**IO17 |** Large Scale Bioinformatics for Immuno-Oncology

## Neoantigens: exercise 1 - 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 access to single-end RNA-seq data from a melanoma patient:

*Patient_1_RNAseq.fastq*

Perform HLA typing with TIminer (function TIminerAPI.**executeOptitype**) using „Patient_1" as subject ID)

**Hint**: you can perform parallel computation using up to 8 CPUs.

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

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

```python
from TIminer import TIminerAPI

TIminerAPI.executeOptitype(inputtype="rna",
    inputFile1="../Input/Patient_1_RNAseq.fastq",
    outputFile="../Output/Patient_1_HLA_typing.txt",
    subjectId="Patient_1",
    threadCount=8)
```

1) What is the resolution of the HLA types in Optitype's output text file (NOT in the coverage plot)?

☐ 2 digits
☑ 4 digits (e.g. HLA-A02:01)
☐ 6 digits
☐ 8 digits

2) Select the HLA types predicted by Optitype for Patient 1 (Exercise 1):

☐ HLA-A23:01

☑ HLA-A02:01

☐ HLA-A23:02

☑ HLA-B50:01

☑ HLA-B44:02

☐ HLA-B35:01

☑ HLA-C06:02

☐ HLA-C04:01

☑ HLA-C05:01

3) Is the HLA-A locus predicted by Optitype for Patient 1 (Exercise 1) homozygous or heterozygous?

☑ Homozygous

☐ Heterozygous

HLA-A02:01    HLA-A02:01    HLA-B50:01    HLA-B44:02    HLA-C06:02    HLA-C05:01

4) Is the HLA-C locus predicted for Patient 1 by Optitype (Exercise 1) homozygous or heterozygous?

☐     Homozygous

☑     Heterozygous

HLA-A02:01     HLA-A02:01     HLA-B50:01     HLA-B44:02     HLA-C06:02     HLA-C05:01

4) The Optitype coverage plot for Patient 1 (Exercise 1) shows that the RNA-seq reads were assigned ambiguously to two regions of the HLA-B locus (see the light-yellow regions).

You can extract the FASTA sequences of the two predicted HLA-B alleles considered by Optitype from:

https://raw.githubusercontent.com/FRED-2/OptiType/master/data/hla_reference_rna.fasta

Use the HLA names reported in the Optitype coverage plot

Then, you can compare the two FASTA sequences with BLAST:

https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch&BLAST_SPEC=blast2seq&LINK_LOC=align2seq

**Hints**: Which is the length of the longest sequence the two HLAs have in common, without mismatches? What is the length of the RNA-seq reads in the FASTA file?
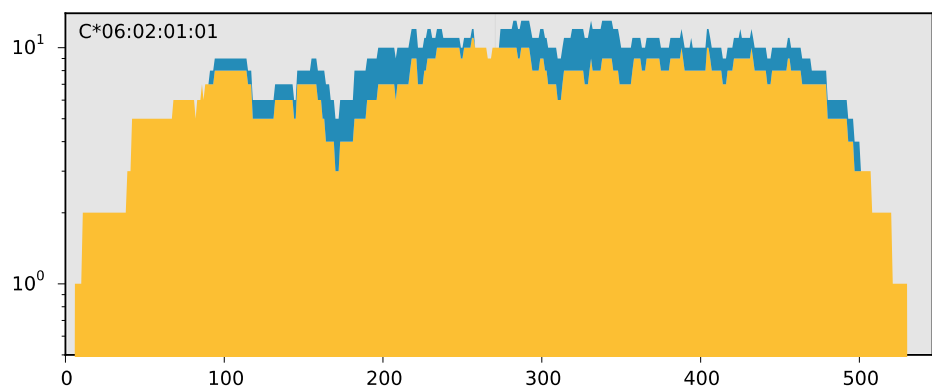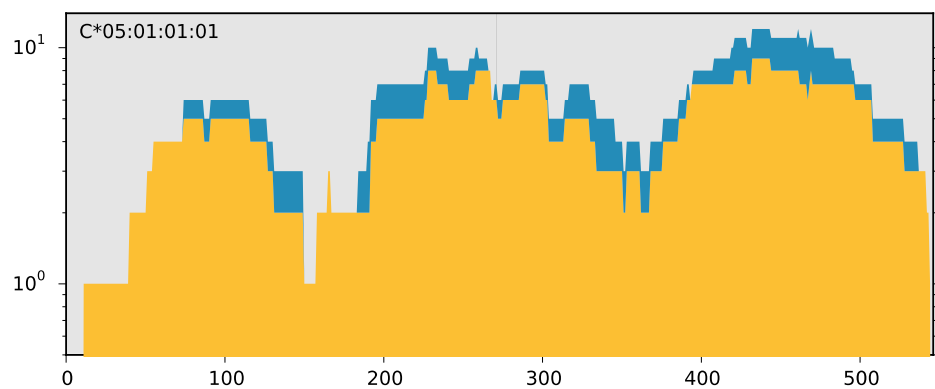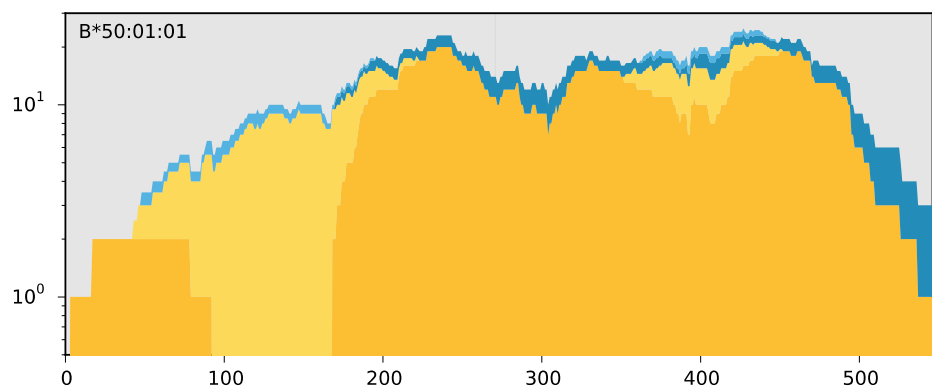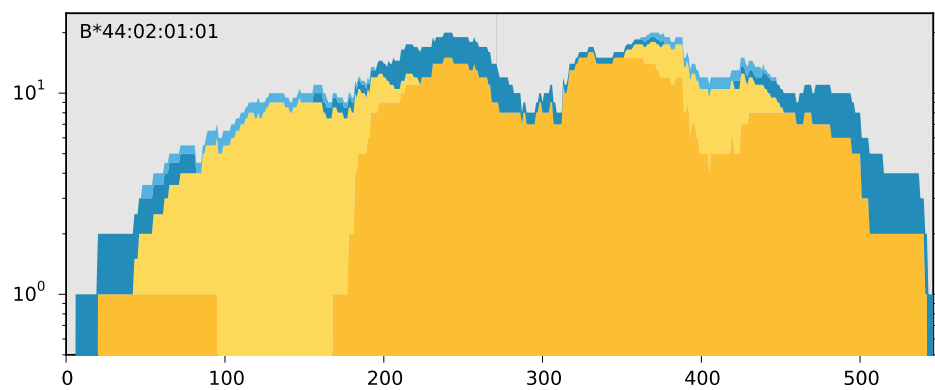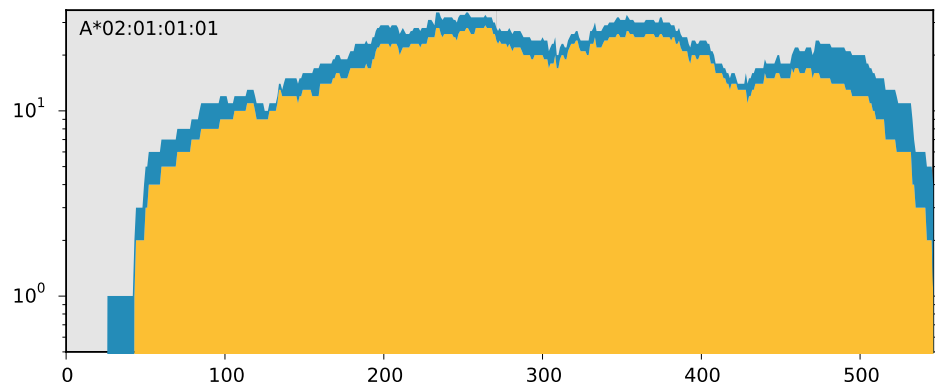
| | | | |
|---|---|---|---|
| ■ paired, no mismatches, unique | ■ unpaired, no mismatches, unique | ■ paired, mismatched, unique | ■ unpaired, mismatched, unique |
| ■ paired, no mismatches, ambiguous | ■ unpaired, no mismatches, ambiguous | ■ paired, mismatched, ambiguous | ■ unpaired, mismatched, ambiguous |

## HLA00341 HLA-B*50:01:01

Sequence ID: Query_24085   Length: 546   Number of Matches: 1

**Range 1: 1 to 546** Graphics

▼ Next Match  ▲ Previous Match

| Score | Expect | Identities | Gaps | Strand |
|---|---|---|---|---|
| 876 bits(474) | 0.0 | 525/549(96%) | 6/549(1%) | Plus/Plus |

```
Query  1    GCTCCCACTCCATGAGGTATTTCTACACCGCCATGTCCCGGCCCGGCCGCGGGGAGCCCC  60
            ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Sbjct  1    GCTCCCACTCCATGAGGTATTTCCACACCGCCATGTCCCGGCCCGGCCGCGGGGAGCCCC  60

Query  61   GCTTCATCACCGTGGGCTACGTGGACGACACGCTGTTCGTGAGGTTCGACAGCGACGCCA  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GCTTCATCACCGTGGGCTACGTGGACGACACGCTGTTCGTGAGGTTCGACAGCGACGCCA  120

Query  121  CGAGTCCGAGGAAGGAGCCGCGGGCGCCATGGATAGAGCAGGAGGGGCCGGAGTATTGGG  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  CGAGTCCGAGGAAGGAGCCGCGGGCGCCATGGATAGAGCAGGAGGGGCCGGAGTATTGGG  180

Query  181  ACCGGGAGACACAGATCTCCAAGACCAACACACAGACTTACCGAGAGAACCTGCGCA-CC  239
            |||||||||||||||||||||||||||||||||||||||||||||||| |||||||| ||
Sbjct  181  ACCGGGAGACACAGATCTCCAAGACCAACACACAGACTTACCGAGAGAGCCTGCGGAACC  240

Query  240  -GCGCTCCGCTACTACAACCAGAGCGAGGCCGGGTCTCACATCATC-CAGAGGATGTACG  297
             |||||   ||||||||||||||||||||||||||||||||| |||| ||||||||||||
Sbjct  241  TGCGCG--GCTACTACAACCAGAGCGAGGCCGGGTCTCACA-CTTGGCAGAGGATGTATG  297

Query  298  GCTGCGACGTGGGGCCGGACGGGCGCCTCCTCCGCGGGTATGACCAGGACGCCTACGACG  357
            ||||||| ||||||||| ||||||||||||||||||||||| | ||||| |||||||||
Sbjct  298  GCTGCGACCTGGGGCCCGACGGGCGCCTCCTCCGCGGGTATAACCAGTTAGCCTACGACG  357

Query  358  GCAAGGATTACATCGCCCTGAACGAGGACCTGAGCTCCTGGACCGCGGCGGACACCGCGG  417
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  358  GCAAGGATTACATCGCCCTGAACGAGGACCTGAGCTCCTGGACCGCGGCGGACACCGCGG  417

Query  418  CTCAGATCACCCAGCGCAAGTGGGAGGCGGCCCGTGTGGCGGAGCAGGACAGAGCCTACC  477
            |||||||||||||||||||||||||||||||||||| ||||||||| ||| |||||||||
Sbjct  418  CTCAGATCACCCAGCGCAAGTGGGAGGCGGCCCGTGAGGCGGAGCAGCTGAGAGCCTACC  477

Query  478  TGGAGGGCCTGTGCGTGGAGTCGCTCCGCAGATACCTGGAGAACGGGAAGGAGACGCTGC  537
            |||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Sbjct  478  TGGAGGGCCTGTGCGTGGAGTGGCTCCGCAGATACCTGGAGAACGGGAAGGAGACGCTGC  537

Query  538  AGCGCGCGG   546
            |||||||||
Sbjct  538  AGCGCGCGG   546
```

from pos. 25, identical region 204 bp long >> read length (76 bp)

from pos. 347, identical region 117 bp long