OPEN ACCESS

Haya: The Saudi Journal of Life Sciences
Abbreviated Key Title: Haya Saudi J Life Sci
ISSN 2415-623X (Print) |ISSN 2415-6221 (Online)
Scholars Middle East Publishers, Dubai, United Arab Emirates
Journal homepage: https://saudijournals.com

Original Research Article

To Analyze Single Nucleotide Polymorphism and Pseudogenes using Genome Wide Association in Diabetes Mellitus

Talib Yusuf A[1*], SH Talib[2], Huzefa S. Bhagat[1], Abdoulaye Diawara[3]

[1]Dept of Biotechnology, Burhani College, Mazgaon, Mumbai, MS, India
[2]Dept of Medicine, MGM Medical College and Hospital, Aurangabad, MS, India
[3]Dept of Bioinformatics, African Centre of Excellence in Bioinformatics (ACE-B), USTTB, Bamako, Mali

DOI: 10.36348/sjls.2021.v06i11.004 | **Received:** 19.08.2021 | **Accepted:** 26.09.2021 | **Published:** 11.11.2021

***Corresponding author:** Dr. Talib Yusuf

Abstract

Diabetes Mellitus type 2 is said to be one of the complex diseases which is caused by complex interplay between genetic, epigenetic and environmental factors, while the major environmental factors, i.e., diet and physical activity level are well known, but the challenge is to identify the genetic factors involves in it. NGS (next generation sequencing) and GWAS (Genome Wide association studies) have led to technical development of genetic variants risked and protection of Type 2 Diabetes Mellitus. NGS which shows the amount of gene which has been expressed and there arrangement of nucleotide bases in the gene fragment which code for protein, also some genes , or a copy of gene which has lost the ability to produce a functional protein, may be due to mutation or inaccurate duplication in the sequence which are termed as Pseudo gene. These expressions of pseudo gene can occur due to SNP's (Single Nucleotide Polymorphism) are DNA sequence variant that occur when a single nucleotide (A, T, C or G) in the genome sequence altered.

Keywords: Single Nucleotide Polymorphism, Pseudogene, Diabetes Mellitus type2, Pathways, Gene Ontology.

1. INTRODUCTION

Pseudo gene can be defined as a copy of gene that has lost the capacity to produce a functional protein; causes can be due to mutation or inaccurate duplication [1]. These pseudo gene are functionless and evolutionary inert, they neither conserved nor removed from the expression [2, 3]. In some scenario, those pseudo genes are not translated into protein [4], are at least transcribed into RNA [5]. These pseudo gene transcripts are capable of influencing the activity of other genes that code for proteins, therefore altering expression and in turn affecting the phenotypic of the organisms [6]. Although more than 99% of human DNA sequence is the same across the population, variants in DNA sequence can provide major impact on how human's genetic arrangements respond to disease; environmental co factors such as bacteria, viruses and toxins, chemicals and drug and other therapies. Sciences Sporadic discovery and characterization of pseudo genes over following 20 years have divided these class into three [7], a) unitary pseudo gene, they are formed when spontaneous mutation occur in a coding gene that abolish either transcription or translation, b) duplicated pseudo gene, it is formed when replication of the chromosome is performed incorrectly, such duplication may lead to formation of functional gene families like HOX gene clusters, but if the part of the gene is not faithfully copied than these can lead to different types of mutations or the loss of promoter or enhancer thus resulting in a non-functional duplication pseudo gene., c) processed pseudo gene, it is formed when an mRNA molecule is reversed transcribed and integrated into a new location in the parental genome, these processed class of pseudo gene are produced from mRNA [8, 9, 46], they usually lack introns and a promoter.

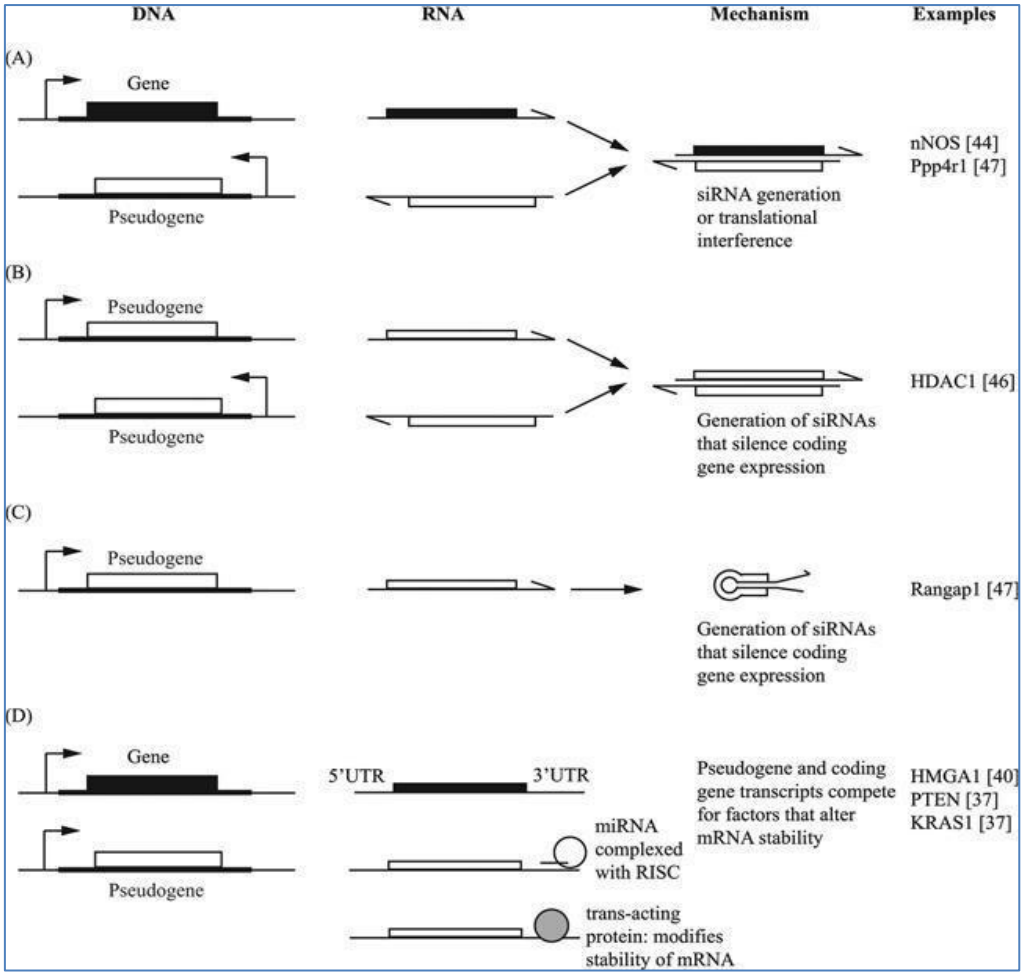


Fig-1: Diagrammatic representation of different mechanism in pseudogenes [46]

The sequencing of a range of genomes, including the human genome, has revealed the extent of pseudo gene abundance [9]. Estimates for the number of human pseudo genes range from 10000 to 20000, making them almost as prevalent as coding genes [10]. The majority of these are processed pseudo genes and fewer than 100 are unitary pseudo genes. The types of genes that produce processed pseudo genes are predominantly highly expressed housekeeping genes or shorter RNAs such as genes encoding ribosomal proteins. Pseudo genes are found in various species including bacteria, plants, insects, and nematode worms, there examples can be extracted from various biological databases [11]. Sometimes pseudo genes can also be named as "Junk DNA" [12] because they lack protein-coding capacity. Some genes that appear to be pseudogenized may in fact code for protein others are genuinely non-coding, but are no means "junk" as they may actually play functional roles [13]. It is noted that apparently some pseudogenes are capable of producing lncRNA (long non-coding RNA). It has been observed RNA was taking a purely intermediary role in the expression of protein from DNA, it is now widely acknowledged that ncRNA can play significant role in the regulation of gene expression. SNPs play an important role in evolutionary segments which make them easier to follow the population and individual studies. SNPs do not cause disease, but they can help determine the likelihood and can correlate the dissimilarity between the classes. Of course, SNPs are not absolute which indicates some relationship with disease development, someone who has inherited 2 alleles IRS4 (insulin receptor substrate), or IRS2 (insulin receptor substrate) may never develop Diabetes Mellitus Type 2, while another who has been inherited 2 alleles may develop, IRS 6 (insulin receptor substrate) is just one gene that has been linked to Diabetes Mellitus.

2. MATERIALS AND METHODS

Three (03) subjects (HS (Healthy Subjects), PDS (Prediabetic Subjects), and DS (Diabetic Subjects)) were selected based on HBA1c glycated hemoglobin (form of hemoglobin which is measured primarily to identify the 3 months plasma glucose concentration) and Body Mass Index (BMI) of MGM medical college and Hospital, Aurangabad. (M.S) with their informed consent, those with morbidity that needed medical attention were excluded from the study. Ethical approval was granted by the Ethics Committee of MGM University of Health Sciences, Kamothe, Navi Mumbai, (M.S). Randomly there 2 ml blood was

extracted mixed with Anticoagulant and proceeded for Genomic DNA isolation. Genomic DNA isolated from Whole Blood treated with Trizol. Total RNA expected from the treatment out of which mRNA is enriched as carries coding region which has been transcribed from DNA can be studied. RNA concentration and purity [260/280] were checked by using Nanodrop Spectrophotometer [Table 1].

Table-1: Concentration and purity

Sample ID	Concentration (ng/µl)	Yield (µg)	A260/280
HS	278	5.5	2.01
PDS	242	4.8	1.89
DS	326	6.5	1.97

Once the purity and concentration by Nanodrop Spectrophotometer of RNA is has been processed to prepare the cDNA with the help of reverse transcriptase, and then proceed to Whole Transcriptome Analysis (WTA). Where only exomes are targeted to find the SNPs or any pseudogenes are expressed. A consensus set of expressed transcripts was obtained and different classes of transcript including novel isoforms are identified. Transcript abundance was estimated by FPKM values (fragments per kilobase million) and TPM (Transcripts per kilobase million) (count up the total reads in a sample and divide that number by 1,000,000, this is the "per million" scaling factor) Table 2.

Table-2: FPKM & TPM value

Samples	Gene ID	FPKM	TPM	Location chr
HS	ENSG230291	0.3796	0.366	12
PDS	ENSG230291	0.8704	0.5815	12
DS	ENSG230291	0.0309	0.0234	12

3. RESULTS AND DISCUSSION

The results based on FPKM, TPM and SNPs found on the transcriptome data obtained after completing the WTA analysis ([list 1], here one nucleic acid G is found mutated in PDS sample while T is found same in HS and DS sample,[Table 3] also transcriptome based values of FPKM and TPM [table 2] found more in PDS (prediabetic subject) as compare to HS and DS (healthy and Diabetic) subjects [graph 1 and graph 2] located on chromosome 12 [table 2]. In chromosome 12 a fragment of 487 base pairs were found in all subjects out of whom one or two single nucleotide polymorphisms is found at the end of that pseudogene after WTA [list 1]. Which concludes that there are various SNPs involved in human chromosome located on different region, these single nucleotide polymorphisms may lead to pseudogenes which will not code for protein in the cell, which lead to incomplete functioning of the body and make more susceptible and prone to various clinical and sub clinical complications. Also, those SNPs found can play vital role in disturbing the metabolic pathways and can give some complication and metabolic disorders.

1) Healthy Subject (HS)

ATGAAGTTTAATCCCTTTGTGACTTCCGACCGAAGCAAGAATCGCAAAAGGCATTTCAAC
GCACCTTCCCACATTCGAAGGAAGATTATGTCTTCCCTTCTTTCCAAAGAGCTGAGACAG
AAGTACAACGAGCGATCCATGCCCATCCGAAAGGACGATGAAGTTCAGGTTGTACGAGGA
CACTACGAAGGTCAGCAAATTGGCAAAGTGGTCCAGGTTTACAGGAAGAAATATGTTATC
TACATTGAATGGGTGCAGCGGGAAAAGGCTAATGGCACAACTGTCCACATAGGCATTCAC
TCCAGCAAGGTGGTTATCACTAGGCTAAAACTGGACAAAGACAGCAAAAAGATCCTTGAA
CGGAAAGCCAAATCTCGCCAAGTAGGAAAGGAAAAGGGCAAATACAAGGAAGAAACAATT
GAGAAGATGGAGGAATAA

2) Prediabetic Subject (PDS)

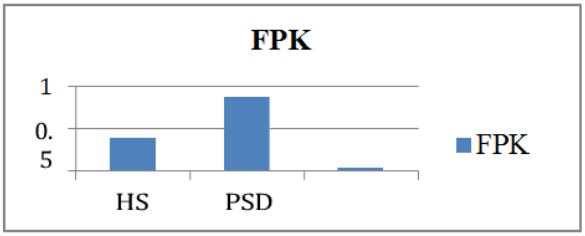
ATGAAGTTTAATCCCTTTGTGACTTCCGACCGAAGCAAGAATCGCAAAAGGCATTTCAAC
GCACCTTCCCACATTCGAAGGAAGATTATGTCTTCCCTTCTTTCCAAAGAGCTGAGACAG
AAGTACAACGAGCGATCCATGCCCATCCGAAAGGACGATGAAGTTCAGGTTGTACGAGGA
CACTACGAAGGTCAGCAAATTGGCAAAGTGGTCCAGGTTTACAGGAAGAAATATGTTATC
TACATTGAATGGGTGCAGCGGGAAAAGGCTAATGGCACAACTGTCCACATAGGCATTCAC
TCCAGCAAGGTGGTTATCACTAGGCTAAAACTGGACAAAGACAGCAAAAAGATCCTTGAA
CGGAAAGCCAAATCTCGCCAAGTAGGAAAGGAAAAGGGCAAATACAAGGAAGAAACAATT
GAGAAGATGGAGGAA

3) Diabetic Subject (DS)

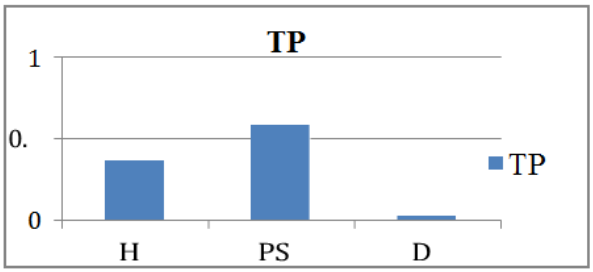
ATGAAGTTTAATCCCTTTGTGACTTCCGACCGAAGCAAGAATCGCAAAAGGCATTTCAAC
GCACCTTCCCACATTCGAAGGAAGATTATGTCTTCCCTTCTTTCCAAAGAGCTGAGACAG
AAGTACAACGAGCGATCCATGCCCATCCGAAAGGACGATGAAGTTCAGGTTGTACGAGGA
CACTACGAAGGTCAGCAAATTGGCAAAGTGGTCCAGGTTTACAGGAAGAAATATGTTATC
TACATTGAATGGGTGCAGCGGGAAAAGGCTAATGGCACAACTGTCCACATAGGCATTCAC
TCCAGCAAGGTGGTTATCACTAGGCTAAAACTGGACAAAGACAGCAAAAAGATCCTTGAA
CGGAAAGCCAAATCTCGCCAAGTAGGAAAGGAAAAGGGCAAATACAAGGAAGAAACAATT
GAGAAGATGGAGGAATAA

List 1 above is Fasta sequences of HS, PDS and DS subject highlighting pseudogenes Table 3: Single Nucleotide Polymorphism of Samples, HS, PDS and DS

sample	SNP
HS	GGAATAA
PDS	GGAGGAA
DS	GGAATAA



Graph-1: Showing FPKM value of HS, PDS and DS



Graph-2: Showing TPM Values of HS, PDS and DS

4. CONCLUSION

Several traits like obesity, insulin resistance and increased levels of pro inflammatory cytokines precede and accompany T2DM. On the other hand, it is widely accepted that T2DM has a hereditary component. Hence, it is natural to expect patients suffering from hereditarily transferred predisposition to T2DM to have one or more SNPs responsible for accelerated progression of underlying traits. The indigenous Aurangabad, Marathwada Region population, having an increased genetic predisposition to develop T2DM, may be considered appropriate for this kind of study. The authors acknowledge certain limitations of this study. Confirmation of common variants in the human genome with modest effects on common disease risk, even if real, need large sample sizes to overcome the influence of many genetic and environmental modifiers. Also, our study subjects consisted of Aurangabad (M.S) ethnicity, and thus, the generalizability to other ethnicities is unknown. Typically, SNPs have been used as markers to search for the real determinant of a disease in linkage disequilibria with it. As previously mentioned, the use of functional SNPs, which may be the real disease determinants, could be an important factor in increasing the sensitivity of association tests. Despite the obvious importance that alterations in the regulation, expression level or splicing of genes can have for the phenotype, these have long been ignored in the most common approaches to finding functional SNPs, which have instead focused more on the possible effect of polymorphisms causing amino acid changes.

ACKNOWLEDGEMENT

I sincerely thanks to MGM medical College and hospital for providing the fund for the said WTA (Whole transcriptome Analysis).

Conflict of interest

The author declares no conflict of interest**.**

REFERENCES

1. Jacq, C., Miller, J., & Brownlee, G. (1977). A pseudogene structure in 5S DNA of Xenopuslaevis. Cell; 12, 109–120.
2. Mighell, A.J., Smith, N.R., Robinson, P.A., & Markham, A.F. (2000). Vertebrate pseudogenes. FEBS Lett. 468; 109–114.
3. Zhang, Z.D., Frankish, A., Hunt, T., Harrow, J., & Gerstein, M. (2010). Identifi cation and analysis of unitary pseudogenes: historic and contemporary gene losses in humans and other primates. Genome Biol, 11; R26.
4. D'Errico, I., Gadaleta, G., & Saccone, C. (2004). Pseudogenes in metazoa: origin and features. Briefings Funct. Genomics Proteomics, 3; 157–167.
5. Zheng, D., Frankish, A., Baertsch, R., Kapranov, P., Reymond, A., Choo, S.W., Lu, Y., Denoeud, F., Antonarakis, S.E., Snyder, M. (2007). Pseudogenes in the ENCODE regions: consensus annotation, analysis of transcription, and evolution. *Genome Res, 17*; 839–851.
6. Ohshima, K., Hattori, M., Yada, T., Gojobori, T., Sakaki, Y. and Okada, N. (2003). Whole-genome screening indicates a possible burst of formation of processed pseudogenes and Alu repeats by particular L1 subfamilies in ancestral primates. *Genome Biol, 4*; R74.

7. Torrents, D., Suyama, M., Zdobnov, E., & Bork, P. (2003). A genome-wide survey of human pseudogenes. *Genome Res, 13*, 2559–2567.
8. Zhang, Z., Harrison, P.M., Liu, Y., & Gerstein, M. (2003). Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome. *Genome Res, 13*, 2541–2558.
9. Zhang, Z., & Gerstein, M. (2004). Large-scale analysis of pseudogenes in the human genome. *Curr. Opin. Genet*. Dev, 14; 328–335.
10. Zhang, Z., Harrison, P., & Gerstein, M. (2002). Identifi cation and analysis of over 2000 ribosomal protein pseudogenes in the human genome. *Genome Res, 12*; 1466–1482.
11. Pink, R.C., Wicks, K., Caley, D.P., Punch, E.K., Jacobs, L., & Carter, D.R. (2011). Pseudogenes: pseudo-functional or key regulators in health and disease? RNA, 17, 792–798.
12. Karro, J.E., Yan, Y., Zheng, D., Zhang, Z., Carriero, N., Cayting, P., Harrrison, P., & Gerstein, M. (2007). Pseudogene.org: a comprehensive database and comparison platform for pseudogene annotation. *Nucleic Acids Res;* 35, D55–D60.
13. Balakirev, E.S., & Ayala, F.J. (2003). Pseudogenes: are they 'junk' or functional DNA? *Annu. Rev. Genet*; 37, 123–151.
14. Caley, D., Pink, R., Trujillano, D., & Carter, D. (2010). Long noncoding RNAs, chromatin, and development. *Sci. World J; 10*, 90–102.
15. T.So, J.Y., Sun, X.H., Kao, T.H., Reece, K.S., & Wu, R. (1985). Isolation and characterization of rat and human glyceraldehyde-3-phosphate dehydrogenase cDNAs: genomic complexity and molecular evolution of the gene. Nucleic Acids Res, 13; 2485–2502.
16. Redshaw, Z., & Strain, A.J. (2010). Human haematopoietic stem cells express Oct4 pseudogenes and lack the ability to initiate Oct promoter-driven gene expression. J. Negat. Results Biomed, 9; 2.
17. Fujii, G.H., Morimoto, A.M., Berson, A.E., & Bolen, J.B. (1999). Transcriptional analysis of the PTEN/MMAC1 pseudogene, psi PTEN. Oncogene, 18, 1765–1769.
18. Kalyana-Sundaram, S., Kumar–Sinha, C., Shankar, S., Robinson, D.R., Wu, Y.M., Cao, X., Asangani, I.A., Kothari, V., Prensner, J.R., Lonigro, R.J. (2012). Expressed pseudogenes in the transcriptional landscape of human cancers. Cell; 149, 1622–1634.
19. Pei, B., Sisu, C., Frankish, A., Howald, C., Habegger, L., Mu, X.J., Harte, R., Balasubramanian, S., Tanzer, A., Diekhans, M. (2012). The GENCODE pseudogene resource. *Genome Biol; 13*, R51.
20. Khachane, A.N., & Harrison, P.M. (2009). Assessing the genomic evidence for conserved transcribed pseudogenes under selection. BMC Genomics, 10; 435.
21. Reymond, A., Marigo, V., Yaylaoglu, M.B., Leoni, A., Ucla, C., Scamuffa, N., Caccioppoli, C., Dermitzakis, E.T., Lyle, R., Banfi, S. (2002). Human chromosome 21 gene expression atlas in the mouse. Nature; 420, 582–586.
22. Elliman, S.J., Wu, I., & Kemp, D.M. (2006). Adult tissue-specifi c expression of a Dppa3-derived retrogene represents a postnatal transcript of pluripotent cell origin. *J. Biol. Chem*; 281; 16–19.
23. Olsen, M.A., & Schechter, L.E. (1999). Cloning, mRNA localization and evolutionary conservation of a human 5-HT7 receptor pseudogene. Gene; 227, 63–69.
24. Lin, M., Pedrosa, E., Shah, A., Hrabovsky, A., Maqbool, S., Zheng, D., & Lachman, H.M. (2011). RNA-seq of human neurons derived from iPS cells reveals candidate long non-coding RNAs involved in neurogenesis and neuropsychiatric disorders. PLoS ONE. 6, e23356.
25. Poliseno, L., Salmena, L., Zhang, J., Carver, B., Haveman, W.J., & Pandolfi, P.P. (2010). A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature; 465, 1033–1038.
26. Chiefari, E., Iiritano, S., Paonessa, F., Le Pera, I., Arcidiacono, B., Filocamo, M., Foti, D., Liebhaber, S.A., & Brunetti, A. (2010). Pseudogene-mediated posttranscriptional silencing of HMGA1 can result in insulin resistance and Type 2 diabetes. Nat. Commun; 1, 1–7.
27. Piehler, A. P., Hellum, M., Wenzel, J. J., Kaminski, E., Haug, K. B. F., Kierulf, P., & Kaminski, W. E. (2008). The human ABC transporter pseudogene family: Evidence for transcription and gene-pseudogene interference. *BMC genomics*, *9*(1), 1-13.
28. Lin, H., Shabbir, A., Molnar, M., & Lee, T. (2007). Stem cell regulatory function mediated by expression of a novel mouse Oct4 pseudogene. *Biochemical and biophysical research communications*, *355*(1), 111-116.
29. Zou, M., Baitei, E. Y., Alzahrani, A. S., Al-Mohanna, F., Farid, N. R., Meyer, B., & Shi, Y. (2009). Oncogenic activation of MAP kinase by BRAF pseudogene in thyroid tumors. *Neoplasia*, *11*(1), 57-65.
30. McCarrey, J. R., & Riggs, A. D. (1986). Determinator-inhibitor pairs as a mechanism for threshold setting in development: a possible function for pseudogenes. *Proceedings of the National Academy of Sciences*, *83*(3), 679-683.
31. Korneev, S. A., Park, J. H., & O'Shea, M. (1999). Neuronal expression of neural nitric oxide synthase (nNOS) protein is suppressed by an antisense RNA transcribed from an NOS pseudogene. *Journal of Neuroscience*, *19*(18), 7711-7720.
32. Hawkins, P. G., & Morris, K. V. (2010). Transcriptional regulation of Oct4 by a long non-coding RNA antisense to Oct4-pseudogene 5. *Transcription*, *1*(3), 165-175.

33. Tam, O.H., Aravin, A.A., Stein, P., Girard, A., Murchison, E.P., Cheloufi, S., Hodges, E., Anger, M., Sachidanandam, R., Schultz, R.M. and Hannon, G.J. Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. Nature (2008) 453, 534–538.
34. Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., Chiba, H., Kohara, Y., Kono, T., Nakano, T. (2008). Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. Nature; 453, 539–543
35. Guo, X., Zhang, Z., Gerstein, M.B., & Zheng, D. (2009). Small RNAs originated from pseudogenes: cis- or trans-acting? PLoSComput. Biol; 5, e1000449
36. Wen, Y.Z., Zheng, L.L., Liao, J.Y., Wang, M.H., Wei, Y., Guo, X.M., Qu, L.H., Ayala, F.J. and Lun, Z.RPseudogene-derived small interference RNAs regulate gene expression in African Trypanosoma brucei. Proc. Natl. Acad. Sci. U.S.A. . (2011) 108, 8345–8350.
37. Ross, J. (1996). Control of messenger RNA stability in higher eukaryotes. Trends Genet; 12, 171– 175.
38. Hirotsune, S., Yoshida, N., Chen, A., Garrett, L., Sugiyama, F., Takahashi, S., Yagami, K., Wynshaw-Boris, A. and Yoshiki, A. (2003). An expressed pseudogene regulates the messenger RNA stability of its homologous coding gene. Nature; 423, 91–96
39. Han, Y.J., Ma, S.F., Yourek, G., Park, Y.D., & Garcia, J.G. (2011). A transcribed pseudogene of MYLK promotes cell proliferation. *FASEB J, 25*, 2305–2312.
40. Bartel, D.P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell;* 136, 215–233.
41. Alimonti, A., Carracedo, A., Clohessy, J.G., Trotman, L.C., Nardella, C., Egia, A., Salmena, L., Sampieri, K., Haveman, W.J., Brog, I, E. (2010). Subtle variations in Pten dose determine cancer susceptibility. Nat. Genet. (2010) 42, 454–458
42. Salmena, L., Poliseno, L., Tay, Y., Kats, L., & Pandolfi, P.P. (2011). A ceRNA hypothesis: the Rosetta stone of a hidden RNA language? Cell; 146, 353–358.
43. Mattick, J.A. (2007). New paradigm for developmental biology. J. Exp. Biol. 210, 1526–1547
44. Kuo, C.H., & Ochman, H. (2010). The extinction dynamics of bacterial pseudogenes. PLoS Genet. 6, e1001050.
45. Pink. R.C., & David, R.F. (2013). Pseudogenes as regulators of Biological functions, *Biochemical Society*, *54*, 103-112.