HHS Public Access

Author manuscript

Nat Methods. Author manuscript; available in PMC 2024 December 01.

Published in final edited form as:

Nat Methods. 2023 December ; 20(12): 1834–1835. doi:10.1038/s41592-023-02056-x.

Chromoscope: interactive multiscale visualization for structural variation in human genomes

Sehi L'Yi[1], **Dominika Maziec**[1], **Victoria Stevens**[1], **Trevor Manz**[1], **Alexander Veit**[1], **Michele Berselli**[1], **Peter J. Park**[1,*], **Dominik Glodzik**[1,*], **Nils Gehlenborg**[1,*]

[1]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA

With the wider adoption of whole-genome sequencing (WGS) for genome analysis, researchers can profile structural variants (SVs)[1] in addition to smaller insertions, deletions and point mutations. Somatic SVs—such as copy number variants, translocations, transposable element insertions, and complex rearrangements—have been shown to play a critical role in cancer development and therapy selection[2,3]. Germline SVs can also perturb disease-associated genes[4], with WGS leading to increased diagnostic yields in rare disease cohorts[5]. However, the size and complexity of the data, combined with the difficulty of obtaining accurate SV calls, pose challenges in the interpretation of SVs, often requiring laborious visual inspection of potentially pathogenic variants using multiple visualization tools. To facilitate the identification of functionally relevant SVs and the characterization of genome-wide SV patterns, efficient browsing of candidate variants and visual assessment at multiple resolutions—from the whole genome to base-pair resolution at the sequencing read level—are vital.

We developed Chromoscope (https://chromoscope.bio), an interactive visualization tool that supports **multiscale** and **multiform** visualizations. Chromoscope enables the user to analyze SVs at multiple scales, using **four main views** (multiscale) (Fig. 1a–d). Moreover, each view uses different visual representations (multiform) that can facilitate the interpretation for a given level of scale. The **multi-sample overview** (Fig. 1a) at the cohort level displays circular summaries of SVs and their types as arches connecting chromosomal breakpoints, using different colors to indicate duplications, deletions, inversions, and inter-chromosomal translocations, as well as copy number variants. With this cohort overview, the user can quickly inspect the patterns across large cohorts and find samples of interest. The **genome view** (Fig. 1b) shows the selected sample as a circular visualization with more details, such as chromosome ideograms, highlighting gain and loss of function mutations in putative cancer driver genes or any other mutations chosen by the user. The **variant view** (Fig.

*Corresponding authors.

Author Contributions

S.L. developed Chromoscope in close collaboration with D.G.; D.G., P.P., and N.G. conceived and supervised the project. D.M., V.S. and D.G. assisted with case studies. T.M. developed the Python package. D.M., M.B. and A.V. suggested critical improvements for the tool. The draft manuscript was written by S.L. and D.G. and all authors edited and gave feedback on the manuscript.

Competing Interests

N.G. is a co-founder and equity owner of Datavisyn. D.G. is a consultant for Repare Therapeutics.

Code Availability

The entire source code is publicly available at https://github.com/hms-dbmi/chromoscope under the MIT license (DOI: 10.5281/zenodo.7665959). The Chromoscope web application is available at https://chromoscope.bio.

1c) displays shorter genomic regions, showing multiple tracks for each variant type in the region and gene annotations. Individual SVs are visualized using arches of varying heights and orientations (upward and downward) and intuitively display chains of SVs (e.g., chromothripsis and chromoplexy) (Fig. 1g). As visual inspection of a variant at the read level is still considered the best approach for *in silico* validation for SNVs and indels, Chromoscope expands this step for SV breakpoints. The **breakpoint view** (Fig. 1d) shows aligned reads around breakpoints and highlights 'clipped' (or 'split') reads spanning breakpoints, making it easy to assess the evidence of each SV. For the seamless SV analysis across scales, all views in Chromoscope are interactive and linked. The user can smoothly zoom and pan, use a genomic region selection tool, search for a gene of interest, and select an SV to instantly display read-level views for an in-depth examination. In contrast to existing tools that do not facilitate navigation and comparisons of data across scales or across samples (e.g., IGV[6] and JBrowse[7] do not support the circular multi-sample overview), the multiscale/multiform nature of Chromoscope facilitates a comprehensive SV analysis within a single tool. A detailed comparison with other tools is provided in Supplementary Table 1.

Chromoscope was developed using the Gosling library,[8] a grammar-based framework that offers building blocks for scalable, interactive, and customizable genomics visualizations. The use of Gosling not only enhances the flexibility of Chromoscope but also helps the user to incorporate Chromoscope into different workflows through the Gosling specification (see documentation, https://chromoscope.bio/docs). For example, Chromoscope can export Gosling specifications (JSON)[8] for usage within the Gosling online editor for further customization. A Chromoscope Python package is also available to visualize SV datasets in computational notebooks (e.g., Jupyter Notebooks).

To make SV analysis easily accessible, Chromoscope is designed as a browser-based client-side application that does not require installation and can be readily configured to load the user's datasets. Chromoscope uses standard file formats, such as VCF (point mutations), BAM (sequence alignment data), and BEDPE (SV calls). The JSON-formatted configuration file (Fig. 1e) specifies URLs of remote or local files. These files can be stored publicly (e.g., in public cloud storage) or privately (e.g., in protected cloud storage or by setting up a local server). For easy sharing of observations, Chromoscope generates a URL address that encodes the loaded data sources, the current visualization state, and viewed genomic locations. We provide default data configurations with a collection of 2,778 cancer genomes across 39 types from the PCAWG consortium[9] (Supplementary Table 2), to showcase the genome interpretation capabilities at scale. In our documentation, we illustrate the capabilities of Chromoscope in seven case studies that are not well supported by existing tools.

Through its unique visualization capabilities for WGS analysis, we anticipate that Chromoscope will accelerate the exploration and interpretation of SVs by a broad range of scientists and clinicians, leading to new insights into genomic biomarkers.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the National Institutes of Health (R01HG011773, U01CA200059).

Data Availability

The demonstration cancer genome data was obtained from the publicly available ICGC/TCGA PCAWG Consortium[9] repository available at https://dcc.icgc.org/releases/. Demonstration data is limited to somatic mutation calls which are publicly available, as opposed to germline mutation calls or raw sequencing reads. Additional data from cell lines was obtained from the SRA archive (project ID:SRP162370), which is public, allowing us to showcase Chromoscope's display of raw sequencing reads.

References

1. Cortés-Ciriano I, Gulhan DC, Lee JJ-K, Melloni GEM & Park PJ Computational analysis of cancer genome sequencing data. Nat. Rev. Genet 1–17 (2021). [PubMed: 33235360]
2. Samsom KG et al. Feasibility of whole-genome sequencing-based tumor diagnostics in routine pathology practice. J. Pathol 258, 179–188 (2022). [PubMed: 35792649]
3. Villani A et al. The clinical utility of integrative genomics in childhood cancer extends beyond targetable mutations. Nat Cancer (2022) doi:10.1038/s43018-022-00474-y.
4. Collins RL et al. A structural variation reference for medical and population genetics. Nature 581, 444–451 (2020). [PubMed: 32461652]
5. 100,000 Genomes Project Pilot Investigators et al. 100,000 Genomes Pilot on Rare-Disease Diagnosis in Health Care - Preliminary Report. N. Engl. J. Med 385, 1868–1880 (2021). [PubMed: 34758253]
6. Robinson JT, Thorvaldsdóttir H, Turner D & Mesirov JP igv.js: an embeddable JavaScript implementation of the Integrative Genomics Viewer (IGV). bioRxiv 2020.05.03.075499 (2020) doi:10.1101/2020.05.03.075499.
7. Diesh C et al. JBrowse 2: A modular genome browser with views of synteny and structural variation. bioRxiv 2022.07.28.501447 (2022) doi:10.1101/2022.07.28.501447.
8. L'Yi S, Wang Q, Lekschas F & Gehlenborg N Gosling: A Grammar-based Toolkit for Scalable and Interactive Genomics Data Visualization. IEEE Trans. Vis. Comput. Graph PP, (2021).
9. ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. Nature 578, 82–93 (2020). [PubMed: 32025007]

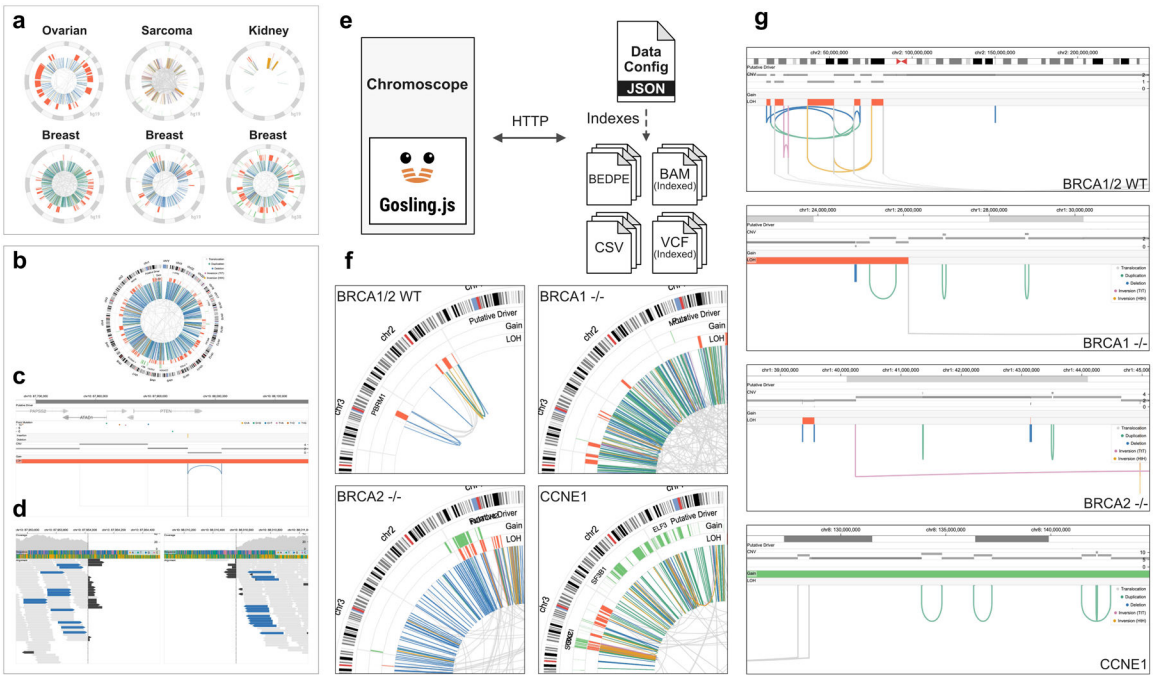

Figure 1.

The Chromoscope interface consists of four views for analyzing structural variation in cancer genomes at multiple scales (a–d). Chromoscope uses a "data config" to load datasets via HTTP requests and does not require setting up a server (e). The data can be either stored privately or publicly (e.g., on local servers or cloud buckets). Chromoscope captures distinct patterns of structural variations and their copy number footprint in samples with different types of chromosomal instability, such as chromothripsis, or associated with loss of BRCA1 −/−, BRCA2 −/−, or CCNE1 amplification (f–g).