

Identifying Variant Positions of TCF7L2 Gene Related to T2DM

K.Vijayalakshmi¹, Prof.M.Padmavathamma²

Research Scholar, Dept. of Computer Science, S.V.University, Tirupati, Andhra Pradesh, India¹

Professor & Head, Dept. of Computer Science, S.V.University, Tirupati, Andhra Pradesh, India²

ABSTRACT: Type2Diabetes Mellitus (T2DM) is associated with impaired insulin secretion. It is the most leading public health problem around the globe, particularly the developing nations in South Asian countries which include India. The pathogenesis of T2DM primarily initiates with inadequacy of pancreatic islet β -cells. Worldwide there are number of Gene Variants that have been associated to T2D. But these genetic variants are differing among the ethnic groups due to different environmental factors and genetic background. By the consequent research by Geneticist, many genes are identified by region wise. Several recent studies by GWAS in Asian Indian populations **TCF7L2** is the first largest gene identified. In this research work we collected tagging SNP protein sequence of TCF7L2 gene (i.e. rs7903146, rs12255372, rs7901695). In this paper we propose DR-VL and GA-VL approaches to identify different transcript variant positions due to single mutation. By identifying these phenotype changes we can predict the disease at earlier and take precautions to reduce the high risk.

KEYWORDS: T2DM, TCF7L2, rs7903146, rs12255372, rs7901695

I. INTRODUCTION

T2DM is an endocrinal metabolic disorder that has become more widespread in these modern years, with significant undesirable effects on human health. T2DM is characterized by insulin resistance (IR) and deficient β -cell function. Usually, T2DM is associated with older age, obesity, family history, previous history of GDM, physical inactivity etc. Almost every organ of the body is affected with long-term complications like blindness, heart and blood vessel disease, stroke, kidney failures, amputations and nerve damage. Simultaneously, Genes play a vital role in the development of Diabetes Mellitus (DM)[5,6]. Insulin secretion is regulated by different hormonal stimuli that activate respective molecular pathway within pancreatic β cells. The most two important physiological stimuli are glucose & incretin. The SNPs of each gene effects on incretin sensitivity or incretin secretion, proinsulin conversions etc. With continuous research on Pathomechanisms of T2DM genes, they identified 27 confirmed potential susceptibility genes for T2DM [2]. In that TCF7L2 is the large, strongly associated gene with 3 different common variants .We are concentrating to identify various transcript position from tagging SNPs of protein sequences of the above with computational approach.

II. RELATED WORK

“Pathomechanisms of Type 2 DiabetesGenes” by Harald Staiger,Fausto Machino et.al Type2 Diabetes Mellitus is a complex metabolic disease that is caused by insulin resistance and β -cell dys-function .And also this disease has an evident genetic component and represents a polygenic disease. It was crucially influenced by the development of affordable high density SNPs. The related genes are identified by two approaches i.e. candidate gene approach and hypothesis free GWA scans and finally different gene variants affecting insulin secretions are identified. “The Genetics of Type2 Diabetes Mellitus: A Review” by Sunitha Singh identified different variants that have been associated to T2DM.By conducting survey on various populations based on genetic factors, various genes are identified that are related to T2DM.Across all the GWAs completed, TCF7L2 clearly shows the larger effect in the development of T2DM. KCNQ1 has been shown to be second largest effect next to TCF7L2. G.R.Chandak.C.S.Janipalli .S.Bhaskar.et.al in their paper “Common variants in the TCF7L2 gene are strongly associated with type 2 diabetes mellitus in the Indian population” mentioned that Indian subjects replicates the strong

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 10, October 2015

association of *TCF7L2* variants with T2DM. It also provides evidence for pathogenesis of T2DM by influencing both insulin secretion and insulin resistance. They mention the most common variants i.e. rs7903146, rs12255372, rs7901695 related to this gene are mostly affecting T2DM. "Gene Expression Analysis for Type-2 Diabetes Mellitus- A study on Diabetes with and without parental history" by Chandrasekhar Vasamsetty et al. diabetes is a group of metabolic disease primarily due to life style factors and genetics. They focus on identifying and classifying genes that cause T2DM with 2 different samples in 2 stages i.e. by Multivariate Outlier Detection and Univariate Outlier Detection. Shital C. Shah, Andrew Kusiak in their paper "Data mining and genetic algorithm based gene/SNP selection" mentioned that genomic studies provide large volumes of data with no. of SNPs ranging into thousands. The analysis of SNPs permits to determine the most significant features that reflect the best interactions between genotypic & phenotypic data. They proposed 2 approaches for identification of significant genes that are related to drug therapy. They are Weighted Decision tree based gene selection (WDTGS) and Genetic algorithm based gene selection (GAGS).

III. GENETICS OF T2DM

Nowadays, considerable progress was made in the identification of Type 2 Diabetes risk genes. And also the recent GWA studies revealed convincing evidence for the contribution of genes to the pathogenesis of T2DM. Several findings indicate that genetics is an important contributing factor for causing this disease. According to the commonly accepted common variant-common disease hypothesis type 2 diabetes is caused by the simultaneous occurrence of common DNA sequence variations in many genes. Each of these DNA alterations is supposed to exert only moderate effects on the affected genes' function and/or expression, but in their sum, these variations confer an increased susceptibility toward the unfavorable environmental factors also. Single nucleotide polymorphisms (SNPs) exchanges of single base pairs that changes approximately 90% of the sequence variation within the human genome. There are two main approaches that lead the search for type 2 diabetes genes. One is candidate gene approach other is hypothesis-free GWA scan. In Candidate gene approach, the common genetic variants examined and these variants' allele frequencies are finally analyzed for being changed in type 2 diabetes cases compared with healthy controls. The positional genetic copies of genes located between or near diabetes-linked chromosomal symbols are compared. But this is difficult due to the non-Mendelian mode of inheritance of human type 2 diabetes; and the size of the chromosomal areas identified in this way that often encompass up to hundreds of genes. However the general diabetes-associated variants in *TCF7L2* on chromosome 10q25.3 (10-23) were recognized by this manual approach and simulated in several populations and ethnicities, and were confirmed in prospective studies of meta-analyses and also furthermore set by GWA studies also. At present *TCF7L2* represents the most realistic diabetes risk gene with an overall allelic relative risk.

IV. TCF7L2 GENE

TCF7L2 variants are the strongest well known genetic risk factors for T2DM. It is a key component and involved in the development of a wide variety of cell lineages and organs. These variants influence T2DM influences in adipogenesis, myogenesis, and pancreatic islet development, as well as in beta-cell survival and insulin secretory granule function. Finally, *TCF7L2* polymorphisms have been associated with impaired insulin secretion, glucose production, and glucose tolerance via direct effects on pancreatic islet beta cells [25]. Although genetic tests for *TCF7L2* could help to predict the incidence and the rate of onset of T2DM. The strongest predictors continue to be positive family history, increased body mass index, increased blood pressure, and increased serum levels of triglycerides, and liver enzymes, all of which precede inception of metabolic syndrome. The analytical power of *TCF7L2* variants disappears with lifestyle modifications and the data suggest that genetic susceptibility to T2DM is determined by *TCF7L2* variants might attest an actionable sign for early intervention and disease prevention.

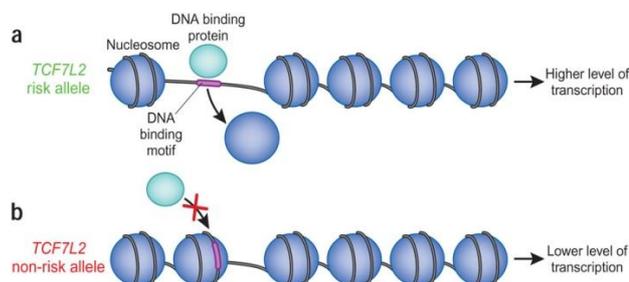


Figure1: TCF7L2 Gene

V. DNA STRUCTURE AND PROTEIN TRANSLATION

DNA, or deoxyribonucleic acid, is the heritable material located in the cell nucleus in humans and almost all other organisms. The information in DNA is stored as a code made up of four chemical bases: adenine (A), guanine (G), cytosine (C), and thymine (T). Human DNA consists of about 3 billion bases, and more than 99 percent of those bases are the same in all people [3]. DNA bases pair up with each other, A with T and C with G, to form units called base pairs. Each base is also attached to a sugar molecule and a phosphate molecule. Together, a base, sugar, and phosphate are called a nucleotide. Nucleotides are arranged in two long strands that form a spiral called a double helix. The structure of the double helix is somewhat like a ladder, with the base pairs forming the ladder's rungs and the sugar and phosphate molecules forming the vertical sidepieces of the ladder.

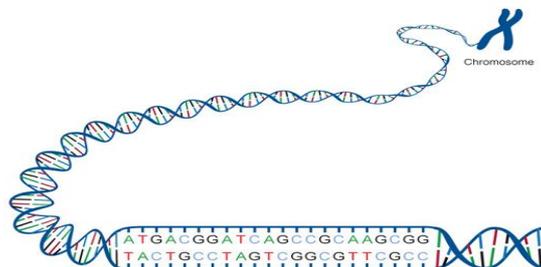


Figure2: DNA Structure

Transcription is the process through which a DNA sequence is enzymatically copied by an RNA polymerase to produce a complementary RNA. Or, in other words, the transfer of genetic information from DNA into RNA. In the case of protein-encoding DNA, transcription is the beginning of the process that ultimately leads to the translation of the genetic code (via the mRNA intermediate) into a functional peptide or protein. Translation is the process that takes the information passed from DNA as messenger RNA and turns it into a series of amino acids bound together with peptide bonds. It is essentially a translation from one code (nucleotide sequence) to another code (amino acid sequence).

VI. PROPOSED WORK

In this research we identified TCF7L2 gene is the most susceptible gene in Asian Indians [1]. According Harald Staiger, Fausto Machino, Andreas Fritsche et.al words Type 2 Diabetes is a multifactorial disorder that is caused by insulin resistance and β cell dysfunction. And they summarize the current knowledge about risk genes and their variant's pathomechanism. With their efforts of research, identified 27 confirmed and potential susceptibility genes of Type2 Diabetes [2]. These are different gene variants that affect insulin secretions and insulin sensitivity of each identified gene. Glucose and Incretin are the two physiological stimuli that regulate insulin secretion that activate respective molecular pathways within the pancreatic β cell. The SNP' of each gene effects on incretin sensitivity, incretin secretions, proinsulin conversions etc and effect on β cell functionality. So, mainly we concentrated on TCF7L2 gene, gathered information related to from NCBI public repository.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 10, October 2015

National Center for Biotechnology Information (NCBI) is part of the United States National Library of Medicine (NLM), a branch of the National Institutes of Health. It has a responsibility for making available the GenBank DNA sequence and a series of databases relevant to biotechnology and biomedicine. In this Gene Expression Omnibus (GEO) is one of the public functional genomics data repositories supporting MIAME-compliant data submissions. This database stores curated gene expression data sets, as well as original Series and Platform records in the Gene Expression Omnibus (GEO) repository. This repository is used to search terms to locate experiments of our own interest.

To assess the method, we used data set of different size. We collected 50 different transcript variants related to rs7903146, rs12255372, rs7901695 SNPs [4]. There we are considering the protein sequence of each gene variant which is having different transcript positions related both diabetic and Non-Diabetic patients. To identify these disparity positions, we propose the following Iterative model as shown below

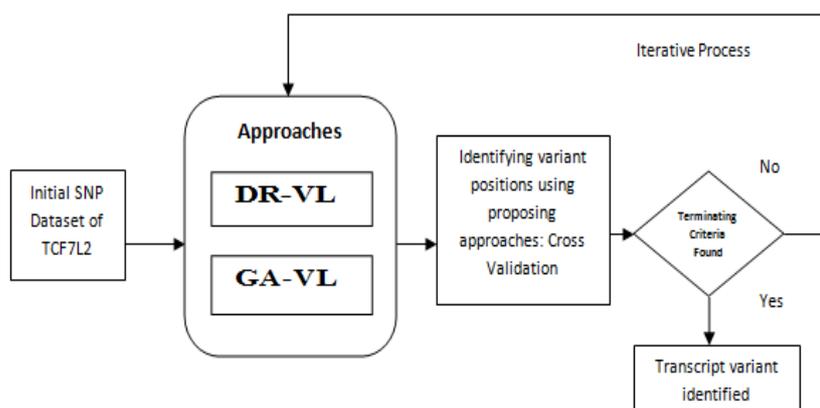


Figure3: Iterative Model to identify Transcript Variant Position

In the proposing model the collected tagging SNP protein sequences are evaluated and variant position changes are identified by following two proposing approaches namely DR-VL and GA-VL. In the evaluation process the data set is underdone 10-Fold-Cross Validation to make training and testing data set.

6.1: Decision Rule Variant Locator (DR-VL):

The data set contains relatively large sequence of normal, diabetic patients. We compared normal tagging SNP with diseased CDS one by one and identified 12 different transcript positions from the collected dataset. To identify those positions we use the following Rule:

```

IF CDS_P=1&& CDS_V=K | CDS_P=12&& CDS_V=L | CDS_P=130&& CDS_V=H | CDS_P=162&& CDS_V=Q |
CDS_P=163&& CDS_V=P | CDS_P=137&& CDS_V=Q....
THEN
    DECISION=" HIGH RISK OF DIABETIC";

ELSE

    DECISION="NORMAL";

```

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 10, October 2015

6.2: Genetic Algorithm Variant Locator (GA-VL):

In this approach we are randomly running data set by dividing into training and testing data using 10-Fold-Crossvalidation. Optimized variant positions are identified by running the following genetic algorithm steps.

- Step1: Randomly generate Test data and training data using 10-fold –cross validation
- Step2: Initialize fittest individual X from CDS as test data
- Step 3: Best individual Y is selected using Tournament selection
 - Step3.1: Apply one-bit crossover operation on X, Y
 - Step 3.2: Evaluate X[i], Y[i] where i=0 to N
 - Step 3.3: If X[i] != Y[i]
 - Identify and Store CDS_V and CDS_L
 - Else Continue;
 - Step 3.4 Do mutation on the results to produce new offspring
- Step4: Repeat until reach termination

Where CDS is tagging Codon Sequence
N is the total no.of individual Codon sequences in data set
CDS_V is Codon variable
CDS_L is Location of the variable

By computing the above steps we get the same positions as well as some other optimized locations using GA-VL.

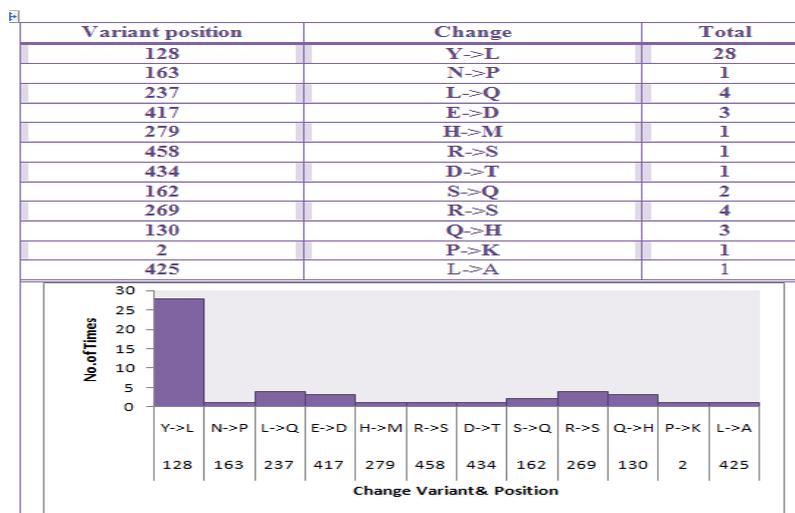
VII. ANALYSIS OF COMPUTATIONAL RESULTS

By performing Meta data analysis of collected information we measure computational accuracy of each model by using the following formula.

$$\text{Accuracy} = \frac{(TN + TP)}{(TN+TP+FN+FP)}$$

[(Number of correct assessments)/Number of all assessments)]

We get more performance accuracy with GA-VL (0.9642) when compared to the DR-VL (0.9264).The total no. of occurrences of each transcript location and changes are tabulated, diagrammatically shown using bar chart as shown below



International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 10, October 2015

VIII. CONCLUSION

We conclude that Genes play an active role in the strengthening of Type2 Diabetes along with rapid changes in life style patterns, environmental changes, increasing physical inactivity. We chosen TCF7L2 gene and its common variant SNPs like rs7903146, rs12255372, rs7901695 to build a model to identify the transcript change in the diabetic patient in both ways using DR-VL and GA-VL. We analysed the testing accuracy of the both developing approaches.

REFERENCES

1. V.Radha & V.Mohan Madras Diabetes Research Foundation & Dr. Mohan's Diabetes Specialities Centre Chennai, India. 2007, Indian J Med Res 125, "Genetic predisposition to Type2 diabetes among Asian Indians", pp 259-274
2. "Pathomechanisms of Type 2 Diabetes Genes" by Harald Staiger, Fausto Machino, Andreas Fritsche et.al., Endocrine Reviews 30: 557-585, 2009.
3. Doria A, Patti ME, Kahn CR 2008 The emerging genetic architecture of type 2 diabetes. Cell Metab 8:186-200
4. Collins FS, Guyer MS, Charkravarti A 1997 Variations on a theme: cataloging human DNA sequence variation. Science 278:1580-1581.
5. G.R. Chandak, C.S. Janipalli, S. Bhaskar et al. 2007 Diabetologia 63-67 "Common variants in the TCF7L2 gene are strongly associated with type 2 diabetes mellitus in the Indian population
6. Florez JC 2008 Newly identified loci highlight cell dysfunction as a key cause of type 2 diabetes: where are the insulin resistance genes? Diabetologia 51:1100-1110
7. Stumvoll M, Fritsche A, Haring HU 2002 Clinical characterization of insulin secretion as the basis for genetic analyses. Diabetes 51(Suppl 1):S122-S129
8. Abate N, Chandalia M, Satija P, Adams-Huet B, Grundy SM, Sandeep S, Radha V, Deepa R, Mohan V 2005 ENPP1/PC-1 K121Q polymorphism and genetic susceptibility to type 2 diabetes. Diabetes 54:1207-1213
9. Grarup N, Urhammer SA, Ek J, Albrechtsen A, Glumer C, Borch-Johnsen K, Jørgensen T, Hansen T, Pedersen O 2006 Studies of the relationship between the ENPP1K121Q polymorphism and type 2 diabetes, insulin resistance and obesity in 7,333 Danish white subjects. Diabetologia 49:2097-2104
10. Moore AF, Jablonski KA, Mason CC, McAteer JB, Arakaki RF, Goldstein BJ, Kahn SE, Kitabchi AE, Hanson RL, Knowler WC, Florez JC 2009 The association of ENPP1K121Q with diabetes incidence is abolished by Lifestyle Modification in the diabetes prevention program. J Clin Endocrinol Metab 94:449-455
11. Grant SF, Thorleifsson G, Reynisdottir I, Benediktsson R, Manolescu A, Sainz J, Helgason A, Stefansson H, et al. 2006 Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. Nat Genet 38:320-323
12. Florez JC, Jablonski KA, Bayley N, Pollin TI, de Bakker PI, Shuldiner AR, Knowler WC, Nathan DM, Altshuler D 2006 TCF7L2 polymorphisms and progression to diabetes in the Diabetes Prevention Program. N Engl J Med 355:241-250
13. Zhang C, Qi L, Hunter DJ, Meigs JB, Manson JE, van Dam RM, Hu FB 2006 Variant of transcription factor 7-like 2 (TCF7L2) gene and the risk of type 2 diabetes in large cohorts of U.S. women and men. Diabetes 55:2645-2648
14. Scott LJ, Bonnycastle LL, Willer CJ, Sprau AG, Jackson AU, Narisu N, Duren WL, Chines PS, Stringham HM, Erdos MR, Valle TT, Tuomilehto J, Bergman RN, Mohlke KL, Collins FS, Boehnke M 2006 Association of transcription factor 7-like 2 (TCF7L2) variants with Saxena R, Gianniny L, Burt NP, Lyssenko V, Giuducci V, Sjögren M, Florez JC, Almgren P, Isomaa B, Orho-Melander M, Lindblad U, Daly MJ, Tuomi T, Hirschhorn JN, Ardlie KG, Groop LC, Altshuler D 2006 Common single nucleotide polymorphisms in TCF7L2 are reproducibly associated with type 2 diabetes and reduce the insulin response to glucose in nondiabetic individuals. Diabetes 55:2890-2895
15. Chandak GR, Janipalli CS, Bhaskar S, Kulkarni SR, Mohankrishna P, Hattersley AT, Frayling TM, Yajnik CS 2007 Common variants in the TCF7L2 gene are strongly associated with type 2 diabetes mellitus in the Indian population. Diabetologia 50:63-67
16. Marzi C, Huth C, Kolz M, Grallert H, Meisinger C, Wichmann HE, Rathmann W, Herder C, Illig T 2007 Variants of the transcription factor 7-like 2 gene (TCF7L2) are strongly associated with type 2 diabetes but not with the metabolic syndrome in the MONICA/KORA surveys. Horm Metab Res 39:46-52
17. Hayashi T, Iwamoto Y, Kaku K, Hirose H, Maeda S 2007 Replication study for the association of TCF7L2 with susceptibility to type 2 diabetes in a Japanese population. Diabetologia 50:980-984
18. Wareham NJ, Franks PW, Harding AH 2002 Establishing the role of gene-environment interactions in the etiology of type 2 diabetes. Endocrinol Metab Clin North Am 31:553-566
19. Grarup N, Andersen G 2007 Gene-environment interactions in the pathogenesis of type 2 diabetes and metabolism. Curr Opin Clin Nutr Metab Care 10:420-426
20. Grant SF, Thorleifsson G, Reynisdottir I, et al. Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. Nat Genet. 2006 Mar. 38(3):320-3. [Medline]
21. Lyssenko V, Lupi R, Marchetti P, et al. Mechanisms by which common variants in the TCF7L2 gene increase risk of type 2 diabetes. J Clin Invest. 2007 Aug. 117(8):2155-63. [Medline]. [Full Text].
22. Schafer SA, Tschritter O, Machicao F, et al. Impaired glucagon-like peptide-1-induced insulin secretion in carriers of transcription factor 7-like 2 (TCF7L2) gene polymorphisms. Diabetologia. 2007 Dec. 50(12):2443-50. [Medline]. [Full Text].
23. Silbernagel G, Renner W, Grammer TB, et al. Association of TCF7L2 SNPs with age at onset of type 2 diabetes and proinsulin/insulin ratio but not with glucagon-like peptide. Diabetes Metab Res Rev. 2011/07. 27(5):499-505.
24. "Identifying Candidate Genes for Type 2 Diabetes Mellitus and Obesity through Gene Expression Profiling in Multiple Tissues or Cells", Junhui Chen, Yuhuan Meng, Jinghui Zhou, et al. Journal of Diabetes Research Volume 2013 (2013), Article ID 970435, 9 pages, <http://www.ncbi.nlm.nih.gov/geo/>, <http://www.ncbi.nlm.nih.gov/gds>.
25. "Gene Expression Analysis For Type-2 Diabetes Mellitus -- A Study On Diabetes With And Without Parental History." by
26. Vasamsetty, Chandra Sekhar, Peri, Srinivasa Rao; Rao, Allam Appa; Srinivas, K.; Someswararao, Chinta, Journal of Theoretical &
27. Applied Information Technology . 2011, Vol. 27 Issue 1, p43-53. 11p.
28. "The Genetics Of Type 2 Diabetes Mellitus : A Review" by Sunita Singh, Journal of Scientific Research, ISSN : 0447-9483, Vol. 55, 2011 : 35-48
29. Data mining and genetic algorithm based gene/SNP selection, Shah SC, Kusiak A. Artificial Intelligence in Medicine. 2004 Jul; 31(3):183-96.