Protein language models expose viral mimicry and immune escape

Dan Ofer and Michal Linial
The Hebrew University of Jerusalem


ddofer@gmail.com
@ddofer


michall@cc.huji.ac.il
@MichalLinial

TL;DR:

Viruses elude the immune system through molecular mimicry, adopting their hosts biophysical characteristics. We show that Protein Language Models (**PLMs**) can differentiate between human and viral proteins.

Understanding where the immune system and models makes mistakes could reveal viral immune escape mechanisms, and insights into immunogenicity.

We find the biological immune system makes mistakes similar to those of computational PLMs, and how to extract insights for these "adversarial" examples using interpretable multimodal error-analysis models.

Results: Human - Virus detection

Task: Is a protein of Human or Viral origin?

(Novel) Dataset: Non-redundant (UniRef90/50), reviewed, Human (18,418) and Vertebra-host virus (6,699) protein sequences

	Model[a]	AUC	Accuracy	Precision	Recall	Log-Loss
	Length only	61.97	78.50	78.50	78.50	0.52
	AA n-grams	91.95	88.49	88.49	88.49	0.28
qLORA Finetuned	ESM2 8M	98.09	94.72	92.15	92.33	0.20
	ESM2 150M	99.26	96.99	95.54	95.48	0.12
	ESM2 650M	**99.67**	**97.86**	**96.85**	96.68	0.09
Static embedding	Linear-T5	99.56	97.57	97.57	97.57	**0.06**
	Tree-T5	99.65	97.7	97.7	**97.70**	**0.06**

[a]AA, Amino Acids; T5, T5 model embeddings from UniProt, used as ML input features

Insights: What characterizes mistakes?

Features of Errors

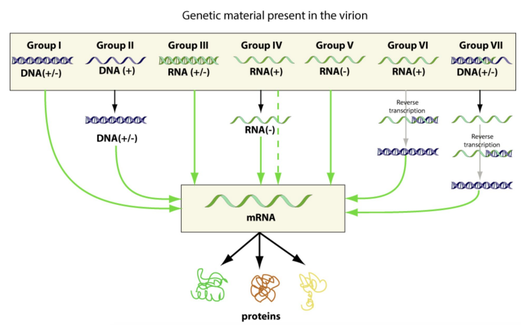
Features	Mistake rate (%)	Lift (x)[a]
"Adaptive immune" keyword	60.5	15.5
Endogenous retrovirus	30	7.7
Oncogene keyword	19.3	4.9
Sequence length < 170	12.1	3.1
Virus	9.4	2.4
Name: "putative"	8.7	2.2
Few keywords (< 8)	8.8	2.2

[a]Lift - frequency of mistakes relative to overall (3.9%)

- Viruses are disproportionately misclassified: 9.48% of viral proteins vs 1.87% humans'
- Endogenous retroviruses are viral sequences embedded in the human genome
- Non-human hosts are easier to classify (evolutionary distance from humans)

Features identified using an autoML model to classify PLM's mistakes, with annotation metadata as inputs (annotations, taxonomy etc)

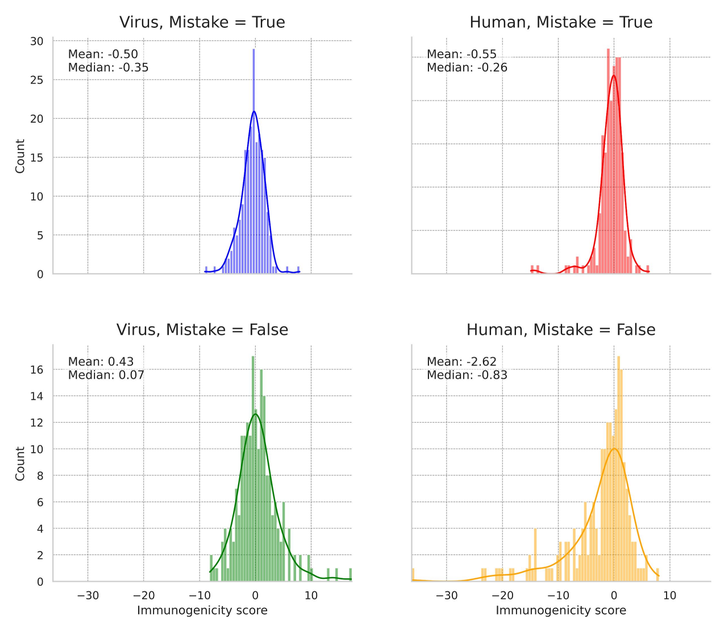
Characteristics of "Escaper" Virii



Genetic materials	Mistake rate (%)	# Proteins
dsDNA-**RT**	34.2	108
ssRNA-**RT**	19.5	666
ssDNA	13.1	129
dsDNA	8.2	4421
ssRNA	8.1	1017
dsRNA	0.8	358

- There are large differences in model mistake rates between major viral genera and families
- Prominently viruses notable for evading the **biological immune system** (immune evasion**)**, that cause long-term or **life-long** diseases including HIV (AIDS), HPV (Papilloma), hepatitis E (liver disease), Orthoherpesviridae (Herpes) and more.

Immunogenicity profiles



- IEDB Immunogenicity scores = propensity of a sequence to elicit an immune response
- Viruses are more immunogenic
- Mistakes are less immunogenic, for viruses AND humans
- The mistakes' immunogenicity profiles are more similar to *each other* than with their ground truth counterparts

Related Work:

"Detecting anomalous proteins using deep representations", NAR Genomics and Bioinformatics, 2024.

Used unsupervised anomaly detection. Tasks included:

- Human/virus - used different data partition, 98.5 rocAUC
- Prions (and novel prion like candidates)
- 3D Disorder and domain segmentation

Conclusions & Future Work

- Protein Language Models are effective, but not perfect at human-virus protein classification
- PLMs make mistakes similar to the biological immune system
- Immunogenic proteins and Viruses adept at evading the immune system have identifiable characteristics
- Multimodal error analysis is an effective DL understanding approach

Code: github.com/ddofer/ProteinHumVir