**CrisprCountsAnalysis (CCA) Input files format**

CCA has two input files per CCA run. If there are multiple test-control sample pairs you want to try per screen, a pair of CCA input files will have to be generate for each run.

IT IS CRITICAL THAT THE TEXT FILES END OF LINES ARE ENCODED AS \n’s (Linux conventions). E.g. see <https://kb.iu.edu/d/acux>.

1. A read-counts file in tab-separated format
   1. The name must be of the form ID.txt, where ID is something short consisting of numbers and letters, with no underscores, spaces, or other special characters.
   2. First column = sgRNA names. These names must all be unique. The names should be contiguous with no spaces, preferably using underscores.
   3. Second column = gene names. It is preferred that standard gene names are used.
   4. Subsequent columns = samples
      1. The names must be just letters, numbers, periods, and underscores. No spaces or special characters. Dashes will cause an undefined behavior.
         1. So instead of BRCA1 -/-, use something like dBRCA1 in the sample name.
      2. Sample names should be human-readable in the following specific format
         1. ScreenID\_Library\_CellLine\_ScreenCondition\_CellLineCondition\_clone\_TtimeReplicate
         2. E.g. RSL1\_TKOv3\_RPE1\_BRCA1\_dBRCA1\_clone1\_T18B
         3. E.g. RSL1\_TKOv3\_RPE1\_BRCA1\_WT\_0\_T0
         4. E.g. RSL1\_TKOv3\_RPE1\_BRCA1\_WT\_0\_T6A
         5. If there is no clone information, use 0 as the clone name.
         6. The condition may be either a gene knock out or a drug name.
   5. Row = sgRNA
   6. Entries = raw count
      1. For a cell, it is the number of counts of sgRNA for that sample.

A screenshot of a cell phone

Description automatically generated

1. A replicate map (repmap) file in tab-separated format, typically named ID.repmap
   1. Header is  
        
      Replicate [tab] Sample [tab] T0  
        
      unless it is a paired chemogenomic run in which case the header is  
        
      Replicate [tab] Sample [tab] T0 [tab] Control\_Pair
   2. First column = sample name
      1. It must be identical to the sample name used in the above file.
   3. Second column = either TEST or CTRL
      1. It can never be anything else other than this.
   4. Third column = T0 sample name.
      1. This has to be identical to the sample name used in the read-counts file.
      2. If the sample is a T0 sample, the first column and this column are identical.
         1. This is how CCA determines it is a T0 sample.
      3. The second column should be CTRL.
      4. While the attached example has only one T0, you can have multiple T0s in the read-count and repmap files.
         1. This is to accommodate different samples having different T0s.
   5. Optional fourth column = control pair
      1. This is only for paired chemogenomic screens.
      2. This is only for TEST samples.
         1. Leave this blank for CTRL samples.
      3. The sample name has to be identical to a CTRL sample name in the read-counts file.
      4. Either all TEST samples must be specified, or none.
      5. In the below example DRUGX is the drug name.

A screenshot of a cell phone

Description automatically generated