



# Generation of possible single-nucleotide variants with a given effect on protein-coding sequence

Student: Oxana Kolpakova

Supervisors: Yuri Barbitov, Institute of Bioinformatics,  
Mikhail Skoblov, Research Center for Medical Genetic



# Introduction

- Public databases such as ClinVar and OMIM collect data on clinically relevant SNPs that show associations between genome and phenotype changes.
- SNPs in the coding regions of the human genome, especially missense substitutions, are the most common cause of genetic pathology.
- However, not all significant SNPs have been identified and described. This complicates the identification of the molecular cause of a genetic disorder in people with a suspected hereditary disease.



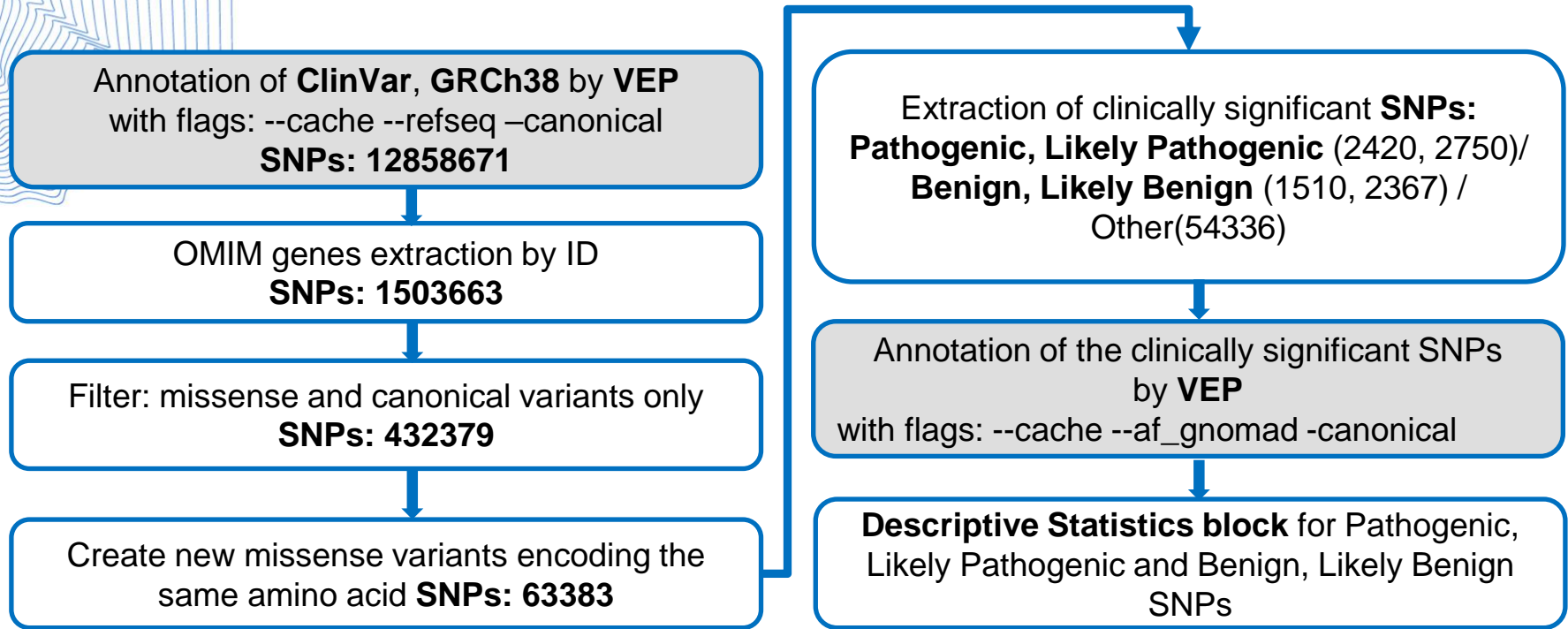




## Aim

- Create a tool generate to pathogenic and benign SNPs for OMIM genes by substitution of 1 nucleotide codon resulting in the same amino acid substitution.
- Our tool expands the list of SNPs for OMIM genes and can be used to improve the molecular genetic diagnosis of hereditary diseases.



# Workflow:



-  our Python script
-  public soft





## Results

- The script work accuracy was validated on coordinates of reference SNPs by Integrative Genomics Viewer, IGV; the script works correctly.
- Generated pathogenic and benign SNPs were not previously described in OMIM and ClinVar DBs (Tab.1)

Table1. Descriptive statistics for sets of clinically relevant SNPs

generated SNPs	benign	likely benign	pathogenic c	likely pathogenic
number	815	1609	876	994
with gnomAD freq	14%	17%	11%	10%
gnomAD freq mean	0.035	0.05	0.005	0.0001
std	0.13	0.13	0.007	0.0002
max	0.62	0.71	0.023	0.002

## Conclusions

- We created a tool to generate pathogenic and benign SNPs by substitution of 1 nucleotide codon resulting in the same amino acid substitution.
- Generated pathogenic and benign SNPs were not previously described in OMIM and ClinVar DBs
- Generated pathogenic and benign SNPs could increase the accuracy of molecular genetic diagnosis
- This tool can be used to create missense SNPs for other gene lists, not just OMIM.



## Future plans

- Expand the capabilities of the tool:
  - add status gnomAD  $>0.05$
  - consider exon/intron boundaries, - DNA strand
  - evaluate the effects of SNPs at splicing boundaries
- Apply the script to real medical data to identify new pathogenic SNPs.

