UPhyloplot2: Visualizing Phylogenetic Trees from Single-Cell RNA-seq Data

Stefan Kurtenbach*, Daniel A. Rodriguez, Michael A. Durante, and J. William Harbour

Bascom Palmer Eye Institute, Sylvester Comprehensive Cancer Center, Interdisciplinary Stem Cell Institute, University of Miami Miller School of Medicine, Miami, FL, USA.
*Corresponding author

Abstract

Recent advances in single cell sequencing technologies allow for greater resolution in assessing tumor clonality using chromosome copy number variations (CNVs), which can be inferred from single cell RNA-seq (scRNA-seq) data using applications such as inferCNV. Inferences regarding tumor clonality are frequently visualized using phylogenetic plots, which previously required time-consuming and tedious manual analysis. Here, we present UPhyloplot2, a python script that generates phylogenetic plots directly from inferCNV output files. The tool is publicly available at https://github.com/harbourlab/UPhyloplot2/.

Introduction

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 is aneuploidy – an abnormal number of chromosomes or chromosomal segments – which often correlates with tumor aggressiveness [3-6]. Further, aneuploidy 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 using applications such as inferCNV [9], HoneyBadger [10], and CaSpER [11], using gene expression patterns to infer CNVs and to cluster cells into putative subclones. This approach for studying tumor clonality has been used successfully by our group and others [8, 12]. Tumor clonality is commonly used to visualize phylogenetic plots, where the length of tree branches is proportional to the number of cells in each subclone. Until now, such visualization required time-consuming and error-prone manual curation. Here we describe a new tool that we have created called UPhyloplot2, which is an

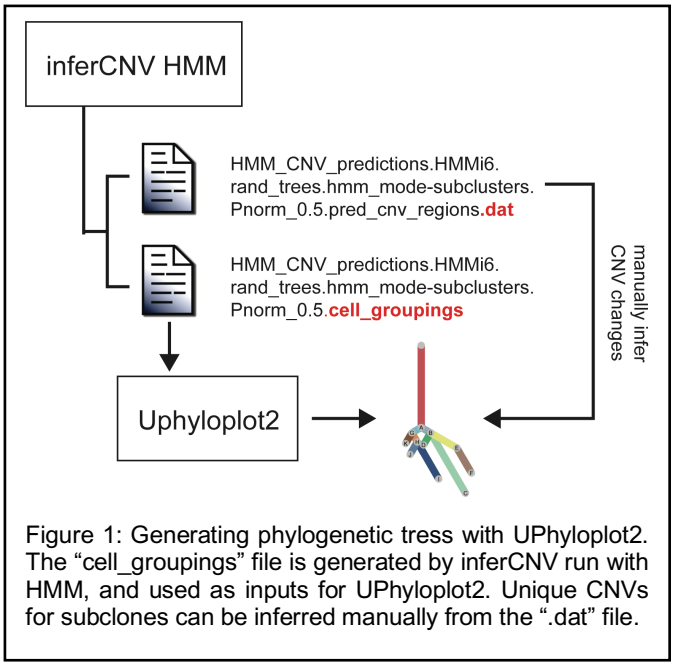

Figure 1: Generating phylogenetic tress with UPhyloplot2. The "cell_groupings" file is generated by inferCNV run with HMM, and used as inputs for UPhyloplot2. Unique CNVs for subclones can be inferred manually from the ".dat" file.

enhanced version of UPhyloplot [13]. This application directly takes inferCNV output files and generates evolutionary phylogenetic plots.

Results

UPhyloplot2 works directly with the output files of inferCNV, plotting evolutionary trees with the length of each branch correlating with the number of cells in the respective subclone. UPhyloplot2 was written in Python 3, making it easy to run on any platform (Figure 1). Uphyloplot2 uses the "HMM_CNV_predictions.*.cell_groupings" file generated by inferCNV (with HMM) to plot. Multiple "cell_groupings" files can be processed at once, to generate one output figure containing one tree per file. Subsequently, unique CNVs can be inferred using the ".HMM_CNV_predictions.*.pred_cnv_regions.dat" file. Four "cell_groupings" file being used as inputs (Figure 2). Output files are true SVG files, allowing for easy editing (colors, lines, branch rotation) in programs like Adobe Illustrator or any other SVG editor.

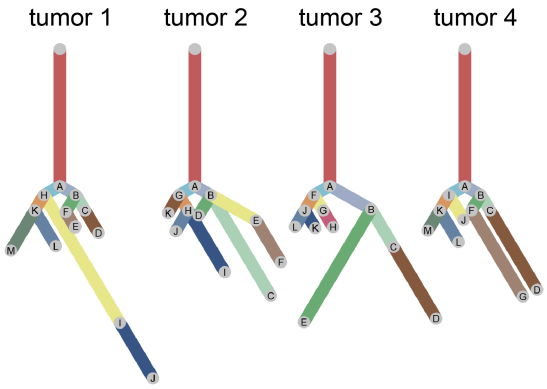

Figure 2. Example plots generated from four cell groupings files at once. Some branches were rotated to avoid overlapping.

Discussion

The python script presented here allows phylogenetic trees of tumor subclones to be plotted from inferCNV output files. The script and further documentation are publicly available at https://github.com/harbourlab/UPhyloplot2/. In contrast to algorithms that estimate molecular time from whole-genome sequencing data using the number of mutations [14], 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 [15, 16], chromoplexy [17] and anaphase catastrophe [18]. It is important to note that our methodology for plotting phylogenetic trees with branch lengths being proportional to the number of cells does not attempt to depict molecular time, but rather, the proportional size of each subclone. New methodologies are also being developed for analyzing single cell CNV and single cell mutation data [19]. In summary, we present an automated tool for generating phylogenetic trees from scRNA-seq data that allows the visualization of tumor subclones and heterogeneity.

References

1. Eddy, J. and N. Maizels, *Gene function correlates with potential for G4 DNA formation in the human genome.* Nucleic Acids Res, 2006. **34**(14): p. 3887-96.
2. Durante, M.A., et al., *Single-cell analysis of olfactory neurogenesis and differentiation in adult humans.* Nat Neurosci, 2020. **23**(3): p. 323-326.
3. Ben-David, U. and A. Amon, *Context is everything: aneuploidy in cancer.* Nat Rev Genet, 2020. **21**(1): p. 44-62.
4. Davoli, T., et al., *Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy.* Science, 2017. **355**(6322).

5. Duijf, P.H., N. Schultz, and R. Benezra, *Cancer cells preferentially lose small chromosomes.* Int J Cancer, 2013. **132**(10): p. 2316-26.
6. Ehlers, J.P., et al., *Integrative genomic analysis of aneuploidy in uveal melanoma.* Clin Cancer Res, 2008. **14**(1): p. 115-22.
7. Yates, L.R. and P.J. Campbell, *Evolution of the cancer genome.* Nat Rev Genet, 2012. **13**(11): p. 795-806.
8. Durante, M.A., et al., *Single-cell analysis reveals new evolutionary complexity in uveal melanoma.* Nat Commun, 2020. **11**(1): p. 496.
9. *inferCNV of the Trinity CTAT Project.*
10. Fan, J., et al., *Linking transcriptional and genetic tumor heterogeneity through allele analysis of single-cell RNA-seq data.* Genome Res, 2018. **28**(8): p. 1217-1227.
11. Serin Harmanci, A., A.O. Harmanci, and X. Zhou, *CaSpER identifies and visualizes CNV events by integrative analysis of single-cell or bulk RNA-sequencing data.* Nat Commun, 2020. **11**(1): p. 89.
12. 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.
13. Field, M.G., et al., *Punctuated evolution of canonical genomic aberrations in uveal melanoma.* Nat Commun, 2018. **9**(1): p. 116.
14. Nik-Zainal, S., et al., *The life history of 21 breast cancers.* Cell, 2012. **149**(5): p. 994-1007.
15. Stephens, P.J., et al., *Massive genomic rearrangement acquired in a single catastrophic event during cancer development.* Cell, 2011. **144**(1): p. 27-40.
16. Cortes-Ciriano, I., et al., *Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing.* Nat Genet, 2020. **52**(3): p. 331-341.
17. Baca, S.C., et al., *Punctuated evolution of prostate cancer genomes.* Cell, 2013. **153**(3): p. 666-77.
18. Galimberti, F., et al., *Anaphase catastrophe is a target for cancer therapy.* Clin Cancer Res, 2011. **17**(6): p. 1218-22.
19. 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.