

# A Visual Data Mining System for Subcellular Location-specific Knowledge Mining from Multi-Omics Data in Cancer

Minwoo Pak<sup>1</sup>, Dabin Jeong<sup>2</sup>, Sungjoon Park<sup>1</sup>, Jeonghyeon Gu<sup>3</sup> and Sun Kim<sup>1,2,4\*</sup>

<sup>1</sup> Department of Computer Science and Engineering Seoul National University, Seoul, Republic of Korea.

<sup>2</sup> Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul, Republic of Korea.

<sup>3</sup> Interdisciplinary Program in Artificial Intelligence, Seoul National University, Seoul, Republic of Korea.

<sup>4</sup> Institute of Engineering Research Seoul National University, Seoul, Republic of Korea.

\* Correspondence: [sunkim.bioinfo@snu.ac.kr](mailto:sunkim.bioinfo@snu.ac.kr)

Cancer dynamics is highly complex that data from a single layer of omics does not give the complete information to study a phenotypic characteristics of a cancer type. While transcriptomic and proteomic abundance carry crucial information of cell states, studies have found that the two omics often do not agree with each other. To better capture the biological context of interest, it is imperative to comparatively analyze them in the context of protein localization and biological pathway. Although there are many data analysis tools for gene expression or protein quantification data, none of them allows researchers to compare different abundance data in the context of protein localization.

In this study, we present ALPACA (A Location-wise Proteome/transcriptome Abundance Comparative Analyzer) a visual data mining system that comparatively analyzes transcriptomic and proteomic abundance data of different cancers in location-wise and biological pathway-specific way. Our system compartmentalizes the whole transcriptome and proteome abundance and visually presents the discrepancies of different cancers using subcellular locations and biological pathways as a filtering and sorting mechanisms. Such filtering and sorting adds biological context of interest to the data analysis to aid the identification of potential biomarkers. ALPACA enables researchers explore the vast search space of numerous cancer types and pathway combinations with little effort and time.

We demonstrate the utility of our system through cases studies of various scenarios. Guided by the quantitative difference metrics that ALPACA provides, we were able to efficiently search through different combinations of cancers and pathways to narrow down on the potential key proteins that can help explain the manifestation of a phenotypic difference between two cancers.