Input format requirements:

factors.tbl must be a tab-separated file, organized as follows (example included in data file):

* first column:
  + firstline is header, ”file”
  + each row is the name of each file as named by the RSEM algorithm (default naming is important to determine isoform vs gene level summary for this analysis)
* second column:
  + firstline is header, “replicate”
  + each row is a string representing which biological replicate corresponds to the data in each file in the first column. These must be matched between the input and IP/antibody samples
* third column:
  + firstline is header, “abTreatment”
  + each row indicates whether or not the sample for the row was treated with antibody pulldown (enrichment for 8OG) corresponding to the file in the row
  + It is REQUIRED that each control be indicated with an entry of “none”, but the indicator for antibody can be any string (i.e. “yes” or “8OG” or “m5C” in case multiple antibody pulldowns were performed, for example)
* fourth column:
  + firstline is header, “exposureCondition”
  + each row indicates the factors that were used for exposure study corresponding to the file in the row
  + It is REQUIRED that each control be indicated with an entry of “control”, but the indicator for other treatments can be any string (i.e. “exposed” or “test” or “ozone-acrolein-methacrolein” or “OAM”)
* fifth column:
  + firstline is header, “exposureLevel”
  + each row indicates the level of exposure corresponding to the file in the row (i.e. “Low” or “High” or “400ppb” or “0.4” or “4”
  + If you only have one exposure level, you must still include this column and just enter all the same value for every row. It can be arbitrary, as it will not be used in DESeq analysis and only in organizing data
  + The variables in this column will be used to organize your data into subsets for each exposure level for enrichment analysis