International Journal of Genetic Engineering 2020, 8(1): 1-6
DOI: 10.5923/j.ijge.20200801.01

Computational Analysis of Single Nucleotide Polymorphism (SNPs) in Human *MYOC* Gene

Amged Mohammed Ibrahim, Afra M. Albakry, Nuha Widat Alla, Mona A. M. Khaeir, Hind. A. Elnasri*

Department of Molecular Biology and Bioinformatics, College of Veterinary Medicine, University of Bahri, Khartoum, Sudan

Abstract Glaucoma is a disease that damages the eye's optic nerve. It usually occurs when fluid builds up in the front part of the eye thus increasing the pressure within the eye and damaging the optic nerve. Among the causes of glaucoma is genetic polymorphisms of *MYOC* gene which can alter the myocilin protein and thus disrupting the regulation of the intraocular pressure which may lead to the disease. This study aimed to analyze nsSNPS in the Myocilin *(MYOC)* gene and the effect they may have on the protein function and structure. SNPs were obtained from the NCBI dbSNP database. The nsSNPs were further analyzed using 8 prediction tools namely GeneMANIA, SIFT, Polyphen-2, PROVEAN, SNPs & GO, PHD SNP, I-Mutant 3.0 and Project Hope. GeneMANIA results showed the association of MYOC gene with 20 other genes and mainly genes sharing the same protein domain. A total of 16 SNPs were predicted to be disease-associated using all software. Three SNPs were found to increase protein stability while 13 SNPs decreased the stability of the protein. In the current study, some SNPs that were previously reported to be associated with glaucoma were also found to be disease related using different software, while other new SNPs were predicted for the first time. In the future, these SNPs can clinically be tested to investigate their association with the disease.

Keywords In silico analysis, *MYOC* gene, Glaucoma, Bioinformatics

1. Introduction

Glaucoma is a complex, heterogeneous ocular disorder with multi factorial etiology characterized by structural damage to the optic nerve, and commonly associated with relatively high intraocular pressure (IOP) [1-2]. It is a leading cause of irreversible blindness worldwide with ~20% of cases occurring secondary to other ocular or systemic diseases [2-4].

Based on anatomical changes in the anterior chamber angle, primary glaucoma may be classified as primary angle closure glaucoma (PACG) or primary open-angle glaucoma (POAG), which may be further subdivided into juvenile open-angle glaucoma (JOAG) and adult onset POAG [1,5]. Glaucoma is a treatable disease if detected early; however, many patients are diagnosed during routine examinations or only following advanced field loss, as glaucoma is typically asymptomatic in the early stages. Therefore, the development of an accurate test for the detection of presymptomatic carriers at risk is important for the management of glaucoma.

* Corresponding author:
hindnasri2017@gmail.com (Hind. A. Elnasri)
Published online at http://journal.sapub.org/ijge

A family history of glaucoma is a well-known risk factor and hence genetic background is considered an important factor for the development of the disease [6-8].

Several genes have been reported to be associated with primary glaucoma including myocilin (MYOC), WD repeat domain, neurotrophin 1, cytochrome P450 family 1 subtype [9-10]. To date, mutations in these genes account for only ~5% of patients with POAG, and the influence of mutations in these genes on patients with PACG remain controversial [11-12].

The *MYOC* gene, is located on chromosome 1q24.3-q25.2. Mutations in the gene are commonly found in juvenile or early adult patients with high IOP although mutation frequencies vary between ethnic groups [13].

Bioinformatics is now playing a key role in different scientific areas. It involves computer sciences, mathematics, and statistics in order to analyze biological data that is being produced through the different sequencing techniques. Bio computing plays a key role in understanding the implication of genomic variations, especially single-nucleotide polymorphisms (SNPs), which represent the most frequent genetic variations in the human genome [14].

SNPs are the single base change in coding or non-coding DNA sequence and are present in every 200-300 bp in human genome [15]. The nonsynonymous SNPs (nsSNPs) are the single nucleotide variations that affect the coding region of the protein and modify the mutated site-encoded amino acid, which may lead to a structural modification of

the mutated protein, and may thus cause function alteration [15].

The aim of the present study was to perform a computational analysis of the nsSNPs in the MYOC gene to identify the possible pathogenic SNPs and the effect they may impose on protein structure and function.

2. Materials and Methods

SNPs in human MYOC gene data wereobtained from The National Center for Biotechnology Information (NCBI) dbSNP database during February 2020. The data obtained was further analyzed using various software.

1- GeneMANIA

GeneMANIA (http://www.genemania.org) is a web interface that helps predicting the function of genes and gene sets, can be used to find new gene members of a pathway or complex. MYOC gene name was entered as an input for GeneMANIA and the results were shown as a diagram showing the genetic interactions, pathways, co-expression, co-localization and protein domain similarity [16].

2- Functional and structural analysis of SNPs

SNPs retrieved from the dbSNP database were analyzed according to the scheme shown in Fig.1.

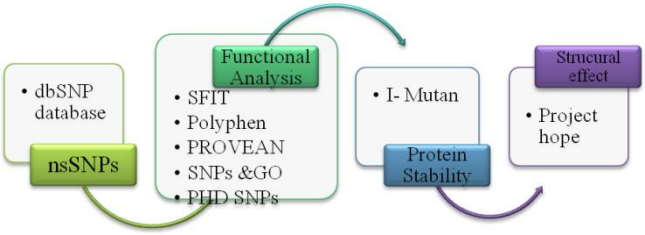


Figure 1. Flow chart for SNP analysis

nsSNPs were analyzed using 7 prediction tools: SIFT, Polyphen-2, PROVEAN, SNPs & GO, PHD -SNP, I-Mutant 3.0 and project hope.

a. Sorting intolerant from tolerant (SIFT)

(http://siftdna.org/www/SIFT_dbSNP.html). It predicts the tolerated and deleterious SNPs and identifies the impact of amino acid substitution on protein function and phenotype alterations. It generates alignments with a large number of homologous sequences, and assigns scores to each residue ranging from zero to one. The input was the rs of the nsSNPs (obtained from the db SNP database) and the results were obtained as either deleterious or tolerated based on the score of 0.05 or less [17].

b. PROVEAN (Protein Variation Effect Analyzer)

(http://provean.jcvi.org/seq_submit.php). It is a software tool which predicts whether an amino acid substitution or indel has an impact on the biological function of a protein. The input was the protein sequence in FASTA format (obtained from Uniprot / Expasy-database) and the amino acid substitution [18].

c. Polymorphism Phenotyping (PolyPhen-2)

(http://genetics.bwh.harvard.edu/pph2/). It is an online bioinformatics program that automatically predicts the consequence of an amino acid change on the structure and function of a protein based on a number of features such as sequence, phylogenetic and structural information. The program searches for 3D protein structures, multiple alignments of homologous sequences and amino acid contact information in several protein structure databases, then calculates position-specific independent count scores (PSIC) for each of the two variants, and then computes the PSIC scores difference between two variants. The higher a PSIC score difference, the higher the functional impact a particular amino acid substitution is likely to have. The nsSNPs that were predicted to be intolerant by SIFT were submitted to Polyphen-2 as protein sequence in FASTA. Then the position for wild type and mutated amino acids were submitted. Prediction outcomes could be classified as benign, possibly damaging or probably damaging, according to the posterior probability intervals (0, 0.2), (0.2, 0.85) and (0.85, 1) respectively [19].

d. SNPs & GO and PHD-SNP

Predicting disease associated variations using GO terms (http://snps.biofold.org/phd-snp/phd-snp.html). SNPs & Go predicts whether the new phenotype derived from a ns SNP is disease related or not (neutral) [20]. The protein sequence was submitted to the program after providing position of the wild and the new amino acid residue. PHD-SNP also shows the same result and it is shown within the same program.

e. Effect of SNPs on Protein Stability

I-Mutant version 3.0 (http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi) was used to predict the effect of the SNPs in the protein stability. I-Mutant basically can evaluate the stability change of a single site mutation starting from the protein structure or from the protein sequences [21]. The input was the protein sequence and position of wild and new amino acid **residue**. The output is classified into decrease or increase stability based on RI, and the DDG value.

3- Investigation of the structural effect

Project hope (HOPE; http://www. cmbi.ru.nl/hope/home) is an automatic mutant analysis server to study the insight structural features of native protein and the variant models. HOPE provides the 3D structural visualization of mutated proteins, and gives the results by using UniProt and DAS prediction servers. The input was the protein sequence and wild type and new amino acids. HOPE server predicts the output in the form of structural variation between mutant and wild type residues and the effect they have on protein structure and hence the function. [22].

3. Results

Fig.2 shows the co-expression. physical interaction,

shared protein domain between various gene and MYOC gene network. Eight genes OLFM1, OLFM2 OLFM3, OLFM4 OLFML1, OLFML2A OLFML2B, OLFML3 were having a shared protein domain. (Appendix 1). These genes are parlogs for MYOC and hence the shared domains.

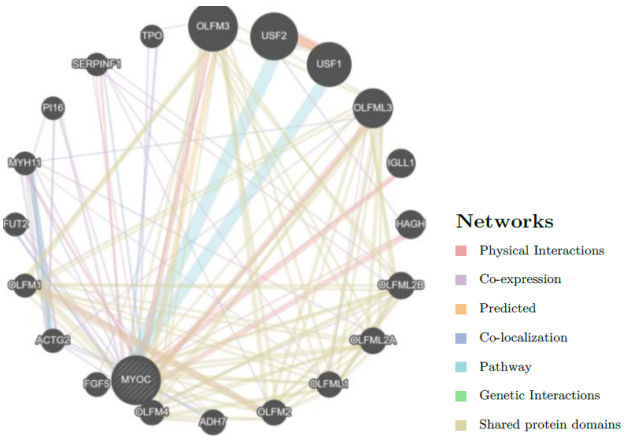


Figure 2. GeneMANIA result for MYOC gene

The total number of nsSNPs obtained from db SNPs was 109 SNPs. Using SIFT software 30 SNPs were found to be deleterious while 65 were found to be tolerated. Ten SNPs did not give any result using SIFT. These deleterious SNPs were further analyzed using Polyphen-2 which showed that 24 SNPs were probably damaging, 5 SNPs possibly damaging, while only one SNP were reported to be benign (Appendix 2). On the other hand PROVEAN results showed that 19 SNPs had a deleterious effect (Appendix 2). Only 17 SNPs were found to be disease related using both SNPs & GO and PHD- SNP (Appendix 2).

The total SNPs predicted to be disease related using the five different software were 16 SNPs as shown in Appendix 3.

Regarding the effect of the SNPs on protein stability, I-mutant results showed that 13 SNPs decreased the protein stability while 3 SNPs (rs74315330, rs74315331 and rs201573718) increased the protein activity

Project hope result for structural analysis:-

Four SNPs were investigated using project hope. They were selected based on polyphen 2 score (score =1) (Appendix 4).

4. Discussion

In this study, investigation of nsSNPs in the MYOC gene was done using different computational software. A total of 16 SNPs were reported to be damaging using five different soft wares.

A study carried among patients in Pakistan, showed the association of different SNPs in the MYOC gene and glaucoma, although with no statistical significance [23]. These SNPs (rs74315328, rs74315330, rs74315332, rs74315334, rs74315336, rs74315338 and rs121909193) were also confirmed in the current study to be disease related. Another SNP with rs74315341 has been reported among Caucasian and Brazilian population to be associated with glaucoma [24] but in this study this SNP was predicted to have a neutral effect using two software namely SNP and Go and PhD- SNP..

Two SNPs were also reported to be disease related in this study and were also detected among Australian population (rs74315330, and rs74315334 [25]. Another SNP rs rs74315329 was reported as an important risk factor among Tasmanian population [26], but it has not been predicted in the present study. Another two SNPs (rs74315328 and rs74315331) were reported to be associated with hereditary glaucoma in the United states [27]. and were also confirmed in this study.

Genetic defects can lead to an altered protein product which can be secreted into the extracellular matrix of the trabecular meshwork causing a severe form of autosomal dominant JOAG associated with very high IOP [9]. The effect of SNPs on protein structure can have different impacts such as increasing or decreasing its activity (as predicted by I Mutant) and hence affect the folding of the protein in the correct manner or affecting binding of the protein with specific types of ions or ligands as predicted by project hope- and can hence affect the function of the protein. Previous studies reported that mutated myocilin become tangled in the cell in its altered form [28].

5. Conclusions

The current study showed that 16 nsSNPs are associated with glaucoma using various computational tools. These mutations can distort the protein stability or it's binding with other ligands and thus affecting its function.

Appendix 1: GeneMANIA Results of MYOC Gene

Gene Name	Explanation	Co-expression	Co-localization	Physical interaction	Pathway	Shared protein domains
ACTG2	actin, gamma 2, smooth muscle, enteric	**Yes**	**Yes**			
ADH7	alcohol dehydrogenase 7 (class IV), mu or sigma polypeptide	**Yes**	**Yes**			
FGF5	fibroblast growth factor 5	**Yes**				
FUT2	fucosyltransferase 2	**Yes**	**Yes**			
MYH11	myosin, heavy chain 11, smooth	**Yes**				

Gene Name	Explanation	Co-expression	Co-localization	Physical interaction	Pathway	Shared protein domains
	muscle					
OLFM1	olfactomedin 1					**Yes**
OLFM2	olfactomedin 2					**Yes**
OLFM3	olfactomedin 3			**Yes**		**Yes**
OLFM4	olfactomedin 4					**Yes**
OLFML1	olfactomedin like 1					**Yes**
OLFML2A	olfactomedin like 2A					**Yes**
OLFML2B	olfactomedin like 2B					**Yes**
OLFML3	olfactomedin like 3			**Yes**		**Yes**
PI16	peptidase inhibitor 16	**Yes**				
SERPINF1	serpin family F member 1	**Yes**	**Yes**			
TPO	thyroid peroxidase	**Yes**	**Yes**			
USF1	upstream transcription factor 1				**Yes**	
USF2	upstream transcription 2				**Yes**	

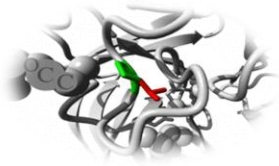
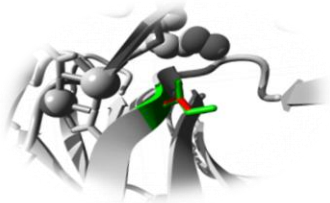
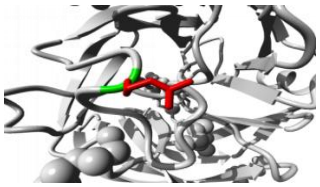
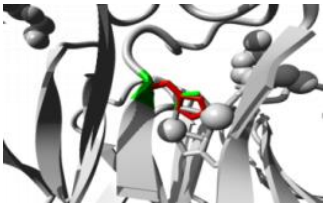
Appendix 2: Result of SNP Analysis Using Various Software

No	SNP	Amino Acid Change	SIFT Prediction	Polyphen 2 Prediction	PROVEAN prediction	SNPs & GO prediction	PHD-SNP Prediction
1.	rs28936694	G399V	Deleterious	probably damaging	Deleterious	Disease	Disease
2.	rs74315328	Y437H	Deleterious	probably damaging	Deleterious	Disease	Disease
3.	rs74315330	P370L	Deleterious	probably damaging	Deleterious	Disease	Disease
4.	rs74315331	I477S	Deleterious	probably damaging	Deleterious	Disease	Disease
5.	rs74315331	I477N	Deleterious	probably damaging	Deleterious	Disease	Disease
6.	rs74315332	N480K	Deleterious	probably damaging	Deleterious	Disease	Disease
7.	rs74315334	G367R	Deleterious	probably damaging	Deleterious	Disease	Disease
8.	rs74315335	Q337R	Deleterious	possibly damaging	Neutral	Disease	Disease
9.	rs74315336	K423E	Deleterious	probably damaging	Deleterious	Disease	Disease
10.	rs74315338	C433R	Deleterious	possibly damaging	Deleterious	Disease	Disease
11.	rs74315339	Q48H	Deleterious	Benign	Neutral	Neutral	Neutral
12.	rs74315340	C245Y	Deleterious	probably damaging	Deleterious	Disease	Disease
13.	rs74315341	G252R	Deleterious	probably damaging	Neutral	Neutral	Neutral
14.	rs121909193	G364V	Deleterious	probably damaging	Deleterious	Disease	Disease
15.	rs121909194	D380H	Deleterious	probably damaging	Deleterious	Disease	Disease
16.	rs142680834	D446Y	Deleterious	possibly damaging	Deleterious	Neutral	Disease
17.	rs143474164	D294A	Deleterious	probably damaging	Deleterious	Neutral	Disease
18.	rs145934417	R296H	Deleterious	probably damaging	Deleterious	Disease	Disease
19.	rs146391864	V329M	Deleterious	probably damaging	Deleterious	Disease	Disease
20.	rs147122394	T325M	Deleterious	probably damaging	Deleterious	Disease	Disease
21.	rs150438494	N420Y	Deleterious	probably damaging	Deleterious	Neutral	Neutral
22.	rs150724391	G31W	Deleterious	probably damaging	Neutral	Neutral	Disease
23.	rs199705804	L152P	Deleterious	probably damaging	Neutral	Neutral	Disease
24.	rs200120115	R126W	Deleterious	probably damaging	Neutral	Neutral	Neutral
25.	rs200208925	V53A	Deleterious	probably damaging	Neutral	Neutral	Disease
26.	rs200968862	P223S	Deleterious	possibly damaging	Neutral	Neutral	Neutral
27.	rs200971340	R128W	Deleterious	probably damaging	Neutral	Neutral	Disease
28.	rs201206951	L334P	Deleterious	possibly damaging	Neutral	Neutral	Disease
29.	rs201573718	R422H	Deleterious	probably damaging	Deleterious	Disease	Disease
30.	rs376735175	K484Q	Deleterious	probably damaging	Neutral	Neutral	Neutral

Appendix 3: Total Number of SNPs Predicted to be Disease Related Using Different Software

No	SNP ID	Amino acid change
1.	rs28936694	G399V
2.	rs74315328	Y437H
3.	rs74315330	P370L
4.	rs74315331	I477S
5.	rs74315331	I477N
6.	rs74315332	N480K
7.	rs74315334	G367R
8.	rs74315336	K423E
9.	rs74315338	C433R
10.	rs74315340	C245Y
11.	rs121909194	D380H
12.	rs145934417	R296H
13.	rs146391864	V329M
14.	rs147122394	T325M
15.	rs201573718	R422H
16.	rs121909193	G364V

Appendix 4: Project Hope Results

SNP ID (Amino acid change and Position)	3D Structure	Effect of amino acid change	Effect in protein Structure
rs74315331 P370L		The mutant residue (Leucine) is bigger than the wild-type residue (Proline)	Loss of interactions with the ligand.
rs74315332 I477S		The mutant residue (Serine) is smaller than the wild-type (Isoleucine) residue	The differences in size between the wild and mutant residue disturb the interaction with the metal-ion: The mutated residue is not in direct contact with a ligand
rs74315336 G367R		The mutant residue (Arginine) is bigger than the wild-type (Glycine) residue.	The difference in properties between wild-type and mutation can easily cause loss of interactions with the ligand
rs137853277 D380H		The mutant residue (Aspartic) is bigger than the wild-type (Histidine) residue	The size difference between wild-type and mutant residue makes that the new residue is not in the correct position to make the same hydrogen bond as the original wild-type residue.

- color key: grey color = the protein, -green color =the wild-residue, red color =the mutant-type

REFERENCES

[1] Foster PJ, Buhrmann R, Quigley HA and Johnson GJ. The definition and classification of glaucoma in prevalence surveys. Br J Ophthalmol 86: 238-242, 2002.

[2] Pakravan M, Yazdani S, Javadi MA, Amini H, Behroozi Z, Ziaei H, Katibeh M, Solaimanizad R, Ghahari E and Yaseri M. A opulation-based survey of the prevalence and types of glaucoma in central Iran: The Yazd eye study. Ophthalmology 120: 1977-1984, 2013.

[3] Casson RJ, Chidlow G, Wood JP, Crowston JG and Goldberg I. Definition of glaucoma: Clinical and experimental concepts. Clin Experiment Ophthalmol 40: 341-349, 2012.

[4] Weinreb RN, Aung T and Medeiros FA. The pathophysiology and treatment of glaucoma: A review. JAMA 311: 1901-1911, 2014.

[5] Goldwyn R, Waltman SR and Becker B: Primary open-angle glaucoma in adolescents and young adults. Arch Ophthalmol 84: 579-582, 1970.

[6] Booth A, Churchill A, Anwar R, Menage M and Markham A. The genetics of primary open angle glaucoma. Br J Ophthalmol 81: 409-414, 1997.

[7] Wiggs JL. Genetic etiologies of glaucoma. Arch Ophthalmol 125: 30-37, 2007.

[8] Khan AO: Genetics of primary glaucoma. Curr Opin Ophthalmol 22: 347-355, 2011.

[9] Janssen SF, Gorgels TG, Ramdas WD, Klaver CC, van Duijn CM, Jansonius NM, et al. The vast complexity of primary open angle glaucoma: disease genes, risks, molecular mechanism and pathobiology. Prog Retin Eye Res. Nov; 37: 31-67. 2013.

[10] Takamoto M, Araie M. Genetics of primary open angle glaucoma. Jpn J Ophthalmol; 58(1)1-15, 2014.

[11] Fingert J: Primary open-angle glaucoma genes. Eye (Lond) 25: 587-595, 2011.

[12] Shastry BS: Genetic susceptibility to primary angle closure glaucoma (PACG). Discov Med 15: 17-22, 2013.

[13] Kwon YH, Fingert JH, Kuehn MH, Alward WL. Primary open- angle glaucoma. N Engl JMed. 360(11): 1113- 24, 2009.

[14] Hind B, Sara B, AL Mehdi K, Imane M, Hicham C, Hassan R, Rachid S, Fouad B, and Abdelhamid B. (2019). Prediction and Structural Comparison of Deleterious Coding Nonsynonymous Single Nucleotide Polymorphisms (nsSNPs) in Human LEP Gene Associated with Obesity. Bio Med Research International. https://doi.org/10.1155/2019/1832084.

[15] Lee JE, Choi JH, Lee JH, Lee MG. Gene SNPs and mutations in clinical genetic testing: haplotype-based testing and analysis. Mutat Res. 2005; 573: 195-204.

[16] Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. J. Nucleic Acids Res. 2010.

[17] Ng PC Heinkoff S. SIFT predcitiong amino acid changes that effect protein function. Nucleic Acic Res, 31: 3812-3814. 2003.

[18] Yongwook C. and Agnes P. Chan.. Sequence analysis PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics 31, 2745–2747. 2005.

[19] Adzhubei, Ivan, Daniel M. Jordan, and Shamil R. Sunyaev. "Predicting functional effect of human missense mutations using PolyPhen‐2." Current protocols in human genetics 76: 7-20. 2013.

[20] Cabrese R, Capriotti E, Fariselli P, Mareellis PL, Cassdio R. Functional annotations improve the predictive score of human diseaes related mutaions in proteins. Human Mutation. 30: 1237-1244. 2009.

[21] Capriotti E, Fariselli P, Casadio R I Mutant 2.0 Prediction of protein stability changes upon mutations from the protein sequence or structure. Nucleic Acids Res. 33: W306-W310. 2005.

[22] Hanka V., Tim AH te Beek, Remko KP Kuipers, Maarten L. Hekkelman, Gert V. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. BMC Bioinformatics, vol. 11, no. 1, p. 548. 2010.

[23] Nazir S, Mukhtar M, Shahnawaz M, Farooqi S, Fatima N, Mehmood R, et al. (2018) A novel single nucleotide polymorphism in exon 3 of MYOC enhances the risk of glaucoma. PLoS ONE 13(4): e0195157.

[24] Povoa CA, Malta RF, Rezende MM, de Melo KF, Giannella-Neto D. Correlation between genotype and phenotype in primary open angle glaucoma of Brazilian families with mutations in exon 3 of the TIGR/ MYOC gene. Arq Bras Oftalmol. 69(3): 289-97. 2006.

[25] Souzeau E, Burdon KP, Dubowsky A, grist S, Usher B, Fitzegerald JT, et al. Higher prevalence of myocilin mutations in advanced glaucoma in comparison with less advanced disease in an Australian disease registry. Ophthalmology. 120(6) 1135-43. 2013.

[26] Hewitt AW, Bennett SL, Richards JE, Dimasi DP, Booth AP, Inglehearn C, et al. Myocilin Gly252Arg mutation and glaucoma of intermediate severity in Caucasian individuals. Arch Ophthalmol. Jan; 125(1): 98- 104. 2007.

[27] Craig JE, Baird PN, Healey DL, McNaught AI, McCartney PJ, Rait JL, et al. Evidence for genetic heterogeneity within eight glaucoma families, with the GLC1A Gln368STOP mutation being an important phenotypic modifier. Ophthalmology Sep; 108(9): 1607-20. 2001.

[28] Liu Y, Vollrath D. Reversal of mutant myocilin non-secretion and cell killing: implications for glaucoma. Hum Mol Genet Jun; 13(11): 1193- 204. 2004.