

Genome analysis

Figeno: multi-region genomic figures with long-read support

Etienne Sollier ^{1,2,*}, Jessica Heilmann¹, Clarissa Gerhauser¹, Michael Scherer ¹,
Christoph Plass¹, Pavlo Lutsik^{1,3}

¹Division of Cancer Epigenomics, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany

²Faculty of Biosciences, Ruprecht-Karls-University of Heidelberg, 69120 Heidelberg, Germany

³Department of Oncology, KU Leuven, 3000 Leuven, Belgium

*Corresponding author. Division of Cancer Epigenomics, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany.
E-mail: etienne.sollier@dkfz.de (E.S.)

Associate Editor: Peter Robinson

Abstract

Summary: The vast amount of publicly available genomic data requires analysis and visualization tools. Here, we present figeno, an application for generating publication-quality Figures for GENOMics. Figeno particularly focuses on multi-region views across genomic breakpoints and on long reads with base modifications. In addition, we support epigenomic data including ATAC-seq, ChIP-seq or HiC, as well as whole genome sequencing data with copy numbers and structural variants.

Availability and implementation: Figeno is available as a python package with both a command line and graphical user interface. It can be installed via PyPI and the source code is available at <https://github.com/CompEpigen/figeno>.

1 Introduction

Recent advances in sequencing have led to the generation of many different types of genomic and epigenomic data, which require visualization tools for facilitating analysis and interpretation. An important requirement for such tools is the support of various data types, for large datasets, and for generating publication-quality images. The integrative genomics viewer (IGV) (Robinson *et al.* 2011) enables interactive exploration of genomics data, but lacks support for some data types like chromatin conformation (e.g. HiC). Furthermore, IGV has limited export capabilities, making it subpar for generating publication-quality figures. Other tools are geared toward a specific data type, e.g. trackplot (Mayakonda and Westermann 2024) for bigwig files, while others like Gviz (Hahne and Ivanek 2016) or PyGenomeTracks (Lopez-Delisle *et al.* 2021) can handle more data types. An important limitation of most genomic visualization software is that they only display one genomic region at a time, which is admittedly sufficient for many use cases but e.g. precludes the visualization of novel chromatin interactions across breakpoints in HiC data. NeoLoopFinder (Wang *et al.* 2021) has a visualization component which enables the visualization of HiC data across breakpoints, but it is not primarily a visualization tool. Third generation sequencing technologies (Oxford Nanopore Technologies, Pacific Biosciences) provide long reads which can be phased to each parental haplotype, as well as base modification information. Most visualization software do not support long reads, while new tools such as methplotlib (De Coster *et al.* 2020) and

methyartist (Cheetham *et al.* 2022) have been specifically designed to display base modifications from long reads, but lack other common features such as support for bigwig and HiC data. Here, we introduce figeno, an application for visualizing genomics and epigenomics data, including long reads, which allows multiple regions to be visualized simultaneously.

2 Figeno: a flexible, versatile, and user-friendly solution for genomic plotting

2.1 Figeno overview

Figeno is a visualization tool for genomic data tracks (Fig. 1A). Figeno's key features are (i) the support of many data types; (ii) multi-region views; (iii) the support for base modifications from long reads; (iv) export into various image formats (see Supplementary Table S1 for a comparison to other visualization software). Of the ten types of tracks that can be visualized with figeno (Fig. 1B), the first three tracks serve general purposes: "chr_axis" specifies the genomic coordinates, "genes" shows annotated genes in the region, and "bed" allows for additional custom genomic annotations. Gene files are provided for the reference genomes hg19, hg38, and mm10, but users can also provide custom annotation files for further reference genomes (in NCBI RefSeq or GTF format). "bigwig" tracks allow for the visualization of epigenetic data types including ChIP-seq or ATAC-seq, and "hic" tracks can be used to visualize HiC data (.cool format). For alignment files in bam format, figeno visualizes

Received: 21 March 2024; Revised: 17 May 2024; Editorial Decision: 29 May 2024; Accepted: 3 June 2024

© The Author(s) 2024. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

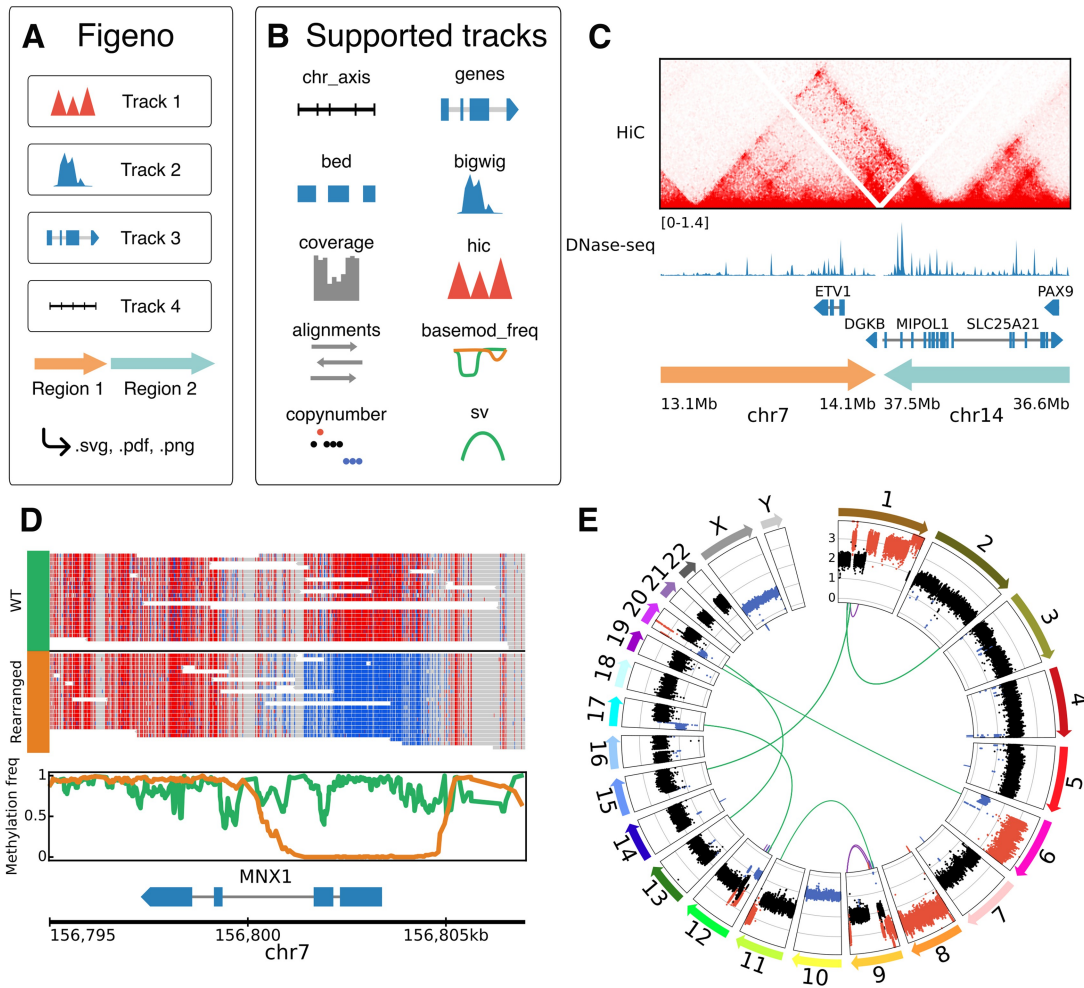


Figure 1. Figeno overview and example outputs. (A) Overview of figeno. (B) List of all 10 track types available within figeno. (C) HiC data around a breakpoint in the LNCaP cell line (Wang *et al.* 2021), and DNase-seq bigwig track from the LNCaP cell line, downloaded from the ENCODE project (ENCSTR000EKT). (D) Allele-specific DNA methylation at the MNX1 locus in the GDM-1 cell line, based on nanopore sequencing data. For the alignments track, red indicates a methylated CpG site and blue unmethylated. (E) Circular plot showing copy numbers and structural variants for the THP-1 cell line [data from the cancer cell line encyclopedia (Ghandi *et al.* 2019)]. Copy-number gains are indicated in red and losses in blue.

either the alignments themselves, the coverage, or base modification frequency (see section long-read support). Finally, figeno supports the visualization of whole genome sequencing data with tracks for copy number and structural variants. The figures can be exported as bitmap (png) or vector (svg, pdf) graphics formats for direct use in publication or further processing with a vector graphics editor. A configuration file in json format serves as input to figeno. In addition to custom generation of the configuration file, we provide a graphical user interface (GUI) for interactive figure configuration (Supplementary Fig. S1).

2.2 Multi-region plotting

Figeno can plot one or multiple regions simultaneously. An important feature is that, although some tracks will be plotted independently for each region, other tracks can show interactions across regions: HiC tracks show chromatin interactions across genomic regions (Fig. 1C), structural variant (SV) tracks can show breakpoints across regions (Fig. 1E), alignment tracks can link multiple alignments from split

reads aligned to different regions (Supplementary Fig. S2), and scales in bigwig and coverage tracks can automatically be adjusted across regions.

2.3 Support for long-read data

Figeno implements features unique to long-read sequencing data, including grouping alignments by haplotype [if the HP tag is set in the bam file, e.g. by WhatsHap (Martin *et al.* 2016)] and coloring by base modification (if the MM and ML tags are provided in the bam file). In addition to the most common base modification (5mC), any base modification (e.g. 5hmC, 6mA) can be visualized. Furthermore, two base modifications can be visualized at the same time (e.g. 5mC and 5hmC). We also support a “basemod_freq” track to specify the base modification frequency at each position in a haplotype-aware manner. The alignments and “basemod_freq” tracks can display allele-specific DNA methylation. The leukemic cell line GDM-1 harbors a translocation between chromosomes six and seven, which activates *MNX1* by enhancer hijacking (Weichenhan *et al.* 2023). By visualizing allele-specific

methylation in this region using figeno, we observed that the wild type allele is methylated at the *MXN1* promoter while the rearranged allele is hypomethylated (Fig. 1D). Finally, reads aligning to different genomic regions (“split reads”) can be visualized by a line connecting all alignments from the same read (Supplementary Fig. S2).

2.4 Customizable figure layouts

The default layout is “horizontal” and results in the regions being arranged horizontally from left to right. This layout is best-suited for most applications, but we also offer several other layouts that can be particularly useful for plotting whole genome sequencing data. First, the “circular” layout can be used to display all regions on a circle, e.g. to show all chromosomes in a circos plot (Fig. 1E). For settings where we only want to look at 2–8 chromosomes with breakpoints between them, we also provide a novel “symmetrical” layout, where the regions are aligned on two rows, but the order of the tracks in the top row is reversed, which is particularly useful when an SV track is displayed between the two rows (Supplementary Fig. S3).

3 Implementation and availability

Figeno is implemented in python and uses matplotlib (Hunter 2007) for plotting. It relies on pysam (Li *et al.* 2009) for reading bam files, pybigwig for reading bigwig files, cooler (Abdennur and Mirny 2020) for reading HiC data, and vcfs for reading vcf files. The GUI was created using the javascript framework React and utilizes a local Flask-based webserver. The time required to generate a figure depends on the number and types of tracks, on the size of the genomic regions being visualized and on the computer being used, but figeno generally only takes a couple of seconds to generate a figure. The code is completely open-source and is available on Github (<http://github.com/CompEpigen/figeno>) under the GPL-3 license; the present manuscript describes figeno version 1.2.0.

4 Conclusion

Taken together, figeno is a rich and user-friendly visualization tool for genomics and epigenomics data, especially for structural rearrangements and long-read data. It supports an extensive collection of input formats (Supplementary Table S1) and provides a GUI for enhanced usability by a broad range of users, from experienced bioinformaticians to beginners. To support them, we also provide extensive and detailed documentation on ReadTheDocs (<https://figeno.readthedocs.io/>).

Acknowledgements

We thank Anna Riedel, Marvin Mayer, Simge Kelekçi, Yunhee Jeong, and Katherine Kelly (Division Cancer Epigenomics, DKFZ, Heidelberg, Germany) for testing figeno. We thank Alexander Vogel (Oxford Nanopore Technologies) for valuable discussion on setting up nanopore sequencing.

Author contributions

E.S. developed figeno and wrote the original manuscript draft. J.H. performed nanopore sequencing of the GDM-1

cell line. C.G., C.P., M.S., and P.L. provided supervision. All authors reviewed and approved the final manuscript.

Supplementary data

Supplementary data are available at *Bioinformatics* online.

Conflict of interest

None declared.

Funding

This work was supported in part by the German Funding Agency (DFG) [SFB 1074 to C.P.] and the Helmholtz Foundation.

Data availability

ONT data was generated specifically for the manuscript. DNA from the GDM-1 cell line was extracted with the QIAamp DNA micro kit. Short fragments were removed with the PacBio SRE XS kit. Library preparation was performed with the ligation sequencing kit SQK-LSK114. The library was sequenced on one PromethION flow cell for 96 h, with a wash and reload after 48 h, which resulted in a coverage of 56x with an N50 of 21.5 kb. Nanopore sequencing data for the AML cell line GDM-1 is available on the SRA (SRA ID: SRR28257102). All other data used in this manuscript were already publicly available: the HiC data of the LNCaP cell line from GEO accession GSE161493 (Wang *et al.* 2021), the DNase-seq data of LNCaP from the ENCODE project (file ID ENCSR000EKT), the WGS data of the THP-1 from the cancer cell line encyclopedia (Ghandi *et al.* 2019) at the SRA with accession SRR8670675. All data required to reproduce the figures in the manuscript are provided at https://github.com/CompEpigen/figeno/tree/main/test_data.

References

- Abdennur N, Mirny LA. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics* 2020;36:311–6. <https://doi.org/10.1093/bioinformatics/btz540>.
- Cheetham SW, Kindlova M, Ewing AD. Methlyart: tools for visualizing modified bases from nanopore sequence data. *Bioinformatics* 2022; 38:3109–12. <https://doi.org/10.1093/bioinformatics/btac292>.
- De Coster W, Stovner EB, Strazisar M. Methplotlib: analysis of modified nucleotides from nanopore sequencing. *Bioinformatics* 2020; 36:3236–8. <https://doi.org/10.1093/bioinformatics/btaa093>.
- Ghandi M, Huang FW, Jané-Valbuena J *et al.* Next-generation characterization of the cancer cell line encyclopedia. *Nature* 2019;569: 503–8. <https://doi.org/10.1038/s41586-019-1186-3>.
- Hahne F, Ivanek R. Visualizing genomic data using gviz and bioconductor. In: Mathé E, Davis S (eds.), *Statistical Genomics: Methods and Protocols*. New York, NY: Springer; 2016, 335–51. https://doi.org/10.1007/978-1-4939-3578-9_16.
- Hunter JD. Matplotlib: a 2D graphics environment. *Comput Sci Eng* 2007;9:90–5. <https://doi.org/10.1109/MCSE.2007.55>.
- Li H, Handsaker B, Wysoker A *et al.*; 1000 Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. *Bioinformatics* 2009;25:2078–9. <https://doi.org/10.1093/bioinformatics/btp352>.
- Lopez-Delisle L, Rabbani L, Wolff J *et al.* pyGenomeTracks: reproducible plots for multivariate genomic datasets. *Bioinformatics* 2021;37:422–3. <https://doi.org/10.1093/bioinformatics/btaa692>.

- Martin M, Patterson M, Garg S *et al.* WhatsHap: fast and accurate read-based phasing. *Biorxiv* 2016. <https://doi.org/10.1101/085050>.
- Mayakonda A, Westermann F. Trackplot: a fast and lightweight R script for epigenomic enrichment plots. *Bioinf Adv* 2024;4: vbae031. <https://doi.org/10.1093/bioadv/vbae031>.
- Robinson JT, Thorvaldsdóttir H, Winckler W *et al.* Integrative genomics viewer. *Nat Biotechnol* 2011;29:24–6. <https://doi.org/10.1038/nbt.1754>.
- Wang X, Xu J, Zhang B *et al.* Genome-wide detection of enhancer-hijacking events from chromatin interaction data in rearranged genomes. *Nat Methods* 2021;18:661–8. <https://doi.org/10.1038/s41592-021-01164-w>.
- Weichenhan D, Riedel A, Meinen C *et al.* Translocation t(6;7) in AML-M4 cell line GDM-1 results in MNX1 activation through enhancer-hijacking. *Leukemia* 2023;37:1147–50. <https://doi.org/10.1038/s41375-023-01865-5>.