Available online at www.scholarsresearchlibrary.com

Scholars Research Library

Der Pharmacia Lettre, 2013, 5 (5):265-270
(http://scholarsresearchlibrary.com/archive.html)

Scholars Research Library
ISSN 0975-5071
USA CODEN: DPLEB4

Prediction Of Deleterious Nonsynonymous Single Nucleotide Polymorphism (nsSNP) of ACVR1 Gene By Computitional Method

VineetVaibhav, Chandan Kumar Gautam, MukundMhadhav, V. Shanthi*

School of Bio Sciences and Technology (SBST), Vellore Institute of Technology, Vellore 632014, Tamil Nadu, India

---

ABSTRACT

Fibrodysplasia Ossificans Progressiva (FOP) is a genetic disorder of connective tissue. Damaging missense point mutations in ACVR1gene, which is accountable for FOP was examined by computational methods. Out of 19 variants of this gene obtained from NCBI; 3 variants namely R258S, G356D and R206H were found to bedeleterious. These variants were reported as probably damaging, deleterious and decreased DDG by PolyPhen 2.0, SIFT and I-mutant 2.0 respectively. The deleterious condition was further verified by usingPHD-SNP, PANTHER and SNPs&GO. Out of these 3 variants, only R206H was reported to be neutral by SNPs&GO. The results of computational analysis were alike to the wet lab result reported in research literatures.

Keywords*:* SIFT, I-MUTANT 2.0, PolyPhen 2.0, SNPs&GO, ACVR1, PHD-SNP

INTRODUCTION

ACVR1 Gene also known as activinA receptor, type I is a protein encrypting gene located on chromosome 2q23-q24. Activins are dimmersignaling proteins, which signal through a heteromer complex of receptor serine kinases(Type 1 and Type 2) [1]. These gene mutations lead to fibrodysplasiaossificansprogressive (FOP), the most severe and crippling disorder of extraskeletal (heterotopic) ossification inhumans [2]. It has been reported that one out of two million people in the world suffer from this disease [3,4]. This genetic disease is characterized by the congenital malformations of the toes and progressive heterotypical ossification in a specific anatomical pattern [5, 6, 7]. BMP (Bone Morphogenetic Protein) is one of the major factors involved in the transformation of Mesenchymal cells inbone and cartilage [8]. Mutation in the ACVR1 gene which codes for a receptor of BMP 7 is responsible for FOP [9]. ACVR1/ALK2 is one of four type I receptors that mediate signaling through BMPs. At birth, individuals with FOP appear normal except for the malformed great toes, which is observed almost in all affected individuals [9]. In the human genome, a single base change called single nucleotide polymorphism (SNP) is the most frequent type of genetic variation [10] [11]. A single-nucleotide polymorphism is a variationoccurring in DNA sequence due to alteration in single nucleotide of a genome. Itusually occurs in the non-coding region of the genome and its distribution pattern is non homogeneous. The density of SNP is determined by several factors, some of the factors are natural selection [12], genetic recombination and mutation rate [13]. When SNP occurs in protein coding regions and causes the change in amino acid coding for corresponding protein, it is called nonsynonymous Single Nucleotide Polymorphisms (nsSNPs). The studies of SNPs are important in the fields of crop, livestock breeding program and genome wide association studies (GWAS). The knowledge of SNPs is helpful in studying

pharmacokinetics, pharmacodynamics[14],Pharmacogenomics [15] and in designing drug best suited for an individual [16]

MATERIALS AND METHODS

2. METHODS

2.1. Analysis of functional consequences of point mutations by SIFT

SIFT server available at the linkhttp://sift.jcvi.org/www/SIFT_dbSNP.htmlwas used for analyzing deleterious single amino acid polymorphisms[17]. SIFT prediction is based on the degree of conservation of amino acid residues in sequence alignments derived from closely related sequences, collected through PSI-BLAST (SIFT http://sift.jcvi.org/). It predicts whether an amino acid substitution will affect protein function. It works by considering the position at which the change occurred and the type of amino acid change [17].

2.2. Structural homology-based method (PolyPhen 2.0) for simulation for functional change in a point mutation

The server PolyPhen 2.0, which is available at http://coot.embl.de/PolyPhen/ was used to analyze the damaged coding nsSNPs. At the structural level, it is considered to be very important to understand the functional activity of the concerned protein [18]. PolyPhen-2.0 is an automatic tool for prediction of the possible impact of an amino acid substitution on the structure and function of a human protein [19]. The prediction is based on a number of sequences, phylogenetic, and structural features characterizing the substitution [20].

2.3. Support vector machine (I-Mutant 2.0) for predicting stability changes caused by SAPs

I-Mutant 2.0 is used to predict the stability of the protein structure due to single point mutation. This prediction is based upon single point mutation starting from the protein structure or sequence. The protein sequences or the PDB code of the protein can be used as input and the value of free energy change or its sign is obtained as the output[21].

2.4. SNPs and GO for identification of disease causing mutation

SNPs & GO method was used to predict whether a variation is related to any disease or not by exploiting the corresponding protein functional annotation [22]. It is available this http://snps.biofold.org/snps-and-go/pages/method.html link. The prediction of PHD-SNP, PANTHER and SNPs&GOis obtained from this server. Analysis by this server predicts the disease causing variant.

RESULTS AND DISCUSSION

TABLE:1 SIFT analysis

SNo.	SNP	Variants	Prediction	**Toleranceindex**
01	rs112489929	G43S	TOLERATED	0.80
02	rs74905152	V132A	TOLERATED	0.34
03	rs71421047	K148N	TOLERATED	0.30
04	rs55957214	S41F	TOLERATED	0.56
05	rs41265129	L184F	TOLERATED	0.14
06	rs34056189	H47Q	TOLERATED	0.29
07	**rs13406336**	**A15G**	**DAMAGING**	**0.01**
08	**rs121912680**	**R258S**	**DAMAGING**	**0.00**
09	**rs121912679**	**G356D**	**DAMAGING**	**0.00**
10	**rs121912678**	**R206H**	**DAMAGING**	**0.00**
11	rs202095604	V434M	TOLERATED	0.23
12	rs201872272	E212D	TOLERATED	0.08
13	rs201452185	V162M	TOLERATED	0.21
14	rs188547477	R160G	TOLERATED	0.12
15	**rs146610930**	**S187L**	**DAMAGING**	**0.01**
16	rs145780526	T468S	TOLERATED	0.98
17	rs145150729	T172S	TOLERATED	0.12
18	rs144048685	K346N	TOLERATED	0.21
19	rs138808563	K112N	TOLERATED	0.53

The SNP Data set
From NCBI; 19 mutational changes for ACVR1 gene, namely G43S, V132A, K148N, S41F, L184F, H47Q, A15G, R258S, G356D, R206H, V434M, E212D, V162M, R160G, S187L, T468S, T172S, K346N and K112N was found responsible for FibrodysplasiaOssificansProgressiva (FOP).

Deleterious single point mutants identified by the SIFT program
Tolerated and damaging variants were identified based on the tolerance index. Amongstthe 19 variants, 5 were found to be damaging means they are havingtolerance index scores of less or equal to 0.05 and rest 14 were tolerated as they are having more than 0.05. Among damaging type; 3(R206H, G356D, R258S) variant was having tolerance index scores equal to 0.00 and rest two A15G and S187L were having tolerance index scores 0.01.

Damaging single point mutations identified by the PolyPhen 2.0 server
ThePolyPhen2.0 predicted the possible impact of an amino acid substitution on the structure and function of a human protein. The output was obtained in the form of a score. The variants were assigned to be benign, probably damaging or possibly damaging based on the score obtained. 5 out of 19 variantswere found to be probably damaging type and rest 14 were found benign. Probably damaging type wereS41F, R258S, G356D, R206H, S187L having scored0.818, 0.998, 1.00,1.00 and 0.999 respectively.

TABLE: 2 PolyPhen 2.0 server analysis

SNo.	rsID	Variants	PolyPhane	Score
01	rs112489929	G43S	BENIGN	0.027
02	rs74905152	V132A	BENIGN	0.068
03	rs71421047	K148N	BENIGN	0.295
04	**rs55957214**	**S41F**	**PROBABLY DAMAGING**	**0.818**
05	rs41265129	L184F	BENIGN	0.037
06	rs34056189	H47Q	BENIGN	0.001
07	rs13406336	A15G	BENIGN	0.016
08	**rs121912680**	**R258S**	**PROBABLY DAMAGING**	**0.998**
09	**rs121912679**	**G356D**	**PROBABLY DAMAGING**	**1.00**
10	**rs121912678**	**R206H**	**PROBABLY DAMAGING**	**1.00**
11	rs202095604	V434M	BENIGN	0.034
12	rs201872272	E212D	BENIGN	0.293
13	rs201452185	V162M	BENIGN	0.106
14	rs188547477	R160G	BENIGN	0.017
15	**rs146610930**	**S187L**	**PROBABLY DAMAGING**	**0.999**
16	rs145780526	T468S	BENIGN	0.002
17	rs145150729	T172S	BENIGN	0.001
18	rs144048685	K346N	BENIGN	0.005
19	rs138808563	K112N	BENIGN	0.000

Identification of functional variants by I-mutant 2.0
Of the 19 variants, 16 were found with decreased stability using I-Mutant 2.0 server. Among these 16 variants, eight variants showed a ΔΔG value <1.0 and eight variants showed a ΔΔG value >-1.0 as depicted in Table. Only two variants were having a ΔΔG value < -2.0.

Out of 16 decreased stability variant three of them (T172S,S41F,G43S) showed positive ΔΔG value. Three of them (H47Q,R258S,R160G) changed their positively charged amino acid to uncharged amino acid, two of the variants (G43S and G356D) changed their non-polar amino acid to polar amino acid, three of the variants (S41F,R160G, S187L) changed their polar amino acid to non-polar amino acid, eleven of variants (V132A,L184F, H47Q, A15G, R258S,R206H, V434M, E212D, V162M,T468S, T172S) didn't get their amino acid changed on the basis of polarity. R206H retained its positive charge and E212D retained its negatively charged amino acids.

Identification of diseased variants by PANTHER
The impact of single point mutation on protein function was further validated by PANTHER. Out of 19 variant predictions for 10 was not found, 3 were disease causing and rest 6 were found to neutral. The 3 variants R258S, G356D and R206H which were disease causing having there score value 0.726, 0.992 and 0.640 respectively. Following 6 variants were found to be neutral L184F, V434M, E212D, S187L, T468S and K346N.

TABLE: 3 I-Mutant 2.0

SNo.	SNP	Variants	NUCLIOTIDE CHANGE	I MUTANT	DDG
01	**rs112489929**	**G43S**	**G/A**	**DECREASE**	**0.17**
02	**rs74905152**	**V132A**	**T/C**	**DECREASE**	**-2.70**
03	rs71421047	K148N	A/T	INCREASE	0.86
04	**rs55957214**	**S41F**	**C/T**	**DECREASE**	**0.10**
05	**rs41265129**	**L184F**	**G/T**	**DECREASE**	**-0.76**
06	**rs34056189**	**H47Q**	**C/G**	**DECREASE**	**-0.32**
07	**rs13406336**	**A15G**	**C/G**	**DECREASE**	**-1.10**
08	**rs121912680**	**R258S**	**G/C**	**DECREASE**	**-2.65**
09	**rs121912679**	**G356D**	**G/A**	**DECREASE**	**-1.14**
10	**rs121912678**	**R206H**	**G/A**	**DECREASE**	**-1.23**
11	**rs202095604**	**V434M**	**G/A**	**DECREASE**	**-1.49**
12	**rs201872272**	**E212D**	**G/C**	**DECREASE**	**-0.31**
13	**rs201452185**	**V162M**	**G/A**	**DECREASE**	**-0.58**
14	**rs188547477**	**R160G**	**C/G**	**DECREASE**	**-1.19**
15	**rs146610930**	**S187L**	**C/T**	**DECREASE**	**-1.39**
16	**rs145780526**	**T468S**	**C/G**	**DECREASE**	**-0.30**
17	**rs145150729**	**T172S**	**C/G**	**DECREASE**	**0.29**
18	rs144048685	K346N	G/T	INCREASE	0.26
19	rs138808563	K112N	A/C	INCREASE	-0.20

TABLE: 5 Panther analysis

SR.NO.	SNP	VARIANTS	NUCLIOTIDE CHANGE	PANTHER	R.I.	SCORE
01	rs112489929	G43S	G/A	N/A	N/A	N/A
02	rs74905152	V132A	T/C	N/A	N/A	N/A
03	rs71421047	K148N	A/T	N/A	N/A	N/A
04	rs55957214	S41F	C/T	N/A	N/A	N/A
05	rs41265129	L184F	G/T	Neutral	4	0.317
06	rs34056189	H47Q	C/G	N/A	N/A	N/A
07	rs13406336	A15G	C/G	N/A	N/A	N/A
08	**rs121912680**	**R258S**	**G/C**	**Disease**	**5**	**0.726**
09	**rs121912679**	**G356D**	**G/A**	**Disease**	**10**	**0.992**
10	**rs121912678**	**R206H**	**G/A**	**Disease**	**3**	**0.64**
11	rs202095604	V434M	G/A	Neutral	6	0.182
12	rs201872272	E212D	G/C	Neutral	3	0.345
13	rs201452185	V162M	G/A	N/A	N/A	N/A
14	rs188547477	R160G	C/G	N/A	N/A	N/A
15	rs146610930	S187L	C/T	Neutral	6	0.199
16	rs145780526	T468S	C/G	Neutral	7	0.127
17	rs145150729	T172S	C/G	N/A	N/A	N/A
18	rs144048685	K346N	G/T	Neutral	4	0.279
19	rs138808563	K112N	A/C	N/A	N/A	N/A

Identification of diseased variants by PHD-SNP

Disease and neutral condition were further investigated by PHD-SNP. 10 out of 19 variants were found disease causing and rests 9 were found neutral. Out of 10 disease causing variant R160G, E212D, R258S, G356D and R206H showing score value above 0.8, which means they were more prone to disease and other variant S187L, A15G, V132A, S41F and L184F showing score value between 0.5 and 0.8. The variant with the highest score is having a relatively higher Reliability Index like R258S, G356D, R206H, V434M and E212D having R.I. value 9, 10, 7 and 7 respectively.

Identification of diseased variants by SNPs&GO

Of the 19 variants, 2 (R258S, G356D) variants with high score of 0.716 and 0.826 were found to be disease causing by this server. Rests 17 variants were found neutral. The reliability index (R.I.) of these two variants were 4 and 7 respectively.

TABLE: 4 PHD-SNP Analysis

SR.NO.	SNP	VARIANTS	NUCLEOTIDE CHANGE	PHD-SNP	R.I.	SCORE
01	rs112489929	G43S	G/A	Neutral	3	0.335
02	**rs74905152**	**V132A**	**T/C**	**Disease**	**1**	**0.534**
03	rs71421047	K148N	A/T	Neutral	3	0.368
04	**rs55957214**	**S41F**	**C/T**	**Disease**	**4**	**0.703**
05	**rs41265129**	**L184F**	**G/T**	**Disease**	**1**	**0.556**
06	rs34056189	H47Q	C/G	Neutral	3	0.373
07	**rs13406336**	**A15G**	**C/G**	**Disease**	**4**	**0.703**
08	**rs121912680**	**R258S**	**G/C**	**Disease**	**9**	**0.96**
09	**rs121912679**	**G356D**	**G/A**	**Disease**	**10**	**0.983**
10	**rs121912678**	**R206H**	**G/A**	**Disease**	**7**	**0.871**
11	rs202095604	V434M	G/A	Neutral	7	0.137
12	**rs201872272**	**E212D**	**G/C**	**Disease**	**7**	**0.841**
13	rs201452185	V162M	G/A	Neutral	3	0.344
14	**rs188547477**	**R160G**	**C/G**	**Disease**	**7**	**0.843**
15	**rs146610930**	**S187L**	**C/T**	**Disease**	**3**	**0.634**
16	rs145780526	T468S	C/G	Neutral	4	0.312
17	rs145150729	T172S	C/G	Neutral	2	0.42
18	rs144048685	K346N	G/T	Neutral	3	0.351
19	rs138808563	K112N	A/C	Neutral	2	0.397

TABLE: 6 SNPs&GO

SR.No.	SNP	Variants	NUCLEOTIDE CHANGE	SNPs&GO	R.I.	SCORE
01	rs112489929	G43S	G/A	Neutral	9	0.074
02	rs74905152	V132A	T/C	Neutral	8	0.102
03	rs71421047	K148N	A/T	Neutral	8	0.097
04	rs55957214	S41F	C/T	Neutral	6	0.194
05	rs41265129	L184F	G/T	Neutral	8	0.85
06	rs34056189	H47Q	C/G	Neutral	9	0.072
07	rs13406336	A15G	C/G	Neutral	7	0.135
08	**rs121912680**	**R258S**	**G/C**	**Disease**	**4**	**0.716**
09	**rs121912679**	**G356D**	**G/A**	**Disease**	**7**	**0.826**
10	rs121912678	R206H	G/A	Neutral	1	0.425
11	rs202095604	V434M	G/A	Neutral	10	0.016
12	rs201872272	E212D	G/C	Neutral	6	0.185
13	rs201452185	V162M	G/A	Neutral	9	0.061
14	rs188547477	R160G	C/G	Neutral	4	0.308
15	rs146610930	S187L	C/T	Neutral	9	0.068
16	rs145780526	T468S	C/G	Neutral	10	0.021
17	rs145150729	T172S	C/G	Neutral	9	0.074
18	rs144048685	K346N	G/T	Neutral	8	0.076
19	rs138808563	K112N	A/C	Neutral	9	0.065

CONCLUSION

After analyzing missense point mutations in ACVR1 by computational methods by the following servers PHD-SNP, PANTHER, SNPs GO, PolyPhen 2.0, SIFT and I MUTANT 2.0 it is found that among all nineteen variants those were analyzed, out of which 5 were found damaging by SIFT server, 5 were found probably damaging by PolyPhen 2.0 server and 16 of them were found less stable and damaging by I MUTANT 2.0. PHD-SNP identified 10 disease causing variants, 3 and 2 disease causing variant is found by PANTHER and SNPs&GO respectively. Three variants namely R258S, G356D and R206H were continuously found damaging, less stable or disease causing as tabulated in table: 7; except in SNPs&GO in which variant R206H is reported neutral. Variants R258S is found disease causing by restriction enzyme digestion and R206H were found disease causing by DNA sequence analysis[23]. ALK2 (G356D) is also reported to induce heterotopic bone formation via activation of a BMP-regulated Smad pathway[24].

TABLE: 7

MUTATIONAL CHANGE	PHD-SNP	PANTHER	SNPs GO	PolyPhen 2.0	SIFT	I MUTANT 2.0	reference
R258S	Disease	Disease	Disease	probably damaging	damaging	decrease	[23]
G356D	Disease	Disease	Disease	probably damaging	damaging	decrease	[24]
R206H	Disease	Disease	Neutral	probably damaging	damaging	decrease	[23]

REFERENCES

[1] NCBI (**2013**) Retrieved 22 september from http://www.ncbi.nlm.nih.gov/gene/90.
[2] F. S. Kaplan, E. M. Shore, J. M. Connor, Connective tissue and its heritable disorders: molecular, genetic, and medical aspects, Wiley-Liss, New York, **2002**, 2, 827–840.
[3] J. M. Connor, *N. Engl. J. Med.*, **1996**, 335,591–593.
[4] R. Smith, *Clin. Orthop.*,**1998**, 346:7–13.
[5] S. Frederick, M. D. Kaplan, In: R. J. Pignolo(Ed), The International Clinical Consortium on FibrodysplasiaOssificansProgressiva, ClinProcIntClin Consort FOP, May. **2011**, 4, Philadelphia, 1–100.
[6] F. S. Kaplan, M. M. Le, D. L. Glaser, R. J. Pignolo, R. E. Goldsby, J. A. Kitterman, *Best Pract. Res. Clin. Rheumatol*, **2008**, 22, 191–205.
[7] J. M. Connor, D. A. P. Evans, *J. Med. Genet.*, **1982**, 19, 35–9.
[8] M. Macias-Silva, P. A. Hoodless, S. J. Tang, M. Buchwald, J. L. Wrana, *J.Biol. Chem.*, **1998**, 273, 25628–25636.
[9] E. M. Shore, M. Xu, G. J. Feldman, D. A. Fenstermacher, T. J. Cho, I. H. Choi, J. M. Connor, P. Delai, D. L. Glaser, M. LeMerrer, R. Morhart, J. G. Rogers, R. Smith, J. T. Triffitt, J. A. Urtizberea, M.Zasloff, M. A. Brown, F. S. Kaplan, *Nat. Genet.*, **2006**, 38,525–527.
[10] M. Cargill, D. Altshuler, J. Ireland, P. Sklar, K. Ardlie, N. Patil, N. Shaw, C. R. Lane, E. P. Lim, N. Kalyanaraman, *Natural Genetics*, **1999**, 22, 231–238.
[11] J. H. Nadeau, *Nature*, **2002**, 420, 517– 518.
[12] L. B. Barreiro, G. Laval, H. Quach, E. Patin, L. Quintana-Murci, *Nature Genetics*, **2008**, 40, 340–345.
[13] Nachman, W.Michael, *Trends in genetics*, **2001**, 17 (9): 481–485.
[14] C. E. Staatz, L. K. Goodman, S. E. Tett, *Clin. Pharmacokin*, **2010**, 49(4), 207-21.
[15] B.A.Zilfalil, *Malays J. Med. Sci*., **2005**, 12(2), 4–12.
[16] Coriell (2013). Retrieved from on 18 september **2013** from http://www.coriell.org/personalized-medicine/dna-genes-and-snps).
[17] C. P. Ng, S. Henikoff, *Nucleic Acids Research*, **2003**, 31, 3812–3814.
[18] V. Ramensky, P. Bork, S. Sunyaev, *Nucleic Acids Research*, **2002**, 30, 3894–3900.
[19] I. A. Adzhubei, S. Schmidt, L. Peshkin, V. E. Ramensky, A. Gerasimova, P. Bork, A. S. Kondrashov, S. R. Sunyaev, *Nat. Methods*., **2010**, 7, 248-249.
[20] A.Ivan, M. Danie, Jordan, R. Shamil, Sunyaev,*Curr. Protoc. Hum. Genet.*, **2013**, 76, 7.20.1-7.20.41.
[21] C. Emidio, F. Piero, C. Rita, *Nucleic Acids Res*., **2005**, 1(33), W306–W310.
[22] R. Calabrese, E. Capriotti, P. Fariselli, P. L. Martelli, R. Casadio, *Hum. Mutat*, **2009**, 30(8), 1237-44.
[23] Ç. E. Yazıcıoğlu, V. Karatosun, S. Kızıldağ, D. Özsoylu, S. Kavukçu, *Gene*, **2013**, 515, 444–446.
[24] F. Toru, K. Kazuhiro, N. Junya, K. Shoichiro, A. Masumi, I. Kenji , J. Eijiro, K.Tetsuo, M.Yuichi, M. Masaru, M. Kohei, N. Konosuke, N. Akira, T. Hiroshi, O. Yasushi, O. Akira, O. Hiromi, O. Ichiro, Y. Tetsuya, H. Nobuhiko, F. Hirokazu, K. Takenobu, *BiochemBiophys. Res. Commun.*, **2008**, 377(3), 905-9.