trackViewer: a Bioconductor package for interactive and integrative visualization of multi-omics data

To the Editor — As high-throughput sequencing's cost continues to decrease and its applications continue to grow, the amount of data and the number of data types produced by this method are expected to constantly expand. To integrate various types of data and generate hypotheses efficiently, researchers increasingly need concise and meaningful depictions of diverse, complex datasets. Several genome browsers and viewers have been developed for the visualization of genomic data[1–3], but the majority of those tools do not have an easy programming interface that can be plugged into a pipeline. In addition, methylation, mutation, and single-nucleotide polymorphism (SNP) data require a special type of plot, called a lollipop or needle plot, to concisely depict the methylation, mutation, and SNP status. Although several tools have been developed to generate stand-alone lollipop plots (reviewed in ref. [4]), all the genome browsers lack this function. In addition, it is difficult to use existing lollipop-plot tools to visualize dense mutation/SNP and methylation data.

Here we describe trackViewer, a Bioconductor package for the visualization of multi-omics data that can be integrated into any analysis pipeline in R. The installation guide, documentation, implementation details, data, and functionalities of trackViewer are provided in Supplementary Notes 1–6. trackViewer can be used not only to visualize coverage and annotation tracks, but also to generate lollipop and dandelion plots that depict sparse and dense methylation/mutation/variant data to facilitate an integrative analysis of diverse datasets[5,6]. The updated trackViewer (versions 1.19.27 and higher) has a web interface in addition to the

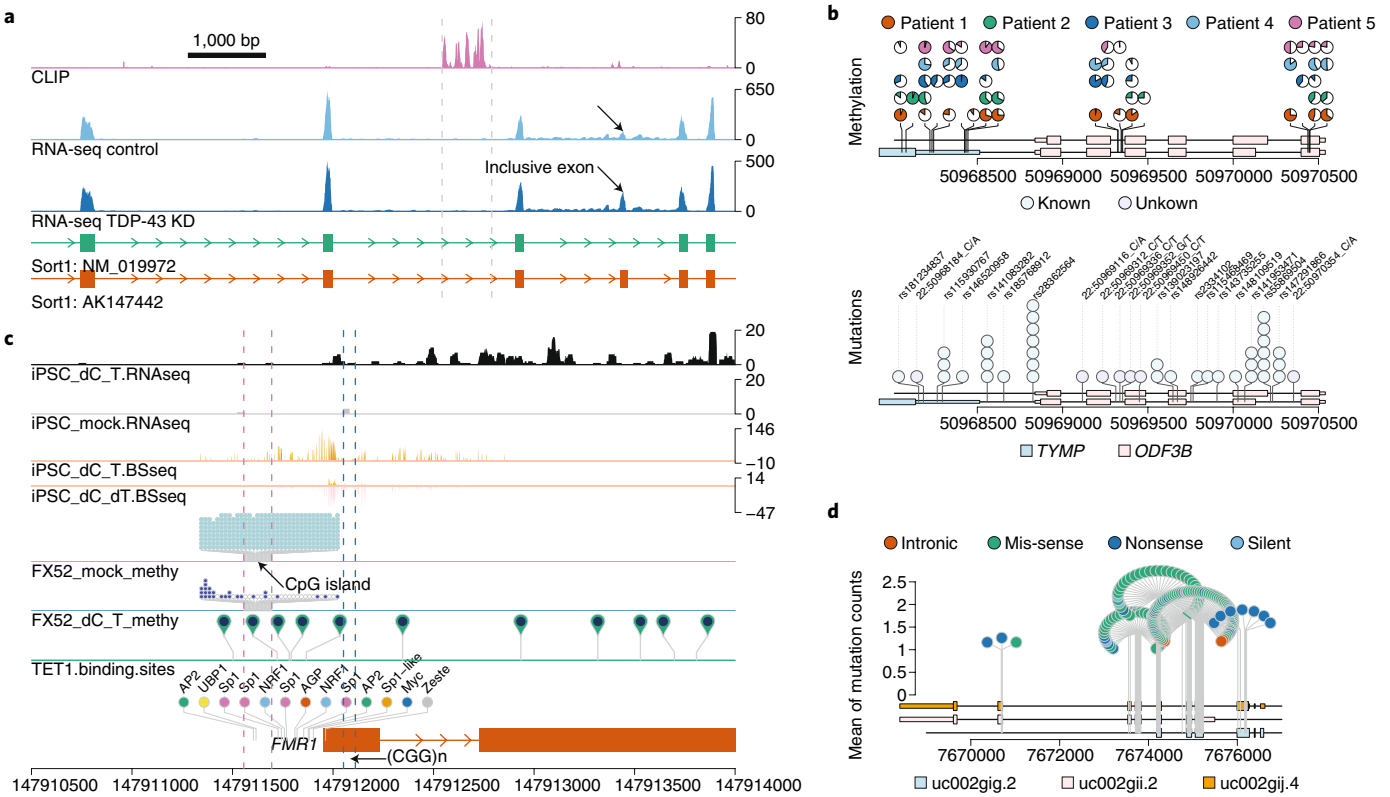

Fig. 1 | Examples to illustrate how trackViewer can be used for integrative visualization of multi-omics data. a, Visualization of RNA-seq and CLIP-seq data along with gene models of *Sort1* using 'viewTracks'. Arrows and dashed lines are used to highlight regions of interest. KD, knockdown. **b**, Lollipop plots of methylation data and mutation data for genes *TYMP* and *ODF3B* from multiple individuals, created with lolliplot. In the methylation plot (top), the white part of a circle indicates the methylated percentage, and the colored part indicates the unmethylated percentage. In the mutation plot (bottom), different colors depict different SNP/mutation events, and the number of circles indicates the number of mutation/SNP events. **c**, Visualization of coverage tracks from multiple types of data, together with lollipop plots of methylation data, binding sites of TET1 and several other TFs, and the gene model of the *FMR1* promoter. CGG repeats in the first exon of *FMR1* are also indicated by a labeled arrow at the bottom. **d**, Visualization of dense mutation data for *TP53* in a dandelion plot created with dandelion.plot.

R programming interface (Supplementary Video 5, Supplementary Note 7). Furthermore, with the 'browseTracks' function, users can generate interactive figures—that is, figures one can easily customize the features of by clicking, dragging, and typing (Supplementary Videos 1–4, Supplementary Note 3).

In Fig. 1, we illustrate how trackViewer can be used to visualize different types of datasets. The simultaneous visualization of RNA-seq and CLIP-seq data along gene models (Fig. 1a) facilitated the discovery that binding of the DNA/RNA-binding protein TARDBP to regions upstream of exon 17 can regulate the expression of exon 18 in *Sort1*. Concise depiction of methylation and mutation/SNP data (Fig. 1b) might help to uncover the correlation between methylation status and mutation status. The simultaneous visualization of methylation status, RNA-seq track, transcription factor (TF) binding sites, and gene models (Fig. 1c) led to the hypothesis that abnormal methylation of CpG islands in the promoter region of *FMR1* may affect TF binding and, subsequently, *FMR1* expression. Sometimes, there are hundreds of mutations located in one gene. To represent such dense distributions in a compact way, trackViewer allows users to generate a dandelion plot (Fig. 1d), with the height of the dandelion indicating the density of the mutations, and the colors representing the mutation types.

As we have been doing for our other Bioconductor packages, such as ChIPpeakAnno, CRISPRseek, motifStack, ATACseqQC, cleanUpdTSeq, GUIDEseq, NADfinder, InPAS, and dagLogo, we will actively maintain trackViewer and respond to users' questions and feature requests under Bioconductor (https://support.bioconductor.org). We envision that more types of genomics data will be generated in the near future, and we will develop functions to convert the new data types to genomic ranges that can be visualized in trackViewer.

Reporting Summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All datasets analyzed during this work are publicly available and are described in detail in Supplementary Note 6. In addition, we have created a Docker image that contains all data and scripts necessary for users to reproduce the figures as either static or interactive (Supplementary Note 2).

Code availability

trackViewer is freely available at https://bioconductor.org/packages/release/bioc/html/trackViewer.html. To make it easy to install and run, we have created a Docker image with trackViewer, RStudio, and all its R and system dependencies already installed (https://github.com/jianhong/trackViewer.documentation). In addition, the Docker image contains all data and scripts necessary for users to reproduce the figures as either static or interactive (Supplementary Note 2). Supplementary Note 7 provides step-by-step instructions on how to install and run the web application of trackViewer, with four examples. ❐

Jianhong Ou[1*] and Lihua Julie Zhu[2,3,4*]
[1]*Department of Cell Biology, Duke University School of Medicine, Durham, NC, USA.* [2]*Department of Molecular, Cell and Cancer Biology, University of Massachusetts Medical School, Worcester, MA, USA.* [3]*Program in Molecular Medicine, University of Massachusetts Medical School, Worcester, MA, USA.* [4]*Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, Worcester, MA, USA.*
**e-mail: jianhong.ou@duke.edu; julie.zhu@umassmed.edu*

Published online: 27 May 2019
https://doi.org/10.1038/s41592-019-0430-y

References
1. Robinson, J. T. et al. *Nat. Biotechnol.* **29**, 24–26 (2011).
2. Wang, J., Kong, L., Gao, G. & Luo, J. *Brief. Bioinform.* **14**, 131–143 (2013).
3. Haeussler, M. et al. *Nucleic Acids Res.* **47**, D853–D858 (2019).
4. Jay, J. J. & Brouwer, C. *PLoS One* **11**, e0160519 (2016).
5. Ben-David, E., Burga, A. & Kruglyak, L. *Science* **356**, 1051–1055 (2017).
6. Huang, D. et al. *Proc. Natl Acad. Sci. USA* **114**, E3149–E3158 (2017).

Acknowledgements
We thank S. Han for editorial and web design assistance, and L. Ni and J. Yu for helpful discussion. L.J.Z. is partially supported by the NIH/NIDDK (grant R01DK076118).

Author contributions
J.O. developed the software and generated Fig. 1 with input from L.J.Z. L.J.Z. drafted the manuscript with input from J.O. Both authors read and approved the final manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41592-019-0430-y.

nature research

Corresponding author(s): Lihua Julie Zhu and Jianhong Ou

Last updated by author(s): April 5th 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☒	☐ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Data used to generate the figures are described in the in the Supplementary Method section. To facilitate the reproduction of the figures, the data and code used to generate the figures are made available as a docker image, which is described in the Supplementary Methods section. VariantAnnotation package 1.28.12 were used to obtain the variant data for generating Figure 1B. No software were used to obtain the data for generating the other figures.
Data analysis	trackViewer is used for generating the figures and it is freely available at http://bioconductor.org/packages/release/bioc/html/trackViewer.html for multiple platforms including Windows, MacOS and UNIX/Linux. To facilitate the reproduction of the figures, the data and code used to generate the figures are made available as a docker image, which is described in the Supplementary Methods section in detail.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data are publicly available with no restriction to access. To generate Figure 1A, TDP-43 cross-linking and immunoprecipitation coupled with high-throughput sequencing (CLIP-seq) and corresponding RNA-seq mapped files were downloaded from gene expression omnibus database (GEO) (GSE27394). To make lollipop plots in Figure 1B, SNPs from a subset of 1000 variants and 50 samples from chromosome 22 were taken from 1000 Genomes stored in VariantAnnotation package

(https://bioconductor.org/packages/release/bioc/html/VariantAnnotation.html) . To generate dandelion plots in Figure 1C, TP53 mutations were downloaded from http://p53.iarc.fr/DownloadDataset.aspx. To create the methylation profile together with gene expression track and TF binding motifs in Figure 1D, data were downloaded from GEO (GSE108577) and http://cisbp.ccbr.utoronto.ca/.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Not applicable since trackViewer is a visualization tool
Data exclusions	Not applicable since trackViewer is a visualization tool
Replication	Not applicable since trackViewer is a visualization tool
Randomization	Not applicable since trackViewer is a visualization tool
Blinding	Not applicable since trackViewer is a visualization tool

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging