Improving Xenograft scRNA-seq Interpretation through Multiplet Filtering in Cancer PDX Models

Abigail Gallegos [1], Jonathon E. Mohl[1,2]

1 Computational Science Program, The University of Texas at El Paso, El Paso, Texas.

2 Department of Mathematical Sciences, The University of Texas at El Paso, El Paso, Texas.

Despite recent advances in genomic studies, aggressive and treatment resistant tumors continue to be a significant challenge due to their high metastatic potential. While there are methods that combine cell lines with patient-derived xenograft (PDX) models and are often found to provide valuable insights for studying these cancer stages, they are also considered to be analytically challenging due to their mixed human-mouse transcriptomes. One of the main issues from this complexity is known as multiplets, this is when two or more cells are captured at the same time within a single droplet generating overlapping and misleading differential expression analyses.

In this study we used a multiplet filtering pipeline on scRNA-seq data from PC3 xenograft models. The PDX samples were generated by engrafting human PC3 tumors into nude mice under three conditions: AB Control (two nude mice with untreated PC3 tumors), EFldCTX (two nude mice with PC3 tumors treated with Id CTX), and PC33LLnud (a PC3 tumor in a nude mouse treated with Id CTX that no longer responds to Id CTX treatment). Starting with 20 463 cells across three samples (AB Control: 6807, EFldCTX: 7356, PC33LLnud: 6300), we implemented a preprocessing pipeline designed for human-mouse xenograft to remove cross-species contamination. Out of the 367 cells that were flagged as doublets by Cell Ranger, Scrublet identified 152 true positive doublets in AB Control, 124 in EFldCTX, and 62 in PC33LLnud, with 26 identified as false positives across all three samples. After the filtering and reprocessing for the UMAP projection in Loupe Browser, it was found that from the 19 starting clusters driven by cross species contamination by implementing this filtering process the clusters obtained reduced to 14 well defined clusters. These findings highlight the importance of multiplet filtering for accurate cluster identification and differential expression analysis in xenograft scRNA-seq data