MULTIVARIATE CORRELATION OF MUTATIONAL SIGNATURE EXPOSURES AND GENE EXPRESSION IN SINGLE-CELL BREAST CANCER

Author: Traian Dobrin
Supervisors: Sara Costa, Ivan Stresec Responsible professor: Joana Gonçalves

1. INTRODUCTION

Cancer is a complex and heterogeneous disease. Understanding this heterogeneity is crucial, as it directly influences treatment response, disease progression, and patient outcomes [2]. One emerging approach to characterizing cancer diversity is through mutational signatures- distinctive patterns of somatic mutations left by different mutagenic processes [1].

2. METHODOLOGY

Data Overview:

- Single-cell RNA-seq and mutational signature exposures from n = 687 cells.
- Mutational signatures inferred via NMF [1], retaining 6 active signatures after filtering.

Gene–Signature Scoring:

Figure 1 shows the gene-signature scoring formula, Figure 2 shows correlations between terms.

- The inner sum ($P_l g,p_j \times U_l p,k_j$) calculates gene g's contribution to canonical component k.
- This is multiplied by the signature loading $V_l s,k_j$ for that component and the correlation of the component.
- We take the absolute value to avoid cancellation of positive and negative contributions.
- Finally, we sum over all canonical components k.

Enrichment analysis

Genes ranked by score were subjected to pre-ranked GSEA in order to find the enriched pathways for each signature.

Pipeline:

Dimensionality reduction, preprocessing statistical correlation, gene-signature scoring and enrichment analysis(Figure 3).

$$S_{g,s} = \sum_k \rho_k \left(\sum_p P_{g,p} \cdot U_{p,k} \right) \cdot V_{s,k}$$

Figure 1. Gene Signature scoring formula

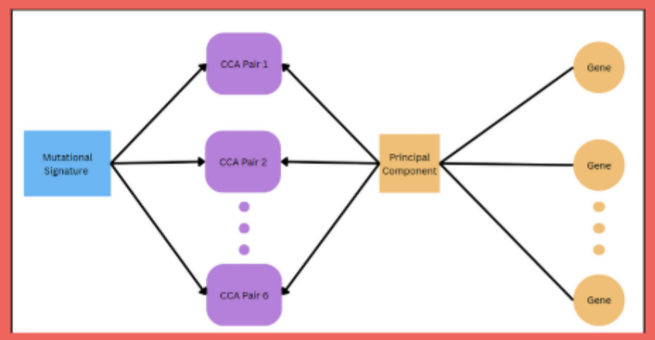

Figure 2. Correlations between terms in scoring formula

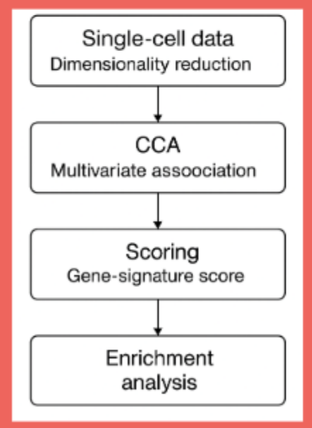

Figure 3. Pipeline description (Figure generated with ChatGPT, OpenAI)

3. RESULTS

Dimensionality Reduction & Association Analysis:

- Applied PCA to gene expression data, retaining 158 components (90% variance).
- Used CCA to identify associations between these components and 6 mutational signatures(Signature contributions in Figure 4).
- Distinct patterns highlight unique gene expression programs for each mutational process (Figure 5).

Gene-Signature Association Scoring:

- Combined PCA and CCA loadings to score gene-signature associations.
- Used absolute values to avoid cancellation, enabling robust gene ranking for each signature.

Figure 4. CCA contributions by mutational signature

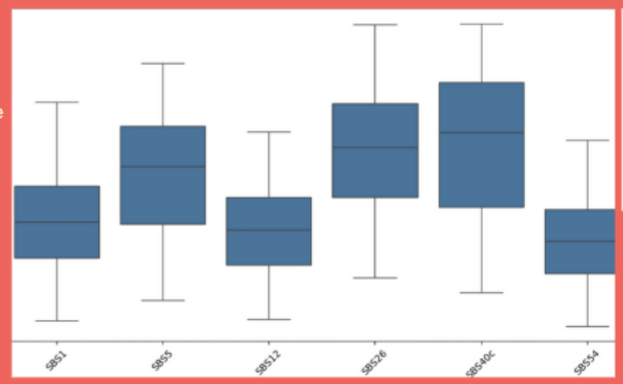

Figure 5. Boxplot of gene score distributions per signature

Signature	Top Enriched Pathways
SBS1	ECM-receptor interaction
SBS5	Focal adhesion, Tuberculosis, SLE
SBS12	ECM-receptor interaction, Focal adhesion, SLE
SBS26	Focal adhesion, ECM-receptor interaction, Protein digestion
SBS40c	Phagosome, SLE, Focal adhesion

Table 1. Enrichment analysis results

Enrichment analysis(Table 1) show: ECM-receptor interaction and focal adhesion pathways are consistently enriched, underscoring their role in tumor progression. Immune-related pathways highlight interactions with the tumor microenvironment, consistent with breast cancer biology [3].

4. CONCLUSIONS & FUTURE WORK

Multivariate analysis linked mutational signatures to distinct gene expression programs, highlighting ECM, adhesion, and immune pathways in breast cancer. Future work will refine gene scoring, include directionality, and expand to larger datasets.

5. REFERENCES

1. Alexandrov, L.B. et al. Signatures of mutational processes in human cancer. Nature, 500(7463), 415–421 (2013).
2. Amgalan, B. et al. Influence network model uncovers relations between biological processes and mutational signatures. Genome Medicine, 15(1), 15 (2023)..
3. Fridman, W.H. et al. The immune contexture in human tumours: impact on clinical outcome. Nature Reviews Cancer, 12(4), 298–306 (2012).