

INRAE



université  
PARIS-SACLAY

# ➤ Convolutional Neural Network for the Classification of SARS-CoV-2 and Primer Creation

Alberto TONDA, Ph.D. (Senior permanent researcher, DR)

*UMR 518 MIA-PS, INRAE, AgroParisTech, Université Paris-Saclay  
UAR 3611, Institut des Systèmes Complexes de Paris Île-de-France*

## ➤ Outline

- Application
- Formalization
- Classification results
- Extracting information from filters
- A new primer



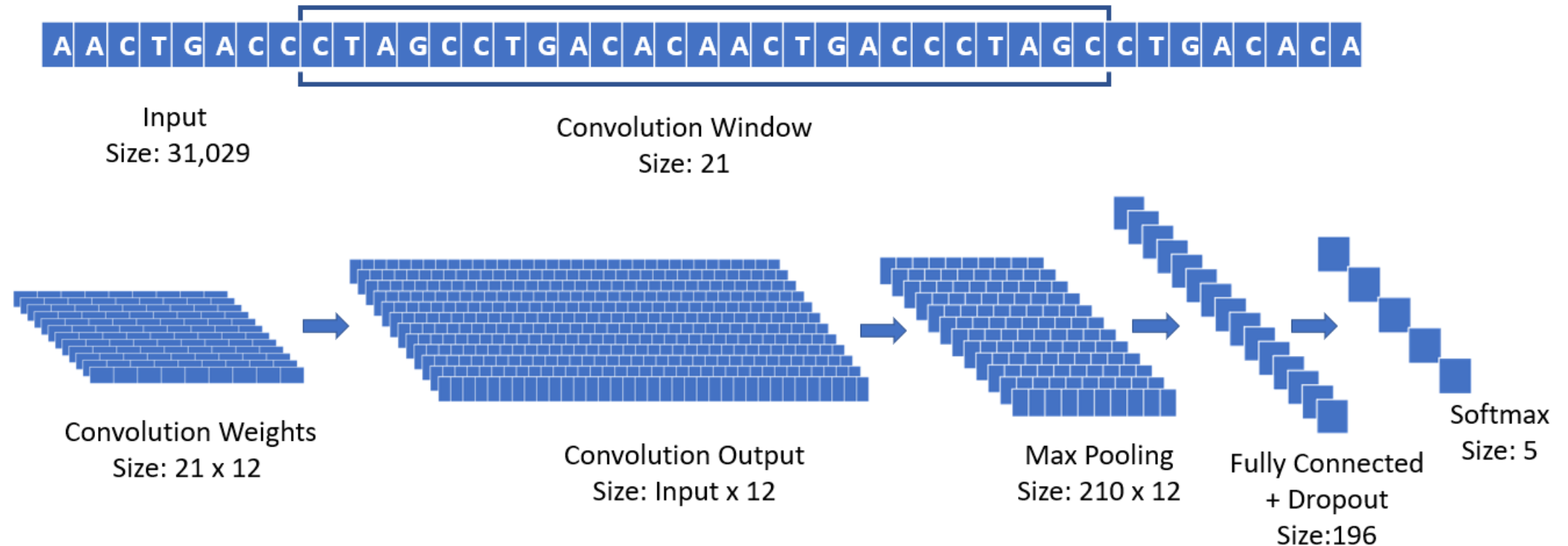
## ➤ Application

- Separate genomic samples of SARS-CoV-2 from others
  - SARS-CoV-2 belongs to the Coronavirus family
  - Lots of “relatives” that are genomically similar
  - For example, the regular flu!
- During the pandemic, collective effort to share data
  - Open repositories with genomic sequences from all the world
  - Fast and reliable genomic sequences from high-throughput



# ➤ Formalization

- Basically a classification problem
  - Input: sequence of variable size (RNA, cDNA)
  - Output: class

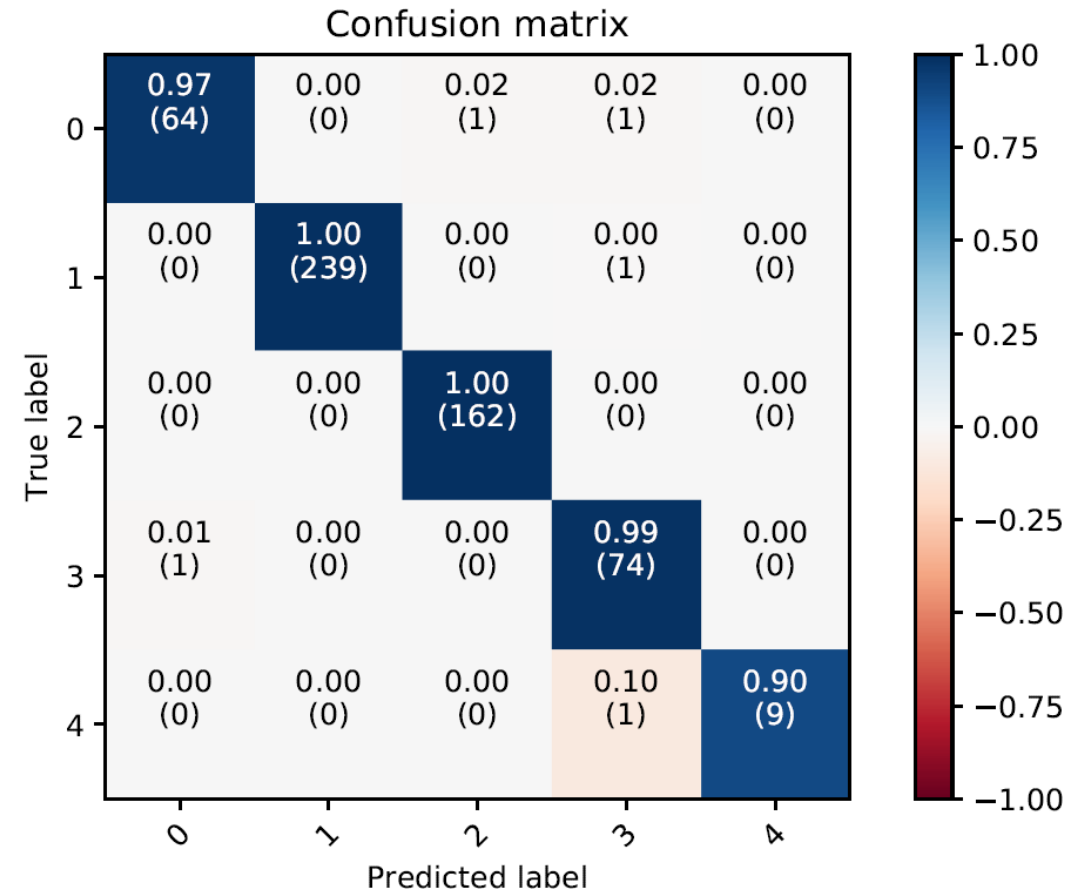
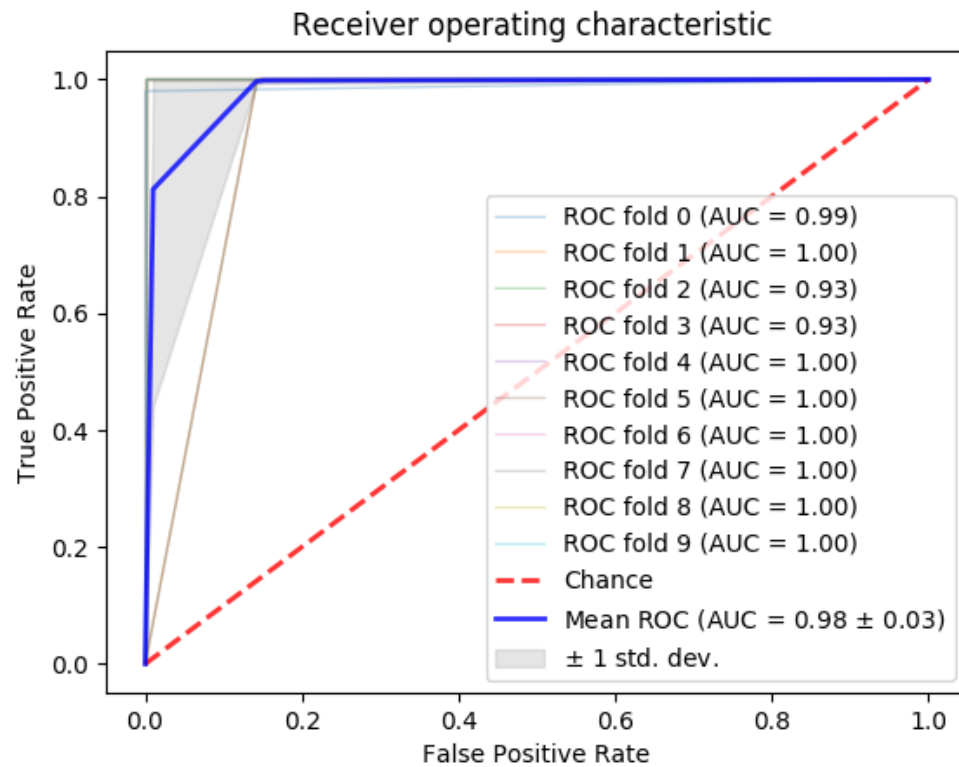


# ➤ Formalization

- Data: a few hundred samples

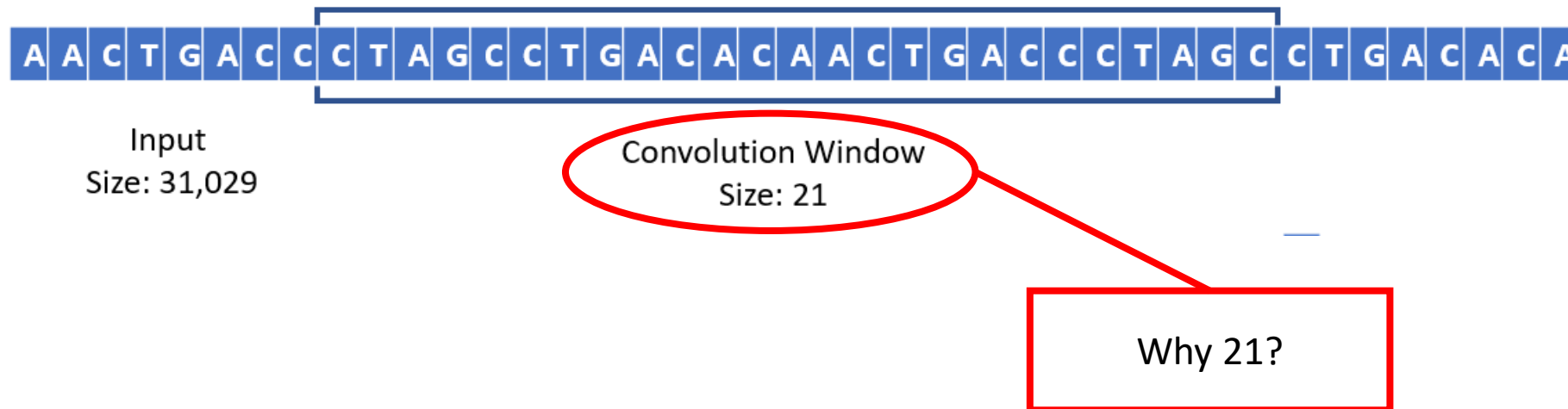
Organism	Label	Number of Samples
SARS-CoV-2	0	66
MERS-CoV	1	236
HCoV-OC43	2	136
HCoV-229E	2	22
HCoV-EMC	2	6
HCoV-4408	2	2
HCoV-NL63	3	58
HCoV-HKU1	3	17
SARS-CoV	4	7
SARS-CoV P2	4	1
SARS-CoV HKU-39849	4	1
SARS-CoV GDH-BJH01	4	1

# ➤ Classification results



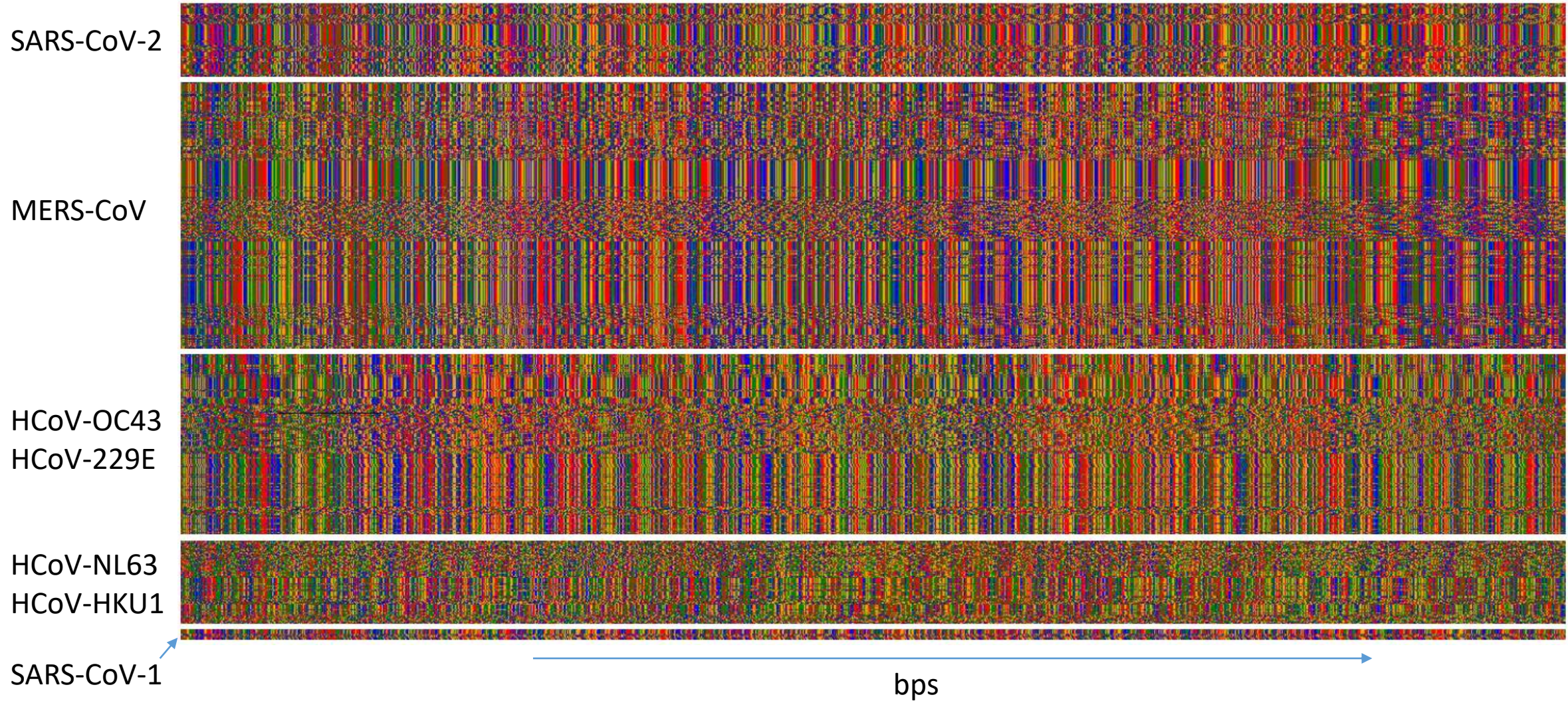
## ➤ Classification results

- Good! But can we do something more?
  - Visualization tell us **where** filters are maximally activated in sample
  - By looking at **recurring patterns** for classification of SARS-CoV-2





# ➤ Filter activations



INRAE

CONVOLUTIONAL NEURAL NETWORK FOR THE CLASSIFICATION OF SARS-CoV-2 AND PRIMER CREATION

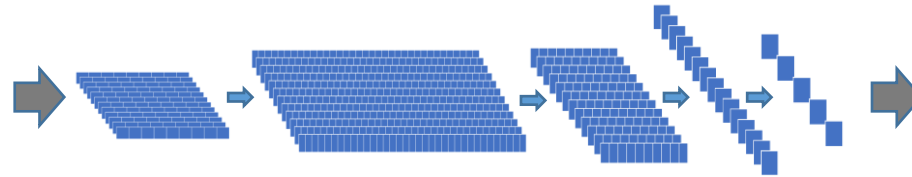
Alberto TONDA, Team EKINOCS, UMR 518 MIA-PS, INRAE, Université Paris-Saclay



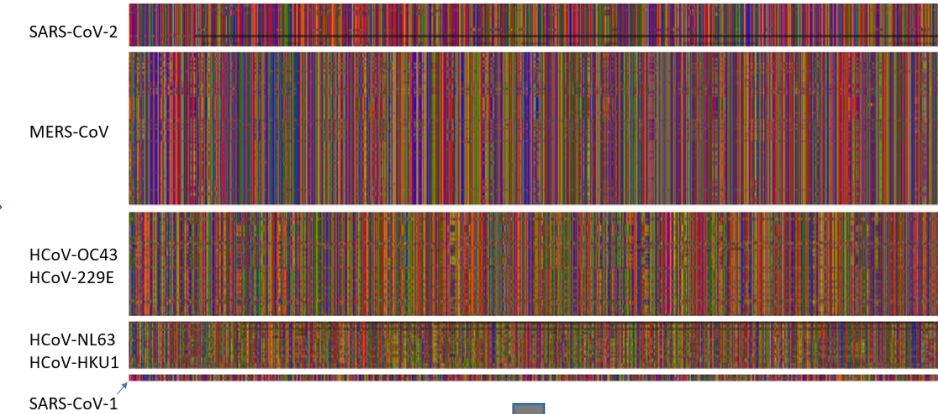
# ➤ Extracting information from filters

```
GATTTAAGTGAGATAGCCAAGCTATTCACTTCCCCTCGTTCTCTGCAGAACT
AAATAAAGCCCTGTTGTTTAGCGTATTGTTGCACCTGTCTGGTGGGATTGTG
CATCTAGGCAGTGGACATATGCTCAACACTGGGTATAATTCTAATTGAATACT
TCGTGTCTCTGTACGTCTCGGTCAACAATACGGTTTCGTCCGGTGCCTGGC
GTCTTTCGTGGCTGGTGTGATCGCGCAAGGTGCGCGCGGTACGTATCGAGCAG
```

Sequencing



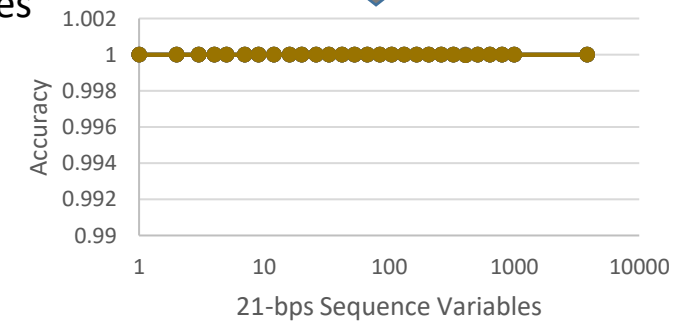
Apply Deep Learning (CNN)



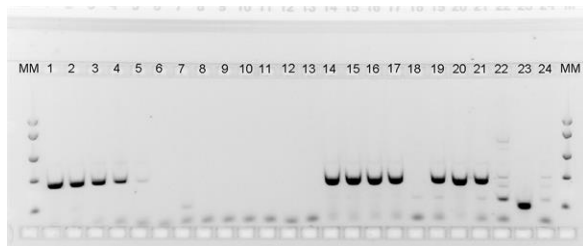
CACGTAGGAATGTCGAACTT  
TGAGCAGTGCCTGACTCAAATC  
CAACTTTTAAACGTACCAATGG  
CTAAAGCATACAATGTAACAC

Feature  
Creation

Get SARS-CoV-2 specific  
bps sequences



Feature Selection



Laboratory Testing

25401	GTTTATGAGA	ATCTTCACAA	TTGGAAGTGT	AACTTTGAAG	CAAGGTGAAA
25451	TCAAGGATGC	TACTCCTTCA	GATTTTGTTC	GCGCTACTGC	AACGATACCG
25501	ATACAAGCCT	CACTCCCTTT	CGGATGGCTT	ATTGTTGGCG	TTGCACTTCT
25551	TGCTGTTTTT	CAGAGCGCTT	CCAAAATCAT	AACCTCRAA	AAGAGATGGC
25601	AACTAGCACT	CTCCAAGGGT	GTTCACTTTG	TTTGCAACTT	GCTGTTGTTG
25651	TTTGTAAACAG	TTTACTCACA	CCTTTTGCTC	GTTGCTGCTG	GCCTTGAAGC
25701	CCCTTTTCTC	TATCTTTATG	CTTTAGTCTA	CTTCTGTCAG	AGTATAAACT
25751	TTGTAAGAAT	AATAATGAGG	CTTTGGCTTT	GCTGGAATG	CCGTTCCAAA
25801	AACCCATTAC	TTTATGATGC	CAACTATTTT	CTTTGCTGGC	ATACTAATTG
25851	TTACGACTAT	TGTATACCTT	ACAATAGTGT	AACTTCTTCA	ATTGTCATTA
25901	CTTCAGGTGA	TGGCACAAAC	AGTCCTATTT	CTGAACATGA	CTACCAGATT

Primer Design

INRAE

CONVOLUTIONAL NEURAL NETWORK FOR THE CLASSIFICATION OF SARS-CoV-2 AND PRIMER CREATION

Alberto TONDA, Team EKINOCs, UMR 518 MIA-PS, INRAE, Université Paris-Saclay

# ➤ Great success!

- It works!

## ➤ Great success...?

- Later, same results obtained with simpler methods
  - But they required more thinking and expert knowledge
  - ML/DL can be good for first approach
  - It also works for SARS-CoV-2 variants!


### scientific reports

Explore content ▾ About the journal ▾ Publish with us ▾

[nature](#) > [scientific reports](#) > [articles](#) > article

Article | [Open access](#) | Published: 22 September 2023

### An Innovative AI-based primer design tool for precise and accurate detection of SARS-CoV-2 variants of concern

[Carmina Angelica Perez-Romero](#), [Lucero Mendoza-Maldonado](#), [Alberto Tonda](#), [Etienne Coz](#), [Patrick Tabeling](#), [Jessica Vanhomwegen](#), [John MacSharry](#), [Joanna Szafran](#), [Lucina Bobadilla-Morales](#), [Alfredo Corona-Rivera](#), [Eric Claassen](#), [Johan Garssen](#), [Aletta D. Kraneveld](#) & [Alejandro Lopez-Rincon](#) 

[Scientific Reports](#) **13**, Article number: 15782 (2023) | [Cite this article](#)

1750 Accesses | 1 Citations | 4 Altmeter | [Metrics](#)



**INRAE**

CONVOLUTIONAL NEURAL NETWORK FOR THE CLASSIFICATION OF SARS-CoV-2 AND PRIMER CREATION

Alberto TONDA, Team EKINOCs, UMR 518 MIA-PS, INRAE, Université Paris-Saclay



## ➤ Questions?

### Bibliography

- Lopez-Rincon, A., Tonda, A., Mendoza-Maldonado, L. et al. Classification and specific primer design for accurate detection of SARS-CoV-2 using deep learning. Sci Rep 11, 947 (2021).  
<https://doi.org/10.1038/s41598-020-80363-5>
- Lopez Rincon A., ..., and Tonda, A. 2021. Design of specific primer sets for SARS-CoV-2 variants using evolutionary algorithms. In Proceedings of the Genetic and Evolutionary Computation Conference (GECCO '21). Association for Computing Machinery, New York, NY, USA, 982–990.  
<https://doi.org/10.1145/3449639.3459359>
- Perez-Romero, C.A., Mendoza-Maldonado, L., Tonda, A. et al. An Innovative AI-based primer design tool for precise and accurate detection of SARS-CoV-2 variants of concern. Sci Rep 13, 15782 (2023).  
<https://doi.org/10.1038/s41598-023-42348-y>

Images and videos: unless otherwise stated, I stole them from the Internet. I hope they are not copyrighted, or that their use falls under the Fair Use clause, and if not, I am sorry. Please don't sue me.