

In this task, you are given obfuscated expression data (file **'Obfuscated\_expression\_quantile\_normalized'**, expression microarray, raw expression values have already been quantile normalized) and we want you to hypothesize what type of medical condition differentiates Cases from Controls.

We have prepared two types of additional resources that may help you:

1. file **'curated\_gene\_disease\_associations\_disgenet'** contains known curated lists of gene-disease associations from the DisGeNet database
2. file **'example\_ppi\_graph\_omnipath\_10k\_nodes'** contains a subset of publication asserted protein-protein interactions (ppi graph) from Omnipath (**extra!**)

You are **NOT!** required to use any of these and can instead rely on any other publicly available resource or online service.

Note, that precise identification of the condition is **less important** than a clear explanation of your thought process. You don't need to write down a report, a notebook(R/Jupyter) with some minimal commentary would suffice.

The estimated time for this task completion is about 2-6 hours, depending on how deep you want to go and your experience. There is no deadline and you are not required to start immediately, but please don't overinvest in this task.