3DChromoTwist: Development of a 3D Chromosome Structure Reconstruction Game for Educational Purposes

Marcin Pawlukiewicz
University of Rhode Island
45 Upper College Rd
Kingston, RI 02881
mapawlukiewicz@uri.edu

Oluwatosin Oluwadare
University of Colorado Colorado Springs
1420 Austin Bluffs Pkwy
Colorado Springs, CO 80918
ooluwada@uccs.edu

Abstract
The 3D structure reconstruction of the chromosome is important so that a better knowledge of chromosome activity can aid us in understanding DNA replication, gene regulation, genome interaction, genome folding, and genome function. The High-throughput Chromosome Conformation Capture (Hi-C) technique incorporates chromosome conformation capture approach with the power of the Next Generation Sequencing technologies to study the 3D chromatin organization. From the Hi-C data, we can infer the contact or interaction frequency (IF) matrix which describes the level of interaction between the chromosome bins. In this work, we propose the development and design of a single-player bioinformatics game called 3DChromoTwist. The project’s objective is to interest non-scientists in learning about three-dimensional (3D) chromosomal structure. A reduced Hi-C contact matrix representing the final, complete structure is provided to players along with a 3D fragment of a chromosome. Players must then solve the puzzle by moving gene loci of 3D chromosomal structures until they form the desired relationships between the folds of the chromosome. We expect that 3DChromoTwist will positively influence the progression of science by introducing a new set of learners to the world of genetics. Ultimately, this will increase public scientific literacy and public engagement with science and technology.

Introduction
High throughput chromosome conformation capture (Hi-C) is a technique for evaluating chromatin’s spatial arrangement in a cell. The technique counts how many interactions there are between genome loci. The genome loci are detectable in a 3D structure when they are close together yet separated by many nucleotides in a linear genome. (Hakim and Misteli 2012).

The process involves crosslinking DNA such that adjoining areas are linked together. The regions can then be quickly identified. When cell genomes are cross-linked with formaldehyde, chromatin crosslinking begins. DNA fragments are then cut with a restriction enzyme, fragments are sealed with a biotin marker, and ligated. Chimeric DNA fragments are then created using reversed crosslinking. After DNA purification, streptavidin is used to pull down the DNA, after which the fragments are sequenced. Hi-C sequencing will generate sequenced pair reads that identify their nucleotides. These are then mapped to a reference genome and filtered for noise, ultimately preprocessed to produce an IF matrix (Sati and Cavalli 2017). Hi-C sequencing results in a comparison of multiple DNA fragments to each other. As a result, it scans the locus of the close pairs and maps their relations onto a N*N matrix, where N is the number of chromosome fragments. In a Hi-C experiment, each element in the matrix comprises a count of reading pairs that connect two homologous chromosome regions. As a result, the chromosomal contact matrix is symmetric and represents all observable interactions between the regions of a chromosome (Lieberman-Aiden et al. 2009). This data can be used in the 3D reconstruction of chromosomes. The reconstruction could be used to study DNA replication, gene regulation, genome interaction, genome folding, and genome function (Oluwadare, Zhang, and Cheng 2018). We can see that the number of relationships between chromosome regions is proportional to the distance in the 3D structure, therefore we can make predictions relying on that information (Lieberman-Aiden et al. 2009). Over the years, many 3D chromosome reconstruction algorithms have been developed. (Oluwadare, Highsmith, and Cheng 2019; MacKay and Kusalik 2020) give a comprehensive review of these methods and their strengths.

The way chromosomes fold helps us to understand data about the intricate connection between chromatin structure, quality movement, and the useful condition of the cell (Woodcock and Ghosh 2010). The genome must be packed properly into the nucleus. Otherwise, serious diseases, such as congenital malformations (ex. fewer or too many fingers) or cancer, may occur (Gorkin 2017). Knowledge of how genes act can help with the reconstruction of the entire genetic or biochemical pathways. This is essential to our understanding of metabolism, signal transduction, and other developmental or psychological processes (Dekker et al. 2002). Folding DNA into chromosomes is crucial as this is something that makes us, us. All the processes in our cell must occur correctly. This prevents tangling and damage during cell division. In the case of pregnancy, if the embryo misses a chromosome, it means that there is some ge-
necic damage, and this can prevent a successful pregnancy (Xie et al. 2021). Folding DNA into chromosomes makes the genetic material fit inside a cell but allows for distant interactions between them. This process is important as this regulates gene activities (Banigan et al. 2020).

Chromosomes consist of tightly packed DNA, wound up in order to condense a vast amount of genetic information into a smaller volume. DNA consists of a double helix, which is then wrapped around histone proteins to create nucleosomes. From there, nucleosomes are coiled into chromatin fiber, then looped to condense the chromatin into the final form of a fully assembled chromosome. Our game takes place at the point nucleosome creation, the point at when the relationships between segments of DNA can be analyzed and used to predict what the final chromosomal structure will be. This is what is calculated then given to solve to players via chromosome segments based on the Hi-C matrices (Alberts et al. 2002).

Chromosomes are part of every nucleus in every cell of our body. Visualization of the folding process is crucial for deeper understanding of the above addressed problems. The game helps not only to understand the processes in our body but also teaches the fundamental knowledge necessary in every biological field. Thus, the game influences the progression of science by equipping non-experts with biological knowledge, engaging them through game-play to contribute to the world of genetics.

The very first game was presented over fifty years ago (Ford 2012). Since then, many new games have appeared. Today, we can count them in millions. Both scientific and non-scientific games played have the element of entertainment. Without this, players will not be interested in the game. Moreover, the game should also be informative by presenting the player with some useful knowledge. Thus, making a game that is simultaneously fun and educational is important.

We believe that 3D structures are an informative and easily understood way to appeal to non-scientific audiences by giving reasonably accurate visualizations. We use current methods and algorithms to make the game as much realistic as possible to enable the possibility for new learners to make new discoveries and contribute to the scientific world.

This paper describes 3DChromoTwist, a game created to educate about chromosomal structure through the use of Hi-C data. The game is directed toward a broad audience and thus our development is adjusted as we expect people from non-scientific backgrounds to participate. The design and easy level are simplistic and, as we proceed to the hard level, more complex. The user can start from any level and move to more advanced levels at later stages as appropriate skills have been developed. Both design and movement of the structures need not be over complicated to reach the desired audience.

**Related Work**

Our inspiration is the popular online multiplayer game “Fold it” (Cooper et al. 2010). The game follows an educational approach and falls in the field of biology. In the game, players learn how protein folding mechanisms work. The design is simplified for non-expert players however, the game has many features. We noticed that the game garnered a lot of popularity over the years and many papers have been published regarding it. As we noticed the huge positive impact of this game, we decided to implement similar solutions into chromosomal structures.

Our work differs by applying the concepts of structure formation to 3D chromosome structure. As opposed to folding parts of a protein to prevent viral attacks within the structure of the protein, chromosome segments are moved around to reconstruct the natural positioning of the chromosomal components into their correct place, allowing the complete chromosome structure and shape to be revealed to the player once a level is successfully solved. This also played using a given Hi-C matrix, a unique component when compared to “FoldIt.” Moreover, there is no current related work in chromosome structure prediction. Thus, this will pioneer a new research direction towards this goal.

**Implementation**

<table>
<thead>
<tr>
<th>Key</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>Hide/ Show the menu.</td>
</tr>
<tr>
<td>L</td>
<td>Surrender.</td>
</tr>
<tr>
<td>O</td>
<td>Hide/ Show the outline of the objects.</td>
</tr>
<tr>
<td>Q</td>
<td>Reset the bin choice.</td>
</tr>
<tr>
<td>R</td>
<td>Hide/ Show the scoreboard (top 5 players).</td>
</tr>
<tr>
<td>+/-</td>
<td>Choose bin with index higher by one.</td>
</tr>
<tr>
<td>-</td>
<td>Choose bin with index lower by one.</td>
</tr>
<tr>
<td>0</td>
<td>Choose a bin with index increased by 10.</td>
</tr>
<tr>
<td>1-9</td>
<td>Pick index from 1 to 9 in the current teens.</td>
</tr>
<tr>
<td>Scroll</td>
<td>Zoom in and out.</td>
</tr>
<tr>
<td>ESC</td>
<td>Quit the game.</td>
</tr>
</tbody>
</table>

The game has been created as a 3D structure where the user can freely move around the objects to obtain a better understanding of the structure with the use of the following keys: W, S, A, and D. While camera motion is controlled with the use of the mouse, holding the keys moves the player...
A dark green color indicates that the correlation between the bins is set up correctly.
A light green, yellow, orange, and red color indicates how close the bins are to their correct positions, from close to far, respectively.

Apart from all this, the player can see a simplified version of the Hi-C data matrix. The data for the chromosome structure has been used from the following paper, "HSA: integrating multi-track Hi-C data for genome-scale reconstruction of 3D chromatin structure" (Zou, Zhang, and Ouyang 2016). We took a hundred points x, y, and z that are loaded onto the level as every game begins. Next, depending on the choice of difficulty, the player receives a random consecutive number of bins that are taken from the loaded data. Each bin is generated in a randomized position. To ease the difficulty of the game, the x coordinates have been randomized with some precision to the original spot. The player can see the Hi-C matrix with the colors (figure 1) defined as presented in table 2. This is an estimation that shows the distance between the current location and the correct final position of the bins. Using Euclidean distance, the program checks the correctness of the correlation between the bins as the player progresses through the level. The colors on the matrix progress towards red as the position of the bins moves away from the correct position.

The distance is measured by the Euclidean distance formula. Every center of the loci (x, y, z) is compared with every other object.

<table>
<thead>
<tr>
<th>Name</th>
<th>Distance</th>
<th>Color</th>
</tr>
</thead>
<tbody>
<tr>
<td>Completed</td>
<td>The correct position of the object with allowed margin of error up to 2 units.</td>
<td>Dark green</td>
</tr>
<tr>
<td>Very Close</td>
<td>Between 2 and 4 units.</td>
<td>Green</td>
</tr>
<tr>
<td>Close</td>
<td>Between 4 and 8 units.</td>
<td>Yellow</td>
</tr>
<tr>
<td>Far</td>
<td>Between 8 and 15 units.</td>
<td>Orange</td>
</tr>
<tr>
<td>Very Far</td>
<td>Above 15 units.</td>
<td>Red</td>
</tr>
</tbody>
</table>

The distance is measured by the Euclidean distance formula. Every center of the loci (x, y, z) is compared with every other object.

The player can pick a bin using the keyboard. The numbers 1 to 9 can pick a bin from the current teens, while 0 increases the bin by 10. The following keys “-” and “=” decrease and increase the current bin by one accordingly. The active bin can be noticed by a color change to red (figure 4). Although the bins have been indexed properly with the order of the x-coordinate, the player needs to figure out the index on its own while playing.

There are two important components in the game that are responsible for motivation and engagement. The score is the first factor. The player gains points as the matrix becomes better and there is a less red and more dark green color on it. The starting matrix is set up as a score of zero. Any deterioration of the matrix will encourage the player make up for this if the player wants to gain points. The scoring algorithm...
Figure 4: Active bin marked on red.

has been adjusted correspondingly, so no matter which matrix the user starts out with, it still achieves the same number of points. Another component is time. The player has a limited amount of time to complete the game. Depending on the level the player chooses, the time increases with difficulty. If the player does not make it on time, the game will end (figure 5). The score will be saved and displayed along with a “Game Over” message. However, if the player is able to complete the structure before time runs out, the score will be increased with bonus points. Depending on the time, the score will be adjusted accordingly. The player will be prompted with a “Victory” message. The player also has the option to surrender. In this instance, there are no bonus points added and the current score will be saved. To complete the level, the player needs to reposition loci according to clues received from the simplified Hi-C data matrix. The matrix is interactive and changes colors as the player changes distance. As loci get closer to their proper destination, the color changes to dark green. However, as the player gets further from the proper final destination, the color changes to red. Once the matrix is fully dark green, the level is complete (figure 6).

Figure 5: End of the game with the solution, level hard.

To make the game more competitive, each level has its own high score board. The top 5 players are always displayed with their scores. The player can show and hide the scoreboard at any point during the game. Once the game is completed, the scoreboard is displayed. Each level of difficulty has its own scoreboard. All the scores are saved into a text file that is secured by a number generated depending on the strings, results, and characters. Therefore, any attempts to change the records manually will result in resetting the file. Similarly, a change in coordinates in the text file will have similar results.

Upon completion of the game, the player, besides the “Game Over” or “Victory” message, will be allowed to see the solution by pressing “ENTER.” Once the key is pressed, the player will see the level’s complete structure as well as the solution next to it (figure 6). The solution represents the entire chromosome structure built out of 100 loci. However, because the player was required only to play with between 7 and 25 loci, all the loci used in the current game and chosen by the randomizing algorithm will be highlighted in red. The solution post-view is the last step in the game (figure 7). The player can choose to close the game by pressing “ESC” on the keyboard and to start the game again, the player is required to rerun the jar file.

Figure 6: End of the game with the solution, level normal.

We expect that the players, with the help of the given solutions in the form of the simplified Hi-C matrices, will be able to solve the puzzle. Naturally, replication of the same
The chromosome is extremely difficult, as there are many possibilities involved in its 3D structure. Hence, the acceptable margin of error is equal to 2 units.

**Conclusion**

The 3DChromoTwist game is playable by anyone using any type of personal computer thanks to the design and simplicity of the 3D software used to develop it. The fact that the results of the 3D chromosomal structure prediction are unknown poses the biggest design challenge. Even the most sophisticated structures offered by the game are uncertain, but the game nevertheless directs the player to them.

The game is very intuitive from the very first level to make it accessible as the player progresses. By making 3DChromoTwist available to users, we hope to demonstrate a fresh method for teaching chromosomal structure prediction and foster a deeper comprehension of the topic. Moreover, the game allows the visualization of these structures in 3-Dimension, giving a realistic perspective of the chromosome behavior. The game is a great complement to today’s textbooks since it makes it easier to understand and picture the structure in real time.

The goal of the game is to gradually increase the level of difficulty while yet maintaining a pleasurable experience for the player. The players’ problem-solving abilities may be enhanced by 3DChromoTwist, which can also assist in the resolution of actual scientific conundrums.

We intend to keep working on 3DChromoTwist. As we believe the game can advance research by introducing a new group of students to the field of genetics, we aim to improve the game design and add new levels throughout time. In the end, this will elevate public interest in science and technology as well as public scientific literacy. If even a fraction of the effort that goes into playing computer games can be directed toward scientific research, we personally believe that scientific advancement is achievable.

**Acknowledgement**

The work reported in this paper is supported by the National Science Foundation under Grant No. 2050919. Any opinions, findings and conclusions or recommendations expressed in this work are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

**References**


