SciMedCentral

JSM Bioinformatics, Genomics and Proteomics

Research Article

In silico Analysis of Single Nucleotide Polymorphisms (Snps) in Human FTO Gene

Marwa Mohamed Osman[1]*, Ahmed Sidahmed Khalifa[2], Alaa Eldin Yousri Mutasim[3], Samia Othman Massaad[1], Mosab Mohamed Gasemelseed[4], Mosab Abdalazim Abdagader[5], Sanabel Alhussein Ahmed[6], Ahmed Mohammed Ahmed[1], Hisham Nouraldayem Altayb[7] and Mohamed Ahmad Salih[1]

[1]*Department of Biotechnology, Africa city of Technology, Sudan*
[2]*Ministry of Health-Khartoum-Sudan*
[3]*Department of microbiology, University of Science and Technology, Sudan*
[4]*Department of medical laboratory Science, Al-Neelain University, Sudan*
[5]*Ministry of Health- Gadarif, Sudan*
[6]*Department of immunology, Endemic Diseases Institute, Sudan*
[7]*Department of microbiology, Sudan University of Science and Technology, Sudan*

***Corresponding author**
Marwa Mohamed Osman, Africa city of Technology-Khartoum, Sudan, Tel: 249912260109; Email: marmarzamaani@gmail.com

Submitted: 08 March 2016

Accepted: 04 April 2016

Published: 07 April 2016

Keywords
- Fat mass and obesity associated protein (*FTO*) gene
- SNPs
- miRNA site
- 3'UTR
- BioGriD

Abstract

Obesity is a common and serious medical problem that adverse health consequences and associated with many serious disease. *FTO* gene plays a main role in obesity. Bioinformatics' analysis of *FTO* gene initiated by Polyphen-2 and SIFT server was used to review 31pathological polymorphisms. Among these 31, 10 pathological polymorphisms were found to be very damaging, with higher Polyphen-2 score of the Polyphen-2 server (=1) and a SIFT tolerance index of 0.000-0.005. Protein structural analysis was done by modeling of amino acid substitutions using Project Hope server for all these pathological polymorphisms. 30 SNPs in the 3'URT containing 63 alleles can be disrupted by a conserved miRNA site. Hence, we hope our results will provide useful information that needed to help researchers to do further study to solve the obesity problem in the future. We considered this study a distinctive one, because there is no researches deal with matter *in silico* studies.

ABBREVIATIONS

ABH: Alpha-Ketoglutarate-dependent dioxygenase; BioGRID: Biological General Repository for Interaction Datasets; DDG: $\Delta\Delta G$ sign; FTO: Fat mass and Obesity associated protein; GWAS: Genome-Wide Association Studies; TI: Tolerance Index; OMIM: Online Mendelian Inheritance in Man; LOF: Loss-of-Function; miRNA: Miro Ribonucleic Acid; nsSNPs: non synonymous Single Nucleotide Polymorphisms; PolymiRTS: polymorphism in micro RNAs and their Target Sites; PolyPhen-2: Polymorphism Pheno typing V2; PSIC: Position-Specific Independent Counts software (Polyphen-2 score); SIFT: Sorting Intolerant from Tolerant; SNPs: Single Nucleotide Polymorphisms; SVM: Support Vector Machine; UTR: Un-Translated Region

INTRODUCTION

Obesity is a common and serious medical problem that adverse health consequences and associated with cardiovascular disease, stroke, type 2 diabetes mellitus, hypertension, dyslipidemia, cancers of the breast, endometrium, prostate, and colon, gallbladder disease, osteoarthritis, respiratory problems, including asthma and sleep apnea, and perhaps depression. In addition to aerobic capacity and the ability to perform physical activities may be hindered by obesity, and this may have implications for physical therapists' interventions. Obesity is a global problem affecting over 400 million adults. It is not only a problem found in the adult population but is also occurring at an increased frequency in children in both the developed and the developing world. It has received both national and international attention because of obesity's detrimental impact on health, the enormous economic burden it imposes, and its increasing prevalence [1-4].

Many factors interact together to play a role in obesity including environmental and lifestyle factors. The rising prevalence of obesity can be partly explained by environmental changes over the last 30 years, in particular the unlimited supply of convenient, highly calorific foods together with a sedentary

lifestyle. Recently the role of multiple genetic polymorphisms in obesity has taken in concern. An estimation of 40 to 70% of the variation in obesity susceptibility observed in the population is due to inter-individual genetic variants [3, 5-7].

Understanding the molecular roots of obesity is an important prerequisite to improve both prevention and management of the disease [4]. From the beginning the advent of the genome-wide association approaches in 2005, genome-wide association studies (GWAS) have identified approximately 2,000 genetic loci with strong associations for more than 300 common traits and diseases [8], this including more than 75 obesity-susceptibility loci [9,10].

Fat mass and obesity associated protein is encoded by *FTO* gene located in chromosome 16 and the cluster of single nucleotide polymorphisms (SNPs) associated with obesity located in the first intron of the gene [11-13]. It is encodes for a nuclear enzymatic protein known as alpha-keto glutarate-dependent dioxygenase belonging to the AlkB homologue (ABH) subfamily of 2-oxoglutarate (2OG) and ferrous iron-dependent oxygenases [14].

Science 2007 and the subsequent years several studies indifferent populations including American, Australian, Korean, Danish, German, and Hispanic American, non-Hispanic Caucasian American, African American and Spanish confirmed that *FTO* variants are associated with obesity [15-25].

In contrast, in a large Palestinian Arab consanguineous pedigree, a homozygous loss-of-function (LOF) mutation in *FTO* was identified. Which nine individuals presented with severe growth retardation and multiple congenital malformations, including microcephaly, severe psychomotor delay, cardio myopathy and characteristic facial features (OMIM 612938) [26].

This finding means that *FTO* is required for normal development of cardiovascular systems and human central nervous and those homozygous loss-of-function (LOF) mutations in the *FTO* gene can lead to an autosomal recessive lethal disorder and multiple defects. However, no other disease-causing mutation in *FTO* has been reported to data [27].

In this study we use different computational methods to identify the *FTO* gene SNPs and the effects of predicted mutation at the proteomic level. We considered this study a distinctive one, because there is no researches deal with matter *in silico* studies.

MATERIAL AND METHODS

FTO gene was investigated in dbSNP/NCBI database (in September 2015. *FTO* gene contained a total of 72908 SNPs; 20711 of them found on *Homo sapiens*; of which 185 were missense, of which 87 were in the coding region, 193 were non-synonymous SNPs (nsSNPs), 187, were in the 3'un-translated region and OMIM_610966. Predictions of deleterious nsSNPs were performed by different software such as SIFT, Polyphen-2 softwares I-Mutant suite and project hope. The FASTA format of the protein and its iso forms (four isoforms) was obtained from Uniprot at Expassy database. The 3D structure of an 87% identical to protein was retrieved from database by using BLAST/NCBI. The protein used as a template is called "Alpha-ketoglutarate-dependent dioxygenase [*Homo sapiens*]" with ID pdb|3LFM|. The SNPS at the 3UTR region were analyzed by Polymirtdatabase. All soft were illustrated in Figure (1).

Bioinformatics data analysis

The BioGRID (Biological General Repository for Interaction Datasets; is an open access database that houses genetic and protein interactions curated from the primary biomedical literature for all major model organism species and humans [28]. SIFT software: "Sorting Intolerant from Tolerant". This is a sequences homology-based tool that presumes that important amino acids will be conserved in the protein family. Hence, changes at well-conserved positions tend to be predicted as deleterious [29]. The cutoff value in the SIFT program is a tolerance index of ≥0.05. The higher the tolerance index, the less functional impact a particular amino acid substitution is likely to have. The server PolyPhen-2(Polymorphism Phenotyping v2) which is available at has been used to analyze the structural damage due to coding nsSNPs which can affect protein functionality. The server is able to calculate a score on the basis of the characterization of the substitution site to a known protein three-dimensional structure. A PSIC score has been calculate for each variant of each site and the difference between them reported. The higher the PSIC score difference is, the higher is the functional impact a particular amino acid substitution is likely to have [30]. I-Mutant Suite is a suite of support vector machine (SVM)-based predictors of protein stability changes according to Gibbs free energy change, enthalpy change, heat capacity change, and transition temperature [31]. The analyses were performed based on protein sequence combined with mutational position and correlated new residue. And the output result of the predicted free energy change (DDG) classifies the prediction into one of three classes: largely unstable (DDG < −0.5 kcal/mol), largely stable (DDG>0.5kcal/mol), or neutral (-0.5≤ DDG≤0.5 kcal/mol). I- Project Hope **s**oftware is an online web service where the user can submit a sequence and mutation. This software collects structural information from a series of sources, including calculations on the 3D protein structure, sequence annotations in Uni Prot and predictions from DAS-servers. It combines this information to give analyze the effect of a certain mutation on the protein structure and will show the effect of that mutation in such a way that even those without a bioinformatics background can understand it [32]. PolymiRTS database 3.0(Polymorphism in microRNAs and their Target Sites) is a naturally occurring DNA variations in microRNA seed regions and microRNA target sites. Integrated data from CLASH (cross linking, ligation and sequencing of hybrids) experiments, PolymiRTS database provides more complete and accurate microRNA–mRNA interactions. The polymorphic microRNA target sites are assigned into four classes: 'D' (the derived allele disrupts a conserved microRNA site), 'N' (the derived allele disrupts a nonconserved microRNA site), 'C' (the derived allele creates a new microRNA site) and 'O' (other cases when the ancestral allele cannot be determined unambiguously). The class 'C' may cause abnormal gene repression and class 'D' may cause loss of normal repression control. So these two classes of PolymiRTS are most likely to have functional impacts. PolymiRTS is available at [33].

All of this software were used in their default setting.

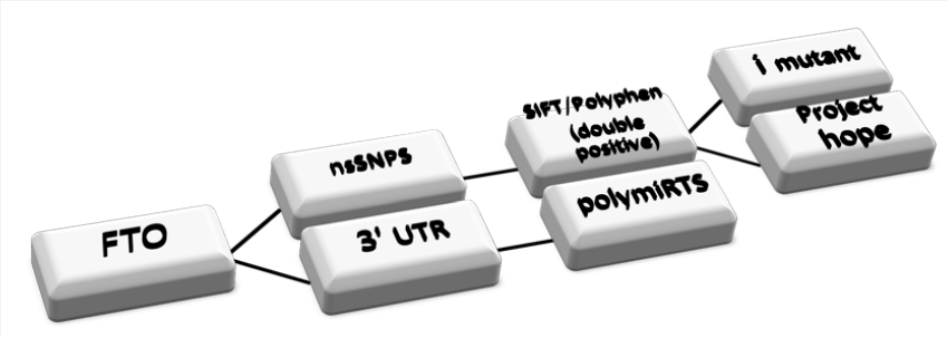


Figure 1 Online Software using in SNPs analysis.

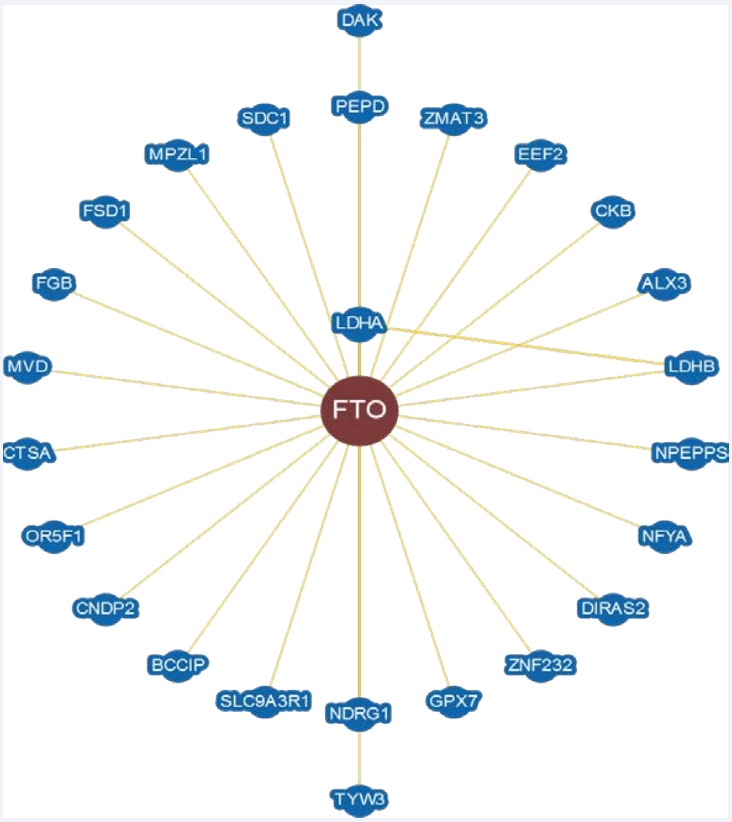


Figure 2 Functional interaction between FTO and its related genes [34].

RESULTS AND DISCUSSION

Alpha-ketoglutarate-dependent dioxygenase has many vital functions: DNA-N1-methyladenine dioxygenase activity, ferrous iron binding, oxidative DNA demethylase activity and oxidative RNA demethylase activity. The gene has 25 physical interactions with other genes (interactors) which they have similar functions and illustrated by using BioGRID and shown in Figure (2) and Table (1).

Deleterious or damaging nsSNPs predicted by PolyPhen and SIFT

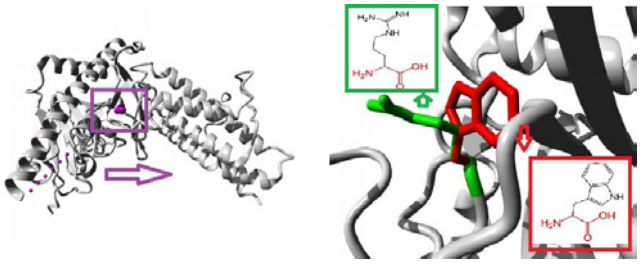
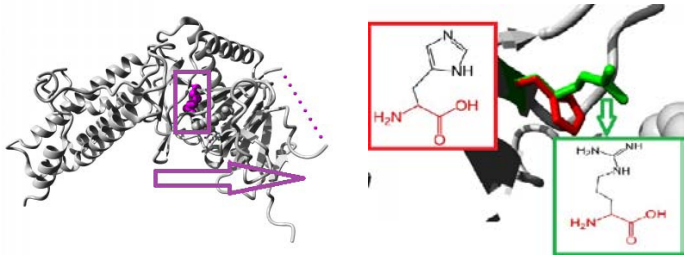
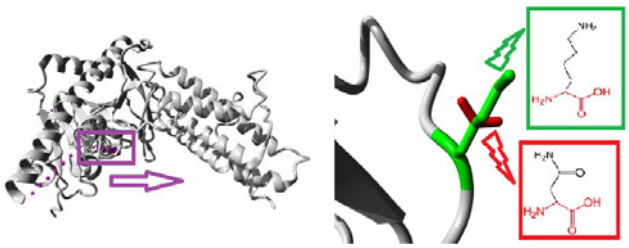
A total number of 193 nsSNPs *FTO* gene were submitted as batch to the SIFT program and PolyPhen-2 server. In the SIFT program, we observed that, out of total 193 nsSNPs; 31 pathological polymorphisms was deleterious and 10 deleterious with low confidence (which was excluded from this study), 9 pathological polymorphisms (rs139577103 → R96H, rs370009039 → K216N, rs373076420 → P288L, rs200152693 → A15S, rs139000284 → R149H and R445H, rs376381270 → R159S, rs370137051 → L193F and L489F) showed a highly deleterious tolerance index (TI) score of 0.00.

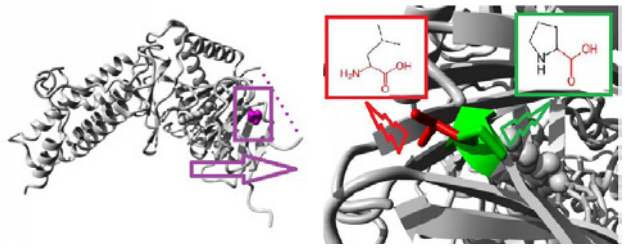
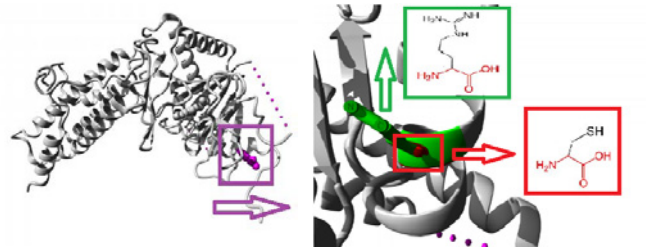
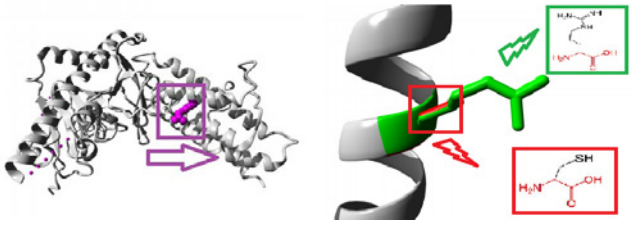
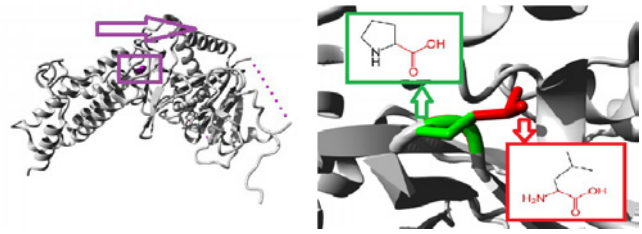
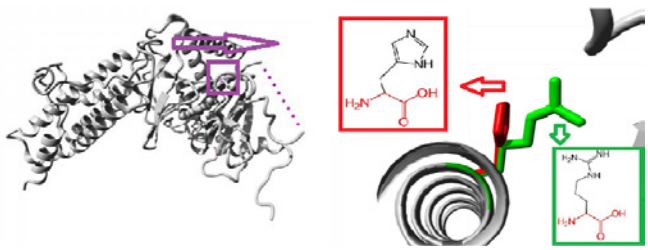
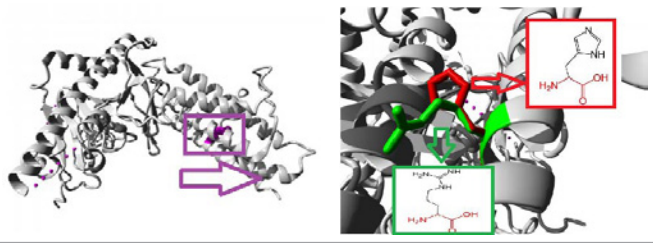
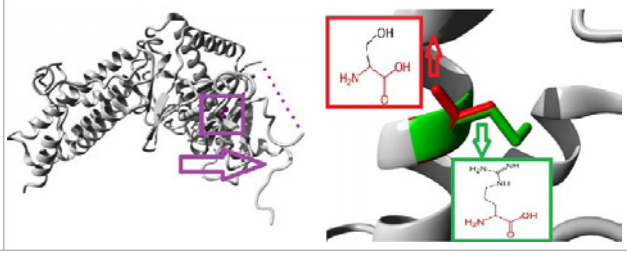
In Polyphen-2 server, our result represented that, 31 pathological polymorphisms were predicted damaging:rs371489995 → L51I,rs61743972 → G182A, rs150450891 → V201I,rs138348216 → M207V and rs377073096 → D350G showed possibly damaging while rs201836578 → K2E,rs140101381 → R80W,

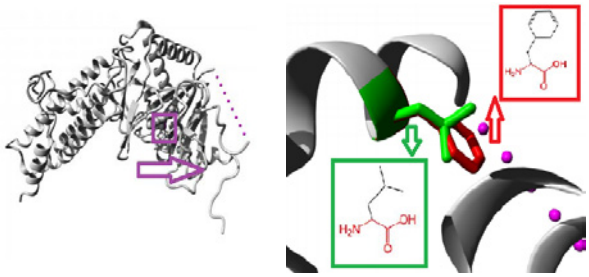
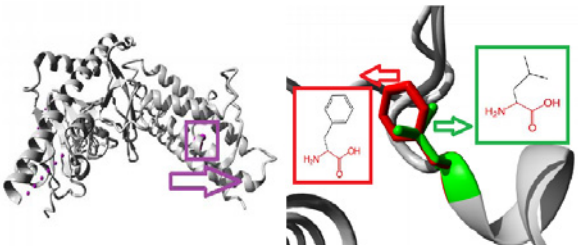
Table 1: Functional interaction between *FTO* and its related genes [34].

Interactor	Role	Organism	Experimental Code Evidence
ALX3	BAIT	H. sapiens	Affinity Capture-MS
BCCIP	HIT	H. sapiens	Co-fractionation
CKB	HIT	H. sapiens	Co-fractionation
CNDP2	HIT	H. sapiens	Co-fractionation
CTSA	BAIT	H. sapiens	Co-fractionation
DAK	HIT	H. sapiens	Co-fractionation
DIRAS2	BAIT	H. sapiens	Affinity Capture-MS
EEF2	HIT	H. sapiens	Co-fractionation
FGB	BAIT	H. sapiens	Affinity Capture-MS
FSD1	BAIT	H. sapiens	Affinity Capture-MS
GPX7	BAIT	H. sapiens	Affinity Capture-MS
LDHA	HIT	H. sapiens	Co-fractionation
LDHB	HIT	H. sapiens	Co-fractionation
MPZL1	BAIT	H. sapiens	Affinity Capture-MS
MVD	HIT	H. sapiens	Co-fractionation
NDRG1	HIT	H. sapiens	Co-fractionation
NFYA	BAIT	H. sapiens	Affinity Capture-MS
NPEPPS	HIT	H. sapiens	Co-fractionation
OR5F1	BAIT	H. sapiens	Affinity Capture-MS
PEPD	BAIT	H. sapiens	Co-fractionation
SDC1	BAIT	H. sapiens	Affinity Capture-MS
SLC9A3R1	HIT	H. sapiens	Co-fractionation
TYW3	BAIT	H. sapiens	Affinity Capture-MS
ZMAT3	BAIT	H. sapiens	Affinity Capture-MS
ZNF232	BAIT	H. sapiens	Affinity Capture-MS

Figure 3: Change in the amino acid sat different positions in *FTO* gene according to nsSNPs.

SNP ID	Picture	Description
rs140101381		change in the amino acid arginine to a tryptophan at position 80
rs139577103		change in the amino acid an arginine to a histidine at position 96
rs370009039		change in the amino acid a lysine to an asparagine at position 216

rs373076420		change in the amino acid a proline into a leucine at position 288
*rs200152693		change in the amino acid an alanine into a serine at position
rs368490949		change in the amino acid an arginine into a cysteine at position 41
rs368490949		change in the amino acid an arginine into a cysteine at position 337
rs372814208		change in the amino acid a proline into a leucine at position 103
rs139000284		change in the amino acid an arginine into a histidine at position 149
rs139000284		change in the amino acid a arginine into a histidine at position 445
rs376381270		change in the amino acid an arginine into a serine at position 159

rs370137051		change in the amino acid a leucine into a phenylalanine at position 193
rs370137051		change in the amino acid a leucine into a phenylalanine at position 489
polymirphism site wild type mutant type*No 3D structure was available for this SNP		

rs113383961 → D89E, rs151263395 → P93R, rs139577103 → R96H, rs147561986 → N143S, rs182784714 → L146M, rs370009039 → K216N, rs373076420 → P288L, rs200152693 → A15S, A311S, rs202007463 → E325V, E29V, rs368490949 → R41C, R337C, rs141978030 → N344D, re372814208 → P399L, P103L, rs143788264 → M400V, M104V, rs139000284 → R149H, R445H, rs376381270 → R159S, R455S and rs370137051 → L193F, L489F showed probably damaging result with score ≥0.996.

Tenpathological polymorphisms with their isoforms (rs143788264 → M1V, rs376527078 → G26,47V, rs145170223 → E5, 34, 55G, rs139000284 → R17, 46, 67H, rs376381270 → R56, 77S and rs370137051 → L90,111F) gave low deleterious prediction confidences by SIFT software were excluded in this study. This exclusion was considered limitation of this study.

The final total number of deleterious or damaging ns SNPs predicted both softwares (SIFT/polyphen-2) were 23 SNPs containing 31 pathological polymorphisms which we considered that the double positive prediction. The results are listed in Table (2).

Prediction of change in stability due to mutation used I-Mutant suite server

I mutant suite server output demonstrated that protein stability with relate free energy has changed due to mutation. all pathological polymorphisms[31] were detected by SIFT/ Polyphen-2 servers and according to Table(3) the results was represented as following: twenty ninepolymorphisms(K → E,L → I,R → W,P → R,R → H,N → S,L → M, G → A,V → I,M → V,K → N,P → L,A → S,E → V,R → C,N → D,D → G,R → S and L → F) predicted a dramatic decrease of the protein stability, while three polymorphisms (D → E,E → V and P → L) predicted increase of stability of FTO protein.

Modeling of amino acid substitution effects due to nsSNPs on protein structure

FTO Protein sequences of the nsSNP submitted to Project Hope revealed the 3D structure for the truncated proteins with its new candidates; in addition, it described the reaction and physiochemical properties of these candidates. According to Sift /Polyphen-2 results in which TOLERANCE INDEX ≤0.005and polyphen-2 score equals 1; we found that 10 pathological polymorphisms from 30 achieved these scores: rs139577103, rs370009039, rs373076420, rs200152693, rs139000284, rs376381270, rs370137051 with TI= 0 while rs368490949, rs140101381 and rs372814208 with TI= 0.001, 0.004 and 0.005respectively. These 10 SNPs were selected to be submitted to the Project Hope software and they indicated pathological polymorphisms change in the amino acids showed in the Figure (3).

Each amino acid has its own specific size, charge, and hydrophobicity-value, Table (4). The original wild-type residue and newly introduced mutant residue often differ in these properties:

In rs370009039, rs139000284, rs376381270, rs368490949 and rs139577103 the mutant residue was smaller than the wild-type residue while in rs373076420, rs200152693, rs372814208, rs140101381, rs370137051 the mutant residue was bigger than the wild-type residue.

The hydrophobicity of the wild-type and mutant residue differed in rs370137051and rs139577103 and we found the mutant residue was more hydrophobic than wild residue in rs140101381, rs200152693, rs368490949 and rs376381270.

In (rs200152693) the hydrophobic interactions, either in the core of the protein or on the surface was lost and the difference in hydrophobicity affected hydrogen bond formation in position 41 in (rs368490949).

There was a difference in charge between the wild-type and mutant amino acid in rs370137051, rs139000284, rs140101381, rs370009039andrs139577103 The charge of the wild-type residue was lost by this mutation which has caused loss of interactions with other molecules in these positions and the difference in charge will disturb the ionic interaction made by the original in position 159 in (rs370137051) and the (R → H)

Table 2: Prediction result of SIFT and Polyphen-2 programs.

SNP ID	Chromosome No./ Coordinate	Nucleotide Change	Amino Acid Change	Polyphen-2 Result	PSIC SD	SIFT Result	Tolerance Index
rs201836578	16/537381	A/G	K2E	PROBABLY DAMAGING	0.997	DELETERIOUS	0.008
rs371489995	16/53859803	C/A	L51I	POSSIBLY DAMAGING	0.891	DELETERIOUS	0.04
rs140101381	16/5385989	C/T	R80W	PROBABLY DAMAGING	1	DELETERIOUS	0.004
rs113383961	16/53859919	T/A	D89E	PROBABLY DAMAGING	0.982	DELETERIOUS	0.011
rs151263395	16/5385993	C/G	P93R	PROBABLY DAMAGING	1	DELETERIOUS	0.009
rs139577103	16/53859939	G/A	R96H	PROBABLY DAMAGING	1	DELETERIOUS	0
rs147561986	16/5386008	A/G	N143S	PROBABLY DAMAGING	0.999	DELETERIOUS	0
rs182784714	16/53860088	C/A	L146M	PROBABLY DAMAGING	1	DELETERIOUS	0.03
rs61743972	16/53860197	G/C	G182A	POSSIBLY DAMAGING	0.939	DELETERIOUS	0.043
rs150450891	16/53860253	G/A	V201I	POSSIBLY DAMAGING	0.679	DELETERIOUS	0.049
rs138348216	16/53860271	A/G	M207V	POSSIBLY DAMAGING	0.985	DELETERIOUS	0
rs370009039	16/538603	A/T	K216N	PROBABLY DAMAGING	1	DELETERIOUS	0
rs373076420	16/53878178	C/T	P288L	PROBABLY DAMAGING	1	DELETERIOUS	0
rs200152693	16/53907733	G/T	A15S	PROBABLY DAMAGING	1	DELETERIOUS	0
rs200152693	16/53907733	G/T	A311S	PROBABLY DAMAGING	0.996	DELETERIOUS	0.02
rs202007463	16/53907776	A/T	E325V	PROBABLY DAMAGING	0.999	DELETERIOUS	0.003
rs202007463	16/53907776	A/T	E29V	PROBABLY DAMAGING	0.997	DELETERIOUS	0.007
rs368490949	16/53913789	C/T	R41C	PROBABLY DAMAGING	1	DELETERIOUS	0.001
rs368490949	16/53913789	C/T	R337C	PROBABLY DAMAGING	1	DELETERIOUS	0.001
rs141978030	16/5391381	A/G	N344D	PROBABLY DAMAGING	0.99	DELETERIOUS	0
rs377073096	16/53913829	A/G	D350G	POSSIBLY DAMAGING	0.896	DELETERIOUS	0.03
re372814208	16/5392282	C/T	P399L	PROBABLY DAMAGING	0.997	DELETERIOUS	0.004
rs372814208	16/5392282	C/T	P103L	PROBABLY DAMAGING	1	DELETERIOUS	0.005
rs143788264	16/53922822	A/G	M400V	PROBABLY DAMAGING	0.985	DELETERIOUS	0
rs143788264	16/53922822	A/G	M104V	PROBABLY DAMAGING	1	DELETERIOUS	0.019
rs139000284	16/53967991	G/A	R149H	PROBABLY DAMAGING	1	DELETERIOUS	0
rs139000284	16/53967991	G/A	R445H	PROBABLY DAMAGING	1	DELETERIOUS	0
rs376381270	16/54145674	G/T	R159S	PROBABLY DAMAGING	1	DELETERIOUS	0
rs376381270	16/54145674	G/T	R455S	PROBABLY DAMAGING	0.999	DELETERIOUS	0
rs370137051	16/54145774	C/T	L193F	PROBABLY DAMAGING	1	DELETERIOUS	0
rs370137051	16/54145774	C/T	L489F	PROBABLY DAMAGING	1	DELETERIOUS	0

PolyPhen-2 result: POROBABLY DAMAGING (more confident prediction) / POSSIBLY DAMAGING (less confident prediction); PSIC SD: Position-Specific Independent Counts software if the score is ≥ 0.5; Tolerance Index: Ranges from 0 to 1; The amino acid substitution is predicted damaging if the score is ≤ 0.05 and tolerated if the score is > 0.05.

Table 3: Prediction result of I-Mutant software.

GENE NAME	SNP ID	RI	DDG	SVM2	WT	MT	AMINO ACID POSITION	TEMP	PH
FTO	rs201836578	6	DECREASE	-0.25	K	E	2	25	7
	rs371489995	6	DECREASE	-1	L	I	51	25	7
	rs140101381	5	DECREASE	-0.23	R	W	80	25	7
	rs113383961	2	INCREASE	-0.45	D	E	89	25	7
	rs151263395	8	DECREASE	-1.06	P	R	93	25	7
	rs139577103	8	DECREASE	-1.32	R	H	96	25	7
	rs147561986	6	DECREASE	-0.21	N	S	143	25	7
	rs182784714	7	DECREASE	-1.09	L	M	146	25	7
	rs61743972	4	DECREASE	-1.01	G	A	182	25	7
	rs150450891	6	DECREASE	-0.57	V	I	201	25	7
	rs138348216	7	DECREASE	-0.79	M	V	207	25	7
	rs370009039	1	DECREASE	-0.34	K	N	216	25	7
	rs373076420	4	DECREASE	-0.3	P	L	288	25	7
	rs200152693	7	DECREASE	-0.65	A	S	15	25	7
	rs200152693	8	DECREASE	-0.5	A	S	311	25	7

	rs202007463	2	DECREASE	-0.06	E	V	325	25	7
	rs202007463	1	INCREASE	0.3	E	V	29	25	7
	rs368490949	5	DECREASE	-0.85	R	C	41	25	7
	rs368490949	4	DECREASE	-0.94	R	C	337	25	7
	rs141978030	8	DECREASE	-0.51	N	D	344	25	7
	rs377073096	3	DECREASE	-1.34	D	G	350	25	7
	rs372814208	1	INCREASE	0.07	P	L	399	25	7
	rs372814208	7	DECREASE	-0.94	P	L	103	25	7
	rs143788264	6	DECREASE	-0.47	M	V	400	25	7
	rs143788264	8	DECREASE	-0.85	M	V	104	25	7
	rs139000284	7	DECREASE	-0.93	R	H	149	25	7
	rs139000284	7	DECREASE	-0.91	R	H	445	25	7
	rs376381270	9	DECREASE	-0.9	R	S	159	25	7
	rs376381270	8	DECREASE	-1.01	R	S	455	25	7
	rs370137051	7	DECREASE	-1.36	L	F	193	25	7
	rs370137051	7	DECREASE	-1.42	L	F	489	25	7

RI: Reliability Index; WT: amino acid in Wild Type; MT: amino acid in Mutant Type; DDG: $\Delta\Delta G$ sign; SVM: Support Vector Machine
DDG value: DG (New Protein)-DG (Wild Type) in Kcal/mole; SVM2 value: DDG < 0: decrease stability, DDG >0 increase stability

Table 4: amino acid prosperities according to result obtained from Project Hope software [34].

SNP ID	Amino Acid Change	Wild Type Properties				Mutant Type Properties			
		Size	Charge	Hydrophobisity	Conservation	Size	Charge	Hydrophobisity	Conservation
rs140101381	R80W	<	+charge	<	occurred with other residues	>	neutral	>	Near highly conserved position
rs139577103	R96H	>	+charge	-	only in this position	<	neutral	-	Near highly conserved position
rs370009039	K216N	>	+charge	-	only in this position	<	neutral	-	Near highly conserved position
rs373076420	P288L	<	-		only in this position	>	-		Near highly conserved position
rs200152693	A15S	<	-	<	-	>	-	>	-
rs368490949	R41C	>	+charge	<	not conserved	<	-	>	Near highly conserved position
	R337C	>	-	<	occurred with other residues	<	neutral	>	Near highly conserved position
rs372814208	P103L	<	-	-	not conserved	>	-	-	Near highly conserved position
rs139000284	R149H	>	+charge	-	not conserved	<	neutral		Near highly conserved position
	R445H	>	+charge	-	only in this position	<	neutral	-	Near highly conserved position
rs376381270	R159S	>	+charge	<	not conserved	<	neutral	>	Near highly conserved position
rs370137051	L193F	<	-	-	occurred with other residues	<	-	-	Near highly conserved position
	L489F	<	-	-	only in this position	<	-	-	Near highly conserved position

Size: >: bigger than; <: smaller than Hydrophobisity: >: more hydrophobic <: less hydrophobic

Table 5: SNPs in miRNA target sites.

dbSNP ID	Ancestral Allele	Allele	miR ID	Conservation	miRSite	Function Class
rs183282528	G	A	hsa-miR-369-3p	7	ccctGTATTATAg	C
			hsa-miR-374a-5p	8	ccctgTATTATAg	C
			hsa-miR-374b-5p	8	ccctgTATTATAg	C
			hsa-miR-5692b	8	ccctgTATTATAg	C
			hsa-miR-5692c	8	ccctgTATTATAg	C

rs141270327	G	G	hsa-miR-4763-5p	2	gGGCAGGCgacag	D
			hsa-miR-6078	2	gggCAGGCGAcag	D
rs78427245	G	G	hsa-miR-4437	4	aggaacGAGCCCA	D
			hsa-miR-4674	4	aggaacGAGCCCA	D
		A	hsa-miR-27a-5p	4	aggaacAAGCCCA	C
rs138932249	G	G	hsa-miR-2355-5p	3	tCTGGGGAAagac	D
			hsa-miR-3124-3p	3	tctggGGAAAGAc	D
			hsa-miR-3679-3p	3	tctGGGGAAAgac	D
			hsa-miR-6830-3p	2	tctgggGAAAGAC	D
		T	hsa-miR-1295b-5p	3	TCTGGGTAaagac	C
			hsa-miR-1912	3	TCTGGGTAaagac	C
			hsa-miR-3130-5p	3	tCTGGGTAaagac	C
			hsa-miR-4482-5p	3	tCTGGGTAaagac	C
			hsa-miR-5591-3p	3	tcTGGGTAAagac	C
rs2665269	G	G	hsa-miR-4659a-3p	9	atcatcGAAGAAA	D
			hsa-miR-4659b-3p	9	atcatcGAAGAAA	D
			hsa-miR-6875-3p	2	atcatcGAAGAAA	D
		A	hsa-miR-6844	2	atcatCAAAGAAa	C
rs2689252	A	A	hsa-miR-4659a-3p	2	catcGAAGAAAgt	D
			hsa-miR-4659b-3p	2	catcGAAGAAAgt	D
			hsa-miR-6875-3p	2	catcGAAGAAAgt	D
rs187896465	G	G	hsa-miR-3074-5p	3	gaGCAGGAAttct	D
			hsa-miR-370-3p	6	gAGCAGGAattct	D
			hsa-miR-6893-3p	6	gAGCAGGAattct	D
		A	hsa-miR-544a	3	gaGCAGAAAttct	C
			hsa-miR-6071	6	gAGCAGAAattct	C
			hsa-miR-6738-3p	3	gaGCAGAAAttct	C
			hsa-miR-6828-3p	3	GAGCAGAAattct	C
			hsa-miR-767-3p	3	GAGCAGAaattct	C
rs58094871	T					
		G	hsa-miR-5008-5p	2	gaagatGGGCCTC	C
rs45630756	G	G	hsa-miR-5092	3	tatttCGTGGATt	D
rs184637142	G	G	hsa-miR-5092	3	tatttCGTGGATt	D
rs116372569	C	C	hsa-miR-3126-5p	2	GTCCCTCAttctt	D
			hsa-miR-4419a	2	gTCCCTCAttctt	D
			hsa-miR-4510	2	gTCCCTCAttctt	D
			hsa-miR-6127	2	gTCCCTCAttctt	D
			hsa-miR-6129	2	gTCCCTCAttctt	D
			hsa-miR-6130	2	gTCCCTCAttctt	D
			hsa-miR-6133	2	gTCCCTCAttctt	D
			hsa-miR-6834-5p	2	gTCCCTCAttctt	D
			hsa-miR-6875-5p	2	GTCCCTCAttctt	D
		A	hsa-miR-3162-5p	2	gTCCCTAAttctt	C
			hsa-miR-7845-5p	2	GTCCCTAattctt	C
rs147993840	G					
		C	hsa-miR-196a-5p	5	ccACTACCTtcgt	C
			hsa-miR-196b-5p	5	ccACTACCTtcgt	C
rs141624910	G	G	hsa-miR-1197	2	agcttcGTGTCCT	D
rs73625209	G	G	hsa-miR-3622a-5p	3	gttaaCGTGCCTc	D

rs187960932	G	G	hsa-miR-1827	7	taacgTGCCTCAg	D
			hsa-miR-3622a-5p	3	taaCGTGCCTcag	D
			hsa-miR-4649-3p	9	taacgtGCCTCAG	D
		T	hsa-miR-3915	7	taacgTTCCTCAg	C
			hsa-miR-3928-3p	6	taacGTTCCTCAg	C
rs76366199	G	G	hsa-miR-25-5p	2	ccCTCCGCCAgtt	D
			hsa-miR-4730	2	cccTCCGCCAgtt	D
			hsa-miR-658	2	cCCTCCGCcagtt	D
		A	hsa-miR-3165	2	cccTCCACCAgtt	C
			hsa-miR-4456	2	ccctCCACCAGtt	C
			hsa-miR-582-3p	7	ccctccACCAGTT	C
			hsa-miR-6515-5p	2	CCCTCCAccagtt	C
			hsa-miR-6797-5p	2	CCCTCCAccagtt	C
			hsa-miR-6880-5p	2	ccCTCCACCAgtt	C
			hsa-miR-7847-3p	2	cCCTCCACcagtt	C
			hsa-miR-8071	2	cccTCCACCAgtt	C
rs76606072	C	C	hsa-miR-185-5p	3	acaTCTCTCCctt	D
			hsa-miR-4306	3	acaTCTCTCCctt	D
			hsa-miR-4644	3	acaTCTCTCCctt	D
			hsa-miR-6731-5p	3	acatCTCTCCCtt	D
			hsa-miR-6760-5p	3	acatcTCTCCCTt	D
			hsa-miR-8085	3	acatCTCTCCCtt	D
		T	hsa-miR-3153	3	acatCTTTCCCtt	C
			hsa-miR-5584-5p	3	acatcTTTCCCTt	C
			hsa-miR-6733-5p	3	acatCTTTCCCtt	C
			hsa-miR-6739-5p	3	acatCTTTCCCtt	C
rs116978290	G	G	hsa-miR-505-5p	8	gaccTGGCTCCtg	D
		A	hsa-miR-4433-3p	6	gacctgACTCCTG	C
			hsa-miR-5702	7	gacCTGACTCctg	C
-		-				
rs147129925	C					
		T	hsa-miR-4731-3p	2	ctttCTTGTGTAg	C
			hsa-miR-4801	2	ctttCTTGTGTAg	C
rs75618873	G					
		A	hsa-miR-7974	2	gttCACAGCCtgc	C
rs80079461	A	A	hsa-miR-3663-5p	2	tAGACCAGgaggg	D
			hsa-miR-4667-3p	10	tagaccAGGAGGG	D
			hsa-miR-601	2	TAGACCAggaggg	D
			hsa-miR-660-3p	2	tagacCAGGAGGg	D
		G	hsa-miR-4783-3p	2	tagACCGGGAggg	C
rs77984007	A	A	hsa-miR-4501	3	ttTCACATAaaat	D
		C	hsa-miR-6509-5p	5	tttcACCTAAAat	C
			hsa-miR-6800-5p	3	ttTCACCTAAaat	C
			hsa-miR-6802-5p	5	tttCACCTAAaat	C
rs113611471	C	C	hsa-miR-564	2	aCGTGCCAccacg	D
			hsa-miR-891a-3p	2	acGTGCCACcacg	D
		A	hsa-miR-1273g-5p	2	acgtgCAACCACg	C
			hsa-miR-656-5p	2	acgtGCAACCAcg	C
rs708277	G	G	hsa-miR-4786-5p	2	aggaTGGTCTCga	D
			hsa-miR-6502-3p	2	aggATGGTCTcga	D
		A	hsa-miR-4793-5p	2	AGGATGAtctcga	C
			hsa-miR-5196-3p	2	AGGATGAtctcga	C
			hsa-miR-5694	2	aggATGATCTcga	C
rs188814210	G	G	hsa-miR-24-3p	2	gcgTGAGCCAccg	D
			hsa-miR-4284	2	gcGTGAGCCAccg	D
		T	hsa-miR-3165	2	gcgtgATCCACCg	C

rs117637522	A	A	hsa-miR-3168	3	cttTAGAACTcag	D
			hsa-miR-4509	5	CTTTAGAactcag	D
			hsa-miR-4682	3	ctttagAACTCAG	D
			hsa-miR-4744	5	CTTTAGAactcag	D
			hsa-miR-4799-5p	6	cTTTAGAActcag	D
		G	hsa-miR-5009-3p	6	cTTTAGGActcag	C
rs139134234	A	A	hsa-miR-4303	2	CTCAGAAcaccca	D
			hsa-miR-4455	2	ctcagaACACCCA	D
			hsa-miR-609	2	ctcagAACACCCA	D
			hsa-miR-6772-5p	2	ctcagaACACCCA	D
		G	hsa-miR-3194-3p	3	ctCAGAGCAccca	C
			hsa-miR-3907	2	ctcaGAGCACCca	C
			hsa-miR-3921	2	CTCAGAGcaccca	C
			hsa-miR-4653-5p	2	CTCAGAGcaccca	C
			hsa-miR-5691	3	ctCAGAGCAccca	C
			hsa-miR-6741-5p	2	ctcagaGCACCCA	C
			hsa-miR-6805-3p	3	ctCAGAGCAccca	C
rs75681401	A	A	hsa-miR-4659a-5p	7	cctcgCATGGCAg	D
			hsa-miR-4659b-5p	7	cctcgCATGGCAg	D
			hsa-miR-4769-3p	3	cctcgcATGGCAG	D
			hsa-miR-6817-5p	3	cctcgcATGGCAG	D
		T	hsa-miR-4480	7	cctcgCTTGGCAg	C
rs79234192	G	G	hsa-miR-3929	2	gaccCAGCCTAtg	D
			hsa-miR-4419b	2	gaccCAGCCTAtg	D
			hsa-miR-4478	2	gaccCAGCCTAtg	D
			hsa-miR-4505	2	gaCCCAGCCtatg	D
			hsa-miR-5589-5p	2	gACCCAGCctatg	D
			hsa-miR-5787	2	gaCCCAGCCtatg	D
		A	hsa-miR-92a-1-5p	2	gaCCCAACCtatg	C
rs45564131	G	G	hsa-miR-3665	5	tgACCTGCAtcac	D
			hsa-miR-657	5	tgACCTGCAtcac	D
		T	hsa-miR-3974	4	TGACCTTcatcac	C
			hsa-miR-493-3p	5	tGACCTTCAtcac	C

D: The derived allele disrupts a conserved miRNA site; **C:** The derived allele creates a new miRNA site

pathological polymorphism will cause an empty space in the core of the protein in position 96.

The mutant residue in rs373676420, rs139000284, rs370137051 and rs370009030 at positions 228,445.489 and 216 was located near a highly conserved position and only wild residue type was found at this position. Mutation of a 100% conserved residue is usually damaging for the protein [32]. Other SNPs (rs140101381, rs368490949 and rs370137051) with positions 80, 41 and 337 respectively; the wild-type residue occurred often at these positions in the sequence, but other residues had also been observed in these positions. The mutant residue was located near a highly conserved position. In position 193, the (L → F) pathological polymorphism occurred at this position was probably not damaging.

The wild-type residue was not conserved in rs368490949, rs376381270, rs149000284 and rs372814208 at position 41,159,149 and 103 respectively while the mutant residue was located near a highly conserved position.

We suggested these variations in amino acid properties "size, charge, and hydrophobicity-value" due to these polymorphisms may produce new version of mutant protein which then may affect functions and structure of the original protein.

Functional SNPs in 3′untranslated regions (3′UTR) predicted by PolymiRTS data base 3.0

SNPs in 3′UTR of *FTO* gene were submitted as batch to PolymiRTS server. The output showed that among 187 SNPs in 3′UTR region of *FTO* gene, 30 SNPswere predicted, while among the 30 SNPs, 75 alleles disrupted a conserved miRNA site and 63 derived alleles created a new site of miRNA. RS76366199SNP contained (C) allele had 8 miRSites as target binging site can be disrupts a conserved miRNA while RS116372569 SNP had (D) allele with 9 miRSites that disrupts a conserved miRNA site, Table (5).

CONCLUSION

In conclusion, our results suggest that the application of computational tools like SIFT, PolyPhen-2, I mutant-3, Project Hope and polymRTS may provide an alternative approach for selecting target SNPs. The *FTO* gene is very important causative factor of obesity was investigated through computational methods and the influence of functional SNPs were evaluated.

In a total of 72808 SNPs in the *FTO* gene, 192 were found to be non-synonymous and 187 were found to be in the 3′untranslated regions, 31pathological polymorphisms were found to be deleterious and damaging by both Polyphen server and SIFT program, 30SNPsin the 3′ UTR were found to be significance. And according to polymirts results of 3′UTR we suggest that if we will disrupted a conserved miRNA sites we will decrease the gene expression and gene will be blocked. Hence, we hope our results will provide useful information that needed to help researchers to do further study to solve the obesity problem in the future.

REFERENCES

1. AT Ali, NJ Crowther. Factors predisposing to obesity: a review of the literature. JEMDSA. 2009; 14: 81-84.
2. Kopelman PG. Obesity as a medical problem. Nature. 2000; 404: 635-643.
3. Racette SB, Deusinger SS, Deusinger RH. Obesity: overview of prevalence, etiology, and treatment. Phys Ther. 2003; 83: 276-288.
4. Yazdi FT, Clee SM, Meyre D. Obesity genetics in mouse and human: back and forth, and back again. PeerJ. 2015; 3: 856.
5. Elks CE, den Hoed M, Zhao JH, Sharp SJ, Wareham NJ, Loos RJ, et al. Variability in the heritability of body mass index: a systematic review and meta-regression. Front Endocrinol. 2012; 3: 29.
6. Maes HH, Neale MC, Eaves LJ. Genetic and environmental factors in relative body weight and human adiposity. Behav Genet. 1997; 27: 325-351.
7. Hubacek JA, Bohuslavova R, Kuthanova L, Kubinova R, Peasey A, Pikhart H, et al. The FTO gene and obesity in a large Eastern European population sample: the HAPIEE study. Obesity. 2008; 16: 2764-2766.
8. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci. 2009; 106: 9362-9367.
9. Day FR, Loos RJ. Developments in obesity genetics in the era of genome-wide association studies. J Nutrigenet Nutrigenomics. 2011; 4: 222-238.
10. Lu Y, Loos RJ. Obesity genomics: assessing the transferability of susceptibility loci across diverse populations. Genome Med. 2013; 5: 55.
11. Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science. 2007; 316: 889-894.
12. Scuteri A, Sanna S, Chen WM, Uda M, Albai G, Strait J, et al. Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits. PLoS Genet. 2007; 3: 115.
13. Dina C, Meyre D, Gallina S, Durand E, Körner A, Jacobson P, et al. Variation in FTO contributes to childhood obesity and severe adult obesity. Nat Genet. 2007; 39: 724-726.
14. Peters T, Ausmeier K, Rüther U. Cloning of Fatso (Fto), a novel gene deleted by the Fused toes (Ft) mouse mutation. Mamm Genome. 1999; 10: 983-986.
15. Hubacek JA, Bohuslavova R, Kuthanova L, Kubinova R, Peasey A, Pikhart H, et al. The FTO gene and obesity in a large Eastern European population sample: the HAPIEE study. Obesity. 2008; 16: 2764-2766.
16. Legry V, Cottel D, Ferrières J, Arveiler D, Andrieux N, Bingham A, et al. Effect of an FTO polymorphism on fat mass, obesity, and type 2 diabetes mellitus in the French MONICA Study. Metabolism. 2009; 58: 971-975.
17. Villalobos-Comparán, M. Teresa Flores-Dorantes, M. TeresaVillarreal-Molina, Maricela Rodríguez-Cruz, Ana C García-Ulloa, LorRoblesena, et al. The FTO Gene Is Associated With Adulthood Obesity in the Mexican Population. Obes. 2008; 16: 2296 – 2301.
18. Kring SI, Holst C, Zimmermann E, Jess T, Berentzen T, Toubro S, et al. FTO gene associated fatness in relation to body fat distribution and metabolic traits throughout a broad range of fatness. PLoS One. 2008; 3: 2958.
19. Cha SW, Choi SM, Kim KS, Park BL, Kim JR, Kim JY, et al. Replication of genetic effects of FTO polymorphisms on BMI in a Korean population. Obesity. 2008; 16: 2187-2189.
20. Price RA, Li WD, Zhao H. FTO gene SNPs associated with extreme obesity in cases, controls and extremely discordant sister pairs. BMC Med Genet. 2008; 9: 4.
21. Wing MR, Ziegler J, Langefeld CD, Ng MC, Haffner SM, Norris JM, et al. Analysis of FTO gene variants with measures of obesity and glucose homeostasis in the IRAS Family Study. Hum Genet. 2009; 125: 615-626.
22. Tönjes A, Zeggini E, Kovacs P, Böttcher Y, Schleinitz D, Dietrich K, et al. Association of FTO variants with BMI and fat mass in the self-contained population of Sorbs in Germany. Eur J Hum Genet. 2010; 18: 104-110.
23. Cornes BK, Lind PA, Medland SE, Montgomery GW, Nyholt DR, Martin NG, et al. Replication of the association of common rs9939609 variant of FTO with increased BMI in an Australian adult twin population but no evidence for gene by environment (G x E) interaction. Int J Obes. 2009; 33: 75-79.
24. González-Sánchez JL, Zabena C, Martínez-Larrad MT, Martínez-Calatrava MJ, Pérez-Barba M, Serrano-Ríos M, et al . Variant rs9939609 in the FTO gene is associated with obesity in an adult population from Spain. Clin Endocrinol. 2009; 70: 390-393.
25. Hunt SC, Stone S, Xin Y, Scherer CA, Magness CL, Iadonato SP, et al. Association of the FTO gene with BMI. Obesity. 2008; 16: 902-904.
26. Boissel S, Reish O, Proulx K, Kawagoe-Takaki H, Sedgwick B, Yeo GS, et al. Loss-of-function mutation in the dioxygenase-encoding FTO gene causes severe growth retardation and multiple malformations. Am J Hum Genet. 2009; 85: 106-111.
27. Daoud H, Zhang D, McMurray F, Yu A1, Luco SM1, Vanstone J, et al. Identification of a pathogenic FTO mutation by next-generation sequencing in a newborn with growth retardation and developmental delay. J Med Genet. 2016; 53: 200-207.
28. Chatr-Aryamontri A, Breitkreutz BJ, Oughtred R, Boucher L, Heinicke S, Chen D, et al."The BioGRID interaction database. Nucleic Acids Res. 2015. 43: 470–478.
29. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. 2003; 31: 3812-3814.
30. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. Nucleic Acids Res. 2002; 30: 3894-3900.
31. Capriotti E, Calabrese R, Casadio R. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. Bioinformatics. 2006; 22: 2729-2734.
32. HankaVenselaar, Tim AH teBeek, Remko KP Kuipers, Maarten L Hekkelman, GertVriend. Protein structure analysis of mutations

causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. BMC Bioinformatics. 2010; 11:548.

33. Jia M, Yang B, Li Z, Shen H, Song X, Gu W. Computational analysis of functional single nucleotide polymorphisms associated with the CYP11B2 gene. PLoS One. 2014; 9: 104311.

34. Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M, et al. BioGRID: a general repository for interaction datasets. Nucleic Acids Res. 2006; 34: 535-539.

Cite this article

Osman MM, Khalifa AS, Yousri Mutasim AE, Massaad SO, Gasemelseed MM, et al. (2016) In Silico Analysis of Single Nucleotide Polymorphisms (Snps) in Human FTO Gene. J Bioinform, Genomics, Proteomics 1(1): 1003.