# ChromoVis: Feature-Rich Layouts of Chromosome Conformation Graphs

Darya Filippova[1,2], Geet Duggal[1,2], Rob Patro[2], and Carl Kingsford[2]

1 – Joint Carnegie Mellon University — University of Pittsburgh
Ph.D. Program in Computational Biology
2 – Lane Center for Computational Biology, Carnegie Mellon University
5000 Forbes Ave, Pittsburgh, PA, USA, 15213
{dfilippo,geet,robp,carlk}@cs.cmu.edu

**Abstract.** Recent chromosome conformation capture (3C) experiments produce pairwise interactions between fragments of DNA that are spatially close in the nucleus of a cell. Visualizing 3C data on the scale of the whole genome allows scientists to gain insight into chromosomal packing in the nucleus of a cell and to generate hypotheses about what kinds of genomic features (e.g. gene locations, DNA accessibility, GC content) correlate with spatial proximity. We introduce the *chromosome layout problem* which seeks a two-dimensional layout of a chromosome conformation graph such that: (1) an entire chromosome can be visualized as a strand of DNA together with its genomic features, and (2) pairwise spatial constraints obtained from biological experiments are respected as much as possible. Our approach treats the chromosome as a self-avoiding string or polymer while attempting to satisfy the constraints. The layouts we produce accurately depict spatial proximities observed in 3C experiments while avoiding undesirable occlusions and minimizing edge crossings.

**Keywords:** chromosome conformation capture, network layout, genetic algorithms

## 1 Introduction

The DNA of eukaryotic organisms is efficiently packed in the cell's nucleus allowing long strands of DNA to fit in a very small volume. The spatial organization of the packed chromatin has been shown to play an important role in modulating gene expression and facilitating long-range gene regulation [3]. This organization has also been implicated in rearrangement mutations that are associated with some cancers [6].

Recently, the experimental techniques of chromosome conformation capture (3C; [10]) and its subsequent refinements have been applied to determine the three-dimensional shape of the genome in several organisms. In a 3C experiment, proteins on chromosomes that are close to each other in their natural folded state are glued together by formaldehyde treatment. Pieces of DNA that are not

linked are cleaved and washed away while DNA pieces that were bonded are retained. The remaining pairs are sequenced and their location in the original genome recorded. This procedure is repeated in parallel over millions of cells thus providing a quantitative measure of how often the two parts of DNA were close to each other.

The result of a 3C experiment is a *chromosome conformation capture*, or *3C, graph* $G = (V, E)$ where nodes $V$ represent fragments of the chromosomes (similar in size), and edges in $E$ connect two nodes $u_i, u_j \in V$, $i \neq j$, if the pair of chromosome fragments represented by $u_i$, $u_j$ were spatially close. Each 3C edge has a weight $w(e_{ij})$ that equals the number of times the interaction between $u_i$ and $u_j$ was observed in the experiment.

Aided by the new experimental protocols, the study of three-dimensional genome structure will continue to assist in identifying additional examples of long-range gene regulation, drive the development of better models of chromosome packing and its folding principles, and lead to a better understanding of the relationships between spatial location, DNA accessibility, and gene expression. However, the recent increase in the volume of genomic structural information has created a need for intuitive, information-rich visualizations of both spatial proximity encoded in 3C graphs and genomic annotations (e.g. genes, regulatory signals, sequence composition). Displaying both the geometric structure and the genomic annotations will help users develop an intuition about the relationship between these two classes of features and will help them pose hypotheses about how spatial position affects the functioning of the genome.

Correlating genome structure with other biological data is a difficult task. The nodes in the graph $G$ have tens and hundreds of annotations: the genes located on that piece of DNA, how often and when those genes are active, what other genes get turned on if the gene becomes active, presence of specific enzymes or binding activity, and so on. These categorical, numerical, and relational data are associated with a genomic coordinate along a chromosome and plotting them against the chromosomal address helps the user develop a mental model when studying these data (see Fig. 1 for an example). In this paper, we discuss visual features desirable for chromosome conformation graph visualization and develop a novel approach to visualizing spatial data along with the linear annotations on the chromosomes. [XXX - expand more on how we solve the problem]

## 2 Related Work

Traditionally, genomic data is visualized in a linear fashion with chromosomal coordinates along the $X$ axis, as in the widely used UCSC Genome Tracks [7] and Galaxy Browser [9] (see Fig.1b). However, the linear coordinates make it hard to incorporate any type of relational information where two separate pieces of a genome are joined by an edge. An edge may represent the fact that a gene BRCA1 at position 10000 regulates gene BRCA2 at position 20000, or that a sequence of nucleotides at [30000, 45000] loops around and is spatially close to a sequence [100000, 110000] affecting its function, or that a protein produced

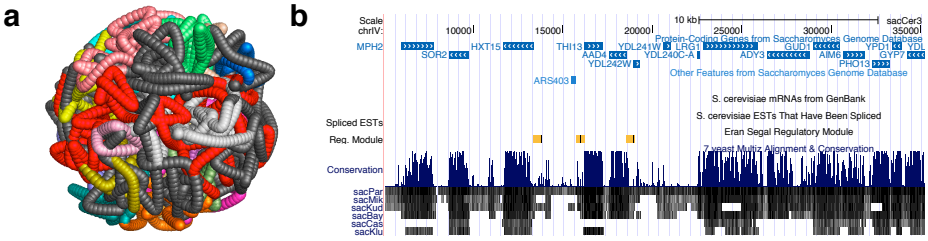


**Fig. 1. A.** A three-dimensional model of a yeast genome based on 3C data [20]. **B.** Linear chromosome visualization in Genome Broswer with multiple annotations (genes and regulatory elements) [7].

by gene BRCA1 catalyzes a protein produced by gene BRCA2 to activate a chemical process.

Another visualization approach to viewing genomic graph data employs a circular layout (Fig. 2a) where the chromosome is laid out along the circumference and 3C edges connect parts of the chromosome that are spatially close. Circos [11] is a widely used tool for such analyses and provides features that has made this layout commonly used for plotting 3C data. The primary drawback of a circular layout is that it provides little infromation about higher-order structural relationships beyond pairwise proximity (e.g. cliques, paths, long-range chromosome looping). Recent work focusing on visualizing matrices of 3C interactions [17, 23] suffer from a similar drawback. In addition, along with other traditional graph layouts [18, 5, 15], graphs with large number of edges become cluttered easily.

A seemingly natural approach for visualizing chromosomal structure is to embed models of the chromasomes into three dimensions. In this approach, chromosomes are represented by a simplified model (such as beads on a string) and the positions of the segments are optimized to avoid overly stretching or compressing the chromatin while attempting to respect distance constraints derived from measured pairwise distances [4, 2, 16]. The various methods for performing these embeddings have been quite successful in producing data-driven models of chromatin structure and have resulted in a number of interactively viewable three-dimensional structures. However, these 3D viewers [1], while useful, have a number of deficiencies as aids to visualization and exploration. The most significant fault is that big portions of the data are occluded from most viewpoints, and the strands of chromosomes in the center of the model (see Fig. 1a) are completely obscured by the strands on the surface. This makes it impossible to quickly grasp correlations between a genomic feature and its spatial location. Interactive exploration of 3D models requires complex, non-intuitive user interaction to support rotation, zoom, translation, and slicing with standard input hardware (i.e., keyboard and mouse).

## 3 Chromosome layout problem

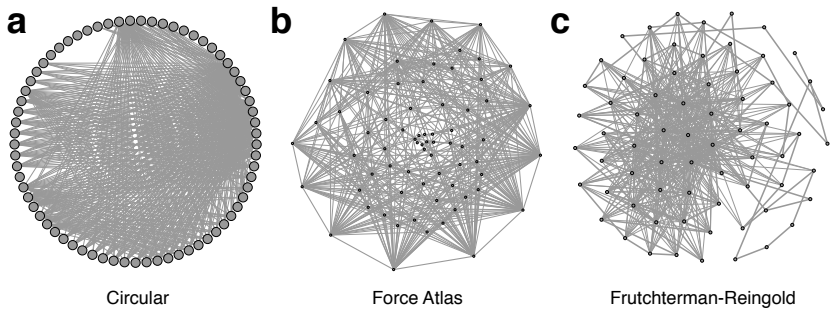


**Fig. 2.** Three layouts for human chromosome 20 from algorithms available in the Gephi .

The most appealing feature of linear and circular representations of the genomic data is that users can see the whole chromosome and zoom in on any part of it at an arbitrary level of detail. The ability to trace the chromosome and gain an overview, then to "zoom and filter" on demand adheres to the widely accepted visualization mantra by Shneiderman [19]. However, tracing the chromosome from the beginning to end is much more difficult when chromosomal edges are not explicitly included in the visualization [17, 18]. To address this issue, we define an *augmented* 3C graph as a 3C graph with additional *chromosomal* edges linking consecutive genomic segments (Fig. 3):

**Definition 1.** *We call $G = (V, E_c \cup E_s)$ an augmented 3C graph where $E_c$ are chromosomal edges connecting nodes $u_i$ and $u_{i+1}$, $1 \leq i \leq |V|$, and $E_s$ are 3C edges that encode spatial proximity of genomically distant fragments $u_i$, $u_j$, $i \neq j$ of a chromosome.*

Edges $E_c$ represent the chromosome by linking nodes in $V$ according to their genomic order. If one were to start at $u_1 \in V$ and follow $E_c$ edges to $u_1$, from $u_1$ to $u_2$ and so on, they would trace the chromosome and end up at the last vertex $u_N$ ($N = |V|$) that corresponds to the last genomic fragment of that chromosome $C$. Edges in $E_s$ represent weighted 3C edges. Given such a graph, we desire a layout that makes it is easy to trace the chromosome from start to end, affords plotting the annotations along the chromosome, and displays spatial relationships in a clear way. We formalize our criteria for the layout below:

*Problem 1.* **Chromosome layout problem.** Given an augmented 3C graph $G = (V, E_c \cup E_s)$ representing a single chromosome $C$, and a set of $T$ annotations on this chromosome, find the layout $L = \{(x_i, y_i) \mid u_i \in V, 1 \leq i \leq N\}$ for $C$ satisfying the following conditions and assuming that $C$ must be displayed with a thickness large enough to account for all annotations in $T$ — we denote this thickness as $\omega_A$:

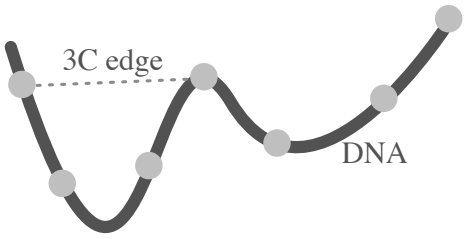


**Fig. 3.** Augmented 3C graph consists of nodes that represent equally spaced cuts of the chromosome (light gray), $E_c$ edges that connect these cuts and represent the chromosome (dark gray curved edges), and $E_s$ edges that indicate spatial proximity (dashed edges).

- **Self-avoiding:** Let $r(e, \omega_A)$ be the rectangle representing edge $e$ in the Euclidean plane of width $\omega_A$. We require that $r(e_1, \omega_A) \cap r(e_2, \omega_A) = \emptyset$ for all $e_1, e_2 \in E_c$ that are not adjacent;
- **Fixed-segment-lengths:** All edges in $E_c$ have uniform length;
- **3C edge distortion:** For every 3C edge $e_{ij} \in E_s$ connecting nodes $u_i$ and $u_j$, the distance between nodes $u_i$, $u_j$ is minimized and is inversly proportional to $w(e_{ij})$
- **3C-avoiding:** Crossings between edges in $E_s$ and those in $E_c$ are minimized;
- **Bounding box:** The layout is completely contained in a circle of radius $R$.

The self-avoiding constraint ensures that a chromosome does not fold onto itself, occluding the data, and that chromosomal edges $E_C$ are given enough space to afford the drawing of multiple annotation tracks of total width $\omega_A$. We require that all chromosomal edges are of the same length to ensure that annotation features plotted along every chromosomal edge are comparable. By requiring that 3C edges are shorter if the frequency of interaction between its endpoints is higher, we are attempting to force certain regions of the chromosome closer together in the layout. Augmented 3C graphs for every human chromosome (except chromosome 4) contained significantly more 3C edges than chromosomal edges ($\mu = 69.7\%$ of all edges with $\sigma = 14.7\%$) resulting in multiple crossings between 3C edges and the chromosomal backbone justifying an additional constraint to minimize such occurrences. Finally, to achieve more compact drawings, we enforce a bounding circle constraint that rewards points for being close to the centroid and penalizes points that are at a distance greater than $R$ from the centroid. To determine the radius of a bounding circle, we use a lenient bound $R = (N-1)D/2\pi$ that in the extreme allows the chromosome to be stretched along the circumference of a circle of radius $R$.

We reformulate Problem 1 as an optimization problem:

$$\text{maximize } f(E_s \times E_c, 0) + g(E_s) \quad , \tag{1}$$

$$\text{subject to } \|u_i - u_{i-1}\| = D \quad \text{where } u_i \in V, 0 < i \leq N, \tag{2}$$

$$\max\| \left( \frac{1}{|V|} \sum_{u_i \in V} u_i \right) - u_j \| \leq R \quad \text{for all } u_j \in V, \tag{3}$$

$$f(\mathcal{E}, \omega_A) = 0, \tag{4}$$

where $\mathcal{E} = \{(e_i, e_j) \mid e_i, e_j \in E_c \text{ and } i < j - 1\}$ and $f(E, \omega)$ is a function that penalizes intersections between pairs of rectangles $r(e_1, \omega), r(e_2, \omega)$ for $(e_1, e_2) \in E$. We take $f(E, \omega)$ to be:

$$f(E, \omega) = \frac{1}{|E|} \sum_{(e_1, e_2) \in E} \delta(r(e_1, \omega), r(e_2, \omega)), \tag{5}$$

where

$$\delta(r(e_1, \omega), r(e_2, \omega)) = \begin{cases} 1 & \text{if } r(e_1, \omega) \cap r(e_2, \omega) = \emptyset \\ -1 & \text{otherwise} \end{cases}. \tag{6}$$

The function $f(E, \alpha)$ achieves its maximum value of 1 when no bounding rectangles for pairs of edges in $E$ intersect. When $\omega = 0$, we set $e(e, 0)$ to be a line segment.

The function $g(E_s)$ measures the total distortion of the set of 3C edges induced by the current layout, and is given by:

$$g(E_s) = 1 + \sum_{e_{ij} \in E_s} \frac{\epsilon(e_{ij})}{\hat{\epsilon}(e_{ij})}. \tag{7}$$

where

$$\epsilon(e_{ij}) = \min(0, t(e_{ij}) - \|u_i - u_j\|) \tag{8}$$

and

$$\hat{\epsilon}(e_{ij}) = \min(0, t(e_{ij}) - D\,|i - j|) \tag{9}$$

and we consider $\epsilon(e_{ij})/\hat{\epsilon}(e_{ij}) = 0$ when $\epsilon(e_{ij}) = \hat{\epsilon}(e_{ij}) = 0$. Above, $t(e_{ij})$ denotes the maximum desirable distance for a 3C edge $e_{ij}$ (i.e. the ideal maximum separation between the endpoints of this edge). The maximum distance achievable in the layout between two genomic fragments $f_i$ and $f_j$ occurs when the chromosome is laid out as a straight line segment between these fragments, and so this maximum attainable distance is simply $D\,|i - j|$. Thus, equation (7) is a measure of the total normalized distortion over all 3C edges given the current layout. Here, we choose only to penalize fragments that are placed farther apart

than $t(e_{ij})$; the fragments may be closer than this target distance without detracting from the objective. Thus, $g(E_s)$ achieves its maximum value, 1, when the endpoints of every 3C edge $e_{ij}$ are separated in the layout by a distance less than or equal to $t(e_{ij})$. We set the maximum desirable distance for an edge to $t(e_{ij}) = x/w(e_{ij})$, where $x$, a user-provided parameter, is the maximum desired distance between the endpoints of a 3C fragment with a weight of 1 (we set $x = 1500$). This target distance function then varies directly as the inverse of the edge weight.

Non-linear constraints (1) and (4) place this problem outside of the class of linear and quadratic problems; additionally, considering that crossing minimization is an NP-hard problem [8], we suspect that Eqns. (1)–(4) are hard to optimize.

## 4 ChromoVis

### 4.1 Solving the optimization problem

We used a genetic computation approach to find layouts satisfying objectives in a relaxed version of Problem 1, where all objectives are normalized and summed together into a single objective. We used the ECJ [13] package to run the optimization and implemented our objectives in Java. We use tournament selection to select the 15 best individuals from each generation to carry over to the next generation. The remaining population for each new generation is generated by breeding the previously selected individuals using a multi-point crossover procedure with the probability of crossover set to 0.25. In each generation, the population contained 30 individuals, and the optimization was run for 500 generations on each chromosome.

In ECJ, every individual is associated with a genome that can mutate with some probability. For Problem 1, we encode a genome as a series of rotation angles $\mathbf{A} = (\theta_1, \ldots, \theta_{N-1})$ where each angle $\theta_i$ determines the rotation for a chromosomal edge $e_{i,i+1}$ relative to a chromosomal edge preceeding it (see Fig. 4). Given $\mathbf{A}$, we reconstructed the Cartesian coordinates for every node $u_i$ in the chain by computing coordinates for all the preceeding nodes, with node $u_1$ always positioned at the origin. By construction, all chromosomal edges $e_{i,i+1}$, $1 \leq i \leq N$, have fixed length, thus satisfying constraint (2). To help satisfy the self-avoiding objective, we also restrict the angles $\theta_i$ in the rotating chain to be $-\pi/3 \leq \theta_i \leq \pi/3$. This results in a smoother drawing that avoids self-intersections and sharp transitions. This approach is similar in spirit to freely rotating chains and self-avoiding random walks [21]. However, our version of the problem additionally attempts to satisfy as many spatial constraints as possible while minimizing crossings between chromosomal and 3C edges with width $\omega$.

When two individuals from the same generation are selected for crossover, their genomes are split at a random location $i$ and the two individuals exchange subsequences of angles from 1 to $i$ retaining $\alpha_{i+1}, \ldots, \alpha_{N-1}$ values. Since the angles are relative, crossover changes the first $i$ angles in the genome and indirectly affects all nodes past $i$ since the Cartesian coordinates for nodes $u_{i+1}, \ldots$

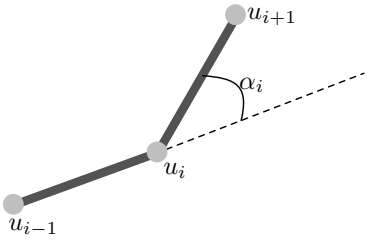


**Fig. 4.** The genome of an individual contains a series of angles where each angle $\alpha_i$ determines a rotation angle of the chromosomal edge $e_{i,i+1}$ relative to the chromosomal edge $e_{i-1,i}$ preceeding it.

depend on the series of angles up to $\alpha_i$. However, this change keeps the structure of the subchain $[u_{i+1}, u_N]$ intact (except for rotating and translating to the new coordinates of the node $u_i$). Our implementation allows multi-point crossover that splits the genome at several locations allowing for greater population diversity. Each angle is mutated independently with probability 0.15 and individual mutations have a similar effect to that of a pivot move in the freely rotating chain simulations [14].

### 4.2 Drawing chromosomes

To make the chromosome the most pronounced feature in the drawing, we weave a splined thick curve through the nodes in $G$. Interpolation also makes the drawing visually appealing and more organic — appropriately so in this biological context. The backbone of the chromosome is easy to point out and makes it easy to trace the chromosome from one end to another (Fig. 5).

## 5 Results and Discussion

### 5.1 3C data

We obtained HiC data (a genome-wide 3C experiment) from Lieberman-Aiden et al. [12] and normalized for experimental biases by the method of Yaffe and Tanay [22]. Since HiC frequencies are inversely related to genomic distance, we further normalized the frequency of an interaction $e_{ij}$ derived from the Yaffe and Tanay method by the mean frequency at the genomic distance for a given $e_{ij}$. The resulting weight $w(e_{ij})$ represents the frequency of observing an interaction $e_{ij}$ as compared to the average interaction frequency for all 3C edges at that genomic distance. We discarded edges with $w(e_{ij}) < 1.5$ for all chromosomes.

### 5.2 Annotations on chromosomes

As proof of concept, we display three annotations along the chromosome: gene locations, DNA accessibility clusters (DNaseI), and GC content. These annotations are all available from Genome Browser [7]. Each annotation is placed

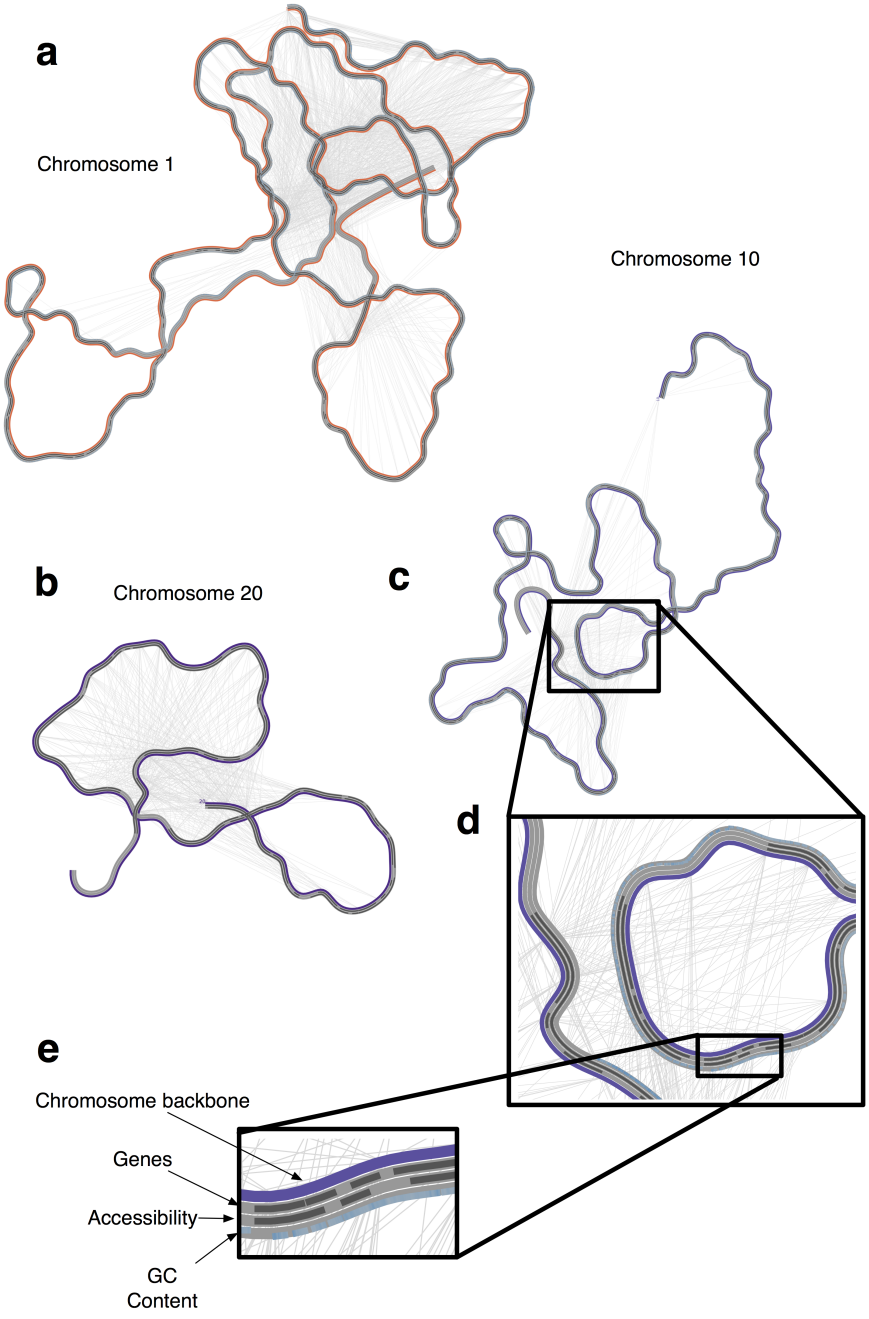


**Fig. 5.** ChromoVis layouts for human (a) Chromosome 1 (b) Chromosome 20 (c) Chromosome 10. 3C edges are shown in gray.

along the chromosome on their separate tracks (Fig. 5e). More annotations can be plotted, however, this may change the layout by trying to accomodate for a thicker chromosomal strand. Continuous annotations, sucj as GC content, are shown as heatmaps, however, their tracks can be made wider to afford the plotting of line or bar charts.

### 5.3 Performance of objectives

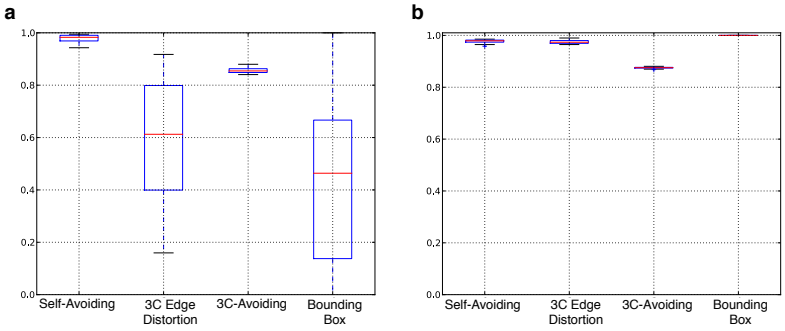


**Fig. 6.** Box plots for the distributions of four of the five objectives for (a) the initial population of layouts (b) and the final optimized population. (The the fixed-length segment objective is satisfied via our representation of a chromosome as a vector of angles on fixed-length segments.)

Our approach combining the rotating chain with the optimization yields near-optimal objective values for all objectives (Fig. 6). The optimization primarily improves the 3C edge distortion while fitting the chromosome in the bounding box. The self-avoiding objective and 3C-avoiding objectives are largely acheived by the rotating chain choosing angles in the range $[-\pi/3, -pi/3]$ since their values are similar in the initial and final populations of the genetic program. The distribution of objectives for the remaining three objectives indicates that the population is heterogeneous, and that the optimization produced several distinct layouts that satisfy our criteria of quality.

### 5.4 Visual identification of spatially close regions

Chromosomal regions that are near each other as determiend by the 3C experiment are also near each other in ChromoVis layouts. For example, all three chromosomes in Figure 5 coil and turn in a way that allows placing densely interacting regions next to one another. At the same time, areas of the chromosomes

not constrained by the 3C edges, are further away from the rest of the chromosome making it easy to differentiate between the packed and unconstrained regions. These examples illustrate ChromoVis's ability to capture higher-order structures while maintaining the readability of the drawing.

## 6 Conclusions

We presented ChromoVis: a new two-dimensional layout for chromosome conformation capture data that simultaneously clearly displays the chromosome and the spatial proximity of chromosomal segments as well as rich genome annotations. Our genetic programming approach, which relaxes the desired optimization objective, achieves good individual objective values while allowing for efficient computation of the layout (usually within [XXX - running] minutes for the largest chromosome on a MacBookPro machine with 2.3 GHz Intel i5 processor with 4Gb of RAM). ChromoVis produces images suitable for publications, large displays or educational posters, and for interactive exploration. While ChromoVis can be directly applied to 3C graphs, the approach is generally applicable to any graph with weighted constraints on a string-like structure: other pairwise associations between parts of the chromosomes, a protein chain with physical contacts, or even a series of correlated point events on a timeline.

## 7 Acknowledgements

This work has been partially funded by National Science Foundation (CCF-1256087, CCF-1053918, and EF-0849899) and National Institutes of Health (1R21AI085376). C.K. received support as an Alfred P. Sloan Research Fellow. D.F. is a predoctoral trainee supported by NIH T32 training grant T32 EB009403 as part of the HHMI-NIBIB Interfaces Initiative.

## References

1. T. M. Asbury, M. Mitman, J. Tang, and W. J. Zheng. Genome3D: a viewer-model framework for integrating and visualizing multi-scale epigenomic information within a three-dimensional genome. *BMC Bioinformatics*, 11:444, January 2010.
2. D. Baù and et al. The three-dimensional folding of the $\alpha$-globin gene domain reveals formation of chromatin globules. *Nature Structural & Molecular Biology*, 18(1):107–114, 2010.
3. G. Cavalli and T. Misteli. Functional implications of genome topology. *Nature Structural & Molecular Biology*, 20(3):290–299, March 2013.
4. Z. Duan and et al. A three-dimensional model of the yeast genome. *Nature*, 465(7296):363–367, 2010.
5. T. M. J. Fruchterman and E. M. Reingold. Graph drawing by force-directed placement. *Software — Practice and Experience*, 21:1129–1164, 1991.

6. G. Fudenberg, G. Getz, M. Meyerson, and L. A. Mirny. High order chromatin architecture shapes the landscape of chromosomal alterations in cancer. *Nature Biotech*, 29:1109–1113, 2011.
7. P. A. Fujita and et al. The UCSC Genome Browser database: update 2011. *Nucleic Acids Research*, 39(Database issue):D876–82, January 2011.
8. M. R. Garey and D. S. Johnson. Crossing Number is NP-Complete. *Journal on Algebraic and Discrete Methods*, 4:312–316, 1983.
9. B. Giardine and et al. Galaxy: a platform for interactive large-scale genome analysis. *Genome Research*, 15(10):1451–5, October 2005.
10. H. Hagège and et al. Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nature Protocols*, 2:1722–1733, 2007.
11. M. Krzywinski and et al. Circos: an information aesthetic for comparative genomics. *Genome Research*, 19(9):1639–45, September 2009.
12. E. Lieberman-Aiden and et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326:289–293, 2009.
13. S. Luke. ECJ: A Java-based evolutionary computation research system, 2013.
14. Neal Madras and Alan D. Sokal. The Pivot Algorithm: a highly efficient Monte Carlo method for the self-avoiding walk. *Journal of Statistical Physics*, 50:109–186, 1988.
15. A. Noack. *Unified Quality Measures for Clusterings, Layouts, and Orderings of Graphs, and Their Application as Software Design Criteria*. PhD thesis, Brandenburg Technical University, 2007.
16. M. Rousseau, J. Fraser, M. Ferraiuolo, J. Dostie, and M. Blanchette. Three-dimensional modeling of chromatin structure from interaction frequency data using Markov chain Monte Carlo sampling. *BMC Bioinformatics*, 12(1):414, 2011.
17. N. Servant and other. HiTC : Exploration of High-Throughput 'C' experiments. *Bioinformatics*, 28(21):2843–2844, August 2012.
18. Y. Shavit and P Lio'. CytoHiC: A Cytoscape Plugin For Visual Comparison Of Hi-C Networks. *Bioinformatics (Oxford, England)*, pages btt120–, March 2013.
19. B. Shneiderman. The eyes have it: a task by data type taxonomy for information visualizations. In *Proceedings 1996 IEEE Symposium on Visual Languages*, pages 336–343, Boulder, CO, 1996.
20. H. Tanizawa and et al. Mapping of long-range associations throughout the fission yeast genome reveals global genome organization linked to transcriptional regulation. *Nucleic Acids Research*, 38:8164–8177, 2010.
21. F. T. Wall, F. Mandell, and J. C. Chin. Self-avoiding random walks subject to external spatial constraints. *Journal of Chemical Physics*, 65:2231–2234, 1976.
22. E. Yaffe and A. Tanay. Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. *Nature Genetics*, 43:1059–1065, 2011.
23. X. Zhou and et al. Exploring long-range genome interactions using the WashU Epigenome Browser. *Nature Methods*, 10(5):375–6, May 2013.