ITMO UNIVERSITY

Saint Petersburg, Russia

Assessment of Unresolved PIK3CA Missense Variants Associated with Cancer by Comparative Genomics

Supervisors: Natalia Petukhova and Dmitrii Bug

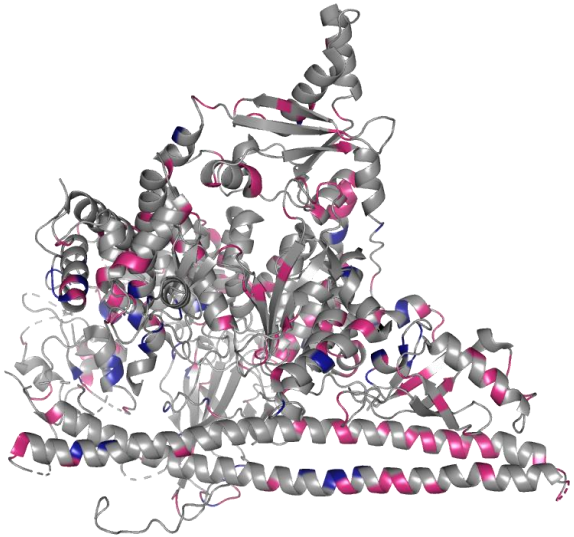
Student: Svetlana Milrud

23.06.2022

Unresolved PIK3CA missense variants

ITMO UNIVERSITY

PIK3CA is gene encoding the **phosphatidylinositol 3-kinase catalytic subunit** (p110a). PIK3CA **missense mutations** have been reported in many **human cancers**



p110a

33%
Benign/Pathogenic variants

67%
Unresolved variants

ClinVar database

Automated tools
(risk estimation based on comparative sequence analysis)

- PolyPhen-2
- PROVEAN
- SIFT
- EVE

Do not have desired level of performance

Project Aim

ITMO UNIVERSITY

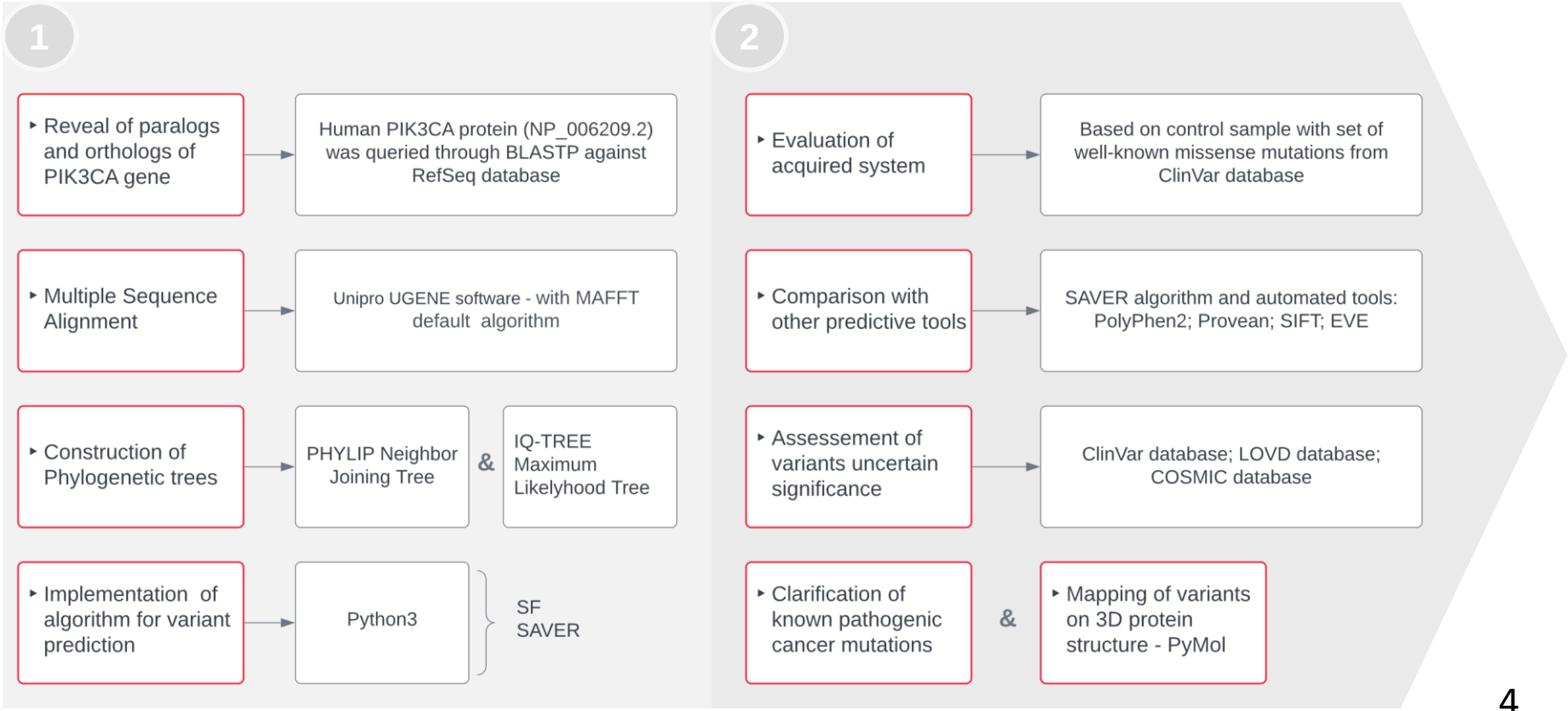
Aim

Determination of precise evolutionary history of the PIK3CA gene in order to evaluate variants of uncertain significance

1. ▸ **Phylogenetic reconstruction** of PIK3CA evolution scenario
2. ▸ **Implementation of algorithm** for prediction of variants of uncertain significance
3. ▸ **Evaluation of the acquired system** on a set of well-known missense mutations
4. ▸ **Comparison of prediction results** with SAVER algorithm and automated tools (PolyPhen2, SIFT, Provean and EVE)
5. ▸ **Assessment of variants of uncertain significance** based on obtained PIK3CA protein map
6. ▸ **Mapping** predicted unresolved PIK3CA **variants** on the protein to understand the **functional domains that were affected**

Project Design

ITMO UNIVERSITY



Saver — https://doi.org/10.1038/gim.2015.208

Results: PIK3CA Orthologs

ITMO UNIVERSITY

Evolutionary history of the PIK3CA gene was reconstructed from representative species across Vertebrata (Actinopterygii, Amphibia, Reptilia, Aves, Mammalia)

Critical step for creating the cleanest possible dataset for assessing the mutations

BLASTP revealed 372 related PIK3CA sequences

The outgroup (65 sequences) was removed based on NJ and ML trees

Results: Algorithm for variant prediction (SF)

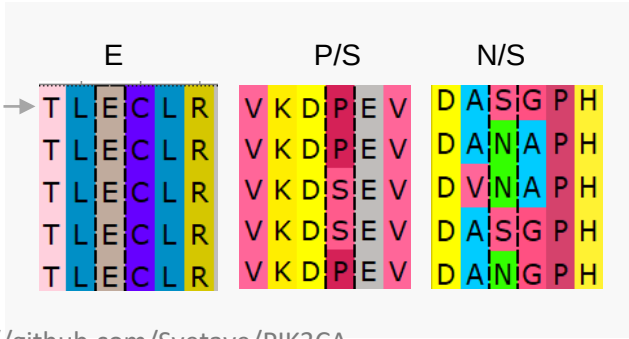
ITMO UNIVERSITY

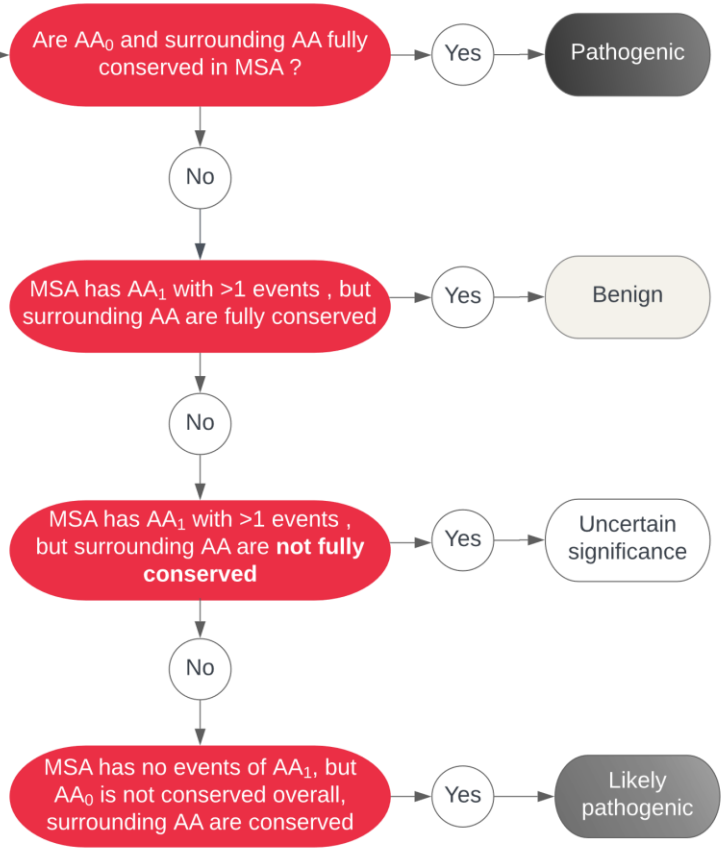
Inputs:

- Multiple sequence alignment
- Variant query

Format:

AA_0-Position-AA_1
e.g **E542K** - substitution of glutamine acid to lysine at position 542



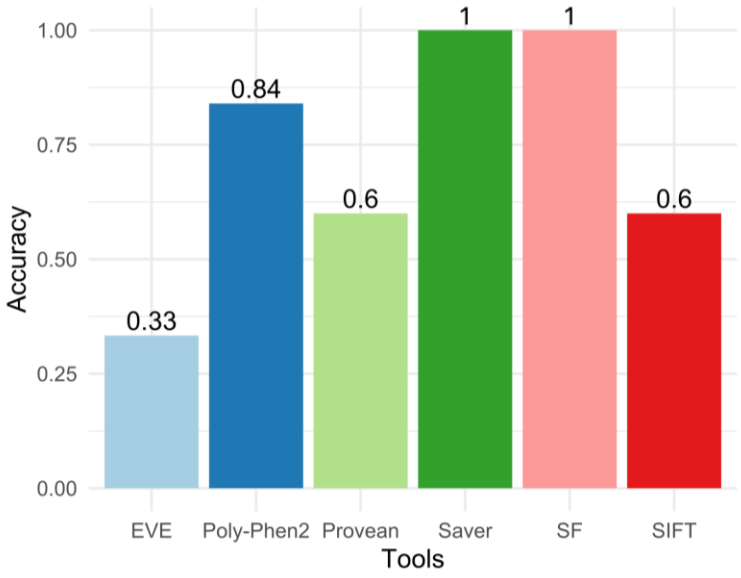


https://github.com/Svetave/PIK3CA

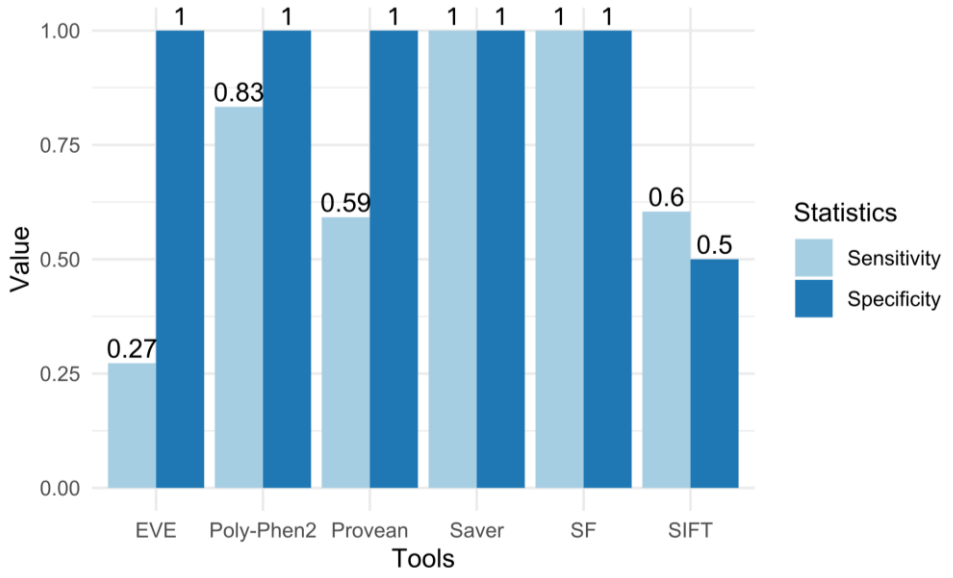
Results: Evaluation of the algorithm and comparison with SAVER and automated tools

ITMO UNIVERSITY

Accuracy



Sensitivity and Specificity

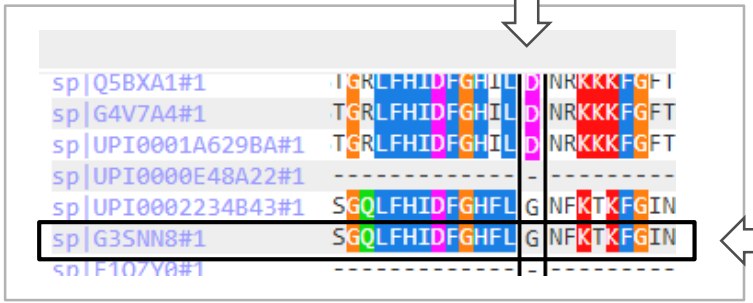


Results: Comparison of SF with automated tools

ITMO UNIVERSITY

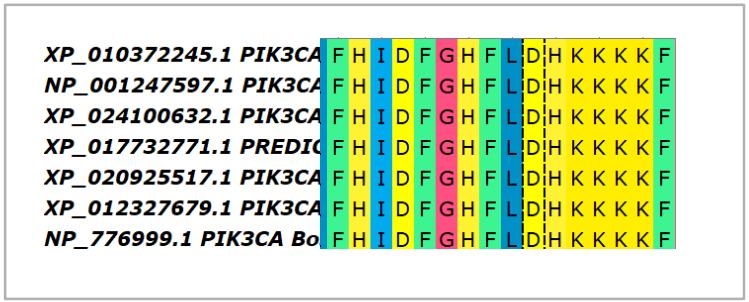
D939G

PolyPhen-2



Paralog

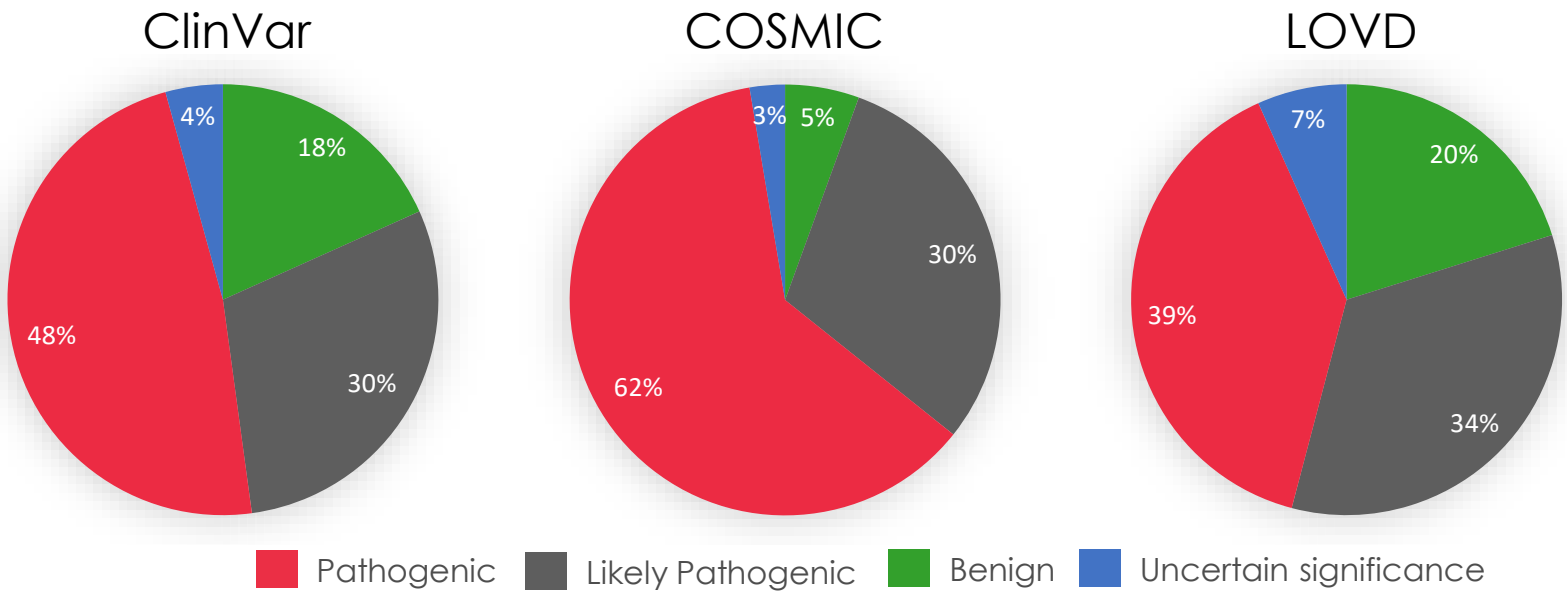
SF



Multiple sequence alignment

Results: Assessment of variants of uncertain significance by SF

ITMO UNIVERSITY

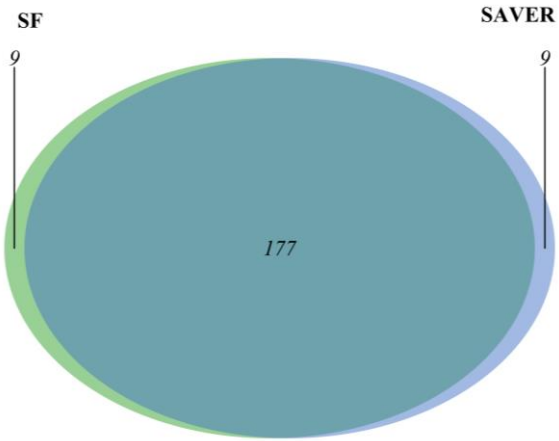


Most of the unresolved PIK3CA missense variants from different databases were identified by SF as **pathogenic**

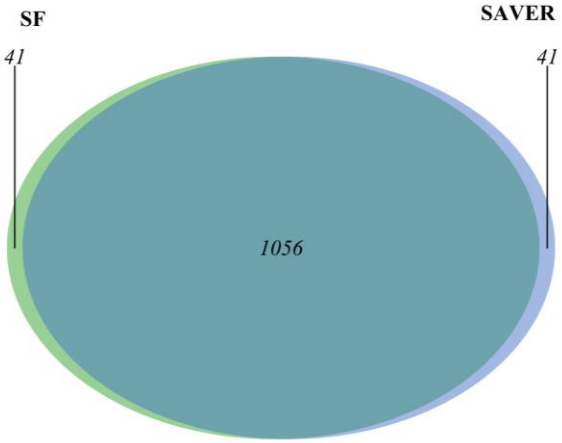
Results: Comparison of SF and SAVER predictions

ITMO UNIVERSITY

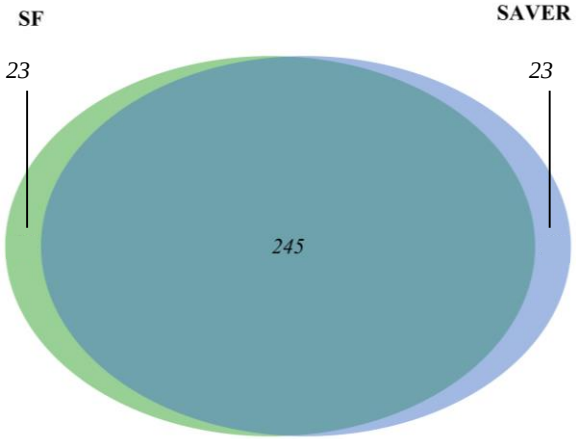
ClinVar



COSMIC



LOVD

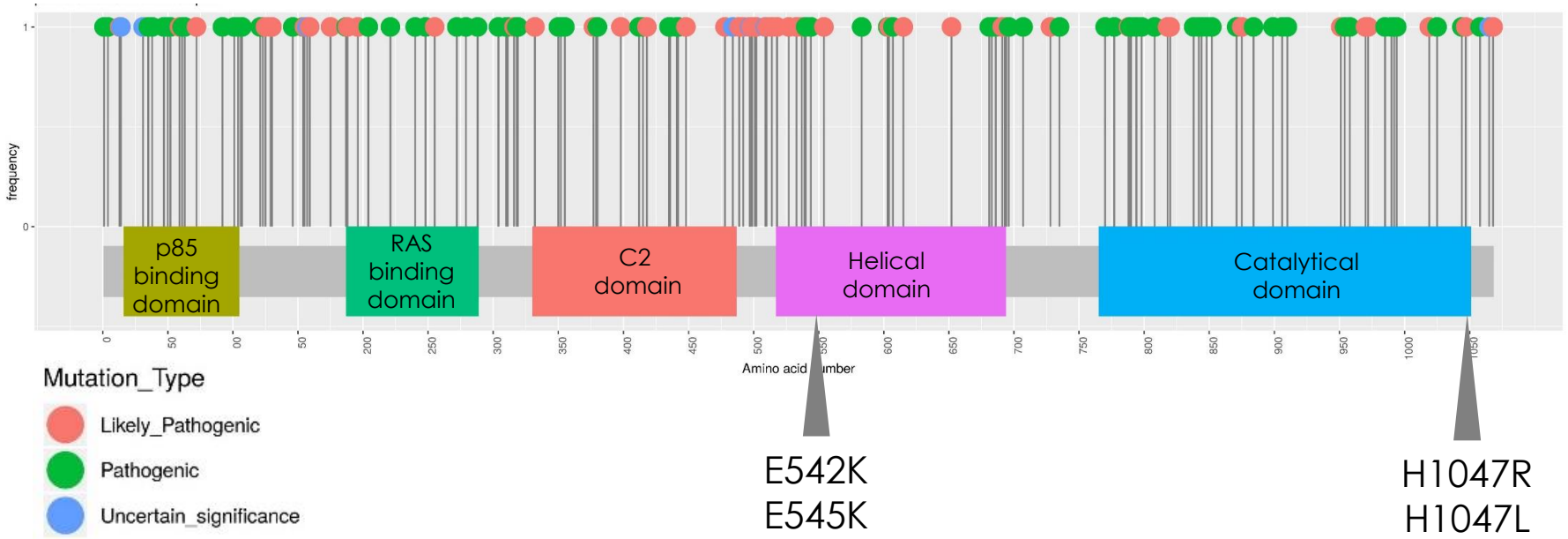


Most of the unresolved PIK3CA variants were **identically predicted** by both algorithms

Results: Mapping of PIK3CA variants on a protein

ITMO UNIVERSITY

SF predicts one of the **hotspot cancer mutations** as Pathogenic, while other three as Likely Pathogenic



Conclusion

ITMO UNIVERSITY

- PIK3CA is a **highly mutable** protein with predominance of pathological mutations
- Accurate evaluation of unresolved variants became possible due to the **implementation of a computational approach** based on comparative sequence analysis
- A clear **improvement in the performance** of our algorithm was achieved by taking into account only function-specific orthologous protein sequences

ITMO UNIVERSITY

Saint Petersburg, Russia

Thank you for attention!

SAVER Algorithm

ITMO UNIVERSITY

Genet Med. 2016 Oct;18(10):1029-36. doi: 10.1038/gim.2015.208. Epub 2016 Feb 18.

Establishing the precise evolutionary history of a gene improves prediction of disease-causing missense mutations

Ogun Adebali [1 2 3], Alexander O Reznik [3 4], Daniel S Ory [5], Igor B Zhulin [1 2 3]

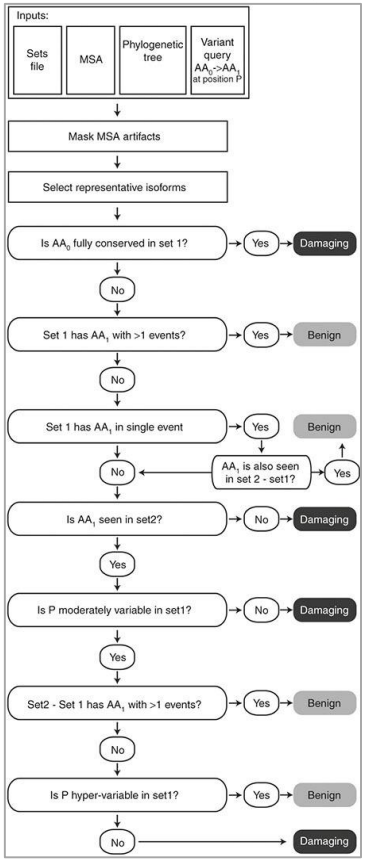
Affiliations + expand

PMID: 26890452 PMCID: PMC4990510 DOI: 10.1038/gim.2015.208

Our project

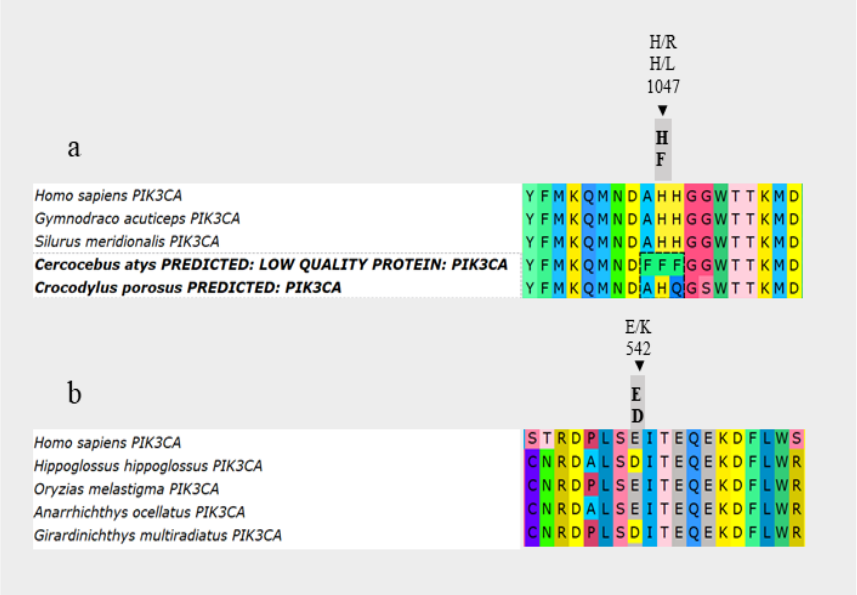
- SET 1: MSA without outgroup
- SET 2: MSA with outgroup

https://github.com/Svetave/PIK3CA



https://doi.org/10.1038/gim.2015.208

Hotspots



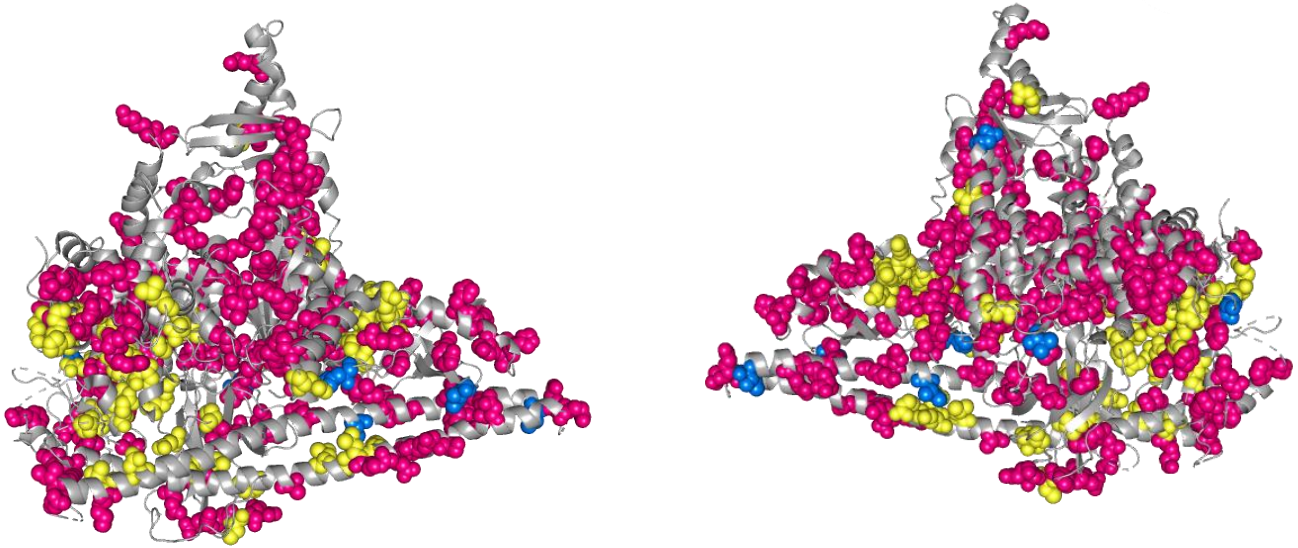
Hotspots mutations

a –SF MSA with H1047R and H1047;

b – SF MSA with E545K

Mapping variants in PyMOL

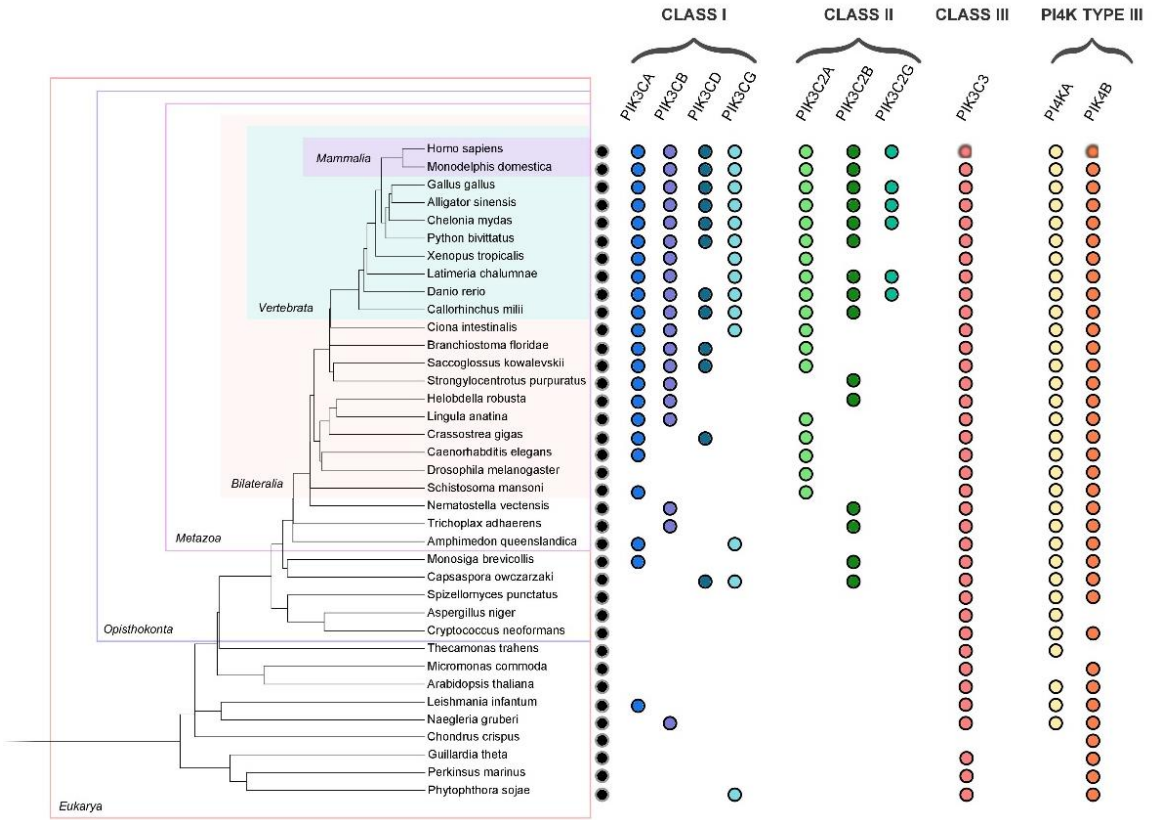
ITMO UNIVERSITY



SF-predicted pathogenic and unresolved PIK3CA missense variants from the ClinVar mapped on 3D protein model (pink – SF Pathogenic and Likely Pathogenic variants; yellow – ClinVar known Pathogenic and Likely Pathogenic variants; blue – SF unresolved variants)

ITMO more than a UNIVERSITY

Paralogs


CLASS I
CLASS II
CLASS III
PI4K TYPE III
PIK3CA
PIK3CB
PIK3CD
PIK3CG
PIK3C2A
PIK3C2B
PIK3C2G
PIK3C3
PI4KA
PI4KB
Mammalia
Vertebrata
Bilateralia
Metazoa
Opisthokonta
Eukarya
Homo sapiens
Monodelphis domestica
Gallus gallus
Alligator sinensis
Chelonia mydas
Python bivittatus
Xenopus tropicalis
Latimeria chalumnae
Danio rerio
Callorhinchus milii
Ciona intestinalis
Branchiostoma floridae
Saccoglossus kowalevskii
Strongylocentrotus purpuratus
Helobdella robusta
Lingula anatina
Crassostrea gigas
Caenorhabditis elegans
Drosophila melanogaster
Schistosoma mansoni
Nematostella vectensis
Trichoplax adhaerens
Amphimedon queenslandica
Monosiga brevicollis
Capsaspora owczarzaki
Spizellomyces punctatus
Aspergillus niger
Cryptococcus neoformans
Thecamonas trahens
Micromonas commoda
Arabidopsis thaliana
Leishmania infantum
Naegleria gruberi
Chondrus crispus
Guillardia theta
Perkinsus marinus
Phytophthora sojae
Tree scale: 100 million years