NG-Circos: next-generation Circos for data visualization and interpretation

Ya Cui1,†, Zhe Cui2,†, Jianfeng Xu2, Dapeng Hao2, Jiejun Shi1, Dan Wang3, Hui Xiao1, Xiaohong Duan4, Runsheng Chen5 and Wei Li1,*

1Division of Computational Biomedicine, Department of Biological Chemistry, School of Medicine, University of California, Irvine, CA 92697, USA, 2Division of Biostatistics, Dan L Duncan Cancer Center and Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, TX 77030, USA, 3Department of Medicine, Division of Cardiology, University of California, Los Angeles, CA 90095, USA, 4ChosenMed Technology (Beijing) Co. Ltd, Beijing 100176, China and 5CAS Key Laboratory of RNA Biology, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China

Received May 15, 2020; Revised August 04, 2020; Editorial Decision August 21, 2020; Accepted August 25, 2020

ABSTRACT

Circos plots are widely used to display multi-dimensional next-generation genomic data, but existing implementations of Circos are not interactive with limited support of data types. Here, we developed next-generation Circos (NG-Circos), a flexible JavaScript-based circular genome visualization tool for designing highly interactive Circos plots using 21 functional modules with various data types. To our knowledge, NG-Circos is the most powerful software to construct interactive Circos plots. By supporting diverse data types in a dynamic browser interface, NG-Circos will accelerate the next-generation data visualization and interpretation, thus promoting the reproducible research in biomedical sciences and beyond. NG-Circos is available at https://wlcb.oit.uci.edu/NG-Circos and https://github.com/YaCui/NG-Circos.

INTRODUCTION

Visualizing increasing volumes of next-generation biological data is critical to the interpretation of such data. Circos plots are circular two-dimensional visual representations that provide a comprehensive solution for presentation and interpretation of multi-dimensional genomic data. Circos (1), the predominant tool for making Circos plots, has been widely used for complex biological data visualization in many studies. However, Circos's outputs are not interactive. Other Circos-derived tools, such as Circosletto (2), CIRCUS (3), J-Circos (4), shinyCircos (5), Rcircos (6), Circleator (7), OmicCircos (8), ggbio (9) are either incapable to produce interactive Circos plots in a web browser or are limited to specific data types. Our previous developed tool, BioCircos.js (10), appears to be the only published software capable of producing interactive Circos plots and has become the state-of-the-art tool in the field (11–12). Nonetheless, BioCircos.js (10) implements only nine functional modules, limiting its scope to perform additional analytical tasks.

To address this weakness, here we developed next-generation Circos (NG-Circos), a JavaScript-based circular genome visualization tool that extends beyond the framework of BioCircos.js (10) to integrate and interpret genomic data types through interactive Circos plots. NG-Circos currently contains 21 modules, enabling various functions that were absent in other tools (including BioCircos.js (10)). By supporting diverse types genomic data types in an interactive browser interface, NG-Circos will accelerate the next-generation data visualization and interpretation, thus promoting reproducible research in biomedical sciences and beyond.

MATERIALS AND METHODS

Implementation of NG-Circos

NG-Circos is written in JavaScript and generates interactive graphics with SVG element based on D3.js (data-driven documents) and jQuery.js. Based on JavaScript, NG-Circos can be used without installing additional packages. After downloading NG-Circos, users can reproduce almost all circular plots drawn by Circos with a web browser. Note that NG-Circos itself is not a web application, but is a library to build interactive Circos plots in web applications.

Implementing image-download function in NG-Circos

The download function in NG-Circos is built using the svg-crowbar.js (https://nytimes.github.io/svg-crowbar/) from The New York Times. NG-Circos now supports the
**RESULTS**

**Workflow of NG-Circos**

NG-Circos has a highly user-friendly workflow. It has three main steps to draw an interactive Circos plot: Step 1 includes drawing chromosomes (or other segments) as the coordinate axes. Step 2 involves adding various data tracks using the relevant modules with high flexibility in module choices (21 modules are currently implemented, Supplementary Table S1). The input data of NG-Circos can be either generated by the supporting Python scripts, or directly through the well-documented JSON data formats. For each module, we provide one example which includes the input data files and all the steps to recreate that example (https://wlcb.oit.uci.edu/modules/). Finally, step 3 incorporates interactive animations, mouse events (Supplementary Table S2) and designing toolboxes for graphic elements. NG-Circos is highly customizable, allowing users to adjust personal settings. We also provide a set of carefully evaluated default settings for each module and provide many demos to make NG-Circos easy to use. In addition, the capability of NG-Circos can be simply broadened by including more functional modules in step 2.

**NG-Circos provides flexible module choices for diverse Circos plots**

The current version of NG-Circos consists of 21 modules (Supplementary Table S1). The combination of modules in NG-Circos allows users to construct diverse types of Circos plots. For example, NG-Circos can reproduce complex published Circos plots (15) by combining ARC, GENE,
HEATMAP, LINK and WIG modules (Figure 1A). Not only can NG-Circos reproduce complex published Circos plots, but also can it renders additional functions such as providing popular interactive Circos plot demos (e.g. Lollipop, Wig and LocusZoom (16) plots) shown in Figure 1B–F (15) (17) (18) (19), that are not seen in other tools. Moreover, we offer more demos in the online website (https://wlcb.oit.uci.edu/NG-Circos) to show the power of this tool: users can easily replace the demo data with their data to produce their own plots. All figures can be download in the SVG and PNG format, in which the SVG format renders users high-quality images that could be further utilized through other applications such as the Adobe Illustrator.

Overall, NG-Circos offers users great flexibility in module choices and Circos plot types.

Case study for interactive data exploration using NG-Circos
Here we present a case study to further illustrate the power of interactive data exploration using NG-Circos. In this case, users can interactively explore driver single nucleotide polymorphisms (SNPs), gene fusions and their impact on protein structure in lung cancer (Figure 2). For example, mouse over events show the SNP frequencies in lung cancer from the Catalogue of Somatic Mutations in Cancer (COSMIC) database (Figure 2B) (20) and the three-dimensional
(3D) protein structure of an EML4-ALK gene fusion (Figure 2C) (21). Remarkably, NG-Circos can also redirect elements (such as SNPs or gene fusions) to external resources. For instance, clicking on a SNP, such as the EGFR T790M variant, opens up a new Protein Data Bank (PDB) database webpage, displaying the T790M variant-affected 3D structure of EGFR (Figure 2D; PDB code: 2JIT) (22). To sum up, NG-Circos serves as a great tool to explore genomic data interactively such that users can extract additional information by mouse hovering and clicking on the plots.

DISCUSSION

Interactive data exploration across diverse data types will certainly promote the next-generation data visualization and interpretation, with some successful examples, such as cBioPortal (23), seen in cancer research. Circos plots are widely used to display voluminous next-generation genomic data, but existing implementations of Circos does not generate interactive outputs, which hinders its usability. To address this issue, NG-Circos provides flexible modules choices for interactive data exploration and diverse Circos plots types. As additional types of genomic data are generated in the future, we will keep updating additional functional modules to extend the power of NG-Circos. We will also actively maintain NG-Circos and respond to inquiries from users. By supporting diverse types of genomic data in an interactive web interface, NG-Circos, we believe, will enhance genomic research in the biomedical field in the future.

SUPPLEMENTARY DATA

Supplementary Data are available at NARGAB Online.

ACKNOWLEDGEMENTS

We acknowledge Tianyi Zang, Yadong Wang and members of the Li lab for constructive discussions and support.

FUNDING

No external funding. Conflict of interest statement. None declared.

REFERENCES