

SOFTWARE

Open Access

Uphyloplot2: visualizing phylogenetic trees from single-cell RNA-seq data



Stefan Kurtenbach^{*†} , Anthony M. Cruz[†] , Daniel A. Rodriguez , Michael A. Durante  and J. William Harbour 

Abstract

Background: Recent advances in single cell sequencing technologies allow for greater resolution in assessing tumor clonality using chromosome copy number variations (CNVs). While single cell DNA sequencing technologies are ideal to identify tumor sub-clones, they remain expensive and in contrast to single cell RNA-seq (scRNA-seq) methods are more limited in the data they generate. However, CNV data can be inferred from scRNA-seq and bulk RNA-seq, for which several tools have been developed, including inferCNV, CaSpER, and HoneyBADGER. Inferences regarding tumor clonality from CNV data (and other sources) are frequently visualized using phylogenetic plots, which previously required time-consuming and error-prone, manual analysis.

Results: Here, we present Uphyloplot2, a python script that generates phylogenetic plots directly from inferred RNA-seq data, or any Newick formatted dendrogram file. The tool is publicly available at <https://github.com/harbourlab/UPhyloplot2/>.

Conclusions: Uphyloplot2 is an easy-to-use tool to generate phylogenetic plots to depict tumor clonality from scRNA-seq data and other sources.

Background

Single cell RNA sequencing (scRNA-seq) has become an important new tool for studying gene expression in individual cells of heterogenous samples. While this technology is still maturing, it is already providing powerful new insights into normal and diseased tissue types [1, 2]. In particular, single cell technology has resulted in great strides in cancer research. A hallmark of cancer cells is aneuploidy and chromosomal copy number variations (CNVs), which often correlate with tumor aggressiveness [3–6]. CNVs can be used to identify subclones of tumor cells and to infer tumor evolution, which can have important clinical implications [7]. Single cell sequencing can be used to analyze subclonal tumor architecture at unprecedented resolution [1, 8]. While single cell DNA sequencing (scDNA-seq) is an emerging technique for

this type of analysis, it is very expensive and yet to be optimized. Alternatively, CNVs can be inferred from scRNA-seq and bulk RNA-seq using applications such as inferCNV [9], HoneyBadger [10], and CaSpER [11]. Following, these applications cluster the inferred CNV patterns, allowing to define discrete subclones and infer tumor evolution. This approach for studying tumor clonality and evolution has been used successfully by our group and others [8, 12]. Tumor evolution is commonly visualized with phylogenetic plots, where the length of tree branches is proportional to the number of cells in each subclone. This, in contrast to plotting the dendrogram files, allows for a simple and intuitive representation of tumor evolution. Until now, such visualization required time-consuming and error-prone manual curation. Here we describe a new tool called Uphyloplot2. This program uses inferCNV output files to generate phylogenetic plots depicting tumor evolution, and also works with any other Newick formatted

* Correspondence: Stefan.Kurtenbach@med.miami.edu
Sylvester Comprehensive Cancer Center, Interdisciplinary Stem Cell Institute, Bascom Palmer Eye Institute, University of Miami Miller School of Medicine, Miami, FL, USA



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

dendrogram files such as those derived from HoneyBADGER and CaSpER (Fig. 1).

Implementation

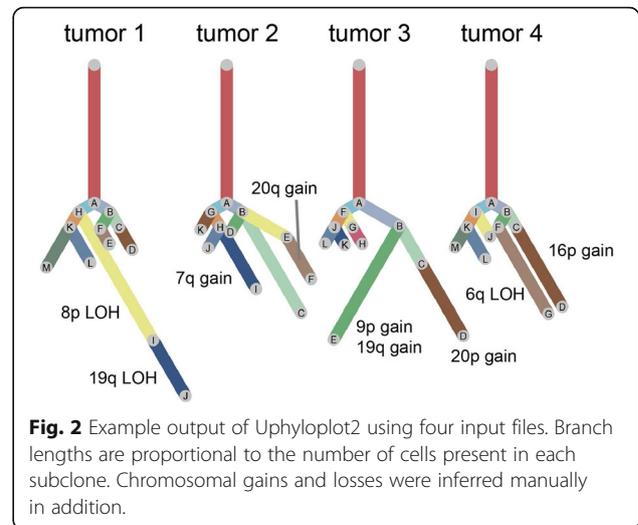
Uphyloplot2 was written entirely in Python 3 to enable pipeline integration, customization, and platform independence.

Availability and requirements

Project name: Uphyloplot2. Project home page: <https://github.com/harbourlab/Uphyloplot2/>. Operating system(s): Platform independent. Programming language: Python. Other requirements: None. License: GNU General Public License v3.0. Any restrictions to use by non-academics: No.

Results

To infer tumor clonality/evolution from scRNA-seq data, we first ran the inferCNV [9] pipeline on four uveal melanoma tumor samples [8] to infer CNVs from RNA-seq and cluster cells into subclones. inferCNV must be run with “HMM” to generate a “HMM_CNVPredictions.*.cell_groupings” file, which contains information on cell clusters. Following, reference cells (normal controls) were removed from that file manually before plotting. Uphyloplot2 can plot multiple trees at once and will plot all files placed in the “Input” directory in one figure. In the example above, we used all four “cell_groupings” files to produce the four phylogenetic trees depicted in Fig. 2. The first branch (seen in red) always has the same length and is introduced to depict the evolution of normal cells to tumor cells. All following

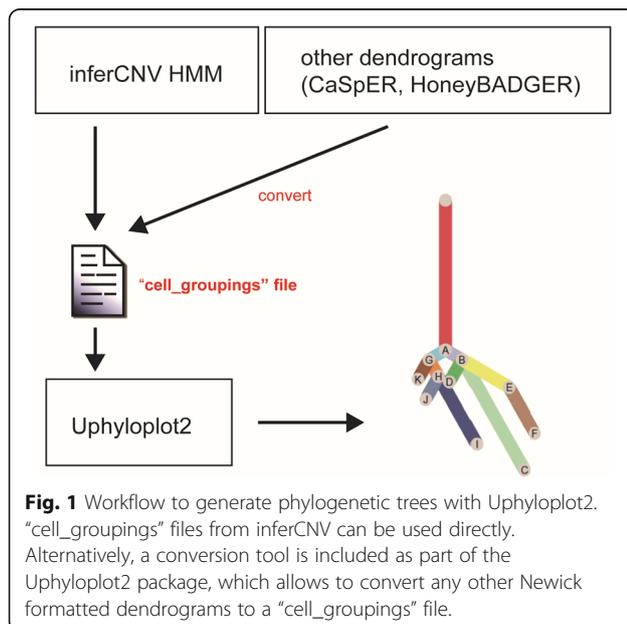


branches are labeled with letters corresponding to distinct tumor subclones. The branch length correlates with the number of cells in the respective subclone. For instance, in tumor 1 most cells are found in cluster “I” and “J”, where “J” is predicted to have directly evolved from “I”. Subsequently, more detailed information on which chromosomal regions were gained and lost for each subclone can be obtained from the “.HMM_CNVPredictions.*.pred_cnv_regions.dat” file. For example, cells in cluster “J” have lost part of chromosome 19q, in addition to the chromosome 8p loss found in cluster “I”. As can be seen in this simple example, sub-clonality of the four tumor samples differs substantially, and indicates the presence of multiple evolutionary branches.

Uphyloplot2 was designed to work directly with the “.cell_groupings” output from inferCNV after removing reference cells. Additionally, Uphyloplot2 can plot user derived, Newick formatted dendrogram files, for instance exported from HoneyBadger, CaSpER, or inferCNV if preferred. Using dendrogram files requires additional processing steps: In brief, using R the dendrogram has to be exported in a “Newick” format. Second, the Uphyloplot2 folder contains a python script called “newick_input.py”, which can be used to convert the Newick file to a “.cell_groupings” file. Once the “.cell_groupings” files are generated, they can be used as outlined above. A detailed user guide is available on the Uphyloplot2 GitHub page.

Conclusions

The python script presented here allows to plot phylogenetic trees of tumor subclones from inferCNV output files and other Newick formatted dendrograms. The output files generated are true Scalable Vector Graphics (SVG) files, enabling easy attribute editing like colors,



lengths, or angles in any SVG editor, while maintaining high resolution. Depending on the datasets, some branches might overlap in the figure, however, these can easily be rotated for visual clarity. In contrast to algorithms that estimate molecular time from whole-genome sequencing data using mutations [13], the use of CNVs to infer clonality and tumor evolution is more complex because some chromosomal segments are selectively altered while others occur through massive genome reorganization such as chromothripsis [14, 15], chromoplexy [16] and anaphase catastrophe [17]. It is important to note that UPhyloplot2 evolutionary plots might not represent molecular time accurately. UPhyloplot2 constructs trees with subclone branch lengths proportional to the number of cells in each subclone. New methodologies are also being developed for analyzing single cell CNV and single cell mutation data [18]. In summary, we present an automated tool for generating phylogenetic trees from scRNA-seq data that allows the visualization of tumor subclones and heterogeneity.

Abbreviations

CNVs: Copy number variations; scRNA-seq: Single cell RNA-seq; scDNA-seq: Single cell DNA sequencing; SVG: Scalable Vector Graphics

Acknowledgements

N/A.

Authors' contributions

SK developed UPhyloplot2, and prepared the manuscript. AMC helped integrating HoneyBadger and CaSpER support. MAD helped with design and manuscript preparation. DAR helped with pipeline generation and data interpretation. JWH helped with overall design, data interpretation, and manuscript preparation. The author(s) read and approved the final manuscript.

Funding

This work was supported by Melanoma Research Foundation Career Development Award (Kurtenbach) and Established Investigator Award (Harbour), National Cancer Institute grant R01 CA125970 (Harbour), A Cure in Sight Jack Odell-John Dagues Research Award (Kurtenbach, Harbour), Bankhead-Coley Research Program of the State of Florida (Harbour), The Helman Family-Melanoma Research Alliance Team Science Award (Harbour) and a generous gift from Dr. Mark J. Daily (Harbour). The Bascom Palmer Eye Institute received funding from NIH Core Grant P30EY014801 and a Research to Prevent Blindness Unrestricted Grant. The Sylvester Comprehensive Cancer Center also received funding from the National Cancer Institute Core Support Grant P30CA240139. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Availability of data and materials

The tool is publicly available at <https://github.com/harbourlab/UPhyloplot2/>, including example data.

Declarations

Ethics approval and consent to participate:

N/A.

Consent for publication

Yes.

Competing interests

None.

Received: 17 December 2020 Accepted: 25 May 2021

Published online: 05 June 2021

References

- Eddy J, Maizels N. Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Res.* 2006;34(14):3887–96.
- Durante MA, et al. Single-cell analysis of olfactory neurogenesis and differentiation in adult humans. *Nat Neurosci.* 2020;23(3):323–6.
- Ben-David U, Amon A. Context is everything: aneuploidy in cancer. *Nat Rev Genet.* 2020;21(1):44–62.
- Davoli T, et al. Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science.* 2017. 355(6322).
- Duijff PH, Schultz N, Benezra R. Cancer cells preferentially lose small chromosomes. *Int J Cancer.* 2013;132(10):2316–26.
- Ehlers JP, et al. Integrative genomic analysis of aneuploidy in uveal melanoma. *Clin Cancer Res.* 2008;14(1):115–22.
- Yates LR, Campbell PJ. Evolution of the cancer genome. *Nat Rev Genet.* 2012;13(11):795–806.
- Durante MA, et al. Single-cell analysis reveals new evolutionary complexity in uveal melanoma. *Nat Commun.* 2020;11(1):496.
- inferCNV of the Trinity CTAT Project.
- Fan J, et al. Linking transcriptional and genetic tumor heterogeneity through allele analysis of single-cell RNA-seq data. *Genome Res.* 2018;28(8): 1217–27.
- Serin Harmanci A, Harmanci AO, Zhou X. CaSpER identifies and visualizes CNV events by integrative analysis of single-cell or bulk RNA-sequencing data. *Nat Commun.* 2020;11(1):89.
- Fricke R, et al. Checklist of the marine and estuarine fishes of New Ireland Province, Papua New Guinea, western Pacific Ocean, with 810 new records. *Zootaxa.* 2019. 4588(1): p. zootaxa 4588 1 1.
- Nik-Zainal S, et al. The life history of 21 breast cancers. *Cell.* 2012;149(5):994–1007.
- Stephens PJ, et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell.* 2011;144(1):27–40.
- Cortes-Ciriano I, et al. Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing. *Nat Genet.* 2020;52(3): 331–41.
- Baca SC, et al. Punctuated evolution of prostate cancer genomes. *Cell.* 2013; 153(3):666–77.
- Galimberti F, et al. Anaphase catastrophe is a target for cancer therapy. *Clin Cancer Res.* 2011;17(6):1218–22.
- Madipour-Shirayeh A, et al., Simultaneous Profiling of DNA Copy Number Variations and Transcriptional Programs in Single Cells using RNA-seq. *bioRxiv.* 2020: p. 2020.02.10.942607.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

