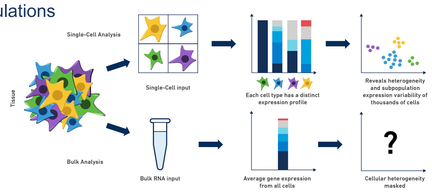
Survival Analysis of Glioblastoma Patients Through Tumor Tissue Deconvolution

Stephen Hwang[1], Juejue Wang[2], Yujie Wei[3]

1. University of California, Santa Cruz, 2. University of Michigan, 3. Wellesley College

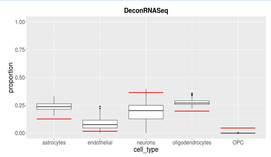
Background

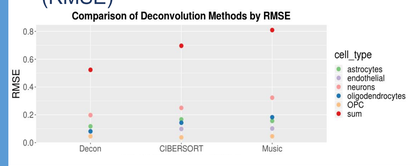
- Glioblastoma multiforme (GBM) is the most aggressive type of brain cancer with a median survival time of 15-16 months and 14k cases in the U.S. each year
- Traditional treatments (surgery, radiation, and chemotherapy) and on-going research concentrate on understanding and estimating patient survival
- Immune cell presence in tumors has been found to be associated with patient survival time
- Single Cell RNA Sequencing:
  - Provides the gene expression profile of individual cells
  - Reveals expression variability of different cell types and subpopulations
- Bulk RNA Sequencing:
  - Provides the average gene expression level of all cells in a tissue; masks the cell type heterogeneity in the tissue



Deconvolution Methods

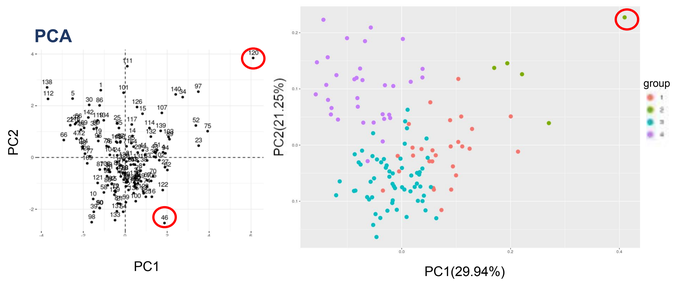
- General outline for deconvolution:
  - Input:
    - Gene signatures (single cell): Expression level of m genes in k cells
    - Mixture (bulk): Expression level of m genes in n tissue samples
  - Output:
    - Cell type composition in the n tissue samples
- Compare deconvolution methods using boxplots and root mean square error (RMSE)





Data Processing

- Remove outliers in the deconvolution result using Principal Component Analysis and by number of non-expressed cell types



Survival Analysis

- K-means clustering:
  - Use log rank and elbow method to decide the optimal number of clusters
  - Use t-SNE dimensionality reduction to visualize clusters in two-dimensional space
- Survival Analysis:
  - Plot the Kaplan-Meier curve for each cluster to get the p-value for log rank test, which compares the survival curve for different clusters
- Fit Cox proportional hazards model to identify statistically significant variables
  - Individual tests on group, B-cell, CD4+ T-cell, CD8+ T-cell, neutrophil, macrophage, dendritic cell (DC) to find which variables are statistically significant for predicting patient survival time

Cox proportional hazards model:

$$\ln\left(\frac{H(t)}{H_0(t)}\right) = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_m X_m$$

Where $H(t)$ is the hazard, $H_0(t)$ is the baseline hazard at time t, which represents the hazard for a person with value 0 for all the predictor variables. $X_1 \ldots X_m$ are the set of covariates, and the beta coefficients measure their respective effect size.This model allows us to examine the effect of different factors on survival time simultaneously.

- Final model fitted with CD8+ T-cells and macrophages

$$\ln\left(\frac{H(t)}{H_0(t)}\right) = 2.19\ \text{T\_cell.CD8} + 1.19\text{Macrophage}$$

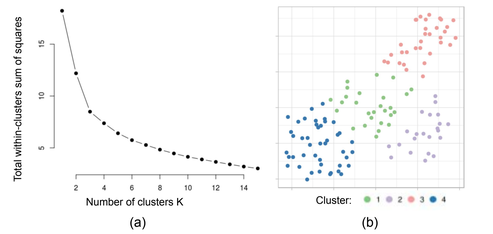


Figure 1: (a) Within-clusters sum of squares for 2 to 14 clusters. (b) Visualization of 4 clusters after t-SNE dimensionality reduction and k-means clustering.

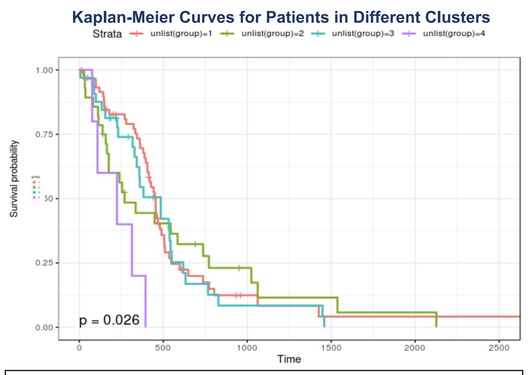


Figure 2: Kaplan-Meier Curves for different clusters of glioblastoma patients. Patients with similar tumor tissue immune cell type compositions are grouped into the same cluster represented by a unique color. A p-value of 0.026 for the log rank test comparison of different Kaplan-Meier curves indicates a significant difference among the survival curves.

Summary of Beta Coefficients for Different Cell Types

Variable	B_cell	T_cell.CD4	T_cell.CD8	Neutrophil	Macrophage	DC
beta	-0.57	0.32	2.7	1.7	2.5	0.44
Hazard ratio (95% CI)	0.57 (0.19-1.7)	1.4 (0.41-4.6)	15 (2.8-86)	5.3 (0.38-76)	13 (1.7-94)	1.6 (0.69-3.5)
p - value	0.31	0.6	0.0018	0.22	0.012	0.28

Comparison of Beta Coefficient for Different Cell Types

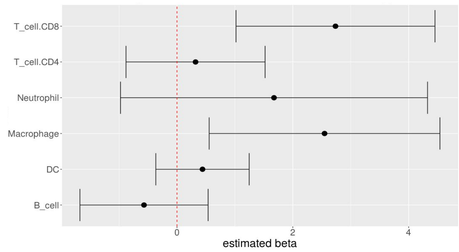


Figure 3: 95% confidence intervals of beta coefficients obtained from univariate Cox proportional hazards model on different immune cell types. The red dash line is 0. A negative beta indicates positive effect on survival time; a positive beta indicates negative effects on survival time.

Results

- A p-value of 0.026 for the log rank test to compare different Kaplan-Meier Curves indicates a significant difference between the survival curves.
- The p-values for the likelihood ratio test, wald test, and score test for the final model are all significant, indicating that the final model is significant. CD8+ T-cells is still significant in the final model while Macrophage fails to be significant in the presence of CD8+ T-cells in the model.
- The hazard ratios of covariates are interpretable as multiplicative effects on the hazard. The hazard ratios for CD8+ T-cells (8.939) and macrophages (3.288) indicate a strong relationship between a higher proportion of these cell types in the patient's tumor tissue and shortened survival time. Holding the other covariates constant, a higher proportion of CD8+ T-cells and macrophages in the patient's tumor tissue is associated with shorter survival time.

Discussion

- Our current results are based on deconvolution output from TIMER due to technical difficulties we incurred while trying to run CIBERSORT and DeconRNASeq. TIMER's distinct output format impedes comparison with the other deconvolution methods, so its accuracy could not be determined
- A possible explanation for inconsistencies with existing literature: More CD8+ T-cells may be sent to attack tumor cells, but their effect to prolong patient survival time may be limited for GBM patients that are severely ill
- Future work:
  - Explore the individual and combined effects of different T-cell subtypes on GBM patient survival time
  - Run CIBERSORT using the phenotype matrix with combined gene expression level of CD4+ T-cells and CD8+ T-cells

Comparison of Beta Coefficient for T_cell.CD8 and Macrophage

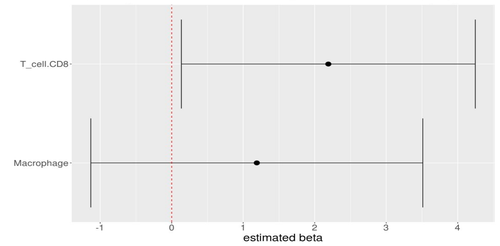


Figure 4: 95% confidence intervals of beta coefficients for CD8+ T-cell and Macrophage obtained from final model. The red dash line is 0. A negative beta indicates positive effect on survival time; a positive beta indicates negative effects on survival time

References:

Azizi, Elham, et al. "Single-cell map of diverse immune phenotypes in the breast tumor microenvironment." Cell 174.5 (2018): 1293-1308.

Gong, Ting, and Joseph D. Szustakowski. "DeconRNASeq: a statistical framework for deconvolution of heterogeneous tissue samples based on mRNA-Seq data." Bioinformatics 29.8 (2013): 1083-1085.

Li, Bo, et al. "Comprehensive analyses of tumor immunity: implications for cancer immunotherapy." Genome biology 17.1 (2016): 174.

Newman, Aaron M., et al. "Robust enumeration of cell subsets from tissue expression profiles." Nature methods 12.5 (2015): 453.

Wang, Xuran, et al. "Bulk tissue cell type deconvolution with multi-subject single-cell expression reference." Nature communications 10.1 (2019): 380.

Acknowledgements:

Xiang Zhou, Ph.D., Matthew Zawistowski, Ph.D., Shiquan Sun, Ph.D., Sarah Gagliano Taliun, Ph.D., Brooke Wolford, Ph.D. candidate, Bhramar Mukherjee, Ph.D., Dan Barker, Big Data Summer Institute faculty and staff

GitHub: