M1430-02-12

Integrating Omics Data Through AI: Predicting Novel Therapeutic Targets for Therapy-Resistant Cancer Patients

Corinne Ramos, Mélodie Boute, Sandra Delebecq, Victor Sénéchal , Mathieu Gaudin

Aliri France SAS, 152 Rue du Dr Yersin, 59120 Loos, France

info@aliribio.com

PURPOSE

Artificial intelligence (AI) is transforming biomedicine by facilitating the thorough analysis of multi-omics data, thereby deepening our grasp of intricate biological systems and the underlying mechanisms of diseases. Our study leverages AI to synthesize diverse omics datasets—including genomics, proteomics, and metabolomics—with the goal of identifying novel therapeutic targets for cancer patients resistant to current treatments. This strategy is designed to enhance the development of personalized medicine and refine treatment approaches, offering new avenues for addressing complex medical challenges.

METHOD(S)

The dataset includes 22 individuals with cervical cancer, each providing a tissue sample. The tissues vary in tumor content and originate from different organs. Measurements were taken on manually selected regions of interest (ROIs). There are two categories: tumor and interface, with the interface referring to the boundary between the tumor and surrounding tissue, classified by specific fluorescent markers. Data distribution is summarized in Table 1. Key clinical data include a patient identifier, a code assigned to each tissue, treatment response (labeled R for Responder and NR for Non-Responder), and the type of tissue from which the biopsy was taken.

Three data modalities were provided:

Proteomics: Contains 5 biomarkers (proteins) with their concentration per square millimeter for each patient.

Transcriptomics: Includes 11,202 genes and their expression levels for each ROI.

Metabolomics: Measures metabolites using mass spectrometry. Data files list mass-to-charge (m/z) ratios and their average intensities. Separate files are available for positive (2,254 m/z) and negative ions (2,548 m/z).

The project involves preparing data through steps such as organizing raw files into a matrix format, ensuring consistency, and normalizing values from different sources to avoid biases. Using a vertical integration approach, DIABLO (with sPLS-DA) is employed to identify minimal molecular signatures for treatment response, focusing on specific "Tumor" and "Interface" ROIs for enhanced spatial resolution.

	Tumor	Interface
ROIs numbers	44	50
Patients numbers	21	17

Table 1. Number of ROIs and patients in the Tumor and Interface categories.

RESULT(S)

Class Separation

The first analysis visualizes each ROI projected by the algorithm onto two components, grouped according to their category: Responder (R) in orange or Non-Responder (NR) in blue. This scatter plot allows for determining which type of data has the highest discriminative power for the task based on the information extracted by the algorithm. This way, a preliminary assessment of the analysis results and the contribution of each modality to the overall distinction between the two groups can be made. Figure 1 shows the four modalities projected onto the two calculated components. Each ellipse contains at least 95% of the samples according to their group affiliation. The more separated the ellipses are, the more distinguishable the data is between each group. Their size indicates the variability present within the samples. A large ellipse signifies that the variance is high between individuals in this space.

Identification of Variables of Interest

Variables of interest are identified based on their contribution to each component, represented by coefficients indicating their weight and importance. The loading plot visualizes this, where bar length shows the significance of a variable, with longer bars indicating higher importance. Bars are ordered from bottom to top according to their relevance in distinguishing classes. The bar color indicates the group in which the variable is more expressed. The coefficient's sign, shown by the bar direction, represents the contribution type left for an inverse relationship and right for a positive one. Figure 2 shows the genes, metabolites and proteins that contribute the most to the two components in tumor ROIs. IDO1 and FoxP3 are the most dominants genes to distinguish the two classes with FoxP3 being the most dominants genes in the Non-Responder group.

Interaction between omics modalities

The relationships between variables are visualized using Clustered Image Maps (CIM), which show molecular signatures for each sample. Low expression is shown in blue, high in red. Variables are in columns, categorized by modality, and samples are in rows, marked by their treatment response.

The created models underwent cross-validation, where samples were tested to compare the model's predictions against actual treatment responses. Although the resulting error rates were acceptable, they could be further optimized. Nonetheless, for both tumor and interface regions, certain variables proved to be significant across different modalities in distinguishing Responders from Non-Responders.

```
$centroids.dist
       comp1      comp2
NR 0.03548387 0.06451613
R  0.50000000 0.36153846
Overall.ER  0.17272727 0.13227273
Overall.BER 0.26774194 0.21302730
```

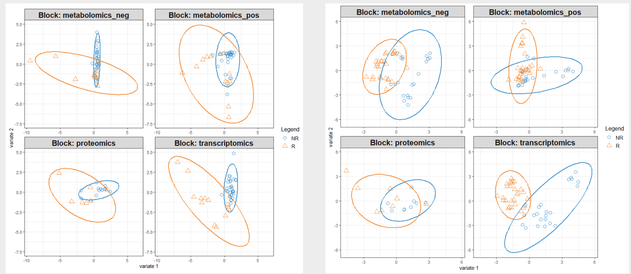

Figure 1. Scatter plot representing samples from the Responder (orange) and Non-Responder (blue) groups with 95% confidence ellipses for each modality block (negative metabolomics, positive metabolomics, proteomics, and transcriptomics) of Tumor and Interface ROIs. The projection axes correspond to components 1 and 2 on the x and y axes, respectively.

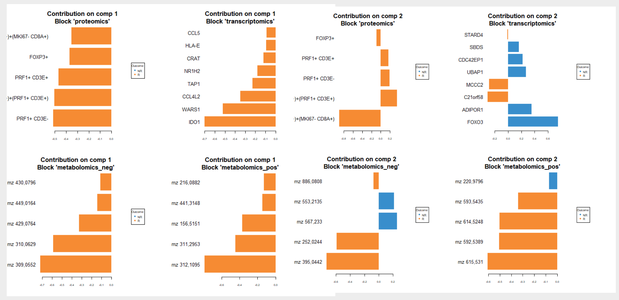

Figure 2. Plot of the contributions of the most relevant variables for components 1 (left) and 2 (right) for the transcriptomic modality of the Tumor ROIs. The variables represented here are genes, proteins and metabolites. The x-axis shows the weight coefficient assigned to each component.

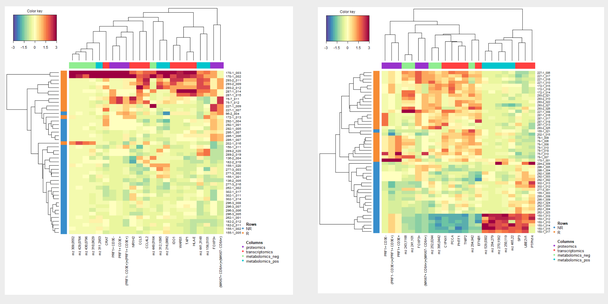

Figure 3. Clustered Image Map showing the expression of each variable based on samples grouped by similarity for the Tumor and Interface ROIs.

CONCLUSION(S)

Our study underscores the substantial impact of integrating multi-omics data with AI analysis in uncovering the complex resistance mechanisms of cancer to immunotherapies, identifying promising new drug targets. By systematically examining metabolomics, proteomics, and transcriptomics datasets from biopsy samples of therapy-resistant cancer patients, we identified significant metabolic disruptions, alterations in protein regulation, and gene expression changes that are crucial for the persistence and survival of these resistant cells. These insights can enable us to identify novel drug targets within these resistance pathways, focusing on critical and previously uncharted mechanisms.

REFERENCE

Argelaguet, R., Velten, B., Arnol, D., Dietrich, S., Zenz, T., Marioni, J. C., Buettner, F., Huber, W., & Stegle, O. (2018). Multi-Omics Factor Analysis-a framework for unsupervised integration of multi-omics data sets. *Molecular Systems Biology*, *14*(6), e8124. https://doi.org/10.15252/msb.20178124.

Ashhurst, T. M., Marsh-Wakefield, F., Putri, G. H., Spiteri, A. G., Shinko, D., Read, M. N., Smith, A. L., & King, N. J. C. (2022). Integration, exploration, and analysis of high-dimensional single-cell cytometry data using Spectre. *Cytometry Part A*, *101*(3), 237-253. https://doi.org/10.1002/cyto.a.24350.

Athieniti, E., & Spyrou, G. M. (2022). A guide to multi-omics data collection and integration for translational medicine. *Computational and Structural Biotechnology Journal*, *21*, 134-149. https://doi.org/10.1016/j.csbj.2022.11.050.

Subramanian, I., Verma, S., Kumar, S., Jere, A., & Anamika, K. (2020). Multi-omics Data Integration, Interpretation, and Its Application. *Bioinformatics and Biology Insights*, *14*, 1177932219899051. https://doi.org/10.1177/1177932219899051.