Chapter 1 - Data Acquisition

We downloaded three datasets for our analysis –– two of which are spatial transcriptomics of high grade serous ovarian cancer and the other one is single-cell sequence data used as mapping reference.

1. Spatial Transcriptomics
   a. Yeh, C.Y., Aguirre, K., Laveroni, O. *et al*. Mapping spatial organization and genetic cell-state regulators to target immune evasion in ovarian cancer. *Nat Immunol* (2024). https://doi.org/10.1038/s41590-024-01943-5 (**Denoted as Xenium**)
   b. Denisenko, E., de Kock, L., Tan, A. *et al*. Spatial transcriptomics reveals discrete tumour microenvironments and autocrine loops within ovarian cancer subclones. *Nat Commun* **15**, 2860 (2024). https://doi.org/10.1038/s41467-024-47271-y (**Denoted as Visium**)
2. Single Cell Transcriptomics
   a. Vázquez-García I, Uhlitz F, Ceglia N, Lim JLP et al. Ovarian cancer mutational processes drive site-specific immune evasion. *Nature* 2022 Dec;612(7941):778-786. PMID: 36517593 (**Denoted as Chromium**)

Xenium Dataset

The Xenium Dataset is derived from a study looking at molecular drivers preventing immune recruitment. They used single-cell spatial and perturbational transcriptomics to elucidate these processes in HGSOC ––– profiling more than 2.5 million cells *in situ* in 120 tumours from 94 patients. This dataset can be accessed in Single Cell Portal Study with accession SCP2640 fir Discovery cohort, SCP2641 for Validation 1 Cohort, SCP2650 for Validation 2 cohort, SCP2644 for Test 1 cohort, and SCP2646 for Test 2 cohort. The perturb-seq could be accessed separately from SCP2707. Below is a table summarising the dataset objects available in rds format:

ObjectName	Class	Count	Data	Scale Data	PCA/UMAP	Cluster	Metadata	Image	Cell Count	Gene Count
PerturbSeq_TYKnu_so.rds	Seurat	Yes	Yes	Yes	No	Yes	Yes	No	18585	13596
ST_Discovery_so.rds	Seurat	Yes	Yes	Yes	No	Yes	Yes	No	491792	979
ST_Test1_so	Seurat	Yes	Yes	Yes	No	Yes	Yes	No	85821	6195
ST_Test2_HGSC1_so	Seurat	Yes	Yes	Yes	No	Yes	Yes	No	542559	1010
ST_Test2_HGSC113_so	Seurat	Yes	Yes	Yes	No	Yes	Yes	No	98482	1000
ST_Test2_HGSC2_so	Seurat	Yes	Yes	Yes	No	Yes	Yes	No	211472	1010
ST_Test2_HGSC8_so	Seurat	Yes	Yes	Yes	No	Yes	Yes	No	294699	1010
ST_Validation1_so	Seurat	Yes	Yes	Yes	No	Yes	Yes	No	447874	280
ST_Validation2_so	Seurat	Yes	Yes	Yes	No	Yes	Yes	No	425578	140

Briefly speaking, the rds-formatted objects does not contain the image nor reduction needed for a successful spatial transcriptomics analysis using Seurat Pipeline. Instead, the team came up with an independent plotting system using EBI image providing the cell segments –– which iteratively colours each part of the plot based on the numbering given to each grid –– like making a pixel graph. Relative code can be accessed from their GitHub page. For our analysis purpose, we used the **ST_Test1_so.rds** considering its sufficient gene counts compared to all other cohorts, which consisted of 62 samples and 85821 cells.

Visium Dataset

The Visium dataset were performed on 8 primary tumour samples from HGSOC patients that were collected during interval de-bulking surgery from HGSOC patients who

underwent taxane-and platinum-based neoadjuvant chemotherapy. Patient number 1, 7, 8 has poor chemotherapy response score, patient 2, 3, and 5 has good response and patient 4 and 6 had partial response to the chemotherapy. Raw data can be accessed from GEO database with accession GSE211956. All 8 samples were used in our analysis.

Chromium Dataset

Whole-genome sequencing, single-cell RNA sequencing, digital histopathology and multiplexed immunofluorescence were performed on tissues collected from of 160 tumour sites across 42 treatment-naive patients with HGSOC. Raw data can be accessed from GEO database with accession GSE180661.