PREDICTING VARIANT DELETERIOUSNESS IN NON-HUMAN SPECIES: TAKING THE CADD APPROACH TO PIG

Christian Groß

WAGENINGEN UNIVERSITY
WAGENINGEN UR

TU Delft

BREED4FOOD

NWO | Applied and Engineering Sciences

1. OBJECTIVE

- ***Develop a method to assign a DELETERIOUSNESS SCORE to variants anywhere in LIVESTOCK GENOMES.***

GTTACTAGTACAT

GTTACTCGTACAT
GTTACTAGTATAT
GTAACTAGTACAT

P(deleterious)

0.15
0.87
0.01

Model Objective

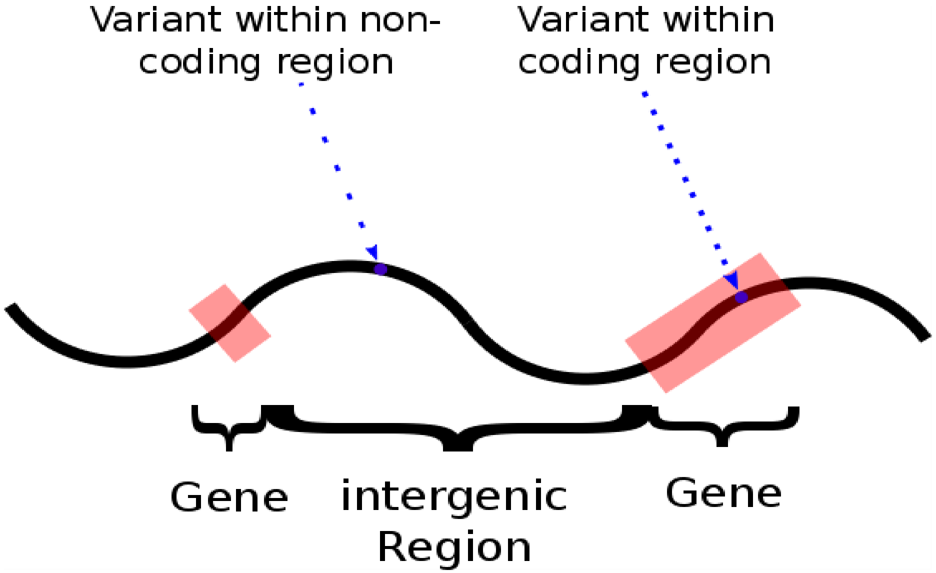
2. METHODOLOGY

Standing on the Shoulders of Giants

- ***Kircher et al., Nature Genetics***

2014

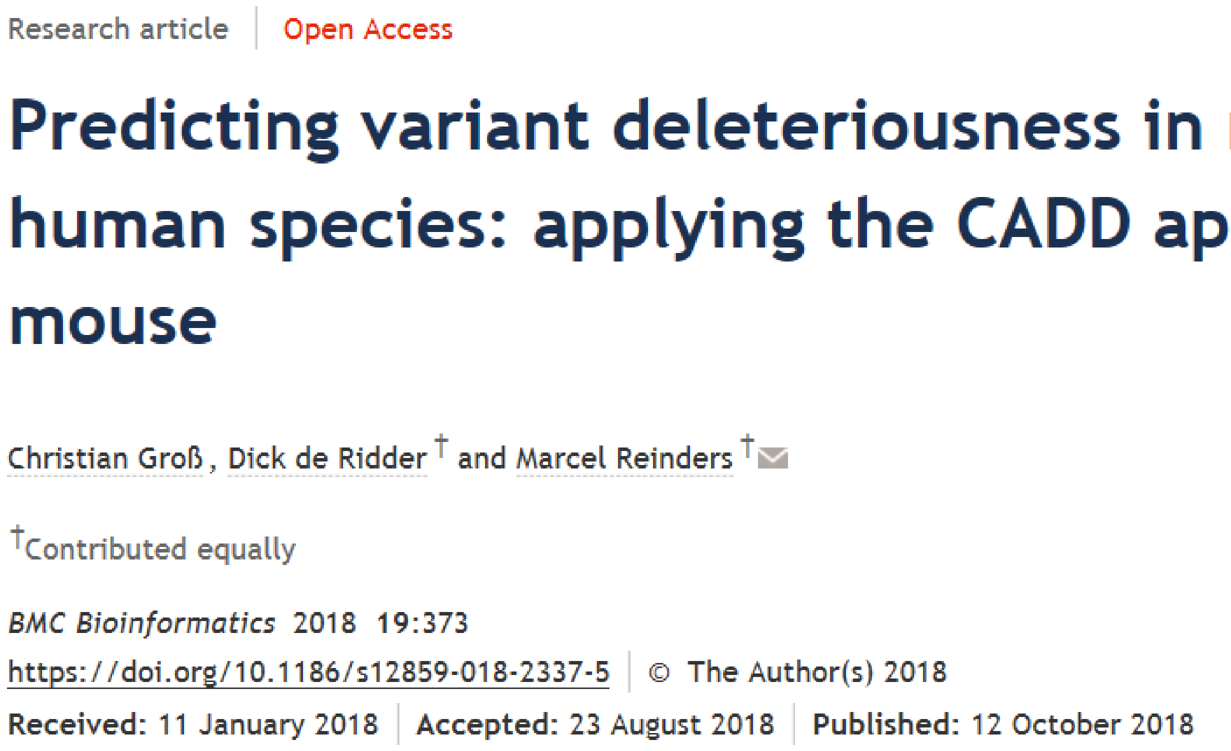
- ***Beyond SIFT, PROVEAN, PolyPhen etc.: one model, one comparable score for variants in coding and non-coding regions***


Variant within non-coding region
Variant within coding region
Gene
intergenic Region
Gene

Combined Annotation Dependent Depletion (CADD)

2. METHODOLOGY

Feasibility Study: mCADD



Research article | Open Access

Predicting variant deleteriousness in non-human species: applying the CADD approach in mouse

Christian Groß , Dick de Ridder [†] and Marcel Reinders [†] ✉

[†] Contributed equally

BMC Bioinformatics 2018 19:373

https://doi.org/10.1186/s12859-018-2337-5 | © The Author(s) 2018

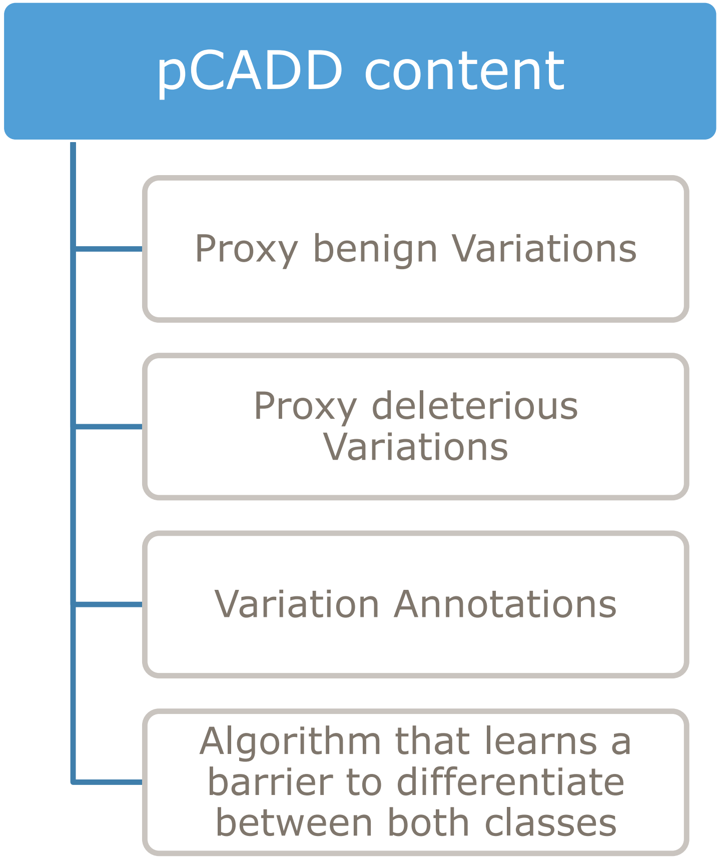
Received: 11 January 2018 | Accepted: 23 August 2018 | Published: 12 October 2018

2. METHODOLOGY

Next step: P(ig)-CADD

2. METHODOLOGY

pCADD – Model outline



pCADD content

Proxy benign Variations

Proxy deleterious Variations

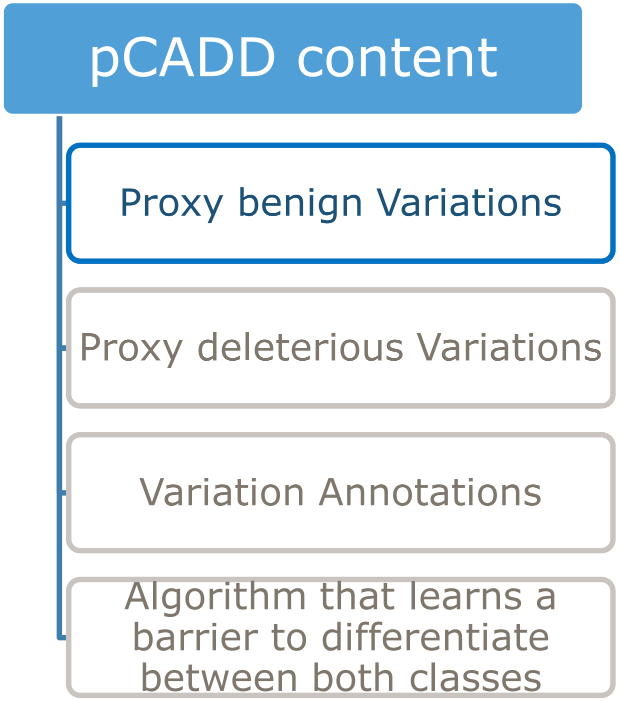
Variation Annotations

Algorithm that learns a barrier to differentiate between both classes

Model outline

2. METHODOLOGY

pCADD – proxy benign variations



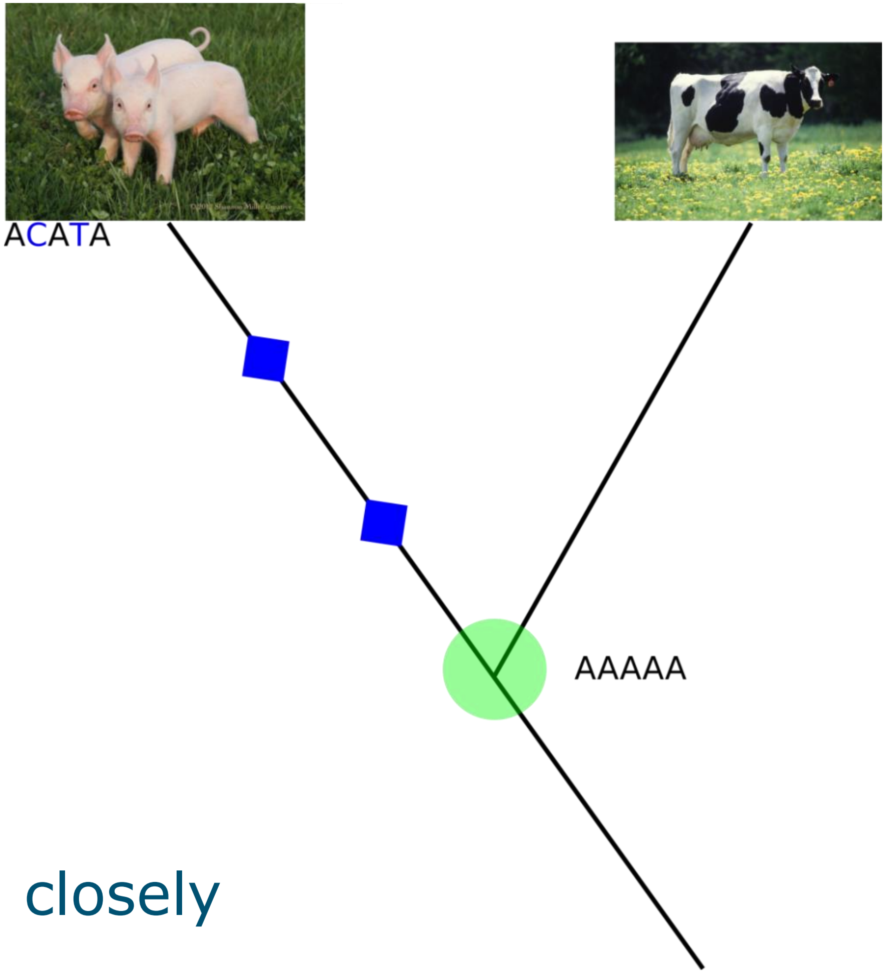
pCADD content

Proxy benign Variations

Proxy deleterious Variations

Variation Annotations

Algorithm that learns a barrier to differentiate between both classes

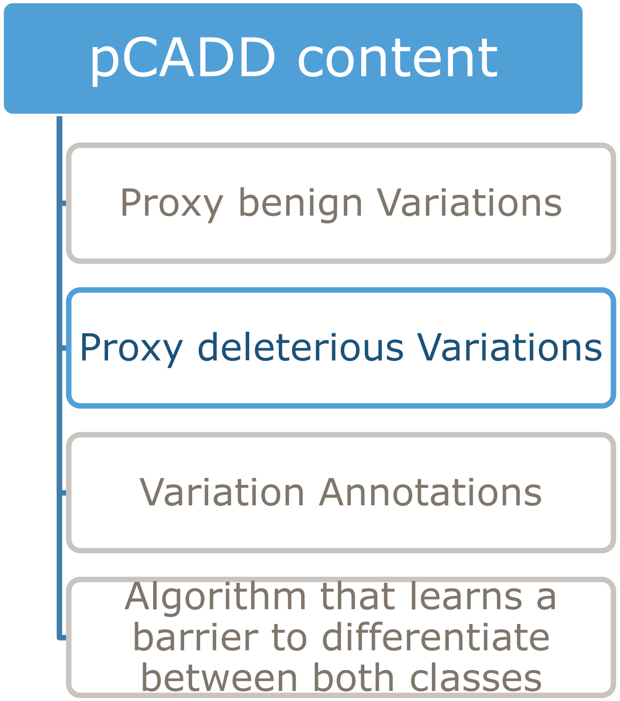


- Infer common ancestor with closely related species

Ancestor inference

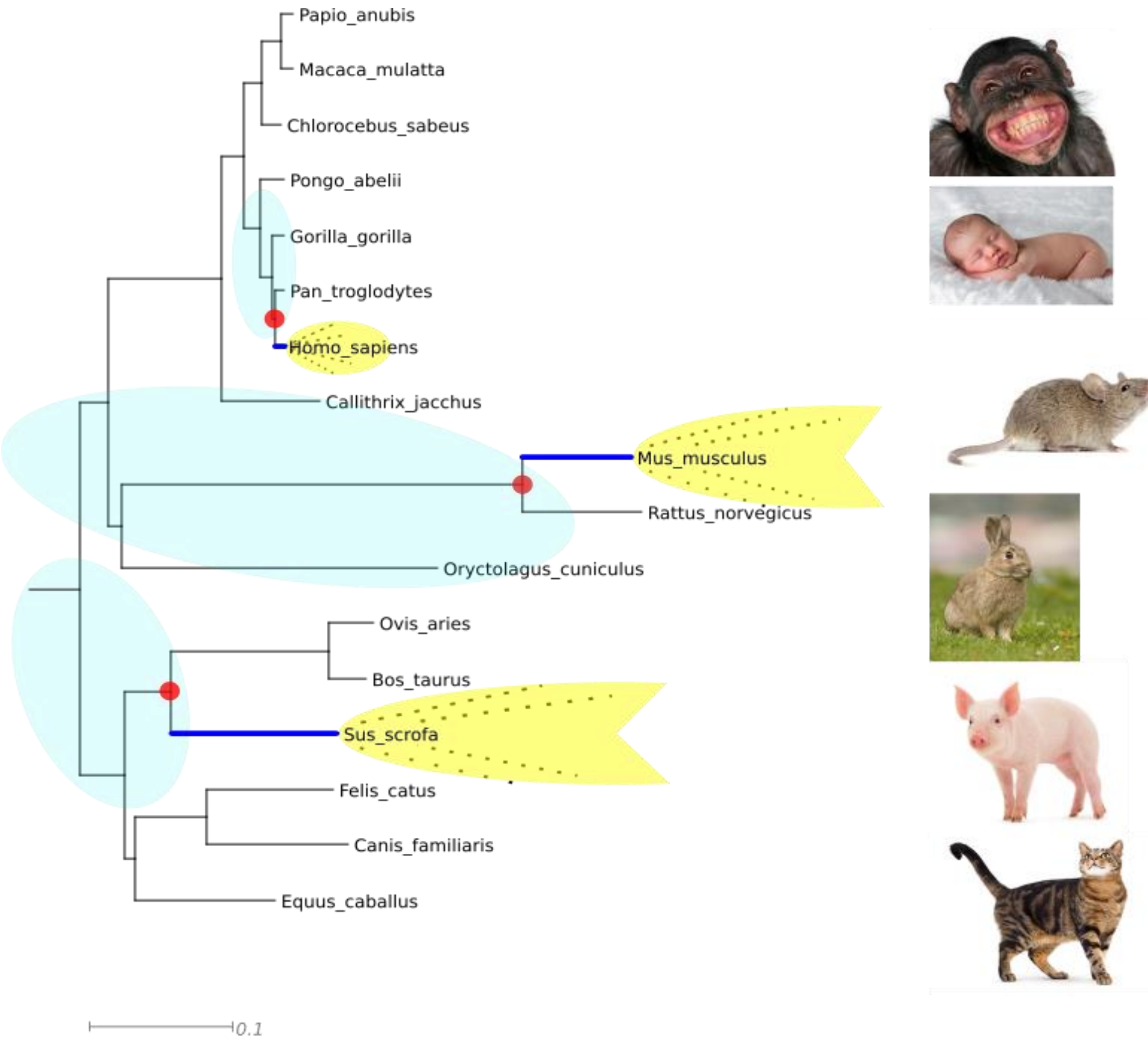
2. METHODOLOGY

pCADD – Simulating SNPs and their constraints



pCADD content

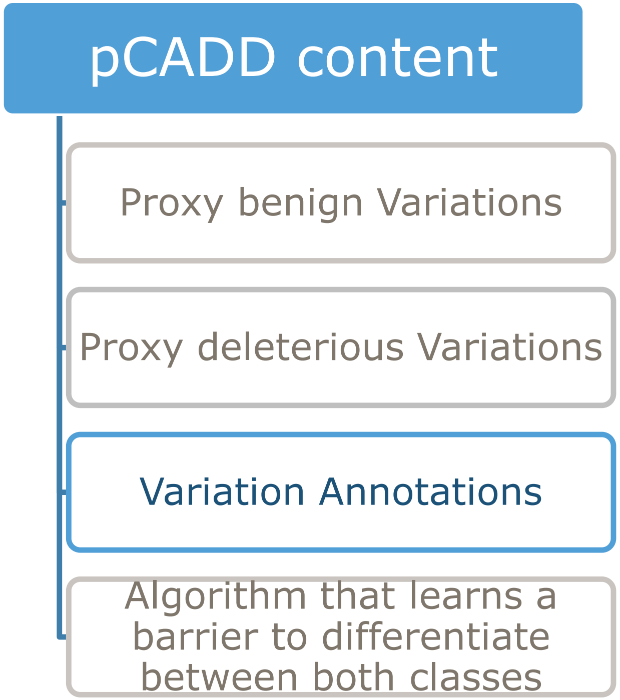
- Proxy benign Variations
- Proxy deleterious Variations
- Variation Annotations
- Algorithm that learns a barrier to differentiate between both classes



Deriving substitution rates

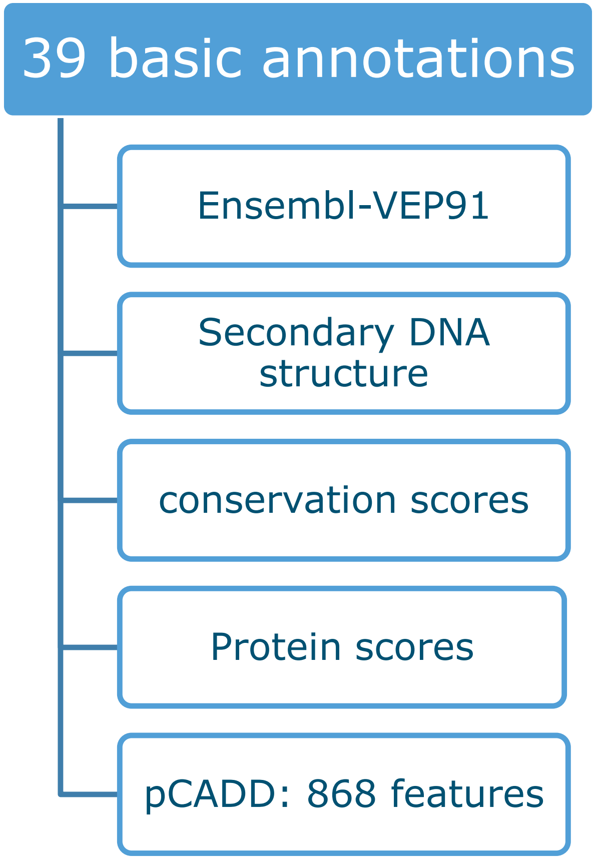
2. METHODOLOGY

pCADD – Variant annotations



pCADD content

- Proxy benign Variations
- Proxy deleterious Variations
- Variation Annotations
- Algorithm that learns a barrier to differentiate between both classes



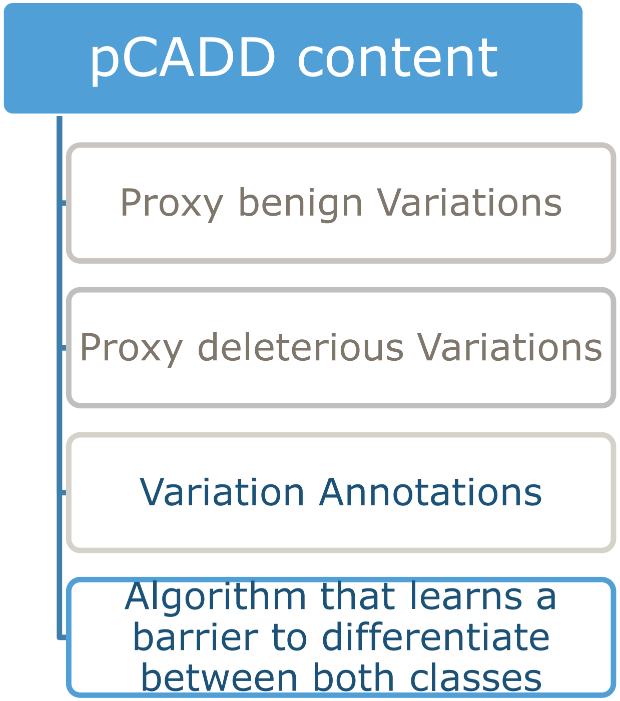
39 basic annotations

- Ensembl-VEP91
- Secondary DNA structure
- conservation scores
- Protein scores
- pCADD: 868 features

Annotation labels

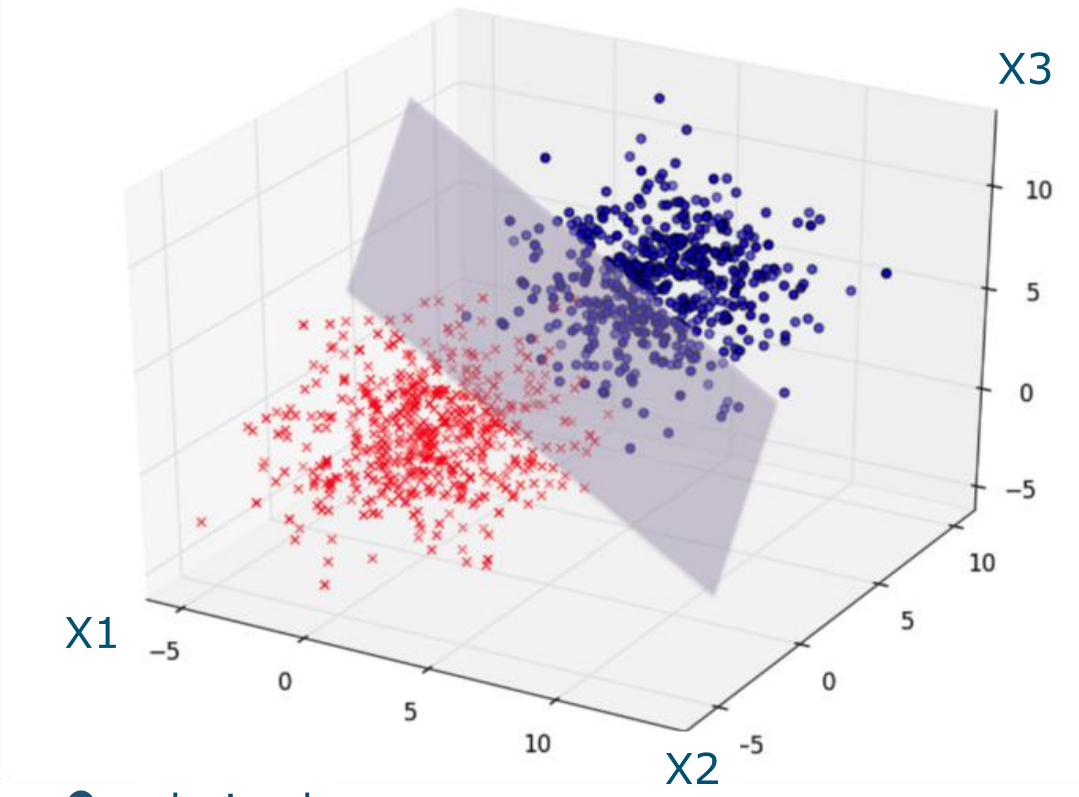
2. METHODOLOGY

pCADD – Generation of the Machine Learning Model



pCADD content

- Proxy benign Variations
- Proxy deleterious Variations
- Variation Annotations
- Algorithm that learns a barrier to differentiate between both classes



● derived

● simulated

▨ Decision boundary

Notes: *X(n)=feature(n)*

In this research more than 3 features were used

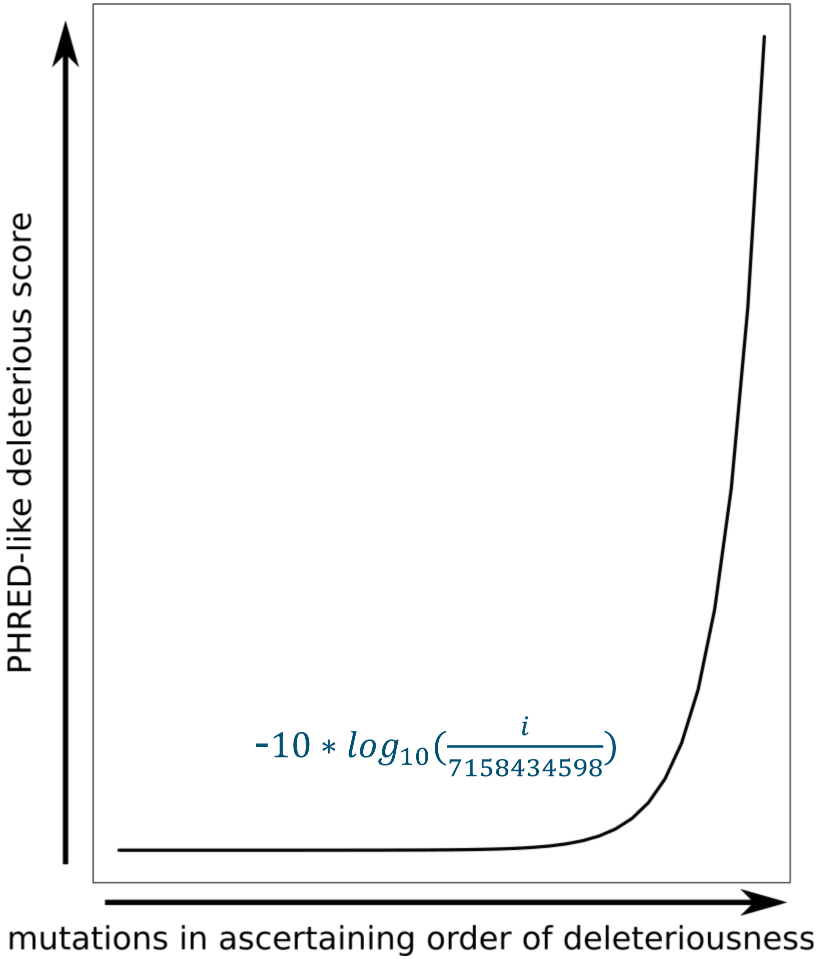
General representation of a Machine learning model

3. METHODOLOGY

pCADD Model Extension - PHRED-like scores

- ***All possible SNPs on chromosome 1-18 and X were generated and annotated (7,158,434,598).***
- ***SNPs were ranked with respect to their deleteriousness.***

Lowest 90%	Lowest 99%	Lowest 99.9%
• PHRED: 0-10	• PHRED: 0-20	• PHRED: 0-30



Hypothetical representation of PHRED-like score distribution

4. Results

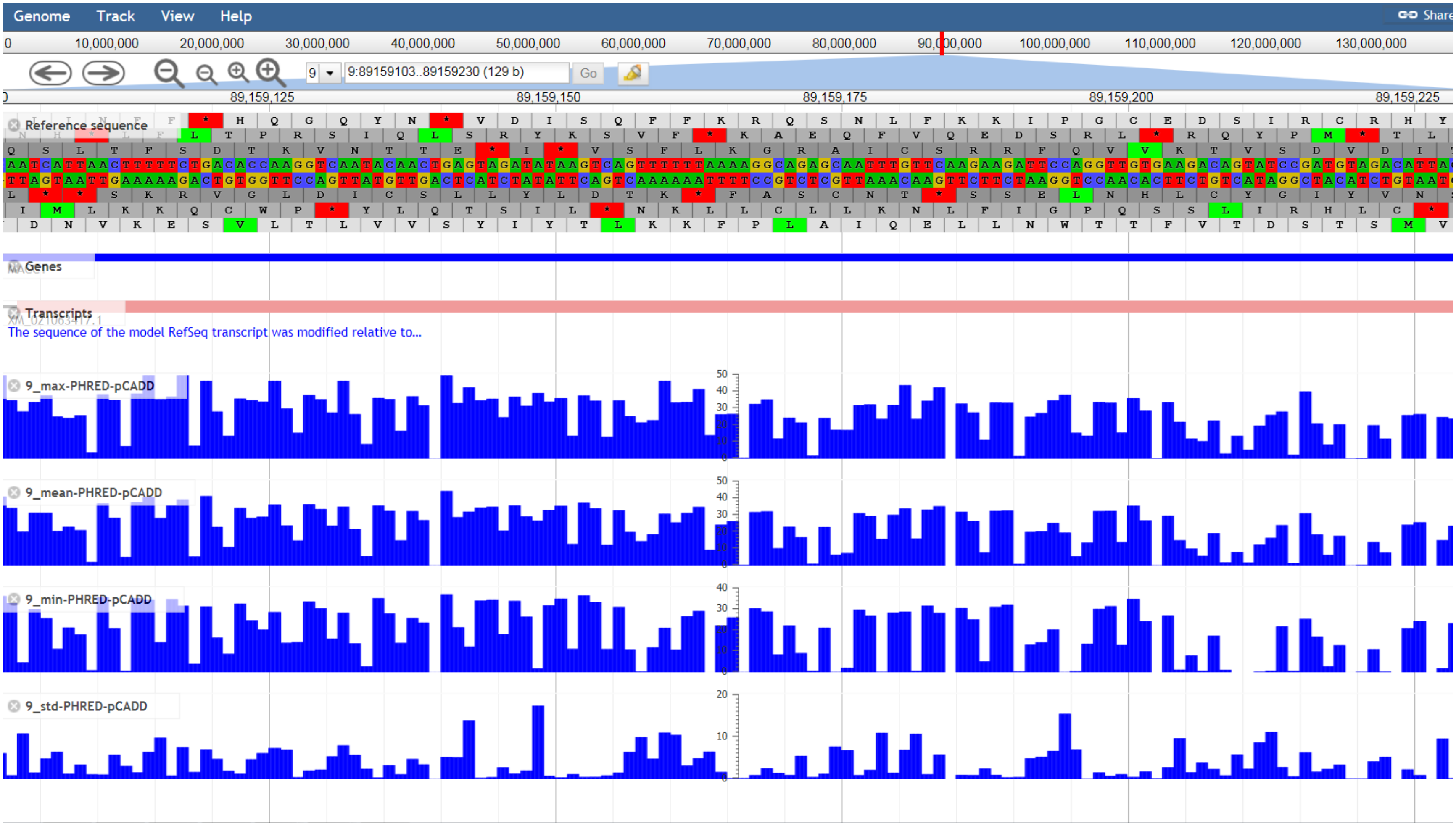
pCADD - Evaluating Known Deleterious Variants

Hap.	Type	SSC	Position	Ref	Alt	Gene	AA change (SIFT)	Raw-score	PHRED-score
DU1	Splice-donor	12	38,922,102	G	A	TADA2A	-	0.95885	21.88258
LA1	Splice-region	3	43,952,776	T	G	POLR1B	-	0.69472	10.14103
LA2	Frameshift	13	195,977,038	C	-	URB1	1961-V/X	NA	NA
LA3	Missense	6	54,880,241	T	C	PNKP	96-Q/R (0.02)	0.9967	29.46386

small set of known variants

4. Results

pCADD – JBrowser Implementation

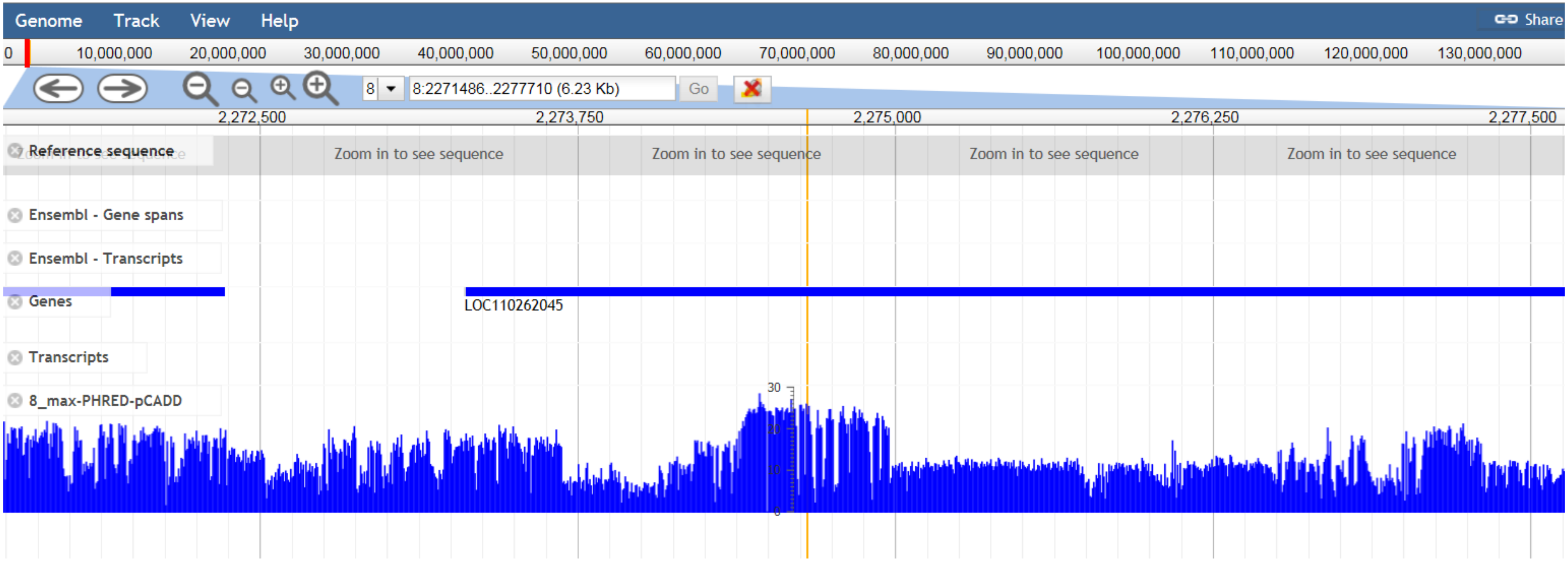


The sequence of the model RefSeq transcript was modified relative to...

MACC1 example of PHRED-scores

4. Results

pCADD – Identification of NCBI genebuild element



Intergenic high-impact, high frequent SNP

6. QUESTIONS?

People to Thank

TU Delft

- Christian Groß
- Marcel Reinders

BREED4FOOD

NWO | Applied and Engineering Sciences

WAGENINGEN UNIVERSITY
WAGENINGEN UR

WAGENINGEN UR
For quality of life

- Dick de Ridder
- Martijn Derks
- Mirte Bosse
- Hendrik-Jan Megens
- Martien Groenen