel. Similarly, we generate a representative image with a green border along with the BED file. These files store, in each row, the chromosome number, initial position, final position, and additional information. This additional information could, for example, include any yellow spheres we might have, the corresponding lines in the BED file would include a third column stating "selected". An example can be seen in step 9 of Fig. 10.

The history section can also be inspected. If the user hovers over one of the images in the history section, we display the number of that ROI on the left and indicate the name of the possible BED file, an enlarged image of that ROI, and an image showing the position of that ROI in the structure with a black sphere (Fig. 5). A click on one of the snapshots loads the saved configuration, as a means to keep track of the current state and to facilitate a certain degree of undo operations.

Tracks, selection, and color on model. The study of chromatin is often complemented with information from tracks: data on epigenetic modifications along the sequence (e.g., methylation and acetylation), that influence gene expression. This information is typically analyzed using 2D charts (i.e., bar charts) in

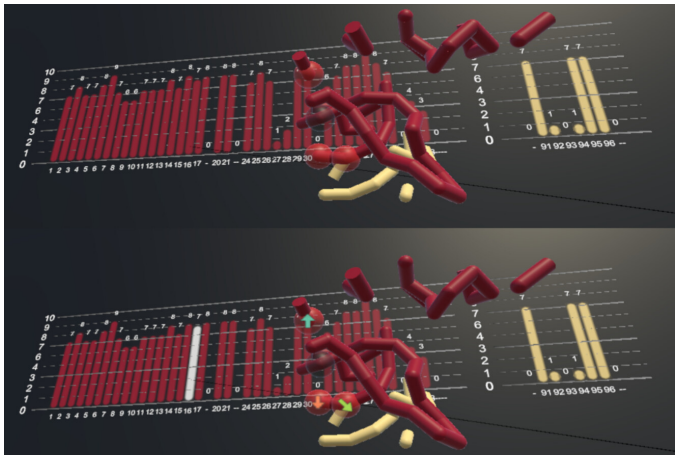

Fig. 8. We can hover over (up) or click (bottom) on the bars to interact with the spheres.

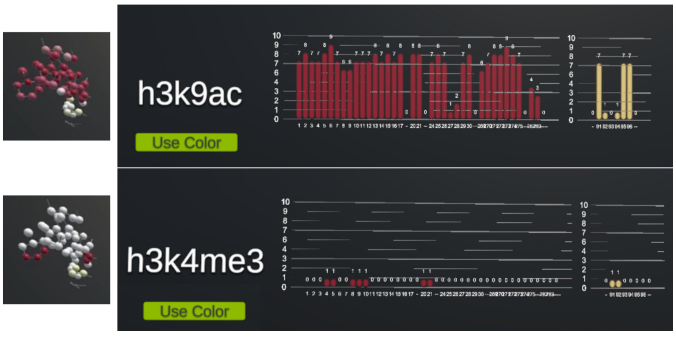

Fig. 9. We can overlay track information over the 3D structure to communicate this information by changing the color with the value of each position. We represent the highest value as a sphere with full color and a value of 0 in white. We show bonds gray and thin to avoid distraction.

desktop applications. In our tool, users can activate the display of these charts with information corresponding to the selected ROI through the console. For our experiment we worked with simulated information of the tracks `h3k4me1`, `h3k4me3`, `h3k4ac`, `h3k27ac`, and `h3k27me3`; for a specific region. In general, the number of tracks of a chromosome can be in the order of hundreds, but the ones we list are the most common scientists use, and we thus fulfill the basic needs of **R5**. The ENCODE project [7] provides such information for many models.

We display these charts on the left wall, and we can interact with them to perform different tasks: Hovering over or clicking on one of the bars has the same effect as doing it on the corresponding sphere (Fig. 8). Coloring the model based on the value of the track in each position, as required by **R6** (Fig. 9).

3.3.5. Example of a Possible Workflow

In Fig. 10 we depict a potential study that could be conducted using our application, from the selection of an ROI to the generation of a BED file.

4. User Study

To evaluate the feasibility and utility of our approach, we first showcased it to one of the experts with whom we discussed the requirements, who confirmed that the tool fulfills the original

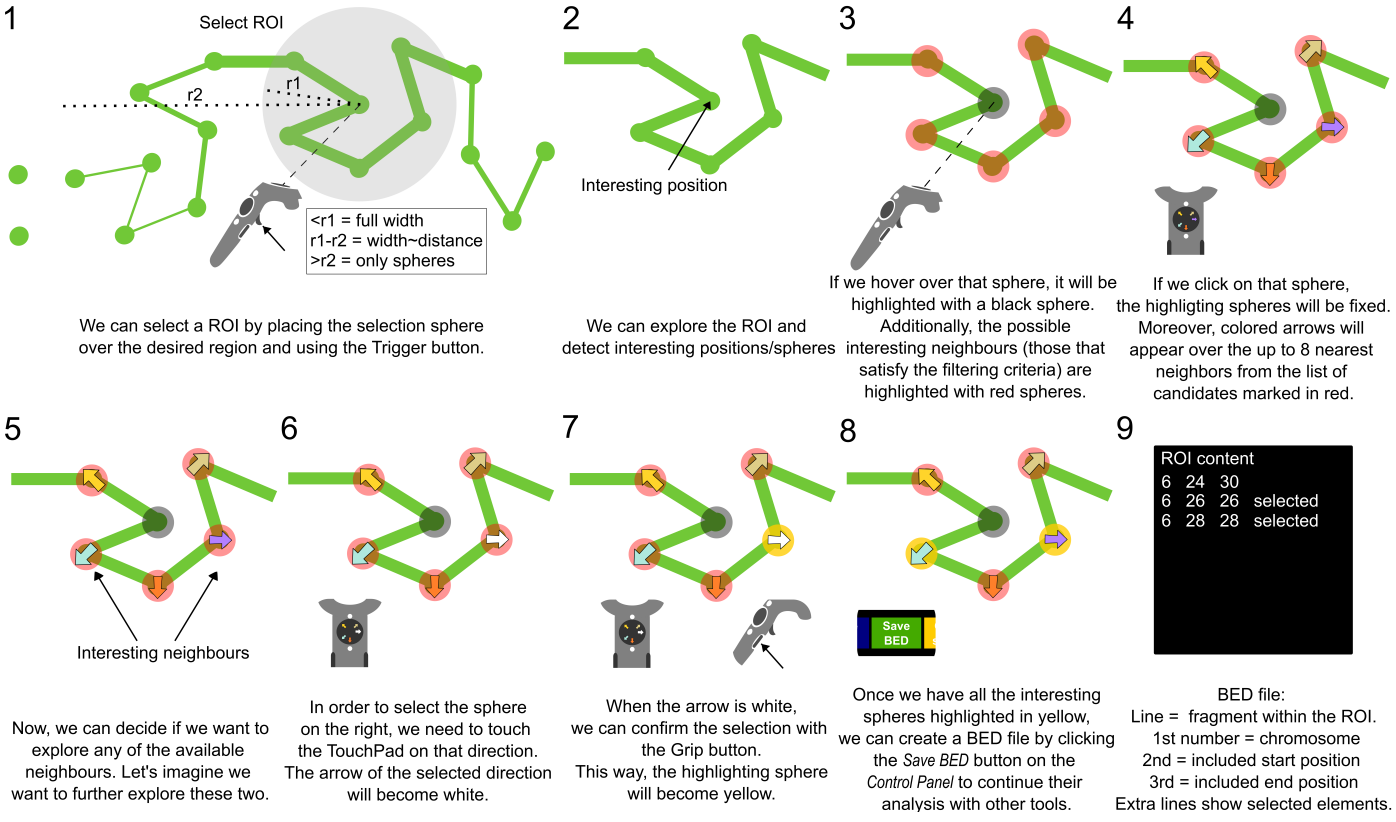

Fig. 10. Example of a possible workflow, from selecting an ROI to exporting a BED file. A sample of how these steps would appear in the application can be found in Appendix E.

requirements and expectations. We then designed a user study to test the applicability of the approach on additional domain experts who did not have any previous experience with our tool.

4.1. Participants

For the study, we asked five unpaid experts to participate. The first acted as a pilot tester for the study with all the defined tasks, before we tested the remaining participants. P1 is working toward a Ph.D. degree and has experience with chromatin for less than five years. His expertise is in informatics, and he had no previous experience with VR devices. Among the other four experts (2 female), one is working toward a Master's degree, two of them are working toward a Ph.D. degree, and one already holds a Ph.D. degree. Their experience with chromatin data varied from less than five years to almost ten. All of them had good or excellent vision and no color deficiency. Their expertise was either in biology (2×), informatics (1×), or both (1×). All of them had only briefly been exposed to or had no previous experience with VR devices. They all use computers daily for their work, and only two of them sometimes play video games.

4.2. Procedure

We conducted both the pilot study and the regular experiments individually, based on the following procedure:

- *Initial introduction:* We provided the expert with the introduction to the experiment, highlighting its objectives and the significance of their participation.
- *Explanation of tasks:* We provided the expert with a general introduction to the tool and the types of tasks to be completed.
- *Consent questionnaire and demographic information:* Basic questionnaires were filled in to ensure informed and ethical participation. More information can be found in Sect. 4.4.
- *Video of application's functions and tutorial:* To provide the experts with identical instructions, we initially presented a video demonstrating the utilization of the application. Afterwards, they performed a step-by-step tutorial to familiarize themselves with the interface.
- *Tasks:* Finally, we asked the expert to carry out specific tasks within the application, utilizing the presented functionality. We describe these tasks in Sect. 4.3 and Appendix A.
- *Final questionnaire:* The post-questionnaire aimed to gather feedback on the overall user experience, the challenges faced, and suggestions for improvement. We provide details on the questionnaires and the results in Sect. 4.4. In addition, we collected spoken comments made by the experts during the sessions. We summarize relevant results in Sect. 4.5.
- *End of the experiment:* We thanked the participants for their time and effort and concluded the session.

The whole experiment lasted ≈ one hour, only the pilot study took ≈ 90 minutes since we discussed both the features and the interactions in detail. The pilot study revealed the need to refine the approach to the tutorial phase, as the controls were difficult to learn. Thus, we extended the time to explore them. Another change was that we asked the participants to follow a think-aloud procedure so that we could better understand their thoughts.

4.3. Tasks

During the experiment, we requested the experts to perform a series of tasks that use the functionality of our tool, which we detail in Appendix A. In summary, we asked the participants for three types of tasks. The first type deals with the manipulation of the structure in the 3D space. It aimed to test their ability to orient themselves within the chromatin structure. The second set of tasks was testing the ability to interact with the individual parts of the chromatin fiber (e.g., their selection). The final set of tasks checked the use of tracks.

4.4. Questionnaires and Interviews

As we noted, we administered two questionnaires during the session, one for consent at the beginning and another for application feedback at the end. In the first one, we asked about gender, visual acuity, color blindness, level of education, area of expertise (such as Biology, Computer Science, or other), prior experience with VR, frequency of using this technology, computer, and video game usage, years of experience with chromatin, and their consent with participating at the experiment.

We divided the second questionnaire into several parts, which we detail in Appendix B. We asked participants to evaluate different aspects of the application regarding the requested tasks. Then, we asked them to evaluate, on a scale from 1 to 10, the mental, temporal, and physical demand, performance, effort, and frustration of the experiment. Finally, we asked the participants to provide us with an overall evaluation of the application, ranging from 1 to 10, indicating their belief that the utilization of this system or an enhanced version of it would enable them to accomplish tasks that cannot be accomplished through any other means. Furthermore, we asked them to specify the tasks and their rationales for accomplishing them and to indicate any supplementary features they would like to have, as well as their willingness to employ the method in their workflow. They were also able to provide any additional remarks they desired.

4.5. Insights

In addition to the questionnaire remarks, we also collected comments they made during the demonstration sessions and the user study, following the *think aloud* protocol. We thus realized that, during those sessions, some participants found relevant insights on the models, that demonstrated the potential utility of the tool. Our reference domain expert, for instance, found a 3D region where three chromosomes seemed to converge as a relevant point that deserved further investigation. Another expert noticed that VR allowed them to notice where some chromosomes were in close proximity much faster than with other tools. This statement is consistent with the demonstrated findings of both [18] and [19], which verified the advantages of VR over desktop applications in tasks requiring detailed spatial awareness.

In addition, our participants revealed several tasks that are required for their workflows, which can be completed using our tool, but cannot be reached with their current tools. These are:

- seeing multiple interactions because Hi-C matrices are only based on pairwise interactions,
- multi-way contacts examination (more than two regions that are in proximity),
- representation of the specific locus in three dimensions,
- identifying regions part of specific environments that are difficult to visualize in 2D, such as multi-way contacts,
- helping to picture the 3D structure of the genome in a way that 2D methods simply cannot achieve,
- assessment tool to demonstrate hypotheses: when exploring data in 2D, experts commonly assume a certain behavior that could be proved by showing this in 3D, and
- provide a general view of the chromatin structure at the genome-wide level: currently, there are no tools that allow experts to have such a genome-wide perspective.

Through our conversations with the experts, we also gathered a set of additional recommendations. Many of these suggestions (details in Appendix D) relate to the UI and its additional features, such as a help menu or a different position of the color legend. Experts would also appreciate the integration of more context information about the displayed structure and data captured by different methods (ATAC, ChIP Seq, etc.). Several ideas, such as the selection of bins above a certain threshold for each of the tracks or seeing the color of two tracks at the same time, led to further adjustments of our design decisions, which we plan to incorporate in future extensions of the tool.

5. Discussion

We now discuss the results of the experiments and the answers to the questionnaires, which we show in Appendix C.

The first notable comment is the fact that, despite the complexity of the controls, especially with a very short training phase, the experience of the participants was quite positive. All the questions related to the interaction (Q1 through Q11) were valued positively by most of the participants. This means that they were confident in the use of the interactions and that the features would suit their workflow (colored questions). Q6 was an exception. This question refers to the selection of individual elements. One participant stated that this option is unnecessary. The other users, though, found it useful.

All participants were enthusiastic and unanimously found our approach to be useful for their work. They verbalized several encouraging comments, such as: “amazing experience, very fun to use, and would love to see this more widespread” or “I liked it a lot, and I believe it is very useful to have a genome-wide perspective of what is happening in the nucleus.” We informed the participants that we intended the tool to assist them in their research workflow, and not as an educational tool. We asked them whether they would incorporate it into their daily work, and all of them responded positively. Nonetheless, they had several constructive suggestions that we we had listed already in the previous section.

Except for the drawbacks mentioned by the participants, we are also aware of other limitations of the current version of our system. The most important one is the learning curve. In its current state, the learning process is longer than the time we dedicated to this in the tutorial. Several participants mentioned that the interface is too complex. For a more extended study, users need to practice with the tool for a longer period than they did for our experiment. This could be solved by compiling a

more detailed tutorial and by adding several extra guided tasks to the experiment.

Another possible limitation of the current state of the application is that we work with a 100Kb resolution model, which was deemed appropriate by all experts. If we wanted to display the model with higher resolution or models with many more elements, however, the render speed would need to be improved.

6. Conclusions and Future Work

In this paper, we presented our design study for the visualization of chromatin data in a VR setup. Our approach, designed in collaboration with domain experts, can complement their current workflow with an extra inspection step that can be used as an initial exploratory task, but also for fast identification and selection of regions of interest. Since we allow the users to export BED files from the selected parts, the output can be seamlessly integrated with other analysis tools such as the Nucleome Browser. As noted by the experts in our user study, our VR approach opens new ways for them to make use of data analysis, such as the exploration of multi-way contacts or studying multiple interactions within the chromatin fiber.

As such, our contribution is the exploration of the possibilities of using an immersive approach to study chromatin data, and the discussion of the different design choices we made. The solutions we decided to implement were assessed by experts who appreciated such tools in addition to their established desktop-based toolkits.

Our current implementation also has some limitations. In particular, in our current design, several functions are not accessible from the virtual environment and are executed from the program's console. We made this decision to maintain better control over the conditions of this experiment due to the numerous functions already handled by the user in the virtual environment. However, in the future, we will consider adding those directly to the control panel. This will require a careful redesign of the interface, as the user cannot be overwhelmed by the controls in the virtual environment. Moreover, our initial distribution of functions in the controllers served us to verify the usability and interest in the proposed functions. However, we want to enhance the user experience redesigning it and extending it to platforms with a single controller or a joystick instead of a touchpad. Similarly, we have the goal to expand the track data available to the user. Furthermore, one of the experts suggested adding a value filter to the tracks so that the 3D inspection would allow looking for neighboring elements with a specific range of values, which could serve to identify regulators or target genes. Currently, it is possible to load a specific region of the model by entering the chromosome number, positions, and radius. In the future, we want to add the option to load regions by entering the names of genes, thus avoiding the step of searching for their positions in a genome browser. Since this happens only once, this is more suitable for console input, due to the limitations of the textual input in VR. In our prototype, the optimization of geometry processing has not been the primary focus of this experiment. In the future we will explore new alternatives to further improve this process. On the other hand, we also aim to enhance the design of the virtual world and the mapping of actions to make them more comfortable and functional for the user. This will also involve the insights that we gained from the experts in the experiment (see Sect. 4.5). Ultimately, we believe that our tool will be able to aid the experts in their daily workflow and give them additional insights that are cumbersome or even impossible to get using the currently available solutions.

Figure images license/copyright

7. Acknowledgements

This project has been supported by PID2021-122136OB-C21 from the Ministerio de Ciencia e Innovación, by 839 FEDER (EU) funds, and The Czech Science Foundation research project no. GA23-05651S. Elena Molina has been supported by an FPU grant from the Ministerio de Ciencia e Innovación.

References

[1] Graph And Chart Unity asset. `assetstore.unity.com/packages/tools/gui/graph-and-chart-78488`; 2017. Accessed: 2023-11-01.

[2] Asbury, TM, Mitman, M, Tang, J, Zheng, WJ. Genome3D: A viewer-model framework for integrating and visualizing multi-scale epigenomic information within a three-dimensional genome. BMC Bioinf 2010;11(1):444:1–444:7. doi:`10/fwpbr5`.

[3] Bannister, AJ, Kouzarides, T. Regulation of chromatin by histone modifications. Cell Res 2011;21(3):381–395. doi:`10/fd63nh`.

[4] Bennie, SJ, Ranaghan, KE, Deeks, H, Goldsmith, HE, O'Connor, MB, Mulholland, AJ, et al. Teaching enzyme catalysis using interactive molecular dynamics in virtual reality. J Chem Educ 2019;96(11):2488–2496. doi:`10/gf74zk`.

[5] Boltsis, I, Grosveld, F, Giraud, G, Kolovos, P. Chromatin conformation in development and disease. Front Cell Dev Biol 2021;9:723859:1–723859:27. doi:`10/gnvwqk`.

[6] Calvelo, M, Pineiro, A, Garcia-Fandino, R. An immersive journey to the molecular structure of SARS-CoV-2: Virtual reality in COVID-19. Comput Struct Biotechnol J 2020;18:2621–2628. doi:`10/k6ss`.

[7] Davis, CA, Davis, CA, Hitz, BC, Sloan, CA, Chan, ET, Davidson, JM, et al. The encyclopedia of DNA elements (ENCODE): Data portal update. Nucleic Acids Res 2018;46(D1):D794–D801. doi:`10/gg79v9`.

[8] Dekker, J, Belmont, AS, Guttman, M, Leshyk, VO, Lis, JT, Lomvardas, S, et al. The 4D nucleome project. Nature 2017;549(7671):219–226. doi:`10/gfkhp8`.

[9] Dekker, J, Rippe, K, Dekker, M, Kleckner, N. Capturing chromosome conformation. Science 2002;295(5558):1306–1311. doi:`10/bq9k7h`.

[10] Dixon, JR, Selvaraj, S, Yue, F, Kim, A, Li, Y, Shen, Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 2012;485(7398):376–380. doi:`10/f3zgv9`.

[11] Djekidel, MN, Wang, M, Zhang, MQ, Gao, J. HiC-3DViewer: A new tool to visualize Hi-C data in 3D space. Quant Biol 2017;5(2):183–190. doi:`10/k6sx`.

[12] Dostie, J, Richmond, TA, Arnaout, RA, Selzer, RR, Lee, WL, Honan, TA, et al. Chromosome conformation capture carbon copy (5C): A massively parallel solution for mapping interactions between genomic elements. Genome Res 2006;16(10):1299–1309. doi:`10/cs522j`.

[13] Durand, NC, Robinson, JT, Shamim, MS, Machol, I, Mesirov, JP, Lander, ES, et al. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell Syst 2016;3(1):99–101. doi:`10/gfkd4z`.

[14] Dwyer, T, Marriott, K, Isenberg, T, Klein, K, Riche, N, Schreiber, F, et al. Immersive analytics: An introduction. In: Immersive Analytics; chap. 1. Cham,: Springer; 2018, p. 1–23. doi:10/kt3b.

[15] Etemadpour, R, Monson, E, Linsen, L. The effect of stereoscopic immersive environments on projection-based multi-dimensional data visualization. In: Proc. IV. Los Alamitos: IEEE CS; 2013, p. 389–397. doi:10/gtss7v.

[16] Goodstadt, MN, Marti-Renom, MA. Communicating genome architecture: Biovisualization of the genome, from data analysis and hypothesis generation to communication and learning. J Mol Biol 2019;431(6):1071–1087. doi:10/gn9t2r.

[17] Goodstadt N., M, Marti-Renom, MA. Challenges for visualizing three-dimensional data in genomic browsers. FEBS Lett 2017;591(17):2505–2519. doi:10/gpf2k5.

[18] Hombeck, J, Meuschke, M, Lieb, S, Lichtenberg, N, Datta, R, Krone, M, et al. Distance visualizations for vascular structures in desktop and VR: Overview and implementation. In: Proc. VCBM. Goslar: EG Assoc; 2022, p. 1–10. doi:10/gtz7k3.

[19] Hombeck, J, Meuschke, M, Zyla, L, Heuser, AJ, Toader, J, Popp, F, et al. Evaluating perceptional tasks for medicine: A comparative user study between a virtual reality and a desktop application. In: Proc. VR. Los Alamitos: IEEE CS; 2022, p. 514–523. doi:10/gtz7k4.

[20] Imakaev, MV, Fudenberg, G, Mirny, LA. Modeling chromosomes: Beyond pretty pictures. FEBS Lett 2015;589(20PartA):3031–3036. doi:10/f7x8cf.

[21] Kent, WJ, Sugnet, CW, Furey, TS, Roskin, KM, Pringle, TH, Zahler, AM, et al. The human genome browser at UCSC. Genome Res 2002;12(6):996–1006. doi:10/fpf5rm.

[22] Kerpedjiev, P, Abdennur, N, Lekschas, F, McCallum, C, Dinkla, K, Strobelt, H, et al. HiGlass: Web-based visual exploration and analysis of genome interaction maps. Genome Biol 2018;19(1):125:1–125:12. doi:10/gd8chx.

[23] Kraus, M, Weiler, N, Oelke, D, Kehrer, J, Keim, DA, Fuchs, J. The impact of immersion on cluster identification tasks. IEEE Trans Vis Comput Graph 2020;26(1):525–535. doi:10/gtnn23.

[24] Kuťák, D, Vázquez, PP, Isenberg, T, Krone, M, Baaden, M, Byška, J, et al. State of the art of molecular visualization in immersive virtual environments. Comput Graph Forum 2023;42(6):e14738:1–e14738:29. doi:10/kt4n.

[25] Le Muzic, M, Mindek, P, Sorger, J, Autin, L, Goodsell, DS, Viola, I. Visibility equalizer cutaway visualization of mesoscopic biological models. Comput Graph Forum 2016;35(3):161–170. doi:10/f8wc84.

[26] Lediaeva, I, LaViola, J. Evaluation of body-referenced graphical menus in virtual environments. In: Proc. Graphics Interface. Mississauga: CIPS; 2020, p. 308–316. doi:10/gt5358.

[27] Li, D, Hsu, S, Purushotham, D, Sears, RL, Wang, T. WashU epigenome browser update 2019. Nucleic Acids Res 2019;47(W1):W158–W165. doi:10/ggfb56.

[28] Li, R, Liu, Y, Li, T, Li, C. 3Disease Browser: A Web server for integrating 3D genome and disease-associated chromosome rearrangement data. Sci Rep 2016;6(1):34651:1–34651:11. doi:10/f8739s.

[29] Lieberman-Aiden, E, van Berkum, NL, Williams, L, Imakaev, M, Ragoczy, T, Telling, A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 2009;326(5950):289–293. doi:10/fnh876.

[30] Marti-Renom, MA, Mirny, LA. Bridging the resolution gap in structural modeling of 3D genome organization. PLoS Comput Biol 2011;7(7):e1002125:1–e1002125:6. doi:10/c6nmbk.

[31] Martin, FJ, Amode, MR, Aneja, A, et al. Ensembl 2023. Nucleic Acids Res 2023;51(D1):D933–D941. doi:10/gq85sp.

[32] Meuschke, M, Voß, S, Gaidzik, F, Preim, B, Lawonn, K. Skyscraper visualization of multiple time-dependent scalar fields on surfaces. Comput Graph 2021;99:22–42. doi:10/gtz7k5.

[33] Molina, E, Vázquez, PP. Two-step techniques for accurate selection of small elements in VR environments. Graph Models 2023;128:101183:1–101183:14. doi:10/k6t8.

[34] Nowotny, J, Wells, A, Oluwadare, O, Xu, L, Cao, R, Trieu, T, et al. GMOL: An interactive tool for 3D genome structure visualization. Sci Rep 2016;6(1):20802:1–20802:8. doi:10/f79sr4.

[35] Nusrat, S, Harbig, T, Gehlenborg, N. Tasks, techniques, and tools for genomic data visualization. Comput Graph Forum 2019;38(3):781–805. doi:10/ggqfk5.

[36] Pettersen, EF, Goddard, TD, Huang, CC, Meng, EC, Couch, GS, Croll, TI, et al. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci 2021;30(1):70–82. doi:10/ghr6mn.

[37] Quinlan, AR, Hall, IM. BEDTools: A flexible suite of utilities for comparing genomic features. Bioinf 2010;26(6):841–842. doi:10/cmrms3.

[38] Rao, SS, Huntley, MH, Durand, NC, Stamenova, EK, Bochkov, ID, Robinson, J, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 2014;159(7):1665–1680. doi:10/xqj.

[39] Reiske, G, In, S, Yang, Y. Multi-focus querying of the human genome information on desktop and in virtual reality: An evaluation. arXiv preprint 2308.13487; 2023. doi:10/k6vz.

[40] Simonis, M, Klous, P, Splinter, E, Moshkin, Y, Willemsen, R, de Wit, E, et al. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture–on-chip (4C). Nat Genet 2006;38(11):1348–1354. doi:10/b2qp9s.

[41] Stevens, T, Lando, D, Basu, S, Atkinson, L, Cao, Y, Lee, S, et al. 3D structures of individual mammalian genomes studied by single cell Hi-C. Nature 2017;544(7648):59–64. doi:10/f9twqq.

[42] Tang, B, Li, F, Li, J, Zhao, W, Zhang, Z. Delta: A new web-based 3D genome visualization and analysis platform. Bioinf 2017;34(8):1409–1410. doi:10/gdcsd9.

[43] Tang, B, Li, X, Li, G, Tian, D, Li, F, Zhang, Z. Delta.AR: An augmented reality-based visualization platform for 3D genome. Innovation 2021;2(3):100149:1–100149:2. doi:10/k6vp.

[44] Todd, S, Todd, P, McGowan, SJ, Hughes, JR, Kakui, Y, Leymarie, FF, et al. CSynth: An interactive modelling and visualization tool for 3D chromatin structure. Bioinf 2020;37(7):951–955. doi:10/k6vt.

[45] Trieu, T, Oluwadare, O, Wopata, J, Cheng, J. GenomeFlow: A comprehensive graphical tool for modeling and analyzing 3D genome structure. Bioinf 2019;35(8):1416–1418. doi:10/gsth8v.

[46] Tsai, CF, Chen, CC, Wu, EHK, Chung, CR, Huang, CY, Tsai, PY, et al. A machine-learning-based assessment method for early-stage neurocognitive impairment by an immersive virtual supermarket. IEEE Trans Neural Syst Rehabil Eng 2021;29:2124–2132. doi:10/mt3k.

[47] Wang, X, Besançon, L, Guéniat, F, Sereno, M, Ammi, M, Isenberg, T. A vision of bringing immersive visualization to scientific workflows. In: Proc. CHI Workshop on Immersive Analytics. 2019,URL: https://hal.science/hal-02053969.

[48] Wang, X, Besançon, L, Rousseau, D, Sereno, M, Ammi, M, Isenberg, T. Towards an understanding of augmented reality extensions for existing 3D data analysis tools. In: Proc. CHI. New York: ACM; 2020, p. 528:1–528:13. doi:10/gm4jj8.

[49] Zhu, X, Zhang, Y, Wang, Y, Tian, D, Belmont, AS, Swedlow, JR, et al. Nucleome Browser: An integrative and multimodal data navigation platform for 4D nucleome. Nat Methods 2022;19:911–913. doi:10/k6vv.

Virtual reality inspection of chromatin 3D and 2D data

Appendix

In this appendix we provide additional material and details that we could not include in the main paper due to space limitations or because it was not essential for explaining our approach.

Appendix A. Tasks

The tasks were divided into three types of use cases: 3D exploration, model manipulation and generation of BED files, and the use of tracks. Some of the steps in these tasks can be seen represented in Figure A.11.

Regarding 3D exploration, the users had to solve the following tasks.

- Move the structure.
- Restart position.
- Answer the question: With which chromosomes does chromosome 1 adjoin?

The following model manipulation tasks were asked:

- Select the region where chromosomes 1, 13, and 15 adjoin.
- Hover over one sphere of chromosome 15 that has as its neighbors those neighbors that have at least one sphere of chromosome 1 and at least one sphere of chromosome 13.
- From here, select one sphere of each chromosome.
- Create a BED file for this ROI.
- Check the effect of decreasing two units of the distance in sequence in the Space-based method.
- Change to Sequence-based method. Select one sphere in chromosome 13. Check the effect of increasing one unit of the distance in the before-sequence option and one in the after-sequence option.

Finally, on the use of tracks, participants were asked to solve the following problems:

- Check how we can load specific locations with the example prepared.
- Check track information of this region.
- Color the model using `h3k9ac` data. Then with `h3k4me3`. Reset colors.
- Select from the `h3k27ac` chart the position with maximum value and two interesting neighbors.

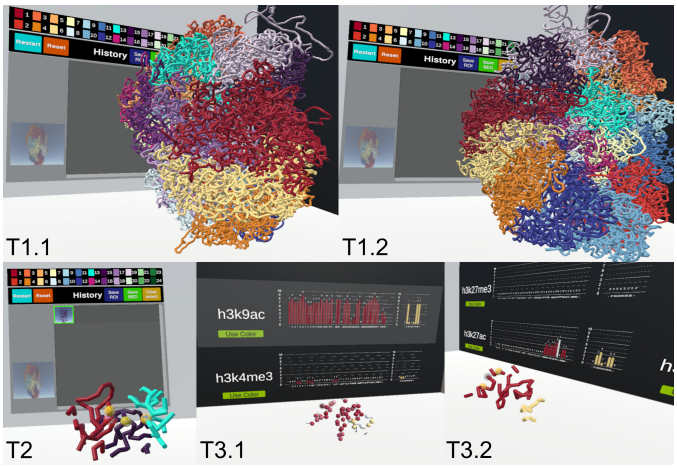

Fig. A.11. Snapshots of some of the steps in the tasks. T1.1 and T1.2 show two possible views to check with which chromosomes chromosome 1 adjoins. T2. One sphere from chromosome 15, one of its possible interesting neighbors from chromosome 1, and one from chromosome 13 selected. T3.1 Model colored using `h3k9ac` data. T3.2. Two interesting neighbors of the position with maximum value in the `h3k27ac` chart selected.

Appendix B. Questionnaires

There were two questionnaires for the users. The first one was for consent and demographic information about gender, visual acuity, color blindness, level of education, area of expertise (such as Biology, Computer Science, or other), prior experience with VR, frequency of using this technology, computer, and video game usage, years of experience with chromatin.

The second questionnaire was divided into several parts. One part asked to evaluate the following features on the Likert scale from 1 to 7 (7 meaning "Completely agree"). Each feature was examined through several questions:

- For the 3D exploration and interaction:
  - **Q1:** The method allows me to better understand the structure than using only 2D elements, such as heatmaps.
  - **Q2:** The method allows me to detect possible interactions between chromosomes or areas in a simpler way than using only 2D elements.
  - **Q3:** As a complement to my usual work methods, such as 2D charts, the method would facilitate my general understanding of the model.
- For the election of Regions of Interest (ROIs), selection of positions, and creation of BED files:
  - **Q4:** The method allows the investigation of interesting areas that would have been more difficult to detect with traditional methods, such as using only 2D data.
  - **Q5:** I would like to have this method as a complement to understanding ROIs that I identify with traditional 2D

methods.

Q6: For me, it is practical to be able to select individual positions using the arrows.

Q7: I would use it to select positions for further analysis.

Q8: This method would be useful to generate BED files to use in other programs to continue the analysis.

- For the use of tracks:

Q9: It is important to be able to view tracks in this environment to perform the analysis.

Q10: I prefer the tracks in this method rather than only in another program with the BED file that I have generated.

Q11: Seeing the model in 3D with the colors set based on the selected track makes it easier to detect interesting positions.

- For the whole application (questions start with "I think that"):

Easiness: The method enables me to do my analysis easily.

Comfort: The method is comfortable to use.

Learnability: It was easy to learn how to use this method.

Understandability: The method would quickly be understood by most people.

Then, we asked them to evaluate on a scale from 1 to 10 (from "Low" to "High") the whole application:

- *Mental Demand:* How mentally demanding was the use of this method?
- *Physical Demand:* How physically demanding was the use of this method?
- *Temporal Demand:* How hurried or rushed was the pace of the use of this method?
- *Performance:* How successful were you in accomplishing what you were asked to do?
- *Effort:* How hard did you have to work to accomplish your level of performance?
- *Frustration:* How insecure, discouraged, irritated, stressed, and annoyed were you?

Finally, the users had to give an overall score from 1 to 10, evaluating if the application would help them carry out a task that they would not be able to perform with their traditional tools. In addition, they were encouraged to suggest new features they would like to add to the application or any other remarks they considered.

Features: Scale 1 to 7					
Question	E1	E2	E3	E4	Avg
3D exploration					
Q1: understand	6	7	7	7	6.75
Q2: interactions	7	6	7	7	6.75
Q3: complement	6	6	6	6	6
ROIs, selection and BED					
Q4: ROIs	6	6	7	6	6.25
Q5: complement	6	6	6	5	5.75
Q6: individual	1	5	7	6	4.75
Q7: would use	4	5	6	7	5.5
Q8: BED	6	5	6	7	6
Tracks					
Q9: view tracks	6	7	7	7	6.75
Q10: tracks here	4	7	7	4	5.5
Q11: colors	6	6	7	6	6.25
General experience					
Easiness	2	6	5	5	4.5
Comfort	2	5	6	6	4.75
Learnability	3	5	6	5	4.75
Understandability	2	5	7	7	5.25

Usage effort: Scale 1 to 10					
Question	E1	E2	E3	E4	Avg
Mental Demand	8	3	4	4	4.75
Physical Demand	8	1	3	1	3.25
Temporal Demand	4	1	3	5	3.25
Performance	7	8	8	8	7.75
Effort	8	2	2	8	5
Frustration	3	2	1	1	1.25

Overall questions					
Question	E1	E2	E3	E4	Avg
Overall score	6	10	8	9	8.25
Would use	Y	Y	Y	Y	Y

Table C.1. Table with final questionnaire answers from the 4 experts and average. Questions in color are phrased as "Would you use. . .," so they refer to the application of a particular feature in the context of each expert's typical workflow.

Appendix C. Answers

In the next Table C.1, we can see the answers to the previously explained questionnaires.

Appendix D. Suggestions

The discussion allowed us to get a set of suggestions to extend or improve the tool. Some of these collected comments and suggestions were:

- "It would be nice to be able to see the BED file in the virtual environment."
- "I would want the chromosome's legend to be updated with the ROI and only see the ones present."
- "There should be a way to hide or show chromosomes from the legend."
- "I miss data about the model in the virtual environment, such as the resolution."
- "I would like to have the color legend in a place where it cannot be covered by the molecule."
- "More context info (chr start end in base pairs) in the 3D layout or in the virtual environment."
- "A little help menu that could pop up to remind you of what the buttons do on both hands."
- "It could be interesting to be able to see the color of 2 tracks at the same time."

- "I would definitely include genes within the selected regions. This will be super useful to see the genes that are in close proximity with other genes."
- "I would include ATAC profiles indicating chromatin accessibility with the tracks."
- "If we are able to integrate all the data from a desired experiment (ChIP, ATAC, genes, and Hi-C modeling) it will be super useful to have all the information in there."
- "It would be great if we could select the bins that are above a certain threshold for each of the tracks, in order to filter the regions out of our interest and for instance keep the tracks with high H3K27ac, which correspond to active enhancers. And then if we have the genes that are in close spatial proximity to this high H3K27ac regions, we can identify regulators or target genes."

Appendix E. Snapshots from the Application

Below, we present how some of the described scenarios would appear within the application.

In Fig. E.12 we show how changing the filtering modes and their parameters would be seen in the application.

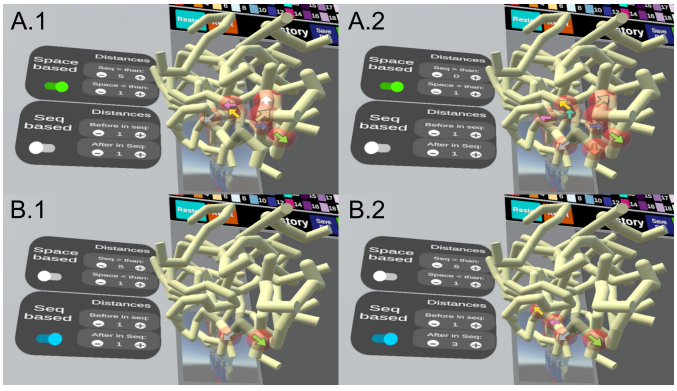

Fig. E.12. Effects of changing filtering approach (A to B) and changing some parameters of each approach (1 to 2) as seen in the application. A.1. Shows the filtering of the neighborhood based on the *space-based* method. A.2. Shows the effect of changing some parameters with this method, in this case, we want to include all the bins in the same sequence. B.1. Uses the same position as before, and filters the model using the *sequence-based* method. B.2. Shows the effect of changing some parameters with this method, in this case, we want to include the three following bins in the same sequence.

In Fig. E.13 we show how some usual steps during an analysis, see Fig. 10, would be seen.

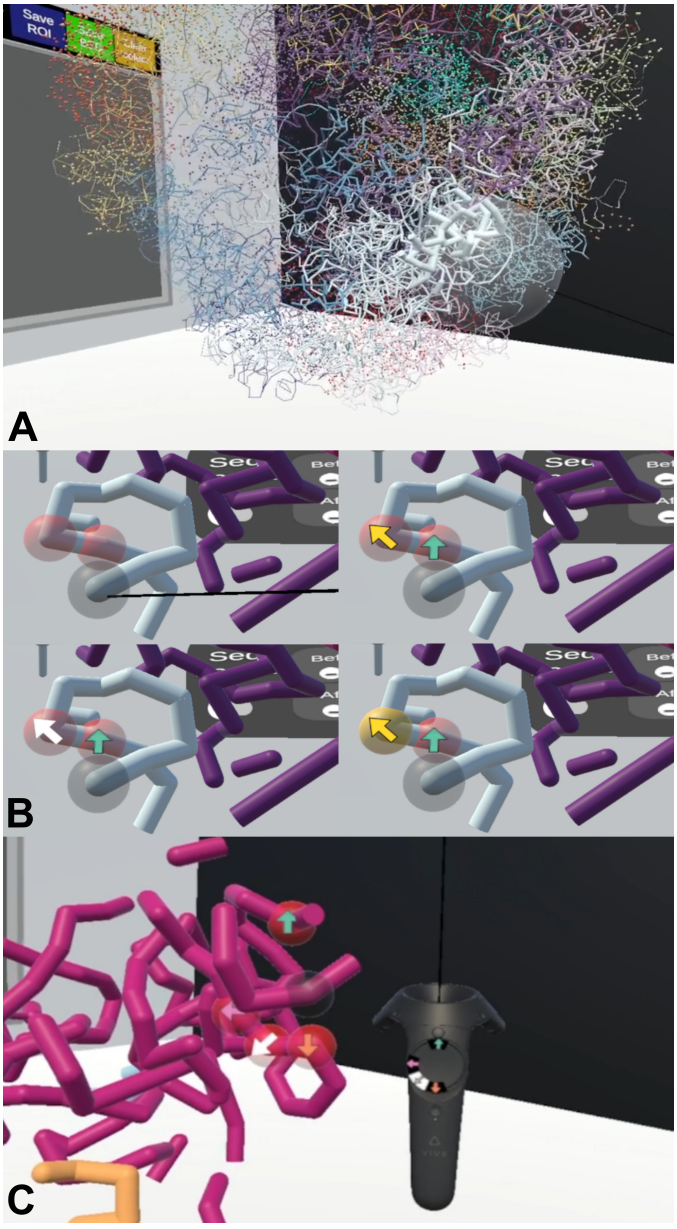

Fig. E.13. Examples of some steps from Fig. 10, as they would appear within the application. A. shows step 1. B. shows steps 3-7. C. shows how the arrows are shown also on the controller.

Figure Images License/Copyright