Integrating Single-Cell and Bulk DNA Sequencing Results for Breast Cancer Subclonal Reconstruction

ELISE STAGAMAN[1], Helena Winata[2], Dan Knight[3], Pedram Razafi[4], Paul Boutros[4,5]

1) BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA , 2) Graduate Programs in Biosciences, UCLA , 3) Department of Human Genetics, David Geffen School of Medicine, UCLA , 4) Memorial Sloan-Kettering Cancer Center, 5) Institute for Urologic Oncology, David Geffen School of Medicine, UCLA

Preparing people to lead extraordinary lives

Introduction

Cancer is caused by an accumulation of mutations, which progressively increase cell proliferation rate as the disease evolves selective advantages[1]. Recent research has demonstrated the impact of mutation timing as a significant driver of clinical outcomes[4]. For this reason, appropriate methods of accurately timing cancer evolution are paramount to understanding disease progression, metastasis, and lethality.

For this study, we obtained multi-sample, bulk tumor DNA sequencing data from a cohort of post-mortem breast cancer patients. Subclonal reconstruction (SRC) analysis was performed to evaluate disease evolution and associated clinical outcomes.

However, SRC relies on statistical assumptions and cannot always accurately infer mutation timing at the cellular level[2]. To mitigate uncertainty, we sequenced mutations of interest using a single-cell panel. This study utilizes one sample from a breast cancer patient to pilot the process of verifying bulk phylogeny with single-cell insight.

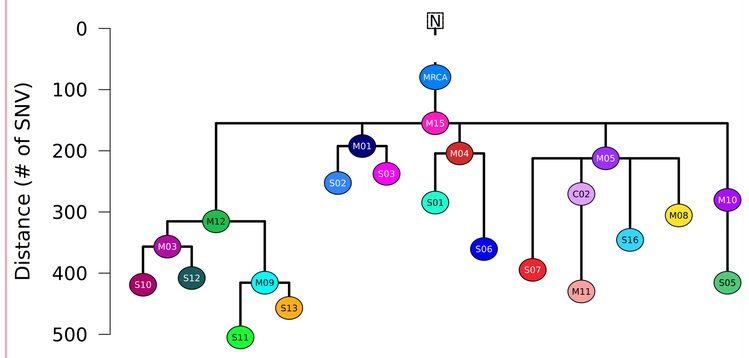


Figure 1: Phylogenetic tree inferred from bulk DNA sequencing

Goals

1. **use single-cell panel to validate bulk phylogeny**
2. **Create 'ground truth' evolutionary tree for benchmarking SRC methods**

Next Steps

1 Correct for Allele Dropouts

Investigate new imputation strategies, directly addressing discrepancies between single-cell and bulk cancer cell fraction (CCF)

2 Comparing Bulk and Single-Cell Phylogeny

Adjust parameters of Likelihood Ratio Test to more accurately predict linear relationships

Methods

Comparing Phylogeny

1 Pair All SNVs

SNV_i	SNV_j
SNV1	SNV2
SNV1	SNV3
SNV1	SNV4

2 Count cells containing each SNV

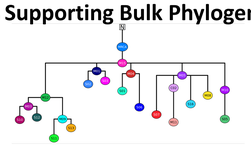
		SNV_j yes	SNV_j no
SNV_i	yes	N_{ij}	N_{i0}
SNV_i	no	N_{0j}	N_{00}

3 Calculate Likelihood of N_{ij}, N_{i0}, N_{0j}, N_{00} given a linear/branched relationship

N_{ij}, N_{i0}, N_{0j}, N_{00} → Observed

Bulk CCF → Expected | linear or branched

4 Do Likelihood Ratio Test + Calculate SNV Pairs Supporting Bulk Phylogeny



2 Methods to Correct for Missing Values in Single-Cell Data

1 KNN Missing Value Imputation

	SNV 1	SNV 2	SNV 3
Cell 1	0	NA	0
Cell 2	NA	NA	0
Cell 3	1	1	1

Find Nearest-Neighbors → Impute

	SNV 1	SNV 2	SNV 3
Cell 1	0	1	0
Cell 2	0	0	0
Cell 3	1	1	1

2 Remove Low-Depth Zeroes

	SNV1	SNV2	SNV3
Cell1	0	NA	0
Cell2	NA	NA	0
Cell3	1	1	1

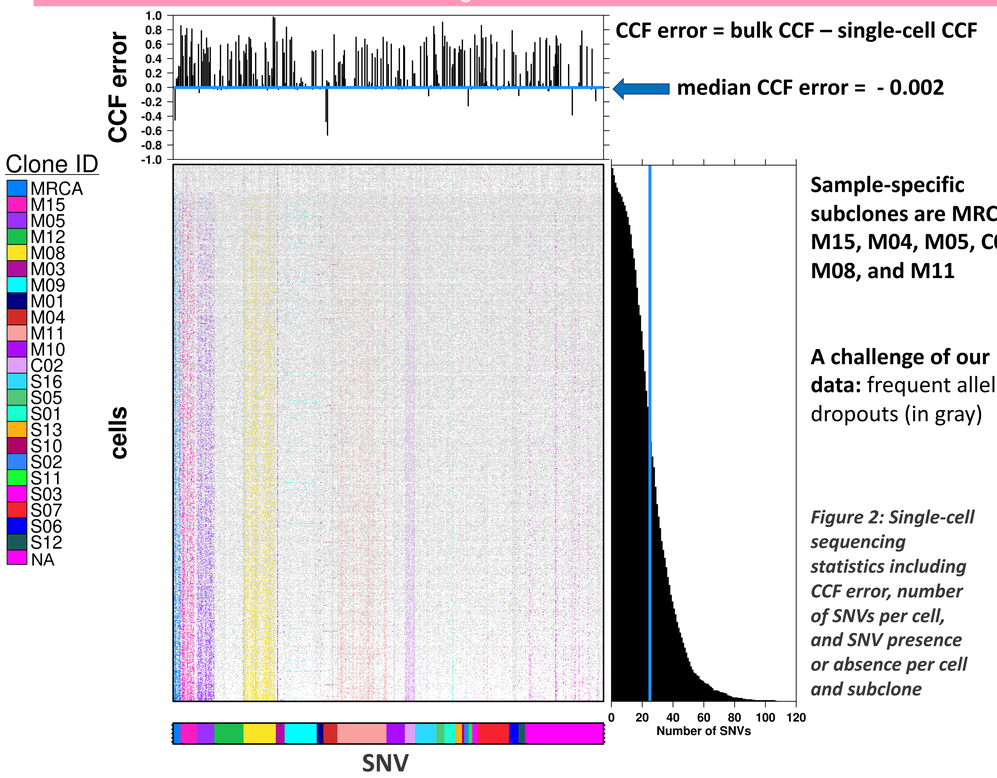
	SNV1	SNV2	SNV3
Cell1	3	NA	1
Cell2	NA	NA	6
Cell3	5	4	1

→

	SNV1	SNV2	SNV3
Cell1	0	NA	NA
Cell2	NA	NA	0
Cell3	1	1	NA

Results

Single-Cell Statistics



CCF error = bulk CCF – single-cell CCF

median CCF error = - 0.002

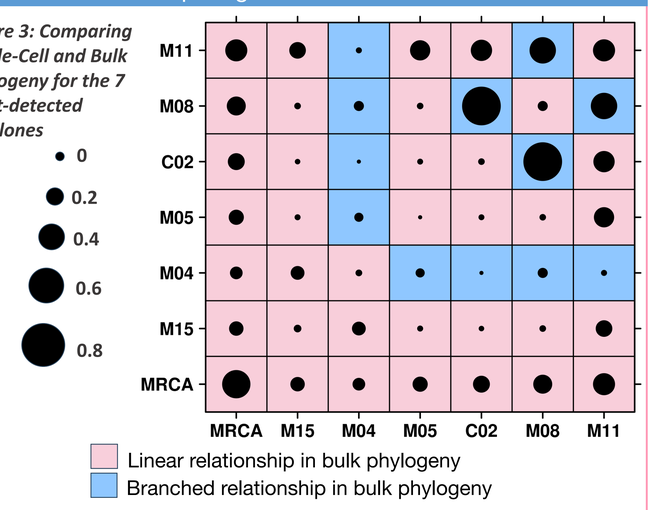
Sample-specific subclones are MRCA, M15, M04, M05, C02, M08, and M11

A challenge of our data: frequent allele dropouts (in gray)

Figure 2: Single-cell sequencing statistics including CCF error, number of SNVs per cell, and SNV presence or absence per cell and subclone

Comparing Branched vs Linear

Figure 3: Comparing Single-Cell and Bulk Phylogeny for the 7 most-detected subclones



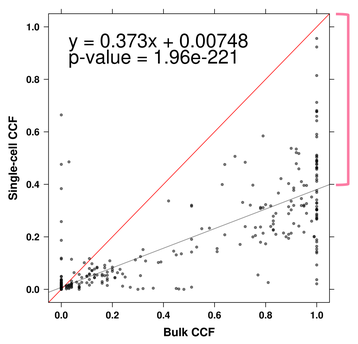
- **Dot size:** percentage of SNV pairs from the 2 subclones indicated that match expected relationship from bulk phylogeny
- Some relationships, such as branching between M08 and C02, were **predicted** by the bulk phylogeny and **supported** by single-cell data
- Some linear relationships, such as C02 and M05, have **little support from single-cell data**

Comparing Single-Cell and Bulk:

CCF:

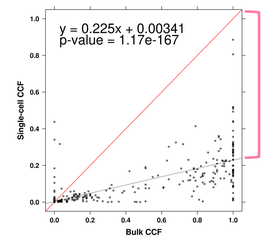
Cancer Cell Fraction

Mutations in the single-cell data show overall lower frequency than in bulk data.



2 Correction Approaches

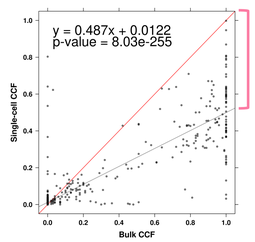
1 Imputation



r = 0.788 MSS = 0.088

Over-imputation of zeroes

2 Filter by Depth



r = 0.879 MSS = 0.043

Elimination of unsupported zeroes

References

1. Nik-Zainal, S., Van Loo, P., Wedge, D. C., Alexandrov, L. B., Greenman, C. D., Lau, K. W., Raine, K., Jones, D., Marshall, J., Ramakrishna, M., Shlien, A., Cooke, S. L., Hinton, J., Menzies, A., Stebbings, L. A., Leroy, C., Jia, M., Rance, R., Mudie, L. J., Gamble, S. J., ... Breast Cancer Working Group of the International Cancer Genome Consortium (2012). The life history of 21 breast cancers. *Cell*, *149*(5), 994–1007. https://doi.org/10.1016/j.cell.2012.04.023

2. Roth, A., McPherson, A., Laks, E. *et al.* Clonal genotype and population structure inference from single-cell tumor sequencing. *Nat Methods* **13**, 573–576 (2016). https://doi.org/10.1038/nmeth.3867

3. Tarabichi, M., Salcedo, A., Deshwar, A.G. *et al.* A practical guide to cancer subclonal reconstruction from DNA sequencing. *Nat Methods* **18**, 144–155 (2021). https://doi.org/10.1038/s41592-020-01013-2

4. Yates, L. R., Knappskog, S., Wedge, D., Farmery, J. H. R., Gonzalez, S., Martincorena, I., Alexandrov, L. B., Van Loo, P., Haugland, H. K., Lilleng, P. K., Gundem, G., Gerstung, M., Pappaemmanuil, E., Gazinska, P., Bhosle, S. G., Jones, D., Raine, K., Mudie, L., Latimer, C., Sawyer, E., ... Campbell, P. J. (2017). Genomic Evolution of Breast Cancer Metastasis and Relapse. *Cancer cell*, *32*(2), 169–184.e7. https://doi.org/10.1016/j.ccell.2017.07.005