

Expression, Interaction, and Role of Pseudogene Adh6-ps1 in Cancer Phenotypes

Martin C Nwadiugwu 

Leonard Davis School of Gerontology, University of Southern California, Los Angeles, CA, USA.

Bioinformatics and Biology Insights
Volume 15: 1–6
© The Author(s) 2021
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/11779322211040591



ABSTRACT: Pseudogenes have been classified as functionless and their annotation is an ongoing problem. The Adh6-ps1—a mouse pseudogene belonging to the alcohol dehydrogenase gene complex (Adh) was analyzed to review the conservation, homology, expression, and interactions and identify any role it plays in disease phenotypes using bioinformatics databases. Results showed that Adh6-ps1 have 2 transcripts (processed and unprocessed) which may have emerged from a transposition and duplication event, respectively, and that induced inversions (Uox gene, In(3)11Rk) involving gene complexes associated with Adh6-ps1 have been implicated in a diverse range of diseases. Adh6-ps1 is highly conserved in vertebrates particularly rodents and expressed in the liver. The top 5 MirRNA targets were Mir455, Mir511, Mir1903, Mir361, and Mir669o markers. While much is unknown about Mir1903 and Mir669o, the silencing of Mir455 and Mir511 is linked with hepatocellular carcinoma (HCC), and Mir361 is implicated in endometrial cancers. Given the identified MirRNA interactions with Adh6-ps1 and its expression in HCC and reproductive systems, it may well have a role in tumorigenesis and disease phenotypes. Nonetheless, further studies are required to establish these facts to add to the growing efforts to understand pseudogenes and their potential involvement in disease conditions.

KEYWORDS: pseudogene, Adh6-ps1, gene expression, gene interaction

RECEIVED: March 19, 2021. **ACCEPTED:** July 26, 2021.

TYPE: Original Research

FUNDING: The author(s) received no financial support for the research, authorship, and/or publication of this article.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Martin C Nwadiugwu, Leonard Davis School of Gerontology, University of Southern California, Los Angeles, CA 90089, USA. Emails: nwadiugw@usc.edu; martintony4u@gmail.com

Background

The alcohol dehydrogenase genes have been identified in human and mouse, and are located on chromosome 3 and 4, respectively. This group of genes has been reported to encode for 5 different classes of enzymes in identical transcriptional orientation, and these enzymes can reverse the oxidation of a handful of primary and secondary alcohols to their related ketones and aldehydes, thereby playing essential roles in alcohol metabolism, homeostasis, and retinoid signaling.^{1,2} The mouse Adh1 gene and human ADH1A, ADH1B, and ADH1C genes encode class I enzymes responsible for ethanol metabolism; class III enzymes are encoded by Adh5 (mouse) and ADH5 (human) genes which are both involved in S-nitrosoglutathione metabolism.^{1,2} Genes encoding class I and III as well as class IV (mouse Adh7, human ADH7) are all involved in retinoid signaling.²

The mapping of Adh gene complex cDNAs to restriction endonuclease digestion of overlapping bacterial artificial clones (BAC) hybridized and isolated to ADH1 cDNA have helped in characterizing and determining its transcriptional direction as Adh7-Adh1-Adh4-Adh5.¹ In deer mouse, blast analysis of Adh1 and Adh4, which were the region of hybridization of ADH2 cDNA, identified class V enzymes namely: Adh6a and Adh6b as well as alcohol dehydrogenase 6 pseudogene 1 (Adh6-ps1), which is not processed into a functional mRNA.¹ Some of the Adh genes have been reported to be expressed in the liver (Adh4, Adh6b) and can be detected in the testes and small intestine (Adh6b).² While much is unknown about pseudogene Adh6-ps1, a phylogenetic tree analysis of the transcriptional orientation suggests that Adh gene complex resulted from a duplication event, and that the gene complex in human

and mouse have at least a gene for each of the 5 ADH classes.¹ Adh6-ps1 possibly belongs to ADH class V and is rather associated with the mouse gene cluster. It belongs to chromosome 3 with coordinates 311508–325678.

The annotation of genes that do not encode functional proteins have been described as an ongoing problem because these copies of genes are disabled by premature stop codons, sequence disablements, and protein domain truncations.³ These genes are commonly referred to as pseudogenes and are of 3 types, namely: processed pseudogenes, duplicated/unprocessed pseudogenes, and unitary/disabled pseudogenes.⁴ The processed genes arise from retrotransposition, which is an mRNA reverse-transcription and genomic DNA reintegration process; they have poly-A tail at the 3' end, extensions such as flanking direct repeats but lack introns and promoters.⁴ The unprocessed genes arise from gene duplication event and have intron-exon structures inherited from their parent gene, while the unitary pseudogenes are like the unprocessed pseudogenes but differ in not being duplicated before disablements.^{4,5} It is biologically important that these genes are well annotated for proper discrimination from those that have protein coding ability, and to provide clues to organismal evolution and history, and show patterns that demonstrate loss of coding ability, transcription, codon, or splice sites.³

Special consideration for studying pseudogenes include their ability to interfere with polymerase chain reaction (PCR) amplification, de novo gene prediction, and other gene-centric studies as well as learning about their mutational tendencies and age.⁶ It is important to understand the function of pseudogenes after a disrupted translation and transcription process as they have been reported to have diverse disablements ranging



from lack of transcription to the presence of frame shifts, repetitive elements, and stop codons.⁶ While many pseudogenes have been evidenced to be actively transcribed, they are not translated due to disablements.

The objective of this study is to review what is known about Adh6-ps1 for possible biological insights, better annotation and profiling, as some pseudogenes have been suggested to play a role in gene regulation by being capable of low coding expression in a state of relaxed selection⁶ and to also produce functional RNAs. This study aims to review the conservation, homology, expression, interactions, and identify any role Adh6-ps1 plays in disease phenotypes.

Methods

Bioinformatics databases were queried and searched between February 5 and March 5, 2021. The characteristics of Adh6-ps1 were obtained using data from GenBank, Mouse Genome informatics (MGI), Ensemble, and National Centre for Biotechnology Information (NCBI) databases.⁷⁻⁹ The phylogenetic tree information, homology, and evolutionary constrain were identified using UCSC Mouse Genome Browser, NCBI BLAST searches, and MGI database⁹⁻¹¹; while possible interactions, expression, and role of Adh6-ps1 were analyzed using Genevestigator and Alliance of Genome Resources.^{12,13} Data from all database searched were used to interpret any likely functions and relevance of Adh6-ps1 in cancer and other diseases.

Results

Characteristics of Adh6-Ps1

Adh6-ps1 gene identity on Genbank is 639769. Adh6-ps1 is highly associated with *Mus musculus* and has 2 splice variants: (a) processed transcript with 4 non-protein coding exons having a nucleotide length of 2789 bps (ENSMUST00000171054.6) and (b) a transcribed unprocessed transcript with nucleotide length of 754 bp (ENSMUST00000196293.2) having 6 non-protein coding exons. The transcribed unprocessed pseudogene (Adh6-ps1-202) has no protein coding potential, while the transcribed or processed transcript (Adh6psi-201) consists of non-protein coding exon variants. This suggests that the processed pseudogene is likely to have emerged from transposition, while the unprocessed transcript possibly arose from a duplication event.⁴

Adh6-ps1 is also known as Adh5ps, Gm7277, predicted gene 7277, predicted gene 639769, RIKEN cDNA 1300002P07, and 1300002P07Rik.¹³ The sequence locus for Adh6-ps1 is Chr3:138374121-138388291 (GRCm38) and the genetic map is Chr3 64.22cM cytoband H2. It is also identified as MGI:3779711 (MGI: <http://www.informatics.jax.org/marker/MGI:1918999>) and on Ensembl as ENSMUG00000090306. It has been sequenced in C57BL, C57BL/6, C57BL/6J, C57BL/6NJ, CAROLI/Eij, SPRET/Eij, and CAST/Eij mouse strains with varying sequence length and types.⁸ The majority of information

provided on Adh6-ps1 in this study was with respect to the C57BL, C57BL/6, and C57BL/6J strains.

Identification of homology and evolutionary constrain

A nucleotide BLAST (BLASTn) was implemented against the “Nucleotide collection” database and optimized for highly similar sequence (megablast) using the accession number NR_033581. The BLAST algorithm identified *M musculus*, *Mus caroli*, *Grammomys surdaster*, and *Rattus norvegicus* as organisms with the closest identity. *Homo sapiens* were, however, not listed. Implementing a reverse BLAST with the top hit accession number of *R norvegicus* (XR_0055000307.1) showed that it is an ortholog of NR_033581 with 89.43% identity. Repeating this process with Adh2 (AJ245750.1) which was part of the initial hit did not output Adh6-ps1 as the reciprocal best hit suggesting a paralogous relationship.

Multiple alignments across 35 vertebrates on Mouse Genome Browser using phyloP, and phastCons—a Hidden Markov Model (HMM), showed high levels of conservation within mouse gene sequences and that Adh6-ps1 is very well conserved in rats. Other vertebrates where it was shown to be conserved include guinea pig, squirrel, rabbit, pika, Human, Rhesus, Chinese pangolin, dog, and sheep. Analyzing Adh6-ps1 via neighbor joining on BLAST tree view also showed high conservation among vertebrates, especially rodents. The result confirms that Adh6-ps1 (NR_033581.1) has 2 transcripts and has a paralogous relationship with *M musculus* clones AC079845.31 and AC079832.18, and Adh2 clone product (AK004863.1) as seen in Figure 1. Investigating AC079845.31 and AC079832.18 showed they seemed to be a more closely related sequence; however, running a BLASTn search for AC079845.31—a *M musculus* complete sequence clone rp23-461a12—yielded no significant similarity result. Furthermore, long interspersed nuclear elements (LINE)—a non-long terminal repeat retrotransposon, were the most common interspersed repeats (Lx5b, Lx5, Lx8) according to the RepeatMaster information strategy implemented on UCSC Genome browser.

Tissue expression and role of ADH6-PS1

The mouse ENCODE transcriptome data¹⁴ on NCBI Gene database shows that Adh6-ps1 is largely expressed in mouse adult male liver with mean RPKM of 0.97 and to a lesser degree expressed in adult subcutaneous fat pad (0.041 RPKM) and adult genital pads (0.018 RPKM). In early mouse development, Adh6-ps1 has been shown to be expressed in connective tissue, liver, and biliary system, and in the reproductive system (Figure 2). These are regions of general endogenous wild-type gene expression of Adh6-ps1 during developments. However, comparison with vertebrate and invertebrate ortholog genes on Alliance of Genome Resource reveals that it is unlikely to have imaginal precursor, therefore, suggesting less likelihood of being involved in the developmental process of invertebrates,

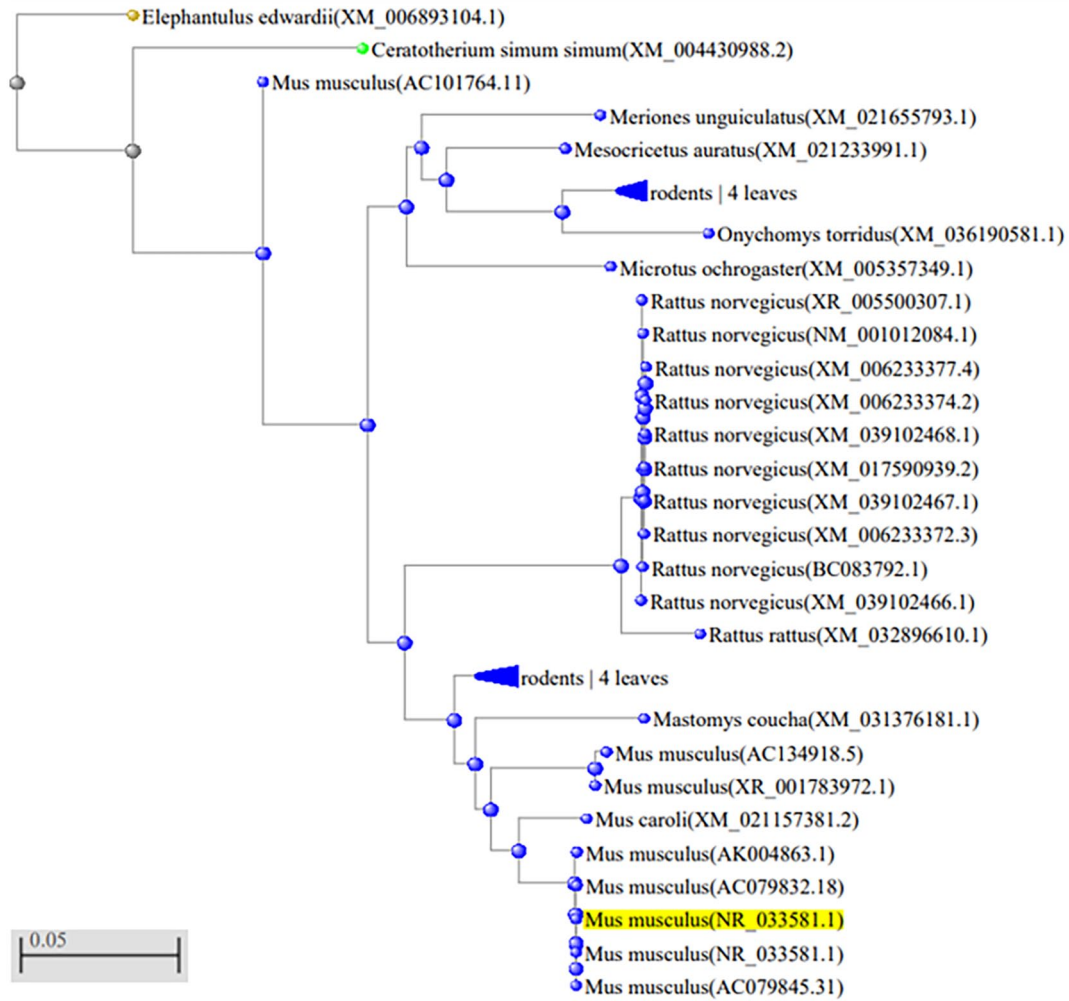


Figure 1. BLAST Tree View. Adh6-ps1 in rodents.

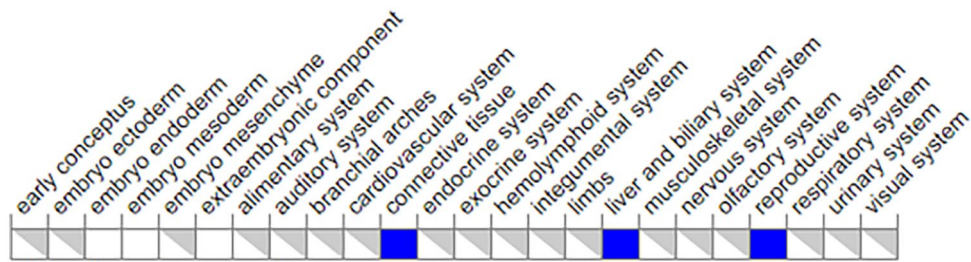


Figure 2. MGI. Tissue expression of Adh6-ps1 during development.

especially drosophila. It is expected that the processed transcript (ENSMUST00000171054.6) constitutes the vast majority of Adh-ps1 transcript expressed in associated tissues and have more genetic interactions because processed pseudogenes are found to be more expressed in genomes of mammals than in genomes of other vertebrates and to occur in poor GC regions.⁶

Relevance to cancer and other diseases

A survey of Adh6-ps1 expression across 124 cell lines in Genevestigator showed moderate expression level in ES-R1—a pluripotent mouse embryonic stem cell line. In cancer,

Adh6-ps1 was found to be moderately expressed in liver hepatocellular carcinoma (HCC) when compared with 17 other cancer types where it showed low expression. In addition, Adh6-ps1 was found to be relatively highly expressed in C57BL/6J/AKRJ strain when tested across 2667 perturbations. This could be due to the popular use of the strain, although the strain have been reported to maximally express most mutations. The expression of Adh6-ps1 across cancers, cell lines, tissues, and perturbations can be seen in Figure 3.

Furthermore, cesium irradiation-induced paracentric inversion of Uox gene located at Chr3, 72.09cM has an inversion that spans 80% of Chr3 involving many genes including

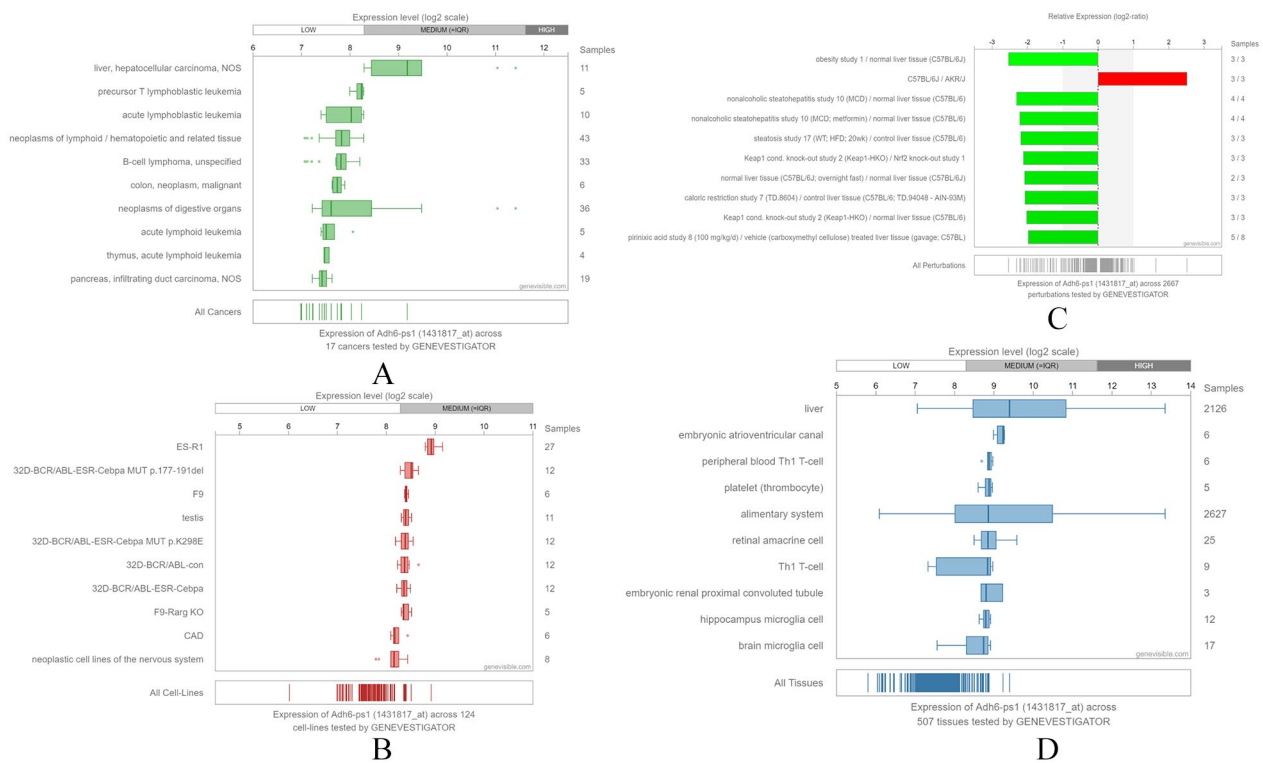


Figure 3. GENEVESTIGATOR. Expression of Adh6-ps1 across: (A) 17 cancers, (B) 124 cell lines, (C) 507 tissues, and (D) 2666 perturbations.

Adh6-ps1. This induced inversion is implicated in hyperuricemia, kidney disease, and phenotypes associated with homeostasis and metabolism, aging, immune and urinary systems.^{8,15} Apart from Uox gene, In(3)11Rk—a chemically induced inversion which produces phenotype that affects the reproductive system—also involves Adh6-ps1.^{8,16} The 2 studies suggest that mutations involving Adh6-ps1 and several other genes within the same chromosomal locus have implications for disease phenotypes; it is however unclear what specific role Adh6-ps1 plays in the disease state and if single mutations to Adh6-ps1 have adverse biological consequences.

While structural and functional gene curation are implemented to identify chromosomal location of elements controlling gene transcription, respectively,¹⁷ this is limited in pseudogene annotation because there are not enough guidance on structural and functional annotation as Adh6-ps1 is non-protein coding.

Interactions of Adh6-ps1

Analysis of Adh6-ps1 on BioGrid revealed no available protein interaction. There was also no gene ontology (GO) process, function, or component listed even after further search on Alliance of Genome Resources.¹³ However, MGI database searches showed that it has several interactions with 245 MirRNA markers. MirRNAs are RNA genes associated with miRNA class; the miRNAs are evolutionarily conserved small RNA molecules lacking coding potential, but play a role in post-transcriptional modification and in regulating cell differentiation, proliferation, and oxidative stress.¹⁸⁻²⁰ The top 5 MirRNA targets are displayed in Figure 4.

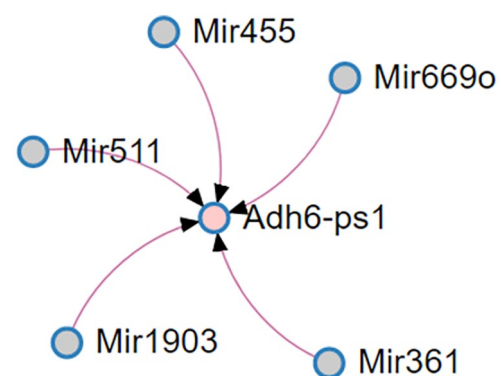


Figure 4. MGI. Top MirRNA interactions with Adh6-ps1.

Out of the 5 MirRNA targets, GO biological process classifies Mir455 and Mir511 to be largely expressed in the alimentary system and to play a role in response to stimulus. As previously stated, Adh6-ps1 is highly expressed in the liver and moderately associated with HCC. It is important to know the role Adh6-ps1 interactions with MirRNAs play in gene regulation because MirRNAs are vital in investigating genes for possible involvement in tumorigenesis, apoptosis, pathogenesis, and disease resistance.²¹

Mir455 and Mir511 were reported to be downregulated in diethylnitrosamine chemically induced hepatocellular carcinoma (DEN-HCC) in rat model.²² The DEN-induced HCC rat model also reported the activation of the notch signaling pathway which plays a role in liver development, lipid metabolism, and lipogenesis.²³ In brief, the interactions between Adh6-ps1, Mir455, and Mir511 may add to the understanding

of HCC because Adh6-ps1 is moderately expressed in HCC, and silencing of Mir455 and Mir511 is linked with HCC.

Using data of 9 solid tumor types from The Cancer Genome Atlas (TCGA) and HCC data from Columbia University Medical Centre, Shen et al²⁴ found that Mir455 was down-regulated with HCC tumor type specificity. In cell lines and tumor tissues of HCC patients, miR-455-5p an alias of Mir455 (<https://www.genecards.org/cgi-bin/carddisp.pl?gene=MIR455>) was reported to be significantly suppressed, but when overexpressed in cell lines it was linked with decreased tumor functions and proliferation, and the suppression of insulin growth factor receptor (IGF-1R).¹⁸ This led to suggestions by Hu et al¹⁸ that miR-455-5p is a potential target for therapeutic interventions in cancer because glycolysis is needed for tumor proliferation, and the suppression of IGF-1R disrupts AKT phosphorylation—a signaling pathway which promotes cell survival and growth, leading to glycolytic enzymes and glucose transporter (GLUT) downregulation.¹⁸ miR-455-5p decreased expression have also been linked with gastric cancer (GC) and colorectal cancer (CRC) and have been found to regulate RAB18 and RAF1 genes which are among its direct targets.^{20,25} The overexpression of miR-455-5p leads to suppression of RAB18 and RAF1 genes and vice versa, and further investigation of their relationship have been suggested as a process that could lead to potential therapeutic intervention strategies for gastric and colorectal cancer treatment.^{20,25}

Meanwhile, it is left to be seen what phenotypical change MirRNAs associations with Adh6-ps1 could lead to as not much is known about the specific role Adh6-ps1 plays in HCC, GC, CRC, and other cancers. Howbeit, there are limited information on the role played by Mir1903 and Mir669o, but some studies have identified Mir361 as a tumor suppressor implicated in endometrial cancers when downregulated.²⁶ This suggests some role for Adh6-ps1 in tumors of the reproductive systems because endometrial cancers are associated with reproductive systems. Unlike HCC, a possible link between Adh6-ps1, Mir361, and diseases of the reproductive systems is even to a lesser extent given that Adh6-ps1 is slightly expressed in reproductive systems.

Taken together, Mir361 is implicated in endometrial cancers and the silencing of Mir455 and Mir511 is linked with HCC, with Mir455 being HCC tumor type specific as well as being associated with GC and CRC. This implies that interactions of MirRNAs with Adh6-ps1 may have a role in tumorigenesis and disease phenotypes.

Conclusions

It has been shown that Adh6-ps1 is highly conserved in vertebrates, particularly rodents. Adh6-ps1 has a processed and unprocessed transcript; the former arose via mRNA reverse transcription and genomic DNA reintegration process, while the later emerged via a duplication event. It is unclear what ancestral functional gene both arose from. Also, the protein structure, functional insights from co-expression, and protein

interactions could not be investigated as Adh6-ps1 was a non-coding protein.

Although pseudogenes are known to be functionless, some have been suggested to play a role in gene regulation.²⁷ Mir455, Mir511, and Mir361 respective expressions in HCC, GC, CRC, and endometrial cancers, and their associations with Adh6-ps1 suggest that Adh6-ps1 may well have a secondary role in tumorigenesis and disease phenotypes. Nonetheless, more experimental studies are required to unpack the specific role Adh6-ps1 plays, and if this is confirmed, it will add to the growing knowledge of pseudogenes and their potential involvement in disease conditions.

Acknowledgements

The author would like to acknowledge Berenice Benayoun for providing useful feedback for the paper.

Author Contributions

MN developed the concepts and wrote the manuscript.

Availability of Data and Materials

All data supporting the study have been duly referenced.

Database

MGI: <http://www.informatics.jax.org/marker/MGI:1918999>
Genevestigator: <https://genevisible.com/cell-lines/MM/Gene%20Symbol/Adh6-ps1>

ORCID iD

Martin C Nwadiugwu  <https://orcid.org/0000-0001-6788-8305>

REFERENCES

1. Szalai G, Duester G, Friedman R, et al. Organization of six functional mouse alcohol dehydrogenase genes on two overlapping bacterial artificial chromosomes. *Eur J Biochem.* 2002;269:224-232. doi:10.1046/j.0014-2956.2001.02642.x.
2. Szalai G, Veres M, Duester G, Lawther R, Lockhart M, Felder MR. Tissue expression pattern of class II and class V genes found in the Adh complex on mouse chromosome 3. *Biochem Genet.* 2008;46:685-695. doi:10.1007/s10528-008-9180-8.
3. Harrison PM. Computational methods for pseudogene annotation based on sequence homology. In: Poliseno, L, ed. *Pseudogenes. Methods in Molecular Biology (Methods and Protocols)*. Vol. 1167. New York, NY: Humana Press; 2014:35-48. doi:10.1007/978-1-4939-0835-6_3.
4. Rouchka CE, Cha EI. Current trends in pseudogene detection and characterization. *Curr Bioinform.* 2009;4:112-119.
5. Zheng D, Gerstein MB. A computational approach for identifying pseudogenes in the ENCODE regions. *Genome Biol.* 2006;7:S13. doi:10.1186/gb-2006-7-s1-s13.
6. Yu Z, Morais D, Ivanga M, Harrison PM. Analysis of the role of retrotransposition in gene evolution in vertebrates. *BMC Bioinformatics.* 2007;8:308.
7. Hubbard T, Andrews D, Caccamo M, et al. Ensembl 2005. *Nucleic Acids Res.* 2005;33:D447-D453.
8. Mouse Genome Informatics Web Site (MGI). *Mouse Sequences Summary Report*. Bar Harbor, ME: The Jackson Laboratory; 2021. <http://www.informatics.jax.org/sequence/marker/MGI:1918999>. Accessed February 27, 2021.
9. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215:403-410.
10. Kent WJ, Sugnet CW, Furey TS, et al. UCSC Genome Browser: UCSC Genome Browser on Mouse Jun. 2020 (GRCm39/mm39) assembly. *Genome Res.* 2002;12:996-1006.

11. Boratyn GM, Schaffer AA, Agarwala R, Altschul SF, Lipman DJ, Madden TL. Domain enhanced lookup time accelerated BLAST. *Biol Direct*. 2012;7:12. doi:10.1186/1745-6150-7-12.
12. Hruz T, Laule O, Szabo G, et al. Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. *Adv Bioinformatics*. 2008;2008:420747. doi:10.1155/2008/420747.
13. The Alliance of Genome Resources Consortium. Alliance of Genome Resources Portal: unified model organism research platform. *Nucleic Acids Res*. 2020;48:D650-D658. doi:10.1093/nar/gkz813.
14. Stamatoyannopoulos JA, Snyder M, Hardison R, et al. An encyclopedia of mouse DNA elements (Mouse ENCODE). *Genome Biol*. 2012;13:418.
15. Cook SA, Akeson EC, Calvano C, et al. Mouse paracentric inversion In(3)5Rk mutates the urate oxidase gene. *Cytogenet Cell Genet*. 2001;93:77-82. doi:10.1159/000056953.
16. Roderick TH. Using inversions to detect and study recessive lethals and detriments in mice. In: de Serres, FJ, Sheridan, W, eds. *Utilization of Mammalian Specific Locus Studies in Hazard Evaluation and Estimation of Genetic Risk*. New York, NY: Plenum Publishing Corp; 1983:135-167.
17. Tello-Ruiz MK, Marco CF, Hsu FM, et al. Double triage to identify poorly annotated genes in maize: the missing link in community curation. *PLoS ONE*. 2019;14:e0224086. doi:10.1371/journal.pone.0224086.
18. Hu Y, Yang Z, Bao D, Ni JS, Lou J. miR-455-5p suppresses hepatocellular carcinoma cell growth and invasion via IGF-1R/AKT/GLUT1 pathway by targeting IGF-1R. *Pathol Res Pract*. 2019;215:152674.
19. Mohr AM, Mott JL. Overview of microRNA biology. *Semin Liver Dis*. 2015;35:3-11.
20. Chai J, Wang S, Han D, Dong W, Xie C, Guo H. MicroRNA-455 inhibits proliferation and invasion of colorectal cancer by targeting RAF proto-oncogene serine/threonine-protein kinase. *Tumour Biol*. 2015;36:1313-1321. doi:10.1007/s13277-014-2766-3.
21. Nwadiugwu MC. Thyroid tumor: investigating MicroRNA-21 gene suppression in FTC and FTA. *Cancer Informatics*. 2020; 19: 1176935120948474. doi:10.1177/1176935120948474.
22. Fornari F, Gramantieri L, Callegari E, et al. MicroRNAs in animal models of HCC. *Cancers (Basel)*. 2019;11:1906. doi:10.3390/cancers11121906.
23. Adams JM, Jafar-Nejad H. The roles of notch signaling in liver development and disease. *Biomolecules*. 2019;9:608. doi:10.3390/biom9100608.
24. Shen J, Siegel AB, Remotti H, Wang Q, Santella RM. Identifying microRNA panels specifically associated with hepatocellular carcinoma and its different etiologies. *Hepatoma Res*. 2016;2:151-162. doi:10.20517/2394-5079.2015.66.
25. Liu J, Zhang J, Li Y, Wang L, Sui B, Dai D. MiR-455-5p acts as a novel tumor suppressor in gastric cancer by down-regulating RAB18. *Gene*. 2016;591:308-315. doi:10.1016/j.gene.2016.07.034.
26. Ihira K, Dong P, Xiong Y, et al. EZH2 inhibition suppresses endometrial cancer progression via miR-361/Twist axis. *Oncotarget*. 2017;8:13509-13520. <https://www.oncotarget.com/article/14586/text/>
27. Xiao J, Sekhwal MK, Li P, et al. Pseudogenes and their genome-wide prediction in plants. *Int J Mol Sci*. 2016;17:1991. doi:10.3390/ijms17121991.