**A standard tool to analyze, interpret and report pathological sequence variants related to recessive diseases in NGS data**

Background

Although individually uncommon in general populations, Mendelian diseases are collectively reported to account for ~20% of infants mortality and ~10% of pediatric hospitalizations. More than 2000(Autosomal and X-linked) recessive disorders have been identified. Prevalence of an offspring affected with a recessive disease is higher among consanguineous couples. Detection of carrier status enables identification of couples with 25% risk of affected offspring. Therefore, preconceptual detection of carrier status and genetic counseling enables prevention of disease and providing them with informed reproductive choices.

However, using the best method to conduct preconception career screening, interpretation and responsible reporting of the results are vital to prevent misinformation and for ethical reasons. Traditional carrier screening methods focus on certain ethnic/geographical populations with a higher prevalence of certain recessive diseases using targeted screening. More recently, lower costs and higher accuracy of NGS based technologies have enabled to test for higher number of conditions using carrier screening gene panels. Targeted analysis with gene panels have been used on untargeted Whole exome sequencing data but only very few studies have been done to identify systemic assessment of Whole exome sequencing data to for preconception carrier screening. Sallevelt. et al.(2017) describes a comprehensive filter method that could be used on NGS data of consanguineous as well as non-consanguineous couples to detect important pathogenic variants causing recessive diseases which is the basis for this tool.

This tool not only can be used in preconception carrier but also in general identification of pathogenic variants in NGS output of patients presenting with mendelian diseases. The advantage in this application is the ability to compare the suggested variant related disease symptoms with the actual symptoms of the patients to confirm the diagnosis.

Functionality

The Input is GATK output vcf file annotated with RefSeq, dbSNP, Clinvar, OMIM and convereted to a tsv. The tool will use a comprehensive filter strategy to identify pathogenic variants responsible for mendelian diseases. The resulting output will show the pathogenic variants and the relevant fields that were collectively assessed.

Technologies used;

1. SQL relational database
2. CGI
3. HTML/CSS

Design and Development

1. HTML/CSS form is presented to the user for the user input vcf file
2. vcf file is converted to a tsv file and is imported to the SQL relational database.
3. Filter methods are used as sql queries in python-based CGI script which performs the queries on the relational database table and produce the output pathogenic variants with the relevant fields.
4. The output will be displayed to the user on the web interface.