



Detection of genetic variants in HLA-DQA1 Gene with thyrotoxicosis patients in Thi- Qar city

Majida G. Maghtoof¹

Manal Badi Salih²

Enas Abdul Kareem Jabbar³

Mohammed Hashim Al Yassiry⁴

mahg5431@yahoo.com1

Enaskareemenaskareem@yahoo.com 3

mohammedhashim7882@yahoo.com4

1,2,3,4 Biology department / college science/ThiQar university

ABSTRACT:

This study was conducted to describe the single nucleotide polymorphism (SNPs) within the exon 2 of HLA-DQA1, a gene that encodes a MHC class II antigen that localized in the surface of many immune cells and they play a central role in the immune system by presenting peptides derived from extracellular proteins, as well as to analyze the consequences of non-synonymous SNPs in this gene using several *in silico* bioinformatics tools in the patients with Hyper thyrotoxicosis. Blood samples were collected, DNA was extracted, and one pair of specific PCR primers was designed to amplify exon 2 of this gene. PCR experiments were optimized and performed, and PCR products were purified and sequenced from both termini. Sequencing results were analyzed, and each observed nsSNP was further analyzed by several computational tools. Out of six observed SNPs, five of them were non-synonymous. It was found that the deleterious effect of these non-synonymous SNPs was varied, and the most damaging mutation was V68F followed by K92N. While the T49S and I99M have a mild parallel effect. However, no clear association was revealed between the observed SNPs of HLA-DQA1 gene and the progression of the Hyper thyrotoxicosis disease. Therefore, the associative pattern of this genetic variation with Hyper thyrotoxicosis disease should be interpreted with caution.

Keywords: Thyrotoxicosis, HLA-DQA1, Gene, *in silico*, polymorphism

INTRODUCTION

Thyrotoxicosis is a condition having multiple etiologies, manifestations, and potential therapies. The term thyrotoxicosis is the term applied when there is excess thyroid hormone in the circulation due to any cause, it can be easily diagnosed by high serum level of thyroxine (T4) and triiodothyronine (T3) and low serum level of thyroid stimulating hormone (TSH) (Maji, 2006). Hyperthyroidism is a form of thyrotoxicosis due to inappropriately high synthesis and secretion of thyroid hormone(s) by the thyroid (Rebecca, 2011). Hyperthyroidism and thyrotoxicosis are hypermetabolic conditions that cause significant morbidity and mortality (Cameron *et al* 2008; Devereaux and Tewelde 2014). Causes of thyrotoxicosis include Graves' disease, it's most common cause of hyperthyroidism, caused by antibodies to the thyroid-stimulating hormone receptor that is stimulatory in nature (Habra and Sarlis, 2005). The prevalence of hyperthyroidism varies between 0.1% and 0.4%. Graves' disease (GD) is the most common cause of thyrotoxicosis and it accounts for 60-80% cases of thyrotoxicosis, other causes of hyperthyroidism include genetic basis (Zaman, 2008), Strong evidence for a genetic basis to thyrotoxicosis comes from family studies demonstrating that a family history of GD has been reported in approximately 50% of patients one of the most abundant genes is HLA, The term HLA refers to the Human Leucocyte Antigen

URL: <http://www.uokufa.edu.iq/journals/index.php/ajb/index>

<http://iasj.net/iasj?func=issues&jid=129&uiLanguage=en>

Email: biomgzn.sci@uokufa.edu.iq



System, which is controlled by genes on the short arm of chromosome six at 6p21.1 to 6p21.3 and covers a region of about 3.6 Mbp depending on the haplotype, and contains at least 150 genes. The HLA loci are part of the genetic region known as the Major Histocompatibility Complex (MHC) is a cluster of genes with immunological and non-immunological functions and present in all vertebrates studied so far. (Kindt et al., 2007; Schulze, 2012). The current study focuses on HLA-DQA1 and relationship with thyrotoxicosis because there is strong evidence that gene plays important role in pathogenicity of thyrotoxicosis (Zeitlin et al. 2008). HLA-DQA1 is a human gene present on the short arm of chromosome 6 (6p21.3) and also denotes the genetic locus which contains this gene, HLA-DQA1 belongs to the HLA class II alpha chain paralogues. (Schmidt, et al., 2007). The aim of this study conducted to describe the single nucleotide polymorphism (SNPs) within the exon 2 of HLA-DQA1.

MATERIALS AND METHODS:

Sampling and DNA Extraction.

The study was conducted on a thyrotoxicosis samples... These samples were diagnosed by endocrinologist at the Diabetes and Endocrines Medical Center / Al-Nassiryhia city. They were of both genders with age range of (18-50). Genomic DNA was isolated using a mammalian genomic DNA extraction kit (Geneaid Biotech, Taiwan). The extracted DNA was evaluated by 0.8% agarose gel electrophoresis in 1X TAE (40 mM Tris acetate; 2 mM EDTA, pH 8.3), and quantified using a nanodrop (BioDrop μ LITE, Biodrop, UK).

PCR Design

Concerning the exon 2 of *HLA-DQA1* genetic locus, one pair of primers was successfully designed using the NCBI primer-BLAST online software (GenBank Acc. No U03675.1). The designed oligonucleotide primer pair is; forward 5'- TACCAGTCTTACGGTCCCTCTG -3', and reverse 5'- GGTAGCAGCGGTAGAGTTGG -3', which designed to amplify 210 bp to partially cover exon 2 of *FST* gene.

PCR Analysis

The PCR reaction was performed using *AccuPower* PCR premix (Cat # K-2012, Bioneer - Korea). Each 20 μ l of PCR premix was contained 1 U of *Top* DNA polymerase, 250 μ M of dNTPs, 10 mM of Tris-HCl (pH 9.0), 30 mM of KCl, and 1.5 mM of MgCl₂. The reaction mixture was completed with 10 pmol of each primer and 50 ng of genomic DNA. Deionized nuclease-free water was added to the final volume (Cat No. C-9011-2. Bioneer - Korea). The optimum annealing temperatures were determined empirically in our extracted DNA template using gradient PCR (ver. Mastercycler-nexus, Eppendorf, 22331Hamburg). The amplification was began by initial denaturation at 94°C for 5 min, followed by 30 cycles of denaturation at 94°C for 30 sec, annealing at 58 °C for 30 sec, and elongation at 72°C for 30 sec, and was concluded with a final extension at 72°C for 5 min. After performing PCR thermocycling, PCR products were verified by electrophoresis on a 1.5% (w/v) agarose gel in 1X TBE buffer (2 mM of EDTA, 89 mM of Tris-Borate, pH 8.3), using a 100-bp ladder (Cat # D-1010, Bioneer - South Korea) as a molecular weight marker for confirmation of the length of the PCR products. Gels were stained with ethidium bromide (0.5 mg/ml) and visualized with gel image documentation system (ChemiDoc, Bio-Rad, USA). All SSCP non-suitable amplicons bands were eliminated.

DNA sequencing and sequencing Analysis

Each unique sample's pattern for the amplified *HLA-DQA1* gene (210bp) fragment was purified from agarose gel by using (EZ EZ-10 Spin Column DNA Gel Extraction Kit,

URL: <http://www.uokufa.edu.iq/journals/index.php/ajb/index>

<http://iasj.net/iasj?func=issues&jid=129&uiLanguage=en>

Email: biomgzn.sci@uokufa.edu.iq



Biobasic. Canada), and sequenced from both ends (Macrogen Inc. Geumchen, Seoul, South Korea). Only clear chromatographs obtained from ABI sequence files were further analyzed, ensuring that the annotation and variations are not because of PCR or sequencing artifacts. The reference sequences of exon 2/*DQA1* gene were retrieved from the NCBI website (<http://www.ncbi.nlm.nih.gov>). The sequencing results of the PCR products were edited, aligned, and compared with their reference sequences using *BioEdit* Sequence Alignment Editor Software Version 7.1 (DNASTAR, Madison, WI, USA).

Protein Structure Designing

The primary structure designing of each FST genotype were begun by mutating the available reference NCBI DNA sequences of exon 2/*HLA-DQA1* gene, by substituting each observed SNP into its reference sequences using *BioEdit* /Lasergene software to represent the non-synonymous twenty sequenced PCR products. The whole amino acid sequences of the humanMHC class II antigen were retrieved online from the NCBI (GenBank: AIC59010.1). The reference nucleic acid sequences, as well as their twenty variants was translated into amino acids in a reading frame corresponds to the reference *HLA-DQA1* amino acid sequences using the *Expasy* online program (<http://web.expasy.org/translate/>). Multiple amino acid sequences alignment was made between the reference exon 2/ *HLA-DQA1* amino acid sequences and its observed genotype using *UniprotKB* software (www.uniprotkb.org), and the results were represented using *BioEdit* /Lasergene software. The three-dimensional structure of the *HLA-DQA1* gene was constructed from the online three dimensional model prediction software; protein prediction engine, RaptorX(<http://raptorx.uchicago.edu/StructurePrediction/predict/>). The virtual proposed changes within its corresponding mutants was performed by using PyMOL-v1,7.0.1 software (www.shrodinger.com).

Finding deleterious nsSNP(s) by the SIFT program

SIFT (sorting intolerant from tolerant) software predicts whether substitution with any of the other amino acids is tolerated or deleterious for every position in the submitted sequence (Patel *et al.*, 2015). The SIFT prediction was given as a tolerance index (TI) score ranging from 0.0 to 1.0, which was the normalized probability that the amino acid change was tolerated. A nsSNP with a TI score of ≥ 0.05 was considered to be deleterious *i.e.* an amino acids with probabilities < 0.05 were predicted to be deleterious. The amino acid sequence of / *HLA-DQA1* along with nsSNPs with corresponding amino acid positions were submitted SIFT program (http://sift.bii.a-star.edu.sg/sift-bin/SIFT_seq_submit2.pl).

Predicting the functional effect of the observed nsSNPs using PolyPhen-2

PolyPhen-2, or Polymorphism Phenotyping version 2, is a tool which predicts possible impact of an amino acid substitution on the structure and function of the protein using straightforward physical and comparative considerations. PolyPhen-2 server is a new development of the PolyPhen tool for annotating coding nonsynonymous SNPs (Adzhubei *et al.*, 2010). This machine learning based method can provide a high throughput prediction tool source of the query nsSNPs (<http://genetics.bwh.harvard.edu/pph2/>). Protein sequence database, wild type, and mutant amino acid position are the input requirements for this server. Prediction outcomes could be classified as probably damaging or benign according to the score ranging from 0 – 1 respectively.

Validation of the functional characterization of nsSNPs(s) by the PANTHER program

The SIFT predicted nsSNPs were validated by PANTHER (protein analysis through evolutionary relationship) program (<http://www.pantherdb.org/tools/csnp>). This tool estimates the likelihood of a particular non-synonymous coding SNP to cause a functional impact on the

URL: <http://www.uokufa.edu.iq/journals/index.php/ajb/index>

<http://iasj.net/iasj?func=issues&jid=129&uiLanguage=en>

Email: biomgzn.sci@uokufa.edu.iq



protein using Hidden Markov Models (HMM) based modeling and evolutionary relationship. It calculates the subPSEC (substitution position-specific evolutionary conservation) score based on an alignment of evolutionarily related proteins. The PSEP (position-specific evolutionary preservation) measures the length of time (in millions of years) a position in current protein has been preserved by tracing back to its reconstructed direct ancestors. The longer a position has been preserved, the more likely that it will have a deleterious effect. The thresholds we chose were: "probably damaging" (time > 450my, corresponding to a false positive rate of ~0.2 as tested on HumVar), "possibly damaging" (450my > time > 200my, corresponding to a false positive rate of ~0.4) and "probably benign" (time < 200my) (Tang and Thomas, 2016).

Prediction of functional impact of the observed nsSNPs using PROVEAN

Further confirmation of the effect of nsSNPs on protein was done using PROVEAN (Protein Variation Effect Analyzer) tool (Choi *et al.*, 2012). PROVEAN tool can predict the impact of an amino acid substitution on the biological function of a protein (<http://provean.jcvi.org/index.php>). This algorithm allows for the best-balanced separation between the deleterious and neutral amino acids, based on a threshold. The default threshold of PROVEAN tool is -2.5, *i.e.*, variants with a score equal to or below -2.5 are considered "deleterious", while variants with a score above -2.5 are considered "neutral".

Prediction of the Severity Effect of the nsSNPs using SNAP2

SNAP2 (Screening for Non-Acceptable Polymorphisms) tool is based on a machine learning device called "neural network". It distinguishes between effect and neutral variants/nonsynonymous SNPs by taking a variety of sequence and variant features into account (Smigielski *et al.*, 2000). The most important input signal for the prediction is the evolutionary information taken from an automatically generated multiple sequence alignment. SNAP2 revealed that if the nsSNP has scored more than zero (> 0) it has a damaging effect and vice versa. Predicting a score (ranges from -100 strong neutral prediction to +100 strong effect prediction), the analysis suggests that the prediction score is to some extent correlated with the severity of effect (<https://www.predictprotein.org>).

Predicting disease – associated SNPs using SNPs&Go

The Predicting disease associated variations using GO terms or SNPs&GO is a support vector machine algorithm based web server, which is used to predict the impact of variations of MHC class II antigen by calculating the functional information such as biological process, cellular component and molecular function, which are arrayed by Gene Ontology (GO) data base (Capriotti *et al.*, 2013; Calabrese *et al.*, 2009). The FASTA sequence of native HBA1 protein and the list of variations were provided as input (S1 File). Probability values >0.5 for each variant was predicted as disease nsSNP.

Investigation of altered protein stability using I-Mutant 2.0

To have a better insight to the stability of the protein caused by mutation, the altered positions were analyzed using I-Mutant 2.0 (Capriotti *et al.*, 2005). The study of the effect of a mutation may alter the stability of the studied protein of interest and might lead to a change in the main characteristics of this protein. I-Mutant 2.0 is a Support Vector Machine-based web server for the automatic prediction of protein stability changes upon single-site mutations (<http://folding.biofold.org/cgi-bin/i-mutant2.0>). The input of the whole sequence of the protein of interest along with the residues change was provided for analysis of DDG value (kcal/mol). The DDG values were computed at temperature 25°C and pH 7.0. If the DDG value is > 0, protein stability decreases and when DDG value is < 0 protein stability increases.



Identification of nsSNPs on ligand binding sites using FTsite server

An FTsite server (<http://ftsites.bu.edu>) was used to predict ligand binding sites of the studied proteins. FTsite accurately identifies binding sites in over 94% of apo-proteins, including the structure-based prediction of protein, the explanation of functional relationships among proteins, and protein engineering (Ngan *et al.*, 2012). The query protein sequence data were submitted as RaptorX-built pdb files.

Analysis of The nature of nsSNPs regions of the studied nsSNPs using ConSurf

ConSurf web-server (consurf.tau.ac.il/) was used to determine the evolutionarily conserved regions of the MHC class II antigen (Ashkenazy *et al.*, 2010). In ConSurf tool, MHC class II antigen sequence homologs were aligned and position-specific scores were calculated using an empirical Bayesian algorithm. The conserved to variable regions were predicted via conservation scores and a coloring scheme and further divided into distinct scales of nine grades.

RESULTS AND DISCUSSION

The rapid progress in our understanding of genetic variation shows that many different types of genetic changes affect complex phenotypes. Indeed, it was known that SNPs are the general form of genetic variations among individuals and are thought to be responsible for the majority of inherited diseases. The focus of this study has been made on the analysis of the SNPs consequences of HLA-DQA1 gene and relationship with thyrotoxicosis because there is strong evidence that gene plays a crucial role in pathogen of thyrotoxicosis.

The polymorphisms of the HLA-DQA1 gene were detected by direct PCR-sequencing strategy. The *in vitro* results indicated the observation of six SNPs (43A>T, 100G>T, 150A>G, 162C>G, 174A>T, 195T>G). These nsSNPs were positioned in variable loci of the HLA-DQA1 gene (Fig. 1). This study indicated that there are four common SNPs that are available in all examined samples whether patients or healthy control samples (100G>T, 162C>G, 174A>T, 195T>G). However, both of the non-common SNPs was found to in both healthy and patient samples. Each observed SNP was further analyzed to identify whether it substitutes the resulting encoded protein. The uniprotKB tool was revealed that five of the observed SNPs are non-synonymous or nsSNPs, which are 49T>S, 68V>F, 88I>M, 92K>N, and 99I>M (Fig. 2). The exact positions of the nsSNPs was localized and represented in the whole encoded protein (Table 1). On the other hand, the consequences of these nsSNPs was evaluated using different publicly available computational algorithms, namely, SIFT, PolyPhen-2, PANTHER, PROVEAN, SNAP2, SNPs&Go, and I-Mutant 2 bioinformatics tools. By comparing the prediction of these methods, the deleterious consequences of these nsSNPs were confirmed in the whole utilized tools collectively. Usually, the SIFT tool is utilized first, which predicts whether an amino acid substitution affects protein function among related genes and domains over evolutionary time (Ng and Henikoff, 2006). In addition, the SIFT predicted characterization of each nsSNP was further validated by PANTHER through investigating the effect of this nsSNP on MHC class II antigen function using HMM-based tool (Tang and Thomas, 2016). Both previous tools were showed that all the observed nsSNPs were tolerated or benign except V68F. Further validation of these nsSNPs was provided through other several tools, such as PolyPhen-2, PROVEAN, SNAP2, and SNPs&Go respectively (Adzhubei *et al.*, 2010; Choi *et al.*, 2012; Smigielski *et al.*, 2000; Capriotti *et al.*, 2013). However, these computational tools that were utilized in the present study have shown slightly different results (Table 2). On the other hand, once the 3-D model was being built, FTsite program was applied. The FTsite program was usually implemented to predict whether each amino acid being mutated is positioned in



a ligand binding site in the protein of interest (Kozakov *et al.*, 2015). The FTsite results indicated that there is no potential ligand binding site for all the observed nsSNPs of the present study.

After applying all the computational tools on all the observed nsSNPs of the present study it was found that the most dangerous nsSNP is V68F that is followed by K92N. In addition, both I88M and I99M may have moderate effect on protein function, while the lowest deleterious nsSNP is T49S. To add another layer of confirmation, the effect of the observed nsSNPs was also analyzed using I-Mutant 2.0, which gives result in the form of effect of mutants on stability of protein with reliability index at pH 7.0 and temperature 25 °C (Capriotti *et al.*, 2005). The I-Mutant 2.0 tool was revealed that all nsSNPs were reduced the stability of the studied protein upon mutation except I99M (Table 3). However, the differences in the variable computational tools are belonging to the algorithmic differences for each utilized tool in the bio-computational analysis. Therefore we would expect the outcomes to occur dissimilar at some point (de Alencar and Lopes, 2010). However, from amino acid sequence retrieved from NCBI, 3D model of this protein was constructed using RaptorX protein modeling tool and visualized in PyMOL. All SNPs were inserted in the native sequence of protein and its consequences were checked using several computational tools. PyMOL was used to locate nsSNPs on protein three-dimensional structure and for analyzing both native and mutant structures. The MHC class II antigen is having 254 amino acid residues. According to RaptorX model (Källberg *et al.*, 2012), from these 254(100%) residues were modeled and 8(3%) positions were predicted as disordered. Secondary structures revealed 25% helix, 31% beta sheet and 43% loop structures. Once the three-dimensional structure of MHC class II antigen is built, conserved and variable regions in MHC class II antigen were predicted using Consurf tool. The conservation scores were calculated by Consurf tool in an evolutionary pattern, in which some positions evolve slowly and are commonly referred to as "conserved", while others evolve rapidly and are referred to as "variable" (Ashkenazy *et al.*, 2010). Though Consurf tool was revealed that the most observed nsSNPs were occupied several degrees of variable regions (Fig. 3). Moreover, all the deleterious and non-deleterious amino acids substitutions were visualized and regarding the three-dimensional structure of MHC class II antigen in each variant (Fig. 4, A). However, two forms of substituted proteins were emerged from the examined samples (Fig. 4, B and C). In both forms, both healthy and patients samples were included. Therefore, the observed nsSNPs weren't involved in the pathogenesis of this disease. However, it is not necessary that all variants have a major deleterious functional impact and some may be well tolerated. However, nsSNPs which are linked to diseases or other phenotypes often have some molecular significance (Johnson *et al.*, 2005). Nevertheless, all the healthy control samples, as in variants no. 2, 5, 13, and 15, weren't shown the 43A>T SNP. Even though, this data weren't sufficient to build an obvious view for such pathogenic phenomenon. In conclusion, the studied MHC class II antigen has several nsSNPs despite of their high degree of variability they didn't shown any clear association with the presence of any suggestive causal SNP. Nonetheless, the utilized bioinformatics tools were also provided a predictive insight for the possible consequences of each observed SNP in both the structure and the biological function of the *HLA-DQA1* gene product. In another study have subsequently replicated association between GD and DR3 in both case-control and family based datasets and gone on to demonstrate association between the DRB1*03-DQB1*02-DQA1*0501 extended haplotype and GD conferring relative risks (RR) for the development of disease between 1.9 and 3.8 (Gough, 2000). Yanagawa *et al.* (1993) have reported a significant increase in DQA1*0501 in USA Caucasians subjects with GD. and proposed that the DQA1*0501 association was independent of DR3 with DQA1*0501 conferring a greater risk. On the other hand, Heward *et al.* (1998) performed the largest case-



control investigation of the association of the class II region with GD in a UK population to date, and provided confirmation in an independent family dataset. The strong association of both DRB1*03 (RR=2.45) and DQA1*0501 (RR=2.26) was found but no independent effect of DQA1*0501 was seen. Due to linkage between DRB1*03, DQB1*02 and DQA1*0501, distribution of the haplotype DRB1*03-DQB1*02-DQA1*0501 between GD and control subjects were analyzed, showing association with GD (RR=2.52) which was confirmed in the family dataset.

Conclusion: this study had been made on the analysis of the SNPs consequences of HLA-DQA1 gene and relationship with thyrotoxicosis, The in vitro results indicated the observation of six SNPs (43A>T, 100G>T, 150A>G, 162C>G, 174A>T, 195T>G) and after applying all the computational tools on all the observed nsSNPs of the present study it was found that the most dangerous nsSNP is V68F that is followed by K92N. In addition, both I88M and I99M may have moderate effect on protein function, while the lowest deleterious nsSNP is T49S. To add another layer of confirmation.

Even though, this data weren't sufficient to build an obvious view for such pathogenic phenomenon. In **conclusion**, the studied MHC class II antigen has several nsSNPs despite of their high degree of variability they didn't shown any clear association with the presence of any suggestive causal SNP

References

- Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., Sunyaev, S.R., 2010. A method and server for predicting damaging missense mutations. *Nat. Methods* 7(4), 248–249. doi: 10.1038/nmeth0410-248
- Ashkenazy, H., Erez, H., Martz, E., Pupko, T., Ben-Tal, N., 2010. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucl. Acids Res.* W529–33. doi: 10.1093/nar/gkq399
- Calabrese R, Capriotti E, Fariselli P, Martelli PL, Casadio R. Functional annotations improve the predictive score of human disease-related mutations in proteins. *Hum Mutat.* 2009; 30(8): 1237–1244. doi: 10.1002/humu.21047 PMID: 19514061.
- Cameron, A.J.; Magliano, D.J.; Zimmet, P.Z.; Welborn, T.A.; Colagiuri, S.; Tonkin, A.M and Shaw, J.(2008).The metabolic syndrome as a tool for predicting future diabetes: the AUSDIAB study. *J Intern Med.* 264:177-186.
- Capriotti E, Calabrese R, Fariselli P, Martelli PL, Altman RB, Casadio R. WS-SNPs&GO: a web server for predicting the deleterious effect of human protein variants using functional annotation. *BMC Genomics.* 2013; 14 Suppl 3(Suppl 3): S6. doi: 10.1186/1471-2164-14-S3-S6 PMID: 23819482.
- Capriotti, E., Fariselli, P., Casadio, R., 2005. I-Mutant 2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucl. Acids Res.* 33, 306–310. doi: 10.1093/nar/gki375
- Choi, Y., Sims, G.E., Murphy, S., Miller, J.R., Chan, A.P., 2012. Predicting the functional effect of amino acid substitutions and indels. *PLOS ONE* 7, e46688. doi: 10.1371/journal.pone.0046688



- de Alencar, S.A., Lopes, J.C.D., 2010. A Comprehensive In silico Analysis of the Functional and Structural Impact of SNPs in the IGF1R Gene. *J. Biomed. Biotechnol.* ID 715139, 8 pages. doi:10.1155/2010/715139
- Devereaux, D. and Tewelde S.Z. (2014) Hyperthyroidism and thyrotoxicosis. *Emerg Med Clin North Am.* ;32(2):277-92
- Gough SC. The genetics of Graves' disease. *EndocrinolMetabClin North Am* 2000; 29:255–66.
- Habra, M. and Sarlis, N. (2005) Thyroid and Aging. *Reviews in Endocrine & Metabolic Disorders*;6:145–154
- Heward JM, Allahabadia A, Daykin J etal. Linkage disequilibrium between the human leukocyte antigen class II region of the major histocompatibility complex and Graves' disease: replication using a population case control and family-based study. *J ClinEndocrinolMetab* 1998; 83:3394–7.
- Johnson M.M., Houck J., Chen C. (2005). Environmental aspects selection for EEE using ANP method. Screening for Deleterious Nonsynonymous Single-nucleotide Polymorphisms in Genes Involved in Steroid Hormone Metabolism and Response, 14; 1326–1329.
- Källberg M, Wang H, Wang S, Peng J, Wang Z, Lu H, Xu J. (2012). Template-based protein structure modeling using the RaptorX web server. *Nature Protocols* 7, 1511-1522,
- Kindt, T. J.; Osborne, B. A. ; Goldsby, R. A. and Kubly, J. editors.(2006). *Kuby Immunology.* 6th edition. W.H. Freeman and Company, New York, NY
- Kozakov, D., Grove, L.E., Hall, D.R., Bohnuud, T., Mottarella, S.E., Luo, L., Xia, B., Beglov, D., Vajda, S., 2015. The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. *Nat. Protoc.* 10(5), 733–755. doi: 10.1038/nprot.2015.043
- Lau M.; Terasaki PI. and Park MS (1994). "International Cell Exchange,(1994). *ClinTranspl:* 467–88. PMID 754-7576
- Maji D. (2006) Hyperthyroidism. *J Indian Med Assoc.* ;104(10):563-4, 566-7.
- Ng PC, Henikoff S (2006): Predicting the effects of amino acid substitutions on protein function," *Annual Review of Genomics Human Genetics*, 7, 61–80. doi: 10.1146/annurev.genom.7.080505.115630
- Ngan, C.H., Hall, D.R., Zerbe, B.S., Grove, L.E., Kozakov, D., Vajda, S., 2012. FTSite: high accuracy detection of ligand binding sites on unbound protein structures. *Bioinformatics* 28, 286–287. doi: 10.1093/bioinformatics/btr651.
- Patel SM, Koringa PG, Reddy BB, Nathani NM, Joshi CG. (2015). In silico analysis of consequences of non-synonymous SNPs of Slc11a2 gene in Indian bovines. *Genomics Data* 5: 72–79. <http://dx.doi.org/10.1016/j.gdata.2015.05.015>
- Rebecca S. ; Henry , B.; David, S.; Jeffrey R.; Klein,I; Laurberg,P.; Victor M.; Scott A.; Douglas S.; Sosa,J.; Marius N. (2011) Hyperthyroidism and other causes of thyrotoxicosis: Hyperthyroidism Management , *EndocrPract.* 17(No. 3)P. e3.



- Schmidt, H.; Williamson, D. and Ashley-Koch, A. (2007). "HLA-DR15 haplotype and multiple sclerosis: a HuGE review.". *Am. J. Epidemiol.* 165 (10): 1097–109.
- Schulze M.S.(2012). The mechanism of peptide exchange by MHC class II molecules Ph.D. thesis. Department of Biology, Chemistry and Pharmacy of Freie Universität Berlin. P. 10-13
- Smigielski EM, Sirotkin K, Ward M, Sherry ST (2000): dbSNP: a database of single nucleotide polymorphisms," *Nucleic Acids Research*, 28(1), 52-355. url: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC102496/pdf/gkd114.pdf>
- Tang H, Thomas PD (2016): PANTHER-PSEP: predicting disease-causing genetic variants using position-specific evolutionary preservation. *Bioinformatics*, 32(14), 2230-2. doi: 10.1093/bioinformatics/btw222
- Yanagawa T, Manglabruks A, Chang YB, Okamoto Y, Fisfalen ME, Curran PG, DeGroot LJ. Human histocompatibility leukocyte antigen-DQA1*0501 allele associated with genetic susceptibility to Graves' disease in a Caucasian population. *J ClinEndocrinolMetab* 1993; 76:1569–74.
- Zaman, M. (2008). Management of thyroid disorders in pregnancy *PJR* 18: 9-11.
- Zeitlin, A. (2010) The role of the HLA region in Hashimoto's Thyroiditis, ph.M. thesis , University of Birmingham.Pag 13-42

Table 1. Nucleotide substitutions and their reference NCBI sequence of the *HLA-DQA1* gene.

No	Nucleotide position in the PCR fragment (210bp)	Nucleotide change in the NCBI reference sequence (U03675.1)	Amino acid position	Amino acid change	Type of SNP
1	31	43A>T	49	T>S	Non-synonymous
2	88	100G>T	68	V>F	Non-synonymous
3	138	150A>G	84	-	Synonymous
4	150	162C>G	88	I>M	Non-synonymous
5	162	174A>T	92	K>N	Non-synonymous
6	183	195T>G	99	I>M	Non-synonymous



Table 2.The *In silico* analysis of the observed nonsynonymous SNPs using several bioinformatics tools.

Mutation	SIFT		PolyPhen-2		PANTHER		PROVEAN		SNAP2		SNPs&Go		
	score	prediction	score	prediction	score	prediction	score	prediction	score	prediction	score	RI	prediction
T49S	0.50	tolerated	0.001	benign	98	probably benign	-1.510	Neutral	60	effect	0.226	5	Neutral
V68F	0.00	affect protein function	0.732	Possibly damaging	324	possibly damaging	-4.430	Deleterious	92	effect	0.842	7	Disease
I88M	0.07	tolerated	0.007	benign	220	possibly damaging	-1.869	Neutral	-8	neutral	0.293	4	Neutral
K92N	0.14	tolerated	0.001	benign	324	possibly damaging	-3.471	Deleterious	31	effect	0.668	3	Disease
I99M	0.18	tolerated	0.000	benign	220	possibly damaging	-1.315	Neutral	-77	neutral	0.136	7	Neutral

Table 3.Investigation of mutant protein stability by I-Mutant 2.0.

Mutation	Reliability index (RI)	DDG value (kcal/mol)	Stability prediction
T49S	1	-2.34	decrease
V68F	9	-0.18	decrease
I88M	8	-0.60	decrease
K92N	4	-0.66	decrease
I99M	7	0.34	increase

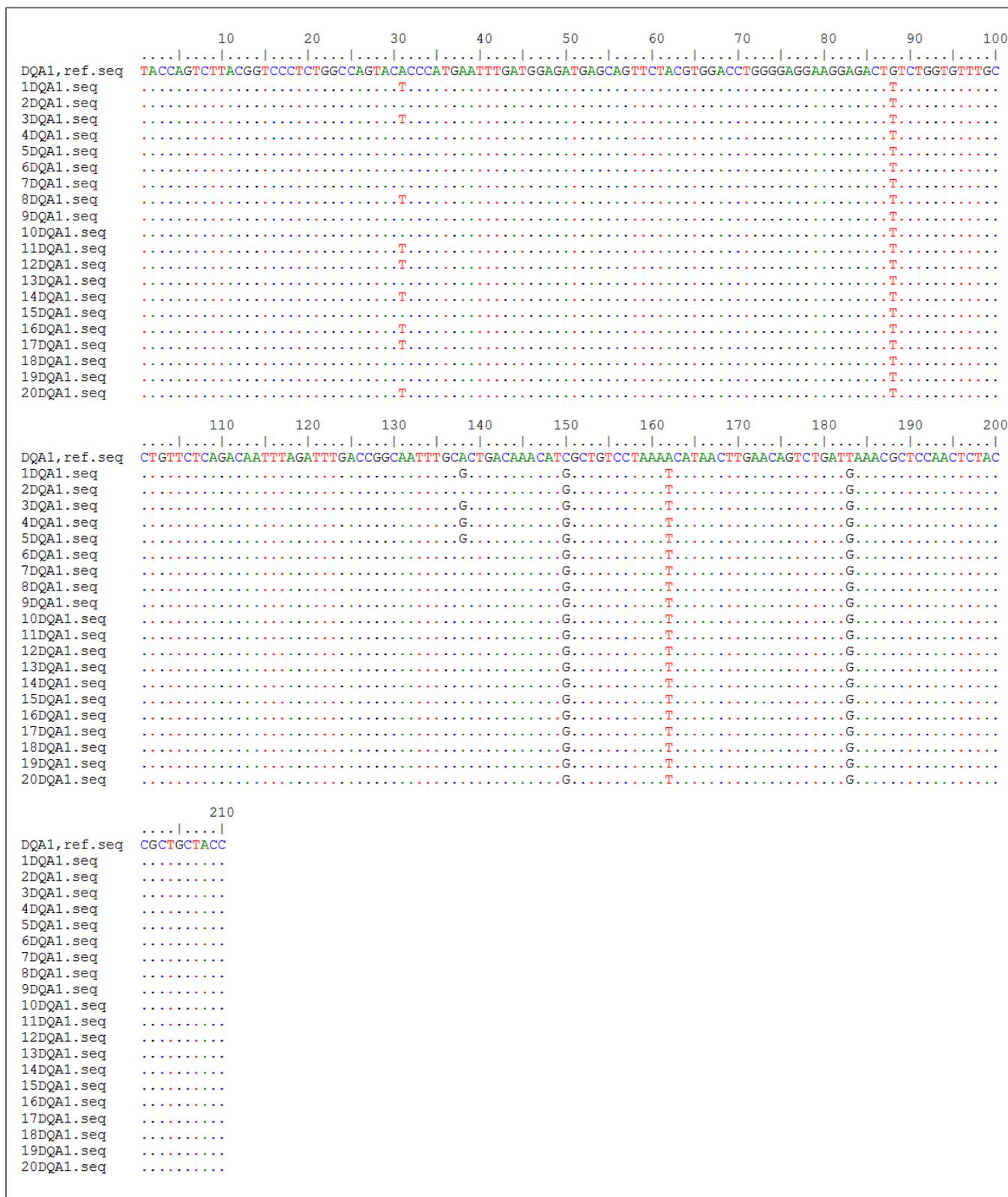


Fig. 1. multiple nucleic acid sequences alignment of DQAI reference sequences (DQA1.ref) compared with its 20 local variants starting from 1DQA1 to 20DQA1. 1, 3, and 5 DQA1 samples are healthy samples while the other samples are patient's samples.



CLUSTAL O(1.2.4) multiple sequence alignment

```

ref.seq. YQSYGPSGQYTHEFDGDEQFYVDLGRKETFWCLPVLQRFRFDRQFALTNIAVLKHNLSLIKRSNSTAAT 70
DQ1_1    YQSYGPSGQYSHEFDGDEQFYVDLGRKETFWCLPVLQRFRFDRQFALTNMAVLNHNLSIMKRSNSTAAT 70
DQ1_2    YQSYGPSGQYTHEFDGDEQFYVDLGRKETFWCLPVLQRFRFDRQFALTNMAVLNHNLSIMKRSNSTAAT 70
DQ1_3    YQSYGPSGQYSHEFDGDEQFYVDLGRKETFWCLPVLQRFRFDRQFALTNMAVLNHNLSIMKRSNSTAAT 70
DQ1_4    YQSYGPSGQYTHEFDGDEQFYVDLGRKETFWCLPVLQRFRFDRQFALTNMAVLNHNLSIMKRSNSTAAT 70
DQ1_5    YQSYGPSGQYTHEFDGDEQFYVDLGRKETFWCLPVLQRFRFDRQFALTNMAVLNHNLSIMKRSNSTAAT 70
DQ1_6    YQSYGPSGQYTHEFDGDEQFYVDLGRKETFWCLPVLQRFRFDRQFALTNMAVLNHNLSIMKRSNSTAAT 70
DQ1_7    YQSYGPSGQYTHEFDGDEQFYVDLGRKETFWCLPVLQRFRFDRQFALTNMAVLNHNLSIMKRSNSTAAT 70
DQ1_8    YQSYGPSGQYSHEFDGDEQFYVDLGRKETFWCLPVLQRFRFDRQFALTNMAVLNHNLSIMKRSNSTAAT 70
DQ1_9    YQSYGPSGQYTHEFDGDEQFYVDLGRKETFWCLPVLQRFRFDRQFALTNMAVLNHNLSIMKRSNSTAAT 70
DQ1_10   YQSYGPSGQYTHEFDGDEQFYVDLGRKETFWCLPVLQRFRFDRQFALTNMAVLNHNLSIMKRSNSTAAT 70
DQ1_11   YQSYGPSGQYSHEFDGDEQFYVDLGRKETFWCLPVLQRFRFDRQFALTNMAVLNHNLSIMKRSNSTAAT 70
DQ1_12   YQSYGPSGQYSHEFDGDEQFYVDLGRKETFWCLPVLQRFRFDRQFALTNMAVLNHNLSIMKRSNSTAAT 70
DQ1_13   YQSYGPSGQYTHEFDGDEQFYVDLGRKETFWCLPVLQRFRFDRQFALTNMAVLNHNLSIMKRSNSTAAT 70
DQ1_14   YQSYGPSGQYSHEFDGDEQFYVDLGRKETFWCLPVLQRFRFDRQFALTNMAVLNHNLSIMKRSNSTAAT 70
DQ1_15   YQSYGPSGQYTHEFDGDEQFYVDLGRKETFWCLPVLQRFRFDRQFALTNMAVLNHNLSIMKRSNSTAAT 70
DQ1_16   YQSYGPSGQYSHEFDGDEQFYVDLGRKETFWCLPVLQRFRFDRQFALTNMAVLNHNLSIMKRSNSTAAT 70
DQ1_17   YQSYGPSGQYSHEFDGDEQFYVDLGRKETFWCLPVLQRFRFDRQFALTNMAVLNHNLSIMKRSNSTAAT 70
DQ1_18   YQSYGPSGQYTHEFDGDEQFYVDLGRKETFWCLPVLQRFRFDRQFALTNMAVLNHNLSIMKRSNSTAAT 70
DQ1_19   YQSYGPSGQYTHEFDGDEQFYVDLGRKETFWCLPVLQRFRFDRQFALTNMAVLNHNLSIMKRSNSTAAT 70
DQ1_20   YQSYGPSGQYSHEFDGDEQFYVDLGRKETFWCLPVLQRFRFDRQFALTNMAVLNHNLSIMKRSNSTAAT 70
*****:*****:*****:***:*****:*****
    
```

Fig. 2. multiple amino acid sequences alignment of DQAI reference sequences (DQA1,ref) compared with its 20 local variants starting from 1DQA1 to 20DQA1. Each highlighted area represents a substituted amino acid.



The conservation scale:

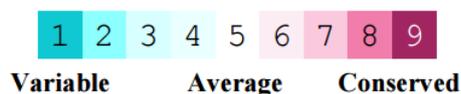


Fig. 3. Unique and conserver regions in MHC class II antigen were determined using Consurf tool. Amino acids were ranked on a conservation scale of 1–9 and are highlighted as follows: blue residues (1–4) are variable, white residues (5) are average, and purple residues (6–9) are conserved. (e) Residues exposed to the surface of the protein are indicated via an orange letter, (b) while residues predicted to be buried are indicated via a green letter. (s) Putative highly conserved and buried structural residues that are demonstrated with a blue letter. (f) Putative functional highly conserved and exposed residues are demonstrated with a red letter. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

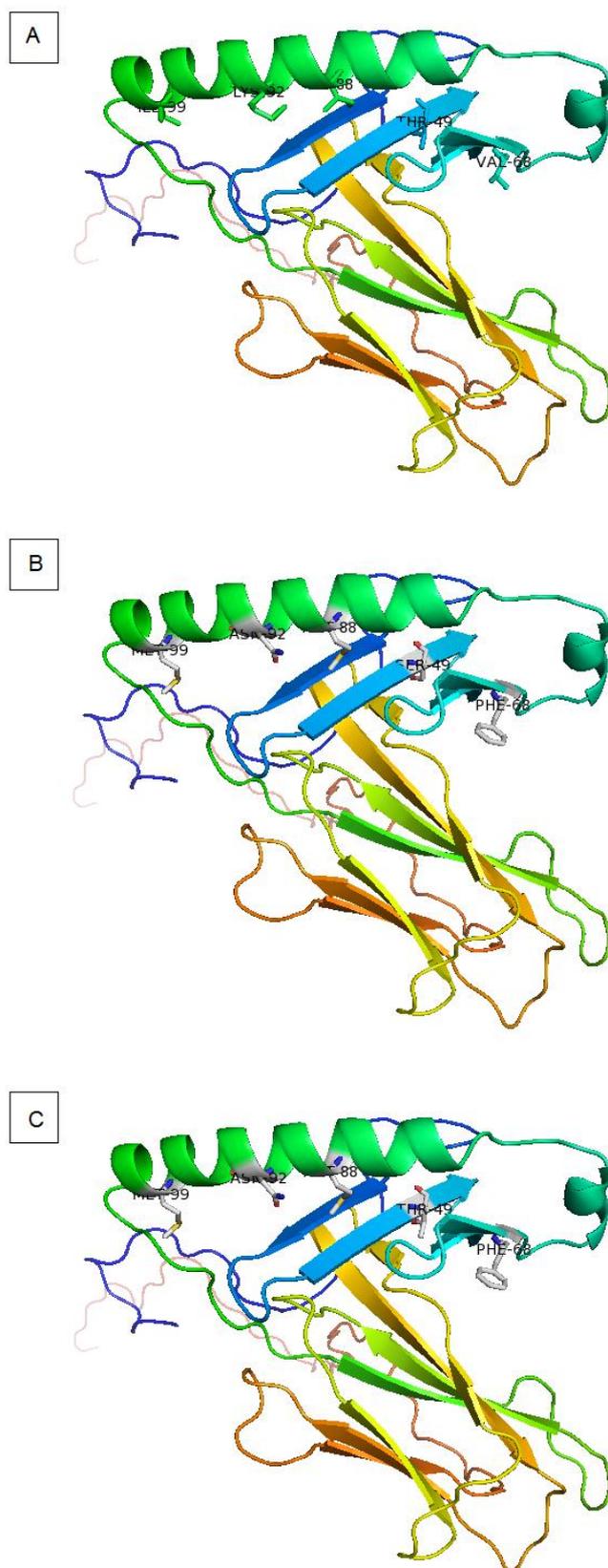
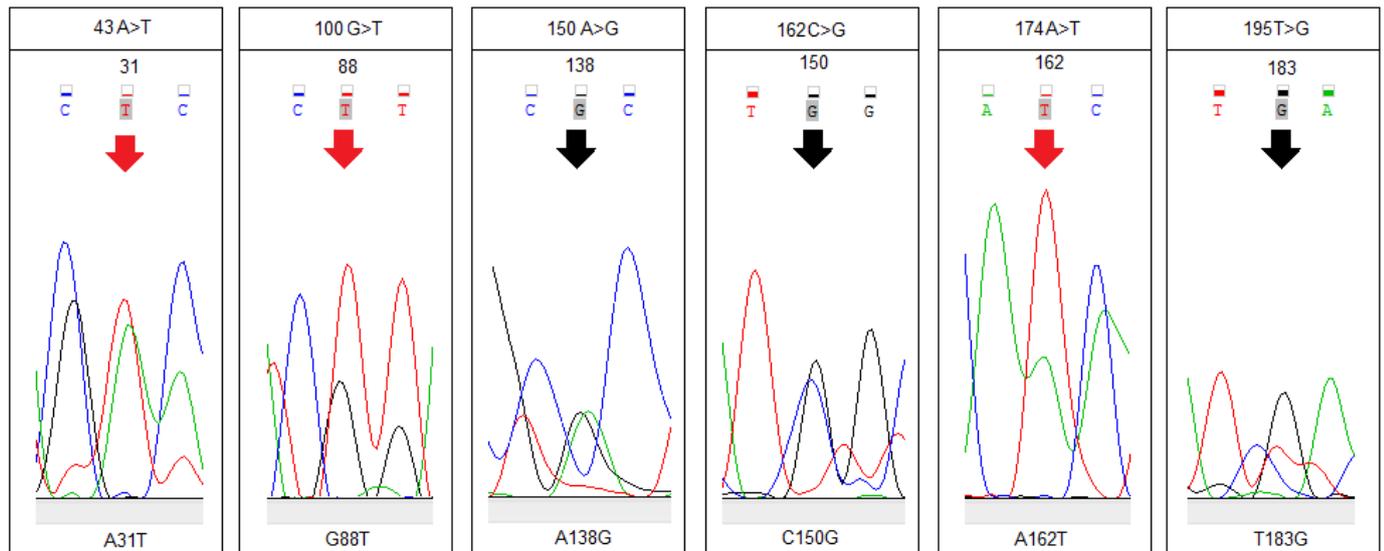


Fig. 4. The substitutions in amino acids and consequent MHC class II antigen structure. A) Native protein, in which, five amino acids were highlighted (T49, V68, I88, K92, I99). B) Substituted protein, in which five amino acid changes were taken place (T49S, V68F, I88M, K92N, I99M). C) Substituted protein, in which four amino acid changes were taken place (V68F, I88M, K92N, I99M).

URL: <http://www.uokufa.edu.iq/journals/index.php/ajb/index>

<http://iasj.net/iasj?func=issues&jid=129&uiLanguage=en>

Email: biomgzn.sci@uokufa.edu.iq



Supplementary Fig. 1. the pattern of the chromatogram of the observed SNPs in the DQA1 gene. The positions of the observed SNPs were highlighted in the bottom of the chromatogram while their positions in the reference genome were highlighted in the upper portion of the chromatogram.