# Spatially Integrated Abstraction of Genetic Molecules

Sarkis Halladjian

université
PARIS-SACLAY

# Spatially Integrated Abstraction of Genetic Molecules

**Thèse de doctorat de l'Université Paris-Saclay**

École doctorale n° 580, Sciences et Technologies de l'Information et de la Communication (STIC)
Spécialité de doctorat: Informatique
Unité de recherche: Université Paris-Saclay, CNRS, Inria, LRI, 91400 Orsay, France
Référent: Faculté des sciences d'Orsay

**Thèse présentée et soutenue à Gif-sur-Yvette, le 14 décembre 2020, par**

**Sarkis HALLADJIAN**

**Composition du jury:**

| | |
|---|---|
| **Alain Denise**<br>Professeur, Université Paris-Saclay, LRI, I2BC | Président |
| **Marc Baaden**<br>Directeur de recherche,<br>Laboratoire de Biochimie Théorique, CNRS | Rapporteur, Examinateur |
| **Michael Krone**<br>Professeur associé, University of Tübingen | Rapporteur, Examinateur |
| **Barbora Kozlíková**<br>Professeur associée, Masaryk University | Examinatrice |
| **Tobias Isenberg**<br>Directeur de recherche,<br>Université Paris-Saclay, CNRS, Inria, LRI | Directeur |
| **Ivan Viola**<br>Professeur associé,<br>King Abdullah University of Science and Technology | Codirecteur |

Thèse de doctorat

NNT: 2020UPASG056

# SPATIALLY INTEGRATED ABSTRACTION OF GENETIC MOLECULES

SARKIS HALLADJIAN

Sarkis Halladjian
*Spatially Integrated Abstraction of Genetic Molecules,*

SUPERVISORS:
Tobias Isenberg
Ivan Viola

Dedicated to the memory of Komitas,
savior of Armenian music,
born in the year the DNA was first discovered.

1869 – 1935

## ABSTRACT

The human genome consists mainly of DNA, a macromolecule consisting of a long linear sequence of bases, tightly packed to fit in the relatively small nucleus. The packing gives rise to multiple hierarchical organizational levels. Recent research has shown that, along with the linear sequence, the spatial arrangement of the genome plays an important role in the genome's function and activity. The visualization of both linear and spatial aspects of genome data is therefore necessary.

In this thesis, we focus on the concept of continuous visual abstraction for multiscale data, applied to the visualization of the human genome. Visual abstraction is a concept inspired by illustrations that makes the job of visual processing simpler, by guiding the attention of the viewer to important aspects.

We first extract characteristics of multiscale data and makes a parallel comparison between genome and astronomical data. The existing differences create the need for different approaches. A common point however is the need for continuous transitions that helps viewers grasp the relationships and relative size differences between scales.

To satisfy the conditions posed by the two aspects of the multiscale genome data, we present two conceptual frameworks, based on the same data.

The first framework, ScaleTrotter, represents the spatial structure of the genome, on all available levels. It gives users the freedom to travel from the nucleus of a cell to the atoms of the bases, passing through the different organizational levels of the genome. To make the exploration of the structure of all levels possible, smooth temporal transitions are used. Even though all the scales are not simultaneously visible, the temporal transition used superimposes two representations of the same element at consecutive scales emphasizing their relationship. To ensure the understandability and interactivity of the data, unnecessary parts of the data are abstracted away with the use of a scale-dependent camera.

The second framework, Multiscale Unfolding, focuses on aspects that are not visible in ScaleTrotter: the linear sequence and a simultaneous overview of all the organizational levels. The data is straightened to unfold the packing that occurs on several levels in a way that conserves the connectivity between the elements. To represent all the available levels, we use smooth spatial transitions between the levels. These spatial transitions are based on the same concept of the temporal transitions of the previous framework, superimposing scales and emphasizing on their relationship and size difference. We introduce

an interaction technique called Multiscale Zliding that allows the exploration of the data and further emphasizes the size differences between the levels.

In each framework, one of either linear of spatial aspect of genome data is sacrificed to emphasize the other. The thesis concludes with a discussion about the possibility of combining the two frameworks, minimizing the sacrifices to explore the two equally important aspects of the genome. In this thesis, we take a step closer to fully understanding the activity of the genome.

# RÉSUMÉ

Le génome humain est principalement constitué d'ADN, une macromolécule constituée d'une longue séquence linéaire de bases, étroitement serrée pour s'insérer dans le noyau relativement petit. L'empaquetage donne lieu à de multiples niveaux hiérarchiques d'organisation. Des recherches récentes ont montré que, parallèlement à la séquence linéaire, l'agencement spatial du génome joue un rôle important dans la fonction et l'activité du génome. La visualisation des aspects linéaires et spatiaux des données du génome est donc nécessaire.

Dans cette thèse, nous nous concentrons sur le concept d'abstraction visuelle continue pour les données multi-échelles, appliqué à la visualisation du génome humain. L'abstraction visuelle est un concept inspiré par des illustrations qui simplifie le travail de traitement visuel, en guidant l'attention du spectateur vers les aspects importants.

Nous commençons par extraire les caractéristiques des données multi-échelles et faisons une comparaison parallèle entre le génome et les données astronomiques. Les défis à relever pour créer une visualisation interactive des données du génome sont fondamentalement différents à plusieurs égards de ceux de l'astronomie. Tout d'abord, les données du génome ont des niveaux d'échelle entrelacés - l'ADN est une molécule extrêmement longue et connectée qui se manifeste à tous les niveaux d'échelle. Deuxièmement, les éléments de l'ADN ne disparaissent pas lorsqu'on zoome, mais les niveaux d'échelle auxquels ils sont observés regroupent ces éléments différemment. Troisièmement, nous disposons d'informations détaillées pour l'ensemble des données et pour tous les niveaux d'échelle, ce qui constitue un défi pour l'exploration visuelle interactive. Enfin, les niveaux d'échelle conceptuels des données sur le génome sont proches dans l'espace d'échelle. Les différences existantes créent le besoin d'approches différentes. Un point commun cependant est la nécessité de transitions continues qui aident les spectateurs à saisir les relations et les différences de taille relative entre les échelles.

Pour satisfaire aux conditions posées par les deux aspects des données génomiques multi-échelles, nous présentons deux cadres conceptuels, basés sur les mêmes données.

Le premier cadre, ScaleTrotter, représente la structure spatiale du génome, à tous les niveaux disponibles. Il donne à l'utilisateur la liberté de voyager du noyau d'une cellule aux atomes des bases, en passant par les différents niveaux d'organisation du génome. Pour rendre possible l'exploration de la structure de tous les niveaux, des transitions temporelles fluides sont utilisées. Même si toutes les échelles

ne sont pas visibles simultanément, la transition temporelle utilisée superpose deux représentations d'un même élément à des échelles consécutives, ce qui met en évidence leur relation. Pour garantir la compréhensibilité et l'interactivité des données, les parties inutiles des données sont extraites à l'aide d'une caméra dépendante de l'échelle.

Le deuxième cadre, Multiscale Unfolding, se concentre sur des aspects qui ne sont pas visibles dans ScaleTrotter : la séquence linéaire et une vue d'ensemble simultanée de tous les niveaux organisationnels. Les données sont redressées pour déplier l'empaquetage qui se produit à plusieurs niveaux de manière à conserver la connectivité entre les éléments. Pour représenter tous les niveaux disponibles, nous utilisons des transitions spatiales douces entre les niveaux. Ces transitions spatiales sont basées sur le même concept que les transitions temporelles du cadre précédent, en superposant les échelles et en mettant l'accent sur leur relation et leur différence de taille. Nous introduisons une technique d'interaction appelée Multiscale Zliding qui permet l'exploration des données et met davantage l'accent sur les différences de taille entre les niveaux. Dans l'ensemble, le dépliage multi-échelles permet aux spectateurs de saisir la composition structurelle de l'ADN, des chromosomes aux atomes, avec des niveaux croissants de dépliage, et peut être appliquée dans l'illustration et la communication basées sur des données ainsi qu'il peut servir de base de visualisation pour les experts du domaine.

Dans chaque cadre conceptuel, l'un des deux aspects linéaire ou spatial des données sur le génome est sacrifié pour mettre l'accent sur l'autre. La thèse se termine par une discussion sur la possibilité de combiner les deux cadres, en minimisant les sacrifices pour explorer les deux aspects du génome qui sont d'égale importance. Dans cette thèse, nous faisons un pas de plus vers la compréhension complète de l'activité du génome.

# PUBLICATIONS

This thesis uses material previously published or submitted for publication as papers. In particular, Chapter 3 is based on item 1 below and Chapter 4 is based on item 2 below.

**First-authored journal publications**

1. **Sarkis Halladjian**, Haichao Miao, David Kouřil, M. Eduard Gröller, Ivan Viola, and Tobias Isenberg. ScaleTrotter: Illustrative Visual Travels Across Negative Scales. *IEEE Transactions on Visualization and Computer Graphics, 2020, 26 (1), pp.654-664.*

2. **Sarkis Halladjian**, David Kouřil, Haichao Miao, M. Eduard Gröller, Ivan Viola, and Tobias Isenberg. Multiscale Unfolding: Illustratively Visualizing the Whole Genome at a Glance. *IEEE Transactions on Visualization and Computer Graphics, 2021.*

**Co-authored journal papers under submission**

3. David Kouřil, Ondřej Strnad, Peter Mindek, **Sarkis Halladjian**, Tobias Isenberg, M. Eduard Gröller, and Ivan Viola. Molecumentary: Producing Scalable Documentaries Using Molecular Visualization and Procedural Narration. *Submitted to: IEEE Transactions on Visualization and Computer Graphics.*

# ACKNOWLEDGMENTS

One thing I learned during my PhD thesis is that you cannot wish for better supervisors than Tobias Isenberg and Ivan Viola, who inspired me during the last three years and who surely became two sources of inspiration I will look for for the rest of my life. Tobias, thank you for being so patient and constantly motivating me. Ivan, thank you for always trying to push me for more. Thank you both for guiding me through the lows and the highs.

My next thanks go to my reviewers, Marc Baaden and Michael Krone. Thank you both for taking the time to read the manuscript and to write the reports. Thank you also to Alain Denise and Barbora Kozlíková for completing the jury. I really enjoyed taking part in the discussions that followed my presentation. If the defense had taken place in normal conditions, you would have gotten delicious food and drinks for your trouble. I owe you.

Special thanks to Meister Eduard Gröller, Haichao Miao and David Kouřil who, along with my two supervisors, constituted the main task force that I was part of during the whole duration of my thesis. Thank you boys, you never hesitated to help me no matter how many times I bothered you with questions. You started as colleagues and ended up as great friends.

I want to thank of course my colleagues at AVIZ, Jean-Daniel Fekete, Petra Isenberg, Pierre Dragicevic, Frédéric Vernier, Katia Evrat, Catherine Plaisant, Tanja Blascheck, Lonni Besançon, Paola Valdivia, Steve Haroz, Gaëlle Richer, Christoph Kinkeldey, Marie Destandau, Xiyao Wang, Natkamon Tovanich, Luiz Morais, Jiayi Hong, Lijie Yao, Mickael Sereno, Alaul Islam, for providing a special fun working environment that I am sure cannot be found anywhere else.

I also want to thank my colleagues at TU Wien who made my stays in Vienna, and the conferences I attended, much more enjoyable.

I now want to thank the people who have supported me for much longer than the duration of my PhD.

Պապ, Մամ, Մարինէ շատ կը սիրեմ ձեզի: Գիտեմ որքան զոհուած էք եւ դեռ կը զոհուիք ինծի համար: Երախտապարտ կը մնամ:

Thank you to all family members who helped me one way or another, specially Maral, Garo, Meghry and Anna.

Thank you to all my friends from before the time I traveled abroad, Najib, Roupina, Milia, Gudji, Marina, Haig, Asbed, Sevag, Varant, Chant, Arek, Nanor, Khajag, Vana and all those I am not mentioning. I miss you tremendously.

Finally, thank you to my friends in France, Anahit, Aram, Arax, Arsen, Berdj, Chantal, Chantig, Hagop, Hovig, Jean-Paul, Meghry, Nareg, Nairi, Naïri, Nency, Njteh, Pegor, Raffi, Sacha, Serge, Tsoline, Viken and everybody else. You were there to either help me focus on work (usually with coffee), or help me take my mind off work. Thank you for becoming my family abroad. Thank you for existing.

# CONTENTS

# LIST OF FIGURES

## LIST OF TABLES

# 1 INTRODUCTION

Human's thirst for a complete understanding of the laws of nature has led them to the exploration of the astronomically big and the infinitesimally small. During one of those exploration journeys 150 years ago, Friedrich Miescher discovered a substance who was later called DNA. The realization that DNA is the hereditary molecule carrying the genetic information came 75 years later and, since then, DNA research has never stopped. Even though the human genome was (almost) fully sequenced in the early 2000's, we are far from a complete understanding of its function and activity.

DNA is a long sequence of genetic code that would expand to 2 m if fully stretched, but somehow fits in the nuclei of human cells of 6 µm diameter. This is made possible because DNA is tightly packed, over and over, which creates several organizational levels. Recent techniques have not only shed light upon this organization, but also provided experts with more insight about the role of the organization of the genome in its function and activity.

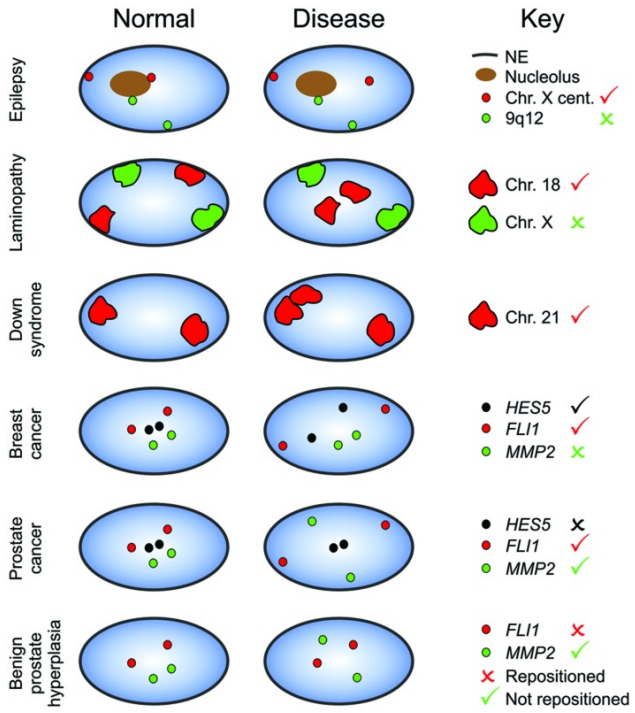



Figure 1: Reorganization of the genome in disease. Certain loci adopt alternative nuclear positions in disease (tick) compared to normal cells, whilst the positions of other loci are conserved in disease (cross). Figure from [88].

Even if two people share the exact same DNA sequence, variations in the structure of the DNA could result in differences. These differences could be harmless such as a difference in eye color, but could also result in more major health concerns such as cancer (Figure 1).

Therefore, experts need the appropriate tools to explore and analyze the gathered data about the 3D structure of the genome, just as patients need to be informed about the cause of their disease. Even high school students are thought about the organization of the genome. In fact, anyone may be curious enough to be interested about the structure of the carrier of genetic information.

## 1.1 SCIENTIFIC ILLUSTRATION

Traditionally, scientific illustrations are used to reflect the findings of science and technology [53]. Scientific illustration takes the viewer to the often unobservable, a category under which fall the human genome and the entire field of molecular biology (Figure 2).

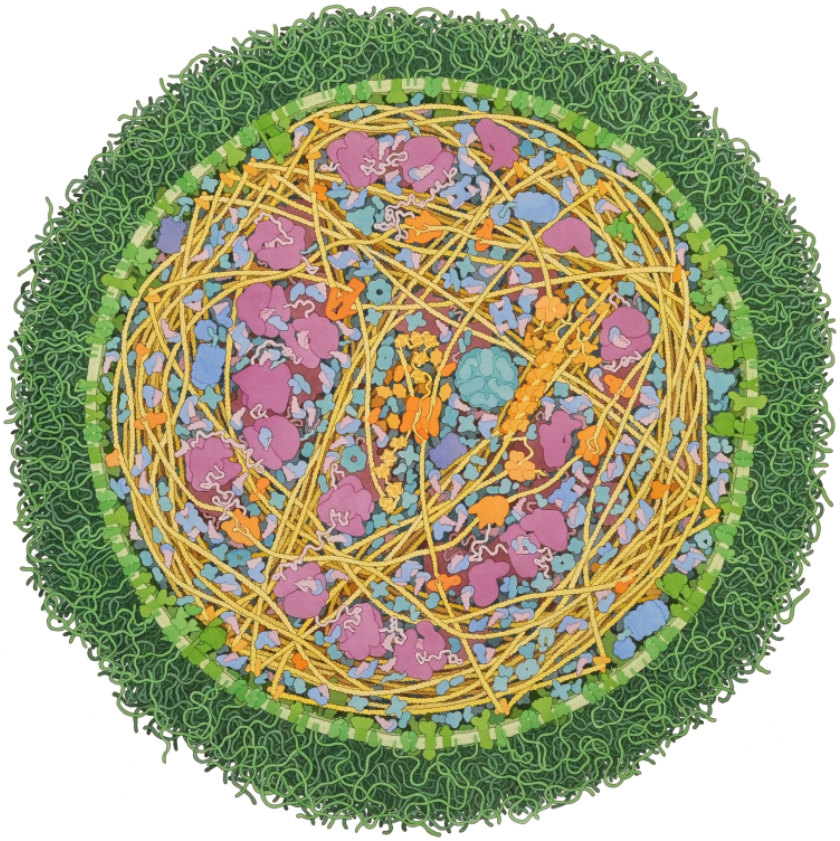


Figure 2: Illustration of an entire mycoplasma cell, about 300 nm in diameter, which is at the small end of the range of observed sizes. Illustration by David S. Goodsell, The Scripps Research Institute. doi: 10.2210/rcsb_pdb/goodsell-gallery-011.

The essence of scientific illustration is the communication of concepts that cannot be conveyed via words [53], because representing data visually amplifies people's cognition [18]. When trying to achieve this goal, scientific illustrators are allowed a certain expressive freedom to remove unnecessary detail and simplify shapes in order to guide the viewer's focus towards the important aspects that will convey the intended concepts. This process of transforming data into visual representations by removing elements non-vital to the goal of the illustration is called visual abstraction.

Traditional scientific illustrations however are typically not interactive and are not entirely dependent of the data. Naturally, they cannot satisfy our goal of exploring 3D models based on data gathered through continuously evolving techniques.

## 1.2 ILLUSTRATIVE VISUALIZATION

On the other hand, visualization tools are usually interactive and flexible enough to adapt to changes in the data they are representing. Even though advances in computer graphics technologies have facilitated the task of visualizing today's huge datasets (including genome data), the inherent capability of visualization to amplify human cognition is no longer enough to make sense of those datasets. Visualization therefore borrows abstraction techniques used in traditional illustrations. These visualizations are referred to as illustrative visualizations [109]. They apply visual abstraction on the data by removing unnecessary detail and keeping in the final visual representation the elements relevant to the purpose of the visualization.

For a visual abstraction to be meaningful, it has to make the job of visual processing simpler so that less of a cognitive processing needs to be invested, for a given purpose or goal, in comprehending the abstracted visual representation to understand the intended aspects of reality [126].

Our goal is to visualize genome data with its multiple organizational levels. The genome is made of the macromolecule DNA, which spans over several orders of magnitude in scale. Therefore, in an attempt to find out what abstractions are meaningful (in other words, what abstractions make the visual processing simpler), we look at two areas: molecular visualization and multiscale visualization, covering together the fundamental properties of DNA visualization.

### 1.2.1 *Molecular visualization*

A lot of previous work has investigated abstraction in the context of molecular visualization [24, 25, 76, 83, 125], even DNA visualization [84, 85, 89, 90]. These approaches use different types of abstractions,

some using level-of-detail techniques, others simplifying the structure of molecules or using different shading techniques.

A common theme that arises though is that most approaches apply seamless transitions between different abstraction stages, a notion that we take and apply in our work. However, the applied abstractions do not cover large scale differences.

### 1.2.2 *Multiscale visualization*

Therefore, we look at work covering multiscale data spanning over several orders of magnitude in spatial scale. A couple of visualization tools exist that represent the 3D multiscale structure of the genome [5, 98]. Contrary to the previous molecular abstractions, these tools do not use smooth transitions, but use sudden switches between the different scales.

Even though other fields such as cartography or astronomy use smooth transitions, we cannot use their techniques because of fundamental differences that exist between the data (we detail these differences in Chapter 3). However, we take inspiration from these fields to provide users with smooth and intuitive navigation in space and scale with smooth transitions.

We also noticed two different types of transitions between different representations: temporal transitions, where the user can transition between different representations but only see one at a time, and spatial transitions, where the user can see different representations applied on different parts of the data at once. We use each of these transition types in each of the frameworks we present.

## 1.3 THESIS STATEMENT AND OVERVIEW

In this thesis, we focus on the concept of continuous abstraction for multiscale data to visualize the human nuclear genome, starting from chromosomes all the way down to the atoms forming the nucleotides. In particular, we focus on the visualization of six scales of the DNA macromolecule and how to integrate the scales in one multiscale visualization while maintaining the relationships between them.

To pursue this goal, we review in Chapter 2 several areas related to this thesis, to contextualize it. We start by detailing the organizational levels of the human genome, and look at existing tools that domain experts use. We then look at techniques used in illustration and at previous attempts of understanding and controlling visual abstraction. Next, we review the use of controlled abstraction in visualization, particularly in molecular visualization. Finally, we review work that investigates the scale axis of the abstraction space, with a focus on transitions between scales and cross scale interactions.

To visualize the 3D genome organization with its multiscale structural units, we describe in Chapter 3 a conceptual framework for an interactive multiscale visualization of genome data. The framework is called ScaleTrotter, because it allows viewers to smoothly transition from the nucleus of a cell to the atomistic composition of the DNA. Because of the complexity of the 3D organization and the several orders of magnitude separating the two extreme scales, the coexistence of the different scales in one scene (while preserving the 3D structure) is ineffective. We therefore make use of temporally-controlled smooth transitions by deploying a scale-dependent camera model, that reduces visual clutter and facilitates interactive exploration (Figure 3). We start the chapter with an analysis of the unique requirements of multiscale representations of genome data.

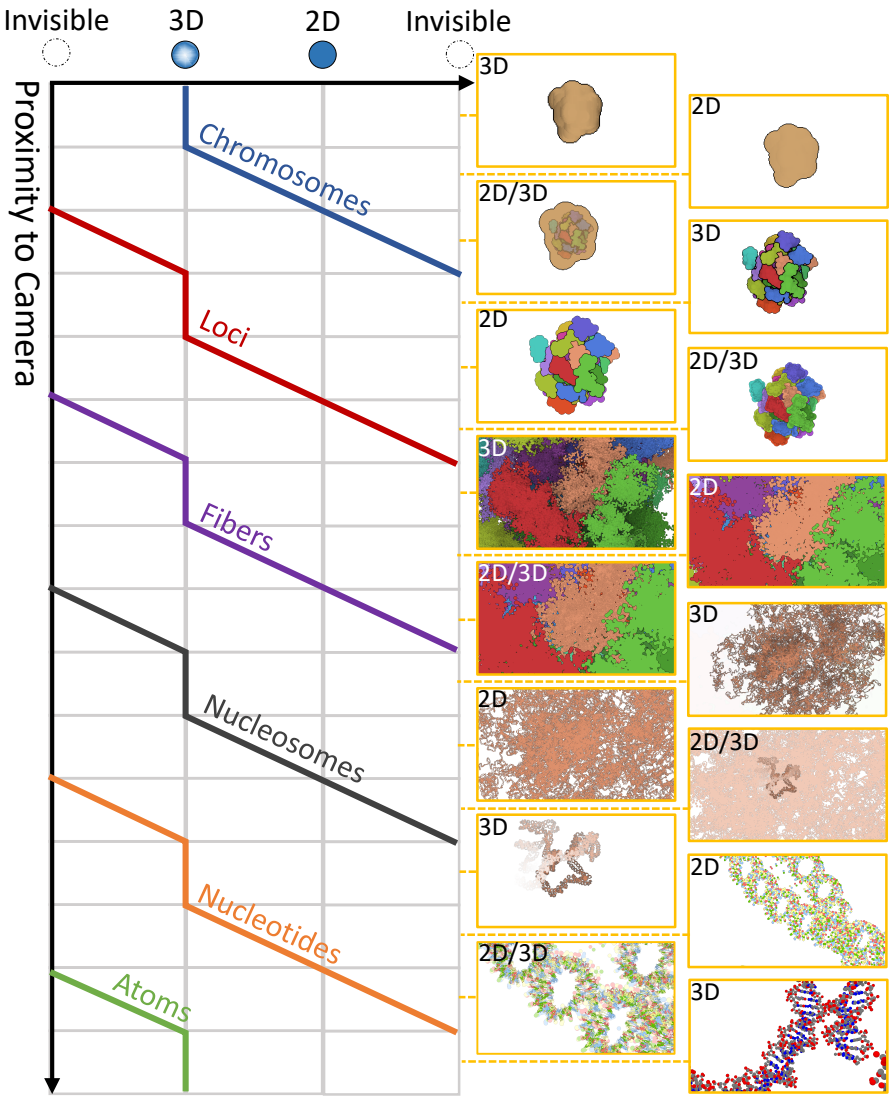



Figure 3: Temporally-controlled smooth transitions traversing the organizational levels of the genome in ScaleTrotter.

Even though the transitions used between scales allow two adjacent scales to be rendered on top of each other (for a brief moment), ScaleTrotter does not allow an overview of all the scales, which makes it harder to grasp the scale differences. In addition, the dense packing of the genome makes it almost impossible to follow sequences at all scales. Therefore, in Chapter 4, we present another framework, Multiscale Unfolding, that allows the interactive creation and exploration of spatially-controlled multiscale visualizations. To emphasize the sequential aspect of the data, we demonstrate how to unfold the data according to their scales. We use spatial transitions between the unfolded scales to create a visualization composed of all the scales simultaneously. We also introduce interaction techniques that allow the exploration of the data while emphasizing the scale differences and attempt to recover lost information about the 3D organization (Figure 4).

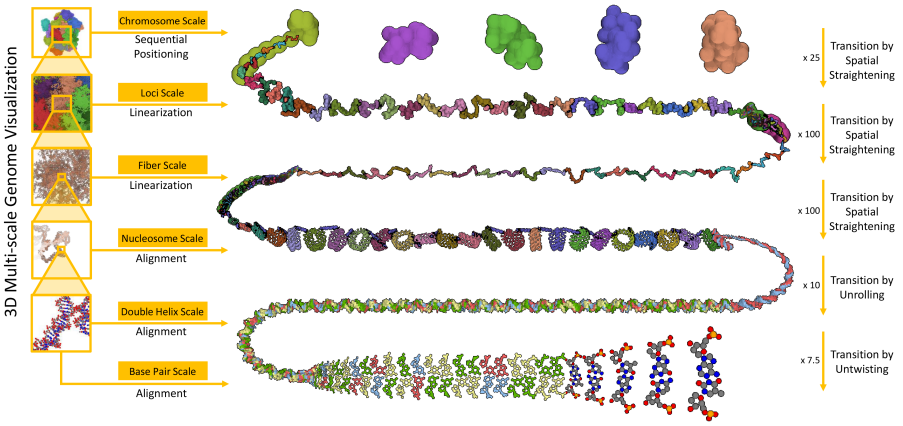



Figure 4: Multiscale unfolding with continuous spatial scale transitions of the 3D human genome.

In each of Chapter 3 and Chapter 4, sacrifices had to be made in order to highlight different aspects of the genome data. However, these aspects are complementary and are both essential for getting a complete understanding of genome function and activity (even though this sounds very ambitious at the moment). We therefore minimize the losses suffered as result of the sacrifices by proposing, in Chapter 5, a framework that combines the previous two. We discuss how smooth transitions from the unfolded model to the 3D model, and vice versa, are possible. We also propose certain functionalities meant to give users control over the scales at which they want to see different parts of the data and the path that the unfolded model occupies. We then discuss how different user groups could benefit from such tools and what data related consideration should be taken into account in the future. We conclude Chapter 5 with what we learned, through this thesis, about abstraction and illustrative visualization.

In the final Chapter 6, we reflect on the work presented in this thesis and our ultimate goal of using abstraction to visualize multiscale 3D data. Based on what we learned from this thesis, we present an abstraction space and a path in that space for the exploration of multiscale data.

# 2 BACKGROUND

This chapter provides an overview of the major topics that this thesis relies on. Specific topics related to Chapter 3 and Chapter 4 are discussed in the Related Work sections of respective chapters.

We start with a quick explanation of the DNA macromolecule, the organization of the genome with its multiscale structural units. We discuss why both the sequential and the structural aspects determine the function and activity of the genome, based on domain literature.

Since we get inspired by illustrations to design our visualizations, we then review visualizations that rely, in their turn, on illustrative techniques. These visualizations are called illustrative visualizations and rely heavily on abstraction. We then review works that give users control over the abstraction, particularly related to the domain of structural biology.

Finally, we review works visualizing multiscale data in general focusing on the transitions between the scales and the cross-scale interactions they support.

## 2.1 THE HUMAN GENOME: FROM NUCLEOTIDES TO CHROMATIN

The story of DNA research starts in 1869, about 150 years ago, when Friedrich Miescher [92] discovered a novel substance he termed "nuclein" inside the cells' nuclei [28] (Figure 5). This term is still preserved in today's name deoxyribonucleic acid, abbreviated to DNA. DNA and protein are the elements forming the chromosomes. In 1929, Pheobus Levene identified that DNA is a string of subunits called nucleotides [28], which are composed of a sugar molecule, the deoxyribose (hence the name deoxyribonucleic acid), attached to a single phosphate group and one of four nitrogen containing bases: adenine (A), cytosine (C), guanine (G), thymine (T) (Figure 6).

In 1944, Avery et al. [6] demonstrated that DNA, not some protein as previously thought, is the hereditary molecule carrying the genetic information. In 1950, Erwin Chargaff [22] found that the nucleotide composition of DNA varies among species, and that within a species the bases in DNA are always present in fixed ratios: the same number of A's as T's and the same number of C's as G's. The latter is known as Chargaff's rule. In 1953, inspired by Chargaff's law, Watson and Crick [130] discovered the double helical structure of the DNA (Figure 7), in which A always pairs with T, and C always with G. The double helix structure solved, among others, the puzzle of how the genetic information could be copied for transmission from cell to cell,


Figure 5: Glass vial containing nuclein isolated from salmon sperm by Friedrich Miescher. The faded label reads "Nuclein aus Lachssperma, F. Miescher" (Nuclein from salmon sperm, F. Miescher). Possession of the Interfakultäres Institut für Biochemie (Interfacultary Institute for Biochemistry), University of Tübingen, Germany; photography by Alfons Renz, University of Tübingen, Germany.

showing an example of how the 3D structure can justify biological activity.

In the early 2000's, the human genome was (nearly) completely sequenced (with small gaps still unresolved to this day; Figure 8). It consists of approximately $3.2 \times 10^9$ nucleotide pairs [2], distributed between 23 pairs of chromosome. In order for the 2 m chain of stretched out nucleotide pairs to fit in a human cell of 6 µm diameter, DNA molecules are highly condensed with the help of proteins.

Nucleosomes are the first level of chromosome packing: 146 nucleotide pairs in double helix form are wrapped around histone proteins (Figure 9). At this initial level of packing, the chromatin is a series of "beads on a string", where the string is the DNA and the bead is a "nucleosome core particle" that consists of DNA wound around a histone. The part of the DNA (string) that connects consecutive nucleosome core particles (beads) is called linker DNA [2] Figure 10.

At the second level of chromosome packing, the nucleosomes are packed on top of one another. For a long time, it was considered that the nucleosomes are packed regularly, consecutively condensing into 30 nm, then 120 nm, then 300–700 nm wide fibers. However, Ou et al. [101] showed that DNA and nucleosomes assemble into disordered chains that have diameters between 5 and 24 nm, with different particle arrangements, densities, and structural conformations (Figure 11). The second level of packing is therefore a less regularly structured 5–24 nm wide fibrous chain of nucleosome core particles.

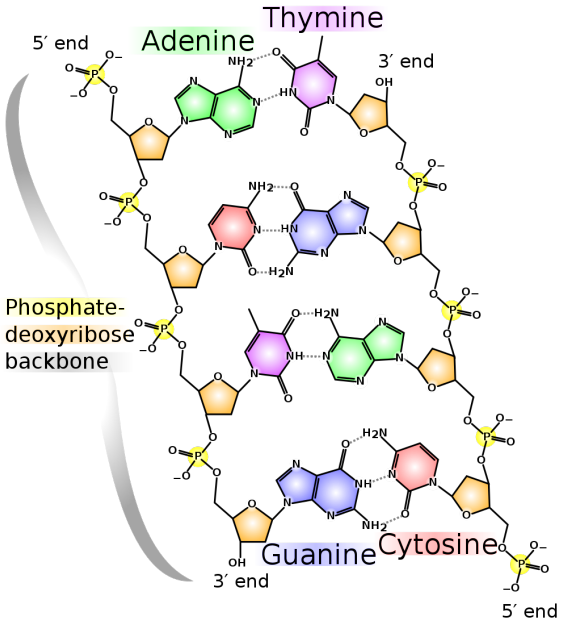



Figure 6: The figure shows the chemical structure of the DNA: the phosphate group (in yellow) and the deoxyribose sugar (in orange) form the sugar-phosphate backbone, from which the bases (i. e., Adenine, Cytosine, Guanine, Thymine) extend. The DNA molecule is shown straightened out; in reality, it is twisted into a double helix like in Figure 7. Illustration by Madeleine Price Ball.

The disordered fiber forms loops, in order to bring relatively distant elements along the linear genome into close spatial proximity [14] (Figure 12, right). Bringing this linearly distant but codependent elements closer triggers processes such as DNA transcription, an important step of gene expression. The chromatin loop forms the third level of organization.

Loops that interact with each other more frequently than others form structural landmarks called topologically associated domains, abbreviated to TADs [31], and thus forming the fourth organizational level (Figure 12, middle).

TADs that interact with each other, even if they are very far on the linear genome, give rise to compartments [14], representing the fifth organizational level (Figure 12, left).

The sixth and final level of the chromosome organization is the coalescence of compartments in the same chromosome, forming chromosomal territories [14]. The joining of territories forms the chromosome.

To summarize, the chromosome is organized hierarchically, where distances in the linear genome are compensated by dense structural folds reducing physical distances, therefore having an important role in the genome activity and function.

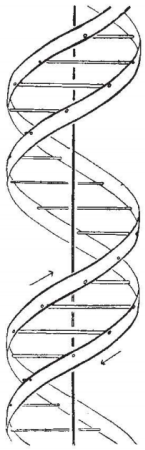


Figure 7: The double helix structure of the DNA. Figure taken from Watson and Crick's article [130] that appeared in the journal Nature on April 25, 1953, and revealed the structure of DNA.

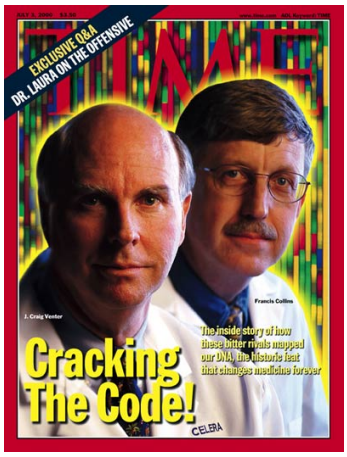



Figure 8: J. Craig Venter and Dr. Francis Collins on the cover of TIME magazine, after they jointly made the announcement of the mapping of the human genome.

The data we currently have and use represents the structure of a human genome on 6 conceptual levels. These conceptual levels might not correspond perfectly to the organizational levels described in the previous paragraphs. However, our frameworks focus on the concept of continuous abstraction for multiscale genome data and are flexible in that the data can be easily replaced with newer data.

Our understanding of higher-order chromatin organization has dramatically expanded thanks to new technologies that detect 3D organizations of chromatin [137]. The role of this organization, in essential biological functions, has increasingly been recognized as important [14]. Genome browsers typically display linear sequences as horizontal tracks, aligned and stacked below the reference nucleotide sequence [46] (Figure 13). Many tools, used by experts, show some

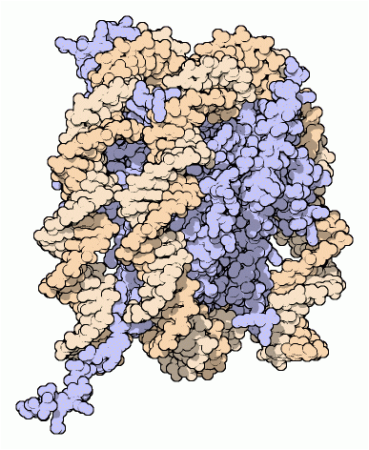


Figure 9: Nucleosome, with DNA (in orange) and histone proteins (in blue). Illustration by David S. Goodsell and the RCSB PDB, licensed under CC-BY-4.0 [44].

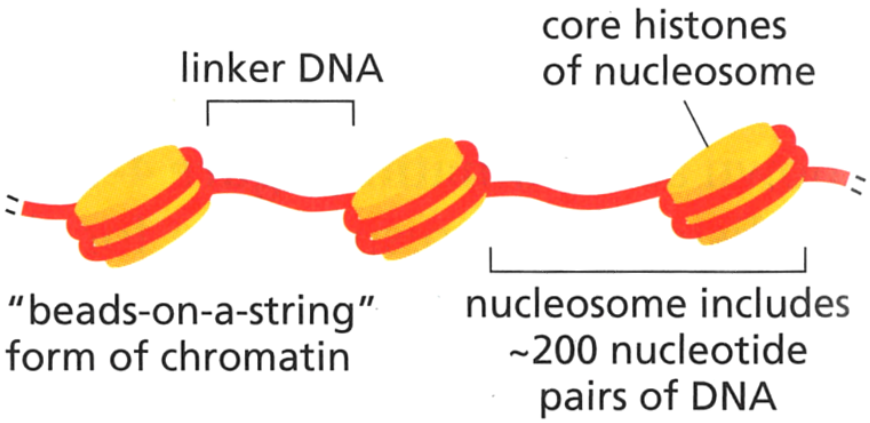



Figure 10: "Beads on a string" formed by nucleosome core particles and linker DNA. Figure from [2].

spatial information by creating visual links between the interacting regions, using matrices, heatmaps or arcs (Figure 14).

However, most of these tools represent a frozen moment/scale in the genome [46], while the data extends over multiple scales and varies in time. 3D representations of the genome model could represent these properties: time-variant and multiscale data. Genome3D [5] (Figure 15) and GMOL [98] (Figure 16) visualize the human genome in 3D based on multiscale data. Both tools, however, do not support a smooth transition between the scales, but discretely switch between them. The transition between the scales is also defined through manual inputs from the user. The tool we present in Chapter 3 provides smooth transitions between the same scales of GMOL [98], based on the camera properties. Our approach adds a missing aspect because it provides a free and smooth navigation of the 3D model, as reported

Figure 11: Stacks of nucleosome forming a disordered chain. Figure from [101].

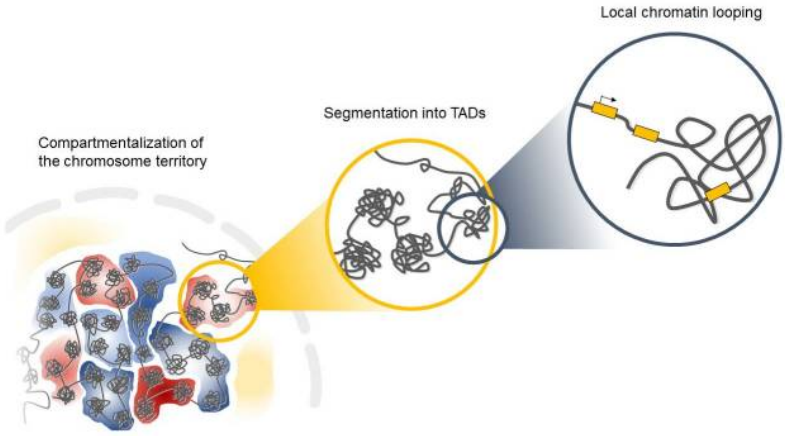



Figure 12: Spatial folding of the disordered chain into loops, TADs, compartments and territories. Figure from [115].

by feedback, and helps users in understanding how one scale relates to the other.

## 2.2 ILLUSTRATIVE VISUALIZATION AND ABSTRACTION

Other than domain experts as we addressed in our discussion so far, many people need to know about the topic and high school students learn about the organization of the carrier of genetic information and how this huge amount of information fits inside the cell nucleus. In fact, the organization of the genome might be of interest to anyone. Since the function of its spatial arrangement has been established [14], a person suffering from a disease, caused by the genome's structure, deserves to be properly informed about the cause of the disease. Even a healthy person can be curious about the genetic code that guides life and its processes. One of the most efficient ways to make scientific concepts, specially those that are not visible in everyday life like the genome, accessible to everyone is through scientific illustrations. For example, textbooks of biology are filled with illustrations.

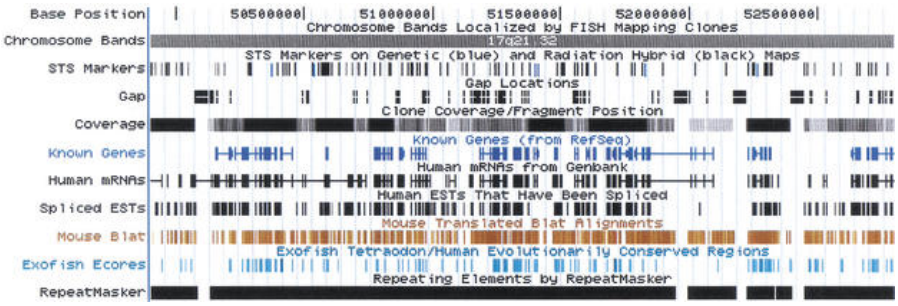



Figure 13: Genome browser showing a region of a chromosome spanning several million bases in horizontal tracks. Figure from [64].

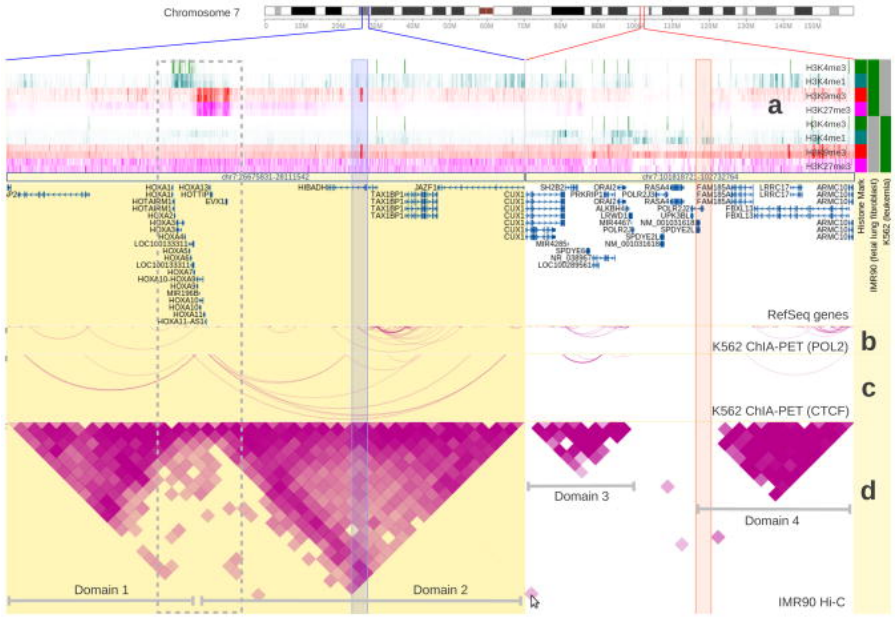



Figure 14: WashU Browser showing interacting regions through matrices and arcs. Figure from [138]

Scientific visualization shares the same principle as illustration in making scientific concepts more accessible to users. Therefore, scientific visualization uses techniques developed in traditional illustrations. These techniques are also referred to as smart visibility techniques [127]. Visualizations that are inspired by techniques used in illustrations are categorized as illustrative visualizations [109]. For example, molecular visualization tools, such as cellVIEW [76] (Figure 17a), take inspiration from illustrations (Figure 17b). Since the DNA is a macromolecule, its visualization is closely related to molecular visualization. Therefore, we also look for inspiration from illustrations.

When making illustrations, illustrators typically simplify the represented models in order to reduce distractions and focus attention on the concepts they are trying to convey. The result of an illustration is therefore not necessarily the representation closest to the reality of

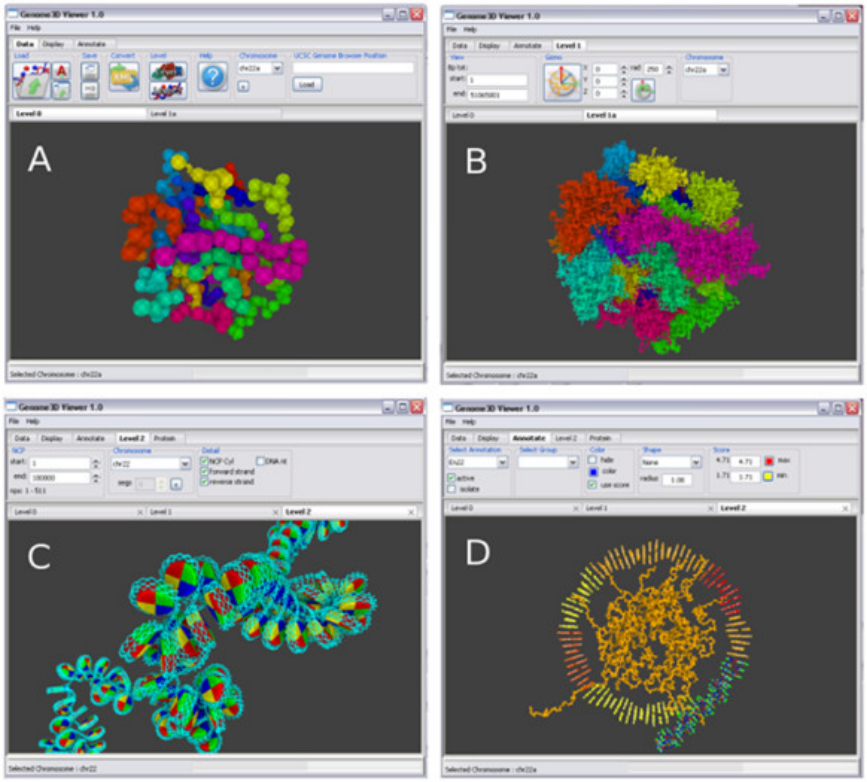



Figure 15: Four screen captures of Genome3D main windows showing progressive "drill-down" views of the same model multi-resolution genomic data. Figure from [5].

the model, but the one that most effectively conveys the intentions of the illustrator. This process of simplification is part of a fundamental concept called abstraction, which can virtually always be found in illustrations.

However, illustrative visualization, other than using expressive techniques to communicate results, has to simultaneously provide insight and allow the exploration of the underlying data, and therefore does not enjoy the same degree of freedom as illustration [109]. Illustrations have to be entirely remade every time the underlying data, or our understanding of the depicted concepts, changes, while illustrative visualizations can adapt to changes in the data provided to them. Another difference is that illustrations are usually not interactive, while interactivity is a main component of visualization [94].

Based on the example set by illustrations, illustrative visualization has to support abstraction and allow total control over it. In consequence, designers of illustrative visualizations have attempted to define and eventually control the use of abstraction. Rautek et al. [109] differentiate between low-level abstractions, focusing on the style in which features of interest are represented, and high-level abstractions,

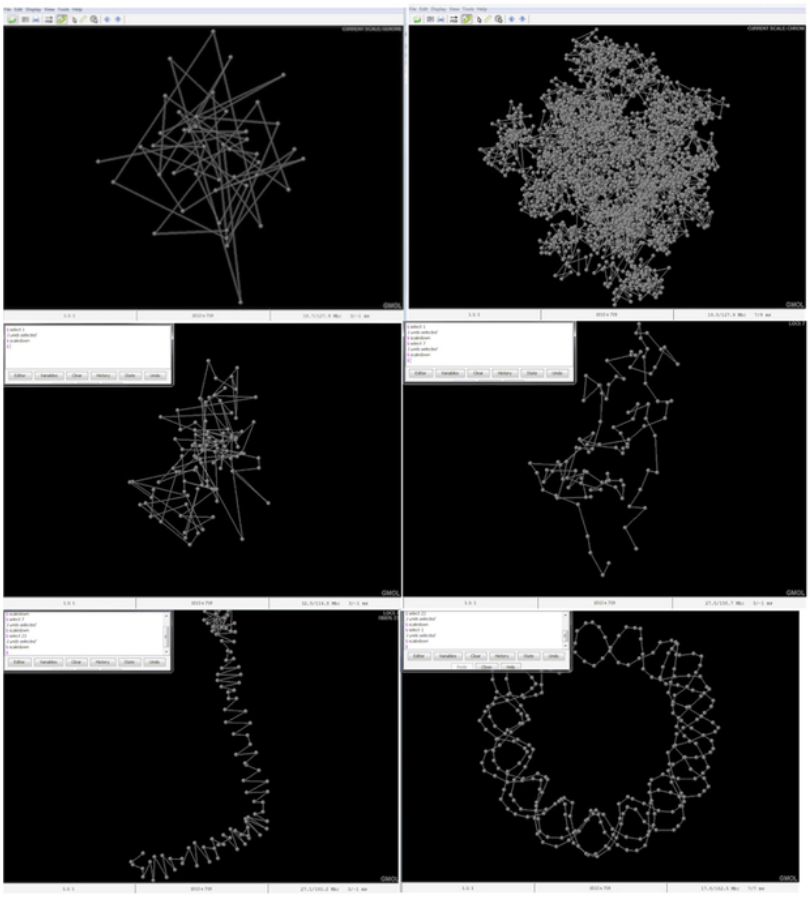
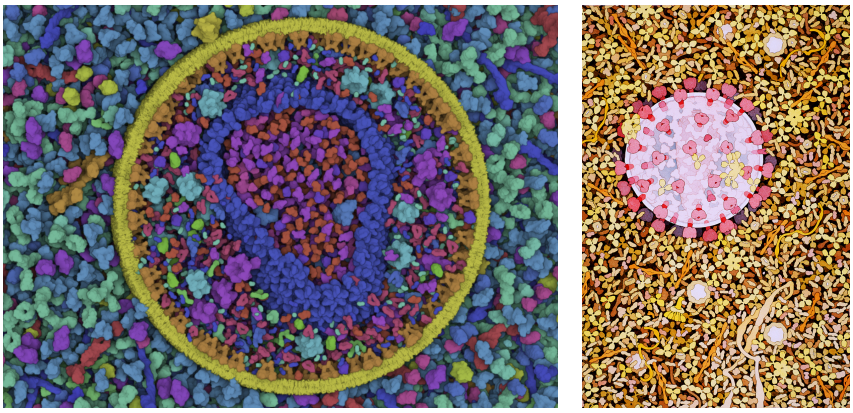


Figure 16: Screen captures of GMOL showing all the different scales. Figure from [98].

focusing on the content of the visualization based on the importance of features of interest.

Viola et al. [126] go further and define abstraction as the process that transforms a source into a less concrete sign of that source. This process comprises multiple steps, each step being a type of abstraction: an initial abstraction transforms a reality into a digital form, a data abstraction that transforms the digital form into data representations, and a visual abstraction transforming data representations into visual representations. What is of a particular interest for this thesis is the last step, the visual abstraction, which transforms a data representation into a visual representation, while intentionally disregarding certain aspects of the data representation. Which aspects are disregarded depends on the purpose of the visual abstraction, which renders it meaningful: the aspects that, if perceived as a stimulus, reduce the cognitive load are preserved, while the others are dismissed.

Based on this definition, abstraction can be controlled through the notion of an abstraction axis, previously introduced by Viola et al. [128]. An abstraction axis is a succession of visual representations, where aspects of a same category are continuously dismissed, while



(a) (b)

Figure 17: (a) Real-time screen-shot of an illustrative cross-section of the HIV virus surrounded by blood plasma. Figure from [76]. (b) Illustration of an HIV under attack by the immune system. Illustration by David S. Goodsell, The Scripps Research Institute.

aiming for the same goal all along [126]. For example, Everts et al.'s [34] abstraction axis controls the contraction of brain fiber tracts and provides smooth transitions between the different states of contraction.

Two abstractions axes that can be independently controlled form an abstraction space. Mohammed et al. [96] create a controllable abstraction space with two axis each controlling the structural abstraction of a type of cell (Figure 18). Other examples [89, 125], that we discuss in Section 2.3, provide control of abstraction spaces created by axes of different abstraction types (Figure 19).

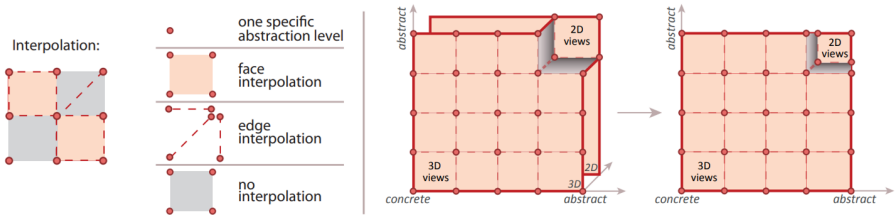



Figure 18: Abstraction space of Abstractocyte, with the transition methods on the left. Figure from [96].

To make an axis controllable, transitions between its successive steps is necessary. The transitions can be either temporally-controlled or spatially-controlled. When the different representation appear continuously (one after the other), occupying the same areas on screen, transitions are temporally-controlled, since they share screen space but not screen time (examples in Figure 20, 21, 22). However, if the different representation appear simultaneously (altogether at the same time), side by side, transitions are spatially controlled, since they share screen time but not screen space (example in Figure 24).

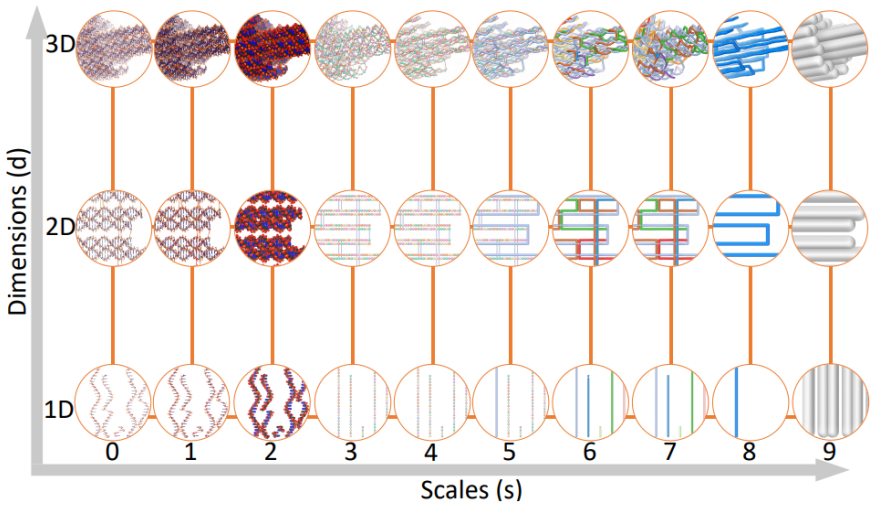



Figure 19: Abstraction space of DimSUM formed by a dimension axis and a scale axis. Figure from [89].

In the following paragraphs, we review previous works related to molecular visualization and multiscale visualization, that present some sort of abstraction axes, whether with temporally-controlled or spatially-controlled transitions.

## 2.3 ABSTRACTION IN MOLECULAR VISUALIZATION

Van der Zwan et al. [125] suggest three separate axes of abstraction. First, they gradually transition from a space fill model of a protein molecule to a ribbon model, passing through five structural abstraction stages (Figure 20). Second, they transition between different stages of spatial perception (Figure 21) and, third, between different stages of illustrativeness (Figure 22).

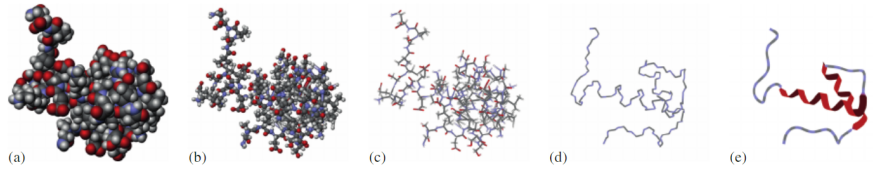



Figure 20: Structural abstraction stages: (a) space fill, (b) balls-and-sticks, (c) licorice, (d) backbone, and (e) ribbon. Figure from [125].

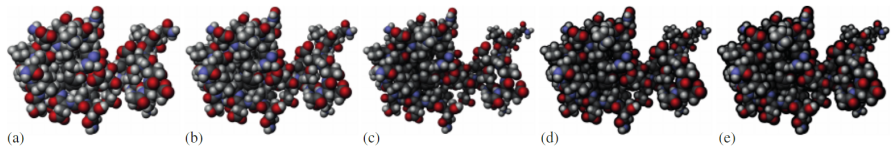



Figure 21: Stages of support of spatial perception: (b)–(c) ambient occlusion and object attenuation and (d)–(e) added halos. Figure from [125].

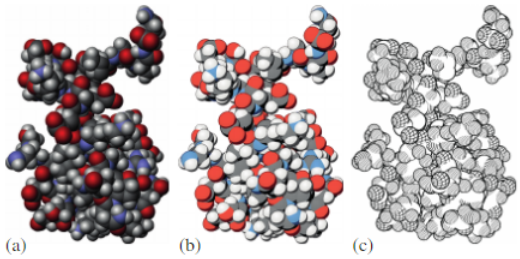



Figure 22: Abstraction along 'illustrativeness': (a) 'photorealistic,' (b) cel shading, and (c) black-and-white. Figure from [125].

Cipriano et al. [24, 25], guided by different high-level tasks, use a series of different surface abstractions. Miao et al. [90] apply seamless temporal transitions between ten semantic representations of DNA nanostructures, with different degrees of abstraction. In this thesis, we are focusing on displaying the organizational levels of the genome that exist at different scales. Our goal is to understand abstraction in the context of multiscale data. The mentioned examples of abstraction do not infer the large scale changes that we would like to have, we could use them however to abstract on each scale individually.

Later, Miao et al. [89] combine their ten semantic representations of DNA nanostructures with another axis which abstracts the dimensions from 3D to 1D (Figure 19). Lindow et al. [84] also abstract along the dimensions to represent DNA and RNA structures (Figure 23). These abstractions are guided by domain specific tasks. We focus on visualizing 3D data at all scales, but abstractions along the spatial dimensions could be envisioned, whether at certain scales or all scales together. These abstractions require a deep understanding of the tasks that experts have to accomplish.

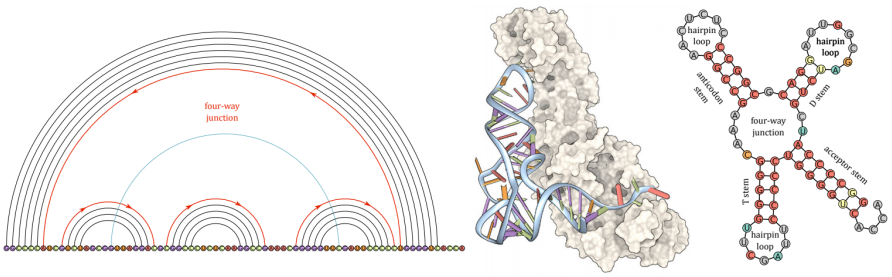



Figure 23: 2D linear model (left), 3D visualization (middle) and 2D graph model (right) of an RNA model. Figure from [84].

These examples feature temporally-controlled transitions between different abstraction stages. As previously mentioned in Section 2.1, the spatial organization of the genome plays an important role in essential biological processes. We therefore want to visualize the spatial structure of all the scales. Temporally-controlled transitions allow us to leave the data intact at each scale, and smoothly transition between scales, as we show in Chapter 3.

However, when staying true to the 3D structure, we sometimes lose the ability to follow patterns because of the tightly packed nature of DNA. After all, DNA is a sequence of $3.2 \times 10^9$ nucleotide pairs, whose order is crucial. To emphasize this linear aspect of our multiscale data, we straighten the data and use spatially-controlled transitions to visualize all scales at once.

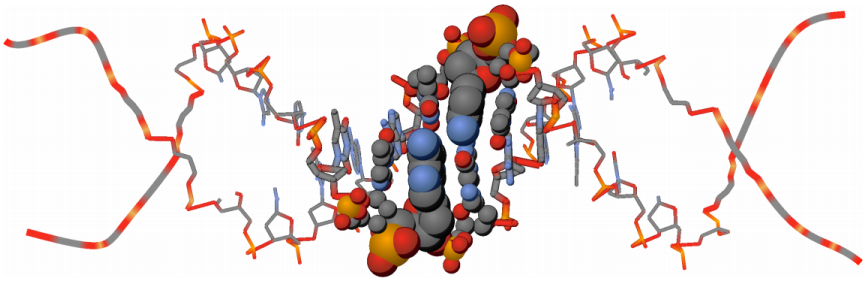


Figure 24: Spatial control of structural abstraction of a DNA double helix. Figure from [85].

Similar abstractions to the ones mentioned in the previous examples feature spatially-controlled transitions. Lueks et al. [85] use similar abstraction stages to van der Zwan et al.'s [125], but with spatial transitions. Parulek et al. [103, 104] use seamless spatial transitions by interpolating between different representations along two abstraction axes. Once again, these abstractions do not imply important scale changes. We could however use them, if necessary, on individual scales.

Since abstraction is a transformation that disregards detail of a source element, level-of-detail (LOD) techniques [8, 35, 76, 103] are great examples of abstraction, whether supporting temporal or spatial transitions. A visual abstraction is guided by the goal of reducing the cognitive load of the user. Some LOD abstractions serve this goal indirectly, because their primary goal is to make the rendering of large molecular structures possible and interactive in real time. This short-term goal itself serves the higher goal of reducing the cognitive charge of the user. We take inspiration from the LOD techniques and remove entire parts of the data when their level of detail is irrelevant to the situation.

## 2.4 ABSTRACTION IN MULTISCALE VISUALIZATION

Scale is a system that links the size of a phenomenon to conventionally defined numerical representations [82]. In this thesis, we focus on visualizing the human genome, which along its organizational levels, covers several orders of magnitude in scale. When representing data that cover multiple scales, in other words multiscale data, one of the main goals of the visualization is commonly helping users grasp and

relate the measures covered in the visualization. This process of interpreting measures by a user involves the construction of mental models of scale and is referred to in psychology as scale cognition or scale sense [23]. Studies in the domain suggest that personal and professional experiences play a critical role in the conceptualization of scale [82, 102, 122]. They also found that scale cognition is refined by visual and kinesthetic experiences and direct interaction with phenomenons of various magnitudes, among students of various schooling levels [122] and experienced professionals [60, 122].

For the majority of laypeople, interactions with genome data are nonexistent and the only visual experiences are through illustrations. Experts might have access to expensive electron microscopes, but even microscopy is unable to distinguish different parts of a chromosome simultaneously or at high resolution [124]. Ou et al. [101] have developed a technique that reveals the chromatin ultrastructure and 3D packing of DNA in both human interphase cells and mitotic chromosomes (Figure 25). However, to study the 3D architecture of genomes, experts rely on chromosome conformation capture techniques which detect the frequency of interaction between any two genomic loci (specific positions) [29, 124].

Visualization of genome data should therefore serve the initial goal of enriching the visual experiences and interactions of both layman users and experts with concepts that are not accessible to the eye. The ultimate goal remains the reduction of the cognitive charge imposed on the target user trying to grasp the measures distributed over several orders of magnitude in scale.

In the scope of this thesis, we discuss multiscale data if it covers several orders of magnitude, in particular from individual atoms ($10^{-10}$m) to the nucleus of a human cell ($10^{-5}$m). Most of the previous work mentioned in Section 2.3 cover a small scale range of scale. Van der Zwan et al. [125] transition between an atomistic representation of a molecule to a ribbon diagram representation. On one side of the abstraction axis, atoms are in the order of $10^{-10}$m while on the other side, ribbons represent the helical structure in the order of $10^{-9}$m. Similarly, even with as many as ten abstraction stages, Miao et al. [90] transition from an atomistic representation to a tubular representation of the structure. These abstractions cover a small range in spatial scale, and therefore do not fit in our description of multiscale data, as they would only cover one or two scales.

Other than covering 4–5 orders of magnitude in physical size, the human genome comprises 7 different conceptual scales, according to the current understanding of genome scientists.

A field that covers more orders of magnitude than the human genome with similar number of conceptual scales is astronomy. Fundamental differences exist however between both datasets. These differences include scale density (number of conceptual levels in the

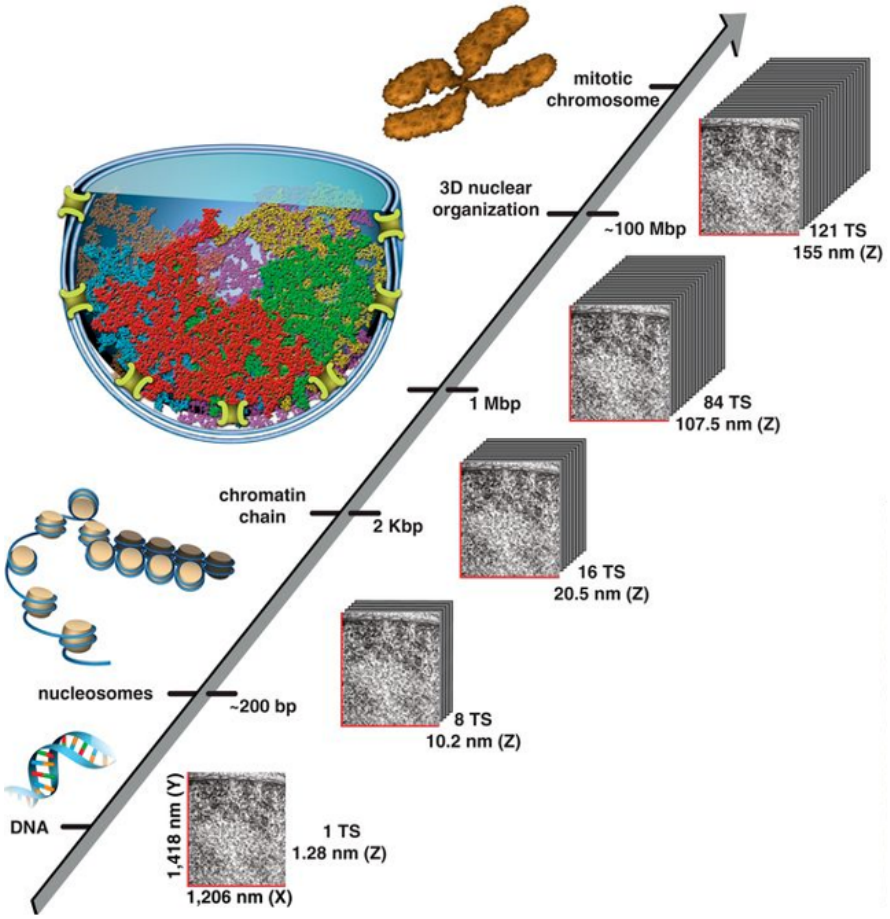



Figure 25: ChromEMT enables the ultrastructure of individual chromatin chains, megabase domains, and mitotic chromosomes to be resolved and visualized as a continuum in serial slices through large 3D volumes. Figure from [101]

covered scales) and physical connection in space across conceptual levels. Compared to DNA, astronomical multiscale data has a similar number of conceptual levels covering a larger spectrum of orders of magnitude. It therefore has a higher scale density. Also, in astronomy unlike the DNA, the data across the conceptual levels is not physically connected. These differences create the need for different approaches. For example, because of the properties of multiscale astronomical data, when transitioning between two levels in astronomy, one level might completely disappear before its higher level appears, leaving an empty space in between. However, this does not happen in DNA visualization and we therefore need appropriate transitions. We further discuss details about these differences in Chapter 3, where we motivate our first approach for a temporally controlled exploration of genome data.

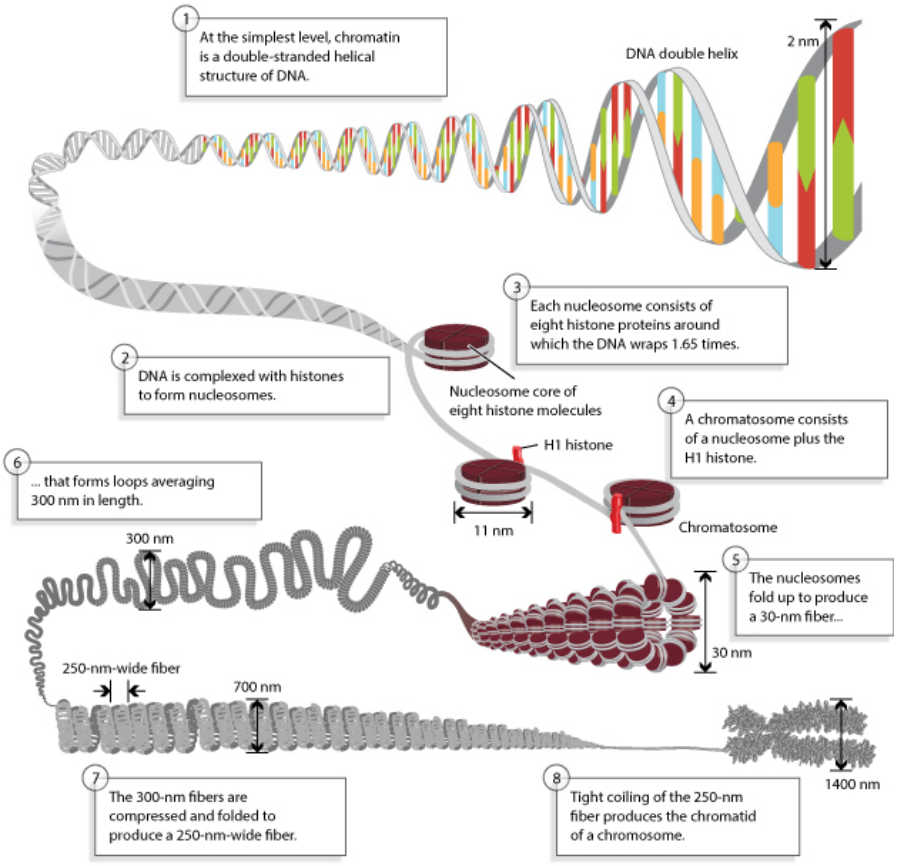



Figure 26: Illustration of the organization levels of DNA with spatial transitions. ©2013 Nature Education Adapted from Pierce, Benjamin. Genetics: A Conceptual Approach, 2nd ed. All rights reserved.

Illustrations use a set of effective techniques to represent multiscale data. Some static illustrations of DNA take advantage of the linear aspect of the data to spatially transition between its different organizational levels (Figure 26). Since DNA is highly condensed and packed, illustrators have to modify of parts of the data in order to show the different organizational levels, which makes it difficult to see their structure.

To make such illustrations possible, the sizes of the levels are modified such that levels that are far in scale from each other are represented in comparable sizes. We therefore lose the notion of relative sizes. Some illustrations use perspective scale transitions in order to represent scale differences more correctly. Other illustrations use techniques such as juxtaposition to represent levels of different sizes, side by side, staying true to the relative difference of size (Figure 27). However, because of the orders of magnitude covered by the genome, displaying all the levels while respecting the relative size difference with correct perspective would mean that the lower scales would not be visible.

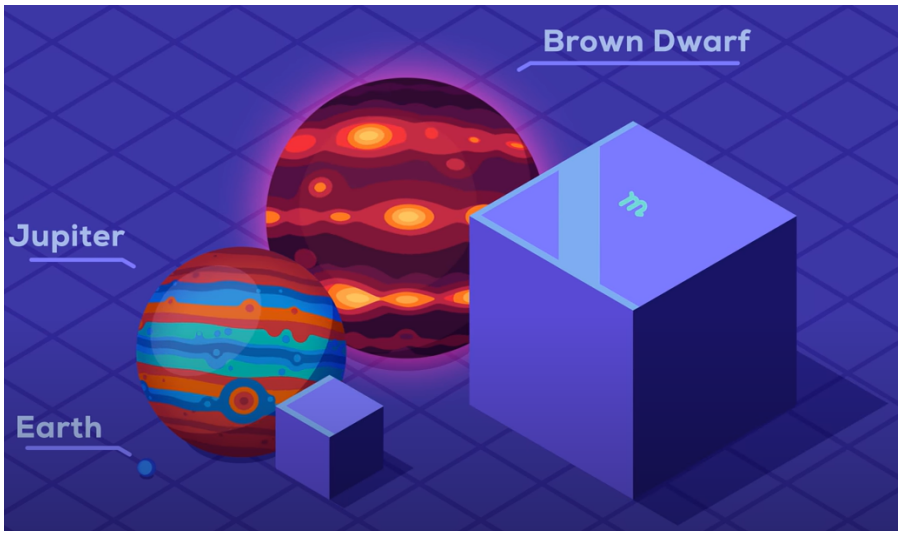



Figure 27: Illustration comparing sizes of celestial bodies using juxtaposition. Still from youtube video https://youtu.be/3mnSDifDSxQ

Illustrations also use a clever arrangement to hide transitions between levels, specially between levels that are shown using different representations or even completely skip certain levels (Figure 28). We cannot use these special arrangements because we want to show all levels. Transitions are also essential for our work as we use them to compensate for the loss of relative size differences. We take inspiration from the spatial transitions of these illustrations to create the framework that we describe in Chapter 4.

Other illustrations, that represent the structures of the organizational levels with higher confidence, use different techniques that do not require modifications of the data, such as close-ups [118] (Figure 12). Animations and interactive experiences covering many orders of magnitude use zooming effects and/or clever image blending to make the changes unnoticeable. We take inspiration from these techniques to create the framework the we present in Chapter 3.

## 2.5 SUMMARY

In this chapter, we explained the different organizational levels of the human genome, according to our understanding at the moment of writing this thesis. Recent discoveries have shown that the 3D organization of chromatin modulates important biological processes. Therefore, if we want to gain full insight over the activity of the genome, understanding the linear sequence of nucleotides is essential but not sufficient. We also have to understand how DNA is packed inside the nucleus: a packing that gives rise to the different organizational levels, controlling certain processes and expanding over 5 orders of magnitude in scale.

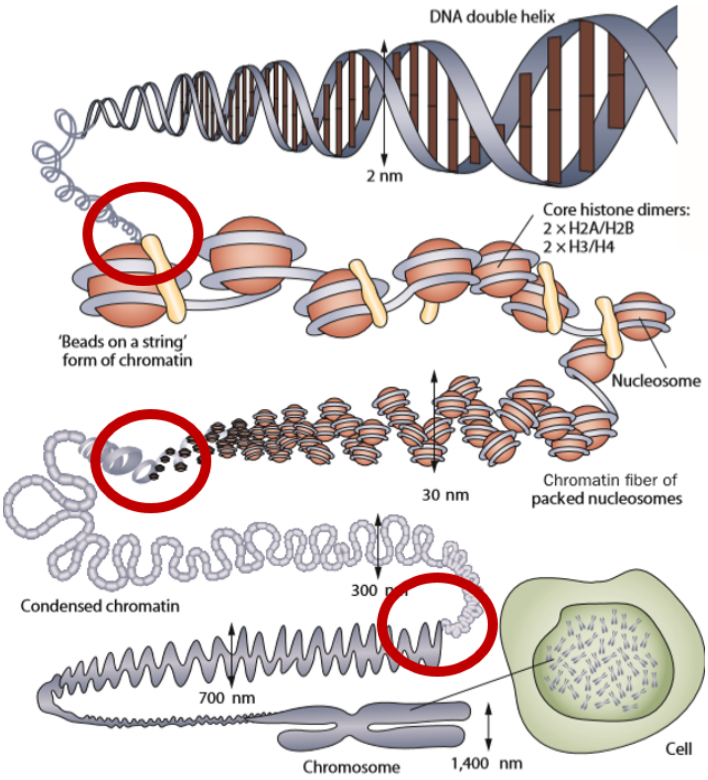
Experts use techniques that gather interaction data at high resolutions, showing the frequency of contact between different regions of the DNA. From the interaction data, they derive a model of the human genome at different scales. However, they lack the tools to visualize this 3D spatial data.

Understanding the packing of the DNA is not only the job of experts. Students or anyone curious enough is entitled to understand such a critical element of life.

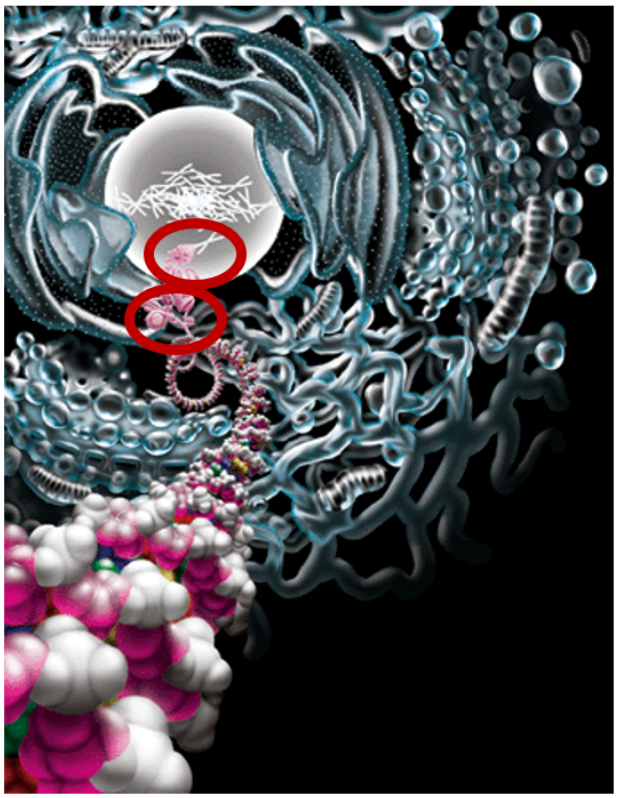
We therefore take inspiration from existing scientific illustrations that already play the role of communicators of scientific concepts to the general audience. Illustrations typically use abstractions to guide the viewer and convey the intended message. However, illustrations do not rely on actual data and are usually not explorable.

In order to design visualizations that are based on data, allow its exploration, and successfully communicate the multiscale aspect of the data, we looked at techniques used in illustrative visualizations and how they control abstraction. We also reviewed previous work in two areas that are complementary for DNA visualization: molecular visualization and multiscale visualization.

Eventually, we came up with two complementary frameworks, each focusing on a different aspect of the multiscale genome data. First, we introduce a framework that focuses on the 3D structure of the multiscale data, using temporally controlled transitions (Chapter 3). However, when we show the full 3D structure, the sequence is not recognizable. Therefore, we provided a second framework that modifies the data, sacrificing aspects of the 3D structure, and emphasizes the sequential aspect of the multiscale data, using spatially controlled transitions (Chapter 4). We noticed that a trade-off was necessary to show both the 3D structural and sequential aspects of the data, and tried to minimize the sacrifices that had to be made.




(a)



(b)

Figure 28: Illustrations of organizational levels of DNA using clever arrangement to hide spatial transitions. The absence of smooth transitions is highlighted in red circles.

# 3

# SCALETROTTER: ILLUSTRATIVE VISUAL TRAVELS ACROSS NEGATIVE SCALES

The understanding of the full function of the genome requires the exploration of both its structural and sequential aspects. As a first step, in this chapter, we use a temporal abstraction axis to explore the spatial structure of the genome. We thus present ScaleTrotter, a conceptual framework for an interactive, multiscale visualization of biological mesoscale data and, specifically, genome data. ScaleTrotter allows viewers to smoothly transition from the nucleus of a cell to the atomistic composition of the DNA, while bridging several orders of magnitude in scale. The challenges in creating an interactive visualization of genome data are fundamentally different in several ways from those in other domains like astronomy that require a multiscale representation as well. First, genome data has intertwined scale levels—the DNA is an extremely long, connected molecule that manifests itself at all scale levels. Second, elements of the DNA do not disappear as one zooms out—instead the scale levels at which they are observed group these elements differently. Third, we have detailed information and thus geometry for the entire dataset and for all scale levels, posing a challenge for interactive visual exploration. Finally, the conceptual scale levels for genome data are close in scale space, requiring us to find ways to visually embed a smaller scale into a coarser one. We address these challenges by creating a new multiscale visualization concept. We use a scale-dependent camera model that controls the visual embedding of the scales into their respective parents, the rendering of a subset of the scale hierarchy, and the location, size, and scope of the view. In traversing the scales, ScaleTrotter is roaming between 2D and 3D visual representations that are depicted in integrated visuals. We discuss, specifically, how this form of multiscale visualization follows from the specific characteristics of the genome data and describe its implementation. Finally, we discuss the implications of our work to the general illustrative depiction of multiscale data.

Main portions of this chapter were previously published at IEEE VIS 2019 [50]. Therefore, any use of "we" in this chapter refers to myself, Haichao Miao, David Kouřil, M. Eduard Gröller, Ivan Viola, and Tobias Isenberg.

## 3.1 INTRODUCTION

The recent advances in visualization have allowed us to depict and understand many aspects of the structure and composition of the living cell. For example, cellVIEW [76] provides detailed visuals for viewers to understand the composition of a cell in an interactive exploration tool and Lindow et al. [84] created an impressive interactive illustrative depiction of RNA and DNA structures. Most such visualizations only provide a depiction of components/processes at a single scale level. Living cells, however, comprise structures that function at scales that range from the very small to the very large. The best example is DNA, which is divided and packed into visible chromosomes during mitosis and meiosis, while being read out at the scale level of base pairs. In between these scale levels, the DNA's structures are typically only known to structural biologists, while beyond the base pairs their atomic composition has implications for specific DNA properties.

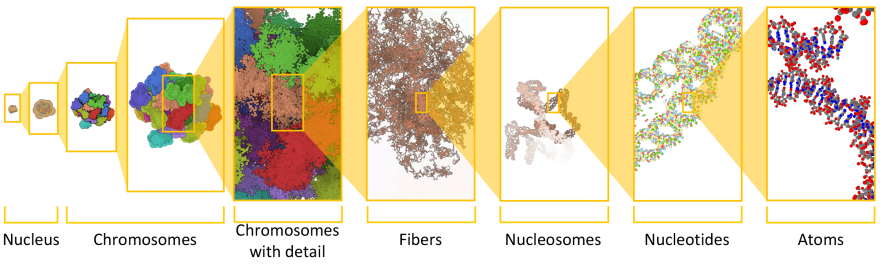



Figure 29: Steps along the voyage into the genomic detail as enabled by ScaleTrotter, showing the semantic scale levels. ScaleTrotter and its *visual embedding* allow us to seamlessly transition between independent representations and interactively explore them.

The amount of information stored in the DNA is enormous. The human genome consists of roughly 3.2 Gb (giga base pairs) [2, 114]. This information would fill 539,265 pages of the TVCG template, which would stack up to approx. 27 m. Yet, the whole information is contained inside the cell's nucleus with only approx. 6 µm diameter [2, page 179]. Similar to a coiled telephone cord, the DNA creates a compact structure that contains the long strand of genetic information. This organization results in several levels of perceivable structures (as shown in Figure 29), which have been studied and visualized separately in the past. The problem thus arises of how to comprehend and explore the whole scope of this massive amount of multiscale information. If we teach students or the general public about the relationships between the two extremes, for instance, we have to ensure that they understand how the different scales work together. Domain experts, in contrast, deal with questions such as whether correlations exist between the spatial vicinity of bases and genetic disorders. It may manifest itself through two genetically different characteristics that are far from each other in sequence but close to each other in

the DNA's 3D configuration. For experts we thus want to ensure that they can access the information at any of the scales. They should also be able to smoothly navigate the information space. The fundamental problem is thus to understand how we can enable a smooth and intuitive navigation in space and scale with seamless transitions. For this purpose we derive specific requirements of multiscale domains and data with negative scale exponents and analyze how the constraints affect their representations.

Based on our analysis we introduce ScaleTrotter, an interactive multiscale visualization of the human DNA, ranging from the level of the interphase chromosomes[1] in the 6 µm nucleus to the level of base pairs (≈ 2 nm) resp. atoms (≈ 0.12 nm). We cover a scale range of 4–5 orders of magnitude in spatial size, and allow viewers to interactively explore as well as smoothly interpolate between the scales. We focus specifically on the visual transition between neighboring scales, so that viewers can mentally connect them and, ultimately, understand how the DNA is constructed. With our work we go beyond existing multiscale visualizations due to the DNA's specific character. Unlike multiscale data from other fields, the DNA physically connects conceptual elements across all the scales (like the phone cord) so it never disappears from view. We also need to show detailed data everywhere and, for all stages, the scales are close together in scale space.

We base our implementation on multiscale data from genome research about the positions of DNA building blocks, which are given at a variety of different scales. We then transition between these levels using what we call visual embedding. It maintains the context of larger-scale elements while adding details from the next-lower scale. We combine this process with scale-dependent rendering that only shows relevant amounts of data on the screen. Finally, we support interactive data exploration through scale-dependent view manipulations, interactive focus specification, and visual highlighting of the zoom focus.

In summary, our contributions are as follows. First, we analyze the unique requirements of multiscale representations of genome data and show that they cannot be met with existing approaches. Second, we demonstrate how to achieve smooth scale transitions for genome data through visual embedding of one scale within another based on measured and simulated data. We further limit the massive data size with a scale-dependent camera model to avoid visual clutter and to facilitate interactive exploration. Third, we describe the implementation of this approach and compare our results to existing illustrations. Finally, we report on feedback from professional illustrators and domain experts. It indicates that our interactive visualization can serve

1 While interphase chromosomes were measured to be approx. 12 µm long, this was after "flattening" and "routine chromosome preparation" [79].

as a fundamental building block for tools that target both domain experts and laypeople.

## 3.2 RELATED WORK

Our work concerns the use of abstraction in illustrative visualization, visual representations of small-scale biology and genome data, and general multiscale data visualization techniques, as we discuss next.

### 3.2.1 *Abstraction in illustrative visualization*

On a high level, our work relates to the use of abstraction in creating effective visual representations, i. e., the use of *visual abstraction*. Viola and Isenberg [128] describe this concept as a process, which removes detail when transitioning from a lower-level to a higher-level representation, yet which preserves the overall concept. While they attribute the removed detail to "natural variation, noise, etc." in the investigated multiscale representation we actually deal with a different data scenario: DNA assemblies at different levels of scale. We thus technically do not deal with a "concept-preserving transformation" [128], but with a process in which the underlying representational concept (or parts of it) can change. Nonetheless, their view of abstraction as an interactive process that allows viewers to relate one representation (at one scale) to another one (at a different scale) is essential to our work.

Also important from Viola and Isenberg's discussion [128] is their concept of *axes of abstraction*, which are traversed in scale space. We also connect the DNA representations at different scales, facilitating a smooth transition between them. In creating this axis of abstraction, we focus primarily on changes of Viola and Isenberg's geometric axis, but without a geometric interpolation of different representations. Instead, we use visual embedding of one scale in another one.

### 3.2.2 *Scale-dependent molecular and genome visualization*

We investigate multiscale representations of the DNA, which relates to work in bio-molecular visualization. Several surveys have summarized work in this field [3, 72, 73, 91], so below we only point out selected approaches. In addition, a large body of work by professional illustrators on mesoscale cell depiction inspired us such as visualizing the human chromosome down to the detail of individual parts of the molecule [45].

In general, as one navigates through large-scale 3D scenes, the underlying subject matter is intrinsically complex and requires appropriate interaction to aid intellection [42]. The inspection of individual parts is challenging, in particular if the viewer is too far away

to appreciate its visual details. Yet large, detailed datasets or procedural approaches are essential to create believable representations. To generate not only efficient but *effective* visualizations, we thus need to remove detail in Viola and Isenberg's [128] visual abstraction sense. This allows us to render at interactive rates as well as to see the intended structures, which would otherwise be hidden due to cluttered views. Consequently, even most single-scale small-scale representations use some type of multiscale approach and with it introduce abstraction. Generally we can distinguish three fundamental techniques: multiscale representations by leaving out detail of a single data source, multiscale techniques that actively represent preserved features at different scales, and multiscale approaches that can also transit between representations of different scales. We discuss approaches for these three categories next.

#### 3.2.2.1 *Multiscale visualization by means of leaving out detail*

An example of leaving out details in a multiscale context is Parulek et al.'s [103] continuous levels-of-detail for large molecules and, in particular, proteins. They reduced detail of far-away structures for faster rendering. They used three different conceptual distances to create increasingly coarser depictions such as those used in traditional molecular illustration. For distant parts of a molecule, in particular, they seamlessly transition to super atoms using implicit surface blending.

The cellVIEW framework [76] also employs a similar level-of-detail (LOD) principle using advanced GPU methods for proteins in the HIV. It also removes detail to depict internal structures, and procedurally generates the needed elements. In mesoscopic visualization, Lindow et al. [83] applied grid-based volume rendering to sphere raycasting to show large numbers of atoms. They bridged five orders of magnitude in length scale by exploiting the reoccurrence of molecular sub-entities. Finally, Falk et al. [35] proposed out-of-core optimizations for visualizing large-scale whole-cell simulations. Their approach extended Lindow et al.'s [83] work and provides a GPU ray marching for triangle rendering to depict pre-computed molecular surfaces.

Approaches in this category thus create a "glimpse" of multiscale representations by removing detail and adjusting the remaining elements accordingly. We use this principle, in fact, in an extreme form to handle the multiscale character of the chromosome data. We completely remove the detail of a large part of the dataset. If we would show all small details, an interactive rendering would be impossible and they would distract from the depicted elements. Nonetheless, this approach typically only uses a single level of data and does not incorporate different conceptual levels of scale.

#### 3.2.2.2 *Different shape representations by conceptual scale*

The encoding of structures through different *conceptual scales* is often essential. Lindow et al. [84], for instance, described different rendering methods of nucleic acids—from 3D tertiary structures to linear 2D and graph models—with a focus on visual quality and performance. They demonstrate how the same data can be used to create both 3D-spatial representations and abstract 2D mappings of genome data. This produces three scale levels: the actual sequence, the helical form in 3D, and the spatial assembly of this form together with proteins. Waltemate et al. [129] represented the mesoscopic level with meshes or microscopic images, while showing detail through molecule assemblies. To transition between the mesoscopic and the molecular level, they used a membrane mapping to allow users to inspect and resolve areas on demand. A magnifier tool overlays the high-scale background with lower-scale details. This approach relates to our transition scheme, as we depict the higher scale as background and the lower scale as foreground. A texture-based molecule rendering has been proposed by Bajaj et al. [8]. Their method reduces the visual clutter at higher levels by incorporating a biochemically sensitive (LOD) hierarchy.

Tools used by domain experts also visualize different conceptual genome scales. To the best of our knowledge, the first tool to visualize the 3D human genome has been Genome3D [5]. It allows researchers to select a discrete scale level and then load data specifically for this level. The more recent GMOL tool [98] shows 3D genome data captured from Hi-C data [124]. GMOL uses a six-scale system similar to the one that we employ and we derived our data from theirs. They only support a discrete "toggling between scales" [98], while we provide a smooth scale transition. Moreover, we add further semantic scale levels at the lower end to connect base locations and their atomistic compositions.

#### 3.2.2.3 *Conceptual scale representations with smooth transition*

A smooth transition between scales has previously been recognized as important. For instance, van der Zwan et al. [125] carried out structural abstraction with seamless transitions for molecules by continuously adjusting the 3D geometry of the data. Miao et al. [90] substantially extended this concept and applied it to DNA nanostructure visualization. They used ten semantic scales and defined smooth transitions between them. This process allows scientists to interact at the appropriate scale level. Later, Miao et al. [89] combined this approach with three dimensional embeddings. In addition to temporal changes of scale, Lueks et al. [85] explored a seamless and continuous *spatial* multiscale transition by geometry adjustment, controlled by the location in image or in object space. Finally, Kerpedjiev et al.

[65] demonstrated multiscale navigation of 2D genome maps and 1D genome tracks employing a smooth transition for the user to zoom into views.

All these approaches only transition between nearby scale levels and manipulate the depicted data geometry, which limits applicability. These methods, however, do not work in domains where a geometry transition cannot be defined. Further, they are limited in domains where massive multiscale transitions are needed due to the large amount of geometry that is required for the detailed scale levels. We face these issues in our work and resolve them using visual embeddings instead of geometry transitions as well as a scale-dependent camera concept. Before detailing our approach, however, we first discuss general multiscale visualization techniques from other visualization domains.

### 3.2.3 *General multiscale data visualization*

The vast differences in spatial scale of our world in general have fascinated people for a long time. Illustrators have created explanations of these scale differences in the form of images (e. g., [131] and [105, Fig. 1]), videos (e. g., the seminal "Powers of Ten" video [32] from 1977), and newer interactive experiences (e. g., [39]). Most illustrators use a smart composition of images blended such that the changes are (almost) unnoticeable, while some use clever perspectives to portray the differences in scale. These inspirations have prompted researchers in visualization to create similar multiscale experiences, based on real datasets.

The classification from Section 3.2.2 for molecular and genome visualization applies here as well. Everts et al. [34], e. g., *removed detail* from brain fiber tracts to observe the characteristics of the data at a higher scale. Hsu et al. [54] defined various cameras for a dataset, each showing a different level of detail. They then used image masks and camera ray interpolation to create smooth spatial scale transitions that show the data's multiscale character. Next, Glueck et al. [41]'s approach exemplifies the *change of shape representations by conceptual scale* by smoothly changing a multiscale coordinate grid and position pegs to aid depth perception and multiscale navigation of 3D scenes. They simply remove detail for scales that no longer contribute much to the visualization. In their accompanying video, interestingly, they limited the detail for each scale to only the focus point of the scale transition to maintain interactive frame rates. Another example of this category are geographic multiscale representations such as online maps (e. g., Google or Bing maps), which contain multiple scale representations, but typically toggle between them as the user zooms in or out. Finally, virtual globes are an example for *conceptual scale representations with smooth transitions*. They use smooth texture transitions to show

an increasing level of detail as one zooms in. Another example is Mohammed et al.'s [96] Abstractocyte tool, which depicts differently abstracted astrocytes and neurons. It allows users to smoothly transition between the cell-type abstractions using both geometry transformations and blending. We extend the latter to our visual embedding transition.

Also these approaches only cover a relatively small scale range. Even online map services cover less than approx. six orders of magnitude. Besides the field of bio-molecular and chemistry research discussed in Section 3.2.2, in fact, only astronomy deals with large scale differences. Here, structures range from celestial bodies ($\geqslant \approx 10^2$ m)[2] to the size of the observable universe ($1.3 \cdot 10^{26}$ m), in total 24 orders of magnitude.

To depict such data, visualization researchers have created explicit multiscale rendering architectures. Schatz et al. [113], for example, combined the rendering of overview representations of larger structures with the detailed depiction of parts that are close to the camera or have high importance. To truly traverse the large range of scales of the universe, however, several datasets that cover different orders of size and detail magnitude have to be combined into a dedicated data rendering and exploration framework. The first such framework was introduced by Fu et al. [37, 51] who used scale-independent modeling and rendering and power-scaled coordinates to produce scale-insensitive visualizations. This approach essentially treats, models, and visualizes each scale separately and then blends scales in and out as they appear or disappear. The different scales of entities in the universe can also be modeled using a *ScaleGraph* [66], which facilitates scale-independent rendering using scene graphs. Axelsson et al. [7] later extended this concept to the *Dynamic Scene Graph*, which, in the OpenSpace system [13], supports several high-detail locations and stereoscopic rendering. The Dynamic Scene Graph uses a dynamic camera node attachment to visualize scenes of varying scale and with high floating point precision.

With genome data we face similar problems concerning scale dependent data and the need to traverse a range of scales. We also face the challenge that our conceptual scales are packed much more tightly in scale space as we explain next. This leads to fundamental differences between both application domains.

## 3.3 MULTISCALE GENOME VISUALIZATION

Visualizing the nuclear human genome—from the nucleus that contains all chromosomal genetic material down to the very atoms that make up the DNA—is challenging due to the inherent organization of the DNA in tubular arrangements. DNA in its B-form is only 2 nm[4]

2 For example 25143 Itokawa, which was visited by the Hayabusa probe.

wide, which in its fibrous form or at more detailed scales would be too thin to be perceived. This situation is even more aggravated by the dense organization of the DNA and the structural hierarchy that bridges several scales. The previously discussed methods do not deal with such a combination of structural characteristics. Below we thus discuss the challenges that arise from the properties of these biological entities and how we address them by developing our new approach that smoothly transitions between views of the genome at its various scales.

### 3.3.1 *Challenges of interactive multiscale DNA visualization*

Domain scientists who sequence, investigate, and generally work with genome data use a series of conceptual levels for analysis and visualization [98]: the *genome* scale (containing all approx. 3.2 Gb of the human genome), the *chromosome* scale (50–100 Mb), the *loci* scale (in the order of Mb), the *fiber* scale (in the order of Kb), the *nucleosome* scale (146 b), and the *nucleotide* scale (i. e., 1 b), in addition to the *atomistic composition* of the nucleotides. These seven scales cover a range of approx. 4–5 orders of magnitude in physical size. In astronomy or astrophysics, in contrast, researchers deal with a similar number of scales:[3] approx. 7–8 conceptual scales of objects, yet over a range of some 24 orders of magnitude of physical size.[4] A fundamental difference between multiscale visualizations in the two domains is, therefore, the ***scale density of the conceptual levels*** that need to be depicted.

Multiscale astronomy visualization [7, 37, 51, 66] deals with ***positive-exponent scale-space***[5] (Figure 30, top), where two neighboring scales are relatively far apart in scale space. For example, planets are much smaller than stars, stars are much smaller than galaxies, galaxies are much smaller than galaxy clusters, etc. On average, two scales have a distance of three or more orders of magnitude in physical space. The consequence of this high distance in scale space between neighboring conceptual levels is that, as one zooms out, elements from one scale typically all but disappear before the elements on the next conceptual level become visible. This aspect is used in creating multiscale astronomy visualizations. For example, Axelsson et al.'s Dynamic Scene Graph [7] uses *spheres of influence* to control the visibility range of objects from a given subtree of the scene graph. In fact, the low scale density of the conceptual levels made the seamless animation of the astronomy/astrophysics section in the "Powers

3 Fu and Hanson [37] provide a nice overview in their Table 1.

4 We only count explicit objects, not distances between objects. We also include smaller asteroids in the order of $10^2$ m. And we use the size of the observable universe at $2 \times 13.8 \cdot 10^9$ light years $= 2.6 \cdot 10^{26}$ m.

5 Positive-exponent scale-space refers to measurement in meters, i. e., everything larger than approx. $1 \cdot 10^0$ m.

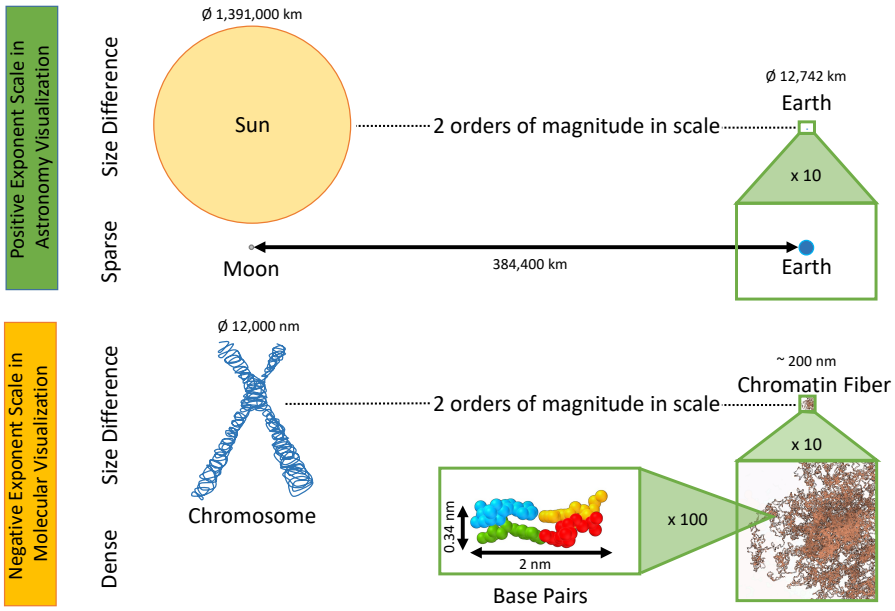



Figure 30: Multiscale visualization in astronomy vs. genomics. The size difference between celestial bodies is extremely large (e. g., sun vs. earth—the earth is almost invisible at that scale). The distance between earth and moon is also large, compared to their sizes. In the genome, we have similar relative size differences, yet molecules are densely packed as exemplified by the two base pairs in the DNA double helix.

of Ten" Video [32] from 1977 possible—in a time before computer graphics could be used to create such animations. Eames and Eames [32] simply and effectively blended smoothly between consecutive images that depicted the respective scales. For the cell/genome part, however, they use sudden transitions between conceptual scales without spatial continuity, and they also leave out several of the conceptual scales that scientists use today such as the chromosomes and the nucleosomes.

The reason for this problem of smoothly transitioning between scales in genome visualization—i. e., in ***negative-exponent scale-space***[6] (Figure 30, bottom)—is that the conceptual levels of a multiscale visualization are much closer to each other in scale. In contrast to astronomy's positive-exponent scale-space, there is only an average scale distance of about 0.5–0.6 orders of magnitude of physical space between two conceptual scales. Elements on one conceptual scale are thus still visible when elements from the next conceptual scale begin to appear. The scales for genome visualizations are thus much denser compared to astronomy's average scale distance of three orders of magnitude.

Moreover, in the genome the building blocks are ***physically connected in space and across conceptual scales***, except for the genome

6 This means, everything of size approx. $1 \cdot 10^{-1}$ m and smaller.

and chromosome levels. From the atoms to the chromosome scale, we have a single connected component. It is assembled in different geometric ways, depending on the conceptual scale at which we choose to observe. For example, the sequence of all nucleotides (base pairs) of the 46 chromosomes in a human cell would stretch for 2 m, with each base pair only being 2 nm wide [4], while a complete set of chromosomes fits into the 6 µm wide nucleus. Nonetheless, in all scales between the sequence of nucleotides and a chromosome we deal with the same, physically connected structure. In astronomy, instead, the physical space between elements within a conceptual scale is mostly empty and elements are physically not connected—elements are only connected by proximity (and gravity), not by visible links.

The large inter-scale distance and physical connectedness, naturally, also create the problem of how to visualize the ***relationship between two conceptual scale levels***. The mentioned multiscale visualization systems from astronomy [7, 37, 51, 66] use animation for this purpose, sometimes adding invisible and intangible elements such as orbits of celestial bodies. In general multiscale visualization approaches, *multiscale coordinate grids* [41] can assist the perception of scale-level relationships. These approaches only work if the respective elements are independent of each other and can fade visually as one zooms out, for example, into the next-higher conceptual scale. The connected composition of the genome does make these approaches impossible. In the genome, in addition, we have a ***complete model for the details in each conceptual level***, derived from data that are averages of measurements from many experiments on a single organism type. We are thus able to and need to show visual detail everywhere—as opposed to only close to a single point like planet Earth in astronomy.

Ultimately, all these points lead to **two fundamental challenges** for us to solve. The first (discussed in Section 3.3.2 and 3.3.3) is how to *visually* create effective transitions between conceptual scales. The transitional scales shall show the containment and relationship character of the data even in still images and seamlessly allow us to travel across the scales as we are interacting. They must deal with the continuous nature of the depicted elements, which are physically connected in space and across scales. The second challenge is a computational one. Positional information of all atoms from the entire genome would not fit into GPU memory and will prohibit interactive rendering performance. We discuss how to overcome these computational issues in Section 3.4, along with the implementation of the visual design from Section 3.3.2 and 3.3.3.

### 3.3.2 *Visual embedding of conceptual scales*

Existing multiscale visualizations of DNA [85, 90, 125] or other data [96] often use geometry manipulations to transition from one scale to the next. For the full genome, however, this approach would create too much detail to be useful and would require too many elements to be rendered. Moreover, two consecutive scales may differ significantly in structure and organization. A nucleosome, e. g., consists of nucleotides in double-helix form, wrapped around a histone protein. We thus need appropriate *abstracted representations* for the whole set of geometry in a given scale that best depict the scale-dependent structure and still allow us to create smooth transitions between scales.

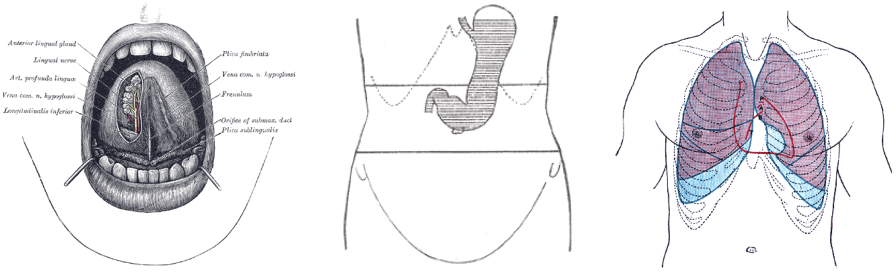


Figure 31: Plates 1013, 1048, and 1216 from *Gray's Anatomy* [47], demonstrate layered composition of multiscale 3D objects by traditional illustrators. The images are in the public domain.

Nonetheless, the mentioned geometry-based multiscale transformations still serve as an important inspiration to our work. They often provide intermediate representations that may not be entirely accurate, but show how one scale relates to another one, even in a still image. Viewers can appreciate the properties of both involved scale levels, such as in Miao et al.'s [90] transition between nucleotides and strands.

Specifically, we take inspiration from traditional illustration where a related visual metaphor has been used before. As exemplified by Figure 31, illustrators sometimes use an abstracted representation of a coarser scale to aid viewers with understanding the overall composition as well as the spatial location of the finer details. This embedding of one representation scale into the next is similar to combining several layers of visual information—or super-imposition [97, pp. 288 ff]. It is a common approach, for example, in creating maps. In visualization, this principle has been used in the past (e. g., [21, 55, 110, 112]), typically applying some form of transparency to be able to perceive the different layers. Transparency, however, can easily lead to visualizations that are difficult to understand [20]. Simple outlines to indicate the coarser shape or context can also be useful [120]. In our case, even outlines easily lead to clutter due to the immense amount of detail in the genome data. Moreover, we are not interested in showing that some elements are *spatially* inside others, but rather that the

elements are *part of a higher-level structure*, thus are *conceptually* contained.

Specifically, we take inspiration from traditional illustration where a related visual metaphor has been used before. As exemplified by Figure 31, illustrators sometimes use an abstracted representation of a coarser scale to aid viewers with understanding the overall composition as well as the spatial location of the finer details. This embedding of one representation scale into the next is similar to combining several layers of visual information—or super-imposition [97, pp. 288 ff]. It is a common approach, for example, in creating maps. In visualization, this principle has been used in the past (e. g., [21, 55, 110, 112]), typically applying some form of transparency to be able to perceive the different layers. Transparency, however, can easily lead to visualizations that are difficult to understand [20]. Simple outlines to indicate the coarser shape or context can also be useful [120]. In our case, even outlines easily lead to clutter due to the immense amount of detail in the genome data. Moreover, we are not interested in showing that some elements are *spatially* inside others, but rather that the elements are *part of a higher-level structure*, thus are *conceptually* contained.

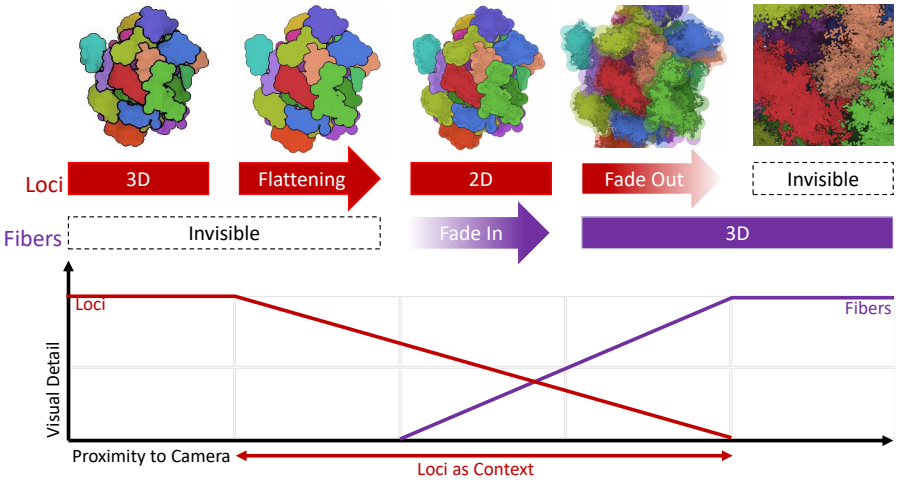
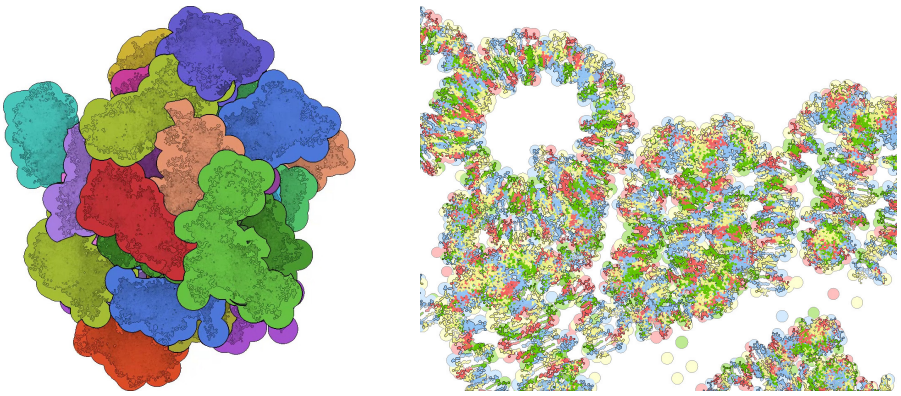



Figure 32: Visual embedding, schematic principle.

We therefore propose ***visual scale embedding*** of the detailed scale into its coarser parent (see the illustration in Figure 32). We render an abstracted representation of the coarser scale first to serve as a context or canvas, and render the representation of the more detailed scale on top of it. The context or canvas should not interfere in its spatial perception with the depiction of the detail because it is typically *surrounding* the next scale. An exemplifying case of how this can lead to perception issues in still images was given by Svetachov et al. [116, Fig. 10]. We thus completely flattened the context as shown in Figure 32 and inspired by previous multiscale visualizations from structural biology [103]. Then we render the detailed geometry of

the next-smaller scale on top of it. This concept adequately supports our goal of smooth scale transitions. A geometric representation of the coarser scale is first shown using 3D shading as long as it is still small on the screen, i. e., the camera is far away. It transitions to a flat, canvas-like representation when the camera comes closer and the detail in this scale is not enough anymore. We now add the representation of the more detailed scale on top—again using 3D shading, as shown for two scale transitions in Figure 33.



(a) (b)

Figure 33: Two snapshots of scale transition views, (a) between the chromosome and the detailed chromosome scales, as well as (b) between nucleotides and detailed nucleotides scales.

Our illustrative visualization concept combines the 2D aspect of the flattened coarser scale with the 3D detail of the finer scale. With it we make use of superimposed representations as argued by Viola and Isenberg [128], which are an alternative to spatially or temporally juxtaposed views. In our case, the increasingly abstract character of rendering of the coarser scale (as we flatten it during zooming in) relates to its increasingly contextual and conceptual nature. Our approach thus relates to *semantic zooming* [106] because the context layer turns into a flat surface or canvas, irrespective of the underlying 3D structure and regardless of the specific chosen view direction. This type of scale zoom does not have the character of cut-away techniques as often used in tools to explore containment in 3D data (e. g., [77, 81]). Instead, it is more akin to the semantic zooming in the visualization of abstract data, which is embedded in the 2D plane (e. g., [132]).

### 3.3.3 *Multiscale visual embedding and scale-dependent view*

One visual embedding step connects two consecutive semantic scales. We now concatenate several steps to assemble the whole hierarchy (Figure 34). This is conceptually straightforward because each scale by itself is shown using 3D shading. Nonetheless, as we get to finer

and finer details, we face the two major problems mentioned at the start of Section 3.3.2: visual clutter and limitations of graphics processing. Both are caused by the tight scale space packing of the semantic levels in the genome. At detailed scales, a huge number of elements are potentially visible, e. g., 3.2 Gb at the level of nucleotides. To address this issue, we adjust the camera concept to the multi-scale nature of the data.

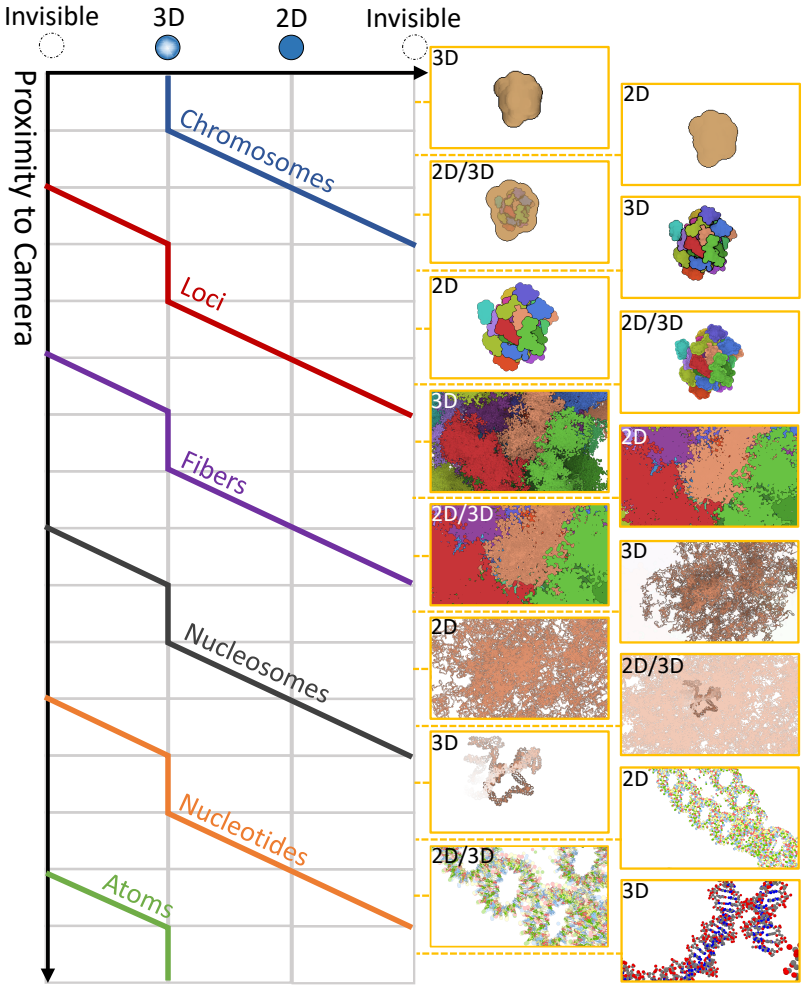



Figure 34: Sequence of visual scale embeddings, based on the data levels.

In previous multiscale visualization frameworks [7, 37, 51, 66], researchers have already used ***scale-constrained camera navigation***. For example, they apply a ***scale-dependent camera speed*** to quickly cover the huge distances at coarse levels and provide fine control for detailed levels. In addition, they used a ***scale-dependent physical camera size or scope*** such that the depicted elements would appropriately fill the distance between near and far plane, or use depth buffer remapping [37] to cover a larger depth range. In astronomy and astrophysics, however, we do not face the problem of a lot of nearby elements in detailed levels of scale due to their loose scale-

space packing. After all, if we look into the night sky we do not see much more than "a few" stars from our galactic neighborhood which, in a visualization system, can easily be represented by a texture map. Axelsson et al. [7], for example, simply attach their cameras to nodes within the scale level they want to depict.

For the visualization of genome data, however, we have to introduce an active control of the ***scale-dependent* data-hierarchy *size or scope*** as we would "physically see", for example, all nucleosomes or nucleotides up to the end of the nucleus. Aside from the resulting clutter, such complete genome views would also *conceptually* not be helpful because, due to the nature of the genome, the elements within a detailed scale largely repeat themselves. The visual goal should thus be to only show a relevant and scale-dependent subset of each hierarchy level. We thus limit the rendering scope to a subset of the hierarchy, depending on the chosen scale level and spatial focus point. The example in Figure 35 depicts the nucleosome scale, where we only show a limited number of nucleosomes to the left and the right of the current focus point in the sequence, while the rest of the hierarchy has been blended out. We thereby extend the visual metaphor of the canvas, which we applied in the visual embedding, and use the white background of the frame buffer as a second, ***scale-dependent canvas***, which limits the visibility of the detail. In contrast to photorealism[7] that drives many multi-scale visualizations in astronomy, we are interested in appropriately abstracted representations through a ***scale-dependent removal of distant detail*** to support viewers in focusing on their current region of interest.

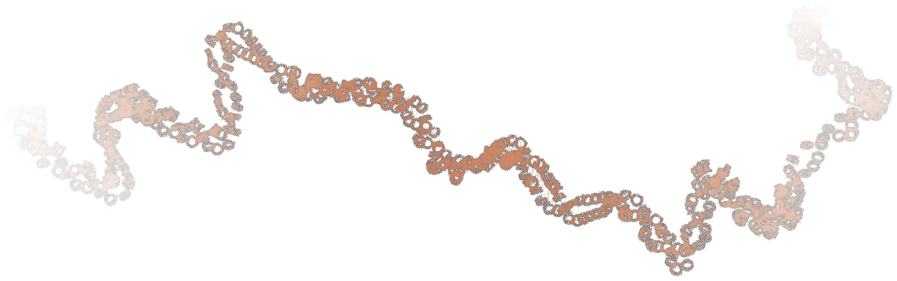


Figure 35: Scale-dependent hierarchy scope realized for nucleosomes by showing five *fiber* locations around the focus and fading out the ends.

## 3.4 IMPLEMENTATION

Based on the conceptual design from Section 3.3 we now describe the implementation of our multi-scale genome visualization framework.

7 Of course, visualizations in astronomy also comprise non-photographic components such as hyperspectral imaging or radio astronomy data, but many scales use depictions based on a photographic camera as their guiding vision.

We first describe the used and then explain the shader-based realization of the scale transitions using a series of visual embedding steps as well as some interaction considerations.

### 3.4.1 *Data sources and data hierarchy*

Researchers in genome studies have a high interest in understanding the relationships between the spatial structure at the various scale levels and the biological function of the DNA. Therefore they have created a multi-scale dataset that allows them to look at the genome in different spatial scale levels [98]. This data was derived by Nowotny et al. [98] from a model of the human genome by Asbury et al. [5], which in turn was constructed based on various data sources and observed properties. For determining the positions of the chromatin fiber, Nowotny et al. used Bancaud et al.'s [9] approach of space-filling, fractal packing. As a result, Nowotny et al. [98] obtained the positions of the *nucleotides* in space, and from these computed the positions of *fibers*, *loci*, and *chromosomes* (Figure 36). They stored this data in their own Genome Scale System (GSS) format and also provided the positions of the *nucleotides* for one nucleosome (Figure 36, bottom-right). Even with this additional data, we still have to procedurally generate further information as we visualize this data such as the orientations of the nucleosomes (based on the location of two consecutive nucleosomes) and the linker DNA strands of nucleotides connecting two consecutive nucleosomes.

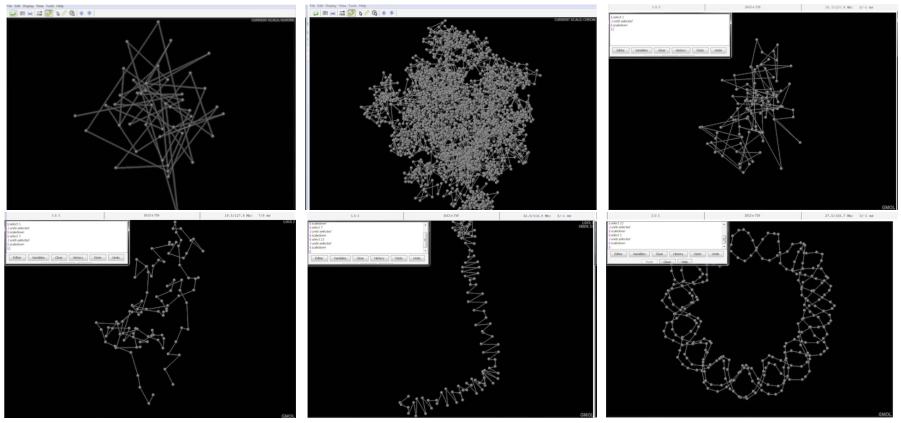
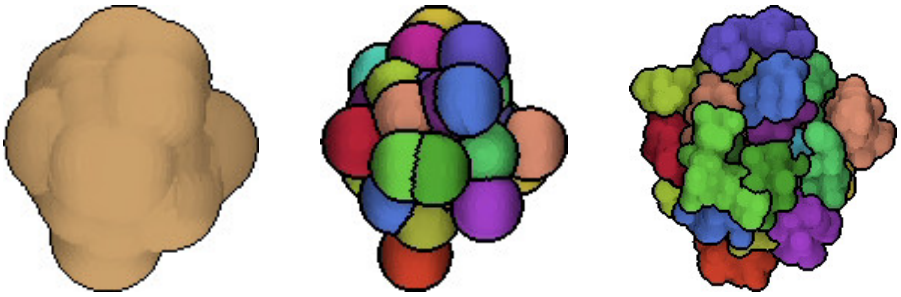


Figure 36: Screenshots from GMOL showing traditional visualizations of the multi-scale genome data by depicting chromosomes, loci of all chromosomes, loci of a single chromosome, fibers, nucleosomes, and nucleotides. Images from Nowotny et al. [98] ().

This data provides positions at every scale level, without additional information about the actual sizes. Only at the nucleotide and atom scales the sizes are known. It was commonly thought that nucleosomes are tightly and homogeneously packed into 30 nm fibers,

120 nm chromonema, and 300–700 nm chromatids, but recent studies [101] disprove this organization and confirm the existence of flexible chains with diameters of 5–24 nm. Therefore, for all hierarchically organized scales coarser than the nucleosome, we do not have information about the specific shape that each data point represents. We use spheres with scale-adjusted sizes as rendering primitives as they well portray the chaining of elements according to the data-point sequence.

With respect to visualizing this multi-scale phenomenon, the data hierarchy (i. e., 100 nucleosomes = 1 fiber, 100 fibers = 1 locus, approx. 100 loci = 1 chromosome) is not the same as the hierarchy of semantic scales that a viewer sees. For example, the dataset contains a level that stores the *chromosome* positions, but if rendered we would only see one sphere for each chromosome (Figure 37b). Such a depiction would not easily be recognized as representing a chromosome due to the lack of detail. The chromosomes by themselves only become apparent once we display them with more shape details using the data level of the *loci* as given in Figure 37c. The locations at the chromosomes data scale can instead be better used to represent the semantic level of the *nucleus* by rendering them as larger spheres, all with the same color and with a single outline around the entire shape as illustrated in Figure 37a.



(a) (b) (c)

Figure 37: Mismatch of the data scale and the semantic scale: the *chromosome* locations are not easily recognized as chromosomes in (b), instead we display them using a single color as in (a) to represent the semantic *nucleus* scale. The chromosomes and their three-dimensional shape are better shown using data from the more detailed *loci* scale (c).

In Table 1 we list the relationships between data hierarchy and semantic hierarchy for the entire set of scales we support. From the table follows that the choice of color assignment and the subset of rendered elements on the screen supports viewers in understanding the semantic level, which we want to portray. For example, by rendering the *fiber* positions colored by *chromosome* we facilitate the understanding of a detailed depiction of a chromosome, rather than that chromosomes consist of several loci. In an alternative depiction for

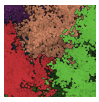
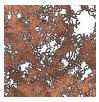
| image | data level | colored by | elements rendered | semantic level | transition |
|---|---|---|---|---|---|
|  | *chromosome* positions | one single color | all chromosomes | nucleus | visual embedding |
|  | *loci* positions | chromosome | all chromosomes | chromosome | visual embedding |
|  | *fiber* positions | chromosome | all chromosomes | chromosome with detail | visual embedding |
|  | *nucleosome* positions | chromosome | focus chromosome | fibers | visual embedding |
|  | *nucleotide* positions | chromosome | nucleosomes of 5 fibers, with links | nucleosomes | color change |
|  | *nucleotide* positions | nucleotide | nucleosomes of 5 fibers | nucleotides | visual embedding |
|  | *atom* positions | nucleotide | nucleosomes of 5 fibers | nucleotides with detail | color change |
|  | *atom* positions | element | nucleosomes of 5 fibers | individual atoms | |

Table 1: Relationship between data and semantic hierarchies.

domain experts, who are interested in studying the loci regions, we could instead assign the colors by *loci* for the *fiber* data level and beyond.

We added two additional scale transitions that are not realized by visual embedding, but instead by color transitions. The first of these transitions changes the colors from the previously maintained chromosome color to nucleotide colors as the *nucleotide* positions are rendered in their 3D shape to illustrate that the *nucleosomes* themselves consist of pairs of *nucleotides*. The following transition then uses visual embedding as before, to transition to *atoms* while maintaining nucleotide colors. The last transition, again changes this color assignment such that the atoms are rendered in their typical element colors, using 3D shading and without flattening them.

### 3.4.2 *Realizing visual scale embedding*

For our proof-of-concept implementation we build on the molecular visualization functionality provided in the Marion framework [94]. We added to this framework the capability to load the previously described GSS data. We thus load and store the highest detail of the

data—the 23,958,240 nucleosome positions—as well as all positions of the coarser scales. To show more detail, we use the single nucleosome example in the data, which consists of 292 nucleotides and then create the $\approx 24 \cdot 10^6$ instances for the semantic nucleosome scale. Here we fully use of Le Muzic et al.'s [76] technique of employing the tessellation stages on the GPU [8], which dynamically injects the atoms of the nucleosome. We apply a similar instancing approach for transitioning to an atomistic representation, based on the 1AOI model from the PDB. To visually represent the elements, we utilize 2D sphere impostors instead of sphere meshes [76]. Specifically, we use triangular 2D billboards (i. e., only three vertices) that always face the camera and assign the depth to each fragment that it would get if it had been a sphere.

If we wanted to directly render all atoms at the finest detail scale, we would have to deal with $\approx 3.2\,\text{Gb} \cdot 70\,\text{atoms/b} = 224 \cdot 10^9$ atoms. This amount of detail is not possible to render at interactive rates. With LOD optimizations, such as the creation of super-atoms for distant elements, cellVIEW could process $15 \cdot 10^9$ atoms at 60 Hz [76]. This amount of detail does not seem to be necessary in our case. Our main goal is the depiction of the scale transitions and too much detail would cause visual noise and distractions. We use the *scale-dependent removal of distant detail* described in Section 3.3.3. As listed in Table 1, for coarse scales we show all chromosomes. Starting with the semantic fibers scale, we only show the focus chromosome. For the semantic nucleosomes level, we only show the focus fiber and two additional fibers in both directions of the sequence. To indicate that the sequence continues, we gradually fade out the ends of the sequence of nucleosomes as shown in Figure 35. For finer scales beyond the nucleosomes, we maintain the sequence of five fibers around the focus point, but remove the detail of the links between nucleosomes.

To manage the different rendering scopes and color assignments, we assign IDs to elements in a data scale and record the IDs of the hierarchy ancestors of an element. For example, each *chromosome* data element gets an ID, which in turn is known to the *loci* data instances. We use this ID to assign a color to the chromosomes. Because we continue rendering all chromosomes even at the *fiber* data level respectively semantic *chromosome with detail* level, we also pass the IDs of the chromosomes to the *fiber* data elements. Later, the IDs of the *fiber* data elements are used to determine the rendering scope in the data levels of *nucleotide* positions and finer (more detail).

For realizing the transition in the *visual scale embedding*, i. e., transitioning from the coarser scale $S_N$ to the finer scale $S_{N+1}$, we begin by alpha-blending $S_N$ rendered with 3D detail and flattened $S_N$. We achieve the 3D detail with screen-space ambient occlusion (SSAO),

8 This technique builds on earlier work by Gumhold [48], Klein and Ertl [67], Reina and Ertl [111], etc.

while the flattened version does not use SSAO. Next we transition between $S_N$ and $S_{N+1}$ by first rendering $S_N$ and then $S_{N+1}$ on top, the latter with increasing opacity. Here we avoid visual clutter by only adding detail to elements in $S_{N+1}$ on top of those regions that belonged to their parents in $S_N$. The necessary information for this purpose comes from the previously mentioned IDs. We thus first render all flattened elements of $S_N$, before blending in detail elements from $S_{N+1}$. In the final transition of *visual scale embedding*, we remove the elements from $S_N$ through alpha-blending. For the two color transitions discussed in Section 3.4.1 we simply alpha-blend between the corresponding elements of $S_N$ and $S_{N+1}$, but with different color assignments.

### 3.4.3 *Interaction considerations*

The rendering speeds are in the range of 15–35 fps on an Intel Core™ PC (i7-8700K, 6 cores, 32 GB RAM, 3.70 GHz, nVidia Quadro P4000, Windows 10 x64). In addition to providing a scale-controlled traversal of the scale hierarchy toward a focus point, we thus allow users to interactively explore the data and choose their focus point themselves. To support this interaction, we allow users to apply transformations such as rotation and panning. We also allow users to click on the data to select a new focus point, which controls the removal of elements to be rendered at specific scale transitions (as shown in Table 1). First, users can select the focus chromosome (starting at *loci* positions), whose position is the median point within the sequence of fiber positions for that chromosome. This choice controls which chromosome remains as we transition from the *fiber* to the *nucleosome* data scale. Next, starting at the *nucleosome* data scale, users can select a strand of five consecutive *fiber* positions, which then ensures that only this strand remains as we transition from *nucleosome* to *nucleotide* positions.

To further support the interactive exploration, we also adjust the colors of the elements to be in focus next. For example, the subset of a chromosome next in focus is rendered in a slightly lighter color than the remaining elements of the same level. This approach provides a natural visual indication of the current focus point and guides the view of the users as they explore the scales.

To achieve the *scale-constrained camera navigation*, we measure the distance to a transition or interaction target point in the data sequence. We measure this distance as the span between the camera location and the position of the target level in its currently active scale. This distance then informs the setting of camera parameters and SSAO passes. After the user has selected a new focus point, the current distance to the camera will change, so we adjust also the global scale parameter that we use to control the scale navigation.

## 3.5 DISCUSSION

Based on our design and implementation we now compare our results with existing visual examples, examine potential application domains, discuss limitations, and suggest several directions for improvement.

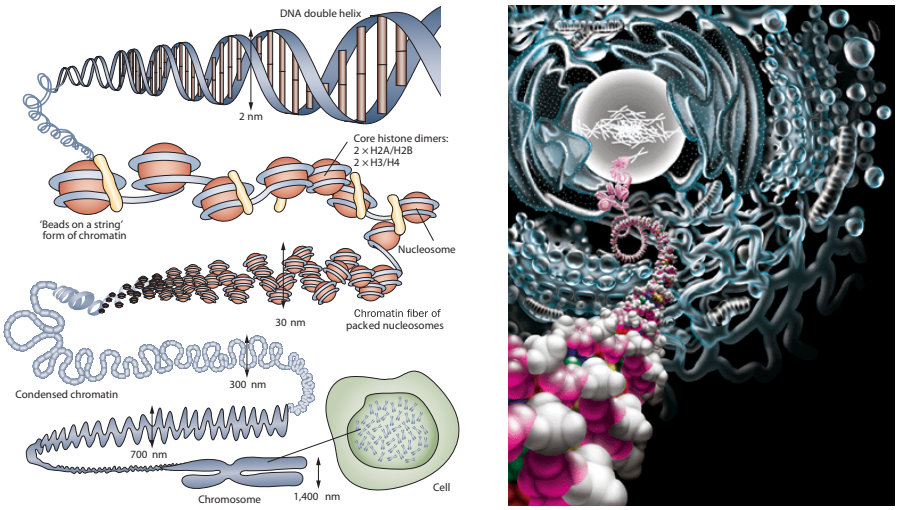
### 3.5.1 *Comparison to traditionally created illustrations*

Measuring the ground truth is only possible to a certain degree, which makes the comparison to ScaleTrotter difficult. One reason is that no static genetic material exists in living cells. Moreover, microscopy is also limited at the scale levels with which we are dealing. We have to rely on the data from the domain experts with its own limitations (Section 3.5.4) as the input for creating our visualization and compare the results with existing illustrations in both static and animated form.

We first look at traditional static multi-scale illustrations as shown in Figure 38; other illustrations similar to the one in Figure 38a can be found in Annunziato's [4] and Ou et al.'s [101] works. In Figure 38a, the illustrators perform the scale transition along a 1D path, supported by the DNA's extreme length. We do not take this route as we employ the actual positions of elements from the involved datasets. This means that we could also apply our approach to biologic agents such as proteins that do not have an extremely long extent. Moreover, the static illustrations have some continuous scale transitions, e. g., the detail of the DNA molecule itself or the sizes of the nucleosomes. Some transitions in the multi-scale representation, however, are more sudden such as the transition from the DNA to nucleosomes, the transition from the nucleosomes to the condensed chromatin fiber, and the transition from that fiber to the 700 nm wide chromosome leg. Figure 38b has only one such transition. The changeover happens directly between the nucleosome level and the mitotic chromosome. We show transitions between scales interactively using our *visual scale embedding*. The static illustrations in Figure 38 just use the continuous nature of the DNA to evoke the same hierarchical layering of the different scales. The benefit of the spatial scale transitions in the static illustrations is that a single view can depict all scale levels, while our temporally-controlled scale transitions allow us to interactively explore any point in both the genome's spatial layout and in scale. Moreover, we also show the actual physical configuration of every scale according to the datasets that genome researchers provide, representing the current state of knowledge.

We also compare our results to animated illustrations as exemplified by the "Powers of Ten" video[9] [32] and a video treating the

9 http://www.eamesoffice.com/the-work/powers-of-ten/




(a) (b)

Figure 38: Artistic depictions of image-spatial scale transitions for genome data. Images from [121]/[105] and © Springer Nature/The American Association for the Advancement of Science, respectively, used with permission.

composition of the genome[10] and created by Drew Berry et al. in 2003. The "Powers of Ten" video only shows the fibers of the DNA double helix curled into loops—a notion that has since been revised by the domain experts. Nonetheless, the video still shows a continuous transition in scale through blending of aligned representations from the fibers, to the nucleotides, to the atoms. It even suggests that we should continue the scale journey beyond the atoms. The second video, in contrast, shows the scale transitions starting from the DNA double helix and zooming out. The scale transitions are depicted as "physical" assembly processes, e. g., going from the double helix to nucleosomes, and from nucleosomes to fibers. Furthermore, shifts of focus or hard cuts are applied as well. The process of assembling an elongated structure through curling up can nicely illustrate the composition of the low-level genome structures, but only if no constraints on the rest of the fibrous structure exist. In our interactive illustration, we have such constraints where we can zoom out and in and where we have restrictions on the locations of all elements coming from the given data. Moreover, the construction also potentially creates a lot of motion due to the dense nature of the genome and, thus, visual noise which might impact the overall visualization. On the other hand, both videos convey the message that no element is static at the small scales. We do not yet show this functionality in our visualizations.

Both static and dynamic traditional visualizations depict the composition of the genome in its mitotic stage. The chromosomes only

10 http://www.wehi.edu.au/wehi-tv/molecular-visualisations-dna

assume this stage, however, when the cell divides. Our visualization is the first that provides the user with an interactive exploration with smooth scale transitions of the genome in its interphase state, the state in which the chromosomes exist most of the time.

### 3.5.2 *Feedback from illustrators and application scenarios*

To discuss the creation of illustrations for laypeople with ScaleTrotter, we asked two professional illustrators for feedback who work on biological and medical visualizations. One of them has ten years experience as a professional scientific illustrator and animator with a focus on biological and medical illustrations for science education. The other expert is a certified illustrator with two years experience plus a PhD in Bioengineering. We conducted a semi-structured interview (approx. 60 min) with them, to get critical feedback [61, 70] on our illustrative multi-scale visualization and to learn how our approach compares to the way they deal with multi-scale depictions in their daily work.

They immediately considered our ScaleTrotter approach for showing genome scale transitions as part of a general story to tell. They missed the necessary additional support for telling a story such as the contextual representation of a cell (for which we could investigate cellVIEW [76]) and, in general, audio support and narration. Although they had to judge our results isolated from other story telling methods, they saw the benefits of an interactive tool for creating narratives that goes beyond the possibilities of their manual approaches.

We also got a number of specific pieces of advice for improvement. In particular, they recommended different settings for when to make certain transitions in scale space. The illustrators also suggested the addition of "contrast" for those parts that will be in focus next as we zoom in—a feature we then added and describe in Section 3.4.3.

According to them, our concept of using *visual scale embedding* to transition between different scalar representations has not yet been used in animated illustrations, yet the general concept of showing detail together with context as illustrated in Figure 31 is known. Instead of using visual scale embedding, they use techniques discussed in Section 3.5.1, or they employ cut-outs with rectangles or boxes to indicate the transition between scales. Our *visual scale embedding* is seen by them as a clear innovation: "to have a smooth transition between the scales is really cool." Moreover, they were excited about the ability to freely select a point of focus and interactively zoom into the corresponding detail. Basically, they said that our approach would bring them closer to their vision of a "molecular Maya" because it is "essential to have a scientifically correct reference." Connected to this point we also discussed the application of ScaleTrotter in genome research. Due to their close collaborations with domain experts they

emphasized that the combination of the genomics sequence data plus some type of spatial information will be essential for future research. A combination of our visualization, which is based on the domain's state-of-the-art spatial data, with existing tools could allow genome scientists to better understand the function of genes and certain genetic diseases.

In summary, they are excited about the visual results and see application possibilities both in teaching and in data exploration.

### 3.5.3 *Feedback from genome scientists*

As a result of our conversation with the illustrators they also connected us to a neurobiologist who investigates 3D genome structures at single cell levels, e. g., by comparing cancerous with healthy cells. His group is interested in interactions between different regions of the genome. Although the spatial characteristics of the data are of key importance to them, they still use 2D tools. The scientist confirmed that a combination of their 2D representations with our interactive 3D-spatial multi-scale method would considerably help them to understand the interaction of sequentially distant but spatially close parts of the genome, processes such as gene expression, and DNA-protein interactions.

We also presented our approach to an expert in molecular biology with 52 years of age and 22 years of post-PhD experience. He specializes in genetics and studies the composition, architecture, and function of SMC complexes. We conducted a semi-structured interview (approx. 60 minutes) to discuss our results. He stated that transitions between several scales are definitely useful for analyzing the 3D genome. He was satisfied with the coarser chromosomes and loci representations, but had suggestions for improving the nucleosome and atomic scales. In particular, he noted the lack of proteins such as histones. He compared our visualization with existing electron microscopy images [100, 101], and suggested that a more familiar *filament-like* representation could increase understandability. In his opinion, some scale transitions happened too early (e. g., the transition from chromosome-colored to nucleotide-colored nucleotides). We adjusted our parametrization accordingly. In addition, based on his feedback, we added an interactive *scale offset* control that now allows users to adjust the scale representation for a given zoom level. This offset only adapts the chosen representation according to Table 1, while leaving the size on the screen unchanged. The expert also suggested to build on the current approach and extend it with more scales, which we plan to do in the future. Similar to the feedback from the neurobiologist, also the molecular biologist agrees that an integration with existing 2D examination tools has a great potential to improve the workflow in a future visualization system.

### 3.5.4 *Limitations*

There are several limitations of our work, the first set relating to the source data. While we used actual data generated by domain experts based on the latest understanding of the genome, it is largely generated using simulations and not actual measurements (Section 3.4.1. We do not use actual sequence data at the lowest scales. Moreover, our specific dataset only contains 45 chromosomes, instead of the correct number of 46. We also noticed that the dataset contains 23,958,240 nucleosome positions, yet when we multiply this with the sum of 146 base pairs per nucleosome we arrive at $\approx$ 3.5 Gb for the entire genome—not even including the linker base pairs in this calculation and for only 45 chromosomes. Ultimately better data is required. The overall nature of the visualization and the scale transitions would not be impacted by the modified data and we believe that the data quality is already sufficient for general illustration and teaching purposes.

Another limitation is the huge size of the data. Loading all positions for the interactive visualization takes approx. two minutes, but we have not yet explored the feasibility of also loading actual sequence data. We could investigate loading data on-demand for future interactive applications, in particular in the context of tools for domain experts. For such applications we would also likely have to reconsider our design decision to leave out data in the detailed scales, as these may interact with the parts that we do show. We would need to develop a space-dependent look-up to identify parts from the entire genome that potentially interact with the presently shown focus sequences. Another limitation relates to the selection of detail to zoom into. At the moment, we determine the focus interactively based on the currently depicted scale level. This makes it, for example, difficult to select a chromosome deep inside the nucleus or fibers deep inside a chromosome. A combination with an abstract data representation—for example with a domain expert sequencing tool—would address this problem.

### 3.5.5 *Future work*

Beyond addressing the mentioned issues, we would like to pursue a number of additional ideas in the future. A next step towards adoption of our approach in biological or medical research is to build an analytical system on top of ScaleTrotter that allows us to query various scientifically relevant aspects. As noted in Section 3.5.2, one scenario are spatial queries to determine whether two genes are located in a close spatial vicinity in case they somehow are related. Other visualization systems developed in the past for analyzing gene expressions can benefit from the structural features that ScaleTrotter offers.

Extending to other subject matters, we will also have to investigate scale transitions where the scales cannot be represented with sequences of blobs. For example, can we also use linear or volumetric representations and extend our visual space embedding to such structures? Alternatively, can we find more effective scale transitions to use such as geometry-based ones (e. g., [85, 90, 125]), in addition to the visual embedding and the color changes we use so far? We have to avoid over-using the visual variable color which is a scarce resource. Many elements could use color at different scales, so dynamic methods for color management will be essential.

Another direction for future research are generative methods for completing the basic skeletal genetic information on the fly. Currently we use data that are based on positions of nucleotides, while higher-level structures are constructed from these. Information about nucleotide orientations and their connectivity is missing, as well as the specific sequence which is currently not derived from real data. ScaleTrotter does not contain higher-level structures and protein complexes that hold the genome together and which would need to be modeled with a strict scientific accuracy in mind. An algorithmic generation of such models from Hi-C data would allow biologists to adjust the model parameters according to their mental model, and would give them a system for generating new hypotheses. Such a generative approach would also integrate well with the task of adding processes in which involve the DNA, such as condensation, replication, and cell division.

A related fundamental question is how to visualize the dynamic characteristics of the molecular world. It would be highly useful to portray the transition between the interphase and the mitotic form of the DNA, to support visualizing the dynamic processes of reading out the DNA, and to even show the Brownian motion of the atoms.

Finally, our visualization relies on dedicated decisions of how to parameterize the scale transitions. While we used our best judgment to adjust the settings, the resulting parameterization may not be universally valid. An interactive illustration for teaching may need parameters different from those in a tool for domain experts. It would be helpful to derive templates that could be used in different application contexts.

## 3.6 CONCLUSION

ScaleTrotter constitutes one step towards understanding the mysteries of human genetics—not only for a small group of scientists, but also for larger audiences. It is driven by our desire as humans to understand *"was die Welt im Innersten zusammenhält"* [what "binds the world, and guides its course"] [43]. We believe that our visualization has the potential to serve as the basis of teaching material about the

genome and part of the inner workings of biologic processes. It is intended both for the general public and as a foundation for future visual data exploration for genome researchers. In both cases we support, for the first time, an interactive and seamless exploration of the full range of scales—from the nucleus to the atoms of the DNA.

From our discussion it became clear that such multi-scale visualizations need to be created in a fundamentally different way as compared to those excellent examples used in the astronomy domain. In this paper we thus distinguish between the *positive-exponent scale-space* of astronomy (looking inside-out) and the negative-exponent scale-space of genome data (looking outside-in). For the latter we provide a multi-scale visualization approach based on *visual scale embedding*. We also discuss an example on how the controlled use of abstraction in (illustrative) visualization allows us to employ a space-efficient superimposition of visual representations. This is opposed to juxtaposed views [128], which are ubiquitous in visualization today.

A remaining question is whether the tipping point between the different types of scale spaces is really approximately 1 meter ($1 \cdot 10^0$ m) or whether we should use a different point in scale space such as 1 mm. The answer to this question requires further studies on how to illustrate multi-scale subject matter. An example is to generalize our approach to other biologic phenomena such as mitotic DNA or microtubules as suggested in Section 3.5.5. If we continue our journey down the negative-exponent scale-space we may discover a third scale-space region. Models of atoms and subatomic particles seem to again comprise much empty space, similar to the situation in the positive-exponent scale-space. A bigger vision of this work thus is to completely replicate the "Powers of Ten" video—the 36 orders of magnitude from the size of the observable universe to sub-atomic particles—but with an interactive tool and based on current data and visualizations.

# 4

# MULTISCALE UNFOLDING: ILLUSTRATIVELY VISUALIZING THE WHOLE GENOME AT A GLANCE

In this chapter, we take the second step towards exploration of the human genome, by addressing aspects that were not the focus of Chapter 3. We use a spatial abstraction axis to explore the linear structure of the genome. We thus present Multiscale Unfolding, an interactive technique for visualizing multiple hierarchical scales of DNA in a single view, which allows viewers to explore genome data at different scales and understand how one scale is spatially folded into the next. The DNA's extremely long sequential structure that is arranged differently on several different scale levels is often lost in traditional 3D representations, mainly due to its multiple levels of dense spatial packing and the resulting occlusion. Furthermore, interactive exploration of this complex structure is cumbersome, requiring visibility management like cut-aways. In contrast to existing temporally controlled multiscale data exploration, we allow viewers to always see and interact with any of the involved scales. For this purpose we separate the depiction into constant-scale and transitioning-scale zones. Constant-scale zones maintain a single scale representation, while still linearly unfolding the DNA. Using inspiration from illustration, transitioning-scale zones transition between adjacent constant-scale zones using scaling, transparency, and level unfolding. We thus unfold the spatial structure of the long DNA macro-molecule, maintain its local organizational characteristics, linearize its higher-level organization, and use spatially controlled, understandable interpolation between neighboring scales. We also contribute interaction techniques that provide viewers with a coarse-to-fine control for navigating within our all-scales-in-one-view visualizations and visual aids to illustrate the size differences. Overall, Multiscale Unfolding allows viewers to grasp the DNA's structural composition from chromosomes to the atoms, with increasing levels of "unfoldedness", and can be applied in data-driven illustration and communication as well as as a basis for visualizations for domain experts.

Main portions of this chapter are from an article accepted for publication in TVCG [49]. Therefore, any use of "we" in this chapter refers to myself, David Kouřil, Haichao Miao, M. Eduard Gröller, Ivan Viola, and Tobias Isenberg.

## 4.1 INTRODUCTION

Exploration of 3D data is essential for all physical sciences, from the smallest phenomena in particle physics to the largest known structures in astrophysics. In many of the sciences in this spectrum (including biology) and applied fields such as medicine and engineering, it is not only important to understand data at a single spatial scale. Instead, many phenomena require a multiscale data analysis. Within visualization, researchers have contributed to this analysis by providing several means for the exploration of multiscale data (e. g., [7, 13, 37, 50, 66, 71, 96, 103, 113]), for many of the mentioned application domains.

To date, most of the proposed data exploration strategies follow a temporal control in the data navigation. At any given time, the visual representation shows a subset of the 3D dataset at a single spatial scale factor, and this scale factor can be changed interactively, i. e., over time. This strategy has the benefit that it allows viewers to experience and explore the actual 3D spatial structure of the data. But, it requires them to examine the data's scale changes sequentially and, thus, requires them to memorize the different spatial scale configurations for comparison. Moreover, if data has a complex spatial configuration at a given scale such as the heavily intertwined DNA then it is difficult to make such a comparison—even if both scales would be shown side-by-side.

An alternative way of exploring multiscale data is to use a spatial control of scale. The visual representation of different scale configurations are shown at different locations in a single image. Illustrators have used this strategy in the past to depict the different scale aspects of multiscale data such as the DNA, anatomy structures, and of astrophysics (e. g., Figure 40). This paradigm has the advantage that it allows us to explain—in particular to non-experts and in general education—the inter-connectedness of the scales as well as (local examples of) their spatial configurations, all in one view. This technique is particularly well suited for phenomena in which the spatial scales are, in fact, physically connected like in genetic macro-molecules (i. e., DNA, RNA).

In this paper we thus take inspiration from static illustrations and present a framework that allows us to interactively create and explore spatially-controlled multiscale visualizations based on captured and simulated genome data (Figure 39). We demonstrate how to unfold the data according to their scales and depict them following chosen image-space paths. Next, we spatially transition from one unfolded scale to the next, with each transition depending on the characteristics of the two involved scales. We also introduce interaction techniques to navigate our new multiscale representation with spatially-controlled abstraction and relate our representations to temporally controlled

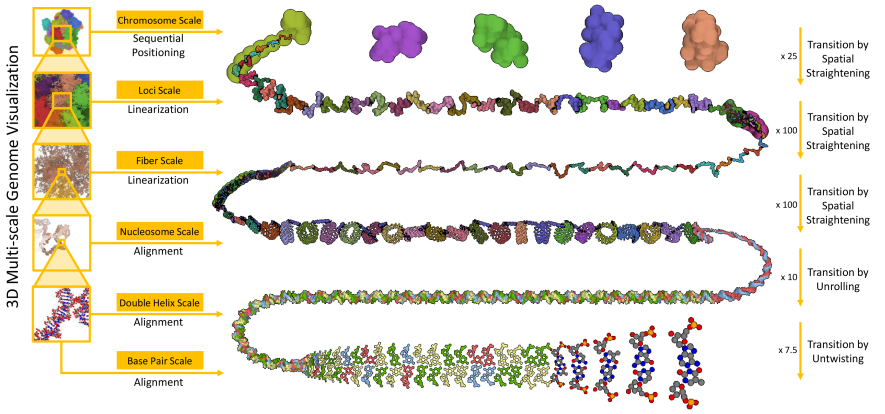



Figure 39: Multiscale unfolding with continuous scale transitions of the 3D human genome, showing chromosomes, loci, fibers, nucleosomes, double helix, and flattened base pairs. Scales 1–4 are composed of details from the next-lower scale, and linearized based on the elements' own positions.

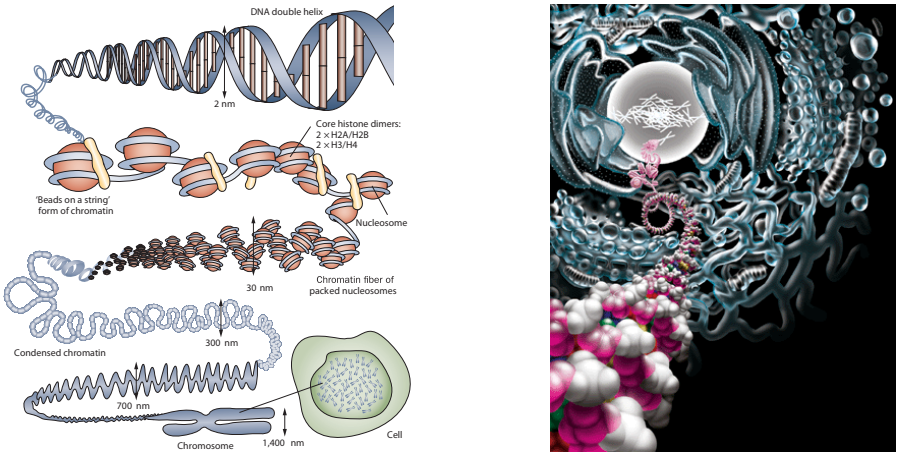
multiscale visualizations. Finally, we report on feedback from professional illustrators and domain experts, which indicates that our interactive visualization facilitates understanding and communication between experts and laypeople. With additional visuals from the data domain (e. g., data-specific bases, gene locations, etc.) overlaid on top of the visualization, it can also act as the basis of a multiscale genome browser.

## 4.2 RELATED WORK

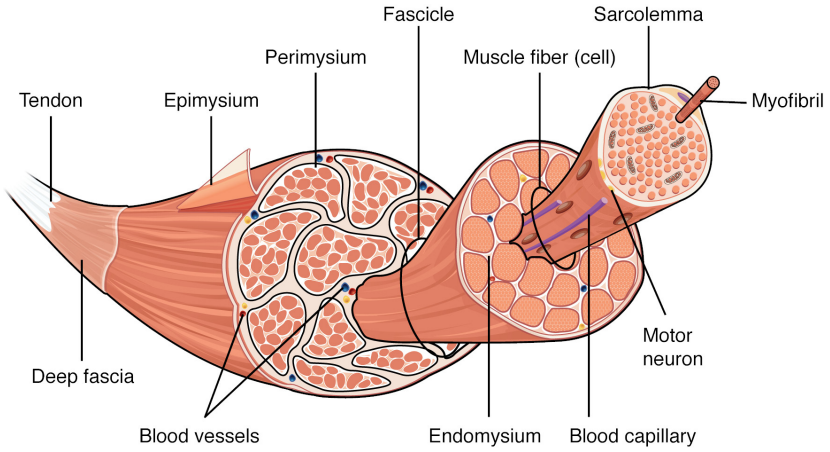
While we focus on genome visualization, our approach relates to past work in several other fields. First, we briefly discuss hand-crafted multiscale visualizations. Next, we review how researchers have realized multiscale visualizations in the past in general, before focusing on the visualization of molecular and genome data. Finally, we relate our work to interactive multiscale approaches.

### 4.2.1 *Hand-crafted illustrations*

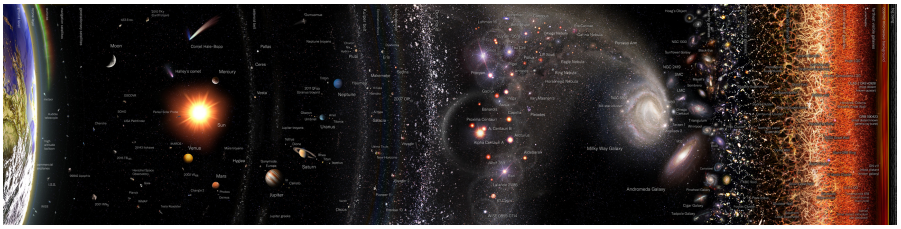
Artists *hand-craft* illustrations to depict linear structures at multiple scales in an unfolded manner. Example illustrations concern the human genome (Figure 40a, Figure 40b), muscle fibers (Figure 40c), or astronomy data (Figure 40d). These illustrations help a viewer to grasp the hierarchical structure, while global spatial characteristics are sacrificed. The artist uses the space as a resource to arrange the linear structure from its smallest to its largest scale representation. To the best of our knowledge, these illustrations are created manually and are usually not directly based on real data. These educational images are inspirational to us. With our work we provide means to




(a) (b)




(c)



(d)

Figure 40: Artistic depictions of image-space scale transitions for (a),(b) genome data, (c) anatomy, and (d) astrophysics. Images (a),(b) from [121]/[105] and © Springer Nature/The American Association for the Advancement of Science, respectively, used with permission. Image (c) by Leeah Whittier (illustrator) and Biga et al. [12] (authors), CC BY-SA 4.0. Image (d) by Pablo Carlos Budassi (cropped), CC BY-NC 4.0.

automatically create them based on real data and also make them interactively explorable.

### 4.2.2 *Spatial transitions of multiscale representations*

Multiscale phenomena are complex as they contain features that are only visible if viewed at a certain resolution. There are various strategies to deal with features at multiple scales. Glueck et al. [41] smoothly transitions between coordinate grids for different scales to facilitate navigation and orientation in the multiscale environment. The method of Everts et al. [34] spatially contracts brain fiber tracts to represent the structure at a higher scale. Although brain fiber tracts have an elongated shape, this approach is not directly applicable to our case, as DNA strands do not share the necessary properties to apply the chosen aggregation strategies. In terms of rendering large complex scenes, Lindow et al. [83] propose a method to exploit the repetitive nature of molecules. Their approach bridges five orders of magnitude in scale. CellView [76] incorporates in a cell visualization a level-of-detail scheme that reduces the number of displayed atoms while the radii of aggregation spheres are increased for distant objects. The method is based on earlier work by Parulek et al. [103]. The CellView approach does not meet our needs, as we depict all levels in one image. Hsu et al. [54] cast non-linearly bent camera rays across several spatial scales. They create smooth transitions of views between different levels of detail and depict the multiscale nature of geometric and volumetric models. *Scalable Insets* are applied for the navigation of a large number of patterns in multiscale visualizations [78]. This technique uses magnified views to visualize small details that would not be visible either due to their size or location, which result from the multiscale nature of the data. Zhang [136] proposed a scale-space animation to visualize the spatial relationship between structures in multiple views. They argue to integrate cross-scale semantic information into an animation to convey the complex multiscale structure. Lueks et al. [85], based on an earlier system by Zwan et al. [125], proposed a dedicated control over the abstraction of molecular systems and thereby enable the creation of seamless spatial transitions. Their method does not deal with the spatial re-positioning of the structure. The DNA-specific approach by Miao et al. [90] provides the user with control over the level of abstraction through animated transitions in order to depict the correspondence between scale representations. Many multiscale visualization methods [34, 85, 90, 91, 125] apply continuous abstraction of the data for reducing the visual complexity. Visual abstraction has been formalized by Viola and Isenberg [128], with a recent update by Viola et al. [126], for the purpose of illustrative visualization. According to their definition of abstraction axes, our work is categorized into an abstraction along the ge-

ometric abstraction axis. Although these techniques offer solutions for certain types of multiscale phenomena, they are not applicable for spatially abstracting heavily tangled structures like the DNA. Spatial transitions are done for depicting multiscale phenomena and also for complexity reduction in general. In medicine, anatomical flattening [74] transforms a 3D object to a 2D representation for different analysis tasks. As an advantage, visual occlusion can be avoided in these 2D representations. Especially vascular structures are similar to DNA. There exist related approaches to spatially transform the elongated shape through straightening [15], flattening [119], and reformatting [95]. We are inspired by these techniques and advance them for genome data.

### 4.2.3 *Visualization of genome data*

Depending on the feature of interest, there are various techniques to represent individual molecules. Kozlíková et al. [72, 73] provide reviews on various molecular visualization techniques. Due to the massive number of DNA molecules, these techniques are only applicable to the atomistic scale in our data. For the specific application of genome data visualization, Nusrat et al. [99] provided taxonomies in this respect. The multiscale character of the DNA presents numerous visualization challenges. Halladjian et al.'s [50] *ScaleTrotter*, for example, bridges several orders of magnitude in scale and enables the users to interactively explore the spatial structure. Due to their visual embedding approach, only one or two scales are visible at a time. The linear arrangement of DNA molecules favors a spatially abstracted representation, where all scales are visible in one image (as in the traditional illustrations in Figure 40). Miao et al. [89] proposed an *abstraction space*, where the strands of a DNA origami structure can be continuously rearranged in space, going from a 3D to a 2D to a 1D representation. The user can adjust the abstraction level, but, only one scale is visible at a time in their approach. A steerable abstraction space [89, 96, 125] enables users to change the scale of a representation depending on interactions with a control panel. Connectomics datasets comprise structures of the nervous system and share similar properties with DNA, due to the tubular shape, large size, and given connections. In this domain, Mohammed et al. [96] continuously abstract the representation of astrocytes and neurites through navigating a specific abstraction space. In general, past visualization techniques for molecular structures and genome data do not offer means to spatially and simultaneously unfold the various scale representations.

### 4.2.4 *Cross-scale interactions*

While we can imitate hand-crafted illustrations, the key argument for us is the interactive exploration in the spatially transformed representations. In our data-driven approach, we enable the user to access the features across different scales. In the past, researchers proposed various techniques for interacting and navigating in multiscale environments. Furnas and Bederson [38] deal with large information scale-space diagrams for understanding multiscale interfaces. In 2D visual spaces, pan, zoom, and bird's eye view are commonly used techniques. For objects of interests that are far apart these methods are ineffective. Javed et al. [58] present the PolyZoom technique, which enables users to build hierarchies of focus regions that can be stacked on each other to depict subsequent scales. Elmqvist et al. [33] proposes a space-folding technique for the same set of problems. Their distortion-based method folds the intermediary space to guarantee visibility for multiple focus regions. McCrae et al. [87] suggest an image-based environment representation to allow users a consistent navigation in multiscale datasets, such as the Earth. There are very different approaches to deal with multiscale phenomena and the data require specific interaction techniques. As our Multiscale Unfolding technique creates a multiscale linearized representation, we have to also provide appropriate interactions for it.

## 4.3 CONCEPT AND OVERVIEW

In our work we focus on DNA data with its unique structural properties. Specifically, DNA is extremely thin and long. If stretched out, the 3.2 Gbp (giga base pairs) [2, 114] of the human genome would be approximately 2 m long [2, p. 179], yet the DNA B-form's double-helix is only 2 nm in diameter [4, 130]—a ratio of $10^9$:1. It is organized in different structural forms at different scale levels for which genome scientists have derived structural data [98]. The involved scales include: the *chromosome* scale, the *loci* scale, the *fiber* scale, the *nucleosome* scale, the *double-helix* scale, and the *nucleotide* scale. With the exception of the 46 chromosomes of the human genome, which are separate entities, the DNA is also physically connected through all the scales. To fit this extremely long chain into the 6 µm cell nucleus, the chain is tightly packed together [2, p. 187]. The result is a complex spatial arrangement of supercoiled [56] DNA strands at the different scale levels. To fully understand the activity and function of a genome, experts have to explore both its sequential chain structure and its spatial structure. Yet it is impossible to understand the sequential chain if looking at it in a traditional three-dimensional genome representation.

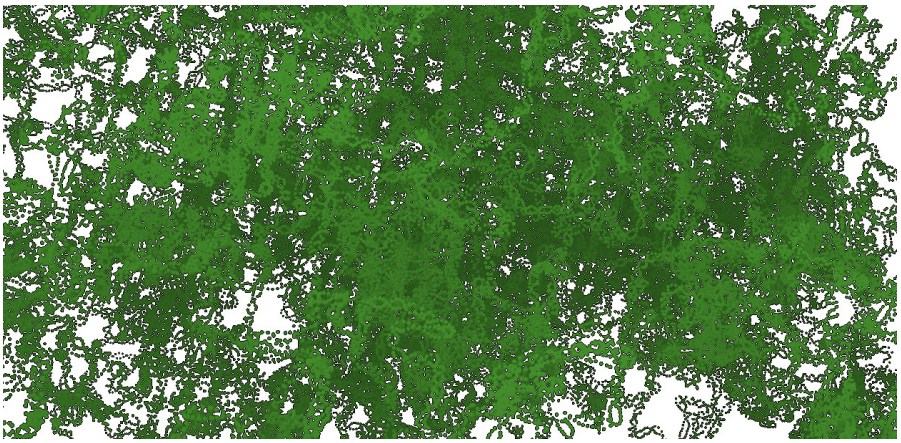


Figure 41: Regular 3D visualization of the nucleosome positions (generated with ScaleTrotter [50]). The genome's linear structure is invisible.

In ScaleTrotter [50] (Figure 41), for instance, only two scales are visible at a given time and user interaction is required to explore one scale and traverse to another one. It is obviously also impossible to understand the spatial structure with only a linear sequence of elements at any scale. We thus have to partially sacrifice the spatial 3D structure to be able to explore both aspects of the genome—as it is a common approach in visualization as we outlined in Section 4.2.2. Also genome experts use tools such as genome browsers that are based on the sequential aspect of the data rather than the spatial structure—yet without spatial aspects.

We thus face three major tasks: (1) making the sequential structure of the DNA visible and understandable across several scales with uncluttered views, (2) providing users with an understanding how the structure is linked to itself across different scale levels, and (3) making such a visualization interactively explorable. A major challenge of existing temporal exploration techniques for multiscale data is that they increase the viewers' memory load and that they are only effective with interaction. It would be necessary to show several or all scale levels in a single view. We take inspiration from professional illustrators (e. g., Figure 40) and the few existing yet limited approaches from the visualization literature [54, 85]. Consequently, we control the depiction of scale based on the individual elements' location in image space.

To realize a spatialized control of the depiction scale (or "zoom level") we use four key concepts. The ***first key concept*** is the ***spatial straightening*** of the sequential structure. The straightening depends on the depicted scale level and is performed in relation to the neighboring scales. In visualization, reformatting has been applied, for example, to flatten 3D structures [74] to resolve visibility issues or to analyze the structures in a simplified way. We also use straightening for these purposes and adapt it to the tremendous scale differences in the genome data, which are of multiple orders of magnitude in

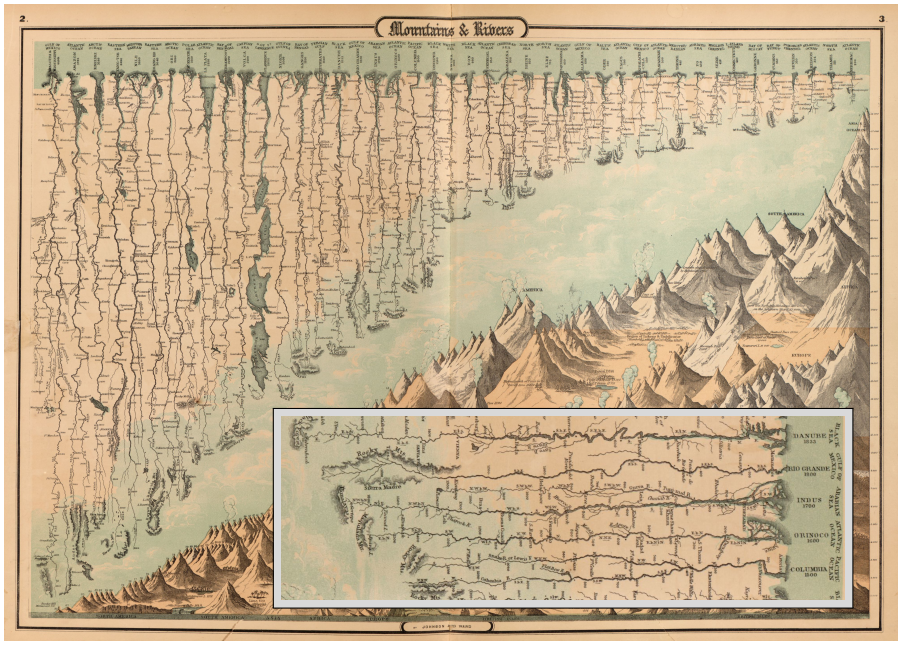



Figure 42: Straightened river illustration. Image from [59, pp. 2–3], ⓢ.

size. Reformatting as an approach to normalize data and make them comparable has already been used, e. g., in geographic illustration more than 150 years ago (Figure 42). Tufte [123, pp. 76–77] describes this illustration mainly as a means to compare river lengths that goes beyond "just another decorated bar chart" because it adds local detail and thus contains more (local) information about the rivers than their length alone. The straightening of the rivers is done in a scale-dependent way where large bends are removed and smaller ones are kept. This is analogous to a frequency-based or wavelet-based analysis where the various frequencies are selectively dampened or removed. In our case, straightening allows us to understand the local 3D structure at a given scale without the "wiggliness" of the genome sequence from a higher scale cluttering our 3D view. The spatial straightening serves as a *visibility management* technique. We realize it by applying data from two neighboring scale levels, using Halladjian et al.'s [50] relationship between data and semantic hierarchies. As we depict, e. g., the semantic level of *fibers*, we render, in fact, the positions of the next-lower, more detailed data level of *nucleosome positions*. We preserve the local detail of the *nucleosome positions* at the lower level, while we straighten and remove the details from the path of the fiber positions at the higher level (Figure 43). We perform straightening in a data-dependent semantic way relative to an upper level, it may involve, e. g., unbending, untwisting, unfolding, and untangling the corresponding lower level, and we describe its realization in Section 4.4.1.

Except for the chromosomes, the size of the DNA does not allow us to show all elements at a specific level after straightening, so we

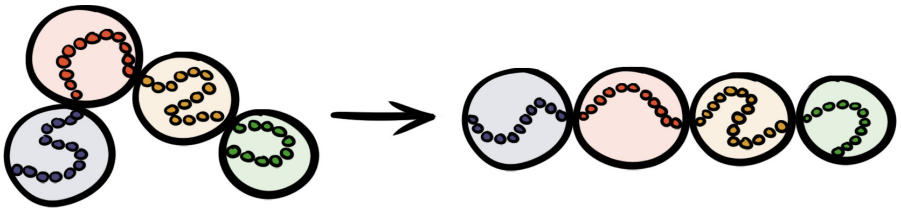


Figure 43: Spatial straightening applied by re-arranging the spatial detail from a more detailed scale (the *straightenee*) to place the positions of the coarser scale (the *straightener*) along a chosen path.

can only show sections. Due to the repetitive nature of the DNA, this is not a problem in many applications and a subsection is sufficient to illustrate the characteristics of a specific level. Nonetheless, we integrate all the scales in a single view and we need to convey how one scale relates to the next one as they are connected spatially and in scale space. We apply the straightening on several scales, but so far we treat each scale as independent from the others. The ***second key concept*** are ***continuous spatial transitions*** between scale representations that illustrate how the structure of a more detailed scale is packed and folded into the next-coarser one. The transition between scales is done through multiple linked views [10, 17, 134] on the spatial 3D data without the need to separate the views from each other. We discuss the realization details of the continuous transitions and their relation to the straightened scale levels in Section 4.4.2.

With these transitions, we combine all scale levels together in a single view, depicting a representative detail section for each of them. We apply this straightening iteratively down the scale hierarchy. In each of the iterations the coarse-scale details are straightened out (i.e., removed) while the fine-scale details are kept. The ***third key concept*** is the ***role change of the scales*** as we move between the scales. For example, going down the hierarchy the fine-scale of the previous iteration becomes the coarse-scale of the current iteration. Our Multiscale Unfolding extends the two-scale strategy of spatial scale transitions as it was previously discussed in the visualization literature. We propose a multiscale strategy applicable as long as data for the spatial straightening and the transition depiction is available from more than two scales. One way of arranging the straightened information in the 2D image plane is a scanline layout as given in Figure 39. To fit all levels into a single view, the scale changes from top to bottom with each row. The traversal direction flips as the levels are connected from one row to the next one. As a result we achieve a compact combined view across all scale levels. We are now able to display all scales of the DNA structure in a single view, with different sections depicting different scale levels. This has the benefit of no longer requiring interaction to understand how the different scales relate to each other. Nevertheless, as the ***fourth key concept*** we provide means to facili-

tate an ***interactive exploration of Multiscale Unfolding***. With interaction viewers particularly benefit from the linked-view character of the visualization. For this purpose we reuse techniques from the HCI literature as we explain in Section 4.5.

## 4.4 SPATIALLY-CONTROLLED MULTISCALE MAPPING

We base our approach on Nowotny et al.'s [98] genome data: 3D locations for the genome for *chromosomes*, *loci*, *fibers*, and *nucleosomes* as well as the *nucleotide* positions for one nucleosome. This dataset is in an interphase configuration, not the often depicted mitotic or "H" chromosome. In this data, each chromosome contains 50–100 loci positions, a locus contains 100 fiber positions, a fiber contains 100 nucleosome positions, and each nucleosome consists of 146 base pairs (bp). We derive the locations of these base pairs from the 292 nucleotide or base positions in the nucleosome. We construct the linkers between two successive nucleosomes based on their orientation, with an average of 46 bp per linker.[1]

It is important to note that the six *data scales* (i. e., positions) are different from the *conceptual scales* that we perceive in a rendering. The reason is that for each element of a given scale we only have a position in space, but no shape or size. We thus represent each element with a sphere. For example, if we render the positions of a sequence of 146 base pairs we perceive the shape of a nucleosome, even though we only render base pairs. So when we depict a conceptual scale, we actually render elements of the next-lower (more detailed) scale.

Below we begin our discussion of Multiscale Unfolding by describing the spatial straightening for the scales such that we can control the placement of a scale independent from its detail (i. e., sub-elements).

### 4.4.1 *Spatial straightening*

The described data, at any given scale, is tightly packed, both in the genome's interphase configuration as we explore it and in the mitotic state as well. In a regular three-dimensional visualization such as ScaleTrotter, the actual linear structure of the genome is difficult or even impossible to see (e. g., Figure 41). We thus spatially straighten the conceptual scale (i. e., the *straightener* level), for instance such that all nucleosomes we perceive can be arranged on a chosen path. We then adjust the positions of the rendered data level (i. e., the *straightenee* level)—in our example the nucleotide positions that make up a nucleosome—such that the connectivity of the straightenee level from before the spatial straightening is maintained also after the transfor-

1 Sources give various ranges for the linker DNA length such as "a few"–80 bp [2] or 20–90 bp [117]. We use 46 bp as the center of the 38–53 bp range cited on the respective Wikipedia article [133], but this number can be adjusted.

mation. In our example the string of nucleotides that forms the nucleosomes needs to remain connected. As each scale has different requirements, we now describe each of them in turn.

**Chromosome scale: Sequential positioning.** We begin with the chromosome scale as the coarsest level of the data. This scale is different from all other scales because the chromosomes are independent structures, and are spatially not connected to each other. We render their local detail (in the form of *loci* positions) without any adjustments and arrange them sequentially in order along the chosen path as shown at the top of Figure 39.

**Loci and fiber scales: Linearization.** The next scales are similar to each other—they are characterized by a tightly packed, intertwined chain—and we thus treat them the same way. We have to consider both the positions of the *straightener*—the conceptual scale which we straighten to follow a given path—and of the *straightenee*—the more detailed scale whose positions we actually render. Basically, we have the coarse-level path and fine-level details that are defined relative to the coarse-level path. Straightening now replaces the coarse-level path by a line, while we keep the fine-level details and apply them relatively to the line.

For this purpose we begin by computing displacement vectors for each of the elements of the straightenee scale (scale $i$). As we show in Figure 44a, we consider each straightenee element (small, filled green, blue, and brown circles) as belonging to a given, single element (large, non-filled circle; only blue shown) of the straightener scale (scale $i-1$). All $N$ blue elements of the straightenee scale, e. g., belong to a single element of the straightener scale (the straightener's blue element: the large non-filled blue circle) whose position we depict with a red dot. We adjust the positions of the first $N/2$ blue straightenee elements based on their old locations and the new positions of straightener's green element and its blue element. Similarly, we reposition the second $N/2$ blue straightenee elements based on their old locations and the new positions of the straightener's blue and brown elements.

Let us consider the example of interpolating between the straightener's blue element and its brown element. We first determine the first and last elements $e_1^i$ and $e_N^i$ in the straightenee sequence (at scale $i$), which belong to the same straightener element $e_k^{i-1}$ (at scale $i-1$, with $k$ being an iterator of this scale), as a base line with length $l_i$ (Figure 44a). We then subdivide this base line into $N-1$ sub-segments such that we have a reference position for each of the $N$ elements of the sequence and center it on the position of the respective straightener element (shown in Figure 44b for the blue sequence). For the second $N/2$ blue straightenee elements we then compute an offset vector $\Delta_{j,k}^i$ ($j$ being an iterator of the straightenee scale $i$) from its corresponding position $f_{j,k} \in [0.5, 1]$ on the base line and its actual

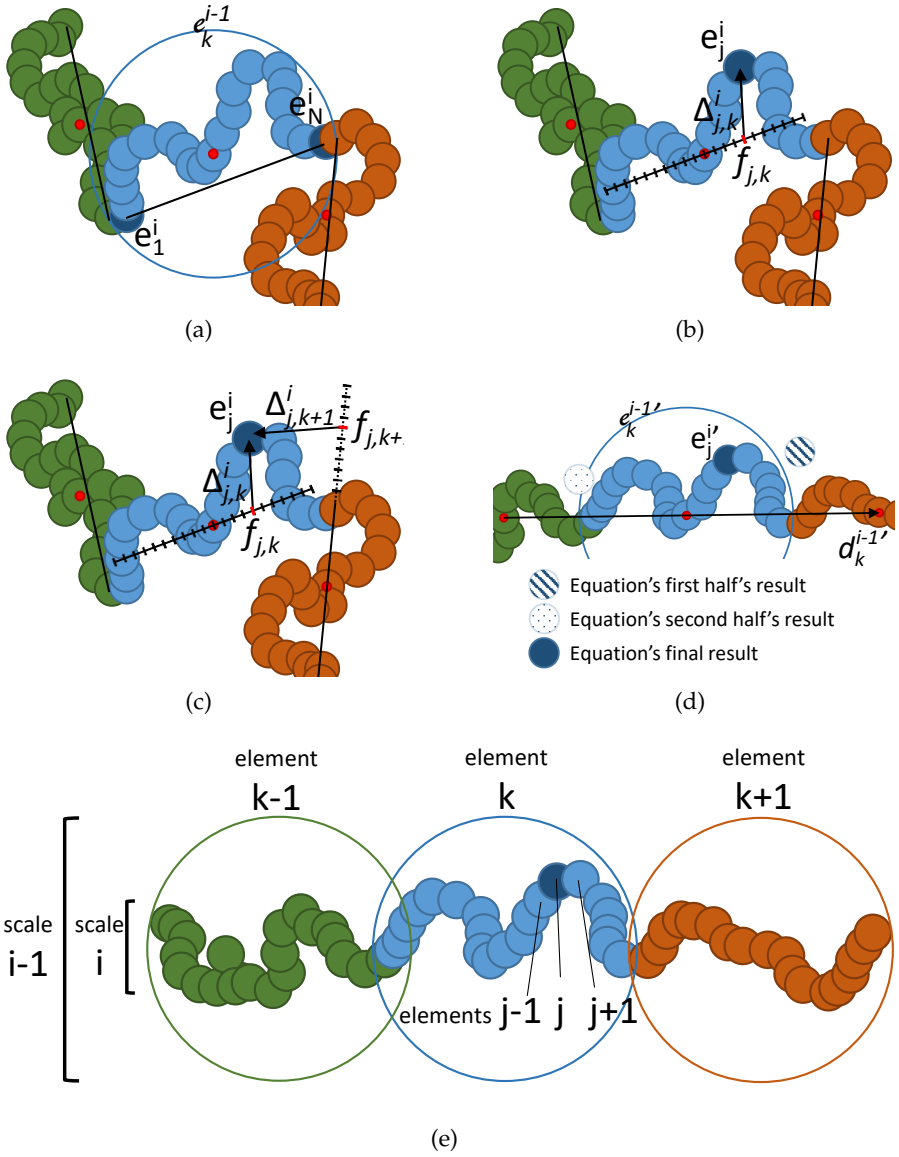



Figure 44: Spatial straightening by linearization for the *loci* and *fiber* scales. In (e) we show a legend of the indices of the formalism we used.

3D position $e^i_j$ before the interpolation (Figure 44b), which represents the influence of the blue straightener element $e^{i-1}_k$ (the influence factor being represented by $f_{j,k}$). However, we also need to consider the influence of the brown straightener element $e^{i-1}_{k+1}$. We thus use the same approach of deriving the base line for the brown straightenee elements, but we extend the baseline after centering by another $N/2-1$ sub-segments to represent the influence $f_{j,k+1} \in [0, 0.5]$ of the brown straightener element. As we show in Figure 44c, we then derive a second offset vector $\Delta^i_{j,k+1}$ for the same blue straightenee element $e^i_j$, but now for the influence $f_{j,k+1}$ of the brown straightener element $e^{i-1}_{k+1}$. The first $N/2$ brown straightenee elements are handled accordingly, and we repeat the process for all pairs of straightener elements. After we thus derived these offset vectors $\Delta^i_{j,k}$ and $\Delta^i_{j,k+1}$ for all $k$ straightenee elements at scale $i$, we first transform them into the basis of the base lines as ${\Delta^i_{j,1}}'$ and ${\Delta^i_{j,k+1}}'$ and then compute, for the elements of the straightener scale $i-1$ after the straightening, the direction vectors $d^{i-1\,\prime}_k = {e^{i-1}_{k+1}}' - {e^{i-1}_{k-1}}'$. We then set ${e^{i-1}_k}'$ and ${e^{i-1}_{k+1}}'$ as the new straightener element positions and derive the new positions of the straightenee elements ${e^i_j}'$ as a straightforward interpolation between the offsets being applied to the new positions:

$$
{e^i_j}' = \begin{cases} f_{j,k}\left({e^{i-1}_k}' + d^{i-1\,\prime}_k(1-f_{j,k}) + {\Delta^i_{j,k}}'\right) + & \\ f_{j,k+1}\left({e^{i-1}_{k+1}}' - d^{i-1\,\prime}_{k+1}(1-f_{j,k+1}) + {\Delta^i_{j,k+1}}'\right) & : \; j > N/2 \\ f_{j,k-1}\left({e^{i-1}_{k-1}}' + d^{i-1\,\prime}_{k-1}(1-f_{j,k-1}) + {\Delta^i_{j,k-1}}'\right) + & \\ f_{j,k}\left({e^{i-1}_k}' - d^{i-1\,\prime}_k(1-f_{j,k}) + {\Delta^i_{j,k}}'\right) & : \; j < N/2 \end{cases}
$$

We show the result schematically in Figure 44d. We apply this process to both the *loci* scale (as a straightener, with the *fiber* positions as straightenee) and to the *fiber* scale (as straightener, with the *nucleosome* positions as straightenee). We show the results of the application of this process in the second and third line in Figure 39.

**Nucleosome scale: Aligning the nucleosomes.** The straightening of the *nucleosome* scale is conceptually simpler than the process for the previous scales because the shape of the nucleosome is known and does not have to be adjusted. The data do not contain information on the orientation of the nucleosomes, they only contain 3D positions. To align the nucleosomes on an arbitrary path we place the nucleosome positions in sequence, and then render the single included nucleosome model at these locations. We orient the model such that each donut-shaped nucleosome's "entrance" connects well to its predecessor and its "exit" connects well to its successor. We then connect each pair of consecutive nucleosomes with a section of double-helix that acts as linker DNA. As the data do not contain information about the length of this linker DNA we use an average of 46 bp per linker as

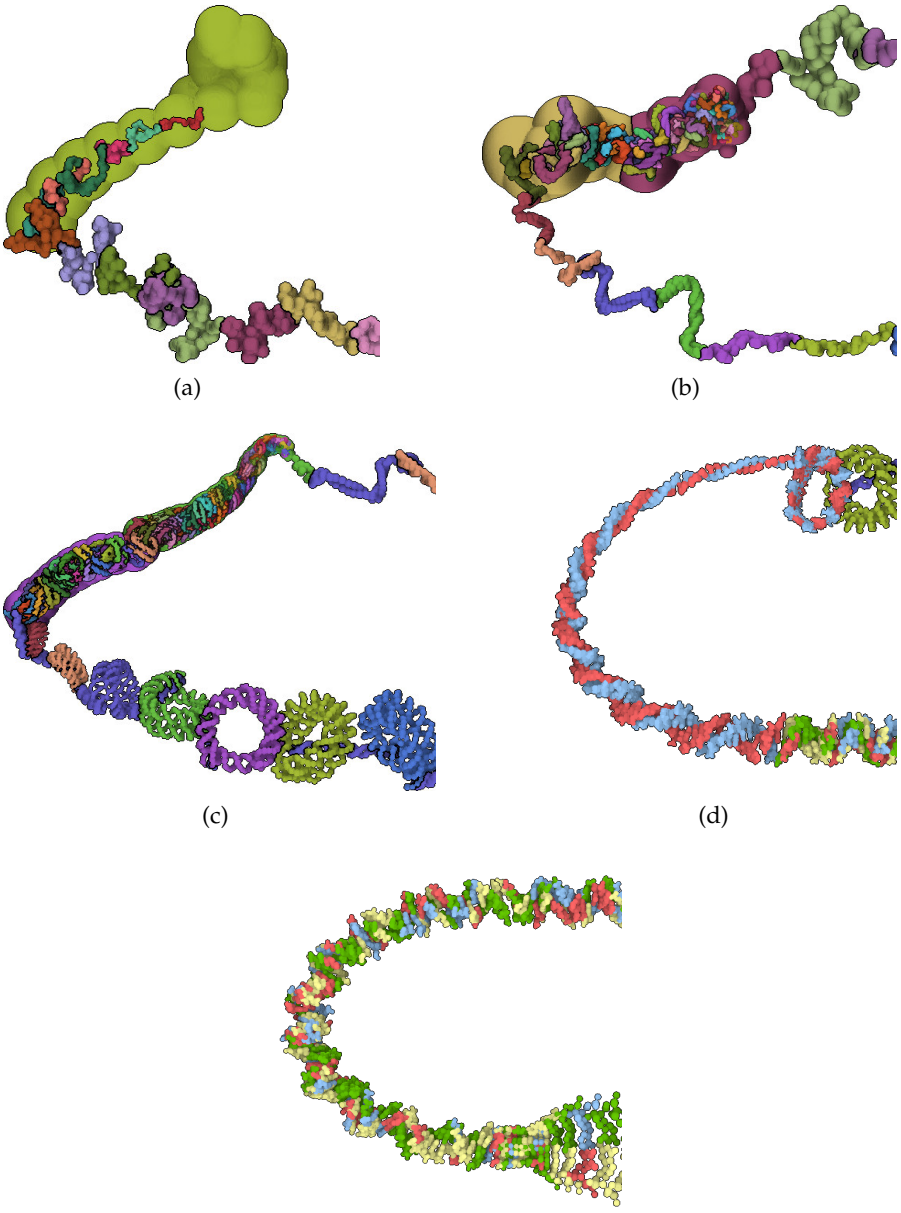
explained above. We show the result in the fourth line from the top in Figure 39.

**Double-helix scale: Unrolling the nucleosome.** At the conceptual double-helix scale we render positions of the atoms, which we derive from the *1AOI* PDB model. From this data we compute the positions of the base pairs. We later match these with the base pair positions in Nowotny et al.'s [98] nucleosome model for the transition between the scales. For the unrolling, we then fix the first base pair's position and then transform the remainder of the model with a rotation such that the connection between the first two base pairs is straigthened. We repeat this process until all 146 bp of the nucleosome have been "unrolled." We show the result of this process in Figure 39 in the fifth line from the top.

**Bases scale: Untwisting the double helix.** In the final spatial straightening stage we "flatten" the double-helix. We calculate the positions of the individual bases from the *1AOI* model and derive vectors that encode the orientation of each base pair. Like with the nucleosome "unrolling," we start by fixing the first base pair and then, base pair by base pair, rotate the model such that the base pair centers all end up in the same plane. Each base's atom arrangement may still be difficult to understand or tell apart, so we introduce a second representation in which we also flatten the atom arrangement of each base, using manually adjusted models for *adenine*, *guanine*, *cytosine*, and *thymine*. Also based the *1AOI* model from the PDB, we thus created two models of each base with Samson Connect, one (A) with the flattened base geometry and one (B) in which also the backbone was rotated. We then interpolate between the stage after the rotation of the base center to the screen plane to the respective A-version and then to the B-version. As we start this transition, we also switch from coloring by bases to coloring by atoms and make base pairs bigger to better show the molecule structure, as shown in the final row of Figure 39.

### 4.4.2 *Spatial transitions*

Any two consecutive spatially straightened scales differ significantly in size and number of elements (with the exception of the last two). We need transition zones from one scale to the next one that show (a) the element size and number difference and (b) illustrate how the more detailed scale "folds" into the coarser one. In addition, we also use these transitions to (c) allow us to change direction such that we can arrange all scale representations in a zig-zag pattern (Figure 46), to form a single multiscale unfolded representation. We connect one *constant-scale zone* with another one by means of a *transition zone*. Elements from the constant zones extend into the transition zone, in which their size is changed according to the respective change of



(a) (b) (c) (d) (e)

Figure 45: Transitions between consecutive scales: (a) between chromosomes and loci, (b) between loci and fibers, (c) between fibers and nucleosomes, (d) between nucleosomes and double-helix, and (e) between double-helix and untwisted double-bases.

spatial scale. We also need to show how the elements of the more detailed scale are located within the elements of the coarser scale. We generally place the elements in the transition zone along a curve that connects the coarser with the more detailed scale as shown in Figure 46. Along this curve, we maintain the connectivity between the elements. We change size and rendering parameters so that we illustrate the containment relationship. Next we describe how we realized these transition zones, again starting from the coarsest representation.

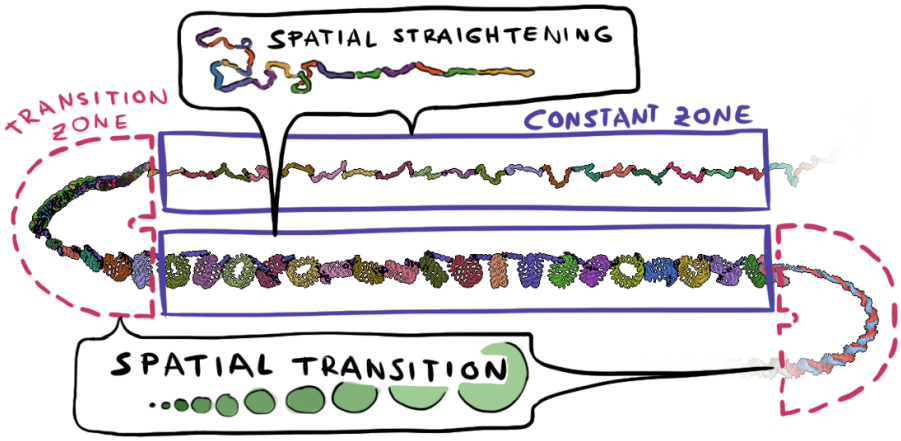



Figure 46: Conceptual arrangement of constant and transition zones.

**Chromosomes-to-loci transition.** In the chromosome scale we render loci positions (50–100 per chromosome), each represented by a sphere. In the next scale we have 100 fiber positions (again each represented by a sphere) for each locus. In each scale, for illustration, we render all spheres that make up one semantic unit using the same color; e. g., all 100 fiber positions are depicted in the same color for each locus as shown in Figure 39. In the spatial transition between two consecutive scales, we fit each color group of a finer scale into a single sphere of the coarser scale. This also means that each color group of the semantic loci scale represents a single (locus) sphere as rendered in the chromosome scale.

Depending on the location of the DNA sequence of the depicted locus scale, a given number of locus spheres remain until the end of each chromosome. While we render the remainder of the locus spheres of the chromosome at the regular data positions in the chromosome scale, we place a fixed number (we use 7) along the upper arc of the transition zone curve, evenly spaced and slowly enlarging in size to illustrate the changing spatial scale (see upper part of Figure 45a). For the second part of the transition, we place the same number of color groups from the loci scale on the lower arc of the curve, slowly decreasing in size (lower part of Figure 45a). At the middle point of the transition curve, we render both a single locus sphere as well as the locus color group that represents the same data—so that their sizes match. To illustrate the "folding of one scale into the next,"

we then continue to render the loci color groups on top of the loci spheres for the upper arc of the transition zone such that the sizes match (i. e., the color groups are shrinking in size towards the top). To ensure that the elements of the more detailed loci scale remain connected, we use the same process as described in Section 4.4.1 for the spatial straightening of the loci scale—just using the transition curve as the straightening path.

**Loci-to-fibers transition.** In contrast to the first spatial transition, for the change from loci to fiber we now need to connect two linearized sequences. For this purpose we dedicate the loci color group that follows in the DNA sequence after the last linearized loci color group as the loci scale's transition group. This group represents a single locus and consists of 100 fiber positions. Each fiber position on the fiber scale, in turn, consists of 100 nucleosome positions which again form a color group on the fiber scale. Like before, we arrange the transition along a curve and place the loci scale's transition group on the upper arc of the curve, and 5 fiber color groups that from the DNA sequence before the depicted constant-zone loci groups. We also again use the basic approach from spatial straightening and adjust their scale such that the elements match when they meet. In the upper part of the transition, we continue to render the increasingly smaller fiber color groups until the start of the loci scale's transition group as shown in Figure 45b. As also can be seen in the figure, the end of this sequence is often hidden due to the spatial arrangement of the loci transition group. As we show in Figure 45b, we further adjust the rendering parameters like the outline transition so that the outline does not become over-emphasized for the loci elements as they are getting smaller towards the top.

Due to interaction, we often have the case that the DNA sequence is arranged such that the linearized part of the loci scale does not end exactly at the end of one of its color groups. In this case we use the remaining part of this last color group and proportionally add another section of the following loci color group to jointly serve as the loci scale's transition group as we show in Figure 45b with the purple and beige color sections at the end of the loci scale. We thus create smooth animations in which the positions of the straightened scales can be continuously adjusted.

**Fibers-to-nucleosomes transition.** The next transition from the fiber scale to the nucleosome scale is similar to the previous one, with the difference that the nucleosome "color groups" are always full nucleosomes, each consisting of 146 base pairs. We follow the approach as before, only ensuring that we also appropriately render the linker sections between two consecutive nucleosomes in the lower part of the transition arcs. At the upper arcs the nucleosomes are placed so close to each other—one nucleosome per sphere of the fiber scale's transition group, which consists of 100 spheres—that we omit the linkers

here as they would not be visible anyway. We show the result of this transition in Figure 45c.

**Nucleosome-to-double-helix transition.** The nucleosome's inherent structure prevents us from using the previous transition styles for switching from nucleosomes to the double-helix. Instead, we take inspiration from the conceptual unrolling of a nucleosome. We fix the end of the transitional nucleosome at the end of the constant-scale nucleosome zone, taking into account its spatial orientation. Depending on needed amount of unrolling, we straighten this part and place it on the transition curve. We adjust the curve such that it always connects to the current point on the transitional nucleosome where the unfolding happens and use colors to indicate the two strands of the double-helix, as shown in the snapshot in Figure 45d.

**Double-helix-to-bases transition.** The final transition between double helix and bases no longer represents a change of spatial scale. Instead, with it we address the issue that in the double helix we cannot easily see all nucleotides in a single view. Due to the helical twist, they are partially occluded by other nucleotides. We use a strand of the double helix with a complete rotation, extracted from our spatially straightend *1AOI* PDB model. In this model we identify the locations of the nucleotides (bases), find the centers between each pair to be the base pair locations, and derive the respective twist (rotation) at each base pair position. For the transition, we remove this rotation based on the position of the structure on the transition zone curve to completely flatten the helix at the end. We show the result of this transition in Figure 45e.

### 4.4.3 *Illustrating scale differences*

Our approach allows us to place the multiscale unfolding representation along an arbitrary curve. For most of our experiments, we decided to use a zig-zag layout with six rows. Figure 46 illustrates this concept schematically and we show the result in Figure 39, where we depict one transitioning chromosome and four complete ones at the top, followed by 20 loci (i. e., 20%–40% of a chromosome), 20 fiber locations (20% of a locus), and 20 nucleosomes (20% of a fiber). Finally, we show two full nucleosomes (292 base pairs) and 39 base pairs for the flattened helix.

In addition, as none of the depicted elements have natural colors due to their small size, we assign random colors to each of the chromosomes from a pool. Next, for all conceptual scales from the loci to the nucleosomes, we assign random colors from the same pool such that each color group shows a single item of the conceptual scale. A single element of the loci scale, e. g., comprises 100 fiber positions, all of which belong to the same color group.

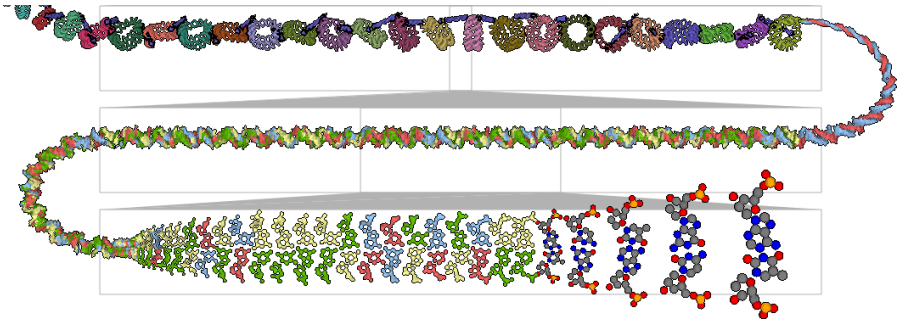


Figure 47: Illustration of the spatial scale ruler. The boxes only indicate the average length difference (based on the number of contained elements), not the actual location or correspondence of sections, because our visualization is sequentially continuous.

These color groups and our chosen number of elements already illustrate the scale differences with respect to the number of elements that are included into a single element of the next-coarser scale. For even a better illustration of the involved spatial scales, we provide the option to add a spatial ruler beneath the scale renderings as shown in Figure 47. This ruler consists of boxes and gray trapezoids that schematically illustrate how much room the full straightened, constant-scale section of a detailed scale take in the next-coarser scale, based on the number of contained elements. The bottom-most base pair scale in Figure 47, e. g., contains the equivalent of four full turns of the double-helix, while the double-helix scale contains 29.2 full turns. These 292 bp in the double-helix scale are the equivalent of two nucleosomes (with 10 bp in a full double-helix turn), which takes 1/10 of the nucleosome scale. For practical applications, however, all these settings can be adjusted as needed.

### 4.4.4 *Implementation details and rendering performance*

We built our implementation using the Marion framework's [94] molecular visualization functionality, augmented with functions to load Nowotny et al.'s [98] GSS genome data. The molecular rendering uses Le Muzic et al.'s [76] GPU-based dynamic atom injection. We inject the respective scales' basic colored rendering shapes as spheres. We use 2D sphere impostors instead of mesh models for the spheres [76]. As the sizes and positions of the depicted elements change in the transition zones, we adjust this data and reload it into the GPU in each frame. On an Intel Core™ PC (i7-7920HQ, 6 cores, 32 GB RAM, 3.10 GHz, nVidia Quadro M2200, on Windows 10 x64) we achieve approx. 6 fps when showing the visualization at a size of 1920 × 1080 pixels.

Typically, we load the chromosome data only at the highest scale (i. e., loci positions). Only for the currently unfolded chromosome, we

also load the data for the remaining scales. It is possible to load all the data at once, but that would take approx. 3 minutes, instead of the mentioned 20 seconds for the details of a single chromosome. The long overall loading time results from the large amount of DNA data with 23,958,240 nucleosome positions for all chromosomes, plus the respective locations at the coarser scales.

## 4.5 INTERACTION FOR MULTISCALE UNFOLDING

Traditional multiscale visualization techniques are often designed to reduce the need for interaction by showing details for all involved scale levels in a single view. The essential benefit of Multiscale Unfolding is that we no longer need to interactively explore different scales and then mentally connect them. The logarithmic representation of the observable universe shown in Figure 4od, for instance, is a multiscale depiction of the massive size and distance differences, without the need to interact with a single part of it. This property also applies to our Multiscale Unfolding of DNA data. It works due to the repetitive nature of DNA, despite the huge amount of data in most of its scales.

While this is true and our design could be used to create still-image illustrations of the genome, we also envision the spatial representations that we do show already to be combined with abstract data depictions at the different levels. This abstract data is unlikely to be repetitive and interaction techniques are essential to be able to explore the whole genome. The possibility to interact with the unfolded scales enables the viewer to explore and analyze the sequential data, while still being able to benefit from the advantages of the two-dimensional sub-views. Below we discuss multiscale navigation techniques for this purpose, treating the first scale with its discrete chromosomes differently from the remaining scales which are spatially continuous through the scale levels.

### 4.5.1 *Chromosome selection and panning*

Most useful for the navigation of our one-dimensional data arrangement is panning along the centerline of our visualization. We enable this for the chromosome level and it works as expected as our visual mapping facilitates a smooth re-positioning. More important, however, for the straightened chromosome scale is to be able to select a new chromosome to "drill into", which we facilitate through a click-based selection. The major challenge here is the amount of data that has to be reloaded due to this single interaction. It takes approx. 20 seconds to load the data for all scales of the chromosome onto the GPU. This waiting time of more than 10 seconds leads to losing a user's attention [19, 93]. As the chromosomes are arranged sequen-

tially, a panning of the scale or multiple successive clicks on the end to get to a particular chromosome are thus practically not feasible. We allow users to pan the chromosome scale, but without affecting the first chromosome that is spatially connected to the following scales. As soon as users have found the correct chromosome, they can select it and only have to wait for the data to be loaded once.

### 4.5.2 *Multiscale Zliding for spatially-specific panning*

Similar to the chromosome scale, we support panning for all of the following scales. We realize this control with a mouse click-and drag interaction, mapping the horizontal offset to the respective translation of the scale on which the interaction was started. While this mapping is linear for the constant-scale zones, it would be non-linear for the transitions. For the sake of intuitiveness, we thus restrict the panning to starting points in constant-scale zones.

We adjust the applied translation speed to the respective scales such that the mouse movement is directly mapped to the chosen scale. We want to avoid having to start a new panning interaction each time we change scales. For this, we take inspiration from the HCI literature and combine ideas from Ramos and Balakrishnan's [107] TLslider widget with the same authors' [108] Zliding technique. We adapt their approaches to the context of our Multiscale Unfolding. The authors used the TLslider to navigate details of a video stream. They arranged video snapshots in an S-shaped curve away from the horizontal to make more room for seeing individual images. Their Zliding technique later generalized this idea and added zoom control derived from pen pressure to 2D positional input. In our case, we only need one-dimensional input (in the x-direction) for panning, freeing the second positional input (the y-direction) for the zoom factor, for which we use the spatial offset.

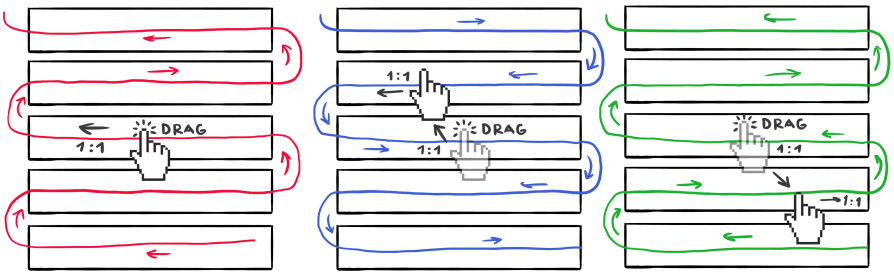



Figure 48: Scale control with Multiscale Zliding. The black annotations indicate mouse movements. The colored paths and arrows indicate the motion of the elements during panning. We determine the gain factor from the hovered mouse position and appropriately increased and decreased through propagation to adjacent scales.

We suggest a novel *Multiscale Zliding* (see Figure 48) that uses the y-distance from the initial click point to control the panning speed, depending on the direction and scale the user hovers while dragging. If we move toward coarser scales, we increase the gain factor [86], while we decrease the gain factor when dragging toward more detailed scales. Regardless of the dragging direction, we adjust the gain factor to match the speed of the hovered scale, if the mouse pointer is within a given distance to its center line, and linearly interpolate in-between these constant-speed zones. An additional challenge is caused by our zig-zag arrangement of scales which results in opposite motion directions for consecutive scales. We flip the direction of the motion as we pass the half-way point between two scales. The ruler as discussed in Section 4.4.3 visually assists this interaction by illustrating the respective interaction zones.

Multiscale Zliding is a one-dimensional multiscale navigation technique that allows users to fluidly control the panning within the long DNA sequence, without having to re-start a click-and-drag interaction in-between. Because of the Multiscale Unfolding's inherent 1D nature, Multiscale Zliding allows us to navigate the data with a standard input device such as a mouse. It could easily be adjusted to a two-dimensional technique by employing, like Ramos and Balakrishnan [107, 108], pressure sensing to be able to use the input's y-component also for spatial control.

### 4.5.3 *Visual mapping adjustments due to interaction*

Multiscale Zliding now allows users to re-position the data in all scales. As a consequence, we had to make two major changes to the previously described visual mappings. The first change affects the transition between the double-helix and its flattened version. The previously described realization of the untwisting works only in a static view. As soon as we interactively move the double helix, fewer or more turns have to be removed depending on how many turns are included in the flattened section (which virtually extends flattened until the very end of the sequence). We fix the orientation of the flattened helix at the lower end of the transition zone so that it always meets the flat constant-scale region of our bottom scale. Conversely, as we pan this bottom scale, the rotation of the helix accumulates and we rotate the double-helix scale appropriately during interaction. We propagate this axial rotation further through the transition zone to the nucleosome scale. At the end of this transition, the double helix is so small in size that we no longer see the turning, so we do not have to propagate it any further.

We make a second major change also at the nucleosome unrolling, which we cannot express as a "normal" unrolling (like that of a garden hose from a rotating spool). The reason is that, similar to the

double-helix, one end of the "nucleosome spool" has to be fixed to the remaining DNA chain. To emphasize the conceptual unrolling, we thus add a twisting transformation to the construction of the nucleosome spiral: While placed at their correct positions when the nucleosome is fully assembled, we space the double bases further apart from each other as the nucleosome is unrolled. This leads to the described animated twisting of the nucleosome which better illustrates its unrolling. Ultimately we "hide" two forms of accumulated twists due to the need to unroll linear structures with fixed ends.

### 4.5.4 *Visually encoding the distance to a clicked item*

While we remove the global 3D spatial configuration to reveal the DNA's linear structure, it still is important because the physical 3D distances between basepairs, in addition to their sequence, also has an effect on the function of the genome (gene expression). To compensate for the loss of the spatial structure of the genome, we thus allow viewers to reveal these distances interactively as shown in Figure 49. By clicking on a given element, we color-map the distances to other elements of the same scale or to elements in other scales. As the scales have vastly different sizes, we adjust the colors to which the distances are mapped to the given scales.

Figure 49: Visual encoding of a given element's actual 3D distances to the remainder of the scales on demand.

## 4.6 DISCUSSION, FEEDBACK, LIMITATIONS

Initially, we set out to solve a problem in illustrative visualization, i. e., how to show all scales of a highly multiscale dataset in one view, without completely separating the scales. As the results of our Multiscale Unfolding approach (e. g., Figure 39) show, we achieved this goal. Furthermore, the new visual representation also leads to several new insights that were beyond our reach before. Below we review our results, report on feedback from domain experts and illustrators, and discuss limitations.

### 4.6.1 *Visual abstraction and generalization*

For the first time we can show the different local spatial arrangements of the different scales of the DNA in its interphase configuration, based on real data and in a single view. Despite using the same straightening method for both the loci and the fiber scales, for example, their configuration is visually different, which also can be seen in the transitions between the scales. Furthermore, through the interactive navigation we can experience and appreciate the huge spatial scale differences in the DNA structure. If we pan a given scale the next-coarser scale barely moves due to the 1:100 relationship of element containment for many of the scales. Similarly, if we move a coarse scale, the lower scales move so fast that this motion is beyond the animation capabilities even of the fastest computer graphics hardware.

Interestingly, with Multiscale Unfolding we perceive all scales without interaction. Also perspective is not necessary. Instead, our explicit spatial transitions demonstrate how one scale relates to and integrates into the next one, similar to traditional illustrations as shown in Figure 40a. Perspective can serve a similar purpose as shown in Figure 40b or as the "emphasized perspective scale transitions" from the VIS literature such as Parulek et al.'s work [103]—here the scale change due to the perspective is emphasized through illustrative visualization mappings. A remaining question concerns the combination of the two approaches and how we would then control the abstraction in the scale transitions. Moreover, the multiscale aspect of our unfolding can be considered as a half of an extremely long one-dimensional fish-eye lens with stair-wise defined *optics* of detail. Of course, it would be possible to define another fish-eye lens profile—a symmetric one, for instance, that would generate effects closer to a traditional fish-eye metaphor. Moving such a lens over the unfolded genome could then interactively reveal the details at an arbitrary place in a sequence, while the rest of the structure is represented with less detail. Ultimately a user-adjustable scale-factor function could assign a specific scale level to each location along the straightened centerline, which would support an arbitrary arrangement of several focus and context regions simultaneously.

Another important question is also how and to which other domains we can generalize our approach, beyond the domain of genome visualization that we demonstrated. As depicted in Figure 40, multiscale phenomena do not only exist in genome data but also, for example, in other scientific fields like anatomy and astronomy. Some of these domains, such as the muscle visualization of Figure 40c, probably only benefit to a limited degree from our approach. They do not comprise large scale differences. Several other data domains exist in which the data is linearly organized as 1D paths, with various levels

of scale. For example, routes (e. g., path of the Tour de France) and GPS-based driving directions could similarly be subjected to Multiscale Unfolding. For the latter, existing abstraction techniques (e. g., [1]) could be employed, as well as the current level of interest for the control of the unfolding. For others, like astronomy (Figure 40d), we could adjust Multiscale Unfolding to use specified paths and a selection of the shown physical property for each scale. The assembly of lakes of the world in Figure 50 could provide inspiration on how to deal with non-linear, discrete, and widely distributed elements. If applied to two- or three-dimensional data, not only our linear interactions (one-dimensional panning) will have to be converted to two-or three-dimensional panning. Also, we will have to use different straightening strategies. For instance, concepts from image morphing (e. g., [11]) could assist the goal, but potentially also the concepts of axial or volumetric deformations [26, 27, 75, 80] and multiresolution curves [36].

The application to 2D or 3D data might also necessitate a more linear arrangement as illustrated in Figure 40d, contrary to our current zig-zag structure. Moreover, astronomy data is characterized by huge empty spaces between elements and between different scales [50]. We could address this problem and the issue of the three-dimensional character by employing a spatial straightening to three-dimensional astronomy data, similar to our aligning of non-connected chromosomes. Planets could be aligned on a path like pearls on a necklace. The further scales would then align stars, galaxies, and so on. A remaining problem of domains like astronomy concerns the uneven density of available data [50]. We know much more about regions close to Earth than we know about the far ends of the observable universe, so an interactive exploration of the entire space as we can facilitate is difficult in these domains.

### 4.6.2 *Feedback, application, limitations, extensions*

To evaluate our approach with experts in this field, we consulted a number of professional illustrators and biologists. Three illustrators from the field of molecular biology (9, 33, and 45 years of post-PhD experience, respectively) specifically commented about our chosen representations, and their feedback allowed us to correct structures, in particular, at the base and double-helix scales. They also recommended to add color-coding based on nucleobases as we now show in Figure 39. They criticized that we do not show histones at the nucleosome scale, which could easily be addressed by adding this element to the used nucleosome model and which we did not yet do since we concentrated on the aspect of controlling the abstraction spatially. They also noted that the transitions lack perspective and are not informative enough about the scale difference. We do not see this

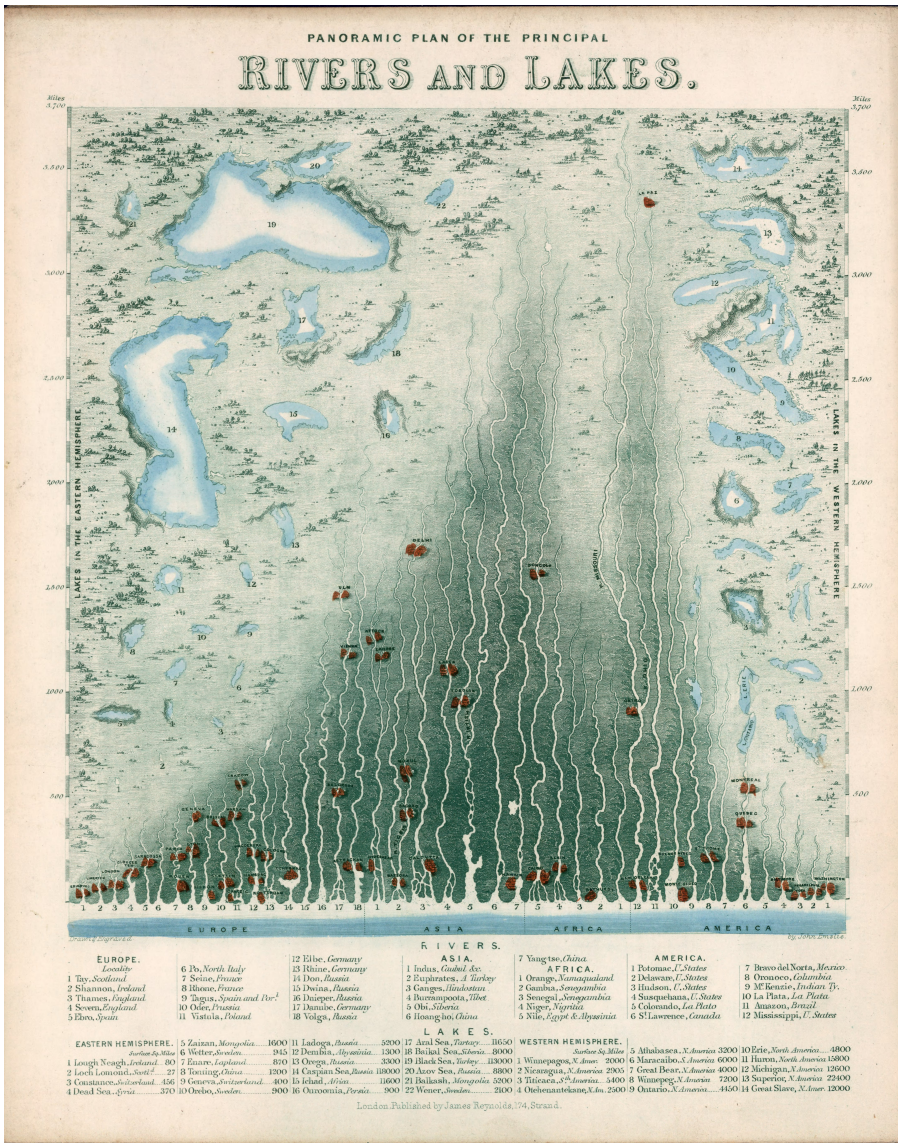



Figure 50: An example of spatial abstraction of straightened rivers and non-linear, discrete lakes in early, hand-drawn atlases. Image: James Reynolds and John Emslie. *Panoramic plan of the principal rivers and lakes.* London, 1851; scan © 2000 Cartography Associates, CC BY-NC-SA 3.0.

as a problem as we are able to introduce perspective at will. Yet it has to be used with care: sufficient perspective to show correct scale differences may imply that only some of the scales can be shown—Figure 40b shows only two or three scales between atoms and chromosomes with perspective transitions. With respect to the application of our approach, they saw potential both for communicating the multiscale aspect of the DNA to laypeople or even biologists and for supporting expert tasks as long as additional data is overlayed on top of the existing visualization.

We also conducted a semi-structured interview with a certified illustrator with three years experience and a PhD in Bioengineering. Considering that she had seen ScaleTrotter before, she preferred the clarity and simplicity of Multiscale Unfolding because, without the 3D structure, she could see all the scales in one view. She stated that this fact would allow her to grasp the differences between non-consecutive scales. She was very interested in the interaction and was keen on exploring the tool herself. Overall, she foresees that our tool can produce aesthetic illustrations and animations for teaching a lay audience. She noted that our tool is based on actual scientific data, which does not only save her research time, but can be also more truthful to nature. She also noted that, "if vector images could be exported from the visualization," she could sketch over them and add her own artistic style to get a final result. If the goal is not to create a publishable illustration, she also sees this tool as a way for domain experts to create illustrations themselves without the need of an illustrator.

Finally, we conducted a semi-structured interview with an expert in molecular biology with 23 years of post-PhD experience. He specializes in genetics and studies the composition, architecture, and function of SMC complexes. He immediately made it clear that this tool can be used to educate first-year students, but for it to be used by domain experts, additional data has to be overlaid on top of the existing structures (noting different types of data for each scale). He found it useful to have a multiscale visualization where, e. g., he can look at a gene on one scale, and at the cluster of genes that the gene belongs to on a higher scale, in one view. He imagined that a version of our tool with additional data would be much easier to use than current tools that rely on diagrams, because it offers more context. He also said that "it would be a dream" to have a tool that combines all of that data. Adding such data requires further work to solve respective data mapping and representation issues as well as to create dedicated interaction techniques.

The molecular biology expert also saw complementarity between this tool and ScaleTrotter, both multiscale genome visualization tool with different approaches. He imagined two windows side by side, with each showing one of both tools. A user could navigate them-

selves in the full 3D structure and then switch to the multiscale straightened view, or use the latter to guide their way in the 3D model. The currently available interactions, however, are limited to a local navigation. A more comprehensive set of interactions would thus be desireable [30, 62, 63, 135], like a global navigation (e. g., "go to *this* location") or data analysis functionality (e. g., "compare *these* two sections").

With respect to technical limitations, while our current rendering speeds allow us to interactively explore the visualization, they are still relatively slow. During interaction we permanently need to update on the GPU the actively shown elements in the constant-scale and transition zones. This requires us to transfer for each frame a lot of data between CPU and GPU. Computing the active sections directly on the GPU could alleviate this problem.

## 4.7 CONCLUSION

On a conceptual level, our work on spatially controlled Multiscale Unfolding extends the state of the art of using abstraction in visualization and provides a rich tool set for the depiction and exploration of multiscale data. Beyond common interactively (i. e., temporally) controlled multiscale visualization techniques, this approach allows us to create convincing still-image visualizations of multiscale data. More importantly for the visualization field, our approach does not require viewers to memorize different scale representation for comparison. It *inherently supports one of the most essential visualization tasks* [16, 40]: data comparison—in our case across scales. Essentially our visual mapping is *a type of multiple linked views* [10, 17, 134], with the fundamental difference to their typical use that our Multiscale Unfolding sections are not *only linked in their response to interactive input but also linked continuously in space*. Our combination of multiscale visualization with linked views extends the visualization technique of *contextualization* [52] to the heavily intertwined DNA data and shows how it can be applied through multiple connected spatial scales. We are advancing the central visualization approach of *unfolding and flattening* [74] into the multiscale domain.

Mulitscale Unfolding also raises many new questions and opens up several new paths for research. With the basic framework in place, e. g., we now want to explore further image-space arrangements apart from the zig-zag layout. Spirals, space-filling lines, or arbitrarily defined paths from artists could be investigated (Figure 51).

Our work so far concentrated primarily on the technical matters of multiscale visualization and on how to control multiscale abstraction. A next step should incorporate further feedback from professional illustrators and genome scientists to improve data representation, artistic control, usability, etc.

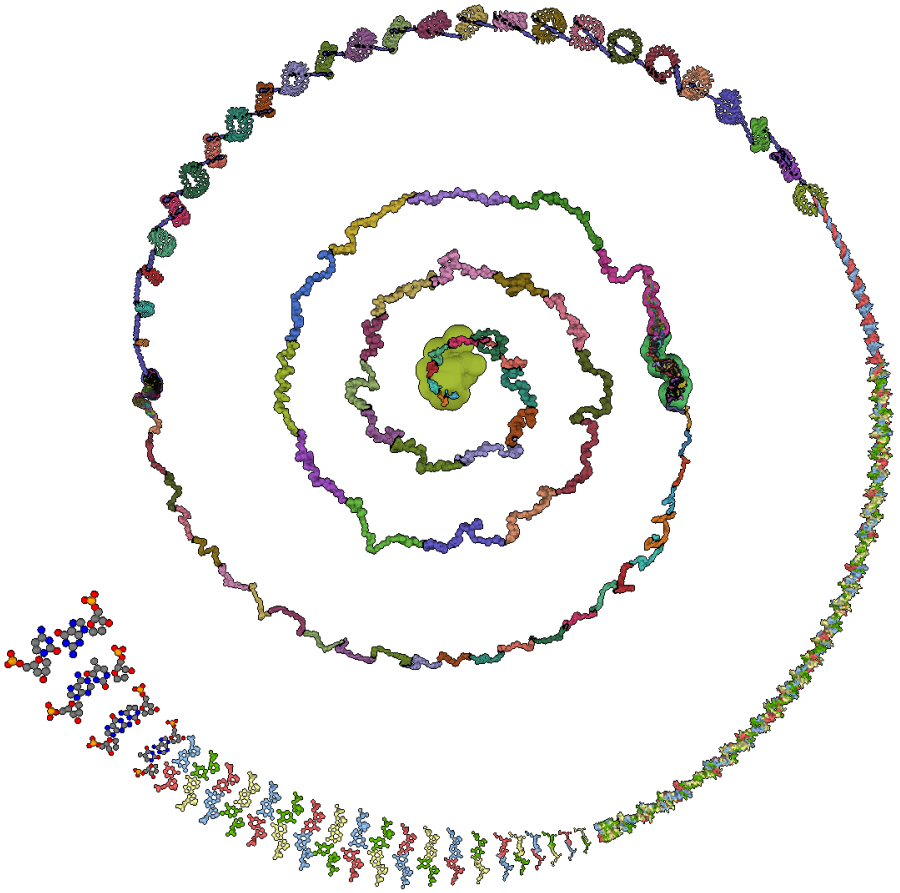


Figure 51: Example for placing our Multiscale Unfolding genome visualization on a spiraling path, from a single chromosome up to the nucleobases and their (flattened) atomic composition.

Another important direction for future work is to further explore interaction possibilities with Multiscale Unfolding to support other fundamental visualization tasks. This concerns the comparisons between different DNA sections on the same or different scales, maybe through an adjustable scale-factor function. Such a function might provide the possibility to control the scope of a given scale and to compare DNA parts that are far apart. Also comparing parts from different chromosomes could develop further beyond depicting two Multiscale Unfoldings side-by-side. Also the interaction with currently non-visible data parts can be addressed. Image-space arrangements other than the zig-zag layout pose questions on how to adjust interaction and navigation, and if Multiscale Zliding can still be used. Further adapting the visual mapping to the specifics of scale interactions can be done. An example is to motion blur fast-moving scales.

An overarching goal is to combine and integrate spatially-controlled Multiscale Unfolding with interactively- or temporally-controlled, traditional multiscale visualization. As we show already in Figure 39, the traditional equivalent of our spatially-controlled is ScaleTrotter [50],

and it works essentially perpendicular to our own visual mapping. Our vision is thus to be able to *pivot* between any selected point, for a given location on the sequence and a given scale (e. g., selected by a double-click), to the exact same data element shown at the same scale and with the same orientation in ScaleTrotter. We discuss about this hybrid framwork with more detail in Chapter 5. The resulting drastic view changes (except of the pivot element) likely require some new animation and visualization techniques to be developed, but together the two techniques would provide viewers with a powerful tool to explore both the spatial and the multiscale aspects of DNA data.

# 5 DISCUSSION

Each of the previous two chapters addresses one of the two important aspects of the human genome. In Chapter 3, we represented the spatial structure of the structural organization of the human genome on several organizational levels, using temporal transitions, while in Chapter 4, we modified the spatial structure to emphasize the linear structure of the genome, using spatial transitions. However, for a complete understanding of the activity of the genome, we need to be able to explore both its linear and spatial structures in a way such that they complement each other. Therefore, we discussed in Section 5.1 the possibility of merging our two frameworks in one and analyzed how seamless transitions between the two representations can be achieved. We also discussed how different user groups, students, illustrators, scientists and generally laypeople, can benefit from such a tool in Section 5.3. In Section 5.4, we discussed characteristics of the data that were not addressed in this thesis. We finished the discussion with Section 5.5, specifying what we learned in this thesis about visual abstraction.

## 5.1 PIVOTING BETWEEN SPATIAL AND TEMPORAL CONTROL OF ABSTRACTION

### 5.1.1 *Context*

The genetic information is encoded in DNA as a one dimensional sequence of nucleobases. Three consecutive nucleobases encode an amino acid, which is the building block of protein. The sequence of nucleobases encoding all the amino acids responsible for a protein is called a gene. Some sequences however do not code for genes, but for regulatory elements (promoter, enhancer, silencer, etc.) that regulate the transcription of neighboring genes. The sequences are neighbors if they are close to each other in the one-dimensional sequence. However, DNA does not exist in cells in a fully stretched state, but rather in an intertwined highly condensed state, giving birth to several organizational levels (Figure 12). Therefore, in any of those organizational levels, sequences that are are far from each other in the linear dimension, could be brought close to each other and thus become neighbors. The regulatory elements can even regulate genes far away in the sequence.

The spatial organization of the genome therefore plays an important role in its activity. It can also be used as a marker of disease as

some genes and chromosomes occupy distinct nuclear positions in diseased cells, compared to their normal counterparts [88] (Figure 1). In terms of spatial arrangement, the genome then presents two important aspects. First, the order in which the nucleobases come one after the other (1D) makes the genes that command the making of proteins. Second, in addition to the sequential distance, the spatial organization (3D) affects the spatial distance between regulatory elements and their target genes. Both aspects play a crucial role in the genome scientist's quest to fully understand the human genome.

### 5.1.2 *The tension between temporal and spatial control of abstraction*

In ScaleTrotter (Chapter 3), we visually represent the 3D structure of the human genome on multiple organizational levels. This allows users (whether experts or laypeople) to explore the spatial structure, and therefore the spatial distances that exist between different regions of the DNA. However, the tight packing makes it impossible to know sequentially how far two spatially close regions are to each other. The problem of not recognizing the sequence arises at the base level, as well as at the higher levels (Figure 52).

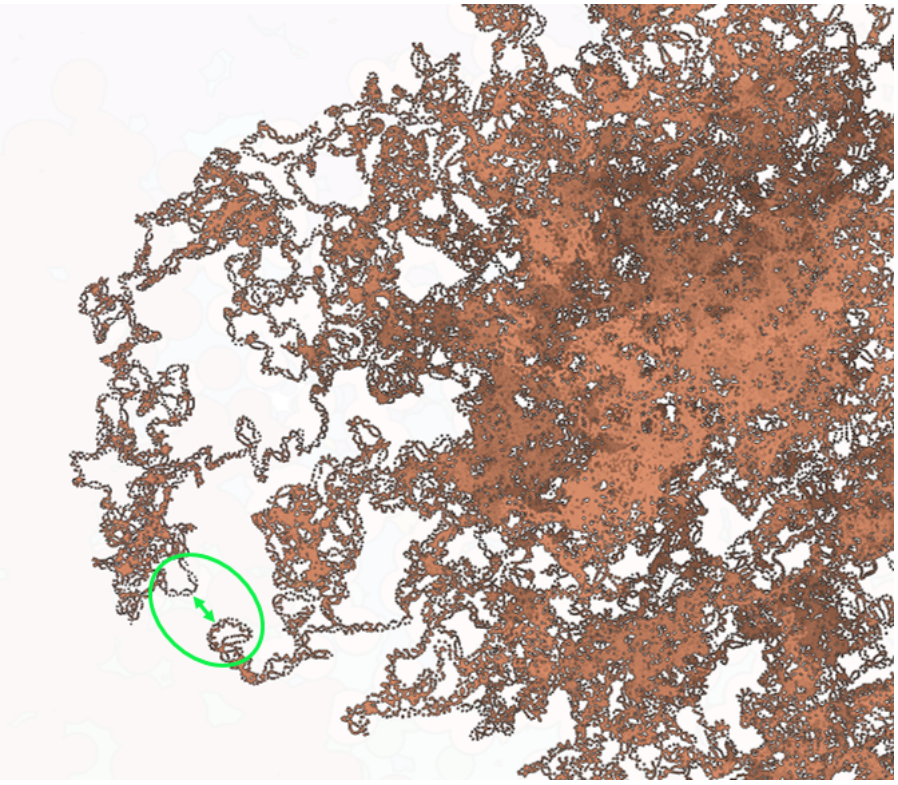


Figure 52: Two regions (shown by the arrow) that are spatially close, but their sequential distance is (almost) impossible to know.

The organizational levels cover such a range in orders of magnitude of space scale that it does not make sense to represent the data on

different levels, while keeping the 3D structure intact. In ScaleTrotter, we use temporal transitions to traverse all the organizational levels, through user interactions. Therefore, exploring the data on multiple levels simultaneously is impossible.

In an attempt to resolve these issues, we presented another framework called Multiscale Unfolding in Chapter 4. In this framework, the focus is on the sequential aspect of the data. We spatially straighten each conceptual level and render the lower level. Our technique ensures that the connectivity of the original data is maintained. We position the different levels on separate rows in a zig-zag fashion. We also modify the sizes of the elements to represent them at acceptable sizes Figure 39. Otherwise, the lowest scale would be $10^4$–$10^5$ times smaller than the highest. To compensate for the sacrifice of the 3D spatial configuration made to reveal the linear structure, we added an interaction that shows through color mapping the spatial distances between the clicked and the other visible elements Figure 49.

### 5.1.3 *A hybrid framework*

In both of the approaches, we made sacrifices to focus on one of the two aspects of genome data: the linear sequence and the spatial organization. These two aspects complimenting each other, we now propose a transition between the two representations, in a framework that combines the previous two.

A user exploring the 3D structure of the data might need to examine the linear sequence. The user then switches the view to the representation where the data is straightened. Vice versa, a user exploring the linear sequence might need to check for other regions that are physically close to the investigated part. We believe that continuous transitions are possible both ways.

In the 3D representation, at some levels, the user already picks an element as the focus element whose lower levels could be investigated further (element in red in Figure 53a). We can extend this concept to all the scales because a focus element is necessary to transition between representations.

At this point, if the user decides to transition to the linear representation, the element in focus would keep its full structure and scale (element in red in Figure 53). In our 3D representation, all the visible elements are already at the same scale. A fixed number of elements sequentially to the right and to the left of the focus element, if visible, would slowly transition to the positions they will occupy in the linear representation, being straightened one by one, starting from the first neghbors of the focus element (elements in blue in Figure 53).

In the 3D representation as it stands in ScaleTrotter, at the nucleosome semantic level, the neighboring elements that would appear in the linear representation at the same scale might not all be visi-

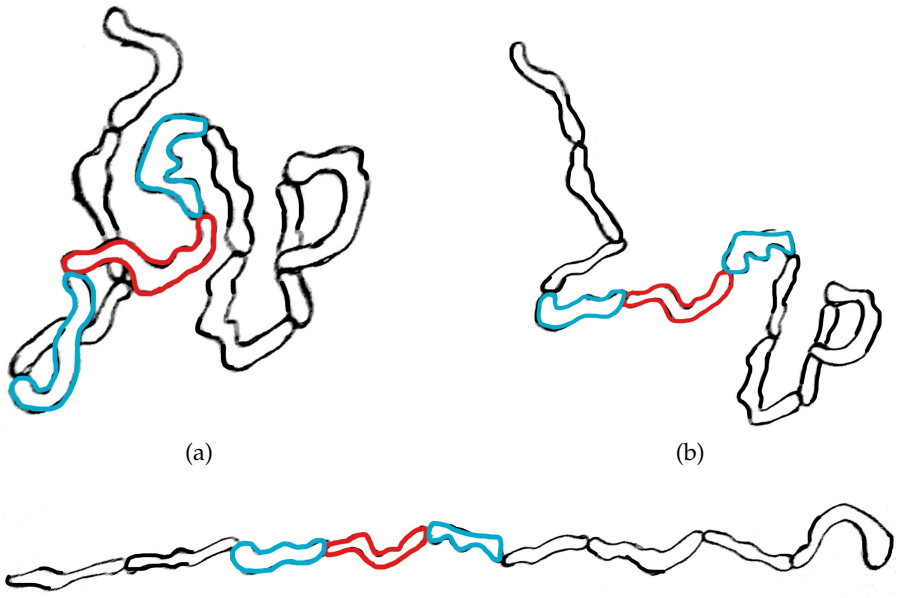



Figure 53: Pivoting around the focus element (red), from the 3D representation (a), to the linear representation (c). (b) shows an intermediate step where, apart from the focus element (red), one element on each side (blue) has been straightened.

ble. When transitioning to the linear representation, these elements can gradually appear in their final positions, or we can change the concept of removing elements in ScaleTrotter in a way that ensures the visibility of the neighbors that would appear in the linear representation. We envision the latter to be the better option, because it would make transitions on all levels similar, without the need of a new "appearing" transition.

A zooming out and a rotation of the camera might be necessary to present the focus element in the center and to fit the fixed number of elements on screen. Techniques for automatic camera control in large biological datasets has been investigated by Kouřil et al. [71].

For the remaining elements that are sequentially further on the left and on the right than the ones that would appear at the same scale as the focus element, several options can be considered. These elements can either be represented as part of their higher level parents or dissected into their lower level children on a row on top or on the bottom of the focus element's. Another option would be to represent them on other rows but still at the same scale. For the highest (resp. lowest) scale, the elements can only be represented at the lower (resp. higher) or same scale. In all possible cases, however, the user should be able to control the scales at which the elements would appear, as it is not mandatory for all scales to appear. We discuss the control of the scales in more detail in the following section.

We consider now that, in the linear representation, the user chooses a different element as element of focus and wants to transition back to the 3D representation. In a similar way, the focus element would stay at the same scale and translate to the center of the screen. All other elements would translate to their respective positions as in the original data, while transitioning into the scale of the focus element (if not at the same scale already). However, this transition is stricter in the sense that the focus element cannot keep the same structure as in the linear representation, because the 3D representation is the one based on the data. The focus element, therefore, has to change its structure and represent the data. The fact that it is in the center of the screen might not be enough to ensure that it can still be perceived in the 3D representation, because it can be occluded in that view. A certain highlighting or change of camera position might be necessary to ensure the visibility of the focus element in the 3D representation.

## 5.2 CONTROL OF THE SPATIAL ARRANGEMENT OF MULTISCALE UNFOLDING

As we already stated in Chapter 1, our goal with the work described in this thesis is to understand how visual abstraction can be controlled. Therefore we want to give users as much control as possible. In ScaleTrotter, we provide interested users, particularly illustrators, with the ability to control the zoom level at which a transition to a different organizational level should happen. We cannot, however, provide control over the positions of the elements because they are based on the data.

### 5.2.1 *Control over the path*

That is not the case however in Multiscale Unfolding, as the data is used to straighten the levels but the elements are placed on a path arbitrarily (in our case, on zig-zag rows covering the screen). However, the optimal path varies according to the task the user wishes to accomplish. Therefore, we envision giving users the freedom to draw the path they wish to see the genome on.

We imagine a drawing mode where the user draws a path by a simple click and drag of the mouse. When the user releases the mouse button, the path is fitted into a smooth spline. The spline provides the user with control points, that divide the spline into different parts (Figure 54). By clicking and dragging a control point, the user moves that particular control point (Figure 55a, Figure 55b). If the mouse interaction happens with a specific keyboard button pressed, the user can move the control point and all the control points that precede (or follow) it (Figure 55c, Figure 55d).

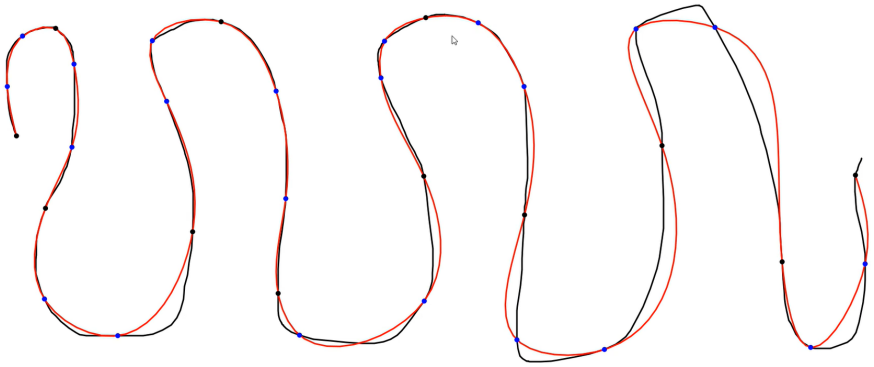


Figure 54: Hand-drawn path by the user (in black), fitted into a smooth spline (in red). Control points are in black and blue.

### 5.2.2 *Control over the scales*

The black control points in Figure 54 divide the spline into separate parts, each for representing elements at a different level. So far, moving from one area to the other in the linear sequence is possible only through the Multiscale Zliding technique we described in Section 4.5.2. If the user wants to see a different region at the lowest scale, he/she has to pan along that scale or a higher scale (to go faster). The user could also transition to the 3D representation, travel in space to the target element, and switch back to the linear representation. A better approach would to be to give the user the ability to immediately decide which part of the data has to be represented at which scale, in the parts separated by the black control points. A transition zone would automatically be created between the levels, even if the levels chosen by the user are not consecutive.

The continuity of the different scales could also be interrupted, leading to the disappearance of transition zones, if the user chooses to see either the same part of the data on several levels, or parts belonging to different chromosomes.

The spline path is accompanied by two parallel splines on either side Figure 55. Those parallel splines are intended to control the size at which the data appears in. Each control point on the parallel lines can be moved perpendicularly away or towards the center path. Both lines move symmetrically from the center path, at fixed distances corresponding to the scales that are available in the data. The further the parallel lines are from each other, the bigger the represented data would be.

The control over both the path and the scales can also be imagined as part of the framework that combines the 3D and linear representations. The concepts discussed in Section 5.2.1 and Section 5.2.2 can be applied on the linear representation side of the framework, and

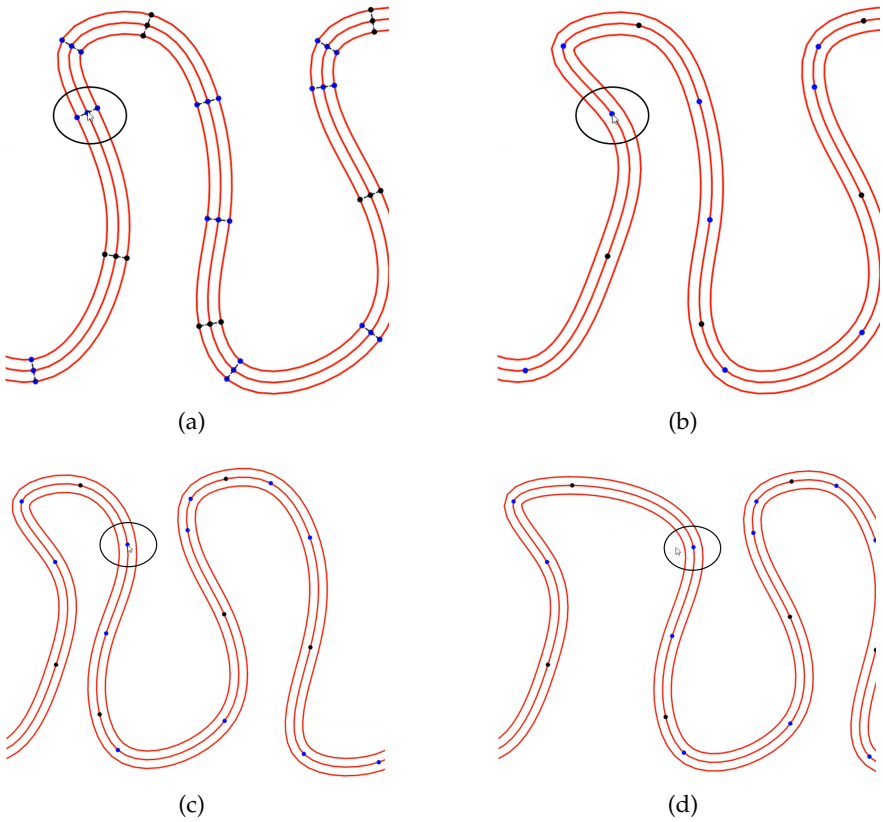


Figure 55: User moving the control points. (a) and (c) the control points are in their original positions. (b) the control point in the black circle moves. (d) all the control points following the point in the top circle move.

disappear in the 3D representation where the conditions of following the data are much stricter.

## 5.3 APPLICATION DOMAINS

We envision that our frameworks can be used by several different user groups. Domain experts would benefit from such a tool combining the spatial and linear structures of genome data. An expert exploring the spatial structure of the data might want to explore the linear sequence of two spatially close regions (for example, the circled regions in Figure 52). He/she could highlight those regions in the 3D structure and transition to the other view, where he/she could investigate the linear sequences of both of these regions. This interaction can also be applied on any regions, to compare their linear sequences. To make the tool more useful to them, additional domain related data (gene locations and names, base names, etc.) (like in Figure 13 and Figure 14) should be overlaid on top of the existing visualizations.

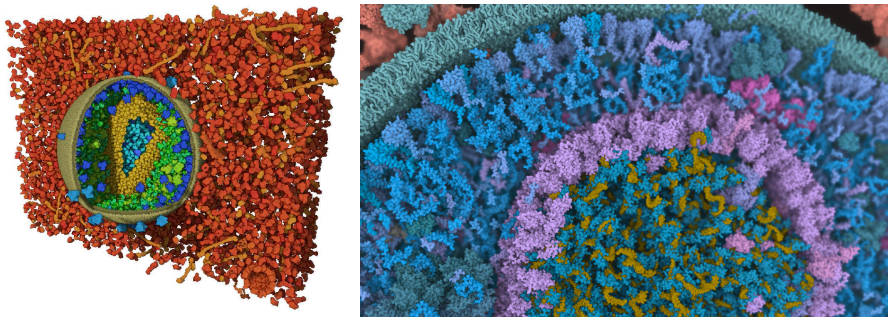
Exploring the genome is not only a task for domain scientists. General laypeople (including high-school students) might be interested in learning the organization and structure of the human genome. Our tools allow them to directly interact with the data, which is not possible with static illustrations or videos. The interaction allows them to get a grasp of the multiscale aspect of the genome. Particularly, the transition zones in our tools allow for a better understanding of how the DNA rolls and folds into the different organizational levels, and how the different levels relate to each other. ScaleTrotter, in addition, can give them a clear understanding of the scale differences and the sheer length and amount of information contained within the genome, along with a clearer understanding of the interphase chromosome configuration, compared to the more traditionally depicted mitotic chromosome (H or X shape) which is shown in virtually all illustrations, including the ones we show in Figure 38.

Finally, as illustrations have traditionally been communicating scientific concepts to the general audience, illustrators can also benefit from our tools. In order to create an illustration, illustrators have to first understand the concepts they want to depict themselves. Our tools can serve as the basis on which experts inform illustrators about the characteristics of the data. As reported in Section 3.5.2, illustrators considered our tool as part of a general story to tell about the cell. Illustrators can also create a draft of an illustration using these tools. They could export images from the visualization, sketch over them and add their own artistic styles to get a final illustration, as reported in Section 4.6.2. With the unfolded representation, they would have the freedom to draw any path, and choose any order of scale levels.

## 5.4 CHALLENGES OF THE DATA

The data we rely on in this thesis is a sample data created with the combination of different data sources by Asbury et al. [5] and transformed into the format that we use by Nowotny et al. [98]. Certain shortcomings are present in the data, the most obvious of which is the incorrect number of chromosomes (compared to a normal genome). However, the tools would still work if we replace the current data with an updated one, in case the ongoing research of the domain allows scientists to capture exact positions of atoms (or at least bases or nucleosomes) inside a given nucleus. Even with simulated data, our visualizations explain well the intended concepts, as do other visualizations of many biological phenomena relying on simulations. Such generated data serves as the basis of many of today's illustrative visualization tools such as, to name just one example, cellVIEW [76], which uses generative approaches (e.g., [68], [69]). Our approach thus allows us to integrate our work with cell visualizations such as the ones generated by cellVIEW (Figure 56). For scientific data anal-

ysis applications by domain experts, of course, we rely on detailed datasets to become available in the future that contain this information.



(a) (b)

Figure 56: Visualizations of generated molecular datasets. (a) from [77]. (b) from [68].

An aspect of the data that we disregard in this thesis is its variability over time. The organization of the genome is dynamic, meaning not only do the positions of elements change, but also their hierarchical belongings. A certain element at a particular level could belong to different elements at the higher level over time. The temporal aspect could be included in future molecular dynamics simulations. This implies that time is another dimension that we should consider in our tools. New smooth transitions might be necessary to travel through time and explore the changes. These transitions could even cover the further condensation of the interphase chromosome that we represent to the mitotic chromosome (i.e., the typical H/X representation depicted in most illustrations today).

## 5.5 ABSTRACTION AND ILLUSTRATIVE VISUALIZATION

While creating the described two frameworks in the context of this thesis, we were inspired by existing illustrations and techniques frequently used in illustration, namely visual abstraction. We applied visual abstraction by transforming, as Viola et al. [126] define, data representations into visual representations.

The genome data that we choose to visualize is divided into multiple scales. Therefore, in Chapter 3, we applied a visual abstraction separately for each scale in the data. We then use smooth temporal transitions to go from the abstract representation of an element at one level to the abstract representation of the element at the next level. The transitions create intermediate abstract representations, that superimpose the previous and the next. Viola and Isenberg [128] introduce the concept of an abstraction axis, as the succession of visual representations that continuously decrease the amount of information at

each step. At first glance, the visual abstractions of this chapter may look like they form a continuous abstraction axis, covering different structural representations of the same model (similarly to Figure 20). However, as the scales cover a large range of orders of magnitude, we have to move the camera and get closer to the model to see the lower levels. By zooming in, we can represent certain elements at the lower levels (more detail), but other elements fall out of its field of view or we intentionally remove them to avoid visual clutter (less context). The visual representations and the transitions connecting them, therefore, are not strictly removing (or strictly adding) information at each step. When traveling from the nucleus to the atoms, we gain information at specific locations (reverse structural abstraction), but at the same time, lose information entirely leave out parts of the data (contextual abstraction). The representations and transitions of this chapter thus rather support the updated definition of an abstraction axis provided by Viola et al. [126], where transitions along an abstraction axis can remove one type of visual detail and replace it with another type. Our abstraction axis is also constructed with a given purpose and application case to illustrate multiscale genome data.

In Chapter 4, we first apply transformations to the data before visually representing it. We abstract the 3D structure of the genome by straightening its elements and placing them on a line next to each other to focus the attention of the viewer on the linear structure of the genome. We also modify the sizes of elements to make them visible at all levels simultaneously. We compensate the size modification with the spatial transitions that, similarly to the temporal transitions, create intermediate abstract representations connecting two levels and displaying the relative size difference between them. In Chapter 4, we demonstrate that abstraction axes cannot only be explored temporally but also spatially. As we follow the elements from the highest level (chromosome) to the lowest (atoms), we continuously unfold the genome and increase the size of the elements to make them visible. Therefore, we get a more abstract 3D structure at each step. However, we also represent elements (or parts of elements) with increasingly more detail until we represent the individual atoms. We have then a spatial abstraction axis, along which some visual information is gained, while another is lost [126].

Generally with multiscale data, it thus seems that a purely mathematical abstraction axis, as defined by Viola and Isenberg [128], typically cannot be identified. Specifically for data covering several orders of magnitude in scale, all the levels typically cannot be examined from one particular point of view, or following one specific arrangement of the data. To explore all the levels, camera movements and/or modification of the data might therefore be necessary. This leads to gaining information in a certain aspect and losing it in a different one, which makes it hard (or potentially even impossible) to tell if informa-

tion in general is lost or gained. Instead, abstraction axes (including ours) are guided by the application and the designer's decisions to convey intended aspects of the data.

We introduced abstraction as a concept that is virtually always present in illustrations, and illustrative visualization as a subfield of visualization that borrows techniques (abstraction being one of those techniques) used in traditional illustrations. Viola and Isenberg [128] define illustrative visualization as a mix of two concepts, abstraction and stylization. These two concepts are not independent from each other, as one can imply the other.

If abstraction is the removal of information, stylization is the collection of expressive techniques that make a visualization illustrative, without removing information. It adds a character to the visualization that makes it immediately recognizable as an illustration, with the intention of getting viewers into the mindset that they are seeing a representation that purposefully illustrates a concept, as opposed to only depicting the data as it is. Visualizations representing concepts that cannot usually be seen with the naked eye, including the visualizations of this thesis, can benefit from this stylization component because the concepts they depict have no "photorealistic" representations. Indeed, color coding the atoms or the bases does not remove information but helps users identify and differentiate between structurally similar elements. Visualizations with spatial abstraction axes also contain the stylization that makes them immediately recognizable as illustrations.

In this thesis, we focused on abstraction, transitions between different abstract representations and therefore abstraction axes. We did not explore different styles and transitions between them. Our representations of the data with spheres and the coloring scheme correspond to a certain style in molecular illustration [44], imposed by the system that we used [94]. The abstractions and transitions are all done following this same style.

We already established that visualizations of multiscale data, covering a large range of orders of magnitude in scale, have to almost always make use of abstraction. If the viewer can tell that the visualization is not a real image but rather an illustration, then the visualization is an illustrative visualization.

# 6 CONCLUSION

The work presented in this thesis takes the first steps towards a tool that allows the full exploration of genome data. The human genome consists of several organizational levels covering 4–5 orders of magnitude in space scale. We propose a couple of approaches, based on the concept of continuous visual abstraction, each focusing on one of the two important aspects of the genome: linear and spatial. We also discuss the possibility of combining the two approaches in a framework that allows the exploration of both aspects.

In the background section (Chapter 2), we detailed the organizational levels of the human genome and explained their role in modulating crucial biological processes. We also looked at existing tools used by experts that typically display linear sequences as horizontal tracks and some spatial information by creating visual links between sequences. Other tools that visualize the multiscale 3D structure lack the smooth transitions between them. Our approaches provide these missing aspects to the users. Next, we reviewed the use of controlled abstraction in visualization, particularly in molecular visualization. The abstractions in these examples could be used on individual scales, but do not infer large scale changes. Therefore, they could not be used with the multiscale genome data, with conceptually different levels. Finally, we reviewed work that investigates abstraction in multiscale visualizations, with a focus on transitions between scales and cross scale interactions. The scale differences in molecular multiscale visualizations were not as large as in the data that we wanted to visualize. Astronomical multiscale data cover large scales, but present fundamentally different characteristics to genome data.

We therefore introduced a new temporal transition in Chapter 3, called visual embedding, that superimposes two representations of the same element at consecutive scales and allows users to understand how the scales relate to each other. We also gave the users the control over at what distance from the camera the transitions should happen. By continuously applying this transition, we allowed the exploration of the full spatial data while seamlessly traversing the organizational levels from the nucleus of the cell to the atoms of the bases. We thus successively transitioned from the abstract representation of an element at a scale to the abstract representation of the element at the next scale, with an abstract representation combining the two in between (Figure 33). To avoid visual clutter and ensure interactivity, we removed unnecessary parts of the data at certain levels.

This approach focused on the spatial structure of the human genome, but the dense packing makes it impossible to follow the linear sequence at all the levels. Therefore, we presented another framework in Chapter 4, that unfolds the organizational levels one by one and straightens the data. This is an abstract representation of an element where the underlying data has been modified, to focus the attention on the linear aspect. Since the 3D structure is already not conserved, this framework gives an overview of all the organizational levels, in contrast with the previous framework. We took abstraction a step further by not respecting the relative sizes of the elements on different scales. To make all levels visible at once, the sizes had to be modified to be made comparable. This modification of the size is compensated by the use of spatial transitions between the scales. The spatial transitions are based on the same concept of visual embedding, that superimposes two representations of an element at consecutive scale, showing the size difference. The size difference is also emphasized to the interaction, where moving elements at a particular scale causes a faster (resp. slower) movement at the lower (resp. higher) scales.

Based on Chapter 3 and Chapter 4, we noticed that in order to show one aspect of the data, we had to sacrifice the other. However, the linear and spatial structures are complementary and both essential for getting a complete understanding of the genome and its activity. We therefore discussed, in Chapter 5, the possibility of combining our two frameworks to minimize the losses suffered in each individual framework. We introduced a smooth pivoting that allows the transition between the 3D structure and the unfolded linear structure of the human genome, and vice versa. We also discussed a couple of methods meant to provide users with more control over the scales and their spatial arrangement.

In the final section of Chapter 5, we discussed what we learned about visual abstraction in the thesis. Our frameworks reinforced Viola et al.'s [126] definition of an abstraction axis, which, contrary to mathematical axes, does not have to incrementally remove information. An abstraction axis can remove a certain type of information while adding another. We also showed that we can explore abstraction axes not only temporally but spatially as well, by introducing new examples of temporal and spatial transitions between abstract representations.

Whether with temporal or spatial exploration of the abstraction axis, the nature of multiscale data extending several orders of magnitude is such that all scales cannot be visible simultaneously without the use of abstraction. We either have to modify the point of view by zooming in, to see detail which implies that other elements will fall out of the field of view. Or we have to modify the data to visualize distant elements or elements of huge size differences simultaneously. In both cases, we encounter abstraction axes where data detail increases

while context decreases. The same is true for maps or astronomical data. The visualization of multiscale data covering several orders of magnitude in scale cannot therefore work without abstraction.

The increase of data detail and decrease of context leads as to an abstraction space created by two independent axes. Multiscale data could be thought of as a hierarchy with elements of the lowest level being its basic components, that combine to form an element at the first higher level and so on until we get to one final element, the root of the hierarchical tree, at the highest scale. The first axis is therefore the axis of degree of data detail. The first (most concrete) point of this axis corresponds to a representation of the highest element with the elements at the lowest level. The last (most abstract) point of the axis corresponds to the representation of the highest element as one shape (or object—one sphere being the genome in our case). The second axis is the amount of context. By context we mean how much of the hierarchical tree is seen in the view. We consider that, if an element appears in the view, all its children are also, by default, in the view (even if they are not individually represented). On this axis, the first (most concrete) point corresponds to a view where the highest element can be seen (no matter at which detail level). The last (most abstract) point corresponds to a view showing only an element of the lowest level. As the genome multiscale data covers many orders of magnitude in scale, it does not make sense to depict the whole genome at the atomic level. Therefore, in the abstraction space formed by these two orthogonal axes, our tools transition from the (most abstract data detail, lowest abstract context) point to the (least abstract data detail, most abstract context) point, creating a diagonal path in the abstraction space (Figure 57).

It typically does not make sense to explore the abstraction space further, outside of this path shown in Figure 57, as it would not allow us to gain much additional insight. For the upper right side of the path, for example, it would mean to represent elements of a level with elements of the higher level (a portion of a sphere in our case), which does not make sense (Figure 58). Exploring the lower left side of the path would require unrealistically large screen sizes. For example, if we represent the whole genome at the atomic scale, on a regular screen, individual atoms cannot be recognized. A way around this would be to use an extremely large high-res wall display, and then the abstraction path would be traversed by the viewer going closer or further from the display. For the case of the tightly packed genome, however, even with the large display, we would still have too much detail to actually understand anything, which could only be solved by adjusting the visualization using perceptually hybrid representations [57].

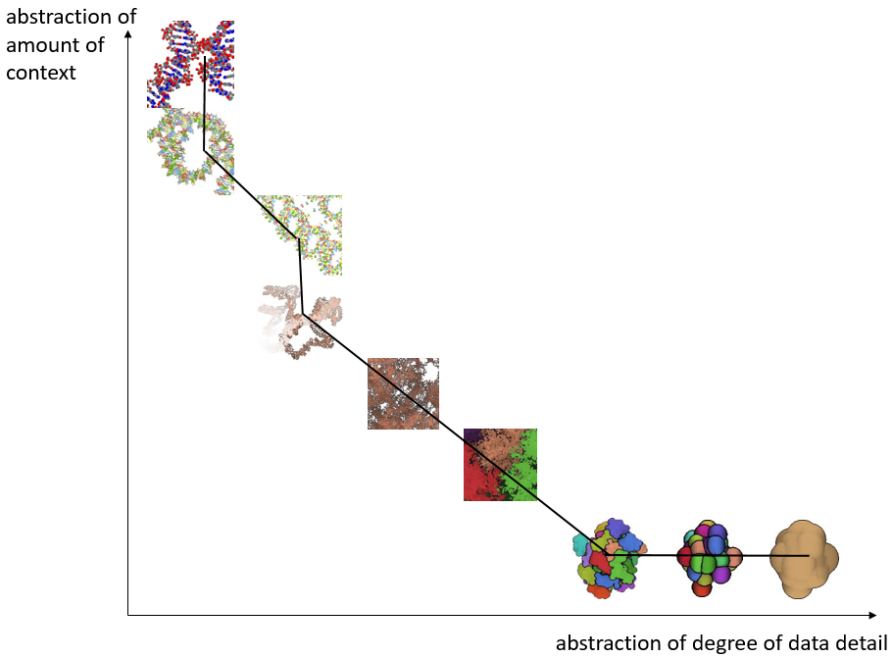



Figure 57: Abstraction space created by two orthogonal axes. The axis "abstraction of degree of data detail" goes from representations where the rendered elements are atoms to representations where the rendered element is the genome. The axis "abstraction of amount of context" goes from a representation of the whole genome to representations of a single atom. The line in black represents the path ScaleTrotter takes in the abstraction space.

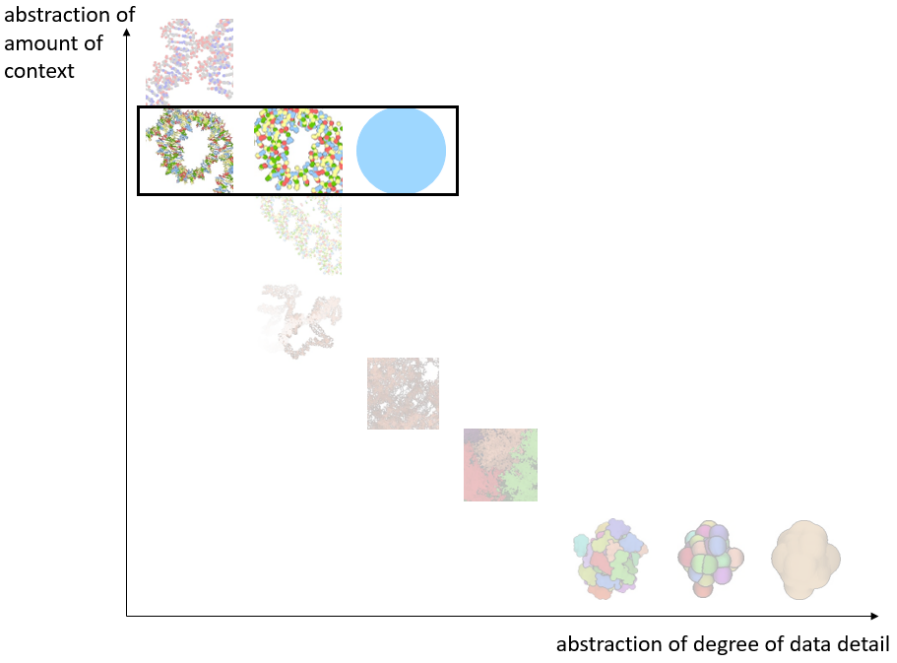



Figure 58: Focus on a horizontal line in the abstraction space. We can see three representations of a nucleosome (fixed point on the amount of context axis), with various degree of data detail. It does not make sense to go further to the right.

# A

## ACKNOWLEDGMENTS

The writing of this thesis and the papers upon which it is based would not have been possible without the participation of several parties that we would like to thank.

We thank the genome scientists who provided the data that our tool relies on and who answered our questions about it. We also thank all experts who provided feedback about our approach and were available for interviews.

Part of this work was funded under the ILLUSTRARE grant by both the Austrian Science Fund (FWF): I 2953-N31 and the French National Research Agency (ANR): ANR-16-CE91-0011-01. The research was further supported by funding from King Abdullah University of Science and Technology (KAUST), under award number BAS/1/1680-01-01 and by funding from ILLVISATION grant by WWTF (VRG11-010).

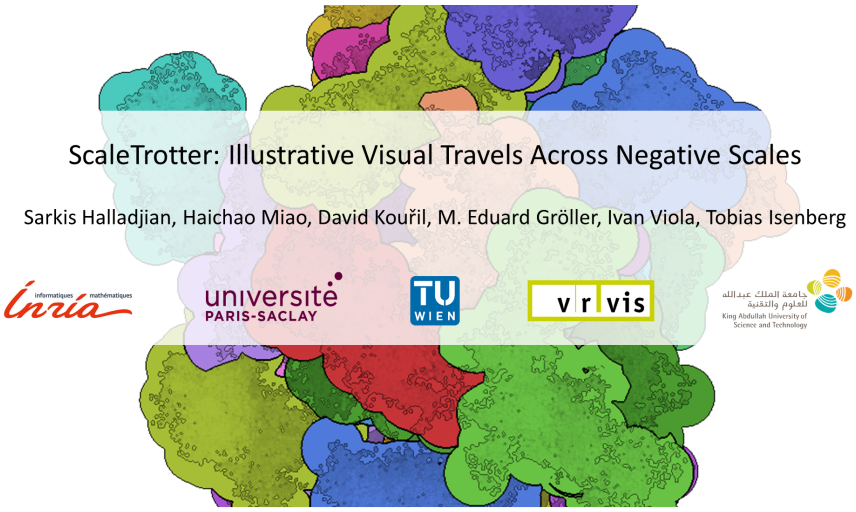
We would like to thank Nanographics GmbH (nanographics.at) for providing the Marion Software Framework.

The papers written during the course of this thesis were partly written in collaboration with the VRVis Competence Center. VRVis is funded by BMVIT, BMWFW, Styria, SFG and Vienna Business Agency in the scope of COMET – Competence Centers for Excellent Technologies (854174), which is managed by FFG.

# B

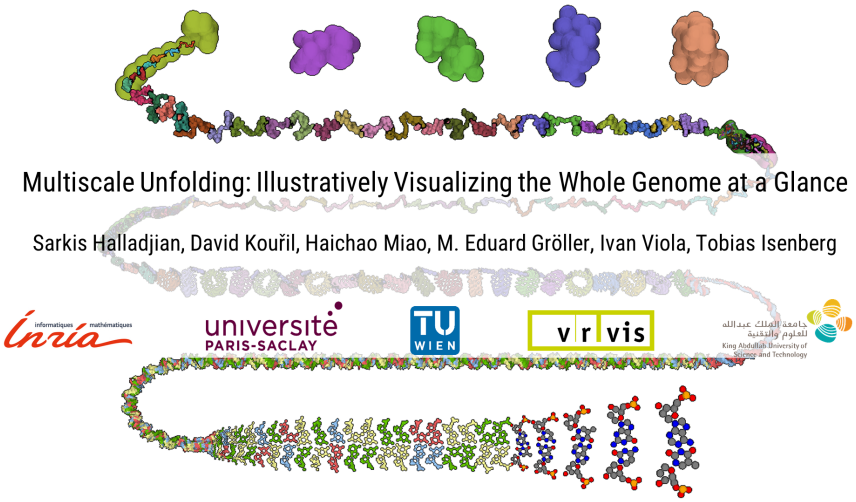
## VIDEO LINKS

### ScaleTrotter: Illustrative Visual Travels Across Negative Scales




Link to video: https://youtu.be/Ira54e3ND74

### Multiscale Unfolding: Illustratively Visualizing the Whole Genome at a Glance




Link to video: https://youtu.be/AGxuxWz56yo

# BIBLIOGRAPHY


[1] Maneesh Agrawala and Chris Stolte. "Rendering Effective Route Maps: Improving Usability Through Generalization." In: *Proc. SIGGRAPH*. New York: ACM, 2001, pp. 241–249. DOI: 10.1145/383259.383286.

[2] Bruch Alberts, Alexander D. Johnson, Julian Lewis, David Morgan, Martin Raff, Keith Roberts, and Peter Walter. *Molecular Biology of the Cell*. 6th. New York: Garland Science, 2015. DOI: 10.3390/ijms161226074.

[3] N. Alharbi, M. Alharbi, X. Martinez, M. Krone, M. Baaden, R. S. Laramee, and M. Chavent. "Molecular Visualization of Computational Biology Data: A Survey of Surveys." In: *EuroVis Short Papers*. Goslar, Germany: Eurographics Association, 2017, pp. 133–137. DOI: 10.2312/eurovisshort.20171146.

[4] Anthony T. Annunziato. "DNA Packaging: Nucleosomes and Chromatin." In: *Nature Education* 1.1 (2008), Article 26. URL: https://www.nature.com/scitable/topicpage/dna-packaging-nucleosomes-and-chromatin-310.

[5] Thomas M Asbury, Matt Mitman, Jijun Tang, and W Jim Zheng. "Genome3D: a viewer-model framework for integrating and visualizing multi-scale epigenomic information within a three-dimensional genome." In: *BMC Bioinformatics* 11.1 (2010), p. 444. DOI: 10.1109/BIBMW.2011.6112510.

[6] Oswald T. Avery, Colin M. MacLeod, and Maclyn McCarty. "Studies of the chemical nature of the substance inducing transformation of pneumococcal types. Induction of transformation by a deoxyribonucleic acid fraction isolated from pneumococcus type III." In: *Journal of Experimental Medicine* 79.2 (Feb. 1944), pp. 137–158. ISSN: 0022-1007. DOI: 10.1084/jem.79.2.137.

[7] Emil Axelsson, Jonathas Costa, Cláudio Silva, Carter Emmart, Alexander Bock, and Anders Ynnerman. "Dynamic Scene Graph: Enabling Scaling, Positioning, and Navigation in the Universe." In: *Computer Graphics Forum* 36.3 (June 2017), pp. 459–468. DOI: 10.1111/cgf.13202.

[8] Chandrajit Bajaj, Peter Djeu, Vinay Siddavanahalli, and Anthony Thane. "Texmol: Interactive visual exploration of large flexible multi-component molecular complexes." In: *IEEE Visualization 2004*. IEEE. 2004, pp. 243–250. DOI: 10.1109/VISUAL.2004.103.

[9] Aurélien Bancaud, Christophe Lavelle, Sébastien Huet, and Jan Ellenberg. "A fractal model for nuclear organization: Current evidence and biological implications." In: *Nucleic Acids Research* 40.18 (July 2012), pp. 8783–8792. DOI: 10.1093/nar/gks586.

[10] Richard A. Becker and William S. Cleveland. "Brushing Scatterplots." In: *Technometrics* 29.2 (May 1987), pp. 127–142. DOI: 10.1080/00401706.1987.10488204.

[11] Thaddeus Beier and Shawn Neely. "Feature-Based Image Metamorphosis." In: *ACM SIGGRAPH Computer Graphics* 26.2 (July 1992), pp. 35–42. DOI: 10.1145/142920.134003.

[12] Lindsay M. Biga, Sierra Dawson, Amy Harwell, Robin Hopkins, Joel Kaufmann, Mike LeMaster, Philip Matern, Katie Morrison-Graham, Devon Quick, and Jon Runyeon. *Anatomy & Physiology*. USA: OpenStax/Oregon State University, 2019. URL: https://open.oregonstate.education/aandp/.

[13] K. Bladin, E. Axelsson, E. Broberg, C. Emmart, P. Ljung, A. Bock, and A. Ynnerman. "Globe Browsing: Contextualized Spatio-Temporal Planetary Surface Visualization." In: *IEEE Transactions on Visualization and Computer Graphics* 24.1 (Jan. 2018), pp. 802–811. DOI: 10.1109/TVCG.2017.2743958.

[14] Boyan Bonev and Giacomo Cavalli. "Organization and function of the 3D genome." In: *Nature Reviews Genetics* 17.11 (2016), p. 661. DOI: 10.1038/nrg.2016.112.

[15] Michelle A. Borkin, Krzysztof Z. Gajos, Amanda Peters, Dimitrios Mitsouras, Simone Melchionna, Frank J. Rybicki, Charles L. Feldman, and Hanspeter Pfister. "Evaluation of artery visualizations for heart disease diagnosis." In: *IEEE Transactions on Visualization and Computer Graphics* 17.12 (Dec. 2011), pp. 2479–2488. ISSN: 1077-2626. DOI: 10.1109/tvcg.2011.192.

[16] Matthew Brehmer and Tamara Munzner. "A Multi-Level Typology of Abstract Visualization Tasks." In: *IEEE Transactions on Visualization and Computer Graphics* 19.12 (Dec. 2013), pp. 2376–2385. DOI: 10.1109/TVCG.2013.124.

[17] Andreas Buja, John A. McDonald, John Michalak, and Werner Stuetzle. "Interactive Data Visualization using Focusing and Linking." In: *Proc. Visualization*. Los Alamitos: IEEE Computer Society, 1991, pp. 156–163. DOI: 10.1109/VISUAL.1991.175794.

[18] Mackinlay Card. *Readings in information visualization: using vision to think*. Morgan Kaufmann, 1999.

[19] Stuart K. Card, George G. Robertson, and Jock D. Mackinlay. "The Information Visualizer, an Information Workspace." In: *Proc. CHI*. New York: ACM, 1991, pp. 181–186. DOI: 10.1145/108844.108874.

[20] Robert Carnecky, Raphael Fuchs, Stephanie Mehl, Yun Jang, and Ronald Peikert. "Smart Transparency for Illustrative Visualization of Complex Flow Surfaces." In: *IEEE Transactions on Visualization and Computer Graphics* 19.5 (May 2013), pp. 838–851. DOI: 10.1109/TVCG.2012.159.

[21] Robert Carnecky, Benjamin Schindler, Raphael Fuchs, and Ronald Peikert. "Multi-layer Illustrative Dense Flow Visualization." In: *Computer Graphics Forum* 31.3 (June 2012), pp. 895–904. DOI: 10.1111/j.1467-8659.2012.03082.x.

[22] Erwin Chargaff. "Chemical specificity of nucleic acids and mechanism of their enzymatic degradation." In: *Experientia* 6.6 (1950), pp. 201–209. DOI: 10.1007/BF02173653.

[23] Fanny Chevalier, Romain Vuillemot, and Guia Gali. "Using concrete scales: A practical framework for effective visual depiction of complex measures." In: *IEEE Transactions on Visualization and Computer Graphics* 19.12 (2013), pp. 2426–2435. DOI: 10.1109/TVCG.2013.210.

[24] Gregory Cipriano and Michael Gleicher. "Molecular surface abstraction." In: *IEEE Transactions on Visualization and Computer Graphics* 13.6 (2007), pp. 1608–1615. DOI: 10.1109/TVCG.2007.70578.

[25] Gregory Cipriano, George N Phillips Jr, and Michael Gleicher. "Multi-scale surface descriptors." In: *IEEE Transactions on Visualization and Computer Graphics* 15.6 (2009), pp. 1201–1208. DOI: 10.1109/TVCG.2009.168.

[26] Sabine Coquillart. "Extended Free-Form Deformation: A Sculpturing Tool for 3D Geometric Modeling." In: *ACM SIGGRAPH Computer Graphics* 24.4 (Sept. 1990), pp. 187–196. DOI: 10.1145/97880.97900.

[27] Sabine Coquillart and Pierre Jancène. "Animated Free-Form Deformation: An Interactive Animation Technique." In: *ACM SIGGRAPH Computer Graphics* 25.4 (July 1991), pp. 23–26. DOI: 10.1145/127719.122720.

[28] Ralf Dahm. "Friedrich Miescher and the discovery of DNA." In: *Developmental Biology* 278.2 (2005), pp. 274 –288. ISSN: 0012-1606. DOI: 10.1016/j.ydbio.2004.11.028.

[29] Job Dekker, Karsten Rippe, Martijn Dekker, and Nancy Kleckner. "Capturing chromosome conformation." In: *Science* 295.5558 (2002), pp. 1306–1311. DOI: 10.1126/science.1067799.

[30] Evanthia Dimara and Charles Perin. "What is Interaction for Data Visualization?" In: *IEEE Transactions on Visualization and Computer Graphics* 26.1 (Jan. 2020), pp. 119–129. DOI: 10.1109/TVCG.2019.2934283.

[31] Jesse R Dixon, Siddarth Selvaraj, Feng Yue, Audrey Kim, Yan Li, Yin Shen, Ming Hu, Jun S Liu, and Bing Ren. "Topological domains in mammalian genomes identified by analysis of chromatin interactions." In: *Nature* 485.7398 (2012), pp. 376–380. DOI: 10.1038/nature11082.

[32] Charles Eames and Ray Eames. *Powers of Ten and the Relative Size of Things in the Universe*. Video. 1977. URL: http://www.eamesoffice.com/the-work/powers-of-ten/.

[33] Niklas Elmqvist, Yann Riche, Nathalie Henry Riche, and Jean-Daniel Fekete. "Mélange: Space Folding for Visual Exploration." In: *IEEE Transactions on Visualization and Computer Graphics* 16.3 (2010), pp. 468–483. DOI: 10.1109/TVCG.2009.86.

[34] Maarten H Everts, Eric Begue, Henk Bekker, Jos BTM Roerdink, and Tobias Isenberg. "Exploration of the brain's white matter structure through visual abstraction and multi-scale local fiber tract contraction." In: *IEEE Transactions on Visualization and Computer Graphics* 21.7 (July 2015), pp. 808–821. DOI: 10.1109/TVCG.2015.2403323.

[35] Martin Falk, Michael Krone, and Thomas Ertl. "Atomistic visualization of mesoscopic whole-cell simulations using ray-casted instancing." In: *Computer Graphics Forum*. Vol. 32. 8. Wiley Online Library. 2013, pp. 195–206. DOI: 10.1111/cgf.12197.

[36] Adam Finkelstein and David H. Salesin. "Multiresolution Curves." In: *Proc. SIGGRAPH*. New York: ACM, 1994, pp. 261–268. DOI: 10.1145/192161.192223.

[37] Chi-Wing Fu and Andrew J. Hanson. "A Transparently Scalable Visualization Architecture for Exploring the Universe." In: *IEEE Transactions on Visualization and Computer Graphics* 13.1 (Jan. 2007), pp. 108–121. DOI: 10.1109/TVCG.2007.2.

[38] George W. Furnas and Benjamin B. Bederson. "Space-Scale Diagrams: Understanding Multiscale Interfaces." In: *Proc. CHI*. New York: ACM, 1995, pp. 234–241. DOI: 10.1145/223904.223934.

[39] Genetic Science Learning Center. *Cell Size and Scale*. URL: https://learn.genetics.utah.edu/content/cells/scale/. Last visited March 2019.

[40] Michael Gleicher. "Considerations for Visualizing Comparison." In: *IEEE Transactions on Visualization and Computer Graphics* 24.1 (Jan. 2018), pp. 413–423. DOI: 10.1109/TVCG.2017.2744199.

[41] Michael Glueck, Keenan Crane, Sean Anderson, Andres Rutnik, and Azam Khan. "Multiscale 3D Reference Visualization." In: *Proc. I3D*. New York: ACM, 2009, pp. 225–232. DOI: 10.1145/1507149.1507186.

[42] Michael Glueck and Azam Khan. "Considering multiscale scenes to elucidate problems encumbering three-dimensional intellection and navigation." In: *AI EDAM* 25.4 (Nov. 2011), pp. 393–407. DOI: 10.1017/S0890060411000230.

[43] Johann Wolfgan von Goethe. *Faust: Eine Tragödie*. Tübingen, Germany: Cotta'sche Verlagsbuchhandlung, 1808. URL: http://www.deutschestextarchiv.de/book/show/goethe_faust01_1808.

[44] David S Goodsell, Shuchismita Dutta, Christine Zardecki, Maria Voigt, Helen M Berman, and Stephen K Burley. "The RCSB PDB "molecule of the month": inspiring a molecular view of biology." In: *PLoS Biol* 13.5 (2015), e1002140. DOI: 10.1371/journal.pbio.1002140.

[45] David S. Goodsell, Margaret A. Franzen, and Tim Herman. "From Atoms to Cells: Using Mesoscale Landscapes to Construct Visual Narratives." In: *Journal of Molecular Biology* 430.21 (Oct. 2018), pp. 3954–3968. ISSN: 0022-2836. DOI: https://doi.org/10.1016/j.jmb.2018.06.009.

[46] Mike Goodstadt and Marc A Marti-Renom. "Challenges for visualizing three-dimensional data in genomic browsers." In: *FEBS letters* 591.17 (2017), pp. 2505–2519. DOI: 10.1002/1873-3468.12778.

[47] Henry Gray. *Anatomy of the Human Body*. 20th. Philadelphia/New York: Lea & Febiger, 1918. URL: https://en.wikipedia.org/wiki/Gray's_Anatomy.

[48] Stefan Gumhold. "Splatting illuminated ellipsoids with depth correction." In: *VMV*. 2003, pp. 245–252.

[49] Sarkis Halladjian, David Kouřil, Haichao Miao, M. Eduard Gröller, Ivan Viola, and Tobias Isenberg. "Multiscale Unfolding: Illustratively Visualizing the Whole Genome at a Glance." In: *IEEE Transactions on Visualization and Computer Graphics* 27 (2021). To appear.

[50] Sarkis Halladjian, Haichao Miao, David Kouřil, M. Eduard Gröller, Ivan Viola, and Tobias Isenberg. "ScaleTrotter: Illustrative Visual Travels Across Negative Scales." In: *IEEE Transactions on Visualization and Computer Graphics* 26.1 (Jan. 2020), pp. 654–664. DOI: 10.1109/TVCG.2019.2934334.

[51] Andrew J. Hanson, Chi-Wing Fu, and Eric A. Wernert. "Very Large Scale Visualization Methods for Astrophysical Data." In: *Proc. VisSym*. Goslar, Germany: Eurographics Association, 2000, pp. 115–124. DOI: 10.2312/VisSym/VisSym00/115-124.

[52] Helwig Hauser. "Generalizing Focus+Context Visualization." In: *Scientific Visualization: The Visual Extraction of Knowledge from Data*. Ed. by Georges-Pierre Bonneau, Thomas Ertl, and Gregory M. Nielson. Berlin, Heidelberg: Springer, 2006, pp. 305–327. DOI: 10.1007/3-540-30790-7_18.

[53] Elaine RS Hodges. *The Guild Handbook of Scientific Illustration*. John Wiley & Sons, 2003.

[54] Wei-Hsien Hsu, Kwan-Liu Ma, and Carlos Correa. "A Rendering Framework for Multiscale Views of 3D Models." In: *ACM Transactions on Graphics* 30.6 (Dec. 2011), Article 131, 10 pages. DOI: 10.1145/2070781.2024165.

[55] Victoria Interrante. "Illustrating Transparency: Communicating the 3D Shape of Layered Transparent Surfaces via Texture." PhD thesis. USA: University of North Carolina at Chapel Hill, 1996. URL: https://cs.unc.edu/newspublications/doctoral-dissertations/abstracts-a-l/#Interrante.

[56] Rossitza N. Irobalieva et al. "Structural diversity of supercoiled DNA." In: *Nature Communications* 6 (2015), 8440:1–10. DOI: 10.1038/ncomms9440.

[57] Petra Isenberg, Pierre Dragicevic, Wesley Willett, Anastasia Bezerianos, and Jean-Daniel Fekete. "Hybrid-image visualization for large viewing environments." In: *IEEE Annals of the History of Computing* 12 (2013), pp. 2346–2355. DOI: 10.1109/TVCG.2013.163.

[58] Waqas Javed, Sohaib Ghani, and Niklas Elmqvist. "Polyzoom: Multiscale and Multifocus Exploration in 2D Visual Spaces." In: *Proc. CHI*. New York: ACM, 2012, pp. 287–296. DOI: 10.1145/2207676.2207716. URL: https://doi.org/10.1145/2207676.2207716.

[59] Alvin J. Johnson and Joseph H. Colton. *Johnson's New Illustrated (Steel Plate) Family Atlas, with Descriptions, Geographical, Statistical, and Historical*. New York: Johnson and Ward, 1862. URL: https://lccn.loc.gov/2006458033.

[60] M Gail Jones and Amy R Taylor. "Developing a sense of scale: Looking backward." In: *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching* 46.4 (2009), pp. 460–475. DOI: 10.1002/tea.20288.

[61] D. F. Keefe, D. B. Karelitz, E. L. Vote, and D. H. Laidlaw. "Artistic collaboration in designing VR visualizations." In: *IEEE Computer Graphics and Applications* 25.2 (Mar. 2005), pp. 18–23. DOI: 10.1109/MCG.2005.34.

[62] Daniel F. Keefe. "Integrating Visualization and Interaction Research to Improve Scientific Workflows." In: *IEEE Computer Graphics and Applications* 30.2 (Mar. 2010), pp. 8–13. DOI: 10.1109/MCG.2010.30.

[63] Daniel F. Keefe and Tobias Isenberg. "Reimagining the Scientific Visualization Interaction Paradigm." In: *IEEE Computer* 46.5 (May 2013), pp. 51–57. DOI: 10.1109/MC.2013.178.

[64] W James Kent, Charles W Sugnet, Terrence S Furey, Krishna M Roskin, Tom H Pringle, Alan M Zahler, and David Haussler. "The human genome browser at UCSC." In: *Genome research* 12.6 (2002), pp. 996–1006. DOI: 10.1101/gr.229102.

[65] Peter Kerpedjiev et al. "HiGlass: Web-based visual exploration and analysis of genome interaction maps." In: *Genome Biology* 19.1 (Aug. 2018), Article 125, 12 pages. ISSN: 1474-760X. DOI: 10.1186/s13059-018-1486-1.

[66] Staffan Klashed, Per Hemingsson, Carter Emmart, Matthew Cooper, and Anders Ynnerman. "Uniview – Visualizing the Universe." In: *Eurographics Areas Papers*. Goslar, Germany: Eurographics Association, 2010, pp. 37–43. DOI: 10.2312/ega.20101005.

[67] Thomas Klein and Thomas Ertl. "Illustrating Magnetic Field Lines using a Discrete Particle Model." In: *VMV*. Vol. 4. 2004, pp. 387–394. DOI: 10.1007/s12650-015-0300-3.

[68] Tobias Klein, Ludovic Autin, Barbora Kozlíková, David S Goodsell, Arthur Olson, M Eduard Gröller, and Ivan Viola. "Instant construction and visualization of crowded biological environments." In: *IEEE Transactions on Visualization and Computer Graphics* 24.1 (2017), pp. 862–872. DOI: 10.1109/TVCG.2017.2744258.

[69] Tobias Klein, Peter Mindek, Ludovic Autin, David S Goodsell, Arthur J Olson, EM Gröller, and Ivan Viola. "Parallel generation and visualization of bacterial genome structures." In: *Computer Graphics Forum*. Vol. 38. 7. Wiley Online Library. 2019, pp. 57–68. DOI: 10.1111/cgf.13816.

[70] R. Kosara, F. Drury, L. E. Holmquist, and D. H. Laidlaw. "Visualization Criticism." In: *IEEE Computer Graphics and Applications* 28.3 (May 2008), pp. 13–15. DOI: 10.1109/MCG.2008.63.

[71] David Kouřil, Tobias Isenberg, Barbora Kozlíková, Miriah Meyer, M. Eduard Gröller, and Ivan Viola. "HyperLabels—Browsing of Dense and Hierarchical Molecular 3D Models." In: *IEEE Transactions on Visualization and Computer Graphics* 27 (2021). To appear. DOI: 10.1109/TVCG.2020.2975583.

[72] Barbora Kozlíková, Michael Krone, Martin Falk, Norbert Lindow, Marc Baaden, Daniel Baum, Ivan Viola, Julius Parulek, and Hans-Christian Hege. "Visualization of Biomolecular Structures: State of the Art Revisited." In: *Computer Graphics Forum* 36.8 (Dec. 2015), pp. 178–204. DOI: 10.1111/cgf.13072.

[73] Barbora Kozlíková, Michael Krone, Norbert Lindow, Martin Falk, Marc Baaden, Daniel Baum, Ivan Viola, Julius Parulek, and Hans-Christian Hege. "Visualization of Biomolecular Structures: State of the Art." In: *EuroVis STARs*. Goslar, Germany: Eurographics Association, 2015, pp. 61–81. DOI: 10.2312/eurovisstar.20151112.061-081.

[74] J. Kreiser, M. Meuschke, G. Mistelbauer, B. Preim, and T. Ropinski. "A Survey of Flattening-Based Medical Visualization Techniques." In: *Computer Graphics Forum* 37.3 (June 2018), pp. 597–624. DOI: 10.1111/cgf.13445.

[75] Francis Lazarus, Sabine Coquillart, and Pierre Jancène. "Axial deformations: An intuitive deformation technique." In: *Computer-Aided Design* 26.8 (Aug. 1994), pp. 607–613. DOI: 10.1016/0010-4485(94)90103-1.

[76] Mathieu Le Muzic, Ludovic Autin, Julius Parulek, and Ivan Viola. "cellVIEW: a tool for illustrative and multi-scale rendering of large biomolecular datasets." In: *Eurographics Workshop on Visual Computing for Biomedicine*. Vol. 2015. NIH Public Access. 2015, p. 61. DOI: 10.2312/vcbm.20151209.

[77] Mathieu Le Muzic, Peter Mindek, Johannes Sorger, Ludovic Autin, David Goodsell, and Ivan Viola. "Visibility Equalizer: Cutaway Visualization of Mesoscopic Biological Models." In: *Computer Graphics Forum* 35.3 (June 2016), pp. 161–170. DOI: 10.1111/cgf.12892.

[78] Fritz Lekschas, Michael Behrisch, Benjamin Bach, Peter Kerpedjiev, Nils Gehlenborg, and Hanspeter Pfister. "Pattern-Driven Navigation in 2D Multiscale Visualizations with Scalable Insets." In: *IEEE Transactions on Visualization and Computer Graphics* 26.1 (Jan. 2019), pp. 611–621. DOI: 10.1109/TVCG.2019.2934555.

[79] Johannes Lemke et al. "The DNA-Based Structure of Human Chromosome 5 in Interphase." In: *The American Journal of Human Genetics* 71.5 (2002), pp. 1051–1059. DOI: 10.1086/344286.

[80] Apostolos Lerios, Chase D. Garfinkle, and Marc Levoy. "Feature-Based Volume Metamorphosis." In: *Proc. SIGGRAPH*. New York: ACM, 1995, 449–456. DOI: 10.1145/218380.218502.

[81] Wilmot Li, Lincoln Ritter, Maneesh Agrawala, Brian Curless, and David Salesin. "Interactive Cutaway Illustrations of Complex 3D Models." In: *ACM Transactions on Graphics* 26.3 (July 2007), Article 31, 12 pages. DOI: 10.1145/1275808.1276416.

[82] Gregory Light, Su Swarat, Eun Jung Park, Denise Drane, Emma Tevaarwerk, and Tom Mason. "Understanding undergraduate students' conceptions of a core nanoscience concept: Size and scale." In: *Proceedings of the First International Conference on Research in Engineering Education*. 2007, pp. 1–13.

[83] Norbert Lindow, Daniel Baum, and Hans-Christian Hege. "Interactive rendering of materials and biological structures on atomic and nanoscopic scale." In: *Computer Graphics Forum*. Vol. 31. 3pt4. Wiley Online Library. 2012, pp. 1325–1334. DOI: 10.1111/j.1467-8659.2012.03128.x.

[84] Norbert Lindow, Daniel Baum, Morgan Leborgne, and Hans-Christian Hege. "Interactive visualization of RNA and DNA structures." In: *IEEE Transactions on Visualization and Computer Graphics* 25.1 (2018), pp. 967–976. DOI: 10.1109/TVCG.2018.2864507.

[85] Wouter Lueks, Ivan Viola, Matthew Van Der Zwan, Henk Bekker, and Tobias Isenberg. "Spatially continuous change of abstraction in molecular visualization." In: *Abstracts of BioVis*. Los Alamitos: IEEE Computer Society, 2011. URL: https://hal.inria.fr/hal-00781520.

[86] Scott I. MacKenzie and Stan Riddersma. "Effects of output display and control-display gain on human performance in interactive systems." In: *Behaviour & Information Technology* 13.5 (1994), pp. 328–337. DOI: http://dx.doi.org/10.1080/01449299408914613.

[87] James McCrae, Igor Mordatch, Michael Glueck, and Azam Khan. "Multiscale 3D Navigation." In: *Proc. I3D*. New York: ACM, 2009, pp. 7–14. DOI: 10.1145/1507149.1507151. URL: https://doi.org/10.1145/1507149.1507151.

[88] Karen J Meaburn. "Spatial genome organization and its emerging role as a potential diagnosis tool." In: *Frontiers in genetics* 7 (2016), p. 134. DOI: 10.3389/fgene.2016.00134.

[89] Haichao Miao, Elisa De Llano, Tobias Isenberg, M Eduard Gröller, I Barišić, and Ivan Viola. "DimSUM: Dimension and scale unifying map for visual abstraction of DNA origami structures." In: *Computer Graphics Forum* 37.3 (June 2018), pp. 403–413. DOI: 10.1111/cgf.13429.

[90] Haichao Miao, Elisa De Llano, Johannes Sorger, Yasaman Ahmadi, Tadija Kekic, Tobias Isenberg, M Eduard Gröller, Ivan Barišić, and Ivan Viola. "Multiscale visualization and scale-adaptive modification of DNA nanostructures." In: *IEEE Transactions on Visualization and Computer Graphics* 24.1 (Jan. 2017), pp. 1014–1024. DOI: 10.1109/TVCG.2017.2743981.

[91] Haichao Miao, Tobias Klein, David Kouřil, Peter Mindek, Karsten Schatz, M. Eduard Gröller, Barbora Kozlíková, Tobias Isenberg, and Ivan Viola. "Multiscale Molecular Visualization." In: *Journal of Molecular Biology* 431.6 (Mar. 2019), pp. 1049–1070. DOI: 10.1016/j.jmb.2018.09.004.

[92] Friedrich Miescher. *Die Spermatozoen einiger Wirbelthiere: ein Beitrag zur Histochemie*. Birkhäuser, 1874.

[93] Robert B. Miller. "Response Time in Man-Computer Conversational Transactions." In: *Proc. AFIPS*. New York: ACM, 1968, pp. 267–277. DOI: 10.1145/1476589.1476628.

[94] Peter Mindek, David Kouřil, Johannes Sorger, Daniel Toloudis, Blair Lyons, Graham Johnson, M Eduard Gröller, and Ivan Viola. "Visualization multi-pipeline for communicating biology." In: *IEEE Transactions on Visualization and Computer Graphics* 24.1 (2017), pp. 883–892. DOI: 10.1109/TVCG.2017.2744518.

[95] Gabriel Mistelbauer, Andrej Varchola, Hamed Bouzari, Juraj Starinsky, Arnold Köchl, Rüdiger Schernthaner, Dominik Fleischmann, Meister Eduard Gröller, and Milos Srámek. "Centerline Reformations of Complex Vascular Structures." In: *Proc. PacificVis*. Los Alamitos: IEEE Computer Society, 2012, pp. 233–240. DOI: 10.1109/PacificVis.2012.6183596.

[96] Haneen Mohammed, Ali K Al-Awami, Johanna Beyer, Corrado Cali, Pierre Magistretti, Hanspeter Pfister, and Markus Hadwiger. "Abstractocyte: A visual tool for exploring nanoscale astroglial cells." In: *IEEE Transactions on Visualization and Computer Graphics* 24.1 (2017), pp. 853–861. DOI: 10.1109/TVCG.2017.2744278.

[97] Tamara Munzner. *Visualization Analysis and Design*. Boca Raton, FL, USA: CRC Press, 2014. DOI: 10.1201/b17511.

[98] Jackson Nowotny, Avery Wells, Oluwatosin Oluwadare, Lingfei Xu, Renzhi Cao, Tuan Trieu, Chenfeng He, and Jianlin Cheng. "GMOL: an interactive tool for 3D genome structure visualization." In: *Scientific Reports* 6.1 (2016), pp. 1–8. DOI: 10.1038/srep20802.

[99] S. Nusrat, T. Harbig, and N. Gehlenborg. "Tasks, Techniques, and Tools for Genomic Data Visualization." In: *Computer Graphics Forum* 38.3 (June 2019), pp. 781–805. DOI: 10.1111/cgf.13727.

[100] Donald E. Olins and Ada L. Olins. "Chromatin History: Our View from the Bridge." In: *Nature Reviews Molecular Cell Biology* 4.10 (Oct. 2003), pp. 809–814. DOI: 10.1038/nrm1225.

[101] Horng D. Ou, Sébastien Phan, Thomas J. Deerinck, Andrea Thor, Mark H. Ellisman, and Clodagh C. O'Shea. "ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells." In: *Science* 357.6349 (2017). ISSN: 0036-8075. DOI: 10.1126/science.aag0025.

[102] Joel D Parker. "Using Google Earth to teach the magnitude of deep time." In: *Journal of College Science Teaching* 40.5 (2011), pp. 23–27.

[103] Julius Parulek, Daniel Jönsson, Timo Ropinski, Stefan Bruckner, Anders Ynnerman, and Ivan Viola. "Continuous levels-of-detail and visual abstraction for seamless molecular visualization." In: *Computer Graphics Forum*. Vol. 33. 6. Wiley Online Library. 2014, pp. 276–287. DOI: 10.1111/cgf.12349.

[104] Julius Parulek, Timo Ropinski, and Ivan Viola. "Seamless visual abstraction of molecular surfaces." In: *Proceedings of the 29th Spring Conference on Computer Graphics*. 2013, pp. 107–114. DOI: 10.1145/2508244.2508258.

[105] Elizabeth Pennisi. "The Human Genome." In: *Science* 291.5507 (Feb. 2001), pp. 1177–1180. DOI: 10.1126/science.291.5507.1177.

[106] Ken Perlin and David Fox. "Pad: An Alternative Approach to the Computer Interface." In: *Proc. SIGGRAPH*. New York: ACM, 1993, pp. 57–64. DOI: 10.1145/166117.166125.

[107] Gonzalo Ramos and Ravin Balakrishnan. "Fluid Interaction Techniques for the Control and Annotation of Digital Video." In: *Proc. UIST*. New York: ACM, 2003, pp. 105–114. DOI: 10.1145/964696.964708.

[108] Gonzalo Ramos and Ravin Balakrishnan. "Zliding: Fluid Zooming and Sliding for High Precision Parameter Manipulation." In: *Proc. UIST*. New York: ACM, 2005, pp. 143–152. DOI: 10.1145/1095034.1095059.

[109] Peter Rautek, Stefan Bruckner, Eduard Gröller, and Ivan Viola. "Illustrative visualization: new technology or useless tautology?" In: *ACM SIGGRAPH Computer Graphics* 42.3 (2008), pp. 1–8. DOI: 10.1145/1408626.1408633.

[110] Peter Rautek, Stefan Bruckner, and Meister Eduard Gröller. "Semantic Layers for Illustrative Volume Rendering." In: *IEEE Transactions on Visualization and Computer Graphics* 13.6 (2007), pp. 1336–1343. DOI: 10.1109/TVCG.2007.70591.

[111] Guido Reina and Thomas Ertl. "Hardware-accelerated glyphs for mono-and dipoles in molecular dynamics visualization." In: *EuroVis*. Citeseer. 2005, pp. 177–182. DOI: 10.5555/2384060.2384089.

[112] Allan Rocha, Usman Alim, Julio D. Silva, and Mario Costa Sousa. "Decal-Maps: Real-Time Layering of Decals on Surfaces for Multivariate Visualization." In: *IEEE Transactions on Visualization and Computer Graphics* 23.1 (Jan. 2017), pp. 821–830. DOI: 10.1109/TVCG.2016.2598866.

[113] K. Schatz, C. Müller, M. Krone, J. Schneider, G. Reina, and T. Ertl. "Interactive visual exploration of a trillion particles." In: *Proc. LDAV*. Los Alamitos: IEEE Computer Society, 2016, pp. 56–64. DOI: 10.1109/LDAV.2016.7874310.

[114] Valerie A. Schneider et al. "Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly." In: *Genome Research* 27.5 (May 2017), pp. 849–864. DOI: 10.1101/gr.213611.116.

[115] Sang-Hyun Song and Tae-You Kim. "CTCF, cohesin, and chromatin in human cancer." In: *Genomics & Informatics* 15.4 (2017), p. 114. DOI: 10.5808/GI.2017.15.4.114.

[116] Pjotr Svetachov, Maarten H. Everts, and Tobias Isenberg. "DTI in Context: Illustrating Brain Fiber Tracts In Situ." In: *Computer Graphics Forum* 29.3 (June 2010), pp. 1024–1032. DOI: 10.1111/j.1467-8659.2009.01692.x.

[117] Heather J. Szerlong and Jeffrey C. Hansen. "Nucleosome distribution and linker DNA: connecting nuclear function to dynamic chromatin structure." In: *Biochemistry and Cell Biology* 89.1 (Feb. 2011), pp. 24–34. DOI: 10.1139/O10-139.

[118] Torin Taerum, Mario Costa Sousa, F Samavati, S Chan, and Joseph Ross Mitchell. "Real-time super resolution contextual close-up of clinical volumetric data." In: *Proceedings of the Eighth Joint Eurographics/IEEE VGTC conference on Visualization*. 2006, pp. 347–354. DOI: 10.5555/2384796.2384849.

[119] Maurice Termeer, Javier Oliván Bescós, Marcel Breeuwer, Anna Vilanova, Frans Gerritsen, and Meister Eduard Gröller. "CoViCAD: Comprehensive Visualization of Coronary Artery Disease." In: *IEEE Transactions on Visualization and Computer Graphics* 13.6 (2007). DOI: 10.1109/TVCG.2007.70550. URL: https://www.cg.tuwien.ac.at/research/publications/2007/termeer-2007-covicad/.

[120] Christian Tietjen, Tobias Isenberg, and Bernhard Preim. "Combining Silhouettes, Shading, and Volume Rendering for Surgery Education and Planning." In: *Proc. EuroVis*. Goslar, Germany:

Eurographics Association, 2005, pp. 303–310. DOI: 10.2312/VisSym/EuroVis05/303-310.

[121] Stephen Tonna, Assam El-Osta, Mark E. Cooper, and Chris Tikellis. "Metabolic memory and diabetic nephropathy: Potential role for epigenetic mechanisms." In: *Nature Reviews Nephrology* 6.6 (June 2010), pp. 332–341. DOI: http://dx.doi.org/10.1038/nrneph.2010.55.

[122] Thomas R Tretter, M Gail Jones, Thomas Andre, Atsuko Negishi, and James Minogue. "Conceptual boundaries and distances: Students' and experts' concepts of the scale of scientific phenomena." In: *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching* 43.3 (2006), pp. 282–319. DOI: 10.1002/tea.20123.

[123] Edward R. Tufte. *Envisioning Information*. Cheshire, CT, USA: Graphics Press, 1990. URL: https://www.edwardtufte.com/tufte/books_ei.

[124] Nynke L Van Berkum, Erez Lieberman-Aiden, Louise Williams, Maxim Imakaev, Andreas Gnirke, Leonid A Mirny, Job Dekker, and Eric S Lander. "Hi-C: A method to study the three-dimensional architecture of genomes." In: *Journal of Visualized Experiments* 39 (2010), e1869. DOI: 10.3791/1869.

[125] Matthew Van Der Zwan, Wouter Lueks, Henk Bekker, and Tobias Isenberg. "Illustrative molecular visualization with continuous abstraction." In: *Computer Graphics Forum* 30.3 (May 2011), pp. 683–690. DOI: 10.1111/j.1467-8659.2011.01917.x.

[126] Ivan Viola, Min Chen, and Tobias Isenberg. "Visual Abstraction." In: *Foundations of Data Visualization*. Ed. by Min Chen, Helwig Hauser, Penny Rheingans, and Gerik Scheuermann. Berlin/Heidelberg: Springer, 2020. Chap. 2, pp. 15–37. DOI: 10.1007/978-3-030-34444-3_2.

[127] Ivan Viola and M Eduard Gröller. "Smart visibility in visualization." In: *Proc. CAe*. 2005, pp. 209–216. DOI: 10.5555/2381219.2381249.

[128] Ivan Viola and Tobias Isenberg. "Pondering the Concept of Abstraction in (Illustrative) Visualization." In: *IEEE Transactions on Visualization and Computer Graphics* 24.9 (Sept. 2018), pp. 2573–2588. DOI: 10.1109/TVCG.2017.2747545.

[129] Thomas Waltemate, Björn Sommer, and Mario Botsch. "Membrane Mapping: Combining Mesoscopic and Molecular Cell Visualization." In: *Proc. VCBM*. Goslar, Germany: Eurographics Association, 2014, pp. 89–96. ISBN: 978-3-905674-62-0. DOI: 10.2312/vcbm.20141187.

[130] James D Watson and Francis HC Crick. "Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid." In: *Nature* 171.4356 (1953), pp. 737–738.

[131] Megan Watzke, Kimberly Arcand, and Katie Peek. *Magnitude: The Scale of the Universe*. Black Dog & Leventhal, 2017.

[132] Chris Weaver. "Building Highly-Coordinated Visualizations in Improvise." In: *Proc. InfoVis*. Los Alamitos: IEEE Computer Society, 2004, pp. 159–166. DOI: 10.1109/INFVIS.2004.12.

[133] Wikipedia. *Linker DNA*. Web site: https://en.wikipedia.org/wiki/Linker_DNA. Visited March 2020. 2020.

[134] Graham Wills. "Linked Data Views." In: *Handbook of Data Visualization*. Ed. by C.-h. Chen, Wolfgang Härdle, and Antony Unwin. Berlin/Heidelberg: Springer, 2008. Chap. II.9, pp. 217–241. DOI: 10.1007/978-3-540-33037-0_10.

[135] Ji Soo Yi, Youn ah Kang, John Stasko, and Julie A. Jacko. "Toward a Deeper Understanding of the Role of Interaction in Information Visualization." In: *IEEE Transactions on Visualization and Computer Graphics* 13.6 (2007), pp. 1224–1231. DOI: 10.1109/TVCG.2007.70515.

[136] Xiaolong Zhang. "Space-scale animation: Enhancing cross-scale understanding of multiscale structures in multiple views." In: *Proc. CMV*. Los Alamitos: IEEE Computer Society, 2005, pp. 109–120. DOI: 10.1109/CMV.2005.16.

[137] Hui Zheng and Wei Xie. "The role of 3D genome organization in development and cell differentiation." In: *Nature Reviews Molecular Cell Biology* (2019), p. 1. DOI: 10.1038/s41580-019-0132-4.

[138] Xin Zhou, Rebecca F Lowdon, Daofeng Li, Heather A Lawson, Pamela AF Madden, Joseph F Costello, and Ting Wang. "Exploring long-range genome interactions using the WashU Epigenome Browser." In: *Nature Methods* 10.5 (2013), pp. 375–376. DOI: 10.1038/nmeth.2440.

## COLOPHON

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*". `classicthesis` is available for both LaTeX and LyX:

https://bitbucket.org/amiede/classicthesis/

université PARIS-SACLAY

ÉCOLE DOCTORALE
Sciences et technologies de l'information et de la communication (STIC)

**Titre:** Abstraction spatiale intégrée de molécules génétiques

**Mots clés:** Visualisation multi-échelle, abstraction visuelle, visualisation illustrative, génome, ADN

**Résumé:** Le génome humain est principalement constitué d'ADN, une macromolécule constituée d'une longue séquence linéaire de bases, étroitement serrée pour s'insérer dans le noyau relativement petit. L'empaquetage donne lieu à de multiples niveaux hiérarchiques d'organisation. Des recherches récentes ont montré que, parallèlement à la séquence linéaire, l'agencement spatial du génome joue un rôle important dans la fonction et l'activité du génome. La visualisation des aspects linéaires et spatiaux des données du génome est donc nécessaire.

Dans cette thèse, nous nous concentrons sur le concept d'abstraction visuelle continue pour les données multi-échelles, appliqué à la visualisation du génome humain. L'abstraction visuelle est un concept inspiré par des illustrations qui simplifie le travail de traitement visuel, en guidant l'attention du spectateur vers les aspects importants.

Nous commençons par extraire les caractéristiques des données multi-échelles et faisons une comparaison parallèle entre le génome et les données astronomiques. Les différences existantes créent le besoin d'approches différentes. Un point commun cependant est la nécessité de transitions continues qui aident les spectateurs à saisir les relations et les différences de taille relative entre les échelles.

Pour satisfaire aux conditions posées par les deux aspects des données génomiques multi-échelles, nous présentons deux cadres conceptuels, basés sur les mêmes données.

Le premier cadre, ScaleTrotter, représente la structure spatiale du génome, à tous les niveaux disponibles. Il donne à l'utilisateur la liberté de voyager du noyau d'une cellule aux atomes des bases, en passant par les différents niveaux d'organisation du génome. Pour rendre possible l'exploration de la structure de tous les niveaux, des transitions temporelles fluides sont utilisées. Même si toutes les échelles ne sont pas visibles simultanément, la transition temporelle utilisée superpose deux représentations d'un même élément à des échelles consécutives, ce qui met en évidence leur relation. Pour garantir la compréhensibilité et l'interactivité des données, les parties inutiles des données sont extraites à l'aide d'une caméra dépendante de l'échelle.

Le deuxième cadre, Multiscale Unfolding, se concentre sur des aspects qui ne sont pas visibles dans ScaleTrotter : la séquence linéaire et une vue d'ensemble simultanée de tous les niveaux organisationnels. Les données sont redressées pour déplier l'empaquetage qui se produit à plusieurs niveaux de manière à conserver la connectivité entre les éléments. Pour représenter tous les niveaux disponibles, nous utilisons des transitions spatiales douces entre les niveaux. Ces transitions spatiales sont basées sur le même concept que les transitions temporelles du cadre précédent, en superposant les échelles et en mettant l'accent sur leur relation et leur différence de taille. Nous introduisons une technique d'interaction appelée Multiscale Zliding qui permet l'exploration des données et met davantage l'accent sur les différences de taille entre les niveaux.

Dans chaque cadre conceptuel, l'un des deux aspects linéaire ou spatial des données sur le génome est sacrifié pour mettre l'accent sur l'autre. La thèse se termine par une discussion sur la possibilité de combiner les deux cadres, en minimisant les sacrifices pour explorer les deux aspects du génome qui sont d'égale importance. Dans cette thèse, nous faisons un pas de plus vers la compréhension complète de l'activité du génome.

**Title:** Spatially Integrated Abstraction of Genetic Molecules

**Keywords:** Multiscale visualization, visual abstraction, illustrative visualization, genome, DNA

**Abstract:** The human genome consists mainly of DNA, a macromolecule consisting of a long linear sequence of bases, tightly packed to fit in the relatively small nucleus. The packing gives rise to multiple hierarchical organizational levels. Recent research has shown that, along with the linear sequence, the spatial arrangement of the genome plays an important role in the genome's function and activity. The visualization of both linear and spatial aspects of genome data is therefore necessary.

In this thesis, we focus on the concept of continuous visual abstraction for multiscale data, applied to the visualization of the human genome. Visual abstraction is a concept inspired by illustrations that makes the job of visual processing simpler, by guiding the attention of the viewer to important aspects.

We first extract characteristics of multiscale data and makes a parallel comparison between genome and astronomical data. The existing differences create the need for different approaches. A common point however is the need for continuous transitions that helps viewers grasp the relationships and relative size differences between scales.

To satisfy the conditions posed by the two aspects of the multiscale genome data, we present two conceptual frameworks, based on the same data.

The first framework, ScaleTrotter, represents the spatial structure of the genome, on all available levels. It gives users the freedom to travel from the nucleus of a cell to the atoms of the bases, passing through the different organizational levels of the genome. To make the exploration of the structure of all levels possible, smooth temporal transitions are used. Even though all the scales are not simultaneously visible, the temporal transition used superimposes two representations of the same element at consecutive scales emphasizing their relationship. To ensure the understandability and interactivity of the data, unnecessary parts of the data are abstracted away with the use of a scale-dependent camera.

The second framework, Multiscale Unfolding, focuses on aspects that are not visible in ScaleTrotter: the linear sequence and a simultaneous overview of all the organizational levels. The data is straightened to unfold the packing that occurs on several levels in a way that conserves the connectivity between the elements. To represent all the available levels, we use smooth spatial transitions between the levels. These spatial transitions are based on the same concept of the temporal transitions of the previous framework, superimposing scales and emphasizing on their relationship and size difference. We introduce an interaction technique called Multiscale Zliding that allows the exploration of the data and further emphasizes the size differences between the levels.

In each framework, one of either linear of spatial aspect of genome data is sacrificed to emphasize the other. The thesis concludes with a discussion about the possibility of combining the two frameworks, minimizing the sacrifices to explore the two equally important aspects of the genome. In this thesis, we take a step closer to fully understanding the activity of the genome.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France