

# An Outcome Weighted Learning Approach for Identifying Clinically Relevant Patient Subgroups from Large-scale Sequencing Data

Arshi Arora, Adam B. Olshen, Venkatraman E. Seshan, and Ronglai Shen

## Introduction

- Unsupervised clustering has been widely used to identify cancer molecular subtypes and associated alterations for discovery of therapeutic targets. Survival analysis is typically done post-clustering. Clustering methods with direct integration of survival information are desired to identify clinically relevant subtypes.
- We developed survClust, a supervised integrative clustering algorithm using a weighted distance approach to delineate prognostic subtypes and associated molecular features across multiple omics platforms (Figure1)

## Results

- survClust identifies more distinct survival subtypes than the integrated subtypes from unsupervised clustering in most TCGA cancer types (Table 1).
- Individual platforms show different prognostic values in different cancer types (Figure 2). Integration provides improvement in prognostic stratification in most cancer types, particularly in HNSC.
- In LGG, survClust leads to more refined prognostic stratification and discovery of additional subgroups beyond IDH-mutant, 1p/19q-codeletion.

Table 1

Cancer Type	Unsupervised Logrank	survClust Logrank	Method
HNSC	4.7(P=0.192)	23.6(P<0.001)	PARADIGM
UCEC	16.1(P=0.001)	43.7(P<0.001)	Manual
LGG	64.5(P<0.001)	267.7(P<0.001)	COCC
LIHC	1.71(P=0.425)	25.3(P<0.001)	iCluster
LUAD	9.69(P=0.084)	10(P=0.002)	iCluster
STAD	3.44(P=0.487)	7.3(P=0.029)	Manual

## Discussion

- We developed a weighted distance method for molecular prognostic stratification and demonstrated the performance in TCGA pan-cancer analysis.
- This approach can also be extended to other clinical outcomes, including treatment response and progression-free survival

## References

- Xing, E. P., Jordan, M. I., Russell, S. J., & Ng, A. Y. (2003). Distance metric learning with application to clustering with side-information. In *Advances in neural information processing systems* (pp. 521-528).
- Hoadley, K. A., Yau, C., Hinoue, T., Wolf, D. M., Lazar, A. J., Drill, E., Shen, R., ... & Akbani, R. (2018). Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. *Cell*, 173(2), 291-304.

