

ggplot2 visualization in action for summarizing large-scale scRNA-seq dataset

Authors: Amani Karnoub¹, Nicole Karnoub¹, Qin Ma², Yuzhou Chang²

1. Millburn High School, Millburn, NJ USA, 07041
2. Department of Biomedical Informatics, Columbus, OH, USA, 43210



Introduction

- The Y chromosome is one of the sex chromosomes that determines male development. The loss of the y chromosome (mosaicism) can occur in some subsets of cells as males age and as a result of various health implications.
 - Loss of y can cause:
 - A weakened immune function of lymphocytes
 - Genetic variability among lymphocytes affects their response to immune challenges.
 - An increased risk of autoimmune diseases, potentially influencing lymphocyte behavior in these conditions.
- Public data is important for this project as it increases sample size and diversity, which enhances the statistics sector and allows for more identification of trends.
 - The statistics as a result can include several factors such as smoking history, sex, disease, tissue, and technology.

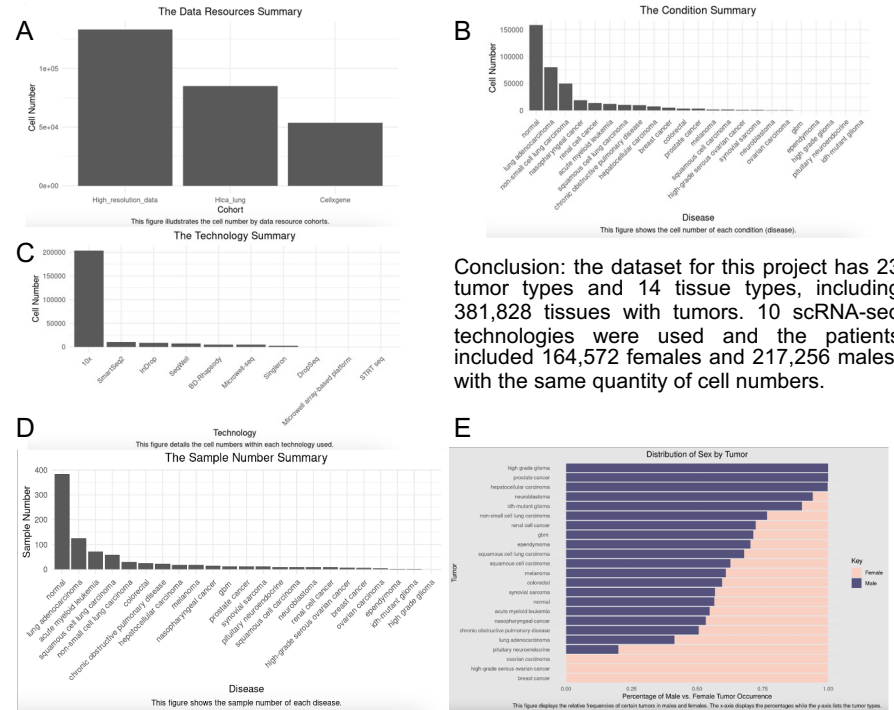
Workflow



We typically began working on a figure by making sure we understood the objective. Then, we would try to learn the code from online sites such as YouTube and ChatGPT before asking for help. Debugging/cleaning came next before we could decorate the figure and generate the final product using ggplot2.

What could have been improved?

Figure demonstration



Conclusion: the dataset for this project has 23 tumor types and 14 tissue types, including 381,828 tissues with tumors. 10 scRNA-seq technologies were used and the patients included 164,572 females and 217,256 males, with the same quantity of cell numbers.

To show *proportion*, figures A, B, C, and D could be turned into pie charts. This would still convey the differences between the respective x-axis data but display a clearer figure as well. Heat maps could also be used for these figures to show the frequencies of the x-axis data compared to the y-axis. This would involve more components and color indicators.