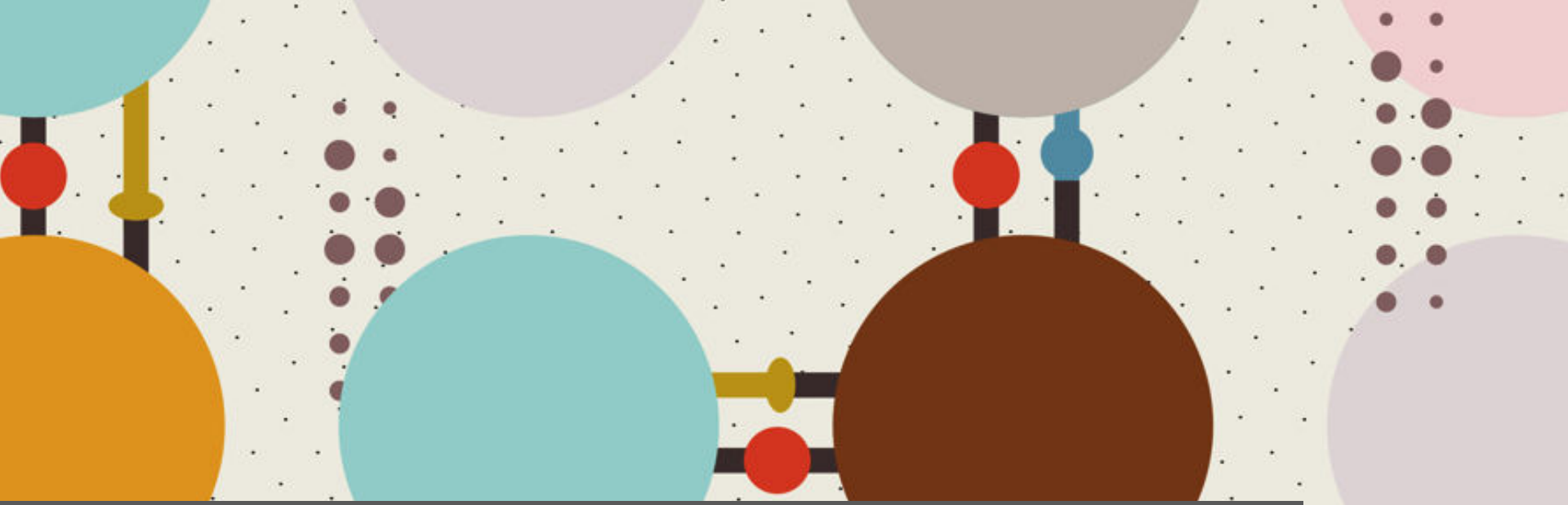


IO17 | Large Scale Bioinformatics for Immuno-Oncology

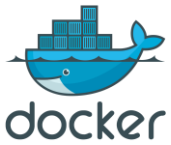
Neoantigens: exercise 3 - Solution

Francesca Finotello, Federica Eduati, and Pedro L. Fernandes

GTPB | The Gulbenkian Training Programme in Bioinformatics
Instituto Gulbenkian de Ciência, Oeiras, Portugal | Sept 19th-22nd, 2017

INSTITUTO GULBENKIAN DE CIÊNCIA

Exercise 3: peptide-MHC binding prediction



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

From Patient_1's data, we have:

- Predicted the HLA types: *Patient_1_HLA.txt*
- Annotated the mutations: *Patient_1_VEP_37_mutations.txt*
- Predicted the sequences of the mutated proteins: *Patient_1_VEP_37_proteins.txt*

Predict the mutated peptides, **8-11 amino acid** long, which bind to the patient's HLAs with TIminer (function TIminerAPI.**executeNetmhcpan**), considering a **IC_{50} cut-off of 500 nM** for the binding affinity.

Important note: To reduce the computational time, we will consider only the first HLA-A type, namely a tab-delimited text file with the same format of the output file of Input/Patient_1_HLA.txt, but with all HLA types equal: *Patient_1_HLA-A.txt*.

Peptide-MHC binding prediction: Python code

```
from TIminer import TIminerAPI

TIminerAPI.executeNetmhcpan(
    inputFile="../Output/Patient_1_VEP_37_mutations.txt",
    mutatedProteinsInputFile="../Output/Patient_1_VEP_37_proteins.txt"
,
    hlaInputFile="../Input/Patient_1_HLA-A.txt",
    outputFile="../Output/Patient_1_NetMHCpan_binders.txt",
    threadCount=8,
    minPeptideLength=8,
    maxPeptideLength=11,
    affinityThresh=500)
```

Peptide-MHC binding prediction: questions

1) How many (binding and non-binding) peptides were predicted by NetMHCpan for Patient 1 (Exercise 3)?

179

Peptide-MHC binding prediction: questions

2) According to the NetMHCpan predictions for Patient 1 (Exercise 3), the binding affinity of the mutated peptide originated from a DNA mutation at position 58,145,430 of chromosome 12 (CDK4 gene) with HLA-A02:01 is stronger than that of the wild-type (i.e. non-mutated) peptide.

TRUE (low IC50/rank = binding high affinity)

Pos	GeneSymbol	Protein	HLA	MutAFF	MutRank	RefAFF	RefRank
12:58145430	CDK4	ENSP00000257904	HLA-A02:01	51.8	0.7	27546.2	33

Peptide-MHC binding prediction: questions

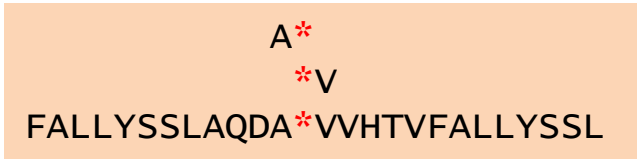
3) A patient is **homozygous for the HLA-A and HLA-B loci**, and **heterozygous for the HLA-C locus** and has **1,000 non-synonymous DNA mutations** affecting protein coding genes.

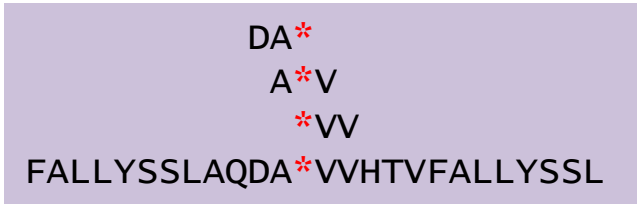
We want to predict the affinity of **8-11 amino acid** long peptides arising from the mutated proteins to the patient's HLAs.

Hint: you can use the **substr**(string, start, stop) function to extract a sub-string in R

How many peptide-HLA pairs do we have to assess (suppose the mutations are in the middle of the proteins and we can always extract a full-length peptide)?

15,2000 = 1,000 * (8+9+10+11) * (1+1+2)





Peptide-MHC binding prediction: questions

4) Write an R code that extracts all mutated peptides of length 9 from the following protein (the mutated amino acid is indicated by "*"):

MISW*VVHTVFLFALLYSSLAQDASPQSEIRAEEIPEGASTLAFVFDVTGSMYDDLVQVI

How many different mutated peptides are extracted? **5**

```
MISW*VVHT
 ISW*VVHTV
  SW*VVHTVF
   W*VVHTVFL
    *VVHTVFLF
MISW*VVHTVFLFALLYSSLAQDASPQSEIRAEEIPEGASTLAFVFDVTGSMYDDLVQVI
```

```r
str <- "MISW*VVHTVFLFALLYSSLAQDASPQSEIRAEEIPEGASTLAFVFDVTGSMYDDLVQVI"
mutpos <- 5
peptidelen <- 9
for (i in 1:peptidelen) {

  pos <- mutpos-peptidelen+i
  peptide <- substr(str,pos,pos+peptidelen-1)
  if (nchar(peptide)==peptidelen) cat(peptide, "\n")

}
```