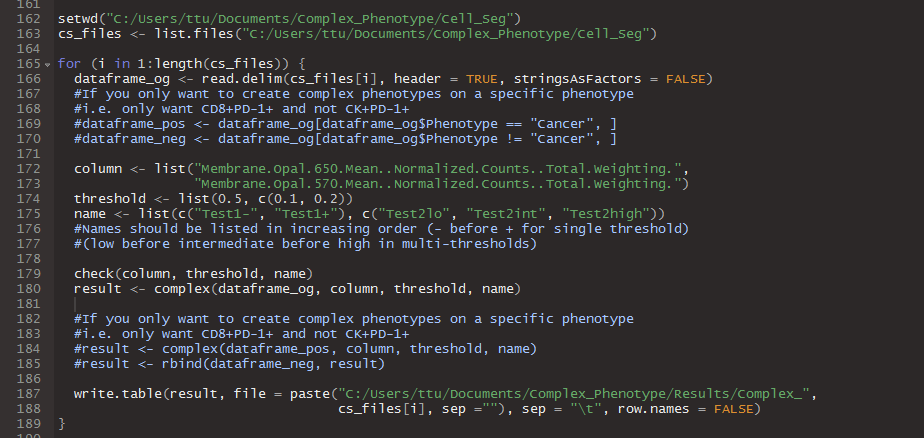
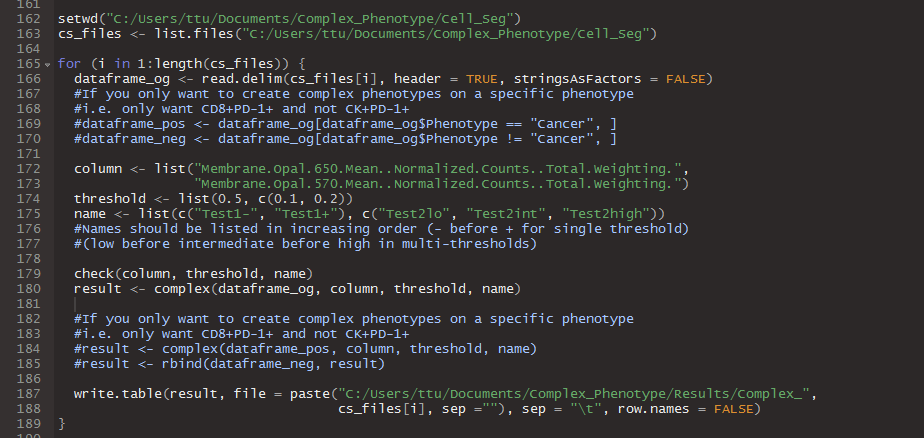
The purpose of this script is to create a complex phenotype based on the thresholds determined by the scoring window in inForm. The script will change the Cell ID and Phenotype columns in the cell\_seg\_data.txt file. The Phenotype column will be changed from the base phenotype (the phenotype as listed in the phenotype column when the cell\_seg\_data.txt file is output from inForm) to the base phenotype + the complex phenotype (added on as determined by the threshold).

**Classic Workflow: Recoordination.R -> Complex\_Phenotype.R -> Phenotype\_Distance\_TT.R**

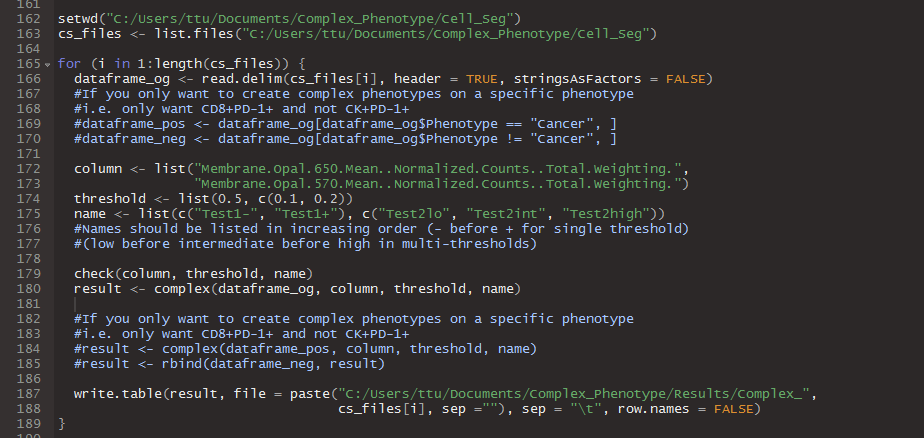
1. The folder **Complex\_Phenotype** should have two subfolders: **Cell\_Seg** and **Results**
   1. Check to make sure the Complex\_Phenotype folder with the two subfolders in it is in the working directory (the stuff inside the setwd() function)
      1. If it is not in the working directory, change your working directory in the areas boxed in red below to wherever the Complex\_Phenotype folder is.
      2. Do not change the “Complex\_” part. Just change everything before the last “/” in the box.

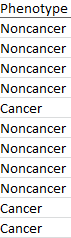


1. The **Cell\_Seg** subfolder should contain **cell\_seg\_data.txt** files. Before running the **Results** subfolder should be empty. This is where your results will go.
   1. Make sure the cell\_seg\_data.txt files are “Tab Delimitated”. If they are not (if they were exported from Spotfire they might be Unicode), ask Anthony or Ting-Fang for a script to change them or you can open them in Excel and “Save As”, “Save as type: **Text (Tab delimited) (\*.txt)**”
2. Change the variables “column”, threshold”, and “name”.
   1. **Column** should be in a string format (in quotes like “”)
   2. **Threshold** should be numbers
      1. The c() is a vector in R. If you put in multiple values of the same data type (i.e. all numbers or all strings) in there separated by commas, it will be grouped together
   3. **Name** should also be in string format
   4. **MAKE SURE THE INDICES LINE UP FOR ALL THREE OF THESE VARIABLES**
      1. The first element of column corresponds to the first element of threshold and the first element of name. Things within c() are all considered part of one element.
         1. Ex. Opal 650 corresponds with 0.5 and c(“Test1-“, “Test1+”)
         2. Ex. Opal 570 corresponds with c(0.1, 0.2) and c(“Test2lo”, “Test2int”, “Test2high”)
      2. Make sure the thresholds within threshold are listed in increasing order within each element.
         1. If there is only one threshold, that is already in order
         2. If there are multiple thresholds (within a c()) make sure the numbers within the c() are increasing in order
      3. Make sure the names within name are listed in increasing order within each element. Stuff within a c() are all within the same element.
         1. “minus” (-) goes before “plus” (+) if there is only one threshold
         2. “low” goes before “intermediate” goes before “high”, etc. if there are multiple thresholds
         3. If n = number of elements in each element of threshold, then name should have n+1 elements for the corresponding element. The script will error out if that requirement is not met
            1. Ex. There are two names in the c() for the first element of name: “Test1-“ and “Test1+” for the one threshold (0.5)
            2. This is because each threshold n creates n+1 bins

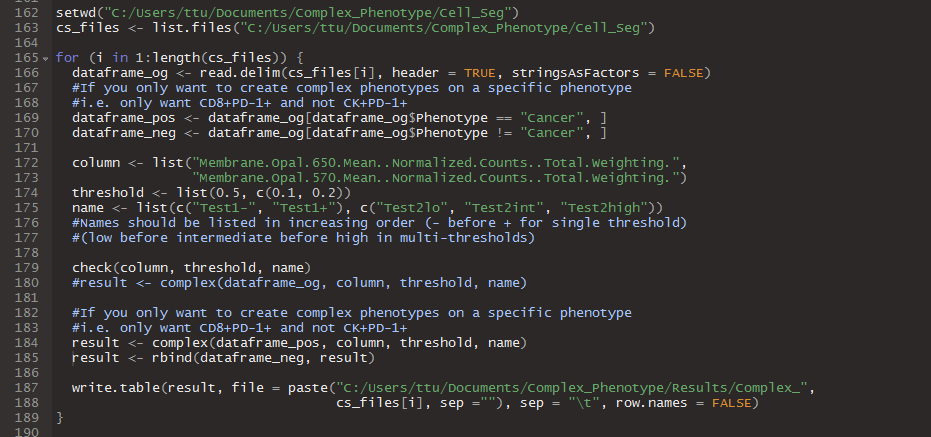


1. The way the script is set up below will create complex phenotypes for every single row regardless of what the base phenotype is. What I mean by base phenotype is what the original phenotype column says when the cell\_seg\_data.txt file is spit out of inForm. In the image below the base phenotypes of the file are “Noncancer” and “Cancer”.
   1. Ex. If you do it this way and you want to complex phenotype with “PD-1+” and “PD-1-“ then every row will be “CancerPD-1+”, “CancerPD-1-“, “NoncancerPD-1+” or “NoncancerPD-1-“





1. If you want to only complex phenotype a particular phenotype like “Cancer” for example, use the “#” to comment out the second of the two lines of code boxed in red above and uncomment the lines boxed in red below. You can change the phenotype being searched for (in the example it is searching for “Cancer”) by change what is in the string in those lines.
   1. Ex. To continue the previous example, if you only want to complex the “Cancer base phenotype with “PD-1+” or “PD-1-“, then all the complex phenotypes will be “CancerPD-1+”, CancerPD-1-“, or “Noncancer”



1. Click “Source” in the top corner of the RStudio Script window and the whole script will run!