Reconstructing the evolutionary dynamics of liposarcoma

Corey Weistuch[1], Rodrigo Gularte Mérida[2], Timour Baslan[3], Nicholas Socci[4], Joseph Deasy[1], Allen Tannenbaum[5], and Samuel Singer[2]

***Abstract*— We have developed a general algorithm that reconstructs the accumulation of DNA copy number alterations (amplifications/deletions) within individual tumors using statistical information gathered from many constituent cells. Such a tool, while general, has the potential to profoundly impact future treatment by revealing the driving events of liposarcoma, a type of cancer characterized by complex copy number alterations and poor patient outcomes. To manage this complexity, we have employed Maximum entropy, a physics-inspired modeling approach widely-used for the identification of complex biological patterns. Integrating this approach with single-cell copy number data, we elucidate the evolutionary distinction between the two primary types of liposarcoma: well-differentiated and dedifferentiated (WD/DDLS).**

Cancer cells are highly variable even within individual tumors. Liposarcomas, for example, are characterized by extensive variability in DNA copy number (CNA) that renders them highly treatment-resistant [1]. What is needed is a framework for reconstructing the accumulation process of these CNA and identifying the critical events that lead to disease progression. Here we provide this framework through the principle of maximum entropy [2], [3], a general and data-driven modeling strategy (Figure 1).

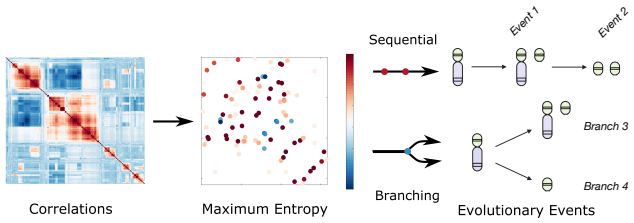

Figure 1. Maximum entropy identifies sparse event signatures from the observed CNA correlations in the data. Positive signatures reveal sequential events while negative ones reveal branching events.

This work was funded by the the Marie-Josée Kravis Fellowship in Quantitative Biology (CW), Breast Cancer Research Foundation (JD,AT), and the NCI SPORE grant in Soft Tissue Sarcoma P50 CA140146 (SS).

[1]Department of Medical Physics, Memorial Sloan Kettering Cancer Center. Email: weistucc@mskcc.org

[2]Department of Surgery, Memorial Sloan Kettering Cancer Center.

[3]Cancer Genetics and Biology Program, Memorial Sloan Kettering Cancer Center.

[4]Marie-Josée & Henry Kravis Center for Molecular Oncology, Memorial Sloan Kettering Cancer Center.

[5] Departments of Computer Science and Applied Mathematics & Statistics, Stony Brook University.

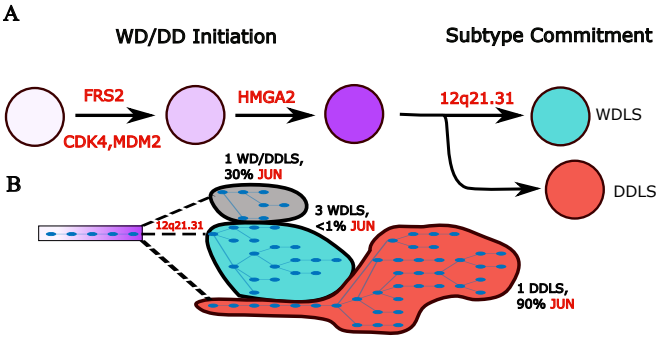

Figure 2. Maximum entropy reconstruction of liposarcoma progression aggregated across several tumors. A.) Simplified diagram of predicted progression events. B.) Detailed map of inferred initiating alterations and subtype-specific marker genes.

To resolve the temporal order of CNA events in liposarcoma, we designed a three stage computational pipeline. First, single-cell copy number estimation is performed using 20000 variable length bins covering the whole genome. Next, maximum entropy is used to rank the order alteration events from CNA frequencies and their covariances [3]. Finally, a lineage tree is constructed from the proposed events in rank-descending order.

Using our pipeline, we reconstructed the individual event histories of 5 distinct WD/DDLS liposarcomas, the two most common subtypes, using single-cell ($N \approx 1400$ cells each) CNA obtained from each tumor. By aggregating the CNA data from each of these tumors, our model resolves early liposarcoma initiation and progression events (Figure 2). In summary, our maximum entropy framework provides a fast, detailed, and automatic reconstruction of the event history of single-cell copy number alterations.

References

[1] A. J. Lazar *et al.*, "Comprehensive and integrated genomic characterization of adult soft tissue sarcomas," *Cell*, vol. 171, no. 4, pp. 950–965, 2017.

[2] C. Weistuch, L. Agozzino, L. R. Mujica-Parodi, and K. A. Dill, "Inferring a network from dynamical signals at its nodes," *PLoS computational biology*, vol. 16, no. 11, p. e1008435, 2020.

[3] G. Marmier, M. Weigt, and A. F. Bitbol, "Phylogenetic correlations can suffice to infer protein partners from sequences," *PLoS computational biology*, vol. 15, no. 10, p. e1007179, 2019.