HEMOGLOBINOPATHIES PREDICTOR

HEMOGLOBIN VECTOR VARIANT MAP AND CLUSTERIZATION

SALVATORE, F. 1,2; BRUNELLO, F. G. 1,2; MARTÍ, M. A. 1,2

1: DEPARTAMENTO DE QUÍMICA BIOLÓGICA, FACULTAD DE CIENCIAS EXACTAS Y NATURALES, UNIVERSIDAD DE BUENOS AIRES, CIUDAD UNIVERSITARIA, PAB. II (CE1428EHA), BUENOS AIRES, ARGENTINA
2: INSTITUTO DE QUÍMICA BIOLÓGICA DE LA FACULTAD DE CIENCIAS EXACTAS Y NATURALES (IQUIBICEN) CONICET. CIUDAD UNIVERSITARIA, PAB. II (CE1428EHA), BUENOS AIRES, ARGENTINA
CONTACT: SALVAA4532@GMAIL.COM

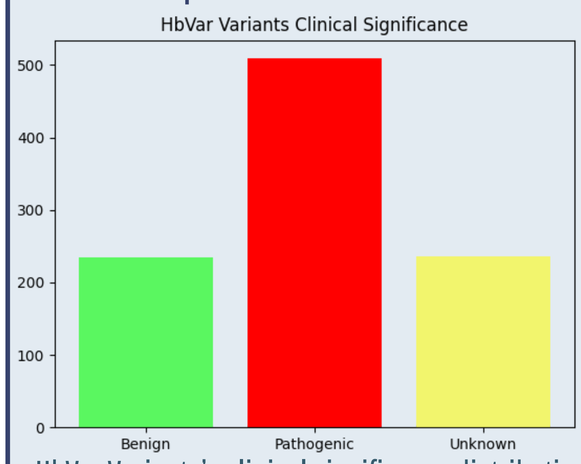
INTRODUCTION

Structural biology and genomics are both essential threads of the bioinformatics, seamlessly intertwined to unravel the profound mysteries of biological systems. Structural biology provides us with the tools to visualize and decipher the three-dimensional arrangement of biological molecules, such as proteins. Genomics has brought to light a profound understanding of the genetic sequences and the molecular architecture that dictate the form and function of an organism.

HbA also known as Human Adult Hemoglobin is a blood protein composed by two pairs of chains α and β of around 145 residues each, shaping a globular and tetrameric protein. It has a heme active site on each subunit which is responsible of intercepting and releasing oxygen atoms used by all our cells to breath.

MATERIALS AND METHODS

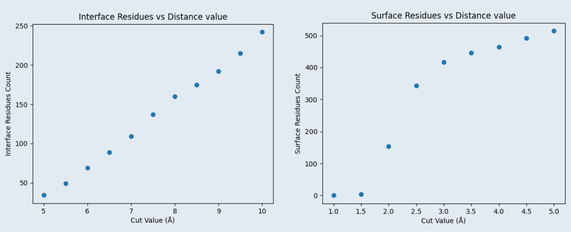
In order to predict the variants's clinical significance a dataset was needed.

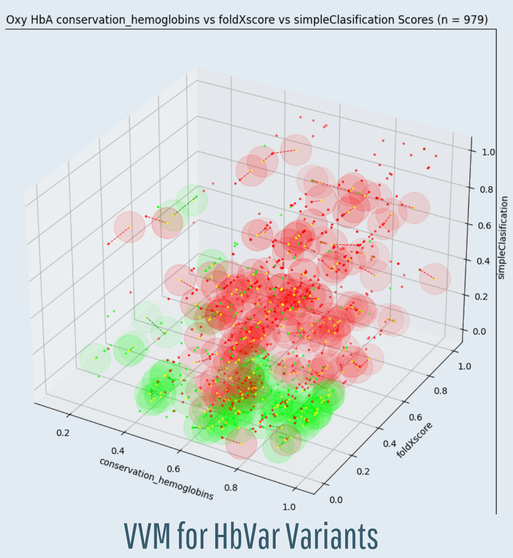


HbVar Variants's clinical significance distribution

Around 1000 variants were scrapped from HbVar database and curated manually. Methods involved in residues structural classification were mainly focused on measuring distances involving each residue's COM.

The VVM method thinks the variants as vectors of N dimensions and plots them as dots.

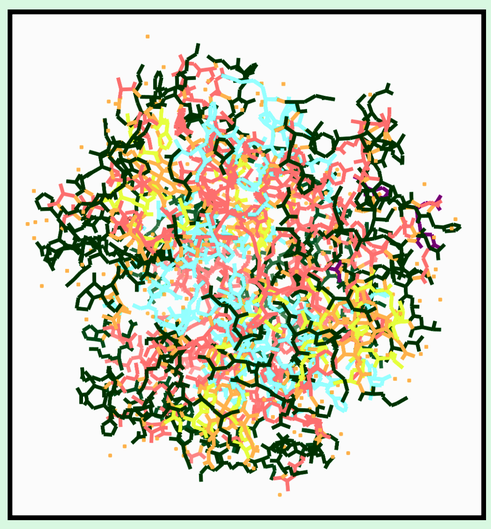




VVM for HbVar Variants

Then it draws an R radius around each unknown variant, counts the amount of pathogenic and benign cured variants inside of it and predicts its clinical significance by frequency. An alternative approach is by only predicting according to the nearest cured variant.

RESULTS

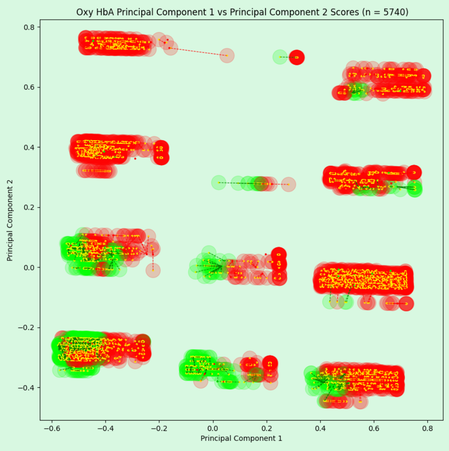


HbA residues coloured by structural classification

First it was necessary to determine each residue's structural classification:

- Surface: 235 (Dark Green)
- Interface: 94 (Sky Blue)
- Core: 196 (Light Red)
- Active Site: 44 (Yellow)
- 2,3 DPG: 5 (Purple)

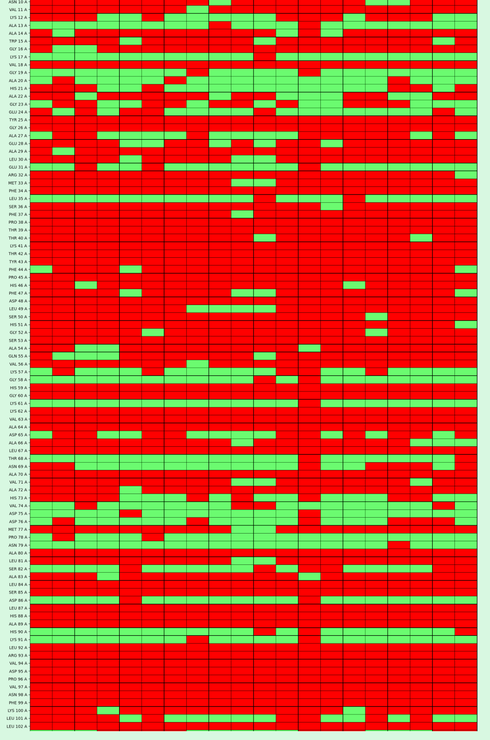
Once that was settled the 20 possible mutations predictions for the 287 residues were made. The predictions that maximizes the accuracy of the model has 5 axis so PCA was needed.



HbA variants predictions by VVM model applying PCA

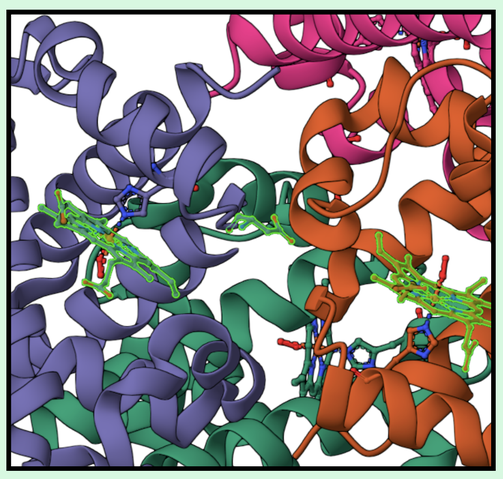
Parameters:

- Conservation Score
- FoldX Score
- Blosum 62 Score
- SASA Score
- Structural Classification



HbA Prediction Heat Map

Out of the 5740 predictions 4025 were predicted as pathogenic and 1715 as benign.



We also aim to explain the importance of some HbA residues. On the left we can see the ASP100 of the B chain (purple) which is an interface residue with its neighbor C chain (orange) but it's also next to both the B and C Heme. All predictions and registered data (7) on this residue are pathogenic and the phenotypes are all O2 high affinity.

This Residue by being on the interface of the protein and also by being so close to two active sites can act as an O2 modulator which is susceptible to any mutation, thus would lean to a pathogenic variant.

In fact this residue is on the 91th percentile on the HBB conservation scores which means it's very high conserved.

ACKNOWLEDGMENTS

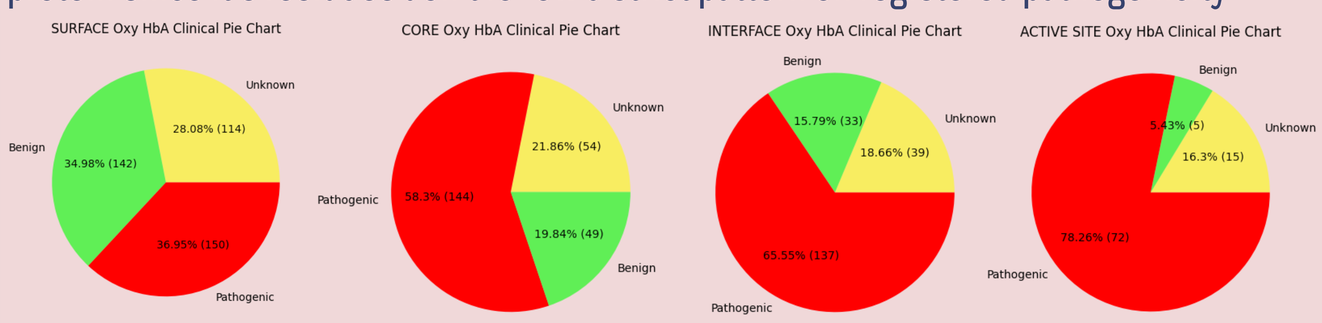
- DEPARTAMENTO DE QUÍMICA BIOLÓGICA, FACULTAD DE CIENCIAS EXACTAS Y NATURALES, UNIVERSIDAD DE BUENOS AIRES, CIUDAD UNIVERSITARIA, PAB. II (CE1428EHA), BUENOS AIRES, ARGENTINA
- INSTITUTO DE QUÍMICA BIOLÓGICA DE LA FACULTAD DE CIENCIAS EXACTAS Y NATURALES (IQUIBICEN) CONICET. CIUDAD UNIVERSITARIA, PAB. II (CE1428EHA), BUENOS AIRES, ARGENTINA

REFERENCES

- BRINGAS M, PETRUK AA, ESTRIN DA, CAPECE L, MARTÍ MA. TERTIARY AND QUATERNARY STRUCTURAL BASIS OF OXYGEN AFFINITY IN HUMAN HEMOGLOBIN AS REVEALED BY MULTISCALE SIMULATIONS. SCI REP. 2017 SEP 7;7(1):10926. DOI: 10.1038/S41598-017-11259-0. PMID: 28883619; PMCID: PMC5589765.
- JORGE SE, BRINGAS M, PETRUK AA, ARRAR M, MARTI MA, SKAF MS, COSTA FF, CAPECE L, SONATI MF, ESTRIN D. UNDERSTANDING THE MOLECULAR BASIS OF THE HIGH OXYGEN AFFINITY VARIANT HUMAN HEMOGLOBIN COIMBRA. ARCH BIOCHEM BIOPHYS. 2018 JAN 1;637:73-78. DOI: 10.1016/J.ABB.2017.11.010. EPUB 2017 DEC 1. PMID: 29199120.
- BUSTAMANTE JP, RADUSKY L, BOECHI L, ESTRIN DA, TEN HAVE A, MARTÍ MA. EVOLUTIONARY AND FUNCTIONAL RELATIONSHIPS IN THE TRUNCATED HEMOGLOBIN FAMILY. PLOS COMPUT BIOL. 2016 JAN 20;12(1):E1004701. DOI: 10.1371/JOURNAL.PCBI.1004701. PMID: 26788940; PMCID: PMC4720485.
- HØIE MH, CAGIADA M, BECK FREDERIKSEN AH, STEIN A, LINDORFF-LARSEN K. PREDICTING AND INTERPRETING LARGE-SCALE MUTAGENESIS DATA USING ANALYSES OF PROTEIN STABILITY AND CONSERVATION. CELL REP. 2022 JAN 11;38(2):110207. DOI: 10.1016/J.CELREP.2021.110207. PMID: 35021073.

CONCLUSIONS

Active Site residue's are all predicted as pathogenic since registered data supports those results. But when predicting surface, interface and core residues predictions are not that accurate and analysis gets more complex, specially on the surface of the protein since it's residues don't follow a strict pattern on registered pathogenicity.



In other words non-active site residues's variants are not as easy to predict and each residue's context should be taken into account, such as it's neighbor residues, it's orientation and interaction inside the globular peptide, it's flexibility once the protein is oxygenated and it's role wether on the functionality or structure of the HbA.

Yet even with it's complexity the VVM model was able to predict with an estimated of 74% accuracy rate which is more than efficient beneath modern proteins pathogenicity predictors (70% - 80%).