IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

COMPUTATIONAL EVALUATION OF HIGH RISK SNPs IN CTHRC1 GENE RESPONSIBLE FOR CARCINOMA

Authors:

Kaishiv Joshi[1], Samonee Ralmilay[1], Simranjeet Kaur[1], Dr. Rajinder Kaur[1]

Affiliations:

[1,2,3,4]Department of Human Genetics, Punjabi University, Patiala, Punjab, India.

ABSTRACT

CTHRC1 gene have a role in vascular calcification of carotid artery lesions. In-vitro and in-vivo studies have been carried out to confirm its connection to cancers especially with esophageal adenocarcinoma and squamous cell carcinoma. We used a variety of computational methods to find nsSNPs that are harmful to the structure and/or function of the CTHRC1 protein and may be causing this disease. Computational analysis was performed by different *in silico* tools including PROVEAN, Poly-Phen-2, I-Mutant, PhD-SNP. Results of our study indicated that 12 SNPs which are found to be highly deleterious: N162D, R146Q, G67R, S142R, N190S, I172T, W214R, R235C, R156G, R235H, S219L and C149Y. This is the first systematic research to use in silico methods to evaluate CTHRC1gene variants, so it will be extremely useful when planning large-scale studies and designing precision medicines.

KEYWORDS

CTHRC1, SNPs, POLYPHEN, PROVEAN, I-MUTANT, PhD-SNP.

INTRODUCTION

CTHRC1 is a glycoprotein that is expressed in the vascular calcifications of carotid artery lesions and helps with vascular remodelling. Esophageal cancer is linked to a number of diseases, including Barret's oesophagus and adenocarcinoma. The paralog of this gene is METTL24 (Methyltransferase Like 24). The CTHRC1 gene is found on chromosome 8q22.3 (Tameda et al., 2014). When compared to normal tissues, the gene is overexpressed in Hepatocellular Carcinoma, resulting in a severe cancer problem. As a result of the gene's role, it is also classified as a new-HCC associated gene. It also aids in the reduction of collagen matrix

deposition (Wang C. et al., 2017).It is involved in the binding of different proteins as well as alveolar bone remodelling. It also encourages osteoblast differentiation and proliferation (Pyagay et al., 2005). This gene is responsible for the growth of the bladder in rat models (Zhu et al., 2018). It is found in normal fibroblasts and plays a significant role in cell migration (Zhao et al., 2018).CTHRC1 aids in the ossification of periodontal ligament stem cells as well. Many diseases linked to tumours are caused by the Collagen Triple Helix Repeat Containing 1 (CTHRC1) gene. The CTHRC1 gene is expressed at higher rates in esophageal tissue that causes esophageal squamous cell carcinoma, according to numerous in-vivo and in-vitro studies (Wang C. et al., 2017). CTHRC1 was overexpressed in cervical carcinoma tissues and Hela cells, according to a study (Zhang et al., 2015), and played a critical role in cell invasion and migration, which was likely mediated by activation of the Wnt/PCP pathway. Colorectal cancer (CRC), Non-Small Cell Lung Cancer (NSCLC), Pancreatic ductal adenocarcinomas (PDAC), Epithelial Ovarian Cancer (EOC), Oral squamous cell carcinoma (OSCC), Breast carcinoma, Hepatocellular carcinoma (HCC), and other tumours such as gastric, cervix, and thyroid tumours are all linked to the CTHRC1 gene. As a result, this gene plays a multifaceted role in a variety of cancers (Jiang et al., 2016). More cervical cancer studies are required to show the basic underlying mechanism of CTHRC1 in cervical cancer progression and metastasis. There has not been any in silico study of the CTHRC1 gene yet. Many nsSNPs found in the CTHRC1 gene that might be associated with various diseases. The wet lab experiments need time and cost. Nowadays, bioinformatics research focuses on the annotation of genes based on algorithms. In order to determine the association of a particular nsSNPs with the disease, various pathogenicity prediction tools have been used. In silico analysis of nsSNPs having deleterious value can be used for disease targets so as to make personalized medicines.

MATERIALS AND METHODOLOGY

SNP Dataset

The dbSNP (NCBI) database (https://www.ncbi.nlm.nih.gov) can be used to find SNPs for the CTHRC1 protein. The FASTA sequence of the CTHRC1 gene was obtained from the Uniprot database (https://www.uniprot.org).

Deleterious SNPs Prediction

Various computational methods were used to investigate the functional implications of the nsSNP in the human CHK2 gene.

I-Mutant

I-Mutant (version 2.0) is a support vector machine (SVM)-based method for predicting protein stability changes as a result of single point mutations (Dabhi & Mistry, 2014). This approach can be used to predict protein stability and the free energy transfer value (DDG). I-Mutant 2.0 is an extremely useful method for predicting whether a mutant protein will influence or not affect the folded protein (Capriotti et al., 2005).

Polymorphism Phenotyping v2 (Polyphen2)

Polyphen v2 **(http://genetic.bwh.harvard.edu/pph2/)** is a useful tool for predicting the effects of amino acid substitutions on the function and stability of human proteins while taking structural and evolutionary considerations into account. It calculates the PSIC (Position-Specific independent score). (Adzhubei et al., 2010).

Predictor of human Deleterious-SingleNucleotidePolymorphisms (PhD-SNP)

PhD-SNP (Predictor of human deleterious single nucleotide polymorphism) server is a Support Vector Machine (SVM) based method to discriminate between neutral and disease-related single point protein variants (Capriotti et al., 2005).

Protein Variation Effect Analyzer(PROVEAN)

PROVEAN is a useful tool for predicting multiple amino acid substitutions, insertions, and deletions (indels) that affect the protein's biological function. As a result, it predicts how amino acid variation affects a protein's function. (Choi et al., 2012). It is an online software available at http://provean.jcvi.org/index.php.

String Database

STRING, a resource method for retrieving interacting proteins, was used to predict the interactions of respective genes or proteins with other proteins. The string-db.org web server can be found at **https://string-db.org/**. (Szklarczyk et al., 2011).

RESULTS

SNP database

The SNPs of *CTHRC1* gene investigated in the present study was retrieved from dbSNP database (dbSNPNCBI: https://www.ncbi.nlm.nih.gov/snp/?term=chek2). Only 106 nsSNP of *CTHRC1gene* were selected for this investigation (Table 1).

Table 1) List of SNPs retrieved from dbSNP

Sr. No.	SNPs	Nucleic Acid substitutions	Amino Acid substitutions	Position
1	rs387907029	A>C	Q44P	44
2	rs6995610	A>G	K11E	11
3	rs35877609	G>T	G205V	205
4	rs36115601	G>A	R193H	193
5	rs74496760	G>A	G203E	203
6	rs76552709	G>T	G168V	168
7	rs111649208	A>G	N162D	162
8	rs141428810	C>A, G, T	R193S	193
9	rs143595509	A>G, T	T228A	228
10	rs145062750	G>A	R156H	156
11	rs145838701	C>G	S97R	97
12	rs145881116	C>A, G, T	R83R	83
13	rs147175675	G>A, C, T	R65Q	65
14	rs147418120	G>A	G216S	216
15	rs148320662	C>T	P183S	183
16	rs149006532	T>C, G	C93R	93
17	rs149662990	A>G	I74V	74
18	rs200589758	C>T	S234F	234
19	rs201319279	T>G	L178W	178
20	rs367999587	G>A	R146Q	146
21	rs369008583	G>A	G67R	67
22	rs370183649	C>G, T	L137V	137
23	rs371703105	C>G	P169R	169
24	rs373203731	C>T	P75L	75
25	rs374084026	C>A, T	A124E	124
26	rs374230888	C>T	P26S	26
27	rs374472802	A>C	S142R	142
28	rs375851229	T>C	I237T	237
29	rs533547163	G>A, C	Q4H	4
30	rs547221151	C>G	P10A	10
31	rs556757186	C>G	P27R	27
32	rs572457431	G>A	G25S	25
33	rs575106689	A>T	S29C	29
34	rs746901634	G>A	V139I	139
35	rs747313008	C>T	P26L	26
36	rs747888820	C>T	R146W	146
37	rs748020101	A>G	N190S	190
38	rs748246741	G>T	A28S	28
39	rs748382570	C>G, T	R65G	65
40	rs748845861	A>G	R150G	150
41	rs748968872	C>G, T	S196C	196
42	rs749448992	T>A	F98Y	98
43	rs749814525	G>C	G5A	5
44	rs750559913	C>T	R132C	132
45	rs750615097	T>C	I172T	172
46	rs751608144	T>C	I204T	204
47	rs751621588	A>G	M185V	185
48	rs752259371	C>G	S18C	18
49	rs752656527	G>T	G207V	207
50	rs752715912	C>T	S232L	232
51	rs752775043	G>C	K39N	39
52	rs753600909	C>G	I213M	213

53	rs753921011	C>T	L19F	19
54	rs754627863	C>T	R13W	13
55	rs754880047	A>G	T188A	188
56	rs754897420	G>C	L140F	140
57	rs755157789	A>G	Q11R	11
58	rs755445506	T>G	F86V	86
59	rs755644928	A>G	K20E	20
60	rs756063602	G>A	D210N	210
61	rs756369288	A>G,T	E96D	96
62	rs756872409	T>C	W214R	214
63	rs758780981	G>A	G25D	25
64	rs759365776	C>A	P69T	69
65	rs759391397	T>C	V209A	209
66	rs759883546	C>T	R12W	12
67	rs760875512	G>A,C	R14H	14
68	rs761116465	T>C	C154R	154
69	rs761452801	A>G	I123V	123
70	rs763118215	T>C	V197A	197
71	rs763322692	C>T	R235C	235
72	rs764245969	G>A	G15D	15
73	rs764595813	C>G,T	R156G	156
74	rs764820491	G>A	A124T	124
75	rs765155238	C>T	P25L	25
76	rs766177216	A>G	R6G	6
77	rs767061518	G>A	R83Q	83
78	rs767275808	G>A	G70E	70
79	rs767599276	C>T	A206V	206
80	rs767759004	G>A	R235H	235
81	rs768079876	C>T	P6L	6
82	rs768153022	G>T	R2L	2
83	rs768415647	G>T	S112I	112
84	rs770323100	C>T	A152V	152
85	rs770966489	T>G	F98L	98
86	rs771540899	C>T	S219L	219
87	rs771665520	C>G	I33M	33
88	rs772240481	G>A	R12Q	12
89	rs772990788	G>A	C149Y	149
90	rs773573431	G>C	V233L	233
91	rs773861122	T>G	C153G	153
92	rs774579022	G>A	A7T	7
93	rs774596955	A>G	E240G	240
94	rs775361593	G>A	C109Y	109
95	rs776145170	T>G	L120R	120
96	rs776981155	G>C	S27T	27
97	rs777739644	G>A	A212T	212
98	rs778614680	G>C	G216A	216
99	rs778869312	T>A	S101T	101
100	rs779328269	T>G	M23R	23
101	rs779641659	G>A	A40T	40
102	rs779752692	T>A,C,G	I123M	123
103	rs780761114	T>C	V17A	17
104	rs781221924	A>G	I189V	189
105	rs781431871	C>G	L21V	21
106	rs781710038	A>G	K87E	87

Prediction of functional nsSNPs in *CTHRC1*

The single nucleotide variants (106) identified by dbSNP analysis were subjected to computational analysis using a variety of methods. Out of 106 SNPs, 35 were found to be benign (36.84 percent), 10 were found to be potentially damaging (10.52 percent), and 50 were found to be potentially damaging nsSNPs using Polyphen 2 study (52.63 percent). Furthermore, the predictions were determined with the help of I-Mutant tool. According to the study performed using the I-Mutant tool, the DDG value of 83 SNPs was found to be less than or equal to 0, indicating that the stability of the protein is decreased. The stability of 23 nsSNPs was found to be were greater than 0 indicating non deleterious SNPs. Using the Provean tool, the 38 nsSNPs (40.86 percent) were found to be deleterious with the deleterious index -2.5 or below this threshhold. The remaining 68 nsSNPs were considered as neutral. All the 106 nsSNPs of the *CTHRC1* gene were further analyzed through PhD-SNP. PhD-SNP is a SVM based classifier which predicts the result through evolutionary information. According to the PhD-SNP study, 39 (37%) nsSNPs were considered as diseased variants, while 67 (63%) were considered to be neutral. Further sorting of nsSNPs according to their energy values indicates only 12 SNPs were found to be extremely deleterious nsSNPs (Table 2). As a result, these non-synonymous SNPs were found to be strongly related to disease progression.

Table 2: Combined table of all the SNPs showing deleterious nature

Sr. N	SNPs	Amino Acid substitutions	I-Mutant Results		PhD-SNP Results	Provean Results		Polyphen 2 Results	
			Stability	DDG		Score	Predictions	Score	Predictions
1	rs111649208	N162D	DEC	-0.39	DIS	-3.53	DEL	0.006	Benign
2	rs367999587	R146Q	DEC	-0.78	DIS	-2.833	DEL	1	PRD
3	rs369008583	G67R	DEC	-1.22	DIS	-4.017	DEL	1	PRD
4	rs374472802	S142R	DEC	-0.46	DIS	-2.5	DEL	0.987	PRD
5	rs748020101	N190S	DEC	-1.06	DIS	-2.617	DEL	0.997	PRD
6	rs750615097	I172T	DEC	-2.07	DIS	-4.375	DEL	0.321	Benign
7	rs756872409	W214R	DEC	-1.39	DIS	-3.7	DEL	0.998	PRD
8	rs763322692	R235C	DEC	-1.61	DIS	-6.367	DEL	1	PRD
9	rs764595813	R156G	DEC	0.23	DIS	-6.5	DEL	1	PRD
10	rs767759004	R235H	DEC	-1.06	DIS	-3.658	DEL	1	PRD
11	rs771540899	S219L	DEC	-0.3	DIS	-2.66	DEL	0.183	Benign
12	rs772990788	C149Y	DEC	-0.81	DIS	-2.733	DEL	0.998	PRD

DEC- Decrease, DIS- Disease, DEL- Deleterious, PRD- Probably Damaging

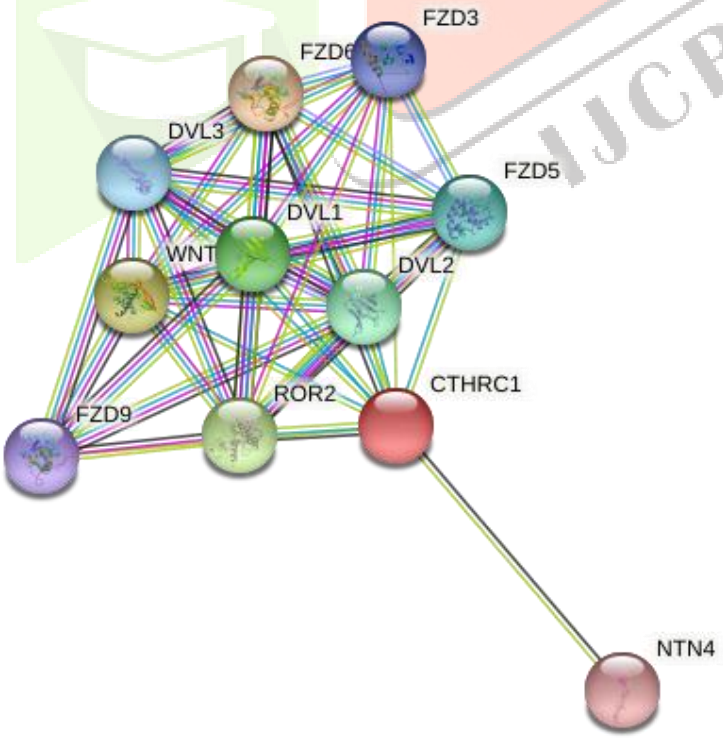

Figure 1. The representation of functional proteins of CTHRC1 gene

Discussion

In most diseases, single nucleotide polymorphisms play a significant role. dbSNPs has identified over 4 million distinct human single nucleotide polymorphisms (SNPs), and 2 percent of the SNPs associated with monogenic diseases are found in the protein coding region, implying that these SNPs may be linked to complex inherited disease traits. The simplest approach is to test the functional consequences of a variant using a functional assay, but this is both expensive and time consuming. As a result, we used a computational approach to analyse SNPs in the CTHRC1 gene, using a variety of in silico methods based on different algorithms. In the NCBI dbSNP (database), 3400 SNPs in the human CTHRC1 gene have been found in non-coding, coding, and regulatory areas. Coding SNPs alter amino acid sequences, altering protein function and potentially increasing disease susceptibility.Some nsSNPs may have a neutral effect on protein function while others may have a major deleterious effect. To analyse the vulnerability of individual SNPs to diseases, it is important to distinguish deleterious SNPs from neutral SNPs.

In present study prediction of *CTHRC1* genetic variants was accomplished by utilizing sequence and structure based bioinformatics tools- PolyPhen 2, PROVEAN, I-Mutant, PhD SNP. With the help of Polyphen 2, I-Mutant tool, Provean, PhD-SNP tools, deleterious SNPs were found to be 50, 83, 38, 39 respectively. 12 SNPs which are found to be highly deleterious nsSNPs were N162D, R146Q, G67R, S142R, N190S, I172T, W214R, R235C, R156G, R235H, S219L and C149Y. Further annotation of CTHRC1 protein was performed with the help of STRING database which was used for the prediction of functional interaction between the CTHRC1 gene and other proteins in the cell. The results predicted from the STRING(Figure 2) identified the functional association of CTHRC1 protein with FZD6 (Frizzled -6 Receptor for Wnt Proteins), WNT3A (Protein Wnt-3A), ROR2 (Tyrosine-protein kinasetransmembranereceptor),DVL1(Segmentpolarityproteindisheveledhomolog DVL-1), DVL2 (Segment polarity protein disheveled homolog DVL-2), FZD5 (Frizzled-5, Receptor for Wnt proteins), DVL3 (Segment polarity protein disheveled homolog DVL-3), FZD3 (Frizzled-3, Receptor for Wnt proteins), FZD9 (Frizzled-9, Receptor for Wnt proteins) and NTN4(Netrin-4).

Since the CTHRC1 variants discovered in this study have never been published before, they must be validated to determine their significance. These findings could help researchers better understand the function of CTHRC1 SNPs in disease susceptibility through laboratory experiments.

CONCLUSION

The present study suggests that in *CTHRC1* gene, out of 106 SNPs, 12 variants found were found to be highly deleterious SNPs. As a result, these nsSNPs can be strongly considered as key candidates in causing diseases associated with CTHRC1 dysfunction, which will aid in drug discovery and development. To investigate the effects of these polymorphisms on protein structure and function, wet lab experiments are needed. The findings of this study may be useful for more scientific research in order to develop better drugs.

ACKNOWLEDGEMENT

We would like to express our sincere gratitude to all students who enthusiastically participated in this research.

REFERENCES

Adzhubei,I.,Schmidt,S.,Peshkin,L.,Ramensky,V.,Gerasimova,A.,Bork,P.,et.al. (2010). A method and server for predicting damaging missense mutations. *Nature Methods*, 7, 248-249. DOI:https://doi.org/10.1038/nmeth0410-248

Capriotti, E., Fariselli, P., & Casadio, R. (2005). I-Mutant2.0: Predicting Stability Changes Upon Mutation from the Protein Sequence or Structure. *Nucleic Acids Research*, 33(Web Server Issue), W306-10. DOI: 10.1093/nar/gki375

Choi, Y., Sims, G., Murphy, S., Miller, J., & Chan, A. (2012). Predicting the Functional Effect of Amino Acid Substitutions and Indels. *Plos One*, 1-13. DOI: https://doi.org/10.1371/journal.pone.0046688

Dabhi,B.,&Mistry,K.(2014).InSilicoAnalysisofSingleNucleotidePolymorphism (SNP) in Human TNF-α Gene. *Meta Gene.*, 2, 586-95. DOI: 10.1016/j.mgene.2014.07.005.

Jiang, N., Cui, Y., Liu, J., Zhu, X., Wu, H., Yang, Z., & Ke, Z. (2016). Multidimensional Roles of Collagen Triple Helix Repeat Containing 1 (CTHRC1) in Malignant Cancers. *Journal of Cancer*, 7(15), 2213-2220. DOI: 10.7150/jca.16539

Pyagay, P., Heroult, M., Wang, Q., Lehnert, W., Belden, J., Liaw, L., et. al. (2005). CollagenTripleHelixRepeatContaining1,aNovelSecretedProteininInjured and Diseased Arteries, Inhibits Collagen Expression and Promotes Cell Migration. *Circ Res.*, 96(2), 261-268.DOI: 10.1161/01.RES.0000154262.07264.12.

Sherry, S., Ward, M., Kholodov, M., Baker, J., Phan, L., Smigielski, E., & Sirotkin, K. (2001). dbSNP: The NCBI Database of Genetic Variation. *Nucleic Acids Research*, 29(1), 308-11. DOI: 10.1093/nar/29.1.308

Szklarczyk,D.,Franceschini,A.,Kuhn,M.,Simonovic,M.,Roth,A.,Minguez,P.,et. al. (2011). The STRING database in 2011: Functional Interaction Networksof Proteins, Globally Integrated and Scored. *Nucleic Acids Research*, 39(Database Issue), D561-8. DOI: 10.1093/nar/gkq973.

Tameda, M., Sugimoto, K., Shiraki, K., Yamamoto, N., Okamoto, R., Usui, M., et.al. (2014). Collagen triple helix repeat containing 1 is overexpressed in hepatocellular carcinoma and promotes cell proliferation and motility. *International journal of oncology*,541-548.

Wang, C., Li, Z., Shao, F., Yang, X., Feng, X., Shi, S., et. al. (2017). High expression of Collagen Triple Helix Repeat Containing 1 (CTHRC1) facilitates the progression of esophageal squamous cell carcinoma through MAPK/MEK/ERK/FRA-1 activation. *Journal of Experimental and Clinical Cancer Research*, 36-84. DOI: 10.1186/s13046-017-0555-8

Zhang,R.,Cao,Y.,Bai,L.,Zhu,C.,Li,R.,He,H.,et.al.(2015).TheCollagenTriple Helix Repeat Containing 1 Facilitates Hepatitis B Virus-Associated Hepatocellular Carcinoma Progression by Regulating Multiple Cellular Factors and Signal Cascades. *Mol Carcinog.*, 54(12), 1554-66. DOI: 10.1002/mc.22229

Zhao, M.-J., Chen, S.-Y., Qu, X.-Y., Abdul-Fattah, B., Lai, T., Xie, M., et. al. (2018). Increased Cthrc1 Activates Normal Fibroblasts and Suppresses Keloid Fibroblasts by Inhibiting TGF-β/Smad Signal Pathway and Modulating YAP Subcellular Location. *Current Medical Science*, 38(5), 894-902. DOI: 10.1007/s11596-018-1959-1

Zhu, B., Ekman, M., Svensson, D., Lindvall, J., Nilsson, B.-O., Uvelius, B., & Swärd, K. (2018). Array profiling reveals the contribution of Cthrc1 to the growth of the denervated rat urinary bladder. *American Journal of Physiology*. Renal physiology, 314(5), F893-905. DOI: https://doi.org/10.1152/ajprenal.00499.2017