Genes & Genomics
https://doi.org/10.1007/s13258-020-00935-7

Online ISSN 2092-9293
Print ISSN 1976-9571

REVIEW

Molecular fossils "pseudogenes" as functional signature in biological system

Rajesh Kumar Singh[1] · Divya Singh[2] · Arpana Yadava[2] · Akhileshwar Kumar Srivastava[3]

Received: 24 October 2019 / Accepted: 3 April 2020

Abstract
Pseudogenes have been known as non-functional molecular relics developed from inactivated genomic mutations while evolution. However, they have a various function at distinct molecular stages (DNA, RNA and protein) in different biological events including in cancer and associated with parental-gene-dependently and parental-gene independently. The interactions of pseudogenes with ancestor genes or other genes alter in their sequences and transcriptional processes. Pseudogene of RNA sequences have multiple functions in post-transcriptional activities as antisense RNAs, endogenous small-interference RNAs, and competing endogenous RNAs. Moreover, it also plays vital roles in controlling of its parent genes and other pseudogenes transcribing into RNA. Pseudogene transcripts make small interfering RNA or decline cellular miRNA level. The current review focuses on pseudogene functional signature in human genome by regulating the gene expression.

Keywords Pseudogenes · Gene expression · Noncoding RNA · Transcription · RNA · Cancer

Introduction

Pseudogenes have been recognised as "relics of evolution", due to the homologous to protein-coding genes but lack protein products (Jacq et al. 1977). It has been noticed that the most of the human pseudogenes do not have functions, however about 20% of them show transcriptional activity (Zhang and Zheng 2014).

The term 'pseudogene' was first given by Jacq and his colleagues explained as a deoxyribonucleic acid (DNA) sequence resembling to the frog 5S ribosomal ribonucleic acid (RNA), which had various deteriorative properties to render the nonfunctional product of RNA (Jacq et al. 1977). After that the word 'pseudogene' has been frequently used for those DNA sequences exhibiting two distinct properties: (a) the similarity of sequences to the functional gene, (b) deteriorative properties of sequence like its (RNA or protein) product is nonfunctional. Apart from above explanations, pseudogenes are also severely attenuated gene copies that could not form a functional product. Nonetheless, many prokaryotic and eukaryotic genomes have not been fully sequenced to identify the plethora of pseudogenes in living organisms for understanding their origins as well as their role in evolution. Instead of their worthless existence, pseudogenes clued about the subsequent event of DNA evolution. Presently, the significant role of pseudogenes in gene regulation has been explored and its potent uses are described to understand the processes in transcriptome evolution.

Noncoding regions of human genome are usually considered as nonfunctional and labelled as "junk" DNA. Recently, many researchers have explained that junk DNA also plays some regulatory roles in biological system. Indeed, pseudogenes are copies of gene lacking coding-sequence e.g. frameshifts and premature stop codons which has similar resemblance to functional genes (Tutar 2012).

The formation of pseudogenes has no any definite evolutionary mechanism and incapable to synthesis protein due

✉ Akhileshwar Kumar Srivastava
akhileshwar.kumar2@gmail.com

Rajesh Kumar Singh
rkszoology@gmail.com

Divya Singh
divya.ds012@gmail.com

Arpana Yadava
ayadava88@gmail.com

[1] Institute of Medical Sciences, Banaras Hindu University, Varanasi 221005, India

[2] Department of Chemical Engineering and Technology IIT, BHU, Varanasi 221005, India

[3] Department of Botany, Banaras Hindu University, Varanasi 221005, India

Published online: 10 April 2020

to these events like premature stop codon, splicing errors, frameshift-causing deletion and insertion. The pseudogenes could be form by these events (Fig. 1): (1) duplication and mutation; (2) processing of retrotransposon by insertion and inactivation of the coding ability (Poliseno 2012). Transcription of pseudogenes depends on the genomic location and it could be processed into short interfering RNAs which control coding sequences via RNAi pathway or they might interact with the promoter of parent genes or act as microRNA decoys to control the parent gene.

Formation of pseudogenes in human genome and its role in evolutionary process as well as in speciation have been widely studied. As describing the following functions of the identified pseudogenes have been explored in several types of human cancer, where pseudogenic DNA, RNA, and proteins have been differentiated the parental gene-related and unrelated functions (Fig. 2) (Poliseno et al. 2015). (1) Exchanging the genomic DNA between pseudogenes and parental genes through gene conversion or homologous recombination which promoter also affects the transcription of parental gene. (2) Pseudogene derived from RNA transcription of sense and antisense affect transcription of the parental gene epigenetically. (3) Pseudogene sense and antisense transcripts of double-stranded RNA is cleaved into endogenous siRNAs. Then esiRNAs influence the expression of parental gene at post-transcriptional level. (4) The transcribed pseudogene RNA in sense orientation compete with parental mRNA for the integrating with microRNAs, RNA-binding proteins (RBPs), and translational machinery. (5) Pseudogene proteins appear highly homologous to parental proteins; however, it expresses in a different level e.g. tissue, cellular, pathophysiological condition. It also carries gain of-function mutations and eventually which alter the function of parental proteins even if they are not completely active. (6) The accumulation of new exons at any side of processed pseudogenes or in the middle of non-processed pseudogenes attributes to differentiate the sequence, hence the function of pseudogenes determines with respect to its parental genes. (7) The mature pseudogene RNAs is transferred into microRNAs or translated into antigenic peptides.

The acquisition of a processed pseudogene within other genes generates several kinds of issues (Fig. 2). When the insertion site is an upstream intron, then processed pseudogene are co-transcribed with its host gene as a non-coding fusion transcript. If a protein undergoes for translation, then it contains only short sequence of pseudogene. The

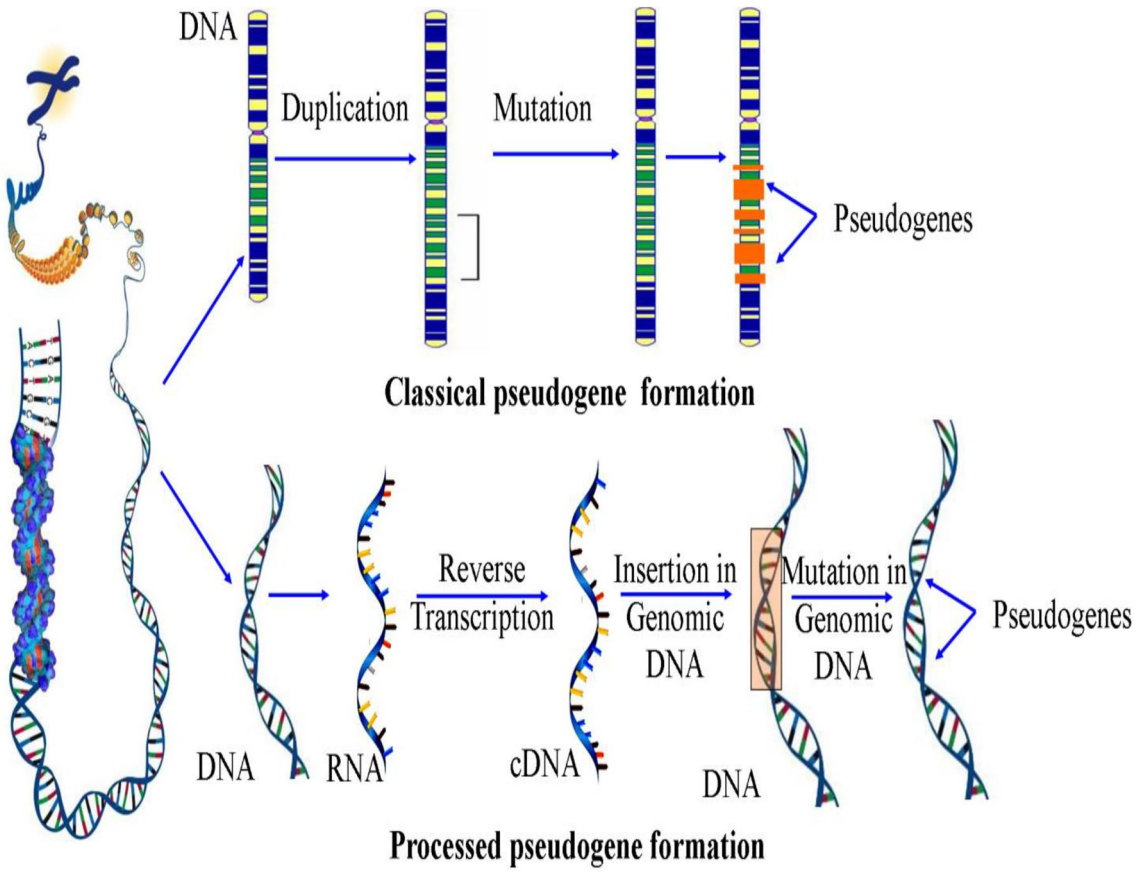

Fig. 1 The mechanism of pseudogenes evolution **a** classical pseudogene formation and **b** processed pseudogene formation

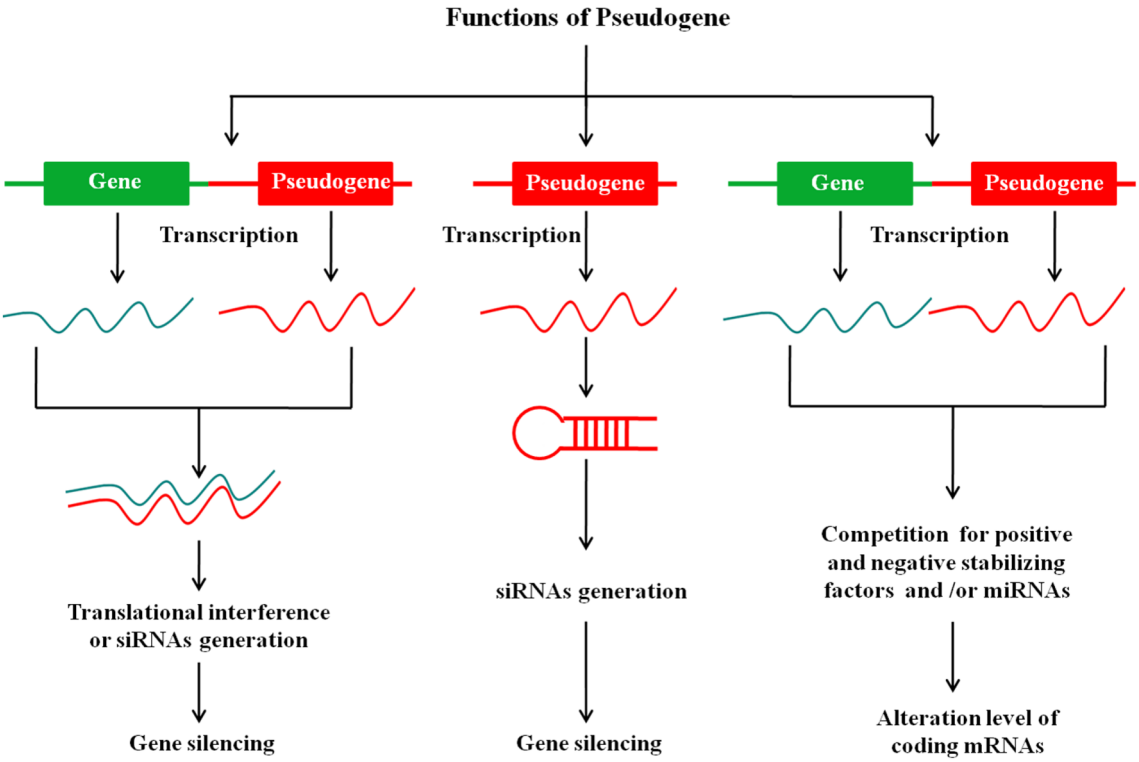

Fig. 2 Diagrammatic sketch of pseudogene functions (Poliseno et al. 2015)

insertions of pseudogenes in upstream introns also alter transcription of host gene through epigenetic silencing. If the insertion site has more downstream intron, then the processed pseudogene is co-transcribed with its host gene as a coding fusion transcript and the translated protein contains sequences of both the gene and the pseudogene. The fusion transcript of the insertion site with 3′-UTR-expressing exon displays an altered post-transcriptional regulation. The accumulation of pseudogene in a coding exon creates an insertional mutagenesis which inhibits the expression of host gene. Pseudogenes and adjacent genes are transcribed jointly through read-transcripts and translated into chimerical proteins. Despite of parental genes, the pseudogene RNAs as source for esiRNAs also affect other unrelated genes.

The pseudogenes are considered as genomic loci which similarity meet to functional genes, although it has not biologically significant because they harbour mutations and interfere their transcriptional and translational activities (Poliseno et al. 2010). The advance research and next-generation sequencing have explored noncoding RNAs, multi-steps functions of pseudogenic DNA, RNA or protein in several cancers. Pseudogenes have significant functions in controlling of transcriptional and post-transcriptional events and also they are capable to produce the novel genes, hence acting as a source for gene renewal (Jeter et al. 2009).

In the last two decades, the different types of regulatory RNAs viz. long RNAs and short RNAs (< 200 nucleotides) have not been identified for protein coding. Now, the long noncoding RNAs (lncRNAs) is classified as intronic, intergenic ncRNAs and natural antisense transcripts (NATs) (Gutschner and Diederichs 2012). The tiling arrays and high throughput sequencing approaches revealed only 1–2% sequences among > 90% of the transcribed human genome have ability to encode proteins (Carninci 2008). These outcomes indicate the ncRNAs may involve highly in biological functional in comparison to protein-coding ones that gives a sign to be annotated of the noncoding transcribed genome.

The contribution of certain pseudogenes has provided the stature of underappreciated molecules a new class of regulatory lncRNAs participating in physiological and pathological events (Wang et al. 2006).

To date around 20,000 pseudogenes have been identified and which are similar to the number of protein-coding genes (~ 27,000) in human (Han et al. 2011). The present knowledge of genes are ambiguous hence several sequences once believed inactive have normally functional RNA genes playing roles in gene silencing either by generating siRNAs or by altering mRNA levels of functional protein-coding gene (Sasidharan and Gerstein 2008). Many studies emphasized on the pseudogene population and its regulatory roles are not able to cover all aspects together. To understand the

adaptation capability of pseudogenes for survival by comparing the genes from different organisms is very attractive. Pseudogenes give a clue of all changes in the genome of specific organism. The current review describes the functional activity of pseudogenes in biological system by regulating the expression of functional genes.

Evolution causing to pseudogenisation

Pseudogenisation is a process of the formation of pseudogene from functional gene during evolution. The most of events for pseudogenisation are neutral that happen by arbitrary deposition of mutations in genes where the functional activities are relaxed and sometime it could be adaptive. The inactivation of older functional gene stimulates the organism fitness to the changing environment or genetic background viz: adaptive pseudogenisation in the human cysteine-aspartic protease 12 (CASPASE12) gene (Wang et al. 2006). CASPASE12 encoding a cysteinyl aspartate proteinase plays role to inactivate the immune response for endotoxins. A null CASPASE12 allele with a point mutation creating a premature stop codon is carried by human populations residing outside of Africa and 10% of the Africans and they have functional allele so far. Epidemiological reports revealed that the null allele is responsible to reduce the cause for severe sepsis (Saleh et al. 2004), indicating that the spreading of null allele might have been stimulated by positive selection (Wang et al. 2006).

Although, pseudogenisation does not create immediate beneficial effect; it opens up for the future evolutionary possibilities or events of future evolutionary pathways. The lack of sarcomeric myosin heavy chain 16 gene MYH16 while human evolution had not beneficial effect during pseudogenisation process and later on it made consequent modifications in the cranial morphology, which was not possible earlier (Stedman et al. 2004). Studies related to pseudogenisation could also be significant dating for evolutionary changes in particular phenotypic traits if the relationship of gene-trait is clear.

The existence of many numbers of a gene copy in human genome produces gene variants that may have some novel genes, but they usually cause to birth for pseudogenes. Pseudogenes could be produced from gene mutations, or random gene duplications, or retrotransposition of processed mRNAs back into the genome. Pseudogenes are classified into three types: (1) unitary pseudogenes, (2) duplicated or unprocessed pseudogenes and (3) processed or retrotransposed pseudogenes.

Unitary pseudogenes are resulted from spontaneous mutations in coding sequence abrogate either transcription or translation of the certain gene. Unitary pseudogenes do not have complete functional counterparts, is referred as ancestral, cognate, or parental genes. Duplicated pseudogenes come from irregular gene duplication causing for the loss of promoters/enhancers or frameshift mutations or premature stop codons and making them inactive, while parental genes remain functional. Duplicated pseudogenes are found around the parental genes. The unitary and duplicated pseudogenes have intron–exon structures, however, the processed pseudogenes from mRNA do not have introns and are reverse-transcribed into DNAs and eventually reverted back into the genome at a new position (Xiao-Jie et al. 2015).

Formation of a pseudogene (lncRNA) from Junk DNA

The word “junk DNA” has been popularised since 2001 at the time of the Human Genome Project. It has been deciphered that only a small portion of the DNA sequences of the human genome encodes amino acid sequences in proteins synthesis and all those sequences in the genome which do not encode amino acid in proteins through genetic code is referred as “junk DNA” (Mathews 2017).

The Encyclopedia of DNA Elements (ENCODE) project has accounted that maximum 75% of the human genome is transcribed into RNAs and only 3% of the human genome encode amino acids for protein synthesis. Due to a long-held protein-centered bias, several genomic sites are transcribed into non-coding RNAs (ncRNAs) was identified as ‘junk’ in the genome associated with transcription regarding as transcriptional ‘noise’ viz lacking biological function (Comings 1972).

Many lncRNAs (long non-coding RNAs) are transcribed from ‘junk’ regions, and even those holding transposons, pseudogenes and simple repeats also have essential regulators with biological importance (Gong and Maquat 2011).

The interaction of gene promoters with transcription factors by action of RNA polymerases switch on for the transcription (Levine and Tjian 2003). Recently, the ncRNA transcripts have been identified within the promoter site of many genes (Core et al. 2008; Seila et al. 2008) and pointed that more complicated regulatory mechanisms should be considered for the study. By using a tiling microarray, it has been focused on the ncRNAs mapping in the transcription start site of 56 cell-cycle-related genes showed high transcriptional activity in the gene promoter site lacking the protein-coding sequences. Among these lncRNA, the non-spliced 1.5-kb ncRNA PANDA was transcribed from 5 kb upstream of the CDKN1A transcription start site and its biological activity was recognised as in the DNA damage response. Although CDKN1A acts as mediators in cell cycle arrest, PANDA enhances cell survival against DNA damage by checking the transcription factor NF-YA from binding the particular promoters

of apoptosis-inhibiting genes (Hung et al. 2011). This suggest that the both cell cycle arrest and anti-apoptotic genes could be induced from the same locus by following DNA damage response and this way a complex network could determine the biological phenotypes. In other study, it has been shown that a promoter-associated lncRNA complementary to the rRNA gene promoter interacts to rRNA gene to form a lncRNA-DNA triplex and this triplex impedes the binding of transcription termination factor 1 to the rRNA gene and selects DNMT3b to silence it (Schmitz et al. 2010).

The junk DNA is not only cause for the loss of activity and also develops a functional treasure in biological systems. The advance researches in lncRNA showed a transcriptional treasure that is derived from the region of junk DNA. However, to date only a small portion of the lncRNAs have been functionally identified, it is believed that functional treasure of lncRNAs would be expanded by many advance technologies for highthroughput screening and functional affirmation. The interactive proteins bind with the transcriptome data could be highly promoted by photoactivatable ribonucleoside increased crosslinking and immunoprecipitation (PAR-CLIP) (Hafner et al. 2010); genomic sites of lncRNAs could be identified by ChIRP and subsequent DNA sequencing (Chu et al. 2011); active motifs of RNA could be determined by RNA-mechanically induced trapping of molecular interactions (RNA-MITOMI) (Martin et al. 2012); RNA dynamics could be assessed by live imaging using engineered fluorescent RNAs (Paige et al. 2011). Due to the presence of large number of lncRNAs in the human genome, it could be targeted first on the disease-associated lncRNAs by approaching to multiple experiments such as expression analysis. The disease-related SNPs are useful markers to explore the function of lncRNAs. Moreover, the occurrence of lncRNAs in such regions are either functionally affected or altered in their expression levels by certain SNP variants and hence it can be the culprits underlying the mechanisms of disease predisposition. Description of such mechanisms requires an elaborative information on lncRNA structure, structure–function relationship and a convenient experimental system to know the small differences.

After exploration of tissue-specific expression patterns and site-specific action of lncRNAs, the drugs targeting lncRNAs have achieved more comprehensive therapeutic potential than conventional drugs. Moreover, the allele-specific regulatory mechanisms of lncRNAs can be used for the controlling of the certain gene expression with minimum adverse effects. Synthetic oligonucleotides having locked nucleic acid modifications with high affinity and specificity are employed for the targeted regulation of lncRNA expression. Unequivocal small molecule chemical compounds targeting to lncRNA could also be examined as candidates to hurdle in the lncRNA-protein interaction, or altering the lncRNA loading onto its target genomic sites (Ling et al. 2015).

The regulatory mechanism in human cells is very complexed, hence it is very difficult to understand the complete disease phenotype by a single molecule. It could be imagined that in a specific genomic locus may have various kinds of intertwined transcripts like protein-coding genes, overlapping intronic and noncoding RNAs in the sense or antisense orientation associated with protein-coding genes. Other complication in regulatory mechanism are also observed in human cells with several isoforms caused by alternative splicing. Hence a loss or gain of a genomic region are immediately observed in cancer and affect to DNA regulatory factors and also in transcriptional region. So, it is very important to understand the genetic aspects like gene locus, adjacent genes, chromatin status and target genomic site, for a functional annotation or therapeutic modifications to check the cancer (Ling et al. 2015).

Functional activity of pseudogenes

Despite parent gene, the pseudogenes regulate the function of other gene expression at transcriptional and post-transcriptional level. The interaction of pseudogene with a gene promoter viz antisense RNA of pseudogenes combines with sense-stranded mRNA of homologous parent gene at the transcriptional level thereby either cease translation or stimulate for siRNAs formation which could suppress the expression of the parent gene. The post-transcriptional events are controlled by pseudogene as functional microRNA decoys, also called as competing endogenous RNA (ceRNA). Eventually, the positive or negative effect on parent gene mRNAs are produced from competitive activity of RNA of parent genes and its homologous pseudogenes for RNA binding proteins (RBPs) and such effects ultimately depend on the functional nature of RBPs. The alteration in levels of pseudogene transcripts lead to changes in the parent gene mRNA levels. Hence, it has been explained the function of pseudogenes as positive or negative regulators of gene expression.

Positive gene regulators

It seems a positive relation of pseudogene with parent gene. A well-known mechanism for such activity of the pseudogene is described as: gene regulation by ceRNA through which the pseudogene transcripts contributes same microRNA response elements (MREs) and compete with the parent transcripts for same microRNAs. This strong evidence explains about a new event of post-transcriptional regulation occurring in a different organisms (Bossi and Figueroa-Bossi 2016; Thomson and Dinger 2016). This regulatory system

shows that the multiple RNA transcripts of pseudogenes and parental genes have the shared MREs for common microRNAs and hence, these transcripts are capable to co-regulate one another. These evidences also suggests that a number of lncRNAs involve in ceRNA regulatory network (Zhang et al. 2013).

Negative gene regulators

As stated above, the stabilized RBPs like HuR interact to AREs to make the mRNA stabilized. If pseudogenes could compete with the parent genes for stabilizing mRNAs, depending on RBPs, then it could be expected that the parent gene mRNA may likely be less efficient so that the mRNA level declines. Such competitive behaviour of pseudogenes function acts as negative gene regulators.

The potential role of pseudogene is described from several evidences explaining that the pseudogenes bind with its functional genes participates in multiple biochemical events of cells. The pair of genes could regulate the overexpression of a functional gene mRNA of a pluripotency-associated transcription factor like Oct4 pseudogene transcript suppresses cell differentiation. The downregulation of Oct4 pseudogene RNA antisense enhances the Oct4 level and its two pseudogenes. These examples provide evidence the pair of gene regulates the expression of functional gene by combining antisense pseudogene transcripts with sense genic transcripts (Han et al. 2011).

Small interfering RNA (siRNA) also plays a vital role in gene expression. It has been reported in mouse oocytes that siRNA repressing the gene expression is formed from the folded pseudogenes transcripts through develops hairpin structures. The reduction of siRNA producer protein (Dicer) decreases the levels of pseudogene-derived siRNAs and stimulates the coding gene mRNAs, which indicated toward the siRNA-dependent regulation.

The most efficient mechanism of pseudogene activity is regulating the mRNA stability along with interference factors. The interaction of transacting molecules with cis-acting sequences of mRNA provides mRNA stability. The resemblance of cis-acting sequences of pseudogenes to functional gene compete for trans-acting molecules, which reduces the mRNA stability along with expression (Pink et al. 2011).

Micro RNAs (miRNA) also influence on the mRNA stability by attaching with 3′ untranslated region of mRNA. The miRNAs decrease the level of gene expression by degrading the mRNA. PTEN is a tumor suppressor and retaining it at definite level playing role in prevention of oncogenesis. The coupled miRNAs regulate both the gene PTEN and pseudogene PTENP1 together. PTENP1 pseudogene interacts with miRNA and decline the level of miRNA in cells, which provokes PTEN to escape from miRNA repression regulation (Pink et al. 2011).

The above evidences explain that pseudogenes silence the gene by playing significant role in translational interference/siRNA generation. Moreover, protein coding mRNA and its related pseudogenes could compete for the stabilizing factors and/or miRNAs, which changes in protein coding mRNA expression levels.

Processed pseudogenes in human genome of both sexes

The large numbers of processed pseudogenes are explored in human genome with approximately 33% higher in X chromosome than in autosomes indicating that the events for human oogenesis take longer period than non-mammalian species. The processed pseudogenes associated with human Y chromosome contains three origin sites (a) inherited chromosome from the ancestor, which develops the Y chromosome, (b) transferred chromosome between the X and Y chromosomes regions of pseudoautosomal and X-transposed sites and (c) retrotransposed of another chromosomes (Maranda et al. 2019).

The processed pseudogenes are produced from genes of the Y chromosome while spermatogenesis and they integrate themselves with many chromosomes. It has been suggested that about 3 times more density of processed pseudogenes are found in the euchromatin of the X chromosome than on the Y chromosome (Maranda et al. 2019).

The lower substrate specificity of L1 reverse transcriptase is peculiar reason for abundance of processed pseudogenes in human genome which causes for the retrotranscription of mRNA into processed pseudogenes and long-lasting production of male and female gametes. Male and female germ cells cause for the large number of production of processed pseudogenes in human genome (Maranda et al. 2019).

Figure 3, the euchromatic region is comprised into following regions: the pseudoautosomal regions, the X-transposed region, the X degenerate regions and the ampliconic regions (Maranda et al. 2019). The pairing and recombination of Y chromosome with X chromosome in course of spermatogenesis at particular regions called as pseudoautosomal regions (PARs). The PARs short homologous regions located in the sex chromosomes, lacking strict pseudoautosomal boundaries which reintegrate at a rate ~ 20 times the genome average (PAR1) and ~ 5 times the genome average (PAR2). The XTR region of the Y chromosome was transposed from the X chromosome to the Y chromosome approx. 3–4 million years ago, after the deviations of humans and chimpanzees were appeared. Veerappa et al. (2013) explains that some parts of XTR also involve in seldom recombination because of gene conversion (Veerappa et al. 2013). The remnants of X-degenerate sequences derived from the autosome lead for evolving of the Y chromosome (Skaletsky et al. 2003).

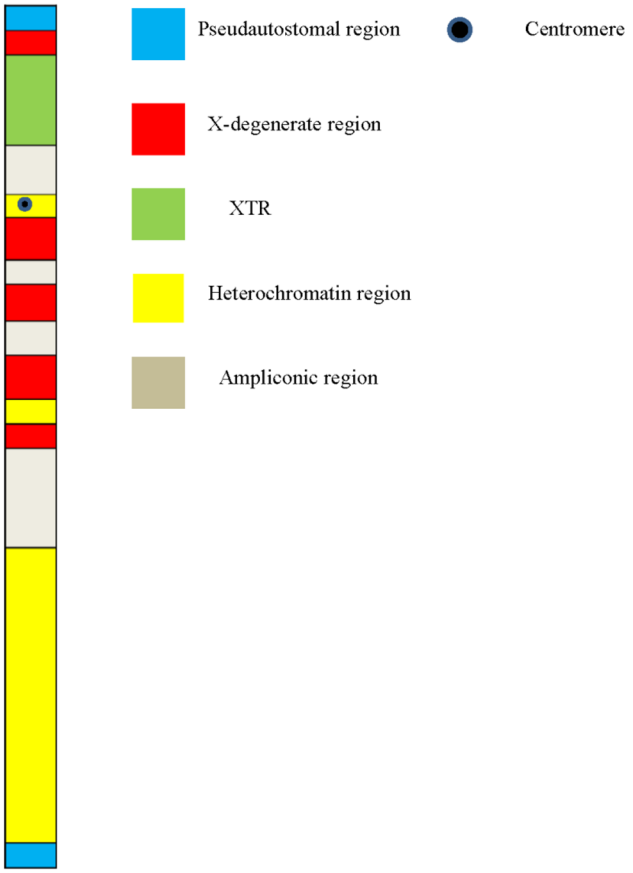

Fig. 3 The picture showing the Y chromosome regions with three heterochromatic and four types of euchromatic sites: the two pseudoautosomal regions, the X-transposed region (XTR), the six X-degenerate regions and the three ampliconic regions (Maranda et al. 2019)

The last categories are ampliconic regions associated with amplified sequences maintaining their very high identity due to with subsequent gene conversions.

The 18 processed pseudogenes developed from the ancestral chromosome are also occurred on the X chromosome in same pattern and leading to the production of the Y chromosome (Table 1) (Skaletsky et al. 2003). The 18 processed pseudogenes have been identified in mainly three regions of chromosome like: three in the pseudoautosomal regions, nine in the XTR and six in X-degenerate regions. The pseudoautosomal regions, PAR1 and PAR2, have one and two processed pseudogenes, respectively. Such frequent recombination between these locations is not able to distinguish of specific pseudogene from one another. Nine extra pseudogenes are explored in the region of XTR. The processed pseudogenes are very similar in the region of the X and Y chromosomes. It has been observed an occasional recombination (gene conversion) between the identified sequences in the XTR region of the X and Y chromosomes. Another six processed pseudogenes on the Y chromosome evolved from the ancestral X chromosome and are differentiated one the basis of their location in three of the four groups remnant of the inversions that hinders the recombination between the X and Y chromosomes.

Table 1 The common processed pseudogenes of X and Y chromosomes (Maranda et al. 2019)

Name of gene	Expanded name of gene	Region
FABP5	Fatty acid binding protein 5	PAR1
USP12	Ubiquitin specific peptidate 12	XTR
RNF19B	Ring finger protein 19B	XTR
SERBP1	Serpine 1 mRNA binding protein 1	XTR
VDAC1	Voltage dependent anion channel 1	XTR
EIF4A1	Eukaryotic translation initiation factor 4A1	XTR
KRT18	Keratin 18	XTR
RPL26	Ribosomal protein L26	XTR
DLGAP5	DLG associated protein 5	XTR
SRI	Sorcin	XTR
GOT2	Glutamic-oxaloacetic transaminase 2	Group 4
ASS1	Argininosuccinate synthase 1	Group 4
RPS24	Ribosomal protein S24	Group 4
AGK	Acylglycerol kinase	Group 4
GAPDH	Glyceraldehyde-3-phosphatedehydrogenase	Group 3
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase	Group 2
AMD1	Adenosylmethionine decarboxylase 1	PAR2
TRPC6	Transient receptor potential cation channel	PAR2

Relation of pseudogenes with parental genes

About 2% protein-coding genes are found in human genome and ~ 95% genes are considered as evolutionarily remnants that is often referred a 'junk DNA'. Many studies explain that most junk DNA is transcribed in several spatiotemporal circumstances. Inspite of tRNA and rRNA, the most transcripts like lncRNA, miRNA, pseudogene RNA and circular RNA are non-coding.

Pseudogenes are often known as relics of their parental genes and lack encoding function causing for mutation in regulatory elements and encoding regions (Salmena 2014). The gene amplification while evolution increased the copy number of gene (Elliott et al. 2013). Some genes might not have been involved for coding protein at specific time causing to various mutations and deterioration into pseudogenes. During regulation of transcription level, lncRNA involves either in silencing of transcription or activating function of genome by recruiting the epigenetic modifiers (Ling et al. 2015) and pseudogenes regulate mostly parental gene expression through binding of shared miRNAs. At the epigenetic level of expression, the pseudogenes regulate the

DNA methylation appearing independently apart from the parental genes. Nonetheless, the DNA methylation level is associated with the status of local DNA micro-environment (Davis et al. 2015). Hence, the pseudogenes are not only caused for revolution and also, they are positively selected to remain and have significant roles in gene expression and regulation.

Pseudogene transcripts give rise non-coding RNA and antisense RNA acting as RNA sponges for miRNA (Chan and Chang 2014). Recently, it has been investigated that the PTEN pseudogene (PTENP1) up regulated the PTEN expression by competing to miRNAs shared with PTEN (Poliseno and Pandolfi 2015) and gave PTEN repression in a DICER-dependent manner (Guo et al. 2016). It is interesting outcomes the two anti-sense lncRNA (asRNAa and asRNAb) occurring on the region of pseudogene PTENP1 also regulate PTEN expression. The PTEN transcription is downregulated epigenetically when asRNAa attaches to PTEN promoter region recruiting DNMT3a and EZH2. However, asRNAb stabilizes PTENP1 by binding to PTENP1 transcript and also upregulates to trans PTEN expression through miRNA and PTENP1 (Johnsson et al. 2013). Anti-sense transcription might have also associated with other pseudogenes. Pseudogene FLT1P1 of VEGF receptor-1 (VEGFR1) controls the function of parental gene VEGFR1, whereas expression of FLT1P1 anti-sense suppresses the expression of both gene VEGFR1 and VEGF-A. The knockdown of FLT1P1 expression inhibits the tumour cell proliferation. The other anti-sense pseudogene transcript of Oct4 also suppress the expression of Oct4, Oct4 and Oct5 pseudogenes (Hawkins and Morris 2010). It has been suggested that the pseudogene transcripts of sense and antisense participate in epigenetic regulation of specific genes.

Also human cytochrome P450 gene CYP2A6 and its pseudogene CYP2A7 compete for binding of miR-126 that enhance the expression of each other (Nakano et al. 2015). These indicate that the pseudogene transcripts also increase the expression of the protein-coding genes through competing with shared miRNAs. The microRNAs targeting 30UTR of mRNA regulate gene expression through repressing the translation and reveals the overexpression 30UTR of pseudogene CYP4Z2P enhanced CYP4Z1 expression (Zheng et al. 2015). Hence, the main function of pseudogene transcripts is to regulate the expression of parental gene by accumulating common miRNAs and releasing expression inhibition.

Moreover, the pseudogene transcripts also suppress the expression of parental genes by targeting other factors. The complexes of Oct4 pseudogene and Oct4P4 with SUV39H1 HMTase and HP1a that suppresses the expression of parental gene Oct4 epigenetically (Scarola et al. 2015). In this way, the pseudogene has similar role as lncRNAs, which forms a complex with other RNAs and genomic modifiers for modulation of DNA transcriptional activity epigenetically.

The transcripts of other pseudogenes encode smaller peptides involved in regulation of parental gene expression. Even if CLRX.1/NOD24 pseudogene NLRP2P lacking complete length of coding region encodes a 45-amino-acid protein is highly homologous to Pyrin-only protein 2 (POP2/PYDC2), a regulator of NFkb. In this way, NLRP2P peptide suppresses transcriptional activity of NFkb and inhibits NF-kb expression (Porter et al. 2014). Several pseudogenes encode a smaller peptide because of earlier stop codons, but it is further required to explore their biological role.

Regulation of oncogenic factors by pseudogene

The high homology of pseudogenes to parental genes facing challenges in studies is in differentiation of their homologous genes from their parental genes and individual genome differences and sequencing. Now a days, several approaches have been explored for pseudogenes differentiations at DNA or RNA level (Karro et al. 2007) and reported multiple functions of pseudogene DNAs, RNAs or proteins in different cancers.

The pseudogenes with certain properties involve in pathogenesis of human cancer and exhibit cancer-specific expression (Kalyana-Sundaram et al. 2012). The occurrences of transcripts of pseudogenes are higher in the cancer tissue than in the normal tissue (Hwang et al. 2005). In some cases, transcripts of pseudogene are detected only in cancer, not in normal tissue.

The inert pseudogenes are subjected to irregular genetic drift, however, the present studies have provided evidences for evolutionary conservation of the pseudogene in different mammalians (Khachane and Harrison 2009). The nonsynonymous to synonymous substitution rate (Ka/Ks) is applied for determination the evolutionary constraint of sequence. The Ka/Ks is less than one when the sequence is under pure selection, equal to one for neutral evolution and greater than one for positive selection. Theoretically, non-functional sequences are kept for neutral selection which Ka/Ks ratios could be equal to one. However, many reports suggested that overlapping of Ka/Ks values between genes and pseudogenes indicating the some pseudogenes are under evolutionary constraint rather than evolving neutrally, but having support as functional units (Bischof et al. 2006).

The dysregulation of pseudogenes act as oncogenic factor for developing cancer such as KRASIP (Poliseno 2012) and also play as major factors in cancer progression (Pink et al. 2011). Like, NANOG and OCT4 transcription factors maintain the pluripotency in embryonic stem cells (Takahashi et al. 2007) and their pseudogenes (NANOGP1 and POU5F1P1) are unfaithfully expressed in human cancers (Cantz et al. 2007). Poliseno et al. (2010) reported that the

pseudogene PTENP1 regulates the expression of tumor suppressor PTEN by interacting microRNA involving in biological events of cancer development (Poliseno et al. 2010). The BRAF pseudogene as a competitive endogenous RNA (ceRNA) developed lymphoma in animal model (Karreth et al. 2015). These evidences provide key insights into the major role of pseudogenes in cancer biology. However, it is limited individual pseudogenes, and expecting that more pseudogenes are involve in cancer programs. So, it requires performing a subsequent analysis in large sample size of patient to explore cancer-related pseudogenes. It has been identified in 13 cancer causing RNA-Seq resource of 293 samples showing correlation of pseudogene expression with cancer progression (Kalyana-Sundaram et al. 2012). Although pseudogenes play role as microRNA sponges and compete with mRNAs to attract microRNAs for interactions and influence the expression of mRNAs (Tay et al. 2014), however the biological characteristic and clinical relevance of pseudogenes act as ceRNAs are not well understood.

Lineage of cancer-specific pseudogene expression

The characterization of pseudogene expression has an analytical challenge held by through very close sequence similarity with its respective coding genes. It has been described a consecutive analysis of pseudogene "transcription" from an RNA-Seq of 293 samples of cancer and normal tissue, surprisingly indicating the wide expression of pseudogenes at genomic level was identified as ubiquitously expressed or lineage and cancer specific (Fig. 4). By integrating these

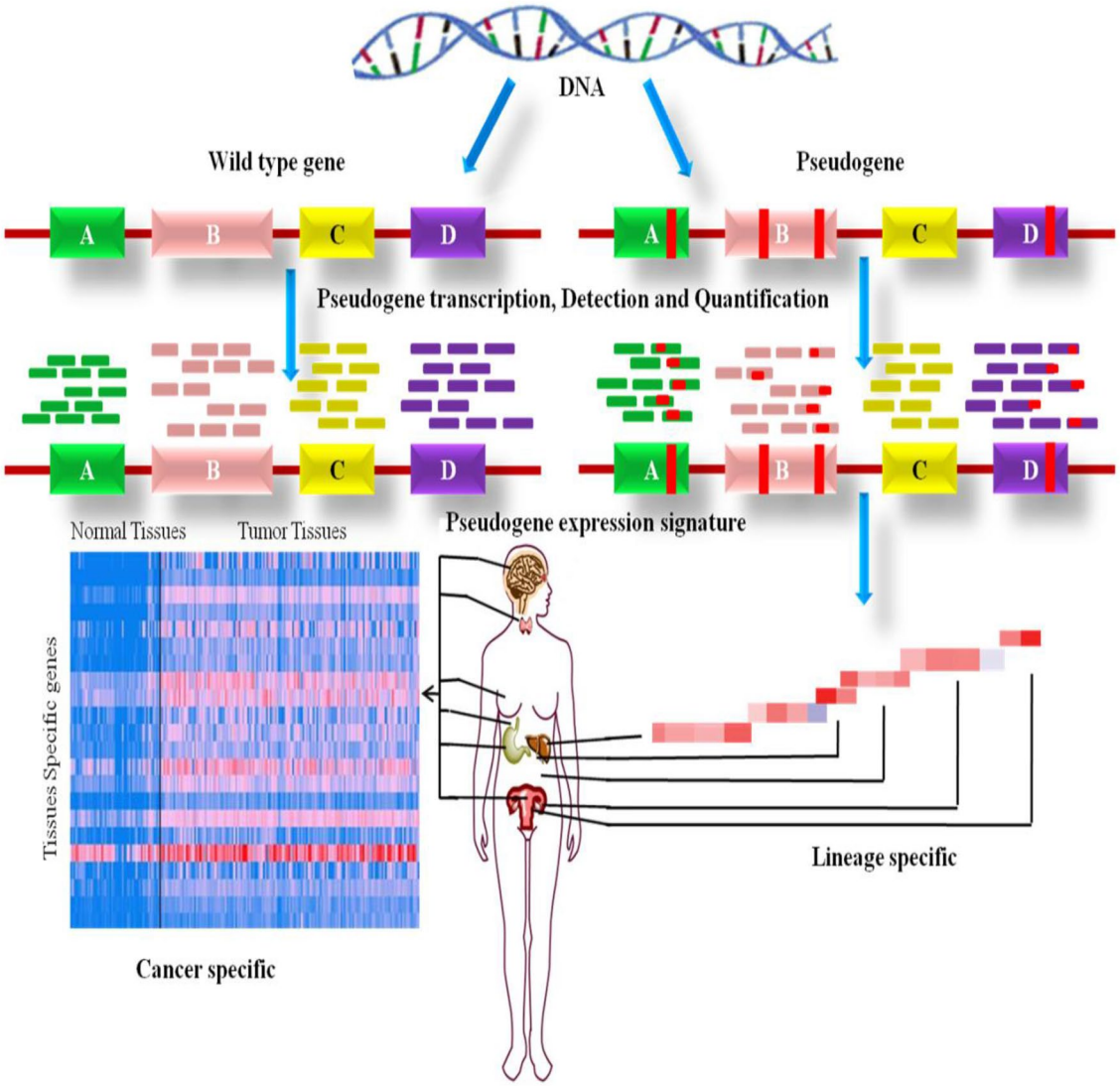

Fig. 4 A comparative presentation of the parental genes (wild-type) with respective pseudogenes on chromosomal site and profiling of lineage-specific pseudogenes expression (Kalyana-Sundaram et al. 2012)

outcomes, it has provided an evidence that transcribed pseudogenes play an essential role in transcriptional landscape of cells which are positioned to induce for cellular differentiation and cancer progression significantly (Kalyana-Sundaram et al. 2012).

The morphology of a growing body indicates the potential role of pseudogenes in controlling the cognate wild type gene expression by acting as a source for endogenous siRNA, competitive inhibitors of translation of wild-type (WT) transcripts, and dominant-negative peptides. Figure 4 shows that the pseudogene transcription has regulated the cognate wild-type gene expression by sequestering miRNAs (Kalyana-Sundaram et al. 2012).

The technical and analytical factors altering the yield of pseudogene transcripts have no significant correlation between similarity of pseudogene-WT pairs and pseudogene yield. The percent similarity is detected from gap penalty and mismatches in sequences. The distribution of the mismatches differentiates the pseudogenes from closely similar wild-type sequences. Hence, three main factors detect the pseudogene transcription by RNA-Seq: (a) level of the pseudogenes expression (b) the depth of RNA sequencing, and (c) complete distribution of mismatches in comparison to the wild-type (Kalyana-Sundaram et al. 2012).

Lineage-specific pseudogene transcripts are highly efficient for the lineage-specific functions which exhibit biological properties as novel resources to differentiate tissue types. In an experiments, the maximum pseudogenes have been detected in both cancer and benign tissues suggesting that cancer samples have transcriptional patterns of pseudogenes in both lineage and cancer specific (Kalyana-Sundaram et al. 2012).

Conclusions and future prospective

The existence of pseudogenes in human genome ascertains that they could have a major role in basic physiology of biological system and disease development. Pseudogenes are understood as 'junk DNA' or 'genomic fossils' that are neither transcribed nor translated into functional proteins. However, some recent studies suggest that the pseudogenes are far more than 'junk' or 'non-functional'. These have vital roles at different levels (DNA, RNA and protein) in health and various severe diseases progression including cancer. The pseudogenes are also considered as a reservoir for gene evolution and protein diversity. Pseudogenes alter the sequences and transcriptional activities by interacting with parental genes or other gene loci. The capability of pseudogenes for encoding proteins mimic or interfere to functions of their parental counterparts which shows no major differentiation between genes and pseudogenes (Ishiguro et al. 2012). Moreover, the close homology of pseudogenes with DNA, RNA and protein sequences produces major hindrance in researches to distinguish between pseudogene and parental gene.

During results interpretations on the functional experiments of pseudogenes, the caution must be exercised because in a non-translated pseudogenes encode for truncated proteins (Zhang et al. 2006). Moreover, some pseudogenes regulate function of their protein-coding cousins that are mediated by noncoding RNAs derived from active pseudogenes. Whereas all pseudogenes do not exhibit the biological functions and, however an unexpected regulatory positive result from the formation of a pseudogene are observed and their effect will be conserved. The major part of pseudogenes has been ignored in the quest to know the biology of health and disease, to date the pseudogene probes are often not available in commercial microarrays. Several evidences stated that dysregulation of pseudogenes produce sever diseases like diabetes and cancer. The above explanations indicate that the evolution of pseudogenes could be integrated with mutation and environmental changes. So, the understanding the evolutionary process of pseudogenes could unravel the several mysterious things associated with alteration of physiology and disease development in biological system of organisms.

Acknowledgements University grant commission (India) has supported financial assistance during research work.

References

Bischof JM, Chiang AP, Scheetz TE et al (2006) Genome-wide identification of pseudogenes capable of disease-causing gene conversion. Hum Mutat. https://doi.org/10.1002/humu.20335

Bossi L, Figueroa-Bossi N (2016) Competing endogenous RNAs: a target-centric view of small RNA regulation in bacteria. Nat Rev Microbiol 14:775–784

Cantz T, Key G, Bleidiβel M et al (2007) Absence of OCT4 expression in somatic tumor cell lines. Stem Cells. https://doi.org/10.1634/stemcells.2007-0657

Carninci P (2008) Non-coding RNA transcription: Turning on neighbours. Nat Cell Biol. https://doi.org/10.1038/ncb0908-1023

Chan WL, Chang JG (2014) Pseudogene-derived endogenous sirnas and their function. Methods Mol Biol. https://doi.org/10.1007/978-1-4939-0835-6_15

Chu C, Qu K, Zhong FL et al (2011) Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. Mol Cell. https://doi.org/10.1016/j.molcel.2011.08.027

Comings DE (1972) The structure and function of chromatin. Adv Hum Genet 3:237–431

Core LJ, Waterfall JJ, Lis JT (2008) Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science 5(80):5. https://doi.org/10.1126/science.1162228

Davis AP, Benninghoff AD, Thomas AJ et al (2015) DNA methylation of the LIN28 pseudogene family. BMC Genom. https://doi.org/10.1186/s12864-015-1487-3

Elliott KT, Cuff LE, Neidle EL (2013) Copy number change: evolving views on gene amplification. Future Microbiol 8:887–899

Gong C, Maquat LE (2011) LncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 39 UTRs via Alu eleme. Nature. https://doi.org/10.1038/nature09701
Guo X, Deng L, Deng K et al (2016) Pseudogene PTENP1 suppresses gastric cancer progression by modulating PTEN. Anticancer Agents Med Chem. https://doi.org/10.2174/1871520615666150507121407
Gutschner T, Diederichs S (2012) The hallmarks of cancer: a long non-coding RNA point of view. RNA Biol 9:703–719
Hafner M, Landthaler M, Burger L et al (2010) Transcriptome-wide Identification of RNA-binding protein and MicroRNA target sites by PAR-CLIP. Cell. https://doi.org/10.1016/j.cell.2010.03.009
Han YJ, Ma SF, Yourek G et al (2011) A transcribed pseudogene of MYLK promotes cell proliferation. FASEB J. https://doi.org/10.1096/fj.10-177808
Hawkins PG, Morris KV (2010) Transcriptional regulation of Oct4 by a long non-coding RNA antisense to Oct4-pseudogene 5. Transcription. https://doi.org/10.4161/trns.1.3.13332
Stedman HH, Kozyak BW, Nelson A et al (2004) Myosin gene mutation correlates with anatomical changes in the human lineage. Nature 428:415–418
Hung T, Wang Y, Lin MF et al (2011) Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. Nat Genet 43:621–629
Hwang SL, Chang JH, Cheng CY et al (2005) The expression of rac1 pseudogene in human tissues and in human brain tumors. Eur Surg Res. https://doi.org/10.1159/000084540
Ishiguro T, Sato A, Ohata H et al (2012) Differential expression of nanog1 and nanogp8 in colon cancer cells. Biochem Biophys Res Commun. https://doi.org/10.1016/j.bbrc.2011.10.123
Jacq C, Miller JR, Brownlee GG (1977) A pseudogene structure in 5S DNA of *Xenopus laevis*. Cell. https://doi.org/10.1016/0092-8674(77)90189-1
Jeter CR, Badeaux M, Choy G, Chandra D, Patrawala L et al (2009) Functional evidence that the self-renewal gene NANOG regulates human tumor development. Stem Cells 27:993–1005. https://doi.org/10.1002/stem.29
Johnsson P, Ackley A, Vidarsdottir L et al (2013) A pseudogene long-noncoding-RNA network regulates PTEN transcription and translation in human cells. Nat Struct Mol Biol. https://doi.org/10.1038/nsmb.2516
Kalyana-Sundaram S, Kumar-Sinha C, Shankar S et al (2012) Expressed pseudogenes in the transcriptional landscape of human cancers. Cell. https://doi.org/10.1016/j.cell.2012.04.041
Karreth FA, Reschke M, Ruocco A et al (2015) The BRAF pseudogene functions as a competitive endogenous RNA and induces lymphoma in vivo. Cell. https://doi.org/10.1016/j.cell.2015.02.043
Karro JE, Yan Y, Zheng D et al (2007) Pseudogene.org: a comprehensive database and comparison platform for pseudogene annotation. Nucleic Acids Res. https://doi.org/10.1093/nar/gkl851
Khachane AN, Harrison PM (2009) Assessing the genomic evidence for conserved transcribed pseudogenes under selection. BMC Genom. https://doi.org/10.1186/1471-2164-10-435
Levine M, Tjian R (2003) Transcription regulation and animal diversity. Nature 424:147–151
Ling H, Vincent K, Pichler M et al (2015) Junk DNA and the long non-coding RNA twist in cancer genetics. Oncogene 34:5003–5011
Maranda V, Sunstrum FG, Drouin G (2019) Both male and female gamete generating cells produce processed pseudogenes in the human genome. Gene. https://doi.org/10.1016/j.gene.2018.10.061
Martin L, Meier M, Lyons SM et al (2012) Systematic reconstruction of RNA functional motifs with high-throughput microfluidics. Nat Methods. https://doi.org/10.1038/nmeth.2225
Mathews CK (2017) Junk DNA: a journey through the dark matter of the genome. FASEB J. https://doi.org/10.1096/fj.201700951
Nakano M, Fukushima Y, Yokota SI et al (2015) CYP2A7 pseudogene transcript affects CYP2A6 expression in human liver by acting as a decoy for MIR-126. Drug Metab Dispos. https://doi.org/10.1124/dmd.115.063255
Paige JS, Wu KY, Jaffrey SR (2011) RNA mimics of green fluorescent protein. Science. https://doi.org/10.1126/science.1207339
Pink RC, Wicks K, Caley DP, Punch EK, Jacobs LCD (2011) Pseudogenes: pseudo-functional or key regulators in health and disease? RNA 17:792–798
Poliseno L (2012) Pseudogenes: Newly discovered players in human cancer. Sci Signal 5:1–14
Poliseno L, Marranci A, Pandolfi PP (2015) Pseudogenes in human cancer. Front Med 2:68. https://doi.org/10.3389/fmed.2015.00068
Poliseno L, Pandolfi PP (2015) PTEN ceRNA networks in human cancer. Methods 77–78:41–50
Poliseno L, Salmena L, Zhang J et al (2010) A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature. https://doi.org/10.1038/nature09144
Porter KA, Duffy EB, Nyland P et al (2014) The CLRX.1/NOD24 (NLRP2P) pseudogene codes a functional negative regulator of NFkappaB, pyrin-only protein 4. Genes Immun 15:392–403
Saleh M, Vaillancourt JP, Graham RK et al (2004) Differential modulation of endotoxin responsiveness by human caspase-12 polymorphisms. Nature. https://doi.org/10.1038/nature02451
Salmena L (2014) Pseudogene redux with new biological significance. Methods Mol Biol. https://doi.org/10.1007/978-1-4939-0835-6_1
Sasidharan R, Gerstein M (2008) Genomics: protein fossils live on as RNA. Nature 453:729–731
Scarola M, Comisso E, Pascolo R et al (2015) Epigenetic silencing of Oct4 by a complex containing SUV39H1 and Oct4 pseudogene lncRNA. Nat Commun. https://doi.org/10.1038/ncomms8631
Schmitz KM, Mayer C, Postepska A, Grummt I (2010) Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. Genes Dev. https://doi.org/10.1101/gad.590910
Seila AC, Calabrese JM, Levine SS et al (2008) Divergent transcription from active promoters. Science. https://doi.org/10.1126/science.1162253
Skaletsky H, Kuroda-Kawaguchl T, Minx PJ et al (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature. https://doi.org/10.1038/nature01722
Takahashi K, Tanabe K, Ohnuki M et al (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell. https://doi.org/10.1016/j.cell.2007.11.019
Tay Y, Rinn J, Pandolfi PP (2014) The multilayered complexity of ceRNA crosstalk and competition. Nature 505:344–352
Thomson DW, Dinger ME (2016) Endogenous microRNA sponges: evidence and controversy. Nat Rev Genet. https://doi.org/10.1038/nrg.2016.20
Tutar Y (2012) Pseudogenes. Comp Funct Genom 2012:424–526
Veerappa AM, Padakannaya P, Ramachandra NB (2013) Copy number variation-based polymorphism in a new pseudoautosomal region 3 (PAR3) of a human X-chromosome-transposed region (XTR) in the Y chromosome. Funct Integr Genom. https://doi.org/10.1007/s10142-013-0323-6
Wang X, Grus WE, Zhang J (2006) Gene losses during human origins. PLoS Biol. https://doi.org/10.1371/journal.pbio.0040052
Xiao-Jie L, Ai-Mei G, Li-Juan J, Jiang X (2015) Pseudogene in cancer: real functions and promising signature. J Med Genet. https://doi.org/10.1136/jmedgenet-2014-102785
Zhang J, Wang X, Li M et al (2006) NANOGP8 is a retrogene expressed in cancers. FEBS J. https://doi.org/10.1002/qua.20952
Zhang Z, Zheng D (2014) Pseudogene evolution in the human genome. https://doi.org/10.1002/9780470015902.a0020836.pub2

Zhang Z, Zhu Z, Watabe K et al (2013) Negative regulation of lncRNA GAS5 by miR-21. Cell Death Differ. https://doi.org/10.1038/cdd.2013.110

Zheng L, Li X, Gu Y et al (2015) The 3′UTR of the pseudogene CYP4Z2P promotes tumor angiogenesis in breast cancer by acting as a ceRNA for CYP4Z1. Breast Cancer Res Treat. https://doi.org/10.1007/s10549-015-3298-2