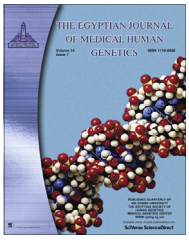
The Egyptian Journal of Medical Human Genetics (2013) **14**, 419–422

Ain Shams University

The Egyptian Journal of Medical Human Genetics

www.ejmhg.eg.net
www.sciencedirect.com



ORIGINAL ARTICLE

In Silico survey of functional coding variants in human AEG-1 gene

Malihe Naderi [a], Roghaye Gharaei [b], Ehsan Soleymani-Nejadian [c], Esmaeil Samadian [d,*]

[a] *Department of Microbiology, Qom branch, Islamic Azad University, Qom 37185-364, Iran*
[b] *Department of Clinical Biochemistry, Faculty of Medicine, Golestan University of Medical Sciences, Golestan, Iran*
[c] *Microbiology Lab, Institute of Forest Protection, Nanjing Forestry University, Longpan Road 159, 210037 Nanjing, China*
[d] *Department of Molecular Medicine, Faculty of Advanced Medical Technologies, Golestan University of Medical Sciences, Golestan, Iran*

Received 27 July 2013; accepted 20 August 2013
Available online 18 September 2013

KEYWORDS

Non-synonymous SNP;
Functional impacts;
AEG-1

Abstract *Background and aims:* Non-synonymous (ns)SNPs represent typical genetic variations that may potentially affect the structure or function of expressed proteins and therefore could have an impact on complex disorders. A computational-based (*In Silico*) analysis has been done to evaluate the phenotypic effect of nsSNPs in human Astrocyte elevated gene-1 (AEG-1), a newly identified candidate in multiple cancers.

Materials and methods: We provide a list of all nsSNPs located in human AEG-1 gene from SNP database. SIFT (Sorting Intolerant From Tolerant), PolyPhen (Polymorphism phenotyping) and FastSNP programs were used in our study.

Results: Out of 32 nsSNPs, alteration rs150644674 (A14V) was predicted to be the most deleterious by both SIFT and PolyPhen servers and nsSNP prioritization by FastSNP software showed that rs1128828 and rs11542044 missenses could have a splicing regulatory role. Besides, functionality of the substitution of rs1128828 (V187F) was predicted by all our used tools.

Conclusions: It could be concluded that these intolerant changes may lie within a functional region of the protein and may affect the stability and folding of AEG-1. These variants are reagents for further protein function and molecular epidemiology studies of cancer susceptibility.

1. Introduction

Astrocyte elevated gene-1 (AEG-1) or 3D3/lyric is newly characterized to be associated with multiple cancers, including hepatocellular carcinoma [1,2], breast carcinomas, malignant gliomas, melanomas [3,4], and neurodegeneration [5]. Also known as human Metadherin and a downstream target mole-

* Corresponding author. Tel.: +98 1714430564; fax: +98 1714430564.
E-mail address: essamadian@gmail.com (E. Samadian)).

Peer review under responsibility of The Egyptian Society of Chest Diseases and Tuberculosis.

ELSEVIER Production and hosting by Elsevier

1110-8630 © 2013 Production and hosting by Elsevier B.V. on behalf of Ain Shams University.
http://dx.doi.org/10.1016/j.ejmhg.2013.08.002

cule of Ha-ras and c-myc [6], it induces anchorage-independent growth and invasiveness of tumor cells with an increased expression of adhesion mediators by activating the NF-κB pathway [6,7] and therefore, it can be a direct regulator of angiogenesis [3]. Moreover, knockdown of *AEG-1* inhibited proliferation of human prostate cancer, neuroblastoma and melanoma cells and induced apoptosis in prostate cancer and neuroblastoma cells [8]. Accordingly, AEG-1 may be a promising target for adjuvant therapy [9].

Human Metadherin mRNA encodes a protein of 582 amino acids with a predicted molecular mass of approximately 64-kDa [2,10], and the amino acid sequences are highly conserved across vertebrates [11,12]. Genomic blast search demonstrated that the *AEG-1* gene consists of 12 exons/11 introns and is located at 8q22 [13].

Single nucleotide polymorphisms (SNPs) are known as the most common changes in the human genome, however only a small group of these variations will be considered as markers, underlying susceptibility to different common human diseases [14]. But, understanding the functions of SNPs can noticeably help to comprehend the genetics of the human phenotypic variations. Among these alterations, non-synonymous (ns)SNPs, especially missense, causing amino acid (aa) substitutions in the encoded protein sequence are greatly thought to have a deleterious effect on the structure and/or function of the related protein [15,16]. Experimental analysis is nearly hard to screen all of the functionally important nsSNPs. Predicting the deleterious nsSNPs for a candidate gene, like *AEG-1*, from the majority of benign (no observable phenotypic impact) nsSNPs has currently received much attention from researchers. Thus computational strategy, based on the biochemical severity of the amino acid substitution and protein sequence and/or structural information, is proposed for the systematic analysis of SNPs [14,17]. In this study, we predicted nsSNPs for *AEG-1* that potentially affect its protein function.

2. Materials and methods

2.1. Data mining

The nsSNPs and their related protein sequences of *AEG-1* gene were retrieved from the National Center for Biotechnology Information (NCBI) database of SNPs, dbSNP (http://www.ncbi.nlm.-nih.gov/projects/SNP) for our computational analysis.

2.2. Analysis of deleterious nsSNPs by SIFT

Sorting Intolerant From Tolerant (SIFT) is based on premise that protein evolution is correlated with protein function. SIFT is a sequence homology-based tool that sorts intolerant from tolerant amino acid substitutions and predicts whether an amino acid substitution in a protein will have a phenotypic

Table 1 SIFT and Polyphen predictions for nsSNPs in human *AEG-1* gene.

#rs ID	Allele change	AA Substitution	Substitution position	SIFT prediction	SIFT score	PolyPhen prediction	PolyPhen score
rs150644674	GCC ⇒ GTC	A [Ala] ⇒ V [Val]	14	DAMAGING	0.00	Probably damaging	0.965
rs140652237	GTG ⇒ ATG	V [Val] ⇒ M [Met]	54	TOLERATED	0.15	Possibly damaging	0.815
rs200211841	TAC ⇒ TGC	Y [Tyr] ⇒ C [Cys]	69	TOLERATED	0.19	Probably damaging	0.990
rs17854373	GCC ⇒ TCC	A [Ala] ⇒ S [Ser]	78	TOLERATED	0.47	Probably damaging	0.995
rs113646142	CCG ⇒ GCG	P [Pro] ⇒ A [Ala]	96	TOLERATED	0.39	Benign	0.000
rs113646142	CCG ⇒ TCG	P [Pro] ⇒ S [Ser]	96	TOLERATED	0.39	Benign	0.000
rs188271601	CTG ⇒ GTG	L [Leu] ⇒ V [Val]	103	TOLERATED	0.31	Benign	0.000
rs144514874	AAG ⇒ AAT	K [Lys] ⇒ N [Asn]	116	TOLERATED	0.25	Possibly damaging	0.939
rs200267294	GAA ⇒ CAA	E [Glu] ⇒ Q [Gln]	134	TOLERATED	0.52	Probably damaging	0.965
rs147831600	GGT ⇒ GCT	G [Gly] ⇒ A [Ala]	138	TOLERATED	0.41	Benign	0.002
rs200266145	AGT ⇒ CGT	S [Ser] ⇒ R [Arg]	179	TOLERATED	0.27	Possibly damaging	0.902
rs1128828	GTT ⇒ TTT	V [Val] ⇒ F [Phe]	187	DAMAGING	0.05	Probably damaging	0.990
rs11998518	AGA ⇒ AAA	R [Arg] ⇒ K [Lys]	199	TOLERATED	1.00	Benign	0.319
rs180867681	CGT ⇒ TGT	R [Arg] ⇒ C [Cys]	205	DAMAGING	0.04	Probably damaging	1.000
rs139649151	GAT ⇒ AAT	D [Asp] ⇒ N [Asn]	213	TOLERATED	0.39	Probably damaging	1.000
rs200350363	CCT ⇒ CGT	P [Pro] ⇒ R [Arg]	222	TOLERATED	0.40	Possibly damaging	0.741
rs11542044	TCC ⇒ CCC	S [Ser] ⇒ P [Pro]	294	TOLERATED	0.29	Benign	0.006
rs150963816	GTT ⇒ TTT	V [Val] ⇒ F [Phe]	307	DAMAGING	0.05	Possibly damaging	0.908
rs140753043	AAG ⇒ GAG	K [Lys] ⇒ E [Glu]	314	TOLERATED	0.67	Possibly damaging	0.882
rs17854374	ACT ⇒ GCT	T [Thr] ⇒ A [Ala]	317	TOLERATED	0.74	Benign	0.000
rs146663706	CGT ⇒ CTT	R [Arg] ⇒ L [Leu]	343	TOLERATED	0.25	Possibly damaging	0.928
rs199900068	TCT ⇒ CCT	S [Ser] ⇒ P [Pro]	351	TOLERATED	0.29	Probably damaging	0.998
rs182729161	CGT ⇒ TGT	R [Arg] ⇒ C [Cys]	370	DAMAGING	0.05	Probably damaging	1.000
rs112966052	GAG ⇒ GGG	E [Glu] ⇒ G [Gly]	400	TOLERATED	0.31	Probably damaging	0.999
rs145881524	CTT ⇒ CCT	L [Leu] ⇒ P [Pro]	437	TOLERATED	0.28	Possibly damaging	0.918
rs138458185	CCA ⇒ CTA	P [Pro] ⇒ L [Leu]	438	TOLERATED	0.29	Benign	0.069
rs200352570	AAT ⇒ ATT	N [Asn] ⇒ I [Ile]	476	TOLERATED	0.15	Probably damaging	0.958
rs143317071	CGT ⇒ CAT	R [Arg] ⇒ H [His]	481	TOLERATED	0.15	Benign	0.025
rs199958820	AGC ⇒ AAC	S [Ser] ⇒ N [Asn]	506	TOLERATED	0.35	Probably damaging	0.989
rs202034424	GTA ⇒ ATA	V [Val] ⇒ I [Ile]	522	TOLERATED	0.19	Benign	0.005
rs201579152	TCT ⇒ TTT	S [Ser] ⇒ F [Phe]	533	TOLERATED	0.07	Possibly damaging	0.955
rs141463674	AAT ⇒ GAT	N [Asn] ⇒ D [Asp]	549	TOLERATED	0.57	Benign	0.003

effect [18]. We used SIFT to detect the deleterious coding nsSNPs and submitted the query in the form of either SNP identifier (ID)s or as protein sequences. SIFT analysis was performed by allowing the algorithm to search for homologous sequences and using the default settings (SWISS-PROT 45 and TrEMBL 28 databases, median conservation score 3.00, remove sequences >90% identical to query sequence). The underlying principle of this tool is that it generates alignments with a large number of homologous sequences and assigns scores to each residue, ranging from 0 to 1. Scores close to 0 indicate evolutionary conservation and intolerance to substitution, while scores close to 1 indicate tolerance to substitution. SIFT scores <0.05 are predicted by the algorithm to be intolerant or deleterious amino acid substitutions, whereas scores >0.05 are considered tolerant. The higher the tolerance index of a particular amino acid substitution, the lesser is its probable impact.

2.3. Simulation for functional changes by PolyPhen

Polymorphism phenotyping (PolyPhen) is an automatic tool for prediction of the possible impact of an amino acid substitution on the structure and function of a human protein [8]. This prediction is based on straightforward empirical rules that are applied to the sequence, phylogenetic and structural information identifying the substitution. Input options for the PolyPhen server are protein sequence, SWALL database ID or accession number together with sequence position with two amino acid alterations. We submitted the query in the form of protein sequence with mutational position and two amino acid variants. PolyPhen searches for 3D protein structures, multiple alignments of homologous sequences and amino acid contact information in several protein structure databases. Then, it calculates position-specific independent count (PSIC) scores for each of the two variants and computes the difference of the PSIC scores of the two variants. The higher a PSIC score difference, the higher functional impact a particular amino acid substitution is likely to have. PolyPhen scores were designated as ''probably damaging'' (0.95–1), ''possibly damaging'' (0.7–0.95), and ''benign'' (0.00–0.31).

2.4. The nsSNP prioritization by FastSNP software

We also used the FastSNP software (http://fastsnp.ibms.sinica.edu.tw) to know if the amino acid changing in a SNP could affect the functionality of the protein. FastSNP uses empirically derived rules to predict that an nsSNP is supposed to strongly affect protein function or structure or to influence protein function or structure lacking any phenotypic effect or the lack of data do not allow this software to make a prediction [19].

3. Results

The human *AEG-1* gene retrieved from dbSNP database contained a total of 32 missense SNPs and their corresponding information including allele changes, amino acid substitutations and substitutation positions are listed in Table 1.

The protein sequences of all 32 nsSNPs were submitted separately to the SIFT program to check its tolerance index. Out of 32 nsSNPs, 5 (15.6%), i.e., rs150644674, rs1128828, rs180867681, rs150963816 and rs182729161 were calculated to be deleterious with a tolerance index score of ⩽0.05, as shown in Table 1. These variants could affect the protein function in the *AEG-1* gene. Remarkably, only substitution A14V (rs150644674) showed a highly damaging impact with an absolute tolerance index score of 0.00. And, the remaining 27 (84.3%) nsSNPs were analyzed to be tolerant in AEG-1.

In PolyPhen analysis, protein sequence with mutational position and amino acid variants related to all the retrieved nsSNPs were submitted as inputs to PolyPhen serve. This algorithm predicted 11 (34.3%) of these variants to be probably damaging, 9 (28.1%) to be possibly damaging, and 12 (36.3%) to be benign substitutions.

It is to be noted that 5 nsSNPs that were observed to be deleterious by SIFT program were also predicted to be highly damaging by PolyPhen, i.e., A14V (rs150644674), V187F (rs1128828), R205C (rs180867681), V307F (rs150963816) and R370C (rs182729161). Furthermore, the highly conserved and deleterious alteration (A14V) was shown to have the higher PSIC value (0.965).

The nsSNP prioritizations by FastSNP program showed 5 missense mutations for *AEG-1* through the interface of Gene Name that are summarized in Table 2. This software scored the higher risks, i.e., 3–4 (Moderate–High) for the changes rs1128828 and rs11542044 that predicted to be non-conservative.

With the exception of rs17854374, the rest of the 4 alterations (rs1128828, rs11542044, rs17854373 and rs11998518) seemed to be able to regulate alternative splicing in *AEG-1* coding sequence. All these potential changes do not alter the protein's functional domains yet. Interestingly, only V187F (rs1128828) that has a splicing regulatory influence was also found to be damaging by two other predictors.

4. Discussion

Several studies in more recent years revealed the multi-functional contribution of AEG-1 in biological processes, though, the three-dimensional structure of AEG-1 has still not been solved, and the functional domains of the protein are not clearly determined [12]. But, sequence analysis of the protein recently showed the presence of putative nuclear localization signals (NLS) between 79 and 91 amino acids (a.a.), 432 and 451 a.a. and 561 and 580 a.a., a transmembrane domain between 51 and 72 a.a., and a lung homing domain between 381 and 443 a.a. residues. Additionally, there is also an N-terminal ''LXXLL'' motif that is enrolled by transcriptional co-activators to interact with transcription factors. The N-terminal 71 amino acids of the protein composed of the transmembrane domain and the 'LXXLL' motif are critical in mediating AEG-1-induced invasion, soft agar growth and NF-κB activation. However, the p65-interaction domain has also been identified in AEG-1 that corresponds to 101–205 a.a. of the protein. The 'LXXLL' motif of AEG-1 might be involved in its interaction with CBP which provides an important relationship of the AEG-1/NF-κB complex to the basal transcriptional machinery [6]. Our *in Silico* analysis of the human *AEG-1* missense mutations demonstrated the functional significance of changes rs150644674, rs1128828, rs180867681, rs150963816 and rs182729161 as predicted to be structurally deleterious and alternative splicing regulatory roles for rs1128828 and

Table 2 FastSNP prioritization results.

#rs IDs	Possible functional effects	Risk	RefSeq mRNA
rs1128828	Missense (non conservative); splicing regulation	Medium–High (3–4)	ENST00000336273
rs11542044	Missense (non-conservative); splicing regulation	Medium–High (3–4)	ENST00000336273
rs17854373	Missense (conservative); splicing regulation	Low–Medium (2–3)	ENST00000336273
rs11998518	Missense (conservative); splicing regulation	Low–Medium (2–3)	ENST00000336273
rs17854374	Missense (conservative)	Low–Medium (2–3)	ENST00000336273

rs11542044. Given that, it may suggest the existence of the functionality of the regions around these positions, especially for substitutions rs1128828 (V187F) and rs150644674 (A14V), has not been established. Due to the different rules for predicting the effect of alterations in these algorithms, the outcomes were found in some ways, dissimilar. However, the variations that overlap the predictions should provide greatest reliability to consider similarly. There is no direct approach of evaluating the accuracy of these predictions made by SIFT and PolyPhen, as it is possible that the algorithms used different data sets. So, even if the predictions by these programs were not completely consistent for this subset of mutants, these *AEG-1* variants still should be regarded as candidates for SNP screening. Furthermore, molecular modeling and experimental investigations can be used to confirm our results and the possible regulatory functions of rs1128828 and rs150644674 will be more understood.

Acknowledgments

Our project has got an approval from the Golestan University of Medical Sciences, Golestan, Iran. We appreciate our colleagues in cancer research worldwide and also, our friends Mr. Mohammad Reza Ashrafi and Syed Abolghasem Ghadami at the Razi University, Kermanshah, Iran, for their useful comments.

Referencess

[1] Yoo BK et al. Molecular mechanism of chemoresistance by astrocyte elevated gene-1. Cancer Res 2010;70(8):3249–58.
[2] Yoo BK et al. Astrocyte elevated gene-1 regulates hepatocellular carcinoma development and progression. J Clin Invest 2009; 119(3):465–77.
[3] Emdad L et al. Astrocyte elevated gene-1 (AEG-1) functions as an oncogene and regulates angiogenesis. Proc Natl Acad Sci U S A 2009;106(50):21300–5.
[4] Yu C et al. Overexpression of astrocyte elevated gene-1 (AEG-1) is associated with esophageal squamous cell carcinoma (ESCC) progression and pathogenesis. Carcinogenesis 2009;30(5):894–901.
[5] Yoo BK et al. Astrocyte elevated gene-1 (AEG-1): A multifunctional regulator of normal and abnormal physiology. Pharmacol Ther 2011;130(1):1–8.
[6] Lee SG et al. Astrocyte elevated gene-1 (AEG-1) is a target gene of oncogenic Ha-ras requiring phosphatidylinositol 3-kinase and c-Myc. Proc Natl Acad Sci U S A 2006;103(46):17390–5.
[7] Emdad L et al. Activation of the nuclear factor kappaB pathway by astrocyte elevated gene-1: implications for tumor progression and metastasis. Cancer Res 2006;66(3):1509–16.
[8] Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. Nucleic Acids Res 2002;30(17):3894–900.
[9] Liu H et al. Knockdown of astrocyte elevated gene-1 inhibits proliferation and enhancing chemo-sensitivity to cisplatin or doxorubicin in neuroblastoma cells. J Exp Clin Cancer Res 2009;28:19.
[10] Li J et al. Astrocyte elevated gene-1 is a novel prognostic marker for breast cancer progression and overall patient survival. Clin Cancer Res 2008;14(11):3319–26.
[11] Sutherland HG et al. 3D3/lyric: a novel transmembrane protein of the endoplasmic reticulum and nuclear envelope, which is also present in the nucleolus. Exp Cell Res 2004;294(1):94–105.
[12] Ying Z, Li J, Li M. Astrocyte elevated gene 1: biological functions and molecular mechanism in cancer and beyond. Cell Biosci 2011;1(1):36.
[13] Wang C, Yang Q. Astrocyte elevated gene-1 and breast cancer (Review). Oncol Lett 2011;2(3):399–405.
[14] Wang LL, Li Y, Zhou SF. Prediction of deleterious non-synonymous single nucleotide polymorphisms of genes related to ethanol-induced toxicity. Toxicol Lett 2009;187(2):99–114.
[15] Bao L, Cui Y. Functional impacts of non-synonymous single nucleotide polymorphisms: selective constraint and structural environments. FEBS Lett 2006;580(5):1231–4.
[16] Chasman D, Adams RM. Predicting the functional consequences of non-synonymous single nucleotide polymorphisms: structure-based assessment of amino acid variation. J Mol Biol 2001; 307(2):683–706.
[17] Jordan DM, Ramensky VE, Sunyaev SR. Human allelic variation: perspective from protein function, structure, and evolution. Curr Opin Struct Biol 2010;20(3):342–50.
[18] Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res 2003;31(13): 3812–4.
[19] Yuan HY et al. FASTSNP: an always up-to-date and extendable service for SNP function analysis and prioritization. Nucleic Acids Res 2006;34(Web Server issue):W635–41.