Journal of Agricultural, Environmental and Veterinary Sciences
Volume (3), Issue (4): 30 Dec 2019
P: 66 - 77

AJSRP
ISSN: 2522- 3364

مجلة العلوم الزراعية والبيئية والبيطرية
المجلد (3)، العدد (4): 30 ديسمبر 2019 م
ص: 66 - 77

In silico analysis of single nucleotide polymorphism (SNPs) in human [BMPR2] gene

Mohamed Saeed Elhassan Mohamed
Hind AbdelAziz Elnasri
Afraa Mohamed Elbakri
Mona Abdelrahman Mohamed. Khaeir
University of Bahri || College of Veterinary Medicine || Sudan

ABSTRACT: Background: (BMPR2) gene is encoded gene and cause pulmonary arterial hypertension. Also, it has major role in regulating the growth the maturation of cells. A (BMPR2) gene contains only 25 SNPs as deleterious SNPs and was analyzed in this study.

Material and methods: 25 SNPs investigated using the NCBI database (htt: / / www.ncbi.nlm.nih.gov/)and the SNPs were analyzed using six prediction tools: SIFT, Polyphen- 2, I- Mutant, PROVEAN, PhD- SNP and Project Hope. 76% SNPs predict Probably Damaging by POLYPHEN software. the protein stability checked by I- MUTANT and 72% SNPs trend to decrease effected. when used SNPs & GO 52% SNPs were diseased.64% SNPs were deleterious by PROVEAN. There are 20 associated genes, 14 genes share the same protein domains and 13 genes similar in their expression when predicted by GENE MANIA software. Using PROJEC HOPE software to predict the structural effect in function.

Result: eight SNPs of 25 SNPs were sharing the same and significant results, so that leads to confirm this result.

Conclusions: eight SNPs, rs137852744, rs137852745, rs137852746, rs137852749, rs137852750, rs137852750, rs143740797and rs374694591 were shown to have highly damaging and cause the pulmonary arterial hypertension disease.

Keywords: SIFT, Polyphen2, BMPR2 gene, pulmonary hypertension, in silico. SNPs.

دراسة وتحليل الجين (BMPR2) عن طريق (SNPs) باستخدام برمجيات الحاسب الآلي

محمد سعيد الحسن محمد
هند عبد العزيز النصري
عفراء محمد البكري
منى عبد الرحمن محمد خير
كلية الطب البيطري || جامعة بحري || السودان

الملخص: الجين المعني بالدراسة والمعروف اختصارًا (BMPR2) هو الجين المسئول عن توازن ضغط الدم في الشريان الرئوي في جسم الإنسان، كما أن له دور كبير في تنظيم نمو الخلايا. ولما يكون هنالك عطب في هذا الجين لا يتمكن جسم الإنسان من أداء دوره بصورة طبيعية، فيصاب عندها بضغط الدم وتحديداً في الشريان الرئوي كما يتأثر نمو الخلايا سلباً.

في هذه الدراسة استخدمنا برمجيات حاسوبية عالية الدقة في تحديد الجزء من الجين والذي تحدث فيه هذه الإشكالات وقمنا بتحديدها. وفي الدراسات السابقة كان هذا الجين هو المتهم أيضاً بهذه المشكلة. لكن لم يتم تحديد الجزء داخل الجين المسؤول عن هذه المشكلة، بالرغم من ظهور بعض هذه الأجزاء في دراسات سابقة.

من مميزات هذه الدراسة استخدام عدد من الخوازميات الحديثة، توفر الجهد والوقت والمال للعمل فقط على جزء معين من الجين وليس كله.

الكلمات المفتاحية: الخوازمية سيفت- الخوارزمية بوليفين- الخوارزمية ايميوتين- الخوازمية سنيب اند قو- الخوازمية جين مينيا- استخدام الخوازميات الحاسوبية- ضغط الدم في الشريان الرئوي.

1. INTRODUCTION

BMPR2 gene (Bone Morphogenetic Protein Receptor Type 2) encodes for a transmembrane serine/ threonine kinase receptor belonging to the transforming growth factor beta (TGF- β) super family (Herpin et al, 2014)[1]. It is specifically recognized by bone morphogenetic proteins (*BMPs*), which are involved in several signaling pathways that regulate cellular differentiation, proliferation and apoptosis *BMPR2* gene is located on chromosome 2q33i, with a total length of about 190 kb. It consisting of 13 exons, 12 introns, and encodes a protein of 1038 amino acids. (Hyun et al, 2005)[2].

Bone morphogenetic protein receptor type 2 spans the cell membrane, so that one end of the protein is on the outer surface of the cell and the other end remains inside the cell.

Mutations in *BMPR2* gene is associated with Pulmonary arterial hypertension (PAH; OMIM #178600, ORPHA 422) (Guillermo et al, 2017)[3]. PAH is a rare, progressive disease that causes the obstruction of precapillary pulmonary arteries. Loss of function or reduction in *BMPR2* expression may be sufficient to develop PAH. It is known that mutations in the genomic code can lead to diseases because the protein can get mislocalized and therefore, cannot carry out its activity properly (Marice, 2010)[4].

Mutational screening of *BMPR2* in PAH patients has been extensively reported. Until recently; more than 450 mutations have been identified in BMPR2 gene (Hironori et al, 2017)[5]. Mutations in this gene have been identified in more than 80% of patients with hereditary pulmonary arterial hypertension (HPAH), although only 20% of the carriers eventually develop the disease (Joshua P et al, 2011)[6]. On the other hand, the frequency of *BMPR2* mutations in idiopathic (IPAH), IPAH patients is much lower, (Erika B.Rosenzweig et al 2008)[7].

The use of in silico tools is becoming nowadays an important technique in studying disease related genes. The recent advances in sequencing techniques are providing big data about the human genomes. The identification of disease- related SNPs derived from large- scale techniques has the potential to create personalized tools for the diagnosis, prognosis and treatment of diseases.

OBJECTIVE

The objective of this study was to investigate the ns SNPs in *BMPR2* gene and their effect in protein structure, function and stability. This will be carried using various bioinformatics' tools.

2. MATERIALS AND METHODS

The first steps for this study was to obtain the non synounomus SNPs (SNPs) in the coding region). They were obtained from NCBI- db SNPs database.(http: / / www.ncbi.nlm.nih.gov/ snp during the year 2018). Protein sequence was obtained from EXPASY and Uniprot database.

2.1 Gene MANIA software

(http: // www.genemania.org) is a web interface that predicts the gene function and finds other genes related to gene of interest. It can also predict the interactions, Pathways, Co- expression, Co-localization between these genes. (Warde- *et al*, 2010)[8].

2.2 SIFT software

(Sorting intolerant from tolerant) SIFT (htt: / / siftdna.org/ www/ SIFT_dbsnp.html). It is an online tool that predicts the tolerated and deleterious SNPs and also the affect of amino acid substitution on protein function. It performs analysis based on different algorithms and it interprets the homologous sequences using the Swiss- Prot (version 51.3) and TrEMBL (version 34.3) (Hu et al, 2012)[9]. It assigns score to each residue ranging from zero to one. The threshold intolerance score for SNPs is 0.05 or less.

2.3 Polyphen- 2 software

(Polymorphism Phenotyping) (http://www.genetics.bwh.harvard.edu/ pph2/).It is used to predict the consequence of an amino acid change on the structure and function of a protein.It searches for 3D protein structures, multiple alignments of homologous sequences.

The input is the protein sequence that was obtained from EXPASy. It estimates the position-specific independent count score (PSIC) for every variant and then determines the difference between them, the higher the PSIC , the higher the functional impact of the amino acid on the protein function may be. (Prediction outcomes could be classified as probably damaging, possibly damaging or benign according to the score ranging from (0–1). (Adzhubei *et al*, 2013)[10].

2.4 PROVEAN software

(Protein Variation Effect Analyzer)(http: / / provean.jcvi.org).Is a software which predicts whether an amino acid substitution has an impact on the biological function of a protein. The input is protein sequence and position of amino acid position. Prediction outcomes could be classified as tolerated or deleterious.

2.5 SNPs & GO, PHD- SNP software

(Single nucleotide polymorphism & Gene Ontology),(http: / / snps.biofold.or g/ phd- snp/ phd-snp.html). It is an accurate method that predicts whether a variation is disease related or not The protein

sequences is submitted in FASTA format (obtained from UniproktB/ Expasy) after submitting the sequence the mutations were submitted in the XPOSY format where X and Y are the wild- type and mutant residues respectively. A score is assigned between zero and one , the probability score higher than 0.5 reveals the disease related effect of mutation PHD- SNP results are presented as part of SNPs& GO output.

2.6 I- MUTANT software

http: / / gpcr2.biocomp.unibo.it/ cgi/ pre dictors/ I- Mutant 3.0/ I- Mutant3.0.cgi).This software was used to check the protein stability (starting from the protein structure or protein sequences) as a result of a mutation (Capriotti et al., 2005)[11]. The output is obtained in the form of protein stability change upon mutation and Gibbs- free energy change (DDG) either increased or decreased stability.

2.7 Project HOPE software

(http: // www.cmbi.ru.nl/ hope/ home) (Hanka et al, 2010)[12]. It is an on line software used to analyze the structural effects of intended mutation, provides the 3D structural visualization of mutated proteins, Gives the results by using Uniprot server. The input is the protein sequence and positions of amino acid substitution. HOPE server predicts the output in the form of structural variation between mutant and wild type. In addition, HOPE builds a report with text, figures, and animations.

3. RESULTS AND DISCUSSON:

3.1 genes – genes interaction

BMPR2 gene was found to be associated with 20 other genes, 14 genes were shared protein domains with *BMPR2* gene (figure 1) and 13 genes are co- expression with *BMPR2* gene (figure 1) predicted by gene MANIA software. The most related gene was *BMPR1A* gene, the mutation in this gene cause juvenile polyposis syndrome. (Warde et al, 2010) [13.]

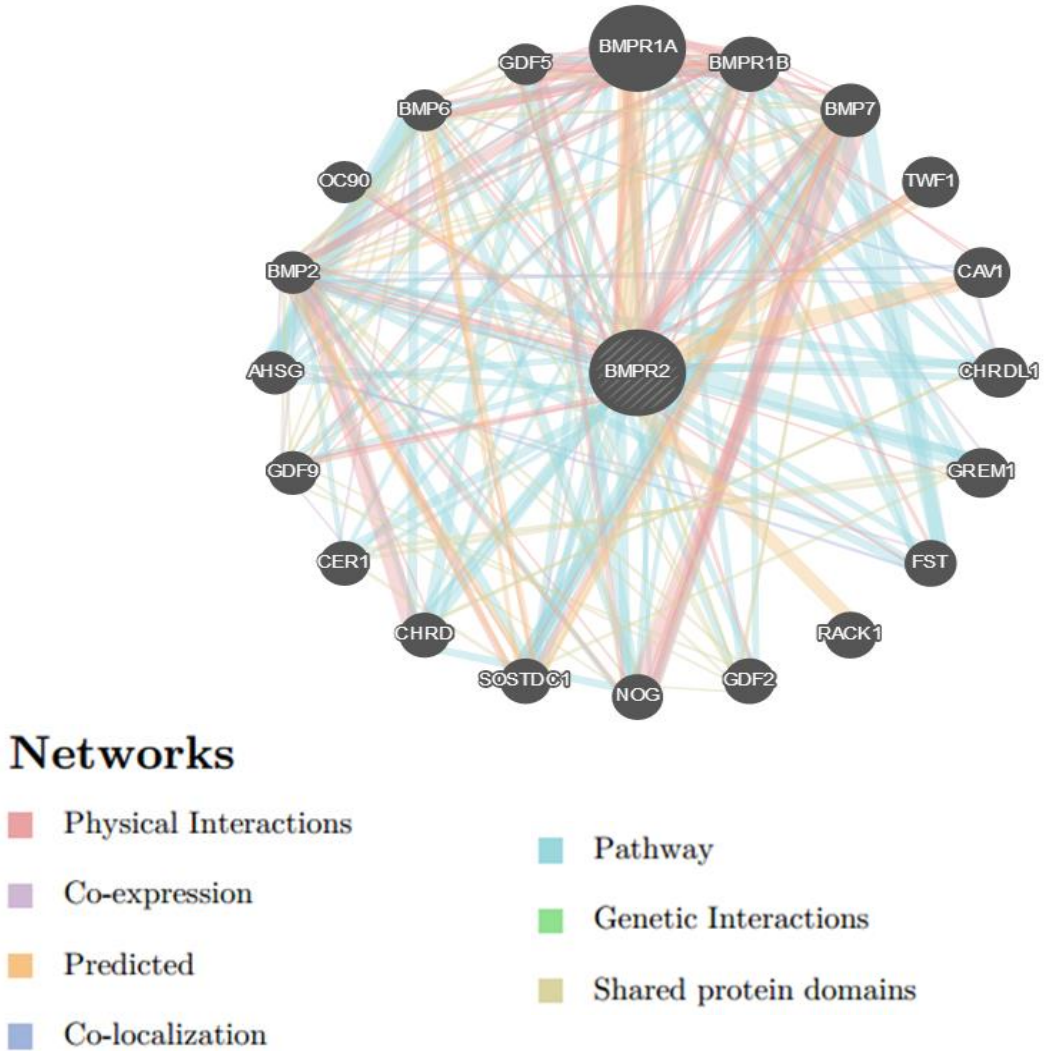

Figure 1: the interaction between *BMPR2* and associated genes.

Table (1) nonsynonymous SNPs prediction with SIFT and polyphen.

DESCRIPTION					SIFT RESULTS		POLYPHEN RESULTS	
SNPs ID	A.A CHANGE	PROTEIN ID	REF ALLELE	ALT ALLELE	PREDICTION	SCORE	PREDICTION	SCORE
rs1048127	G828R	ENSP00000363708	G	A	Deleterious	0.015	Possible Damaging	0.868
rs2229778	A494V	ENSP00000363702	C	T	Deleterious	0.004	Probably Damaging	1.000
rs137852743	C118W	ENSP00000363708	T	G	Deleterious	0.000	Probably Damaging	1.000
rs137852744	C347Y	ENSP00000363708	G	A	Deleterious	0.001	Probably Damaging	1.000

DESCRIPTION					SIFT RESULTS		POLYPHEN RESULTS	
rs137852745	D485G	ENSP00000363708	A	G	Deleterious	0.000	Probably Damaging	1.000
rs137852746	R491W	ENSP00000363702	C	T	Deleterious	0.000	Probably Damaging	1.000
rs137852749	R491Q	ENSP00000363702	G	A	Deleterious	0.000	Probably Damaging	1.000
rs137852750	C123S	ENSP00000363702	T	A	Deleterious	0.000	Probably Damaging	1.000
rs137852750	C123R	ENSP00000363702	T	C	Deleterious	0.000	Probably Damaging	1.000
rs143740797	I246N	ENSP00000363702	T	A	Deleterious	0.005	Probably Damaging	1.000
rs144356403	D545G	ENSP00000363708	A	G	Deleterious	0.032	Probably Damaging	0.988
rs147943794	T621M	ENSP00000363708	C	T	Deleterious	0.013	Probably Damaging	1.000
rs148257675	Y743C	ENSP00000363708	A	G	Deleterious	0.012	Probably Damaging	0.985
rs148682262	R225H	ENSP00000363708	G	A	Deleterious	0.02	Probably Damaging	0.945
rs148770894	R886C	ENSP00000363708	C	T	Deleterious	0.020	Probably Damaging	1.000
rs199915496	Y589C	ENSP00000363708	A	G	Deleterious	0.027	Probably Damaging	0.999
rs199954814	A305V	ENSP00000363708	C	T	Deleterious	0.020	Possible Damaging	0.930
rs200339485	R237C	ENSP00000363702	C	T	Deleterious	0.002	Possible Damaging	0.930
rs201781338	R873Q	ENSP00000363708	G	A	Deleterious	0.020	Possible Damaging	0.952
Continue table (1): nonsynonymous SNPs prediction with SIFT and polyphen.								
rs201938348	S818F	ENSP00000363708	C	T	Deleterious	0.015	Probably Damaging	0.966
rs368027047	I257T	ENSP00000363702	T	C	Deleterious	0.004	Benign	0.412
rs369545039	T295P	ENSP00000363708	A	C	Deleterious	0.018	Probably Damaging	0.998
rs374694591	R266T	ENSP00000363	G	C	Deleterious	0.010	Probably	0.993

DESCRIPTION					SIFT RESULTS		POLYPHEN RESULTS	
		08					Damaging	
rs375492148	I742S	ENSP00000363708	T	G	Deleterious	0.007	Probably Damaging	0.993
rs377763312	P864L	ENSP00000363708	C	T	Deleterious	0.009	Benign	0.090

3.2 SIFT and POLYPHON software

The total number of SNPs for this gene were (322) SNPs, SIFT software showed that, 145 SNPs were in 3UTR side, nine SNPs in 5UTR side, five SNPs lincRNA, 120 SNPs were tolerated, five SNPs were deleterious but with low confidence and 38 SNPs were non- synonymous polymorphism and found to be deleterious. But the unrepeated SNPs were 25 SNPs and had been analyzed in this study.

Among polyphen software result indicate that four SNPs were possibly damaging and one SNP was benign, 20 SNPs were Probably Damaging with high score (1.000- 0.993) so these SNPs were significant using polyphen software.

3.3 Predictions by confirming software

When used SNPs & GO software 13 of 25 SNPs were diseased with (5- 7) reliability index and high degree of probability (0.821- 0.774). while 21 of 25 SNPs were diseased by using PhD software. The stability mutational effect on protein prediction checked by I- MUTANT software showed that 19 SNPs were decrease with negative DDG value, that leaded to confirm the instability of these SNPs. 17 SNPs were deleterious predicted by PROVEAN software.

Table (2) shows SNPs & GO, PHD, I- MUTANT and PROVEN results.

SNPs ID	SNPs & GO			PhD			I- MUTANT			PROVEN
	Prediction	RI	probability	Prediction	RI	probability	Prediction	RI	DDG	Prediction
rs1048127	Disease	1	0.569	Disease	5	0.743	Decrease	9	-1.29	Neutral
rs2229778	Neutral	3	0.341	Disease	1	0.528	Increase	3	0.08	Deleterious
rs137852743	Disease	7	0.851	Disease	7	0.875	Increase	1	0.21	Deleterious
rs137852744	Disease	7	0.854	Disease	6	0.812	Decrease	5	-0.59	Deleterious
rs137852745	Disease	7	0.840	Disease	2	0.621	Decrease	9	-2.11	Deleterious
rs137852746	Disease	6	0.805	Disease	5	0.756	Decrease	5	-0.36	Deleterious
rs137852749	Disease	5	0.774	Disease	3	0.668	Decrease	8	-0.67	Deleterious
rs137852750	Disease	7	0.861	Disease	7	0.858	Decrease	1	-0.35	Deleterious
rs137852750	Disease	7	0.861	Disease	7	0.858	Decrease	1	-0.35	Deleterious
rs143740797	Disease	6	0.821	Disease	7	0.864	Decrease	9	-2.15	Deleterious
rs144356403	Neutral	3	0.371	Disease	3	0.657	Decrease	4	-1.33	Neutral

SNPs ID	SNPs & GO			PhD			I- MUTANT			PROVEN
	Prediction	RI	probability	Prediction	RI	probability	Prediction	RI	DDG	Prediction
rs147943794	Neutral	9	0.074	Neutral	4	0.289	Increase	0	0.14	Neutral
rs148257675	Neutral	3	0.334	Disease	4	0.698	Decrease	1	-0.13	Neutral
rs148682262	Neutral	3	0.334	Neutral	4	0.287	Decrease	8	-1.89	Neutral
rs148770894	Neutral	2	0.402	Disease	5	0.736	Increase	6	-0.14	Neutral
rs199915496	Disease	0	0.514	Disease	5	0.765	Increase	2	-0.38	Deleterious
rs199954814	Neutral	5	0.234	Neutral	1	0.432	Increase	2	0.29	Deleterious
rs200339485	Disease	1	0.562	Disease	3	0.635	Decrease	4	-1.39	Deleterious
rs201781338	Neutral	4	0.321	Disease	3	0.647	Decrease	4	-0.53	Neutral
rs201938348	Neutral	5	0.269	Disease	4	0.715	Decrease	3	-0.48	Neutral
rs368027047	Disease	0	0.512	Disease	5	0.731	Decrease	5	-0.76	Deleterious
rs369545039	Neutral	5	0.239	Neutral	1	0.812	Decrease	6	-1.92	Deleterious
rs374694591	Disease	6	0.811	Disease	3	0.669	Decrease	6	-1.01	Deleterious
rs375492148	Neutral	5	0.367	Disease	3	0.771	Decrease	6	-0.93	Neutral
rs377763312	Neutral	5	0.233	Disease	4	0.676	Increase	4	0.73	Deleterious

- RI: reliability index.
- DDG: unfolding free energy change.
- The highlight result shows the significant SNPs.

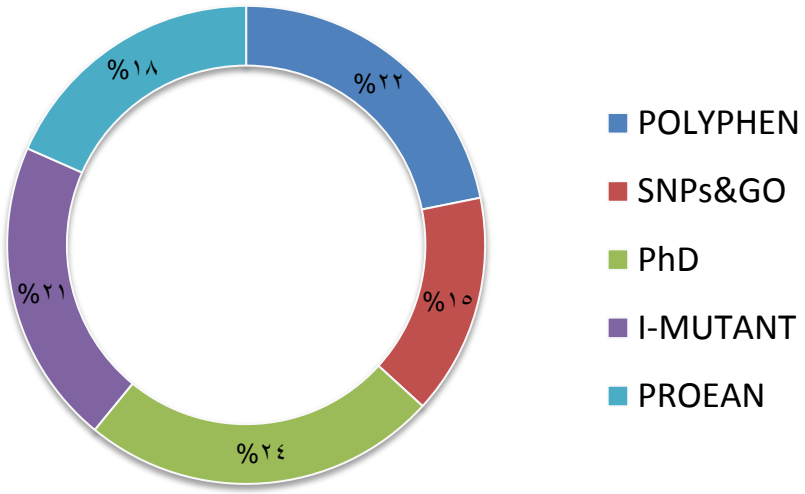

figure (2) illustrates the percentage of significant SNPs in several softwares.

3.4 Predictions by Project Hope:

A/ G mutation: (rs137852749), the mutant causes differences in charge because the residue charge is lost by this mutation, the mutant residue is smaller than wild- type, the mutation will cause an empty space in the core of the protein.

In (rs137852750), the mutant residue introduce a charge which can lead to protein folding problems, the mutant residue is bigger than wild- type, the mutant will cause loss of hydrophobic interaction in the core of the protein.

The rs137852746 substitution forms a Tryptophan; the residue charge is lost by this mutation, the mutant residue is bigger than the wild- type and probably will not fit, the mutation will cause loss of hydrogen bonds in the core of the protein and as a result disturb correct folding.

The rs137852750 substitution Alanine into Valine, the mutant residue is bigger than the wild-type and probably will not fit.

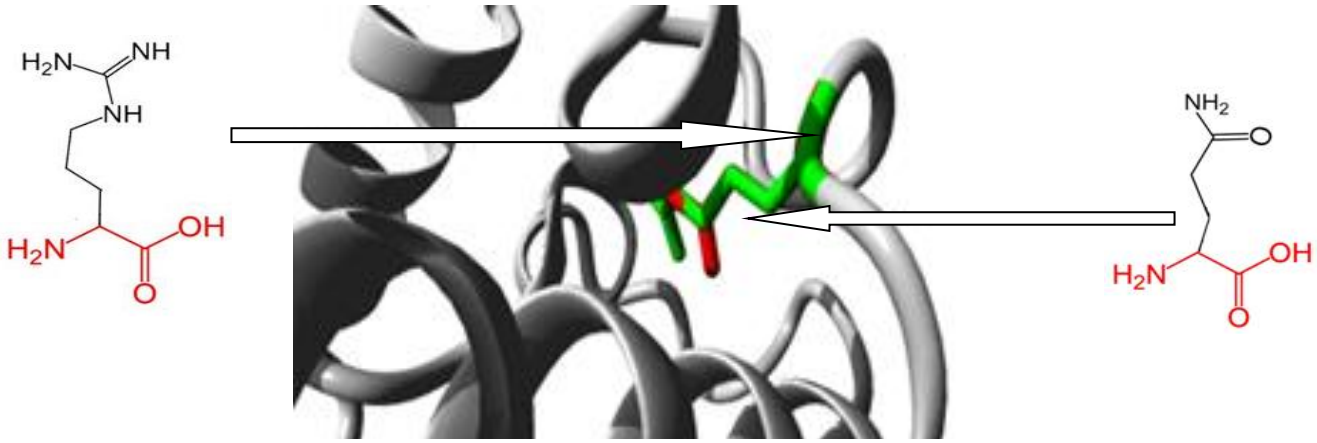

Figure (3) using project hope: rs137852749 Substitution Arginine into Glutamine at position 491, the main protein core is shown in gray color while the wild type and mutated residues are shown in green and red colors respectively.

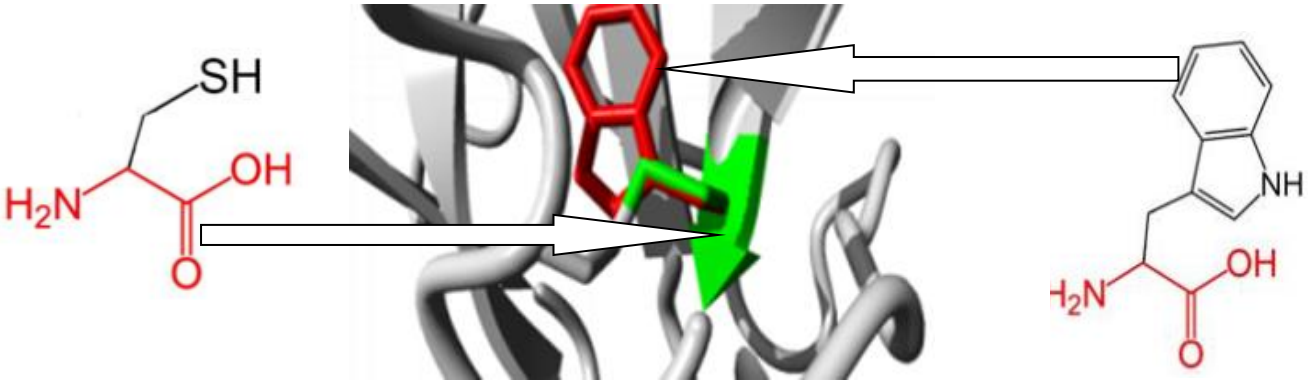

Figure (4) using project hope: rs137852743 substitution cysteine into tryptophan at position 118, the main protein core is shown in gray color while the wild type and mutated residues are shown in green and red colors respectively.

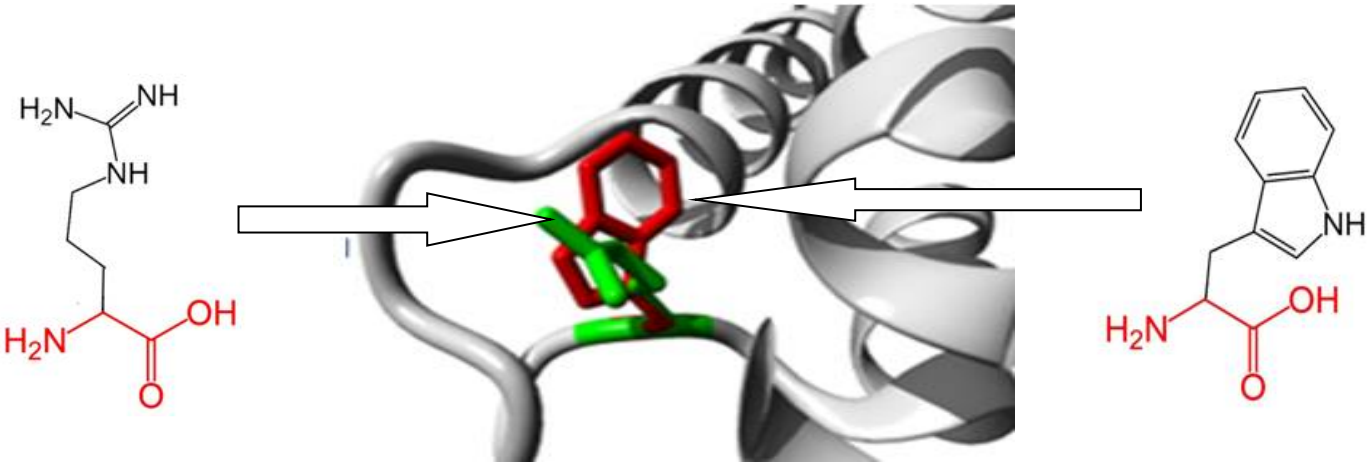

Figure (5) using project hope: rs137852749 Substitution Arginine into Tryptophan at position 491, the main protein core is shown in gray color while the wild type and mutated residues are shown in green and red colors respectively.

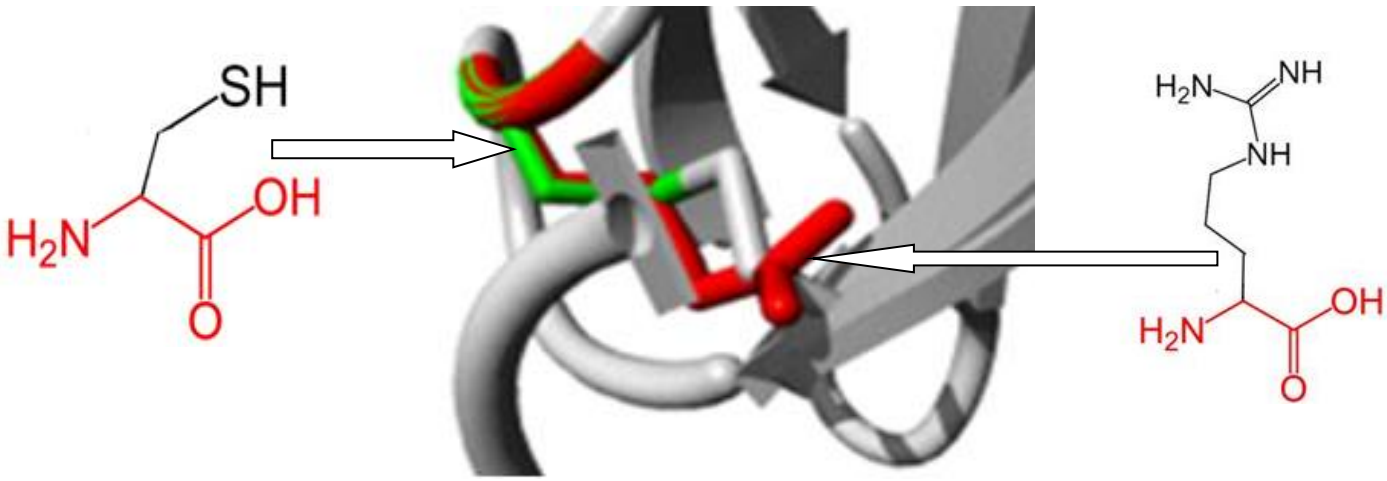

Figure (6) using project hope: rs137852750 substitution cysteine into Arginine at position 123, the main protein core is shown in gray color while the wild type and mutated residues are shown in green and red colors respectively.

According to the previous study the rs137852749 showed the same significant result (Sanna ν et al, 2015)[14]

4. Conclusion

Based on the overall results from this study, the significant SNPs were shared eight significant SNPs that confirming result.

The rs137852744, rs137852745, rs137852746, rs137852749, rs137852750, rs137852750, rs143740797and rs374694591 were shown to have highly damaging should be cause the pulmonary arterial hypertension that the mutations considered important in causing this disease.

References:

1- Herpin, et al. Transforming growth factor- β- related proteins: an ancestral and widespread superfamily of cytokines in metazoans.2004;(28)

2- Hyun,et al. Critical molecular switches involved in BMP- 2- induced osteogenic differentiation of mesenchymal cells.2005;(366) 51- 57.

3- Guillermo Pousada, Vincenzo Lupo, Sheila Cástro- Sánchez, María Álvarez- Satta, Ana Sánchez-Monteagudo, Adolfo Baloira, Carmen Espinós, and Diana Valverde Molecular and functional characterization of the BMPR2 gene in Pulmonary Arterial Hypertension Sci Rep. 2017; 7: 1923.

4- Newman J H, Trembath R C, Morse J A et al.. Genetic basis of pulmonary arterial hypertension: current understanding and future directions. J Am Coll Cardiol. 2004; 43(12, Suppl S) 33S- 39S

5- Joshua P. Fessel, James E. Loyd, Eric D. Austin et al.The Genetics of Pulmonary Arterial Hypertension in the Post- BMPR2 Era. 2011: 10.4103/ 2045- 8932.87293

6- Maricel G. Kann. Advances in translational bioinformatics: computational approaches for the hunting of disease genesBrief Bioinform. 2010 Jan; 11(1): 96–110.

7- Erika B.Rosenzweig, et al Clinical Implications of Determining BMPR2 Mutation Status in a Large Cohort of Children and Adults With Pulmonary Arterial Hypertension. 2008: (27)- 6,: 668- 674

8- Warde- Farley, D., et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic acids research.(2010) ; 38(suppl_2): W214- W22.

9- Hu J, Ng PC. Predicting the effects of frame shifting indels Genome Biol. 2012; vol. 13 pg. R9.

10- Adzhubei, I., et al. Predicting functional effect of human missense mutations using PolyPhen-2. Current protocols in human genetics: (2013) ; 7.20. 21- 27.20. 41.

11- Capriotti E, Fariselli P, Casadio R. I- Mutant 2.0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res. 2005; 33: 306–310.

12- Hanka Venselaar, Tim AH, Beek, Remko KP Kuipers,Maarten L Hekkelman and Gert Vriend. Protein structure analysis of mutations causing inheritable diseases. An e- Science approach with life scientist friendly interfaces. BMC Bioinformatics. 2010; 11: 548.

13- Warde- Farley, D., et al. The Gene MANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic acids research.(2010) ; 38(suppl_2): W214- W220.

14- Sanna Vattulainen, Joonas Aho Pertteli, Salmenperä Siina, Bruce Jonna, Tallila Massimiliano, Gentile Marja, Sankelo Tarja, Laitinen Juha, W. Koskenvuo, Tero-Pekka, Alastalo Samuel, Mylly kangas. Accurate genetic diagnosis of Finnish pulmonary arterial hypertension patients using oligonucleotide-selective sequencing.2015: 10.1002- 147.