Artificial intelligence in genomics medicine

Francisco Requena

Clinical Bioinformatics lab - Imagine Institute

September 24, 2021

Université de Paris

imagine
INSTITUT DES MALADIES GÉNÉTIQUES

Outline

Introduction

- Motivation
- Challenges

Real impact

- CADD score
- FDA-approved AI/ML medical devices

Clinical Bioinformatics Lab

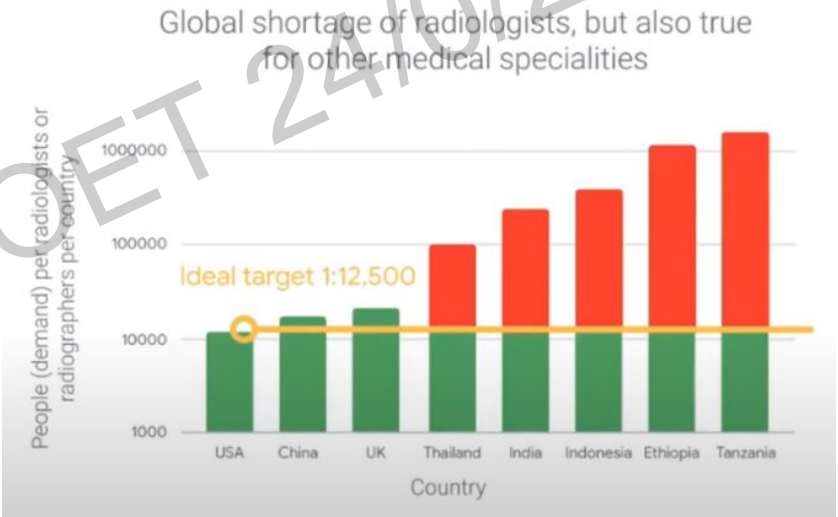
- Single Nucleotide Variants (SNVs)
  - **NCBoost**
- Copy Number Variants (CNVs)
  - **CNVscore**

In the context of patients with rare diseases

Introduction

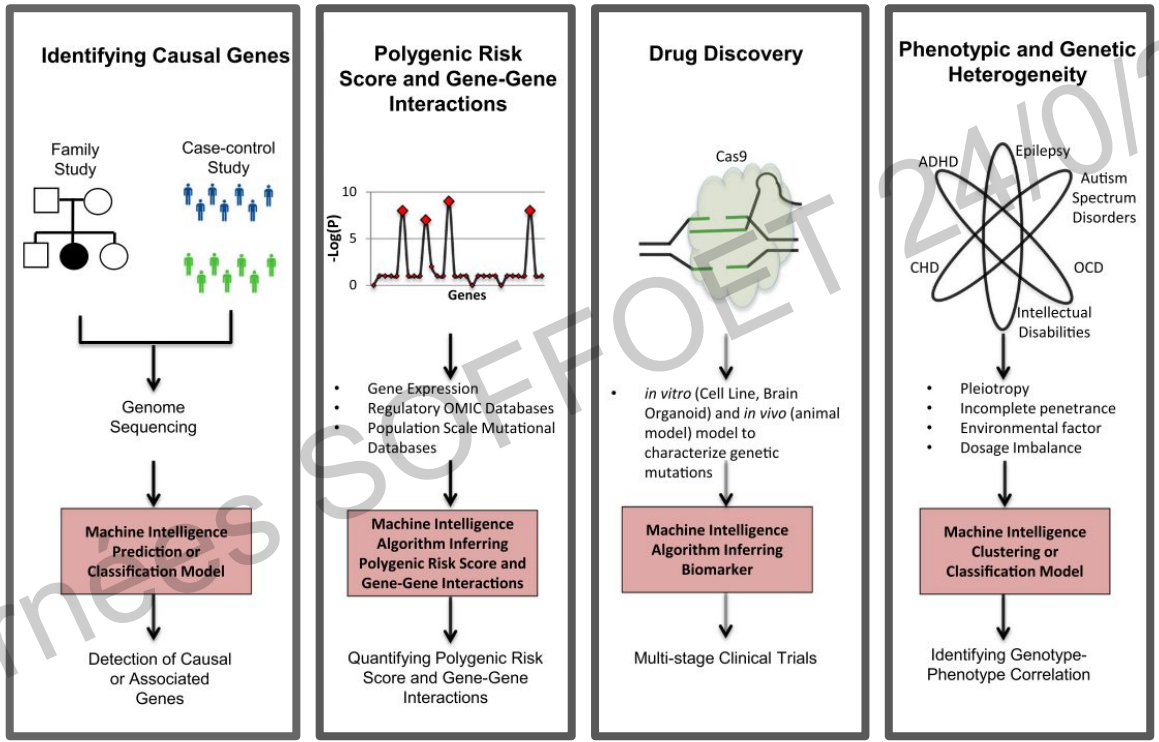
Motivation: The expanding scale and inherent complexity of biological data

- Huge amount of biomedical data
- Human subjectivity
- Lack of expertise



Silvestrin, C. 2016. 1 EUROPE'S LOOMING RADIOLOGY CAPACITY CHALLENGE A COMPARATIVE STUDY C. SILVESTRIN

Precision medicine and artificial intelligence



Uddin, M., Wang, Y. and Woodbury-Smith, M., 2019. Artificial intelligence for precision medicine in neurodevelopmental disorders. *NPJ digital medicine*, *2*(1), pp.1-10.

Identification of disease-causing variants

Supervised machine learning for the identification of pathogenic variants

- Supervised learning
- Unsupervised learning
- Reinforcement learning
- ...

Supervised machine learning for the identification of pathogenic variants

Supervised learning

Unsupervised learning

Reinforcement learning

...

Identification of pathogenic variants

Pathogenic variants

Non-pathogenic variants

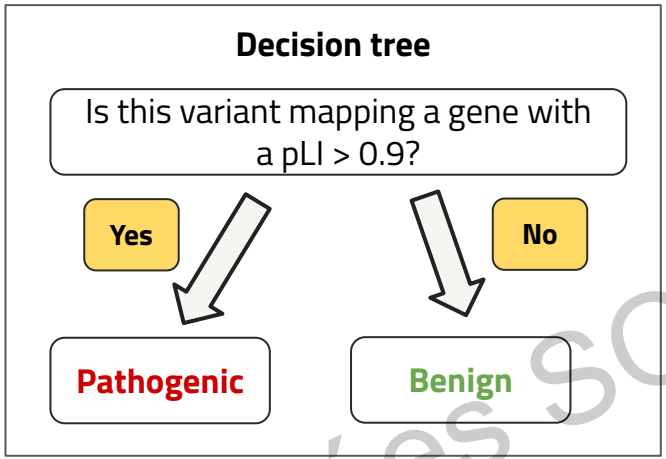
Annotation

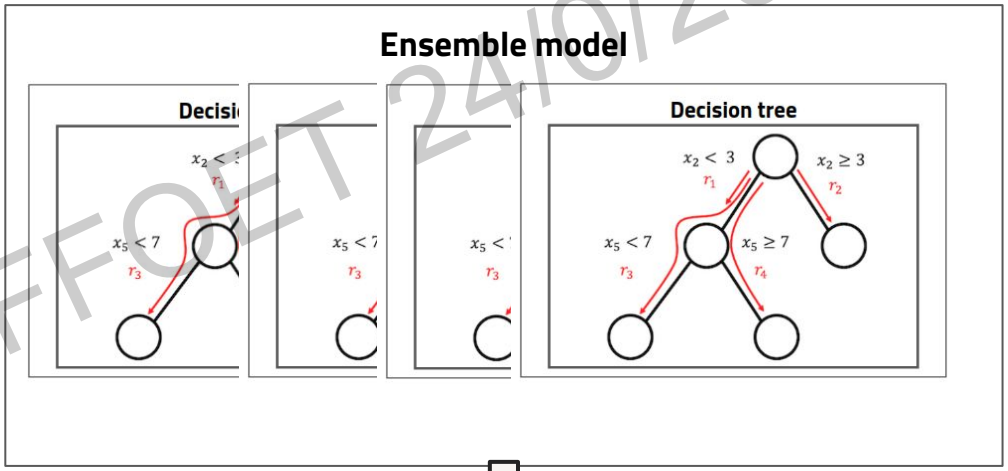
Features

Training

ML model

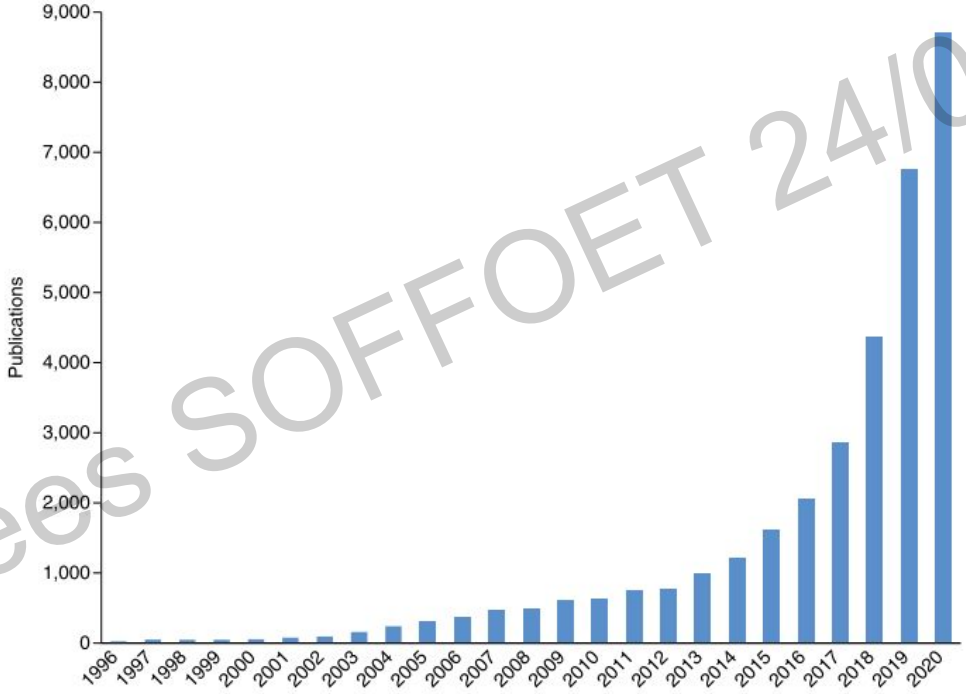
Decision trees are the main element of many ML models





Random Forests, XGBoost...

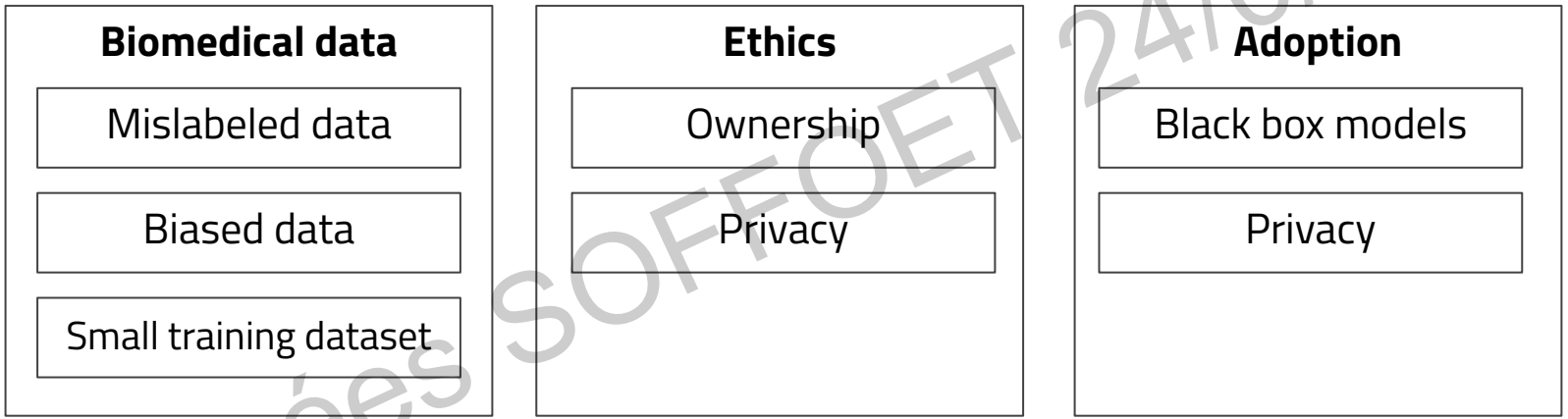
Exponential increase of ML publications in biology



Walsh, Ian, et al. "DOME: recommendations for supervised machine learning validation in biology." *Nature Methods* (2021): 1-6.

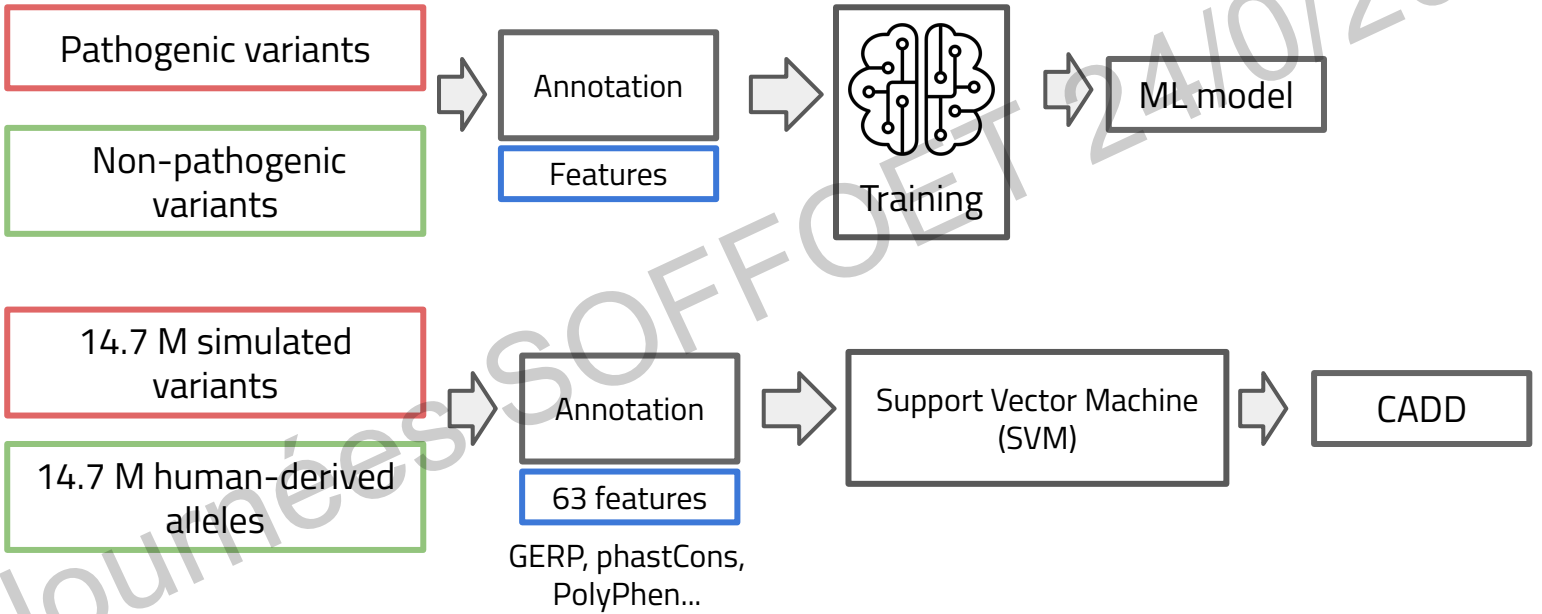
High expectations

Challenges in Machine Learning for Health



Real-world examples

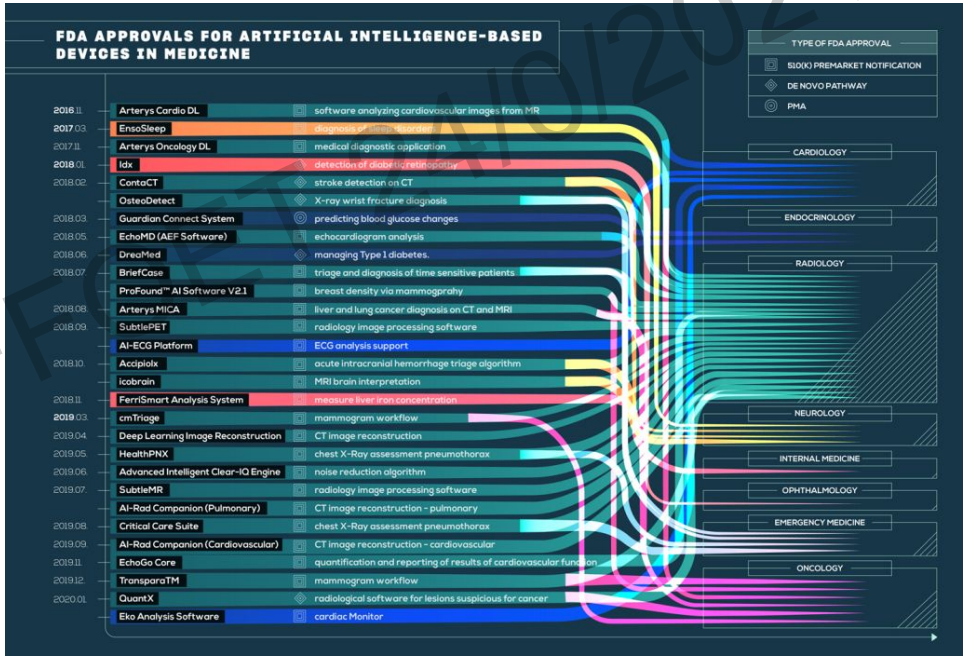
CADD: ML-based score broadly used for the identification of pathogenic variants



- Kircher, Martin, et al. "A general framework for estimating the relative pathogenicity of human genetic variants." Nature genetics 46.3 (2014): 310-315.

64 FDA-approved AI/ML medical devices

- **Radiology** - 30 (46.9%)
- **Cardiology** - 16 (25.0%)
- **Internal Medicine/General** - 10 (15.6%)



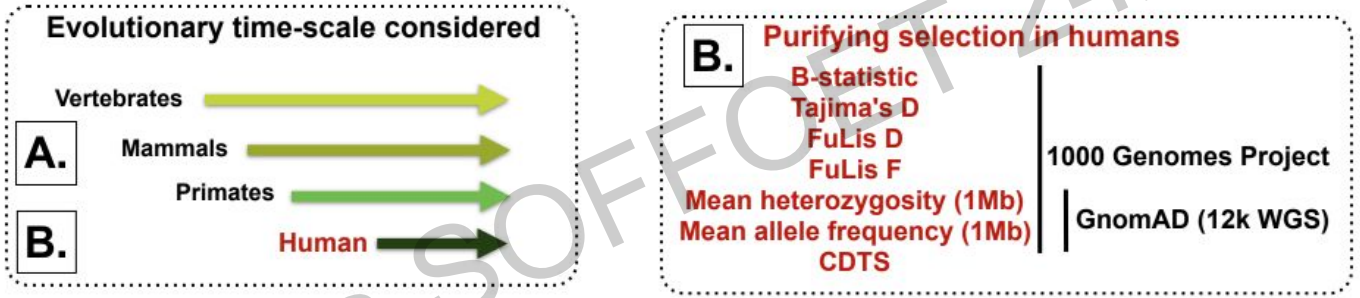
Benjamens, S., Dhunnoo, P. and Meskó, B., 2020. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *NPJ digital medicine*, *3*(1), pp.1-8.

NCBoost

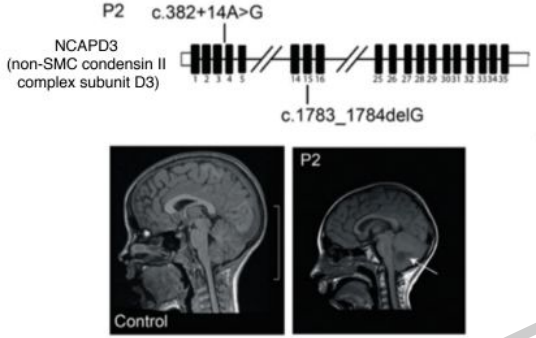
Non-coding Single Nucleotide Variants (SNVs)

Caron, Barthélémy, Yufei Luo, and Antonio Rausell. "**NCBoost classifies pathogenic non-coding variants in Mendelian diseases through supervised learning on purifying selection signals in humans.**" *Genome biology* 20.1 (2019): 1-22.

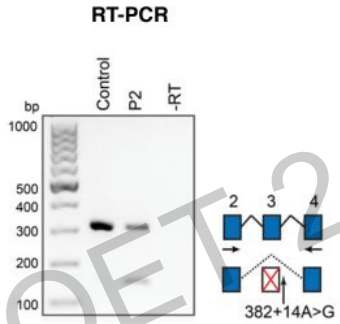
Extraction of a comprehensive set of purifying selection signals

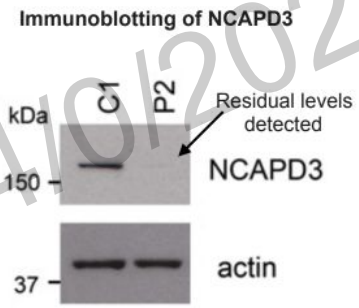


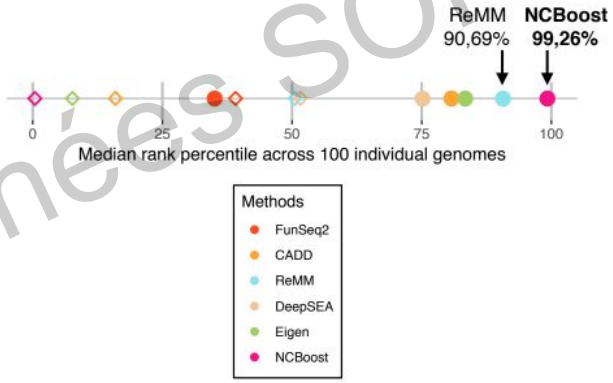
Case study: NCAPD3 intron variant associated with autosomal recessive primary microcephaly-22



Martin et al. Genes Dev. 2016; 30(19):2158–72







Evolutionary scale	Conservation score	Rank Percentile against common non-coding SNVs (%)
Vertebrates	verPhCons	41,45
Mammals	mamPhCons	42,82
Primates	priPhCons	84,65
Humans	meanMAF_1000G	**95,08**
	meanHet_1000G	**94,49**
	TajimasD_YRI_pval	**93,46**
	ncGERP	**95,17**

CNVscore

Copy number variants (CNVs)

Limitations of the clinical interpretation of CNVs

The French network of cytogeneticists and molecular geneticists (Achro-Puce) reported in 2019 a total of 16 993 aCGH tests prescribed to pediatric patient.

~11.8% pathogenic or likely pathogenic CNVs.

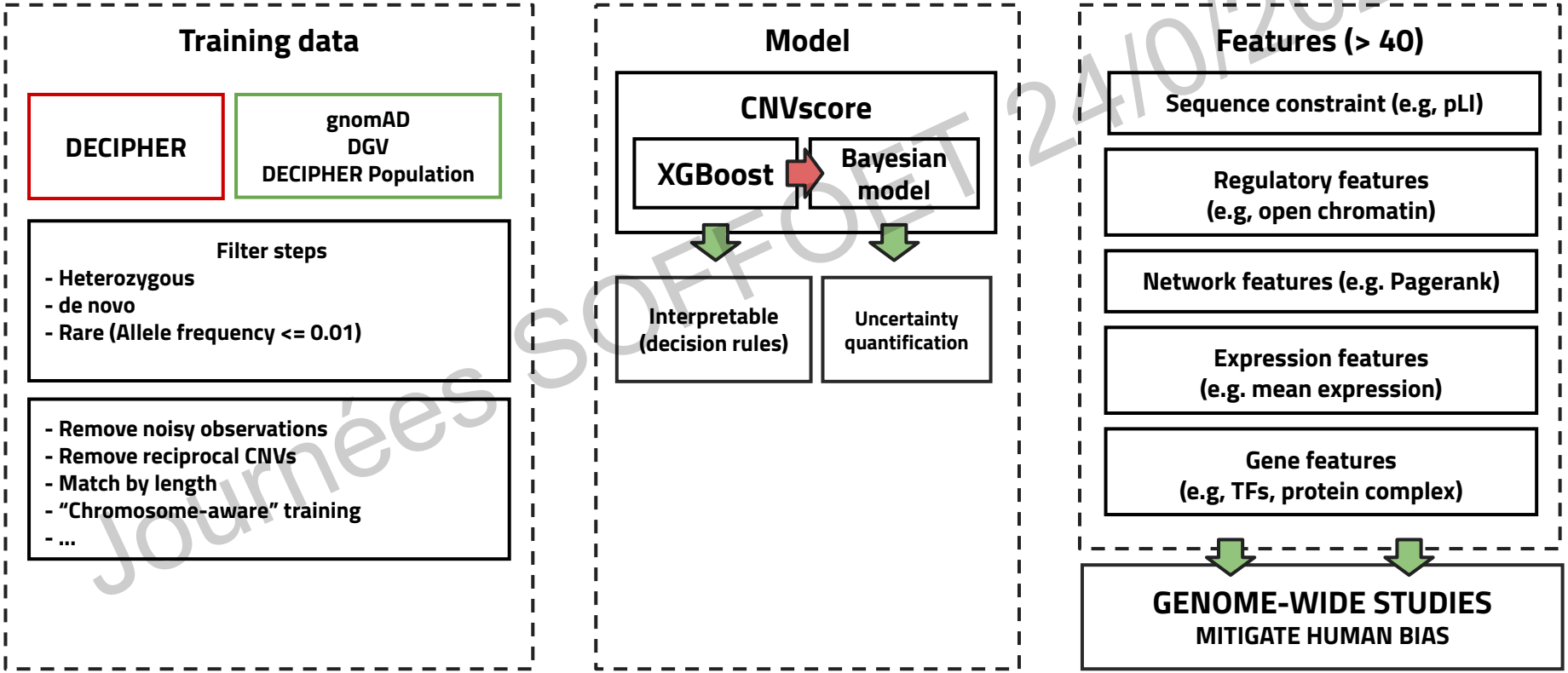
~8.2% variants of unknown significance (VUS).

Limitations of the clinical interpretation of CNVs

CNVs **NOT** mapping:

- Disease-associated genes
- Previously reported pathogenic CNVs

CNVscore: an interpretable machine-learning model for the identification of pathogenic CNVs



Summary

- Potential of artificial intelligence in health
- Challenges
- Examples
- NCBoost
- CNVscore

Acknowledgments

Barthélémy Caron
Francisco Requena
Antoine Favier Stefani
Dritsa
Yufei LUO
Members of the Clinical Bioinformatics lab

Devo-Decode Consortium / MSD Avenir

Chloé-Agathe Azencott (Paris MINESTech)

Fabienne Jabot-Hanin (Imagine)
Patrick Nitschké (Imagine)

Valéry Malan (Necker/Imagine) **Serge Romana** (Necker/Imagine) **Jean-Paul Bonnefont** (Necker)
Fédération de génétique médical GH-Necker
Vincent Cantagrel (Imagine)
Nicolas Garcelon (Imagine)