**Homework 4**

Amit Gabay, 206040768

1. Explain what is the tumor immune microenvironment (TIME) and its significance in cancer therapy. (5-10 lines)

The tumor immune microenvironment (TIME) is a complex ecosystem containing adaptive and innate immune cells with tumor-promoting and anti-tumoral roles.  
TIME plays a critical role in cancer development, progression, and control. The molecular and cellular nature of it influences disease outcome by altering the balance of suppressive versus cytotoxic responses in the vicinity of the tumor.  
Knowing classes of TIME helped to identify initiation of tumors, response, and [therapy](https://www.sciencedirect.com/topics/medicine-and-dentistry/therapeutic-procedure). It will allow prediction of [immunotherapy](https://www.sciencedirect.com/topics/medicine-and-dentistry/immunotherapy) response by revealing therapeutic targets.

1. Compare bulk and single-cell RNA sequencing (scRNA-seq). Discuss the advantages and disadvantages of each method, particularly in the context of studying TIME. (8-15 lines)  
   Let’s say you have a group of people with lung cancer. The goal of your research is to understand what’s happening on a molecular level. **With bulk RNA sequencing, you can compare the results of the patients with lung cancer with those that are healthy**. This will provide an overview of the average differences in gene expression. For certain scenarios, this has proven to be sufficient. For instance, when discovering biomarkers for cancer, or to study the biology of diseases. However, more often than not, the answers may lie behind certain cell types. This requires looking at the expression of genes in individual cells instead of an average representation. In that case, **single-cell sequencing provides the potential to find molecular differences which are only linked to specific cell types.**So, although bulk RNA sequencing is less expensive, less labour-intensive and less time-consuming than single-cell sequencing, in some cases one will prefer to use scRNA-seq depends on the specific research questions and desired level of resolution required to unravel the complexities of the cells.
2. Discuss the rationale behind classifying breast cancer (BC) into subtypes. Mention two different approaches for BC subtype classification. (5-15 lines)  
   BS is heterogeneous disease characterized with distinct molecular subtypes and varied clinicopathological features. Subtyping allows for a more refined understanding of this heterogeneity, enabling clinicians to identify specific molecular pathways driving tumor growth and progression. In addition, studies show that different BS subtypes responses in a different way to therapy, so, understanding subtypes behavior is important for planning treatment strategies and predicting clinical outcomes.

Breast cancer subtyping by IHC and gene expression profiling has contributed to a clinically relevant classification.

1. What are immune cell type signatures? What are they used for? Which method measures enrichment using those signatures and how? (5-15 lines)

Immune cell type signatures are patterns of gene expression or protein markers specific to different immune cell types. This patterns serve as biomarkers for disease diagnosis and allows to understand the composition and functional states of immune cells within the tumor microenvironment that is crucial for predicting treatment response, prognosis, and developing immunotherapeutic strategies.

GSEA (gene set enrichment analysis) is a computational method used to determine whether an a priori defined gene set shows a statistically significant difference between two biological states. With GSEA, we can determine whether specific immune cell types are enriched or depleted within a sample compared to control samples and by doing so, we can uncover the involvement of immune cell subsets in diseases, or in treatment responses.

1. In the tutorials, we used a computational tool that calculate immune (and other) cell types enrichment scores. What is that tool? Describe the output it provides and how it can be useful for this study. (5-15 lines)  
   xCell is a computational tool that performs cell type enrichment analysis from gene expression data for 64 immune and stroma cell types. xCell is a gene signatures-based method learned from thousands of pure cell types from various sources. This method applies a novel technique for reducing associations between closely related cell types, particularly from bulk tumor samples which allows researchers to reliably portray the cellular heterogeneity landscape of tissue expression profiles.   
   with this information, researchers can gain insights into the composition of the tumor microenvironment, including the presence of cancer immune cell subtypes, which can be valuable for understanding tumor-immune interactions, predicting patient outcomes, and guiding immunotherapy strategies against the breast cancer.
2. Describe these subtypes (BC-ImH, BC-ImM, and BC-ImL) based on their immune signature enrichment scores. (3-5 lines)  
   3 immune subtypes of BC, termed BC-ImH, BC-ImM, and BC-ImL, was identified consistently in 6 datasets which had high, medium, and low immune signature scores, respectively.  
   BC-ImH displayed a significantly better survival prognosis than BC-ImL.
3. What clustering algorithm was used by the authors to define those BC subtypes? Suggest another clustering algorithm that could be employed in this context. Explain how you would choose the optimal number of clusters. (3-6 lines)  
   Based on the enrichment scores of 28 immune cell types, they **used hierarchical clustering** to eventually identify those 3 subtypes of BC.  
   Another clustering algorithm that could be employed is **K-means clustering**.  
   For choosing the optimal we can use silhouette method. By calculating K silhouette coefficient and select the K with the highest average silhouette coefficient indicating the best separation between clusters while maintaining cohesion within clusters.
4. Explain the difference between disease-free survival (DFS) and overall survival (OS). (3-6 lines)  
   DFS: The time interval between the end of the treatment and the recurrence of the cancer (or progression of it).   
   OS: The time interval between the beginning of treatment (or diagnosis) and patient death (not necessarily from cancer).   
   DFS focuses specifically on disease while OS provides a broader perspective, taking into account not only disease-related outcomes.
5. Describe in detail **Figure 1**. In your answer mention: What type of visualizations did they use? What do the columns, rows and values represent? Explain the different legends and how they are used to interpret the results. (10-20 lines)

The visualizations used in Figure 1 are 5 heatmaps, representing the enrichment scores of 28 immune cell types across five breast cancer transcriptomic datasets (TCGA-BRCA, METABRIC, GSE24450, GSE2034, and GSE11121).  
The columns of each heatmap represent BC tumor samples (each column is a sample).  
The rows represent the cells types, and the values represent the enrichment scores of immune cell types, which indicate the degree of presence of specific immune cell types in the different samples and BC subtypes (blue – low presence, red – high presence).  
The first 8 rows (the legends) of each heatmap depicts the segmentation of the data by specific subtypes of breast cancer: (1)PAM50 subtype (a breast cancer classification system), (2) is the cancer a Triple-Negative Breast Cancer or not (It's a specific subtype of breast cancer), (3) whether a breast cancer tumor expresses hormone receptors or not, (4) HER2 protein status (a protein that can promote the growth of cancer cells when it is overexpressed), (5) Tumor purity (the proportion of cancer cells within the tumor sample), (6) Immune score and (7) stromal score (level of immune infiltration and stromal content within tumor samples), (8) subtype – BC-ImH, BC-ImM ,or BC-lmL.  
The legends allow us to compare different regions of the heatmap and identify patterns, in the data, for example, our data is perfectly segmented by BC subtype, but maybe we can find a sub-cluster that belongs mainly to another specific legend, e.g. TNBC.

1. Based on the top heatmaps in **Figure 1**, is the PAM50 subtype classification consistent with the three immune subtypes (BC-ImH, BC-ImM, and BC-ImL)? **(Yes/No)**No
2. Look at **Figure 2A**, why were the authors interested in comparing the enrichment PD-L1 signature among the different subtypes? (3-6 lines)  
   PD-L1 is an immunosuppressive marker, which is able to delay the activity of immune cells against the cancer cells. The authors wanted to know the expression levels of the PD-L1 across the different subtypes, asses the severity of immunosuppression in each of them, infer how the PD-L1 effects it, and extract insights regarding the course of effective treatment. For example, if a certain subtype of BC was to show high levels of PD-L1 expression and severe immunosuppression, a treatment that where to target PD-L1 pathways could be an effective approach.
3. Look at **Figure 3A** and answer **TRUE** or **FALSE** for the following statements (alpha = 0.05):

* In the **GSE24450** dataset, there is a significant difference in the **overall survival** between BC-ImL and BC-ImM. False
* In the **GSE24450** dataset, there is a significant difference in the **disease-free survival** between BC-ImL and BC-ImM. True
* In the **METABRIC** dataset, there is **no** significant difference in overall survival between BC-ImH and BC-ImL. False
* In the **GSE2034** dataset, most of the BC-ImH patients are censored within the first 100 months. False

1. In the last results section “Immunological Classification of BC Single Cells” the authors analyze the enrichment of four immune pathways in a scRNA-Seq dataset of breast cancer cells. Why did they use immune pathways instead of 28 cell type signatures (as done in Figure 1)? (3-6 lines)  
   The authors chose to do so because the immune signatures they were assessing are expressed in tumor cells themselves, meaning they needed to work with scRNA-seq data. As we learned in the lecture and tutorial, single-cell sequencing allow the examination of gene expression patterns **within each cell**, instead of focusing on the overall enrichment of immune cell types, which may not capture the heterogeneity of the tumor as accurately as the single-cell does.
2. Describe shortly the results in **Figure 8A** and explain how they are used to validate the results of the new subtype classification. (5-10 lines)  
   Figure 8A, presents us with the hierarchical clustering of 317 tumor cells from 10 BC patients, based on the enrichment scores of 4 immune-related pathways.   
   As we can see in the heatmap, this clustering classifies the tumor cells into the 3 subgroups based on their immune characteristics, and they used to validate the new subtype classification by demonstrating that the identified subtypes have distinct immune profiles.   
   Meaning this figure confirms that the new subtype classification accurately stratifies BC tumors based on their immune characteristics.
3. Write your conclusions from this study. You can describe and mention other results from the paper, say your opinion and be critical, and suggest ideas for improvements or future experiments/analysis that can be done. \*

\* There is no length restriction for this question. Here you should express your understanding and perspective about this study.

The study’s main goal is to identify immune subtypes of breast cancer by analyzing both bulk tumor and single-cell samples.   
As we saw throughout questions 1-14, the researchers performed clustering analysis of data from 6 different datasets: 5 bulk tumor datasets – MWTABRIC, TCGA-BRCA, GSE24450, GSE2034, GSE11121 and one single-cell dataset - GSE75688, based on the enrichment scores of 28 immune cell types.   
They identified 3 BC subtypes: BC-ImH, BC-ImM, and BC-ImL – each reflects high, medium, and low immune signature scores respectively.  
  
Our conclusions from the paper are:

* Main conclusion: The new 3 subtype classification is valid.
* As the study concluded, BC-ImH had the highest TMB and predicted neoantigens, while BC-ImL had the highest SCNA scores.   
  TMB is positively correlated with anti-tumor immune response and SCNAs is negatively correlated with it, which tells us that the 3 different subtypes of breast cancer have different levels of genetic mutations, so the way the immune system responds to the cancer cells can be different as well.
* The study found correlations between immune subtypes and: hormone receptor status, HER2 expression, and TNBC. TNBC and HER2+ tumors were more immunogenic, while HR+ tumors were less immunogenic.  
  Moreover, certain HER2+ or HR+ BC patients within the BC-ImH type could (maybe) benefit from immunotherapy in addition to TNBC patients.
* The expression levels of the PD-L1 marker were higher in BC-ImH, suggesting potential responsiveness to immune checkpoint inhibitors.

All of the above strengthen the assumption that the 3 subtypes classification is indeed valid.

* BC-ImH has a significantly better survival rate than BC-ImL (in multiple datasets), which means that tumors with a high immune response have better clinical outcomes than those with a lower immune response.

Strengths:

* The study analyzed multiple datasets, which increases validity.
* The study used both bulk-tumor and single-cell sequencing, which increases robustness.
* **All** datasets and analysis technics support the classification of BC into the 3 subtypes.
* The study provides a good starting point for future studies, that could explore the translation of these findings into clinical practices and develop personalized treatment for BC patients!

Point of improvement:

* Although the study did use some different data analysis techniques and data presentation methods (such as hierarchical clustering, survival analysis, heatmaps, box plots etc), as we sae in questions 5 and 7, they could have used some more diverse methodologies, like preforming K-means in addition to hierarchical clustering or using Xcell in addition to SSGSEA.   
  each technique offers unique insights into different aspects of the data, which can help strengthen the final conclusion of the study.
* The article notes that the authors only use transcriptomic data, we searched a little bit online and found some other types of data, such as roteomics and epigenomics data, which can be used in research of that kind.  
  We believe that integrating transcriptomic data with other types of data could provide a more comprehensive understanding of the immune landscape in BC.
* Analyzing existing data is very efficient, but we think that conducting functional experiments can strengthen the interpretation of the findings.

Overall, we believe the study has achieved its goal of Identifying breast cancer immune subtypes, which we believe has the potential to contribute valuable insights regarding the immune landscape of BC.