IO17 | Large Scale Bioinformatics for Immuno-Oncology
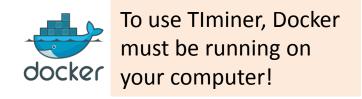
Neoantigens: exercise 4 - Solution

Francesca Finotello, Federica Eduati, and Pedro L. Fernandes

INSTITUTO
GULBENKIAN
DE CIÊNCIA

To use TIminer, Docker must be running on your computer!

We have predicted a list of candidate neoantigens, i.e. mutated peptides binding to the patients HLA-A molecule: *Patient_1_NetMHCpan_binders.txt*

We also have access to normalized RNA-seq expression data from the same patient (assume the data are already in the format required by Timiner, i.e. log2(TPM+1)): *Patient_1_normalized_expression.txt*

Use TIminer (function TIminerAPI.**filterNeoantigenFile**) to filter only the peptides arising from genes with TPM > 1.

Once you get the results, answer the questions related to this exercise at:

https://b.socrative.com/login/student/

Neoantigen prioritization: Python code

```python
from TIminer import TIminerAPI

TIminerAPI.filterNeoantigenFile(
    neoantigenInputFile="../Output/Patient_1_NetMHCpan_binders.txt",
    geneExpressionInputFile="../Input/Patient_1_normalized_expression.txt",
    filteredNeoantigenOutputFile="../Output/Patient_1_candidate_neoantigens.txt",
    expressionThreshold=1)
```

1) How many peptides are select by TIminer for Patient 1 after the filtering (Exercise 4)?

154

2) When running neoantigen filtering for patient 1 (Exercise 4), a warning message reports that the mutated gene KMT2D is not present in the file of normalized expression.

Sometimes, genes are not missing from expression data, but are saved with different gene symbols/names/IDs.

Search for "**KMT2D**" on the website of the HUGO Gene Nomenclature Committee (HGNC) and retrieve **synonyms** (i.e. alternative symbols that have been used to refer to the KMT2D gene) and **previous symbols** (i.e. previous symbols used for KMT2D the gene that have now been dismissed): http://www.genenames.org.

Then, look for these alternative symbols in the file of normalized expression to extract the **expression of KMT2D**.

According to its normalized expression, would KMT2D have passed the **threshold** for neoantigen filtering?

Previous symbol: MLL2 → log2(TPM+1)=18.87 > 1 TPM → **Yes**