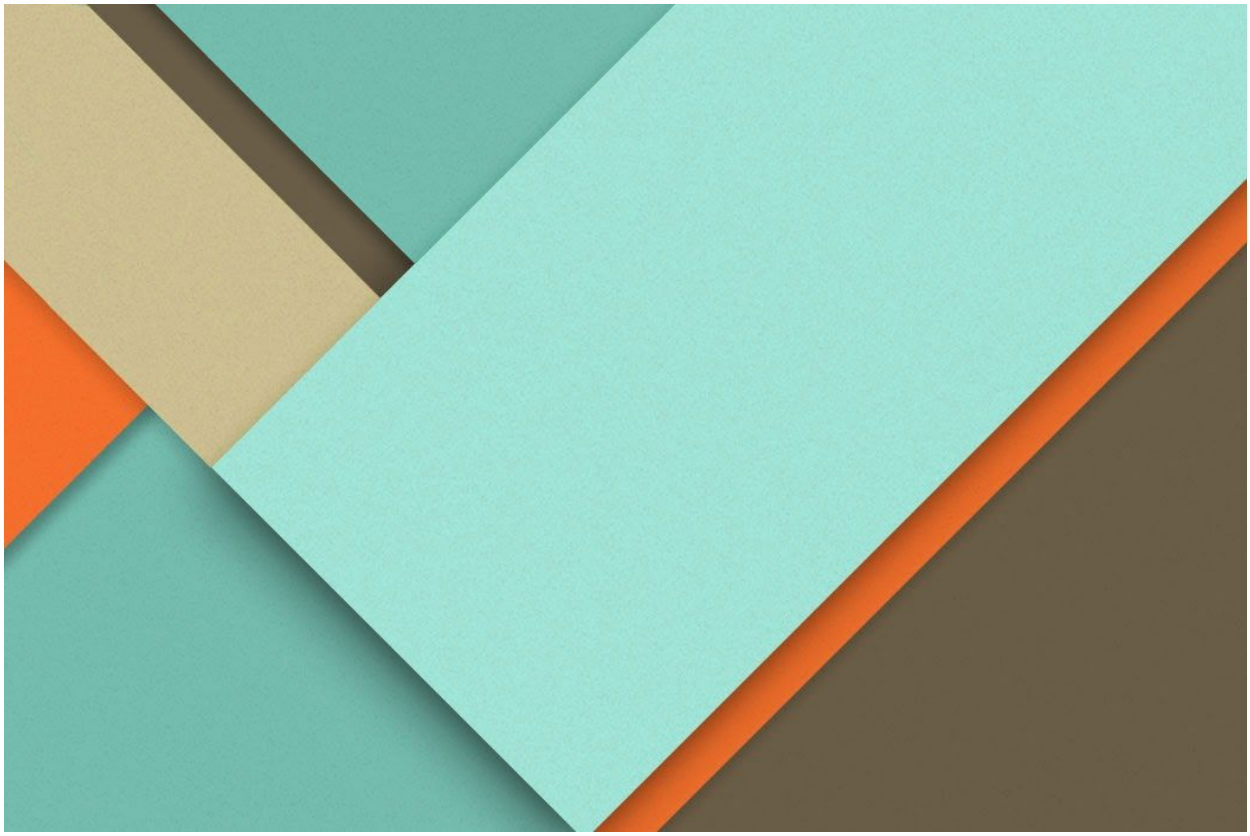

Functional Analysis of Genetic Variants Associated with Type II Diabetes using In-Silico Tools

Team Members

Ashutosh Sanodia (B18BB007)

Shanit Nagre (B18BB032)

We both have worked collectively on theory and results.

Supervisor :
Dr. Pankaj Yadav

Introduction:

In our B-tech Project named Functional Analysis of Genetic Variants Associated with Type II Diabetes using In-Silico Tools , we have studied the genetic aspects related to type II diabetes and have analysed the genes responsible for increasing the risk of type 2 diabetes using in-silico tools. We have observed the mutations in these important genes and have used these substitutions for analyzing the effects of mutation in these genes and their role in type 2 diabetes. Type 2 diabetes (diabetes mellitus) is a metabolic disease that causes sugar to build up in the bloodstream. Type 2 diabetes (T2D) typically results when insulin secretion from the islets fails to keep pace with increasing insensitivity to the action of circulating insulin on its target tissues. Unlike T1D, where the genetic risk is mostly concentrated in the HLA region, the genetic component of T2D risk is not concentrated in one region and appears to be the result of the interaction of multiple genes scattered all across the genome.Small effects of these multiple genes or large effects of few genes can play a role in development of diabetes.

Genes which are observed to have large impact on increasing the risk of type 2 diabetes are **TCF7L2** , **ABCC8** ,**KCNQ1** and **CAPN10** (cause complex metabolic disorder with increased risk of cardiovascular disease). TCF7L2 is said to affect insulin secretion and glucose production. ABCC8 helps regulate insulin and CAPN10 is associated with type 2 diabetes risk in Mexican-Americans.Variations in these genes i.e mutations in their sequence has resulted in increased risk of type 2 diabetes. Studies reported that the rs2237892 and rs2237895 polymorphism in **KCNQ1** has been implicated in T2D risk.

Both genetic and environmental factors are said to be responsible for causing of type II diabetes , obesity is said to be a major cause.Studies have revealed that calpain 10 (*CAPN10*) and transcription factor 7-like 2 (T-cell specific, HMG-box) (*TCF7L2*) that were reliably identified as being associated with T2D. Very little is known about the genetics of type 2 diabetes. Several risk variants are being identified in genome wide studies . These variants still explain a relatively small proportion of the observed heritability. Several studies have found that a risk score based on traditional risk factors (BMI, family history, age, sex, HDL, triglycerides, *etc*.) consistently outperforms any set of genetic markers. We have gone into depth analysis of these 4 genes responsible for increased risk of type 2 diabetes.

In-Silico Tools are the tools which are used to predict the effect of different mutations or substitutions on a particular gene or protein.

SIFT : Uses Fasta Sequence and substitutions of interest as input.

POLYPHEN 2 : Uses Protein Identifier and position and substitutions as input.

FATHMM : Uses Protein Identifier , Substitutions , Predicted Algorithm and Phenotypic associations as input .

METHODOLOGY / WORK PLAN :

We used three In-silico Tools for analyzing the genes and below are the steps used to analyze genes in different In-silico Tools:

- **SIFT** : SIFT Sequence provides SIFT predictions for a given protein FASTA sequence.
- **FASTA sequence :** FASTA format is a text-based format for representing either nucleotide sequences or peptide sequences, in which base pairs or amino acids are represented using single-letter codes. A sequence in FASTA format begins with a single-line description, followed by lines of sequence data.
    - Steps Followed :
        - Taken Input as FASTA sequence for ex. The FASTA sequence for TCF7L2 is >sp|Q9NQB0|TF7L2_HUMAN Transcription factor 7-like 2 OS=Homo sapiens OX=9606 GN=TCF7L2 PE=1 SV=2 MPQLNGGGGDDLGANDELISFKDEGEQEEKSSENSSAERDLADVKSSL VNESETNQNSSS…. Much base pairs ahead.
        - Entered the substitutions in the form of XYZ where X is the initial base pair, Y is it's position and Z is the base with which we want to replace X.
        - Submitted to get the result with a score , telling whether a mutation is DAMAGING or TOLERATING.

- **POLYPHEN 2** : A tool which predicts possible impact of an amino acid substitution on the structure and function of a human protein using straightforward physical and comparative considerations.
    - Steps Followed:
        - Entered Protein ID/Identifier or Protein sequence in FASTA format as an input.
        - Entered Position where base is to be changed.
        - Selected the base to be replaced and from whom to be replaced.
        - Submitted to get the result of substitutions.

- **FATHMM** : It is a tool used to analyze Protein missense variants.
    - Steps Followed:
        - Entered the input in the form of **<protein> <substitution> ,** where <protein> is the Protein ID/Identifier and <substitution> is the amino acid substitution in XYZ format mentioned above . Multiple substitutions can be entered on a single line and should be separated by a comma.

- Predicted Algorithm we selected was weighted as it gives more precise results as compared to unweighted.
- Selected Phenotypic Associations as Gene Ontology so that we know phenotypic consequences of these mutations and Submitted for Result.

RESULTS / OUTCOME :

In Silico Tools that we used for analysis of these genes are SIFT , POLYPHEN 2 and FATHMM. For all 3 genes namely TCF7L2 , ABCC8 and KCNQ1 we have done depth analysis with substitutions available in the uniprot database. Inputs that these tools took were FASTA sequence , substitution of interest and position of substitutions. As an output we got results indicating a score (can be positive or negative) , a prediction of whether the Mutation is Tolerating or Damaging. Every Tool Provided Different Format of results , so we created an excel sheet for every set of results for different genes from different tools.

Link for Excel Sheet containing Result

These charts are for Polyphen 2 others are mentioned in Excel Sheet Link

GENE : TCF7L2

Subs	E322A	D16A	E17A	I19A	F21A	E24A	E26A	E28A	E29A	L48A	K320R
Pred.	DAM.	DAM.	DAM.	DAM.	DAM.	DAM.	DAM.	DAM.	DAM.	DAM.	DAM.

GENE : ABCC8

Substitution	L582V	C435R	H1023Y	R1182Q	R1379C
Prediction	DAMAGING	TOLERATED	TOLERATED	TOLERATED	DAMAGING

GENE : KCNQ1

Substitution	S27A	V324L	T327V	I328L	F340C	A590W	L602A	I609A
Prediction	DAM.	TOL.	TOL.	DAM.	DAM.	DAM.	TOL.	DAMA.

We hereby conclude our reports with all the detailed results in the excel sheet (link given above). Thank You.

CITATION : https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3746083/