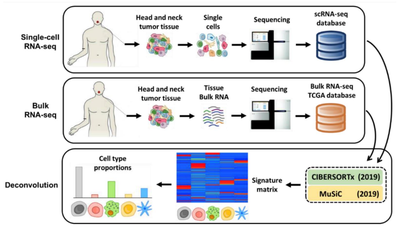
Insights into Breast Cancer from deconvolution of bulk RNA Sequencing Data

Yves Greatti[1], Reem Shamma[1], Lingzhu Shen[1]
[1]Johns Hopkins University, Department of Biomedical Engineering

Abstract

Deconvolution of bulk RNA sequencing data uncovers heterogeneity in cell-type composition and subtypes across breast cancer patients, shedding light on potential molecular mechanisms of cancer development and progression.



Introduction

- Bulk RNA sequencing cannot distinguish between different cell types or states, but single-cell RNA sequencing (scRNA-seq) can identify and profile the transcriptome of individual cells.
- Cancer cells produce cytokines and chemokines that attract immune cells to the tumor microenvironment (TME). Tumor-infiltrating immune cells may physically destroy tumor cells, improving clinical prognosis, but persistent activation of the immune system can lead to chronic inflammation, promoting tumor growth.
- Recent studies show that accounting for the heterogeneity of immune cell infiltration can result in more accurate tumor subtype predictions and survival analyses.
- Breast cancer (BRCA) is highly variable in prognosis, and the TME's heterogeneity can significantly impact patient outcomes.
- We identify the cells present within the TME of breast cancer: immune cells (monocytes, T-cells), myeloid cells (macrophages, monocytes), hormone sensing, basal, alveolar, fibroblast and vascular and lymphatic cells.
- We build and fine-tune a decision-tree regression model to predict the ESTIMATE-immune score of the TCGA cohort from the cell type composition, and reveal the major cell type contributors of the immune response
- From the cell proportions, we analyze the different cancer pathways underregulated or overregulated.
- We assess the impact of specific cell types on the overall survival (OS) and disease-free survival (DFS) of cancer patients.

Materials and Methods

- TCGA-BRCA project consists of data from 1,111 cancer patients and 113 disease-free control patients. Data were collected using TCGAbiolinks and TCGAWorkflow packages.
- Two scRNA-seq datasets:
  - GSE176078[1]: 52 patients, 430,000 cells and different breast cancer subtypes (4 TNBCs, 4 BRCA1 TNBCs, 6 HER+ tumors)
  - GSE161529[2]: 26 primary tumors from three major subtypes of breast cancer (11 ER+, 5 HER2+, and 10 TNBC), 9 major cell types, 49 cell subtypes identified and normal tissue
- Estimate the proportions of different immune cell subpopulations:
  - MUlti-Subject SIngle Cell (MUSiC2)[3]: use scRNA-seq data to deconvolute the bulk RNA seq data
- Determine the impact of various immune cell types on the immune response
  - XGBoost using ESTIMATE immune score (75% for train, 25% for test)
- Identify oncogenes from the estimated cell proportions:
  - PROGENy: gene set enrichment analysis (GSEA) and pathway analysis
  - Decoupler: predicts the activity of transcription factors and pathways within a sample population
- Investigate the potential correlation between cellular fractions and clinical outcomes:
  - Kaplan-Meier survival analysis with log-rank test
  - Hazard ratios from Cox proportional hazards regression models

Results

Deconvolution of Immune Cells From RNA-Seq Data

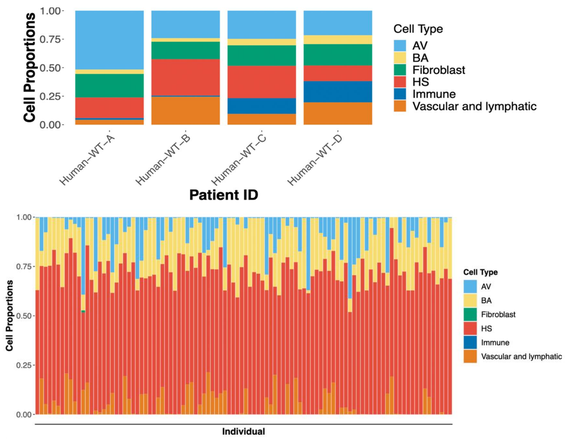


Figure 1: Cell type proportions from single-cell deconvolution using MUSiC2 of normal patients. **Fig 1A.** Cell type proportions from four individuals from the GSE1611529 dataset (top). **Fig 1B.** Cell types of normal patients in our TCGA BRCA cohort deconvoluted using GSE1611529 (bottom).

The disease-free control patients in our cohort had high proportions of HS, AV, and vascular and lymphatic cells, which is consistent with what was observed in normal GSE1611529 patients

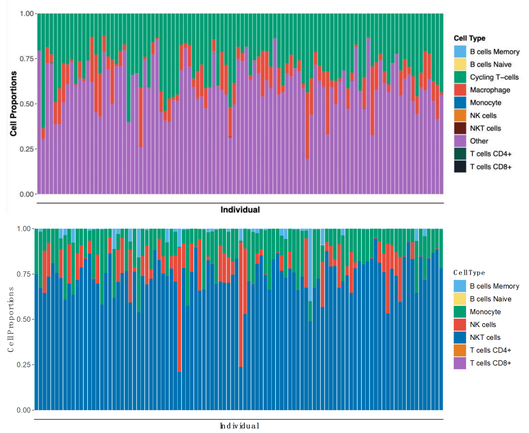


Figure 1: Cell type proportions from single-cell deconvolution using MUSiC2 of tumor patients. **Fig 2A.** Cell type proportions of tumor patients in the TCGA BRA cohort before excluding overrepresented cell types (top). **Fig 2B.** Cell type proportions in TCGA BRA cohort after excluding overrepresented cell types as stacked bar plots (middle) and violin plot (bottom).

In GSE17078 tumor patients, a significant presence of immune cells, such as macrophages, monocytes, and T cells (CD4+ and CD8+), was observed, with similar proportions found in the deconvolution output of TCGA breast cancer patients. When excluding the most overrepresented cell types, macrophages and non-immune cells, the adjusted ratios allowed us to observe a substantial amount of NK and NKT cells, as well as memory B cells to a lesser extent.

Evaluation of Immune Cells in the Immune Response

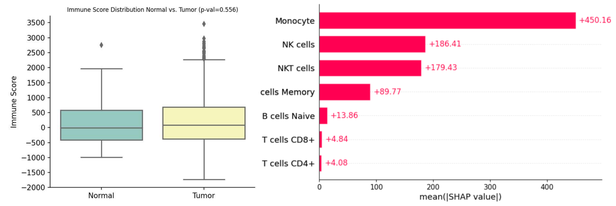


Figure 1A: Immune score distribution across normal and tumor patients. **Figure 1B:** Distribution of mean absolute of SHAP values for each immune cell type.

Cell Type	Coefficient	P-value
B Cell Memory	2.773	0.728
B cells Naive	20.458	0.167
T cells CD8+	-106.369	< 0.001
T cells CD4+	42.236	0.230
NK cells	25.289	0.007
NKT cells	-744.481	< 0.001
Monocytes	1155.076	< 0.001

Table 1: Model coefficients and p-values.

Immune scores quantitatively measure the impact of immune cells on the immune response. Decision-tree based regression model is used to predict the ESTIMATE immune score of the tumor patients. Higher proportions of CD8+ T cells and NKT cells reduce the immune score, indicating that they contribute less to immune response, while higher proportions of monocytes and NK cells increase immune scores and have a higher impact on immune response. Additionally, we utilized Shapley values to estimate the contribution of each cell type to the immune system. Our findings revealed that monocytes, NK, and NKT cells were the most impactful contributors to the immune score among tumor patients.

Pathway Analysis across cell types in normal and BRCA patients

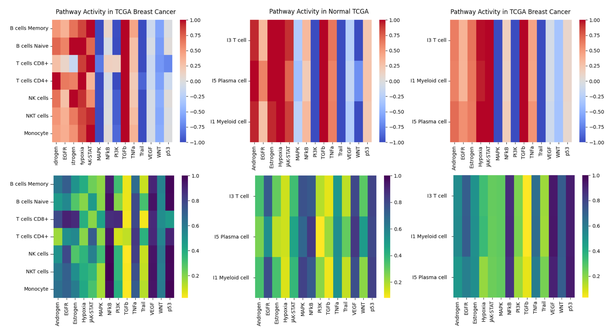


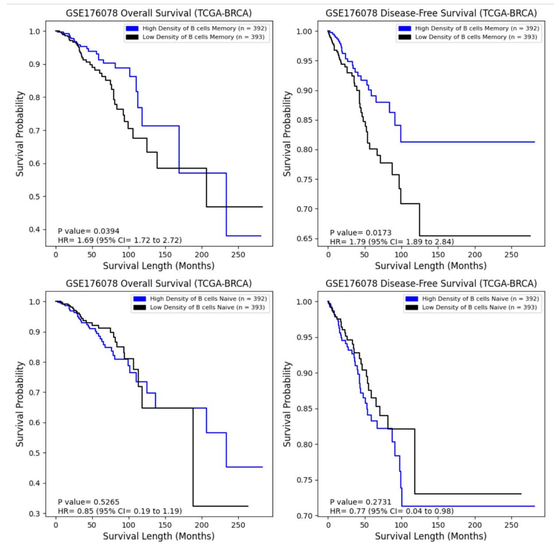
Figure 7: Pathway inference performed on the TCGA BRCA cohorts using the results obtained from deconvolution analysis (top). Immune cell activity levels across different pathways for TCGA normal and tumor patients (middle and bottom).

Most immune cells have high activity in transforming growth factor-β (TGF-β). Lower activity in pathways that induce apoptosis, such as TRAIL (Tumor necrosis factor (TNF)-related apoptosis-inducing ligand), and MAPK pathway. Potential biomarkers for breast cancer immunotherapy: ID1, ID3, COM, PMEPA1, SMAD7 (TGF-β) and RHEBL1, SMIM3, GPR18, RAB37, RNF175 (TRAIL)

Results

Correlation between cell type and clinical outcomes

Tumor patients with high levels of memory B cells had significantly lower OS and significantly higher DFS when compared to patients with low levels of memory B cells. Significant differences in survival were not observed for varying levels of naïve B cells. High levels of CD8+ T cells showed lower DFS that was near significant (p-val = 0.07). Similarly, high levels of NKT cells had significantly lower DFS, while low levels of NKT cells showed higher OS that was near significant (p-val = 0.07).



Conclusion

Our study utilized a deconvolution algorithm to analyze bulk RNA-seq data from the TCGA BRCA cohort and identified cell-type-specific gene expression signatures associated with patient survival. We observed that HS, BA, and AV cells are the most populous cell types in normal vs. control individuals, while the relative proportion of immune cells increases in BRCA patients. Cycling T-cells and macrophages are the predominant immune cell types in BRCA tumors. Pathway analysis reveals significant upregulation of the TGF-β pathway across nearly all immune cell types. Our survival analysis showed conflicting results, suggesting opposing roles of the immune system in tumor development. Overall, our study provides insights into the heterogeneous population of cell types and subtypes present in breast tumors, informs their potential roles in cancer-related pathways, and highlights the need for further research into the complex interactions between immune cells and cancer cells in the TME.

References

[1]Fan, J., Lyu, Y., Zhang, Q., Wang, X., Li, M., & Xiao, R. (2022). MuSiC2: cell-type deconvolution for multi-condition bulk RNA-seq data. Briefings in bioinformatics, 23(6), bbac430. https://doi.org/10.1093/bib/bbac430
[2]Wu, S. Z., Al-Eryani, G., Roden, D. L., Junankar, S., Harvey, K., Andersson, A., Thennavan, A., Wang, C., Torpy, J. R., Bartonicek, N., Wang, T., Larsson, L., Kaczorowski, D., Weisenfeld, N. I., Uytingco, C. R., Chew, J. G., Bent, Z. W., Chan, C. L., Gnanasambandapillai, V., Dutertre, C. A., ... Swarbrick, A. (2021). A single-cell and spatially resolved atlas of human breast cancers. Nature genetics, 53(9), 1334–1347. https://doi.org/10.1038/s41588-021-00911-1
[3]Pal, B., Chen, Y., Vaillant, F., Capaldo, B. D., Joyce, R., Song, X., Bryant, V. L., Penington, J. S., Di Stefano, L., Tubau Ribera, N., Wilcox, S., Mann, G. B., kConFab, Papenfuss, A. T., Lindeman, G. J., Smyth, G. K., & Visvader, J. E. (2021). A single-cell RNA expression atlas of normal, preneoplastic and tumorigenic states in the human breast. The EMBO journal, 40(11), e107333. https://doi.org/10.15252/embj.2020107333