Starting from fastq files provided by Novogene:

1. Fastqs need to run through cellranger count to align reads to the reference genome (link provided here <https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/using/tutorial_ct>)

* User will need to input appropriate reference genome and path to fastqs
* Of note: for biological replicates cellranger does provide an aggregation function that reduces batch effect. The relevance of using this function I believe would vary from experiment to experiment depending on tissue availability and cost effectiveness but I provided a link here as a consideration (<https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/using/aggregate>)

1. Cellranger will output a directory with numerous files. The ones I find most relevant are the web\_summary.html file that reports cell number, reads per cell and sequence saturation (example attached here <file:///Users/maxzinter/Desktop/Data/scRNA_Seq/Ingolf_Bach_Fall_2022/WT_01/web_summary.html>) and then the filtered\_feature\_bc\_matrix which will be assigned as a Seurat object for further analysis.
2. Now in R studio: load the Seurat package and assign the filtered\_feature\_bc\_matrix as a Seurat object. I included an example Seurat script in the email (“Concise\_IPN\_Stress\_Fall\_2022.R”) and the basic Seurat vignette can be found here (<https://satijalab.org/seurat/articles/pbmc3k_tutorial.html>). For analysis of datasets using a control and experimental condition, I prefer to use the integration vignette found here (<https://satijalab.org/seurat/articles/integration_introduction.html>) which will allow for differential expression across cell types.
3. From here I included several stopping points where user input will likely be required, these are also annotated in the “Concise\_IPN\_Stress\_Fall\_2022.R” script:
   1. USER INPUT 1: QC/dead cell removal: User will need to input the threshold for percent mitochondrial reads and features. Of note: there is a more complex QC process published here (<https://github.com/stuberlab/Hashikawa-Hashikawa-2020>) included for reference but unnecessary for the first round. Useful outputs here would be the feature scatter plots for mitochondrial reads and nFeature\_RNA.
   2. USER INPUT 2: Determining principal components and dimensionality of the dataset to be used for clustering of the dataset. This could likely be an output or significance cutoff from the JackStraw Plot. Useful output here would be a umap/tsne projection of the data.
   3. USER INPUT 3: Annotation. Here I use Alfred’s “get\_cluster\_markers.py” script using the csv output from the FindAllMarkers Seurat function. The user should update the IPN\_markers file attached in the email with the appropriate cell type markers and their corresponding cell type. This will then be used to annotate clusters. Useful outputs here would be the annotated umap/tsne projection, a violin plot/dotplot of relevant markers across clusters.
   4. USER INPUT 4: If differential expression testing is of interest, the user should then specify what cell type they would like to have differential expression testing performed on across conditions. Seurat’s FindMarkers function can be used for this and offers a range of differential expression testing methods ( I prefer using the default Wilcoxon rank sum test). Useful outputs here would be a volcano plot of DEGs with maybe top/bottom 10 genes labeled, a csv file listing all DEGs across conditions and volcano plots for relevant genes across conditions (these genes could be selected by the user). Additionally, what I have found helpful when collaborating with other labs in the department is to give them access to the rds file of the Seurat object and a line of code allowing them to inpout genes of interest and project them as volcano plots across condition or cell type.