BIOMEDICAL INFORMATICS

www.bioinformation.net
Volume 17(12)

Research Article

BIOINFORMATION
Discovery at the interface of physical and biological sciences

OPEN ACCESS GOLD

Aberrations in SMAD family of genes among HNSCC patients

E. Thariny, A.S. Smiline Girija, A. Paramasivam & J. Vijayashree Priyadharsini

Saveetha Dental College, Saveetha Institute of Medical and Technical Sciences (SIMATS), Saveetha University, Chennai, India; Corresponding author Dr. J. Vijayashree Priyadharsini, E-mail: vijayashreej.sdc@saveetha.com.
Communicated by T. Lakshmi - E-mail: lakshmi@saveetha.com

Received June 17, 2021; Revised September 27, 2021; Accepted September 27, 2021, Published December 31, 2021

DOI: 10.6026/973206300171113

Declaration on Publication Ethics:
The author's state that they adhere with COPE guidelines on publishing ethics as described elsewhere at https://publicationethics.org/. The authors also undertake that they are not associated with any other third party (governmental or non-governmental agencies) linking with any form of unethical issues connecting to this publication. The authors also declare that they are not withholding any information that is misleading to the publisher in regard to this article.

Author responsibility:
The authors are responsible for the content of this article. The editorial and the publisher have taken reasonable steps to check the content of the article in accordance to publishing ethics with adequate peer reviews deposited at PUBLONS.

Declaration on official E-mail:
The corresponding author declares that official e-mail from their institution is not available for all authors

This article is part of a special issue on Dental Biology

Abstract:
Head and neck cancer is a debilitating disease with several etiological factors. One of the main etiologies to be noticed is the alteration, which is either caused by genetic or environmental factors. Therefore, it is of interest to assess the effect of genetic alterations, especially the non-synonymous mutations of the *SMAD* gene family and its possible association with HNSCC. Data shows a significant novel mutation in the *SMAD* gene family in association with head and neck squamous cell carcinoma (HNSCC), which would aid in better diagnosis and treatment planning for cancer.

Keywords: Cancer, mutation, SMAD gene, genetic alteration, head and neck, pathogenicity

Background:
Head and neck cancer is a morbid lethal malignancy. It is a cancer-causing growth that is present in the mouth, nose, throat, larynx, sinuses, or salivary glands **[1]**. Squamous cell carcinomas are more prevalent in human cancer with few therapeutic options such as chemotherapy and radiation. It is the second most common form of cancer in India among males **[2]**. Head and neck cancer is noted for more than 650000 cases and 330000 deaths annually which are recorded all over the world. In the United States, about 3% of cancers are head and neck cancer **[3]**. Overall the occurrence of head and neck cancer accounts mainly in oral cancer, which is 11.4 per 100000 persons per year worldwide. Members of the SMAD family include eight different SMAD in mammalian cell populations, which are classified into three categories: the first category includes the Receptor-regulated or regulatory *SMAD* gene (R-SMAD), which includes Smaad1, SMAD2, SMAD3, SMAD5, SMAD6, and SMAD7. The second category includes common SMAD (Co-SMAD), i.e SMAD 4.Third category includes inhibitory or anti-SMAD (I-SMAD), which areSMAD6 and SMAD7 **[4]**. The function of the SMAD gene is to deliver extracellular signals from TGF-beta ligands to the nucleus leading the activation of downstream gene transcription to regulate cell growth and division process. Ligand-induced activation of TGF- β family receptors with intrinsic serine/threonine kinase activity triggers the phosphorylation of receptor-regulated SMADs. The *SMAD* genes are expressed according to the categories, regulatory SMAD and common SMAD are located in the cytoplasm, but it gets accumulated once the TGF- β signaling is passed, where they can bind to DNA and regulate transcription. Inhibitory SMAD is predominantly found in the nucleus, where they can act as a direct transcriptional regulator **[5]**. The main function of the *SMAD* gene is to regulate the growth progression any change in the signaling pathway can mislead the outcome of the function. Alterations in the *SMAD4* gene were found to be most often associated with head and neck cancer among the *SMAD* gene family **[6]**. Therefore, it is of

interest to assess the effect of genetic alterations, especially the non-synonymous mutations of the *SMAD* gene family and its possible association with HNSCC.

Materials and method:

Data source:

The source of the patient's data was obtained from the cBioportal database **[7,8]**. This database contains an exhaustive collection of HNSCC case details from different cohorts. The TCGA, Firehose legacy data set constituted a total of 528 head and neck squamous cell carcinoma cases in which sequencing and copy number alteration data were presented for 512 tumor samples. A complete profile of mutated, amplified, deleted genes was in stock for each and every case in the dataset. The demographic details of the cases have been provided in **(Table 1)**. A complete list of essential genes related to the SMAD gene family was derived from the "HUGO Gene Nomenclature Committee at the European Bioinformatics Institute "(www.genenames.org/data/) database. User-defined queries based on these genes were submitted to the cBioportal database and the resultant Oncoprint data was used for further analysis.

Oncoprint data analysis:

The Oncoprint data provides information on the frequency distribution of variations in each of the genes selected, type of variation, changes in the protein-coding amino acids, gene amplification, deletions, insertions, frameshifts, splice site mutations, etc. These details can be used to detect (a) derive a putative association between the disease phenotype and genotype, (b) identify the variations in less understood pathways or genes, and (c) identify any novel variations which can be associated with the disease phenotype **(Table 2) (Figure 1)**.

Protein stability analysis:

I-Mutant v3.0 is a support vector machine (SVM)-based tool for the automatic prediction of protein stability changes upon single point mutations. The software's predictions are based on the protein sequence. The free energy change (DDG) predicted by I-Mutant 3.0 is based on the difference between unfolding Gibbs free energy change of mutant and native protein (kcal/mol) **[9] (Table 3).**

PROVEAN analysis:

Protein variation effect analyzer was used to predict whether the single nucleotide substitutions (non-synonymous variants) in the protein affect the protein function **[10,11] (Table 3).**

gnomAD analysis:

gnomAD v2.1.1 dataset consists of a collection of 125,748 exomes and 15,708 genomes from human sequencing studies. This data was used to verify whether the missense variants found in the HNSCC data is prevalent in other individuals for whom the sequencing of gene data is available **[12]**.

UALCAN analysis:

UALCAN is a comprehensive, user-friendly, and interactive web resource for analyzing cancer OMICS data. The gene expression profile of genes of the SMAD family and the survival curve analysis for the same was performed with the TCGA dataset in UALCAN database **[13]**.

Table 1: Demographic details of patients analyzed in the present study (as obtained from the cBioportal site)

Gender	Male (n = 386) Female (n = 142)
Mutation count	6-3181
Diagnosis of age	19-90 years
Smoking status	Smokers: 515 Data not available: 12 Unknown: 1
Alcohol history	Yes – 352 No – 165 Data not available: 11
Neoplasm Histologic grade	Grade 1: 63 Grade 2: 311 Grade 3: 125 Grade 4: 7 Grade GX: 18 Data not available: 4
Race category	White: 452 African: 48 Asian: 11 American Indian or Alaska native: 2 Data not available: 15

* Neoplasm histologic grade available for 452 patients in the TCGA provisional dataset

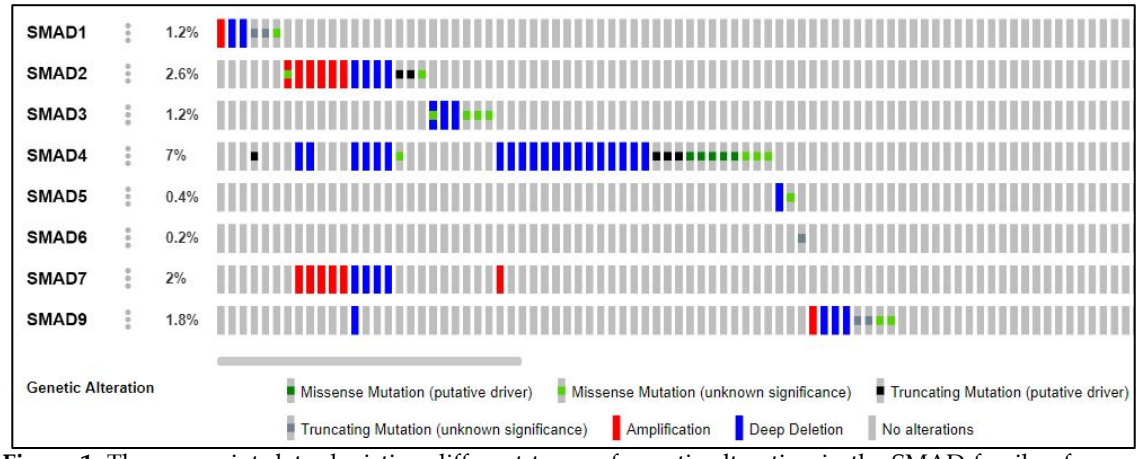


Figure 1: The oncoprint data depicting different types of genetic alteration in the SMAD family of genes. The *SMAD4* was found to

have the highest alteration level of all genes examined (7%). *SMAD2* and *SMAD7* genes were found to have numerous sites of amplification.

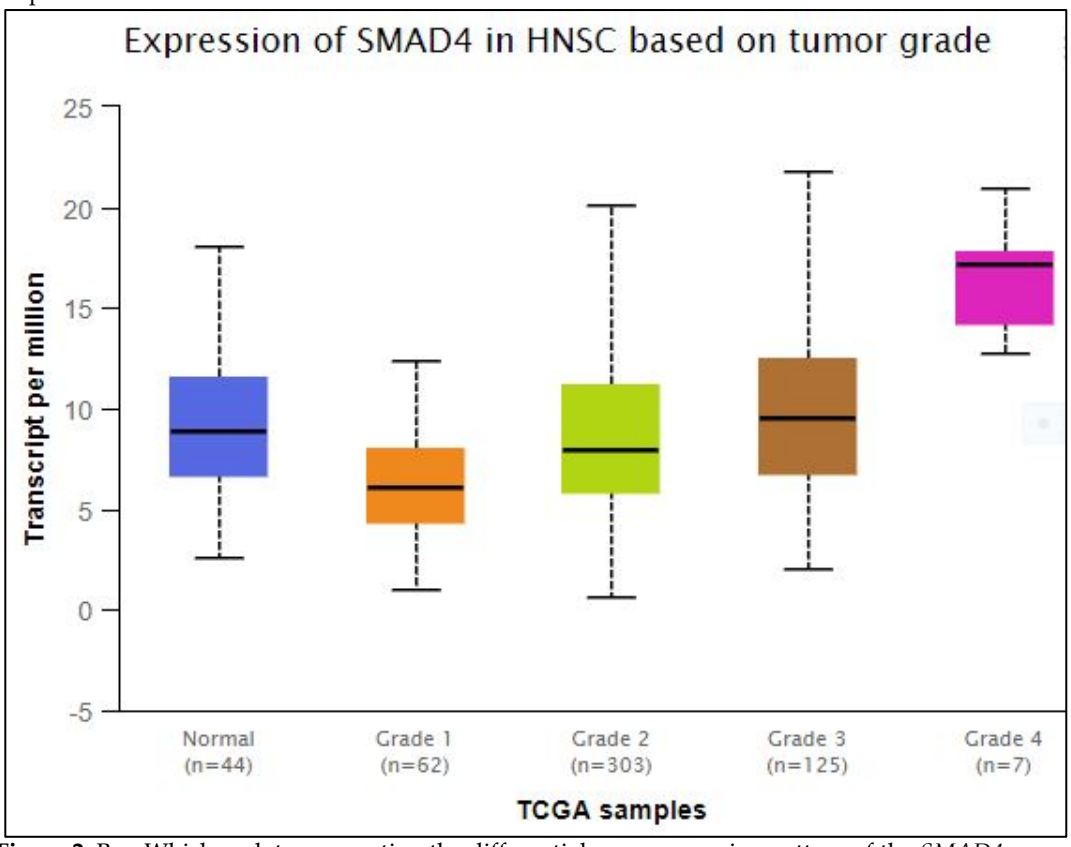


Figure 2: Box-Whisker plot representing the differential gene expression pattern of the *SMAD4* gene across different tumor grades. The X-axis represents the different grades of HNSCC example samples from the TCGA data set and the Y axis denotes the transcripts per million values. A significant difference in the gene expression profile was observed between normal vs grade 1 ($p = 1.49 \times 10^{-4}$), normal vs grade 3 ($p = 9.5 \times 10^{-12}$), normal vs grade 4 (1.11×10^{-6}), grade 1 vs grade 2 ($p = 2.166 \times 10^{-4}$), grade 1 vs grade 3 ($p = 2.90 \times 10^{-7}$), grade 1 vs grade 4 ($p = 4.39 \times 10^{-8}$), grade 2 vs grade 3 ($p = 7.44 \times 10^{-3}$), grade 2 vs grade 4 ($p = 5.77 \times 10^{-4}$) and grade 3 vs grade 4 ($p = 1.98 \times 10^{-2}$). A p-value of less than 0.05 is considered to be significant.

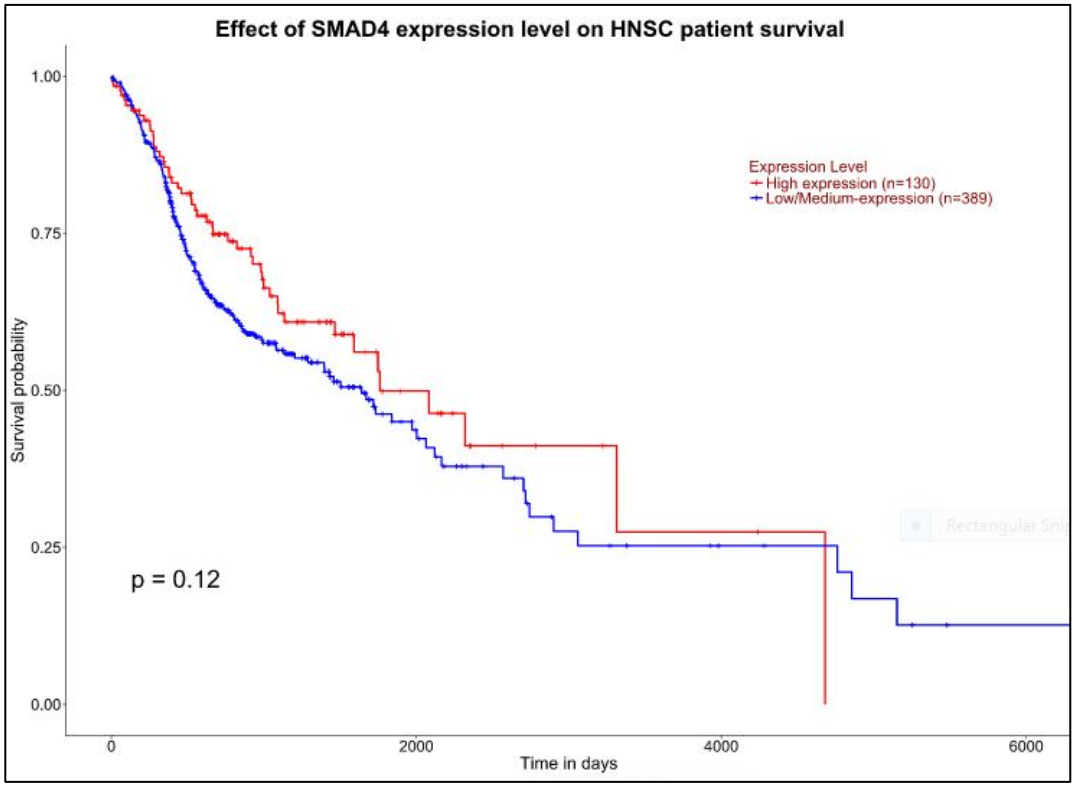


Figure 3: Kalplan-Meier plot showing the association of altered SMAD expression with HNSCC patients' survival. The X-axis represents time in days and Y-axis represents survival probability in HNSCC patients. The red line corresponds to high-level

expression and the blue line corresponds to low/ medium level expression of SMAD (0.12). A p-value of more than 0.05 is not considered to be significant.

Table 2: The frequency of genetic alterations, cytogenetic location of the gene, protein encoded by genes, variant allele frequency in tumor sample and population data as obtained from gnomAD for the *SMAD* gene family.

Gene	Protein	Alteration	Loci	Percentage of alteration
SMAD1	Small Mothers Against Decapentaplegic 1	Amplification Deep Deletion Truncating mutation (US) X259_splice W29* Missense mutation L139V	4q31.21	1.2
SMAD2	Small Mothers Against Decapentaplegic 2	Amplification Deep Deletion V398A M327I R182* Q455*	18q21.1	2.6
SMAD3	Small Mothers Against Decapentaplegic 3	Deep deletion R420H S416F S264Y C64S	15q22.33	1.2
SMAD4	Small Mothers Against Decapentaplegic 4	Deep deletion R361H A118V W99C Q366K R97C S368C P544L P298S A66F Q248* S242* Q461* Q450*	18q21.2	7
SMAD5	Small Mothers Against Decapentaplegic 5	Deep Deletion P456L	5q31.1	0.4
SMAD6	Small Mothers Against Decapentaplegic 6	Truncating mutation (US) D318=	15q22.31	0.2
SMAD7	Small Mothers Against Decapentaplegic 7	Amplification Deep Deletion	18q21.1	2
SMAD9	Small Mothers Against Decapentaplegic 9	Amplification Deep Deletion P185S R20G R284* F385Sfs*24	13q13.3	1.8

Table 3: Protein stability and pathogenesis of variants identified in *SMAD* family of genes

Gene	Alteration	I-Mutant Score	Prediction	PROVEAN Score	Prediction
SMAD1	L139V	- 0.81	Decrease stability	- 2.648	Deleterious
SMAD2	V398A	-1.05	Decrease stability	- 3.765	Deleterious

	M327I	- 2.01	Decrease stability	- 1.417	Neutral
SMAD3	R420H	- 1.42	Decrease stability	- 0.549	Deleterious
	S416F	- 0.69	Decrease stability	- 5.139	Deleterious
	S264Y	- 0.73	Decrease stability	- 5.697	Deleterious
	C64S	- 0.42	Decrease stability	- 9.383	Deleterious
SMAD4	R361H	- 1.02	Decrease stability	- 4.887	Deleterious
	A118V	- 0.76	Decrease stability	- 3.617	Deleterious
	W99C	- 1.67	Decrease stability	- 11.75	Deleterious
	Q366K	- 0.56	Decrease stability	- 2.257	Neutral
	R97C	- 1.61	Decrease stability	- 7.133	Deleterious
	S368C	- 1.63	Decrease stability	- 4.553	Deleterious
	P544L	- 0.84	Decrease stability	- 7.605	Deleterious
	P298S	- 1.03	Decrease stability	- 1.988	Neutral
	A66F	- 0.50	Decrease stability	- 3.719	Deleterious
SMAD5	P456L	- 1.27	Decrease stability	- 9.217	Deleterious
SMAD9	P185S	- 1.54	Decrease stability	0.139	Neutral
	R20G	- 1.14	Decrease stability	- 4.126	Deleterious

Result and Discussion:

Oncoprint analysis

Oncoprint analysis revealed the presence of deep deletions, amplifications, truncated, synonymous and non-synonymous variants. A few of them were found to be putative drivers. *SMAD4* was found to harbor the highest frequency of alterations among all the genes analyzed (7%). Most of the alterations in *SMAD4* were of deep deletion and missense type **(Figure 1).** Further, the variants observed in the present study were compared to the non-synonymous variants in the gnomAD database to identify whether the variant is novel or reported in the general population. The comparative analysis identified a few reported SNPs such as *rs762012589*in *SMAD3* and *rs553369182* in *SMAD9* genes. In addition, several putative drivers precipitated by loss of function of genes such as Q455*, R182* of *SMAD2* gene and Q248*, S242*, Q461* and Q450* of *SMAD4* gene were also identified. Apart from these alterations R361H has been predicted to be oncogenic with a significant loss of function, while, A118V, W99C, Q366k, and R97C were statistically significant hotspots predicted to be oncogenic **(Table 2).**

Protein and pathogenicity analysis

The stability of proteins harboring variants as assessed by I-Mutant showed decrease in stability upon substitution with the nonsynonymous variant. PROVEAN analysis predicted M327I of *SMAD2*, Q366K and P298S of *SMAD4* and P185S of *SMAD9* to be neutral, whilst all other variants were found to be deleterious. W99C of *SMAD4* gene was observed to show the lowest free energy value of -1.67 and lowest score of -11.75 depicting highly deleterious consequences among all the other variants **(Table 3).**

Gene expression and survival curve analysis

Since *SMAD4* gene harbored the highest frequency of gene alteration especially deep deletions, further gene expression analysis was warranted to assess the effect of deletions. Differential expression of *SMAD4* was found in different grades of the tumor as assessed using the TCGA data set in the UALCAN platform (**Figure 2).** Subsequent survival curves analysis although insignificant (p-value = 0.12) revealed that a higher level of *SMAD4* expression provided a better survival advantage to the HNSCC patients when compared to those with low-level expression **(Figure 3).**

Varying rates of SMAD gene mutation has been detected worldwide in relation to head and neck cancer **[14]**. Defect in *SMAD*signaling can result in TGF-β resistance, leading to dysregulation of cell growth. This dysregulation stems into different forms cancers including pancreatic, colon, breast, and lung, oral and prostate cancer. *SMAD4was* first termed deleted in pancreatic cancer locus 4 (DPC4) **[15–17]**. Transforming growth factor β (TGF-B) is a vital component regulating the epithelial cell proliferation, cell division, immune function, and angiogenesis. Since, TGF-β signaling maintains epithelial homeostasis, any dysfunction in TGF-β signaling pathway can lead to malignancies **[18]**. Any abnormal alteration in *SMAD* gene causes defect in TGF-beta pathway, which results in hyper proliferation, reduced apoptosis, and increased genomic instability **[19]**. In order to compensate this, there is an abnormal increase in TGF-beta production by tumor epithelial cells, which further promotes tumor growth, and metastasis by increasing angiogenesis and inflammation in tumor stromal cells. It is noted that overexpression of *SMAD7* causes oral epithelial dysplasia **[20].**

The loss of *SMAD4* gene protein causes a high degree of instability in tumor epithelium. So, alteration in the *SMAD* gene leads to HNSCC tumorigenesis by blocking the growth prognosis and programmed cell death (apoptosis) which is usually controlled by TGF-β signaling pathway **[21]**. The oncoprint data analysis revealed deep deletions observed in the *SMAD4* gene. It is noted that *SMAD* gene also has a role in epithelial-mesenchymal transition in which TGF-β functions as a transcriptional repressor of E-Cadherins, which are activated by the *SMAD4* gene. Germline mutation of *SMAD4* causes juvenile polyposis syndrome **[22]**. Recent studies have found that frequent deletion at loci 18q where the *SMAD4* gene protein is the present and heterozygous loss of *SMAD4* causes cancer **[23].** It is found that nearly 61.12% of oral squamous cell carcinoma is related to the loss of *SMAD4* protein **[24]**. *SMAD1* alteration includes melorheostosis and osteopoikilosis **[25].** This disease condition is associated with protein metabolism and the Th1 differentiation pathway **[26]**. It was analyzed that *SMAD2* over expression of tumor-derived missense mutation was found to promote TGF-beta mediated invasion of MDCK (Madin-Darby canine kidney) cells **[27].** The MH1 domain was frequently mutated in *SMAD2* and *SMAD4* alteration in pancreatic cancer. *SMAD 7* expression was evident in oral dysplasia and also this specific gene acts a

key negative regulator of the TGF-beta signaling pathway **[28].** Interestingly, it has been demonstrated that one of the red complex pathogen *Porphyromonas gingivalis* is capable of promoting the progression of esophageal squamous cell cancer through TGF-betadependent Smad/YAP/TAZ signaling **[29].** Limitations such as (a) the population or the data set is representative of a predominant group of individuals from a specific location which might not represent the cases observed throughout the world, (b) the habits of individuals differ in different geographical locations which may affect signaling pathways other than SMAD to precipitate the disease, (c) the identified variants have to be screened in other populations so as to arrive at conclusive evidence on the role of SMAD proteins in the pathogenesis of HNSCC should be noted.

Acknowledgement:
The authors thank all the consorts and groups who were involved in the compilation of data from patients for public use. We also thank all patients who provided consent for sharing data for research use.

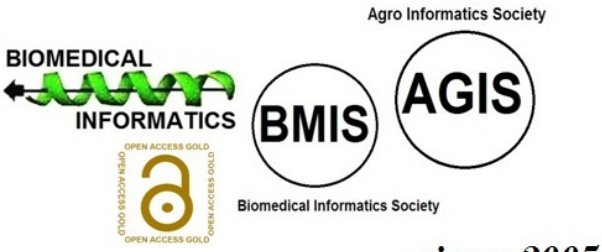
Author contributions:
All authors have contributed towards data analysis, drafting and critically revising the paper, gave final approval of the version to be published, and agree to be accountable for all aspects of the work.

Conflict of Interest: The authors declare no competing interests.

References:

[1] Divaris K *et al.* Cancer Causes Control 2010 21:567. [PMID: 20049634]
[2] Sung H *et al.* CA Cancer J Clin. 2021 71:209. [PMID: 33538338]
[3] Stransky N *et al.* Science 2011 333:1157. [PMID: 21798893]
[4] Troncone E *et al.* Int J Mol Sci. 2021 22:3922. [PMID: 33920230]
[5] Hata A *et al. Mol Med Today.* 1998 4:257. [PMID: 9679244]
[6] Xie W *et al. Oncol Res.* 2003 14:61. [PMID: 14649540]
[7] Cerami E *et al. Cancer Discov*. 2012 2:401. [PMID: 22588877]
[8] Gao J *et al. Sci Signal.* 2013 6:l1. [PMID: 23550210]
[9] Capriotti E *et al. Nucleic Acids Res.* 2005 33:W306. [PMID: 15980478]
[10] Choi Y *et al.* PLoS One 2012 7:e46688. [PMID: 23056405]
[11] Choi Y Chan AP. *Bioinformatics*. 2015 31:2745. [PMID: 25851949]
[12] Karczewski KJ *et al. Nature.* 2020 581:434.
[13] Chandrashekar DS *et al. Neoplasia* 2017 19:649. [PMID: 28732212]
[14] Samanta D *et al. Front Biosci (Landmark Ed)*. 2012 17:1281. [PMID: 22201803]
[15] Mehrvarz Sarshekeh A *et al. PLoS One* 2017 12:e0173345. [PMID: 28267766]
[16] Neapolitan R *et al. BMC Cancer* 2015 15:516. [PMID: 26169172]
[17] Itoh S *et al. European Journal of Biochemistry* 2000 267:6954.[PMID: 11106403]
[18] White RA *et al. Oncogene* 2010 29:5437. [PMID: 20676130]
[19] Qiu W *et al. Cancer Letters* 2007 245:163. [PMID: 16478646]
[20] Chen Y-K *et al. Clin Oral Investig* 2013 17:921. [PMID: 22669485]
[21] Lin L-H *et al. Frontiers in Oncology* 2019 9. [PMID: 31867281]
[22] Malkoski SP *et al. FEBS Lett* 2012 586:1984. [PMID: 22321641]
[23] Muro-Cacho CA *et al. Clin Cancer Res* 2001 7:1618. [PMID: 11410498]
[24] Han G *et al. Cell Biosci* 2011 1:41. [PMID: 22204491]
[25] Wordsworth P *et al. Calcif Tissue Int* 2019 104:530. [PMID: 30989250]
[26] Korc M. *J Clin Invest* 2009 119:3208. [PMID: 19841540]
[27] Mangone FRR *et al. Mol Cancer* 2010 9:106. [PMID: 20462450]
[28] Danielsson K *et al. J Oral Pathol Med* 2010 39:639. [PMID: 20618616]
[29] Qi YJ *et al. PLoS Biol.* 2020 18:e3000825. [PMID: 32886690]
[30] Bornstein S *et al. J Clin Invest* 2009 119:3408. [PMID: 19841536]

Edited by P Kangueane
Citation: Thariny *et al.* Bioinformation 17(12): 1113-1119 (2021)


Agro Informatics Society
BIOMEDICAL
INFORMATICS
BMIS
AGIS
Biomedical Informatics Society
since 2005
BIOINFORMATION
Discovery at the interface of physical and biological sciences

indexed in


PMC
PubMed
EBSCO
INDEXED IN
EMERGING SOURCES CITATION
(Web of Science)
CLARIVATE ANALYTICS
WEB OF SCIENCE
Web of Science Group
Crossref
doi
ResearchGate
publons