Profiles of expressed mutations in single cells reveal patterns of tumor evolution and therapeutic impact of intratumor heterogeneity by **Trisicell**

Farid Rashidi Mehrabadi*, Kerrie L. Marie*, Eva Pérez-Guijarro*, Salem Malikić*, Erfan Sadeqi Azer, Howard H. Yang, Can Kızılkale, Charli Gruen, Huaitian Liu, Michael C. Kelly, Christina Marcelus, Sandra Burkett, Aydın Buluç, Funda Ergün, Maxwell P. Lee, Glenn Merlino†, Chi-Ping Day†, S. Cenk Sahinalp†

NATIONAL CANCER INSTITUTE

INDIANA UNIVERSITY

Introduction

Advances in single-cell RNA sequencing (scRNAseq) technologies uncovered an unexpected complexity in solid tumors, underlining the relevance of intratumor heterogeneity for cancer progression and therapeutic resistance. Available tumor phylogeny reconstruction methods cannot computationally handle a large number of cells and mutations present in typical scRNAseq datasets. In addition, there are no principled methods to assess distinct subclones observed in inferred tumor phylogenies and the genomic alterations that seed them.

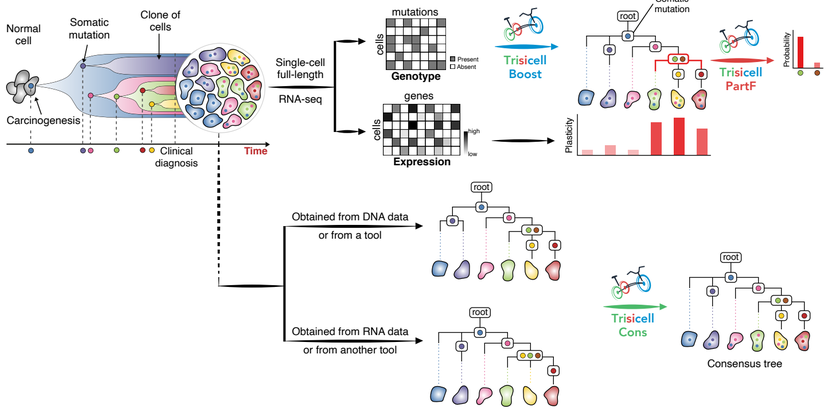
Trisicell (Triple-toolkit for single-cell tumor phylogenetics), is a new computational toolkit for scalable tumor phylogeny reconstruction and evaluation from scRNAseq (as well as single-cell genome or exome sequencing) data. Trisicell utilizes expressed SNVs and Indels to infer evolutionary relationships between genomic alterations and the cells that harbor them.

A comparison of the phylogenies built on three distinct types of sequencing datasets from a mouse melanoma model reveal that Trisicell can be robustly and scalably used on scRNAseq data. These datasets were derived from 24 distinct clonal sublines from the B2905 mouse melanoma model, each grown from a single cell into a genomically uniform population of cells. Specifically, each clonal subline was subjected to

1. High coverage bulk whole exome sequencing (bWES).
2. Single cell RNA sequencing (scRNAseq) through the Smart-seq2 protocol - on ~8 distinct cells sampled from each subline.
3. High coverage bulk whole transcriptome sequencing (bWTS) on a selected set of 11 sublines, which were implanted in syngeneic C57BL/6 mice - each with 3 replicates.

Trisicell was also applied to a larger, preclinical, dataset involving B2905 melanoma cell line implanted in syngeneic C57BL/6 mice, treated with anti-CTLA-4 immune checkpoint inhibitor.

Method



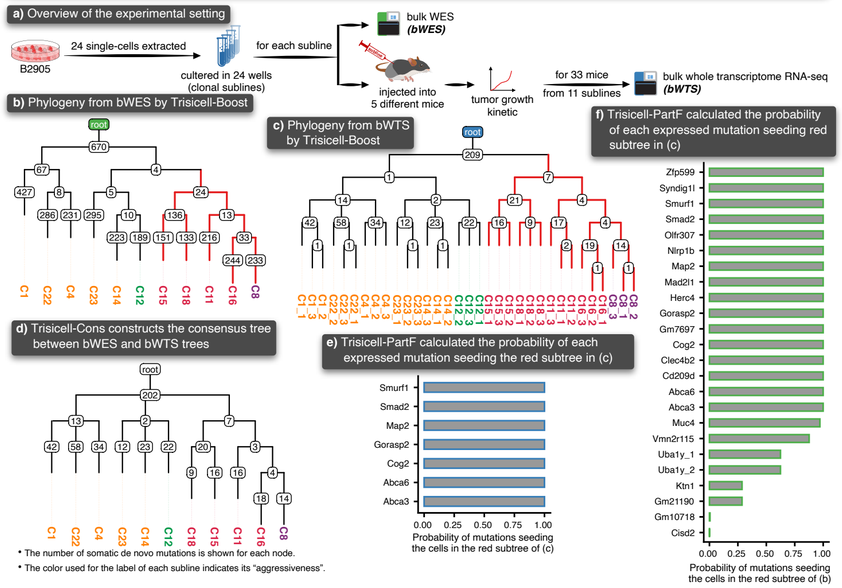
Trisicell is comprised of three computational methods of independent but complementary aims and applications:

1. Trisicell-PartF employs a localized sampling strategy to compute the probability of any user-specified set of cells forming a subclone seeded by one or more (again user-specified) mutations.

2. Trisicell-Cons is devised to compare two or more phylogenies (more specifically cell-lineage trees) derived from the same single-cell data, and build their consensus tree by "collapsing" the minimum number of their edges so that the resulting trees are isomorphic.

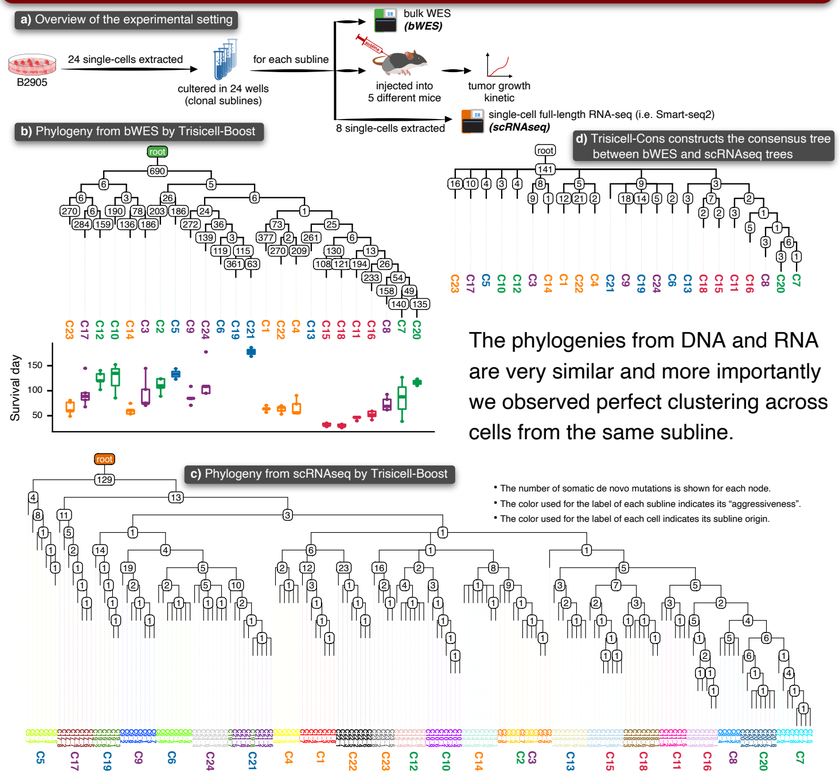
3. Trisicell-Boost is a booster for phylogeny inference methods allowing them to scale up and handle large and noisy scRNAseq datasets.

Validation 1



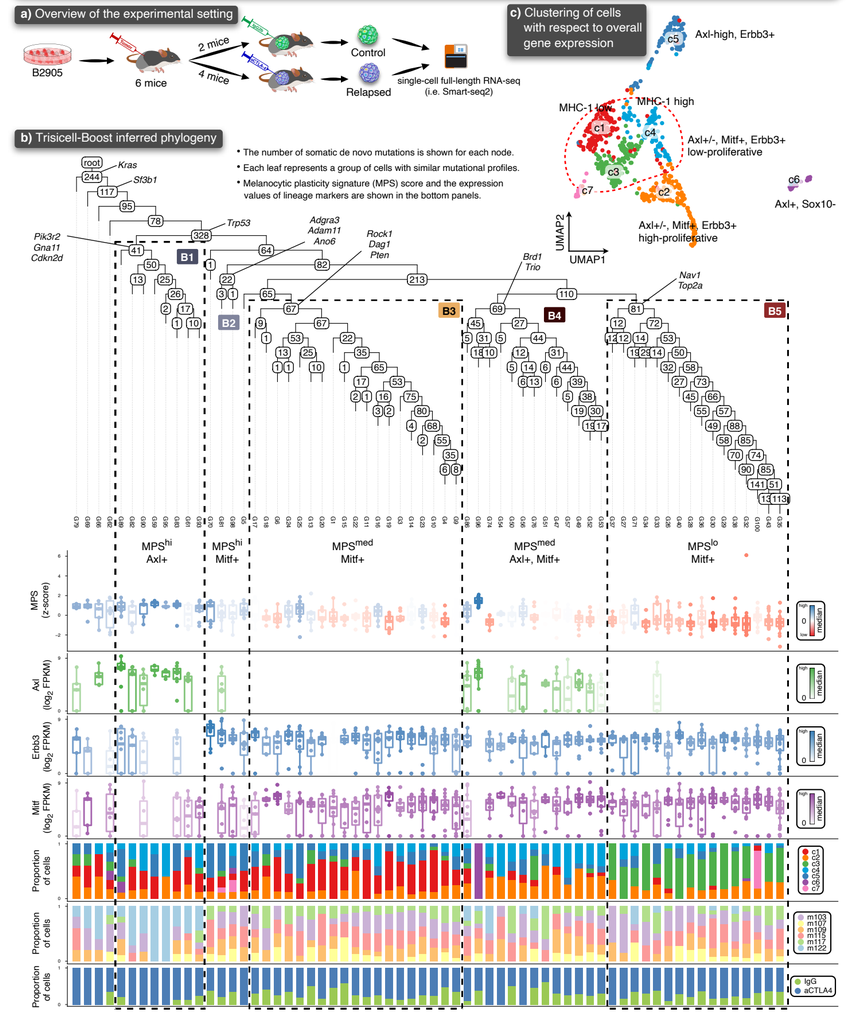
The resulting phylogenies from DNA and RNA are in strong agreement. They cluster the most rapidly growing sublines in vivo in a distinct clade. The probabilities of specific mutations seeding this subclone offer further evidence for the correctness of not only the topology of the inferred phylogeny but also of its mutational placements.

Validation 2



The phylogenies from DNA and RNA are very similar and more importantly we observed perfect clustering across cells from the same subline.

Results



We observed that each subclone with a distinct set of novel somatic mutations is strongly associated with a distinct developmental status. Moreover, each subclone had developed a specific ICB-resistance mechanism. Our new results suggest that subclonal selection may not happen in every tumor responding to therapy. The gene expression clustering implies that the UMAP clusters are only weakly associated with developmental status.

Conclusion

Our results demonstrate the potential of phylogenetic analysis in studying intratumoral heterogeneity and tumor development. This can be achieved through the innovative use of scRNAseq data, which have been primarily used for identifying gene expression differences across distinct cell types. The ability of Trisicell to build large scale, robust tumor phylogenies, to identify and evaluate distinct subtrees and seeding mutations greatly expands possible applications of scRNAseq data including subclonal driver mutation discovery and lineage tracing.

Trisicell is available at https://trisicell.rtfd.io