PICUS: Pointed Interpretation of Clinical Variant Significance

Barış Salman[1,2], Sibel Uğur İşeri[2]

[1]Gen-Era Diagnostics, Bioinformatics Department

[2]İstanbul University, Aziz Sancar DETAE

Introduction

Analyses of genetic variations have identified molecular basis for over 5000 diseases through the past decades [1]. Growing knowledge on inherited diseases and gradual decrease in the cost of next generation sequencing (NGS) applications have made NGS based genetic tests feasible for genetic diagnosis. However, the emergence of NGS in the era of clinical genetics comes with various challenges. One of those challenges is assessing the significance of genetic variations in the human genome. American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) guide has been widely used as standard for sequence variant interpretation regarding 28 criteria [2] based on various attributes of the given variant (Table 1). However, there is significant variance when it comes to implementations of those criteria among different institutes and genome analysts. We have therefore set out to automatize ACMG/AMP criteria via developing a novel bioinformatics tool, namely 'picus'[3].

Table 1: ACMG/AMP criteria breakdown in 8 categories

	Pathogenic	Benign
Population Data	PS4 PM2	BA1 BS1 BS2
Comp. and Pred. Data	PP3	BP4 BP7
Functional Data	PVS1 PS3 PM1 PM4 PP2	BS3 BP1 BP3
Segregation Data	PP1	BS4
De Novo Data	PS2 PM6	
Allelic Data	PM3	BP2
Other Database	PS1 PM5 PP5	BP6
Other Data	PP4	BP5

Keywords: Next generation sequencing, medical genetics, bioinformatics, automation of ACMG/AMP criteria

Result & Discussions

Picus v0.0.4 can automate interpretation of 13 ACMG/AMP criteria in order to classify variants based on significance. There is still need for further manual interpretation in order to incorporate individual of family specific metrics such as segregation data and de novo status of variants.

Research projects have the advantage of selecting desired number of samples from well phenotyped cohorts. Pace of the study is mostly set by publishing papers and various genomic and other experimental methods can be used when assessing a variants significance. More related individuals can be gathered from the family to increase the statistical power (Table 2).

Table 2: Research vs Routine

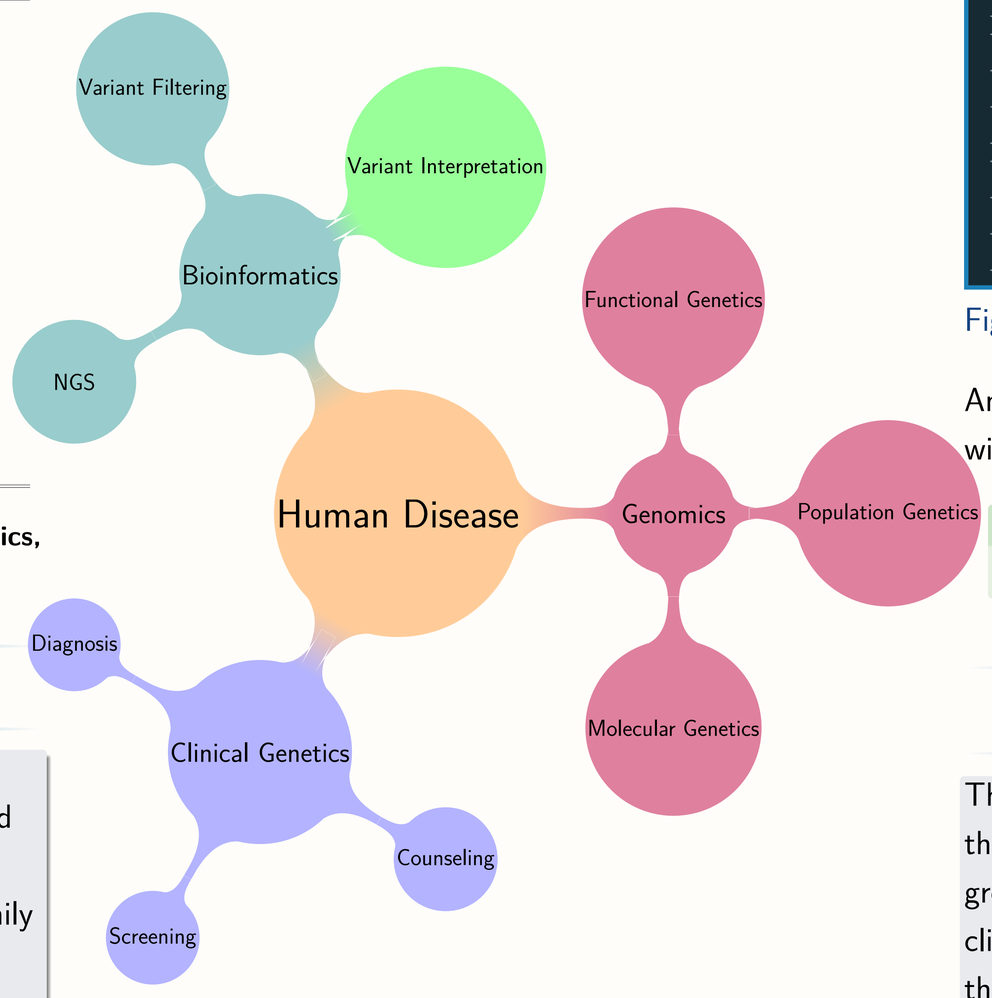
	Research	Routine
Pace	Slow	Fast
Sample Volume	Low	High
Methods	Various	Restricted
Families	Large	Small

In contrast most of the above cannot or may not be carried out in clinical setting. Because;

- Patients are phenotypically heterogeneous.
- There are no cap in number of patients.
- Mostly it is not possible to include any members further than nuclear family members.
- In vivo/vitro methods are restricted since it takes extra time, energy, budget and a team of specialized people.

Software packages that streamline any static and repetitive part of clinical process saves time that can be used to better evaluate the situation. Picus streamlines the process of evidence collection and allows more time to be spent evaluating variants based on their idiosyncratic features (Figure 1). As a result aiming to lower the number of variants of unknown significance that cause a halt in clinical process and adverse psychological effects on patients[4].

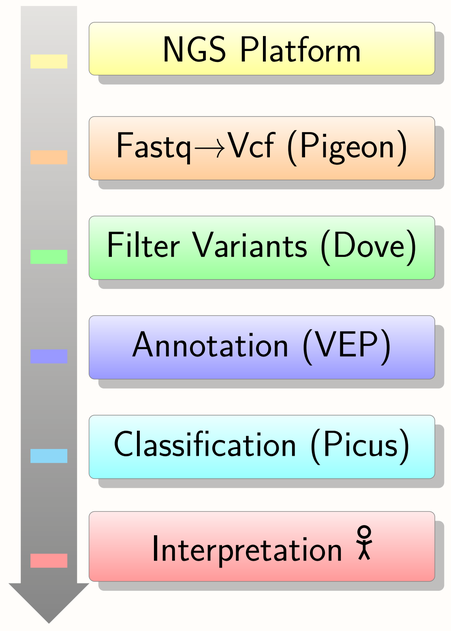
Figure 1: Solving human disease



Material & Method

Picus is developed using Python with the Pandas[5] module. After processing raw fastq data into variant call format (vcf), vcf is further annotated using Ensembl Variant Effect Predictor[6] (VEP) tool (Diagram 1). Parameters such as disease and allele frequency, domain structure of proteins, in silico prediction metrics and documented variant data retrieved from ClinVar have been used for automatized processing of ACMG/AMP criteria prediction. As a result, each variant can be classified as pathogenic, likely pathogenic, benign, likely benign or uncertain significance as suggested by ACMG/AMP.

Diagram 1: Sequence variant discovery workflow



How to use

Picus can be run anywhere with a python3 interpreter and can be installed with pip (See project page[7]). Help screen can be viewed with command 'picus -h' that summarizes parameters (Figure 2).

```
bar@barsdeb:~$ picus -h
usage: picus [-h] [-v] [--version] -i INPUT [-o OUTPUT] [-t TEMPLATE]
             [-l LOGO] [-r REPORT_DATA] [-c COMMAND] [-g GROUP] [-k KEEP]

optional arguments:
  -h, --help            show this help message and exit
  -v, --verbose         WIP.Write what is happening to stdout.
  --version             show program's version number and exit
  -i INPUT, --input INPUT
                        input annotation file.
  -o OUTPUT, --output OUTPUT
                        output annotation file.
  -t TEMPLATE, --template TEMPLATE
                        template file
  -l LOGO, --logo LOGO  custom logo
  -r REPORT_DATA, --report-data REPORT_DATA
                        custom logo
  -c COMMAND, --command COMMAND
                        command to run when compiling in other format
  -g GROUP, --group GROUP
                        group variants based on given parameter
  -k KEEP, --keep KEEP  keep intermediary file while creating report
```

Figure 2: Picus help screen

Analysis can be carried out after Vcf is VEP annotated with following command.

Usage example

picus -i sample.VEP.tsv -o sample.VEP.picus.tsv

Summary & Conclusions

There is high demand for customized bioinformatic tools that could be used in the medical genetics area due to growing number of NGS based genetic tests under clinical setting. Picus aims to help genome analysts in this concept via increased automation in ACMG/AMP criteria.

References

[1] https://omim.org/statistics/entry

[2] Richard S., et. al., Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology *Genetics in Medicine*, 2015

[3] http://classics.mit.edu/Virgil/aeneid.7.vii.html

[4] Makhnoon S., et. al., Experiences of patients seeking to participate in variant of uncertain significance reclassification research, *Journal of Community Genetics*, 2019

[5] McKinney W., Data Structures for Statistical Computing in Python, *Proceedings of the 9th Python in Science Conference*, 2010

[6] McLaren W., The Ensembl Variant Effect Predictor, Genome Biology, 2016

[7] https://github.com/barslmn/picus