Journal of Rawalpindi Medical College (JRMC); 2016;20(4):269-274

Original Article

Association of ADIPOQ Non-Synonymous SNPs with Type 2 Diabetes-A Bioinformatics Study

Abdul Rauf [1], Muhammad Bilal [1], Farhan Rasheed [2], Shazia Shamas [3], Qasim Shahzad Butt [4], Aadil Hussain [4], M Zubair Mehbood [4]

1.Department of Pathology, Nawaz Sharif Medical College, University of Gujrat;2. Allama Iqbal Medical college, Lahore; 3. Department of Zoology, University of Swabi;5.Department of Biochemistry & Molecular Biology, University of Gujrat

Abstract

Background: To find out association of nsSNPs in adiponectin gene with Type 2 Diabetes.

Methods: In this study Bioinformatic analysis was done by using various bioinformatic tools to identify the nsSNPs present in ADIPOQ gene. The nsSNPs were further checked for their deleterious nature through different computational algorithms.

Results: This study identified 19 nsSNPs, out of which three most deleterious were selected i.e., rs372597136 (P47L), rs79645624 (G90S) and rs62625753 (R112H). Characterization and 3D models of superimposed structures of wild type and mutated protein showed the change occur due to presence of nsSNP.

Conclusion: Identified SNPs should be now checked in Pakistani population for their association with type-II diabetes. This study can provide platform for clinicians to design novel studies to identify association of genetic risk factors for different complex diseases.

Key Words: ADIPOQ Non-Synonymous SNPs,Type 2 Diabetes Mellitus ,Bioinformatics

Introduction

Diabetes mellitus type 2 is a metabolic disorder which occurs by hyperglycemia due to decreased insulin secretion, its action or both.[1] Type 2 diabetes usually present in patients with age above 35-40 years or sometimes below 30 in cases where there is high prevalence of diabetes in population. According to International Diabetes Foundation (IDF) the estimates for year 2011 were 366 million people affected with diabetes and for 2012 it were 371 million people globally. These include China (the most populous country) with 92.3 million people, India with 63 million people and United States with 24.1 million people. Moreover, in low and middle-income countries 4 out of 5 people suffered from diabetes.[2] In Pakistan type 2 diabetes is high ranging and it's 5.2 million which is between 7.6%-11% of population and it is estimated that in 2030 it will go up to 14 million (15% population). On diabetes prevalence list the rank of Pakistan is 7 and it falls in top 10 countries with people who range in 20-79 years.[3]

Adiponectin is the most important of all the adipokines, produced by the adipocytes, which have insulin sensitivity effect. It stimulates the oxidation of fatty acids with decreased level of triglycerides, and also enhance glucose metabolism in liver.[4] In type 2 diabetes plasma level of adiponectin is lowered and thus causes insulin resistance.[5] The location of ADPIQ gene is chromosome 3q27 with the size of 17kb and consisting 3 exons and 2 introns.[6]

Different clinical studies confirmed the hypothesis that various factors are involved in decreasing adiponectin level which ultimately result in type 2 diabetes, one of them is the presence of single nucleotide polymorphism (SNP) in gene.[7] These SNPs may comprise intronic, exonic and promoter regions of the gene. SNPs that present in exonic regions may or may not affect the protein sequence and are characterized as synonymous and non-synonymous. In this study computational tools have been applied to study nsSNPs in ADIOQ gene which can be associated with predisposition of Type-2 diabetes.

Material and Methods

For gene information of association ADIPOQ gene all the information was retrieved from ENSEMBLE v76. The gene has two transcripts and out of which the canonical transcript was selected. Then Variant Call Format (VCF file) for this transcript was collected from 1000 genomes project which was then run in snpEff-v4.1 software based on human genome assembly GRCh37.71. Then this data was used to find missense mutations present in ADIPOQ gene. snpEff results showed 19 missense mutations which were then run in different software like MutPred, PROVEAN, SIFT, polyphen-2 and I-Mutant 3.0 to check the functional stability of the obtained mutations. After confirmation through software, the most deleterious nsSNPs were

selected which were predicted through all the software.

Protein Variation Effect Analyzer (PROVEAN v1.1.3.) software tool was used . It predicts whether an amino acid substitution or indel has an impact on the biological function of a protein. The URL to access this tool is http://provean.jcvi.org. It works by clustering the sequences with query sequence using BLAST by global alignment. The closely related sequences will then generate the prediction score. The threshold score is -2.5 below this the protein is considered to have a deleterious effect. The input given was protein sequence and amino variations.Output was obtained in the form of scores and predictions.

Sorting Intolerant From Tolerant (SIFT v.1.03) was used to predict protein function based on sequence homology and physical properties of amino acids. The URL to access the tool is http://sift.bii.a-star.edu.sg/. It needs the protein sequence and amino acid substitution. Protein sequence is aligned by using PSI-BLAST and based on homology the score is calculated which gives the presence of substitution at particular site. Scores <0.05 are considered as in tolerated. The input given to the software was e-mail address, protein query sequence and substitution of interest. The parameters were set as default. Output obtained was in the form of scores.

Polymorphism Phenotyping v.2.2.2 (Polyphen-2) was used to predict change in structure and function due to the presence of amino acid substitution. The URL to access the tool is http://genetics.bwh.harvard.edu/pph2/. This prediction is based on parameters like sequence, phylogenetic and structural information characterized by substitution. The predictions are based on two types of data sets. The First is HumDiv, which was compiled from all damaging alleles with known effects on the molecular function, together with all the homologs that are considered as non-damaging. The second, HumVar consisted of all human disease-causing mutations from UniProtKB, together with common human nsSNPs without annotated involvement in disease, which were treated as non-damaging. All the information needed was protein query sequence, amino acid substitution and its position. Possibly Damaging (may or may not affecting the structure or function) and Benign (don't affect the function of protein). Polyphen scores ranged from 0-1. If score was near to 1, nsSNPs were designed as probably damaging.

I-mutant 3.0 web based tool used for the automatic prediction of protein stability changes upon single-site mutations. The URL to access the tool is http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi. The tool was trained on a data set derived from ProTherm. Input given was in the form of protein query sequence and substitution occurred in the protein. It predicted the stability change upon single site mutation starting either from the protein structure or the protein sequence. In either case I-Mutant3.0 predicted the direction of the free energy change (DDG) and its value (+/- DDG). DDG value with positive sign was the indication of mutated protein with the high stability.

After screening through all the software based on scores and predictions three most deleterious nsSNPs were selected (Table 1).

Table 1: Most Deleterious nSNPs

Mutation	ID	Polyphen	SIFT	Provean	I-Mutant	Allele change
Pro47Leu	rs372597136	Probably	T.031	Deleterious	Increase	CCG⇒CTG (275)
Gly90Ser	rs62625753	Probably	I.0.0	Deleterious	Decrease	GGT⇒AGT (403)
Arg112His	rs79645624	Benign	I.0.02	Deleterious	Decrease	CGC⇒CAC (470)

Modelling & Characterization: I-TASSER (Iterative Threading ASSEmbly Refinement) is a classified method for the protein structure and function prediction.http://zhanglab.ccmb.med.umich.edu. is the URL to open the server. It allows the users to make predicted high-quality 3-D structures and their functions through amino acid sequences. The 3-D structure of the protein was unknown and I-TASSER was able to build a structure. For each target protein, I-TASSER reported five best models through pair-wise structural similarity which correspond to five biggest clusters. Top 5 models predicted by I-TASSER were further verified through SAVES (The Structure Analysis and Verification Server) which verified through three programs: PROVE, ERRAT, VERIFY_3D.

PyMOL was used as downloaded software. It is used to produce high quality protein 3D structures. The input given is normal protein and mutated protein sequences. The output through PyMOL is then in superimposed 3D form which shows the change occurring due to substituted amino acid.

Results

ENSEMBLE-v76 revealed two transcripts of ADIPOQ gene, which give coding sequence of 244aa. Out of

which canonical transcript (ENST00000444204) was selected to produce a VCF file from 1000 genomes project, which was then run in snpEff-v4.1 software to find missense mutations. The results of snpEff- v4.1 software showed different types of variations like 3 prime UTR, 5 prime UTR, downstream, upstream, synonymous and non-synonymous variant. They are named according to the regions in which they are present (Table 2). There are total 481 variants and which can be differentiated by its types. These include 456 SNPs, 12 insertion, 11 deletion and 2 mixed variants. (Table 3). The total number of variants in ADIPOQ gene can also be differentiated by impact such as low, moderate and modifier and their relative percentages are 6.8%, 3.3% and 89.8% respectively (Table 4). Moreover, they can also be categorized by functional class as synonymous and non-synonymous. Total numbers of non-synonymous variants present are 57 which take 63.33% of total variations and the total count of non-synonymous variants is 33 which take 36.66% of total variations (Table 5). The missense variants were verified for functional stability through different software based on different algorithms. Provean v1.1.3 results showed whether amino acid substitution has an impact on protein function or not. The threshold score is -2.5 and below this the protein is considered to have a deleterious effect. Total 19 non-synonymous mutations were verified through Provean v1.1.3 and out of which 10 were considered as neutral as their score was above -2.5. The remaining 9 deleterious non-synonymous variants were: Pro47Leu, Arg55Cys, Arg55Leu, Gly90Ser, Pro91Ser, Arg112His, Arg131His, Ile164Thr and Asp209Gly (Table 6). SIFT-v.1.03 showed protein function based on sequence homology and physical properties of amino acids. The score was given which showed the presence of substitution at particular site. Scores <0.05 were considered as in-tolerated. Out of 19 non-synonymous mutations 11 were considered as tolerated and 8 as in-tolerated. The in-tolerated variants are Gly5Arg, Leu9Gln, Gly90Ser, Asp209Gly, Ile164Thr, Arg112His, Arg131His and Ala161Val (Table 7). Polyphen-2 (v.2.2.2) was used to predict change in structure and function due to the presence of amino acid substitution. This prediction was based on parameters like sequence, phylogenetic and structural information characterized by substitution. Polyphen-v.2.2.2 scores ranged from 0-1. If score was near to 1, nsSNPs were designed as probably damaging. When checked through polyphen-v.2.2.2 out of 19 non-synonymous variants 9 were benign (which don't affect the function of protein) and 10 were come in the category of probably damaging (affecting the structure or function). The probably damaging variants were: Leu9Gln, Pro47Leu, Arg55Cys, Arg55L, Gly90Ser, Pro91Ser, Arg131His, Ile164Thr, Asp209Gly and Glu220Gln (Table 8). Mutant 3.0 web based tool used for the automatic prediction of protein stability changes upon single-site mutations. The tool was trained on a data set derived from ProTherm. It predicted the stability change upon single site mutation starting either from the protein structure or the protein sequence. In either case I-Mutant3.0 predicted the direction of the free energy change (DDG) and its value (+/- DDG). DDG value with positive sign was the indication of mutated protein with the high stability. After confirmation through I-mutant3.0 there were only 3 variants whose energy was increased. And remaining 16 variants decrease their energy values. The 3 variants were: Thr22Ile, Pro47Leu, and Pro91Ser (Table 9). The comparison table made for all non-synonymous SNPs with their rs id. It has results of all prediction tools, and specific allele change (Table 10). After comparison of all prediction tools 3 most deleterious non-synonymous SNPs were selected randomly (Table 11;Figure 1). In Modelling & Characterization, I-TASSER (Iterative Threading Assembly Refinement) is a classified method for the prediction of protein structure and function. The 3-D structure of the wild type protein is unknown and I-TASSER is able to build a structure. For each target protein, I-TASSER reports five best models and further verified through SAVES (The Structure Analysis and Verification Server) which verifies through three programs: PROVE, ERRAT, VERIFY_3D. After then 3D models for mutated protein were designed by same method (Table 12;Figure 2). 3D structures of wild type and mutated models show the pathogenicity of nsSNPs. The first nsSNP is pro47leu in which the two residues differ in charge, hydrophobicity, and size. The mutant residue which is leucine is bigger than the wild type which is proline. Through superimposed 3D structure the wild type residue which is proline is very rigid and therefore it helps to make a specific confirmation to the backbone. And due to mutation this confirmation is disrupted (Fig 3). In nsSNP gly90ser the two residues differ in charge, size and hydrophobicity. According to 3D modelling the mutant residue is bigger in size than wild type and this change causes the presence of bumps. It can be visualized in 3D modelling that Glycine is so flexible to make torsion angles and for serine this is not possible to make these angles. Mutation will change the confirmation of local

Table 2. Variants by type and region identified in ADPIQ gene type region

Type (alphabetical order)	Count	Percent
3 prime UTR variant	396	20.776%
5 prime UTR variant	3	0.157%
Downstream gene variant	130	6.821%
Inframe deletion	3	0.157%
Intron variant	971	50.944%
Missense variant	57	2.991%
Non coding exon variant	133	6.978%
Sequence feature	96	5.037%
Splice region variant	4	0.21%
Synonymous variant	33	1.731%
Upstream gene variant	80	4.197%

Table 3.Number of variants by type In ADIPQ Gene Identified by SnpEff

Type	Total
SNP	456
MNP	0
INS	12
DEL	11
MIXED	2
INTERVAL	0
Total	481

Table 4. Variants by impact in ADIPQ gene identified by SnpEff

Type	Count	Percent
Low	130	6.821%
Moderate	63	3.305%
Modifier	1,713	89.874%

Table 5. Number of variants by functional class in ADIPQ gene identified by SnpEff

Type	Count	Percent
Non-synonymous	57	63.333%
Synonymous	33	36.667%
Exon	226	11.857%
Intron	971	50.944%
UTR	396	20.776%

Table 6. Type of variants based on threshold Score

Classification	Count
Neutral	10
Deleterious	9

Table 7. Type of Variants Based on Threshold Score

Classification	Count
Tolerated	11
In-tolerated	8

Table 8. Type of Variants Based on Threshold Score

22	Count
Benign	9
Probably damaging	10

Table 9. Type of Variants Based on Energy Values

Classification	Count	Percent
Increase	3	15.78%
Decrease	16	84.21%

Table 10. Number of Nonsynonymous Variants and their Probability

Mutation	ID	Polyphen	SIFT	Provean	I-Mutant	Allele change
Gly5Arg	rs201248773	Benign	I.0.00	Neutral	Decrease	GGA⇒AGA (97+148)
Leu9Gln	rs114155159	Probably	I.0.00	Neutral	Decrease	CTG ⇒ CAG (110+161)
Thr22Ile	rs201223375	Benign	T.0.19	Neutral	Increase	ACT ⇒ ATT (149+200)
Gly38asp	rs144448520	Benign	T.0.61	Neutral	Decrease	GGT ⇒ GAT (197+248)
Pro47Leu	rs372597136	Probably	T.031	Deleterious	Increase	CCG ⇒ CTG (224+275)
Arg55Cys	rs138227502	Probably	T 0.06	Deleterious	Decrease	CGT ⇒ TGT (247+298)
Arg55Leu	N/A	Probably	T 0.35	Deleterious	Decrease	CGT⇒CTT (299)
Gly90Ser	rs62625753	Probably	I.0.0	Deleterious	Decrease	GGT⇒AGT (352+403)
Pro91Ser	rs200130041	Probably	T 0.12	Deleterious	Increase	CCC ⇒ TCC (355+406)
Tyr111His	rs17366743	Benign	T 0.58	Neutral	Decrease	TAC ⇒ CAC (415+ 466)
Arg112His	rs79645624	Benign	I.0.02	Deleterious	Decrease	G⇒A (419+470)
Arg131His	rs78685763	Probably	I.0.02	Deleterious	Decrease	CGC ⇒ CAC (476+527)
Ala161Val	rs113716447	Benign	I.0.02	Neutral	Decrease	GCC ⇒ GTC (566+617)
Ile164Thr	rs185847354	Possibly	I.0.01	Deleterious	Decrease	ATC ⇒ ACC (575+626)
Asp209Gly	rs199733477	Probably	I.0.00	Deleterious	Decrease	GAC ⇒ GGC (710+761)
Glu220Gln	rs183590709	Possibly	T.0.26	Neutral	Decrease	GAG ⇒ CAG (742+793)
Arg221Ser	rs138773406	Benign	T.0.35	Neutral	Decrease	CGT ⇒ AGT (745+796)
His241Pro	rs141205818	Benign	T.0.31	Neutral	Decrease	CAT ⇒ CCT (806+857)
Asn244Ile	rs529727905	Benign	T.0.15	Neutral	Decrease	AAC ⇒ ATC (815+866)

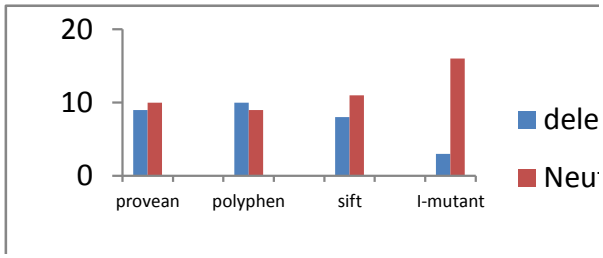


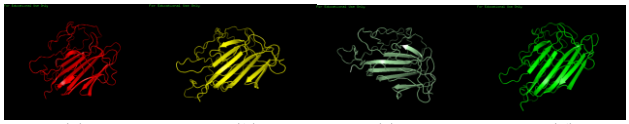
Fig 1 :Relative Count of nsSNPs Through Prediction Tools showing representation of all the non-synonymous variants confirmed by prediction tools. Blue bars indicate deleterious variants while red bar shows neutral variants.

Table 11. Most deleterious nsSNP

Mutation	ID	Polyphen	SIFT	Provean	I-Mutant	Allele change
Pro47Leu	rs372597136	Probably	I.0.031	Deleterious	Increase	CCG ⇒ CTG (224+275)
Gly90Ser	rs62625753	Probably	I.0.0	Deleterious	Decrease	GGT⇒AGT (352+403)
Arg112His	rs79645624	Benign	I.0.02	Deleterious	Decrease	G⇒A (419+470)

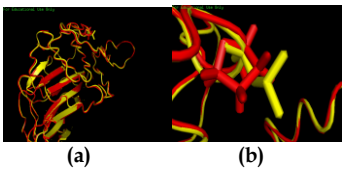
Table 12.Verification results of wild-type and mutated Protein

perification Software	Wild Type Model	Amino Acid Change		
		P47L	G90S	R112H
Verify-3D	74.03%	96.31%	82.38%	83.2%
ERRAT	84.02%	78.632%	76.27%	78.646%



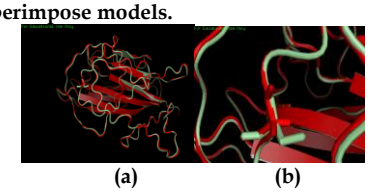
(a) (b) (c) (d)

Fig 2. 3D model design by using I-TASSER (a) Wild-type adiponectin protein (b) 3D model of mutated P47L protein (c) 3D model of mutated G90S protein (d) 3D model of mutated R112H protein



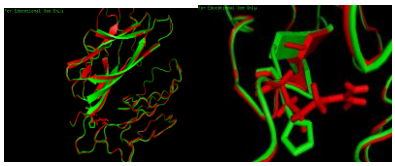
(a) (b)

Fig 3. superimpose image of P47L mutated model over wild type. Red colour represents wild type model while yellow color represents mutated model (b) magnified view of amino acid change in superimpose models.



(a) (b)

Fig 4. Superimpose image of G90S mutated model over wild type. Red colour represents wild type model while blue color represents mutated model (b) Magnified view of amino acid change in superimpose models.



(a) (b)

Fig5. Superimpose image of R112H mutated model over wild type. Red colour represents wild type model while green color represents mutated model (b) Magnified view of amino acid change in superimpose models.

Discussion

Adiponectin a protein, which is synthesized in adipose tissue, plays a significant role for hyperglycemia and dyslipidemia, and in inflammatory mechanisms which lead to a great risk of atherosclerosis in diabetes subjects.[8] The increased risk of inflammation, hyperglycemia and dyslipedimia has been attributed with the increased risk of coronary heart diseases (CHD).[9]Adiponectin is helpful in lipid metabolism for instance in the regulation of high HDL, cholesterol and low triglycerides.[10] Moreover, Adiponectin inhibits the TNF-α in both production and action.[11] Previous studies have shown that the adiponectin is predictor of insulin resistance and its action.[12] There is a strong correlation established in humans between the level of adiponectin and insulin sensitivity.[13]

SNP at +45T>G is the most common variant of exon 2 which is also associated with adiponectin level, obesity, insulin resistance and type 2 diabetes.[14] The previous studies show the association of adiponectin levels with promoter haplotypes which are -11391G>A and -11377C>G.[15] This investigation projected that altered serum adiponectin level is dependent on genetic variation which is shown to occur in gene promoter region in different population like German population.[15]

Human Genome Project along with HapMap has opened new perspectives for computational biologists. There is a large data of human variations and SNPs data which have been generated through HapMap is currently widely used in basic and applied research especially in pharmaceutical industry to design population based personalized medicine. In this study, we have tried to use genomic data through computational algorithms and validate it through wet lab techniques. By using canonical transcript of ADIPOQ, 19 nsSNPs have been identified by SnpEff, which are further evaluated by using different computational algorithms for their deleterious nature. Among 19 non-synonymous SNPs, three nsSNPs have been selected for this study, which are declared deleterious by all algorithms used in prediction software. Interestingly, in our study all three nsSNPs,rs372597136, rs79645624, rs62625753, have

been seen to be associated with type 2 diabetes. rs372597136 nsSNP results Proline change with leucine residue (P47L) at 47 position in ADIPOQ gene and could be associated with T2D. . This notion is also supported by analyzing properties of wild type and mutated residues. The mutant residue which is leucine is bigger than the wild type which is proline. Proline is very rigid and therefore it helps to make a specific confirmation to the backbone and due to mutation this confirmation is disrupted.

Conclusion

Identified SNPs should be now checked in Pakistani population for their association with type-II diabetes.

References

1. 1.Walston J, Silver K, Bogrdus C, Knowler WC. Time of onset of non insulin dependent diabetes mellitus. New Eng J Med 1995; 333: 346-49
2. Whiting DR, Guariguata L, Weil C. IDF Diabetes Atlas.: Global estimates of the prevalence of diabetes for 2011 and 2030. Diabetes Res Clin Pract 2011; 94: 311-21
3. Jafar TH, Levey AS, White FM . Ethnic differences and determinants of diabetes and central obesity among South Asians of Pakistan. Diabet Med 2004; 21: 716-23.
4. Yamauchi T, Kamon J, Waki H. The fat-derived hormone adiponectin reverses insulin resistance associated with both lipoatrophy and obesity. Nat Med2001; 7: 941–46.
5. Weyer C, Funahashi T, Tanaka S. Hypoadiponectinemia in obesity and type2 diabetes: close association with insulin resistance. J Clin Endocrinol Metab.(2001); 86:1930–35
6. Vionnet N, Hani El-H, Dupont S, Gallina S. Search for type 2 diabetes-susceptibility genes in French whites: evidence for novel susceptibility locus for early onset diabetes on chromosome 3q27-qter.Am J Hum Genet 2000; 67:1470–80.
7. Daimon M, Oizumi T, Saitoh T. Calpain 10 gene polymorphisms are related, not to type 2 diabetes, but to increased serum cholesterol. Diabetes Res Clin Pract 2002; 56:147–52.
8. Schulze M B, Rimm E B, Shai I, Rifai N.Relationship Between Adiponectin and Glycemic Control, Blood Lipids, and Inflammatory Markers in type 2 diabetes. Diabetes Care 2004; 27:1680–87
9. Haffner SM, Lehto S, Ronnemaa T, Pyorala K . Mortality from coronaryheart disease in subjects with type 2diabetes and in nondiabetic subjects with andwithout prior myocardial infarction.N Engl J Med 1998; 339:229–34
10. Hotta K, Funahashi T, Arita Y . Plasma concentrations of a novel, adipose-specific protein, adiponectin, in type 2 diabetic patients. Arterioscler Thromb Vasc Biol 2000; 20:1595–99
11. Yokota T, Oritani K, Takahashi I, Ishikawa J N. Adiponectin, a new member of the family of soluble defense collagens, negatively regulates the growth of myelomonocytic progenitors and the functions of macrophages. Blood 2000; 96: 1723–32.
12. Weyer C, Funahashi T, Tanaka S. Hypoadiponectinemia in obesity and type2 diabetes: close association with insulinresistance and hyperinsulinemia. J Clin Endocrinol Metab 2001; 86:1930–35
13. Chandran M, Phillips SA, Ciaraldi T. Adiponectin: more than just another fat cell hormone? Diabetes Care 2003; 26:2442–50
14. Heid IM, Wagner SA, Gohlke H . Genetic architecture of the APM1 gene and its influence on adiponectin plasma levels and parameters of the metabolic syndrome in 1,727 healthy Caucasians. Diabetes 2006; 55: 375–84.
15. Vasseur F, Helbecque N, Dina C. Single-nucleotide polymorphism haplotypes in the both proximal promoter and exon 3 of the APM1 gene modulate adipocyte-secreted adiponectin hormone levels and contribute to the genetic risk for type 2 diabetes. Hum Mol Genet 2002; 11:2607–14..