3. **MEDIUM DIFFICULTY:** Single cells, transcription factor factor activity proxy, and post-translational modifications. The activity of a transcription factor is in principle detectable by observing the expression level of its targets. Indeed, we usually assume that genes expressed similarly are often controlled by a same regulator, and therefore also the reverse should be true. In this project, we want to introduce a proxy for a transcription factor activity based on the expression level of genes regulated by that transcription factor. The regulator-target relationships can be found in specialized databases, such as regulondb for coli, yeastract for yeast... For a positive regulator, its activity can be taken as being a function (linear? sigmoidal?) of the expression level of its targets (mean? median? max?). Now, the question is to show the degree at which transcription factor activities are correlated to the corresponding transcription factor gene expression level. Our null hypothesis is that indeed, the abundance of the transcript is also a proxy of the activity. This can be somehow expected, but plausibly not for all transcription factors. Many for instance are active only in phosphorylated form, meaning that the transcript can also be constitutively expressed; the activity will depend on phosphorylation however, and therefore it will be correlated to the activity of the corresponding kinase. In this way, one could derive a general model to relate the activity of a TF and its transcript's abundance, and in case the correlation is significant, to identify those transcription factors significantly deviating from the general behavior.

   Data source
   https://www.nature.com/articles/s41598-022-12463-3
   https://github.com/SBRG/precise-db/tree/master/data
   file to download: log_tpm_norm.csv

   and for the regulatory network:
   https://regulondb.ccg.unam.mx/datasets

4. **MEDIUM DIFFICULTY:** DNA replication introduces copy number variations that are detectable in RNA-seq, especially at the single cell level. The extent of this signal depends on the growth rate, being visible when the division time (=ln2/growth_rate) is shorter than the time required to replicate the chromosome (speed of the DNA polymerase ~1000nt/s, length of one chromosome arm ~2E+6, time required ~1/2h, around 40min),which is possible only in mero-oligoploid species, like *E. coli* or *B. subtilis*. These copy number variations produce a clear pattern in gene expression in single cell data (see for instance https://www.biorxiv.org/content/10.1101/2022.10.22.513359v1). Since at any given moment, a population of *E. coli* is characterized by a certain distribution in the number of active replication forks per cell, this means that copy numbers are potentially different in different cells, depending on the number of replication forks and the total DNA content.