# Improve clinical diagnosis of rare genetic disorders with **GEMVAP** a **GE**ne-specific **M**issense **VA**riant **P**redictor framework

Pr Guillaume Smits, MD PhD

Marfan Meeting
5th of October 2019

# Missense variant interpretation is challenging

DNA HAS ALL YOU CAN ASK FOR .
DNA HAS ALL YOO CAN ASK FOR .

DNA HAS ALL YOU .
DNA HAS ALY OUC ANA SKF OR.

DNA HAS ALL LOU CAN ASK FOR .

# Predictors evaluate the pathogenicity of variants

|  | Predictor 1 | Predictor 2 | Predictor 3 |
|---|---|---|---|
| Variant 1 | Damaging | Pathogenic | Damaging |
| Variant 2 | Neutral | Pathogenic | Damaging |
| Variant 3 | Neutral | Benign | Neutral |

. . .

# From gene-agnostic variant predictors...



ROC curves for top predictors
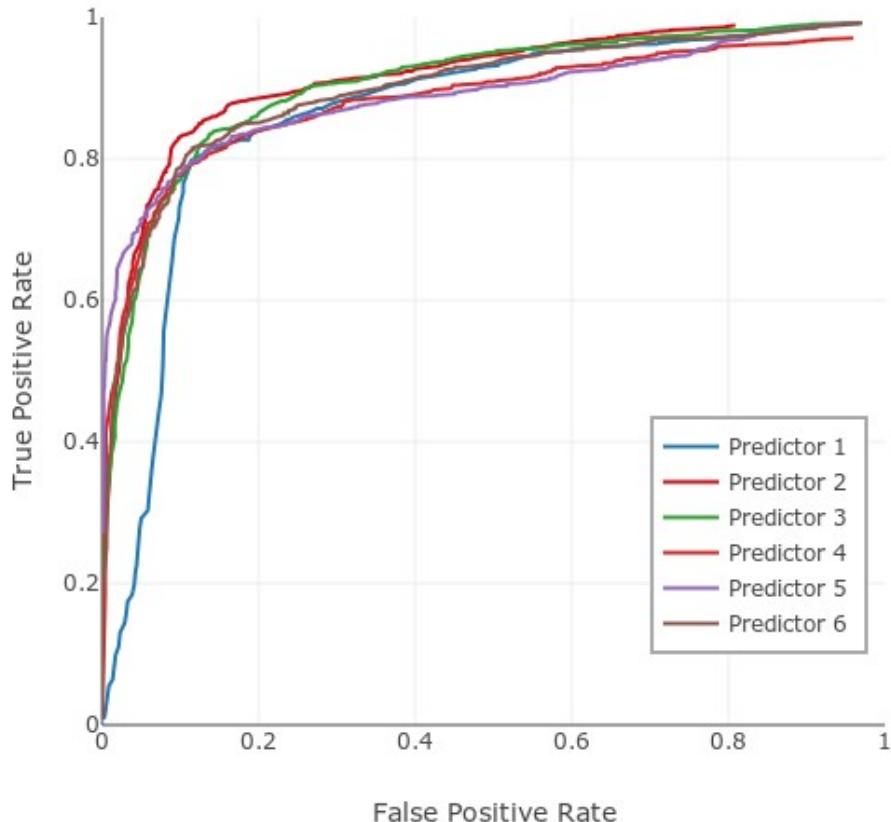
# …to a gene-specific variant predictor



ROC curves for top predictors: optimization
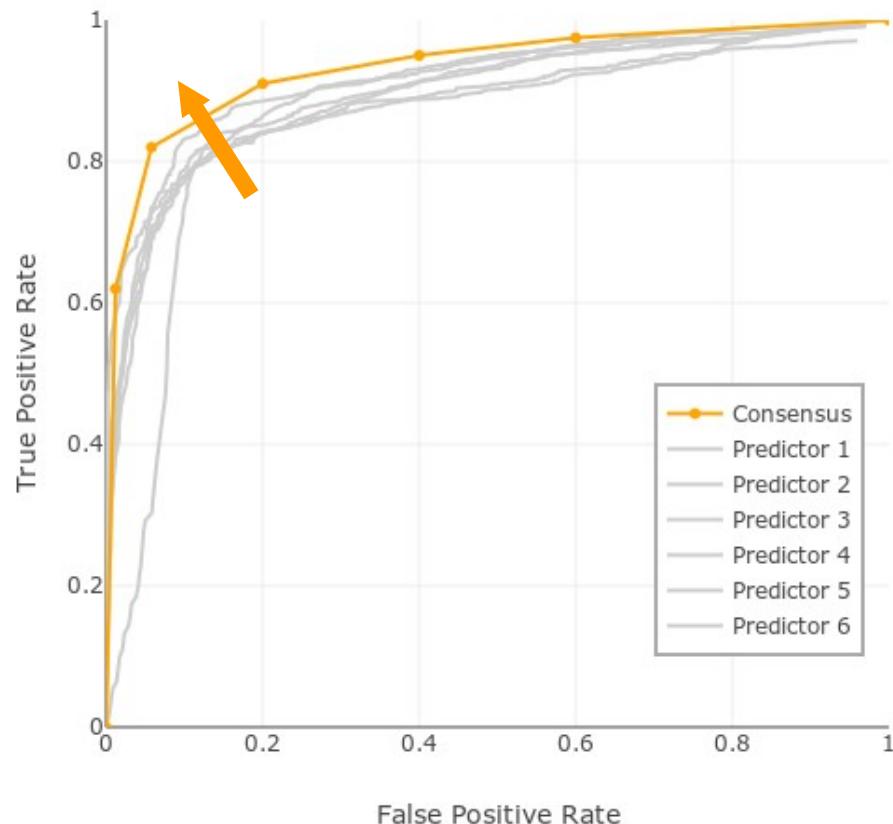
# The GEMVAP framework
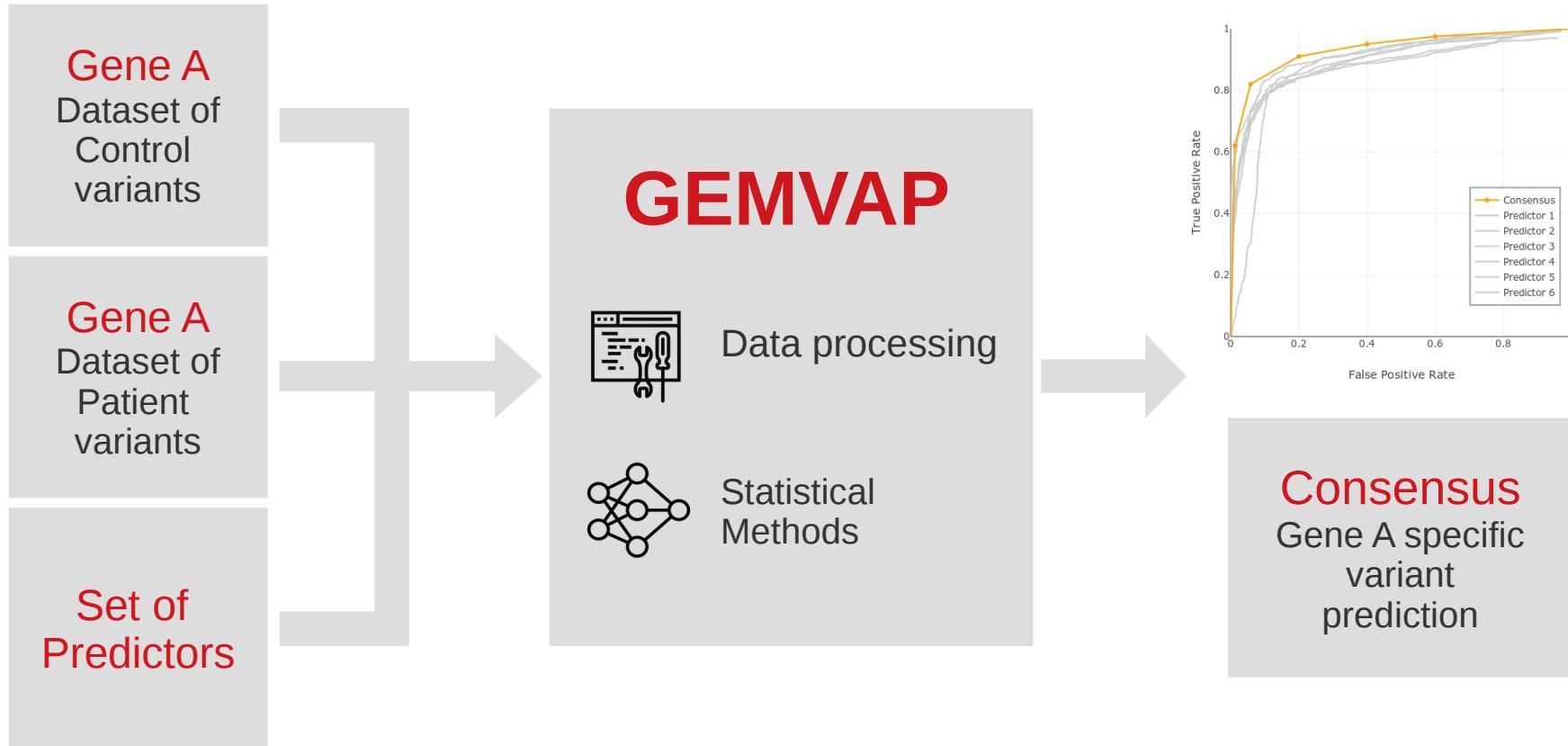
# Collection of FBN1 datasets

| | Pathogenic | Intersection | Control |
|---|---|---|---|
| Size | 1868 | 188 | 976 |
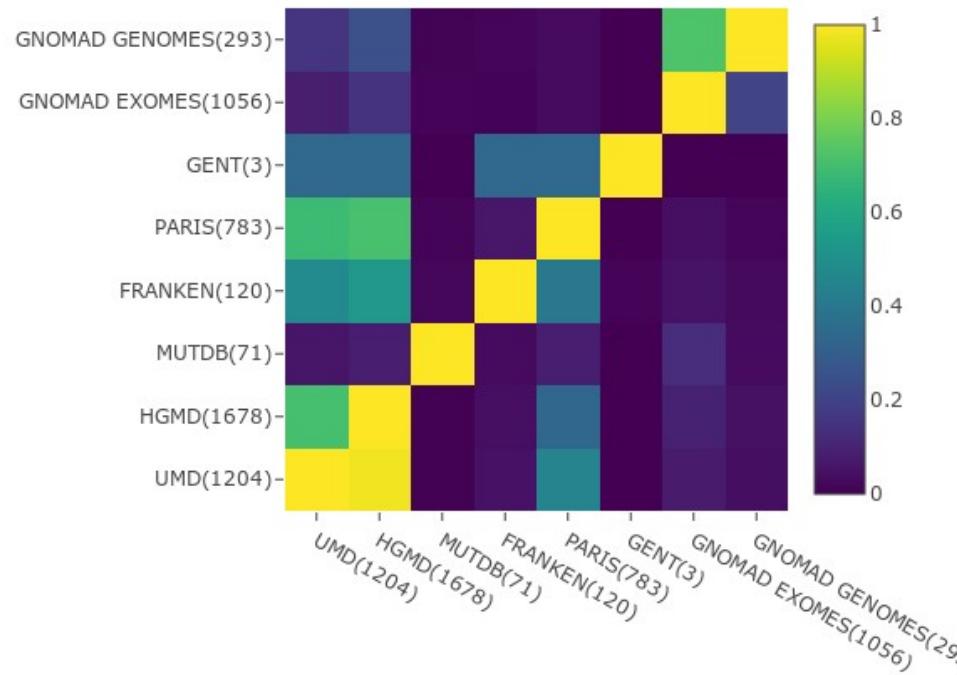
# Tolerance of FBN1 amino acids to mutation



Variants count per reference amino acid
(Training set with Cysteines)
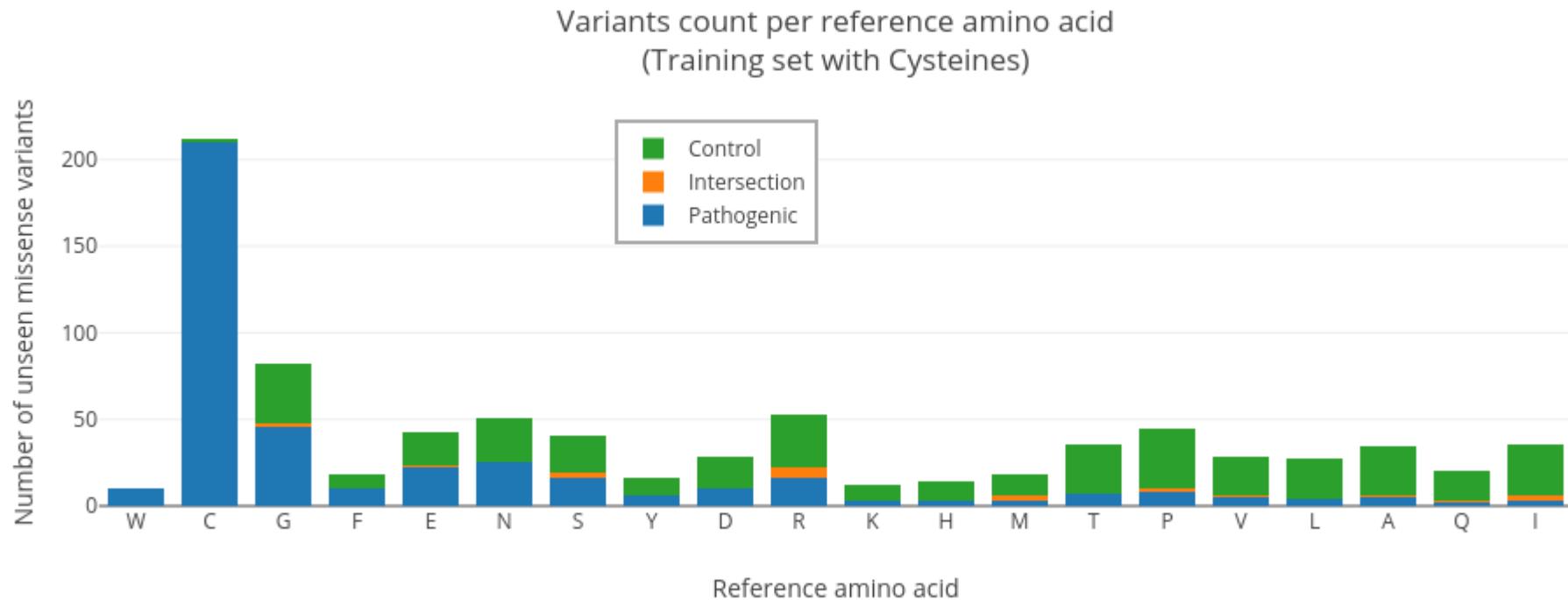
# TopN predictors for each training set

| Missense | Missense & Cysteine | Cysteine |
|---|---|---|
| ■ SIFT_converted_rankscore | ■ M-CAP_rankscore | ■ PROVEAN_converted_rankscore |
| ■ REVEL_rankscore | ■ MutationAssessor_rankscore | ■ M-CAP_rankscore |
| ■ MetaSVM_rankscore | ■ REVEL_rankscore | ■ MutationAssessor_rankscore |
| ■ M-CAP_rankscore | ■ MetaSVM_rankscore | |
| ■ MutationAssessor_rankscore | ■ SIFT_converted_rankscore | |
| | ■ MetaLR_rankscore | |
| | ■ MutPred_rankscore | |
| | ■ PROVEAN_converted_rankscore | |
| | ■ VEST4_rankscore | |
| | ■ SIFT4G_converted_rankscore | |
| | ■ DEOGEN2_rankscore | |
| | ■ Eigen-raw_coding_rankscore | |
| | ■ Eigen-PC-raw_coding_rankscore | |
| | ■ fathmm-XF_coding_rankscore | |
| | ■ FATHMM_converted_rankscore | |
| | ■ Polyphen2_HVAR_rankscore | |
| | ■ MutationTaster_converted_rankscore | |
| **5** | **17** | **3** |

# GEMVAP FBN1 top5 predictors model



Variant count per level of consensus
(pathogenic ROC-based threshold, unseen missense)
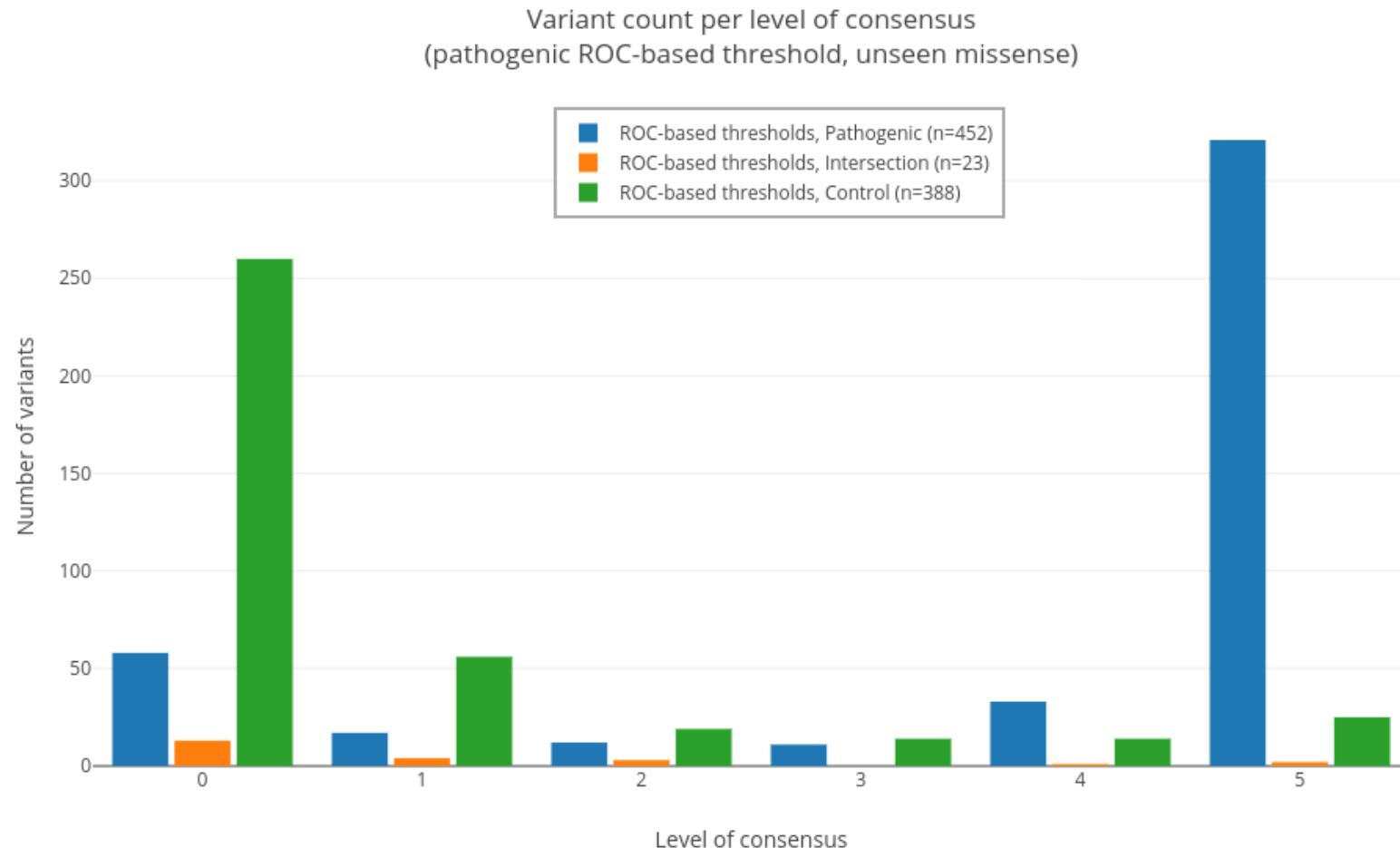
# GEMVAP FBN1 top5 whole gene prediction
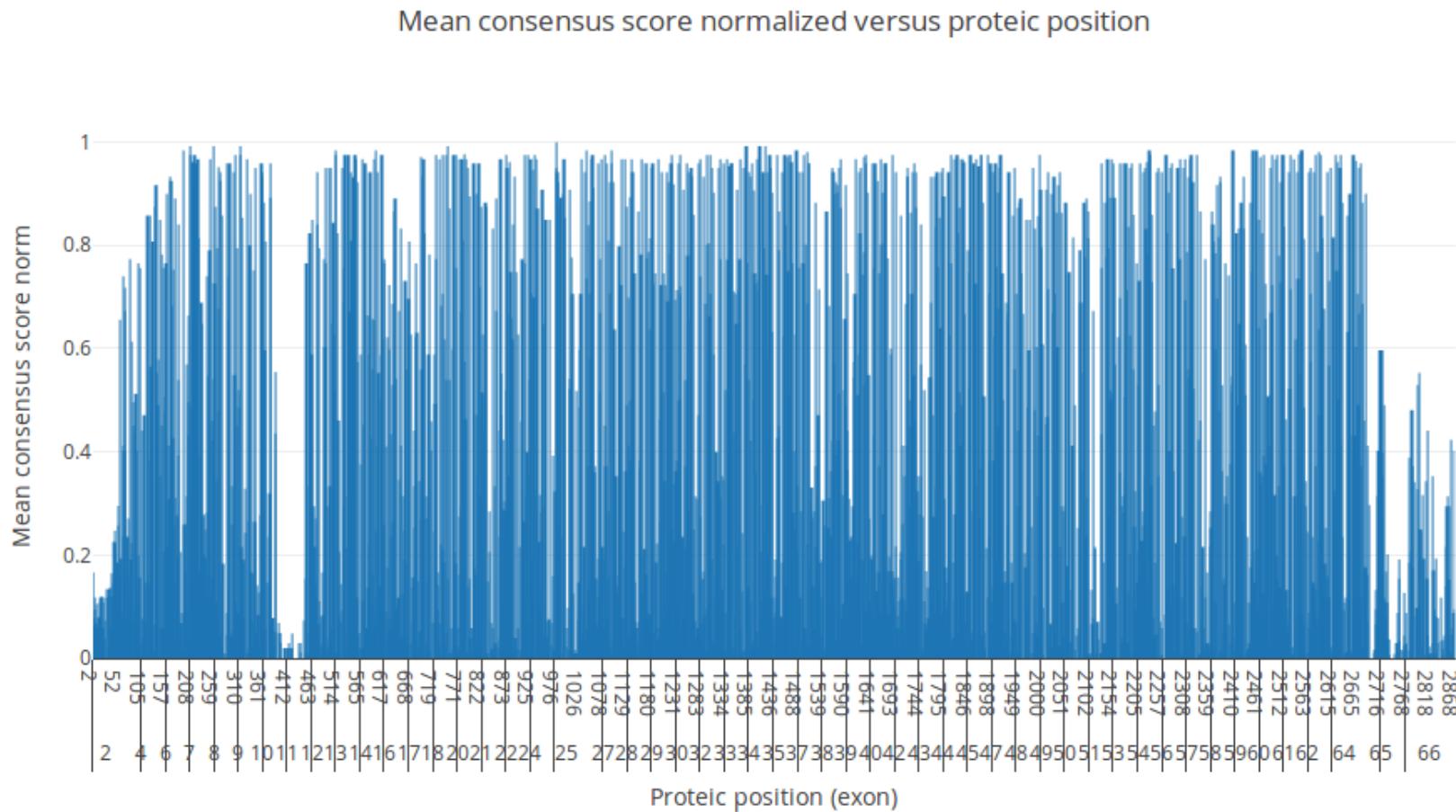


Mean consensus score normalized versus proteic position

# Acknowledgements

Fondation 101 Génomes
FONDATION PRIVÉE
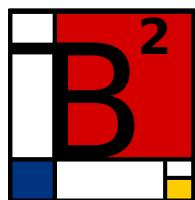
Interuniversity Institute of Bioinformatics in Brussels

THE BELGIAN KIDS' FUND FOR PEDIATRIC RESEARCH

ULB Center of Human Genetics