

Pseudogene in cancer: real functions and promising signature

Lu Xiao-Jie,¹ Gao Ai-Mei,² Ji Li-Juan,³ Xu Jiang³

¹Department of Gastroenterology, Shanghai East Hospital, Tongji University, School of Medicine, Shanghai, China

²Department of Pharmacy, The Fifth People's Hospital of Shanghai, Fudan University, Shanghai, China

³Department of Rehabilitation, The Second People's Hospital of Huai'an, Huai'an, China

Correspondence to

Ji Li-Juan, Department of Rehabilitation, The Second People's Hospital of Huai'an, Huai'an 223301, China; (188@whu.edu.cn), or Xu Jiang (luzg88@163.com)

LX and GAM contributed equally and are co-first authors.

Received 26 September 2014
Revised 13 October 2014
Accepted 18 October 2014

ABSTRACT

Pseudogenes were initially regarded as non-functional genomic fossils resulted from inactivating gene mutations during evolution. However, later studies revealed that they play a plethora of roles at multiple levels (DNA, RNA and/or protein) in diverse physiological and pathological processes, especially in cancer, both parental-gene-dependently and parental-gene-independently. Pseudogenes can interact with parental genes or other gene loci, leading to alteration in their sequences and/or transcriptional activities. Pseudogene-derived RNAs play multifaceted roles in post-transcriptional regulation as antisense RNAs, endogenous small-interference RNAs, competing endogenous RNAs and so on. Pseudogenic proteins can mirror, mimic or interfere with the functions of their parental counterparts. Herein, we discuss the general aspects (origination, classification, identification) of pseudogenes, focus on their multiple functions in cancer pathogenesis and prospect the potentials they hold as molecular signatures assisting in cancer reclassification and tailored therapy.

INTRODUCTION

The word 'pseudogene' was first introduced by Jacq *et al*¹ in 1977, when a copy of the 5S rRNA gene was discovered in *Xenopus laevis*, with 5'-end truncation and 14-bp mismatches that render it non-functional. Since then, numerous pseudogenes have been discovered in organisms from prokaryotes to eukaryotes. In human genome, there are about 11 000 pseudogenes,² exceeding half the number of protein-coding genes.

Traditionally, pseudogenes are considered as genomic loci that resemble real genes, yet are biologically inconsequential because they harbour mutations that abrogate their transcription or translation.³ Resultantly, they were once regarded as 'junk genes', 'relics of evolution' or 'genomic fossil'.⁴⁻⁵ Recently, however, with the aid of next-generation sequencing and research advance in non-coding RNAs, multilayered functions of pseudogenic DNA, RNA or protein have been discovered in multiple cancers. Pseudogenes play important roles in transcriptional and post-transcriptional regulations and also have the potential to evolve into novel genes, thus serving as a reservoir for gene renewal. Moreover, a small handful of pseudogenes have been reported to retain or regain protein-coding properties, and the resultant pseudogenic proteins/polypeptides mirror or interfere with the functions of their parental counterparts in tumorigenesis.⁶⁻¹¹ In this review, we discuss the identification, classification, functions and clinical relevance of pseudogenes in cancer, with recent advances and future perspectives.

ORIGINATION AND CLASSIFICATION OF PSEUDOGENES

The existence of more than one copy of a gene in human genome allows the production of gene variants which may generate novel genes in some contexts, whereas they give birth to pseudogenes in others. Pseudogenes can derive from gene mutations, or unfaithful gene duplications, or retrotransposition of processed mRNAs back into the genome. Accordingly, pseudogenes can be categorised into three types: (1) unitary pseudogenes (figure 1A), (2) duplicated or unprocessed pseudogenes (figure 1B) and (3) processed or retrotransposed pseudogenes (figure 1C).

Unitary pseudogenes are generated when spontaneous mutations in a coding gene abolish either transcription or translation of that gene. As a result, unitary pseudogenes lack the fully functional counterparts (termed 'ancestral genes', 'cognate genes' or 'parental genes') as the other two types of pseudogenes do. Duplicated pseudogenes derive from unfaithful gene duplication, resulting in the loss of promoters/enhancers or frameshift mutations or premature stop codons, thus rendering them non-functional, whereas their parental genes remain functional. Duplicated pseudogenes are often located within the vicinity of their parental genes. Both the unitary pseudogenes and the duplicated pseudogenes retain intron-exon structures. On the contrary, processed pseudogenes lack introns because they are originated from mRNAs that are reverse-transcribed into DNAs and then integrated back into the genome at a new location.

IDENTIFICATION AND CANCER-SPECIFIC EXPRESSIONS OF PSEUDOGENES

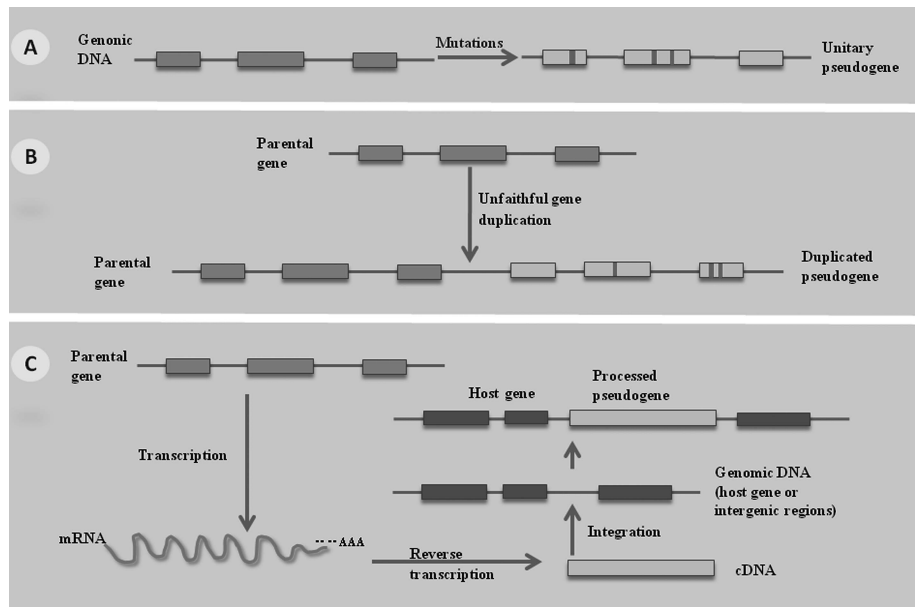
Due to their high homology to parental genes, a major challenge faced by pseudogenes studies is how to distinguish them from their parental genes, with individual genome differences and sequencing errors further complicating the matter. In recent years, multiple approaches have been developed for this purpose at DNA level¹²⁻¹⁴ or, for expressed pseudogenes, at RNA level.¹⁴⁻¹⁶

Pipelines established to identify pseudogene DNA include PseudoPipe,¹⁷ the Human and Vertebrate Analysis and Annotation (HAVANA) method,¹⁸ PseudoFinder and RetroFinder.¹⁹ These pipelines have now been integrated into a consensus platform called ENCYClopedia Of DNA Elements (ENCODE), the most comprehensive database for pseudogenes at present.^{20 21}

Previously, approaches to identify pseudogene RNA were quite limited, mainly relying upon incongruent gene expression platforms, such as public mRNA and Expressed Sequence Tag databases, Cap

To cite: Xiao-Jie L, Ai-Mei G, Li-Juan J, *et al*. *J Med Genet* Published Online First: [please include Day Month Year] doi:10.1136/jmedgenet-2014-102785

Figure 1 The origin and classification of pseudogene. (A) Gene mutations that abrogate transcription and/or translation turn that 'gene' into 'unitary pseudogene'. (B) Duplicated pseudogenes derive from unfaithful (uncompleted or mutated) gene duplications, with the original copy remaining fully functional (parental gene), whereas the duplicated copy becomes pseudogene. (C) Processed pseudogenes arise from mRNA that is reverse transcribed into cDNA and then integrated into another gene (host gene) or intergenic region. That is why they are often located distant from their parental gene and lack introns.



Analysis Gene Expression libraries or gene identification signature-paired end tags.²² In 2012, Shanker Kalyana-Sundaram and his colleagues¹⁵ developed a bioinformatics pipeline to detect pseudogene transcriptions based on next-generation sequencing data of 293 samples (13 cancer and normal tissue types) and identified 2082 pseudogene transcripts, among which 154 are highly tissue-specific and 218 expressed only in cancer samples (178 expressed in multiple cancers, while 40 were single cancer-specific). Of them, a breast cancer-specific pseudogene, ATP8A2- ψ , was selected to be validated by Taqman assays, and the result shows strong concordance to bioinformatics analysis, with ATP8A2- ψ expression found to be restricted to breast tumours with luminal histology. Moreover, subsequent overexpression and knockdown experiments *in vitro* indicated an oncogenic role of ATP8A2- ψ .¹⁵ Similarly, two recent studies by Shen *et al*²³ and Pan *et al*²⁴ reported that the polymorphism of pseudogenes POU5F1P1 rs10505477 and E2F3P1rs9909601, are correlated with patients' prognosis of gastric cancer and liver cancer, respectively. Another example of clinically significant pseudogene comes from POU5F1B, a processed pseudogene located adjacent to MYC on human chromosome 8q24, which is a reliable prognostic marker for patients with stage IV gastric cancer and shows oncogenic role both *in vitro* and *in vivo*.²⁵

Recently, Han *et al*¹⁶ developed a similar computational pipeline and detected 9925 pseudogene transcriptions in 2808 samples across seven cancer types from The Cancer Genome Atlas (TCGA) RNA-seq data. Of the detected pseudogene transcripts, many are tissue and/or cancer-specific. Moreover, this study for the first time systematically revealed the potential of pseudogenes as prognostic and subtype biomarkers in cancers. Tumour subtypes based on pseudogene expression profiles showed high concordance with molecular subtypes based on other omic data such as mRNA expression, miRNA expression, DNA methylation and somatic copy number variation (SNV). Moreover, in kidney cancer, subtypes based on pseudogenes showed stronger prognostic power than those based upon mRNA, miRNA or other molecular data.¹⁶

The detection efficacy of these RNA-based bioinformatics pipelines are determined by three factors: (1) pseudogene expression level (highly expressed pseudogenes are more likely to be detected), (2) coverage depth of RNA sequencing (the deeper the coverage, the more sensitive the detection) and (3)

mismatch distribution patterns of between pseudogene and parental gene (eg, mismatches accumulated in a small stretch of sequence are more likely to be detected than mismatches scattered over long stretches). Future progression in sequencing accuracy and coverage depth will facilitate the researches in this regard.

FUNCTIONS OF PSEUDOGENES IN CANCER

Pseudogenes were once regarded as functionally inert and subject to random genetic drift. However, studies in recent decades have piled up evidences for pseudogene evolutionary conservation across different mammals.^{26–28} The non-synonymous to synonymous substitution rate (K_a/K_s) is usually applied to determine whether a sequence is under evolutionary constraint. Generally speaking, K_a/K_s is less than one if the sequence is under purifying selection, equal to one if it is evolving neutrally and greater than one if under positive selection. Theoretically, non-functional sequences should be under neutral selection, and their K_a/K_s ratios are expected to be equal to one. However, it has been reported that K_a/K_s values between genes and pseudogenes overlap greatly, suggesting that some pseudogenes are under evolutionary constraint rather than evolving neutrally, lending support to their roles as functional units.^{29–31}

In recent years, multilayered functions of pseudogene DNAs, RNAs or proteins have been reported in diverse cancer types.

Functions of pseudogene DNAs

Pseudogene DNAs can function via gene conversion, homologous recombination, exonisation or insertional mutations, of which the former two events often occur between pseudogene and parental gene, whereas, the latter two are usually between pseudogene and host gene (figure 2A–E).

Gene conversion

Gene conversion is a process in which one DNA sequence replaces a homologous sequence such that the sequences become identical after the conversion (figure 2A). Theoretically, gene conversion from pseudogene to parental gene provides an

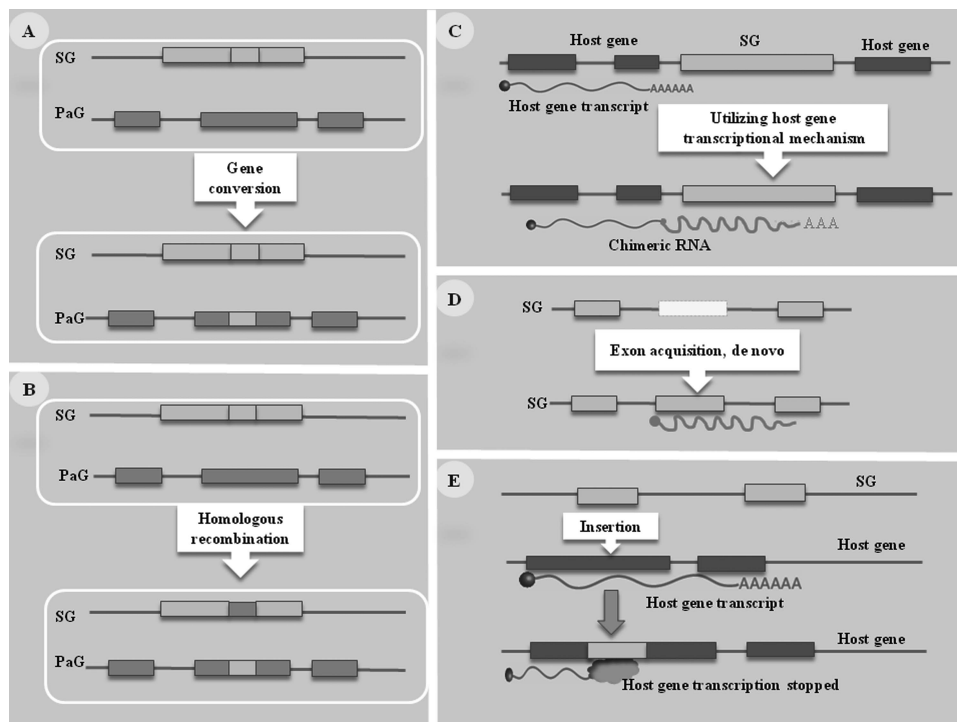


Figure 2 Functions of pseudogene DNA. (A) Gene conversion, a process that part of parental gene sequence is replaced by that of its pseudogene. (B) Homologous recombination, a process in which pseudogene and parental gene exchange DNA sequences. (C, D) Exonisation: pseudogene uses transcriptional mechanism of host gene for transcription, giving birth to a chimeric RNA (C), or, in other cases, obtains exons de novo (D). (E) Insertional mutation, pseudogene inserts into the promoter/exons of its host gene and abolishes the latter from expression. SG, pseudogene; PaG, parental gene.

ideal chance for oncogene activation and/or tumour suppressor gene inactivation.

Conversions from pseudogene to parental gene have been reported in several human diseases.^{32 33} Cytochrome P450 2A6 (CYP2A6), for example, enzyme metabolising precarcinogens including nicotine, has a pseudogene called CYP2A7. Gene conversion from CYP2A7 to CYP2A6 results in the generation of CYP2A6*1B, which has higher nicotine metabolism activity in vivo and thus influences cigarette consumption as well as smoking-induced lung cancer risk.³⁴

Homologous recombination

Pseudogene and parental gene can exchange DNA sequence with each other, a process called homologous recombination (figure 2B). The breast and ovarian cancer-susceptibility gene BRCA1, for example, has a pseudogene called PsiBRCA1 lying upstream of BRCA1. In two families with breast and ovarian cancer, homologous recombination took place between BRCA1 intron 2 and PsiBRCA1 intron 2, which results in a 37-kb deletion, deprives BRCA1 of its promoter and initiation codon, rendering it non-functional. The DNA recombination between BRCA1 and its pseudogene represents a new mechanism for oncosuppressor gene inactivation in cancer.³⁵

Exonisation

Pseudogenes can use the transcriptional mechanism of host genes for transcription (figure 2C), or, in some cases, obtain exons de novo (figure 2D). These processes are called exonisation.

Recently, a pioneering study conducted by Susanna L Cooke and her team¹⁴ identified 42 somatically acquired pseudogenes in 14 out of 629 primary cancer samples and 3 out of 31 cancer cell lines via bioinformatic analysis followed by PCR validation.

Among these pseudogenes, 16 were subsequently analysed to investigate the effect of their insertion sites on their expressions. Of these, none of the 10 pseudogenes inserted into intergenic regions are expressed, one of the three pseudogenes in introns is expressed, while two pseudogenes in 3' UTRs are both expressed. This result indicates that pseudogenes inserted in introns or 3' UTRs are capable of harnessing the transcriptional mechanism of host gene to be expressed, while pseudogenes in intergenic regions are far less likely to be expressed because they lack host genes and, thus, usable transcriptional mechanisms.

KLK3 1P, an unprocessed pseudogene of KLK3, has five exons. Exons 3 and 4 are duplicate copies of KLK1 exon 2, while the other three are 'exonised' de novo from interspersed repeats. KLK3 and KLK3 1P are both regulated by androgen. Interestingly, unlike KLK3, whose protein level is increased in the serum of patients with prostate cancer, the expression level of KLK3 1P step down from normal prostate epithelium cells to localised primary cancer cells to metastatic cells,³⁶ indicating that KLK3 1P may function independently of its parent gene in prostate cancer.

Insertional mutation

Apart from aforementioned functions, a newly discovered function of pseudogene DNA in cancer is that it inserts into the promoter/exons of the host gene and abolishes the latter from expression (figure 2E). Here, we call this process as 'Insertional Mutation'. In lung adenocarcinoma cell line NCI-H2009, a pseudogene called PTPN12 was reported to insert into Exon 1 of MGA, a possible oncosuppressor gene encoding a MAX-interacting protein. This insertion deletes the promoter and exon1 regions of MGA and renders it unexpressed.¹⁵ Given that tumour evolves with accumulation of mutations, insertional inactivation of oncosuppressor gene

by pseudogene may represent a new layer of genetic mutations during tumorigenesis.

Functions of pseudogene RNA

Though only a minor fraction of pseudogenes are transcribed, pseudogene transcripts play diverse roles in post-transcriptional regulation (figure 3A–F). Apart from the traditional roles as antisense RNA or endogenous small-interference RNA (endo-siRNA or esiRNA), they can also function as endogenous competitors for miRNA, for RNA-binding protein (RBP) or for translational machinery. And, in some cases, chimeric RNAs can form between pseudogenes and genes, with their functions still to be clarified.

As antisense RNA

Pseudogene RNA that is transcribed in antisense can combine directly with the parental sense mRNA to inhibit its translation (figure 3A). For example, neural nitric oxide synthase (nNOS) mRNA hybridises with an antisense nNOS pseudogene transcript, forming double-strand RNA–RNA duplex and resulting in nNOS translation suppression.³⁷ However, to date, there is still lack of a validated example that pseudogene RNAs function in this manner in cancer.

Antisense pseudogene RNAs can also function as endo-siRNA, which is shown below.

As endo-siRNA

Some pseudogene transcripts can be processed into endo-siRNAs (figure 3B–D). There are two major sources of pseudogene-derived endo-siRNAs. One is from hybrid double-

stranded RNAs (dsRNAs) composed of sense and antisense RNAs involving pseudogene (figure 3B, C), and the other from the inverted repeat region of pseudogene that is transcribed into hairpin-shaped RNA (figure 3D). These hairpin-shaped RNAs or hybrid dsRNAs can be sliced by Dicer (a ribonuclease protein) into smaller fragments known as endo-siRNAs that are subsequently separated into single strands and incorporated into the RNA-induced silencing complex, degrading target mRNAs, a process called RNA interference.³⁸

For example, in mice oocyte, a pseudogene called Au76 is transcribed into long-hairpin RNA and diced into siRNA, regulating expression of its parental gene (Rangap1).³⁹ Again in mice, pseudogene of Hda1 can be transcribed both in sense and in antisense, which then anneal to each other to form dsRNA that is sliced into siRNAs, regulating the expression of its parental gene.⁴⁰

In hepatocellular carcinoma, pseudogene-derived endo-siRNAs have been reported.⁴¹ ψ PPM1 K is a partial retrotransposed pseudogene with inverted repeats transcribed into long-hairpin RNA that is processed into two endo-siRNAs, which target and inhibit the expression of the parental gene (PPM1 K) and another gene (NEK8), leading to altered mitochondrial activation and decreased cancer cell proliferation, respectively, suggesting an oncosuppressive role both parental-gene-dependently and parental-gene-independently.

As competing endogenous RNA

In recent years, a newly discovered RNA regulatory mechanism called competing endogenous RNAs (ceRNA) has been the hotspot in cancer research.^{3 42–46} ceRNAs refer to RNAs that

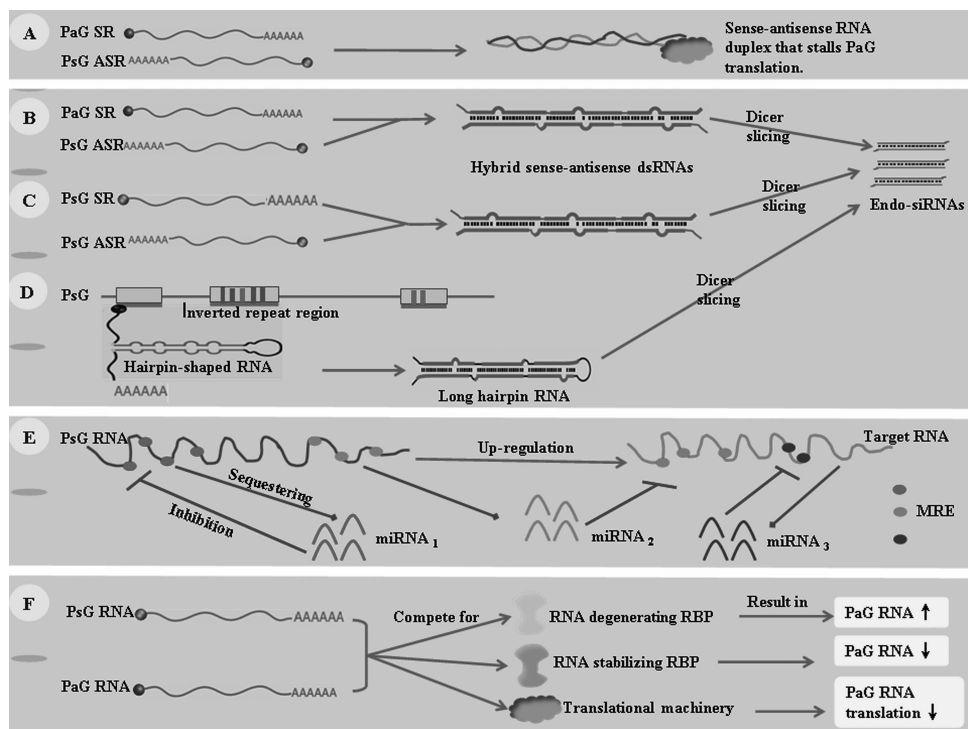


Figure 3 Functions of pseudogenic RNA. (A) Antisense pseudogenic RNA combines with the parental sense mRNA to stall its translation. (B, C, D) Pseudogene-derived endo-siRNAs come from dsRNA hybrid between pseudogenic antisense RNA and parental (B) or pseudogenic (C) sense RNA, or from hairpin-shaped pseudogenic RNA (D), all of which are then sliced by Dicer into endo-siRNAs, triggering RNA interference mechanism. (E) Competing endogenous RNA: when pseudogene RNA and target RNA both harbour MRE that binds to miRNA₂, the former can bind and sequester miRNA₂, freeing the latter from miRNA₂ mediated inhibition, upregulating its level. It is a reciprocal process. (F) Competition between pseudogenic and parental RNA for RBPs or translational machinery alters level or translational activity of the parental RNA. PsG, pseudogene; PaG, parental gene; ASR, antisense RNA; SR, sense RNA; MRE, miRNA response elements; RBP, RNA binding protein; dsRNA, double-stranded RNA; endo-siRNAs, endogenous small-interference RNAs.

share miRNA response elements (MRE) and, therefore, regulate each other's expression by competing for the same pool of miRNAs (figure 3E). Theoretically, any RNA that contains MRE can serve as ceRNAs, or RNA sponges, including pseudogene RNAs. Due to the fact that pseudogene RNAs harbour many of the same MREs as their parental RNAs, they are perfect candidates to form ceRNA pairs with their parental RNAs. Additionally, pseudogene RNAs as ceRNAs regulating expressions of genes other than parental genes have also been reported.

For example, pseudogene OCT4-pg4 transcript was reported to function as a ceRNA to regulate the expression of its parental gene OCT4 by competing for miR-145 in liver cancer. Moreover, the expression level of OCT4-pg4 is significantly relevant with patients' prognosis. Subsequent bimolecular experiments suggested oncogenic role of OCT4-pg4 in HepG2 cell line.⁴⁶

PTENP1, a pseudogene of the famous tumour suppressor gene phosphatase and tensin homolog (PTEN), was found to act as ceRNA both parental-gene-dependently and parental-gene-independently. PTENP1 was found to increase cellular levels of PTEN mRNA in prostate cancer through competitively binding to miR-17, miR-19, miR-21, miR-26 and miR-214 families, freeing PTEN mRNA from miRNA-induced suppression.³ Intriguingly, however, in PTEN knockout cancer cells, PTENP1 showed oncosuppressive role as well, suggesting its oncosuppressive role is at least partially parental-gene-independent. Subsequent study revealed that PTENP1 knockdown leads to reduced levels of p21 in cancer cells.³ Considering that p21 is a target of the miR-17 family and that PTENP1 sequesters miR-17 families, it is reasonable to infer that PTENP1 sequesters miR-17 and reverses miR-17-mediated p21 suppression.

Given the ubiquitous ceRNA network in post-transcriptional regulation^{3 42-46} and the prevalent existence of pseudogene in human genome,² it is sensible to expect an increasing number of pseudogene-involved ceRNA networks identified in cancers.

As competitors for RBP or translational machinery

Due to the high similarity in sequence, pseudogene RNAs can also compete with their parental counterparts for RBPs or translational machinery, and thus exert a regulatory role on the latter (figure 3F).

The effects of competition between pseudogenic and parental RNA for RBPs depend on functions of the RBP. For an

RNA-stabilising RBP, it would lead to reduced parental RNAs. Conversely, for a RNA-degenerating RBP, it would lead to parental RNAs upregulation (figure 3F). For example, MYLKP1, a pseudogene of myosin light chain kinase (MYLK) gene that encodes non-muscle and smooth muscle myosin light chain kinase (smMLCK) isoforms, inhibits parental RNA expression and thus promotes cancer cell proliferation. Subsequent mechanism research revealed that coexpression of MYLKP1 with smMLCK leads to decreased mRNA stability of smMLCK, suggesting competition may exist between this pair of pseudogenic and parental RNA for RNA-stabilising RBPs.⁴⁷

Upon competition for translational machinery, it will result in decreased translation of parental RNAs (figure 3F). ψ Cx43, for example, is a pseudogene of connexin43 (Cx43), which encodes a protein involved in intercellular communication and tumour pathogenesis. In breast cancer, ψ Cx43 inhibits Cx43 translation since the former binds to the translation machinery more efficiently than the latter. Knockdown of ψ Cx43 leads to increased levels of Cx43 mRNA and protein and thus increased cellular sensitivity to chemotherapeutics.⁴⁸

As chimeric RNAs

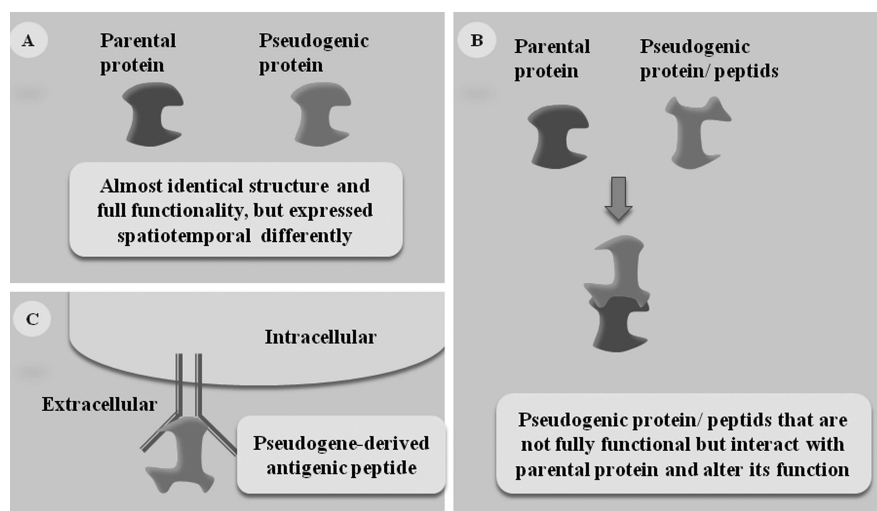
Various chimeric RNAs have been identified in multiple cancers recently, with some of them expressed cancer-specifically.⁴⁹⁻⁵¹ Herein, chimeric RNA refers to an RNA sequence that is transcribed partially from pseudogenes and partially from other genes, but is fused together as a whole (figure 2C).

For example, a chimeric RNA transcript composed of the first two exons of KLK4 and the last two exons of pseudogene KLKP1 has been identified in prostate cancer.^{15 52} This chimeric RNA was highly expressed in 30%–50% of prostate cancer tissues, with barely any expression in benign prostate or other tissues, suggesting a cancer type-specific and tissue-specific expression pattern. However, whether this chimeric RNA can be translated into protein, or how it functions in prostate cancer, remain unclear.

Functions of pseudogenic protein

By definition, pseudogenes are gene loci harbouring premature stop codons, indels or frameshift mutations that abrogate their translation.³ In reality, however, though the majority of pseudogenes have lost protein-coding ability, a small handful of processed pseudogenes retain or regain this ability. The first pseudogenic protein was discovered in 2002, namely PGAM3, a

Figure 4 Functions of pseudogenic protein. Pseudogenic protein can be fully functional (A), partially functional through interaction with parental protein (B) or as antigenic peptide (C).



protein coded by a processed pseudogene in primate white blood cells.⁵³ Later on in 2004, pseudogenic protein in breast cancer cell lines was identified, and the protagonist is the aforementioned pseudogene ψ Cx43.⁴⁸ ψ Cx43 is translated into protein that is highly homologous to Cx43 protein and exhibits growth-suppressive behaviour similar to Cx43 protein.

NANOGP8, one of the 11 pseudogenes of NANOG gene that plays key roles in embryonic stem cell self-renewal, encode a protein detected by anti-Nanog antibody in OS732 cell (human osteosarcoma cell line) and HepG2 (human liver cancer cell line).⁵⁴ And later on in prostate cancer,⁵⁵ the amino acid sequence of NANOGP8-encoded protein was identified and proven to be highly identical to that of NANOG protein. In this study, NANOGP8 was found to be the major source of NANOG RNAs, the abundance of which was correlated with the number of CD44-positive cancer stem cells. And accordingly, RNA interference-mediated NANOG knockdown inhibited tumour development, both in vitro and in vivo.⁵⁵ Though mRNA and protein derived from NANOGP8 and NANOG are almost identical, their expression patterns are somewhat different. For example, both NANOGP8 and Nanog were transcribed in HepG2, whereas only NANOGP8 was transcribed in OS732.⁵⁴ In another study,⁵⁶ NANOG was found to be dominantly expressed in SW620 colon cancer cell line, while NANOGP8 was the major form in two other colon cancer cell lines: HT29 and HCT116. These studies suggest that certain pseudogenes can encode protein with almost identical functions to those of their counterpart proteins, but are expressed in different patterns (figure 4A).

However, not all pseudogenic proteins are fully functional. BRAF is a serine/threonine kinase that is involved in mitogen-activated protein kinase (MAPK)-signalling cascade and mutated in multiple human cancers. Its pseudogene, BRAFP1, located on chromosome Xq13, has many stop codons that abrogate it from translation into a fully functional protein. However, the longest open reading frame of BRAFP1 can encode a 244 amino acid polypeptide, which has high-sequence homology with the CR1 domain of wild-type BRAF protein, and interacts with the latter, thus activating MAPK pathway, exerting an oncogenic role in thyroid tumours. Intriguingly, BRAFP1 RNAs were more frequently detected in samples without BRAF mutation, indicating that either of these two events is sufficient to drive tumorigenesis.⁵⁷ In this case, though pseudogenic protein is not fully functional, it can influence the activity of parental protein and thus play a role in tumorigenesis (figure 4B).

Certain pseudogenic proteins or short peptides derived from open-reading frames, on the other hand, are recognised by the human immune system as 'antigens' (figure 4C). Examples of this kind have been reported in melanoma⁵⁸ and in sarcoma.⁵⁹ Though theoretically 'self', cancer cells can produce proteins that are spatiotemporally inappropriate, thus being recognised by the immune system as 'non-self'. Research on pseudogenic antigens is still in its infancy, but holds promise to give rise to new tumour markers or therapeutic targets.

CONCLUSION AND PERSPECTIVE

It is reported that pseudogenes outnumber half the protein-coding genes in the human genome.² The prevalent existence of pseudogenes indicates that they may play a vital role in basic physiology and disease progression. Traditionally, pseudogenes were viewed as 'junk DNA' or 'genomic fossils', due to the fact that they are either not transcribed or not translated into functional proteins. However, studies in recent decades indicate that they are far more than merely 'junk' or 'non-functional'. In fact,

they play a plethora of roles at multiple levels (DNA, RNA and protein) both in health and in disease, and especially, in cancer. As stated above, pseudogenes represent a reservoir for gene evolution and/or protein diversity. Pseudogenes can interact with parental genes or other gene loci, altering their sequences and/or transcriptional activities (figure 2A–E). Upon being transcribed into RNA, pseudogenes take on a diversity of post-transcriptional regulatory roles in cancer, such as antisense RNA, endo-siRNA, ceRNA, chimeric RNA and RBP and/or translational machinery competitors (figure 3A–F). Moreover, the discoveries of a small handful of pseudogenes capable of encoding proteins^{54–56} that mirror, mimic or interfere with the functions of their parental counterparts (figure 4A–C) have blurred the distinction between genes and pseudogenes. Therefore, in the term, pseudogene, 'pseudo' implies sequence variances compared with parental gene, not indicating pseudo-function. Though sequence mutations render them 'pseudo'-relative to parental genes, many of them perform real and indisputable functions in physical and pathological processes.

A main obstacle in pseudogene research comes from the close homology of DNA, RNA and/or protein sequences between pseudogene and parental gene that render it hard to distinguish the former from the latter. In the past, pseudogenes were considered as nasty 'noises' that would interfere with the detection of their parental genes. As a result, various endeavours were made to rule out rather than identify pseudogenes. With the advent of the next-generation sequencing era, massive multi-omic data are available at online databases, such as TCGA, Gene Expression Omnibus, International Cancer Genome Consortium, ENCODE and so on, which greatly facilitate the research on cancer genome, epigenome and proteome. In this background, many pseudogenes are newly identified with the aid of bioinformatics analysis, especially at RNA level.^{14–16}

Reclassification of cancers within and across tissue-of-origin based on multi-omic data is a hotspot in cancer research nowadays. Pioneering studies have already reclassified several common cancers, providing independent prognostic power and potential therapy-guiding values.^{60–61} Recently, pseudogenes are also applied to act as molecular signatures for cancer subtyping. In a study¹⁶ analysing pseudogene expression patterns in six cancer types, many pseudogenes were found to be expressed differentially among tumour subtypes. Pseudogene-based tumour subtypes show strong concordance with subtypes based on other omic data, including mRNA, miRNA, DNA methylation and SNV. In kidney cancer, pseudogenes even show stronger prognostic power than other molecular signatures. Therefore, it is reasonable for us to predict that, in the future, pseudogene signature, along with other molecular signatures, will pave the way for cancer reclassification and tailored therapy.

Contributors JL-J and XJ conceived the idea and collected literature; LX-J and GA-M read through and analysed relevant literature; LX-J and GA-M wrote the manuscript while JL-J and XJ revised it.

Funding Shanghai Science and Technology Committee Yangfan Program (14YF1411900).

Competing interests None.

Provenance and peer review Commissioned; externally peer reviewed.

REFERENCES

- 1 Jacq C, Miller JR, Brownlee GG. A pseudogene structure in 5S DNA of *Xenopus laevis*. *Cell* 1977;12:109–20.
- 2 ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012;489:57–74.

- 3 Polisenio L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* 2010;465:1033–8.
- 4 Proudfoot N. Pseudogenes. *Nature* 1980;286:840–1.
- 5 Zhang Z, Harrison PM, Liu Y, Gerstein M. Millions of years of evolution preserved: A comprehensive catalog of the processed pseudogenes in the human genome. *Genome Res* 2003;13:2541–58.
- 6 Zhang J, Wang X, Li M, Han J, Chen B, Wang B, Dai J. NANOGP8 is a retrogene expressed in cancers. *FEBS J* 2006;273:1723–30.
- 7 Ishiguro T, Sato A, Ohata H, Sakai H, Nakagama H, Okamoto K. Differential expression of nanog1 and nanogp8 in colon cancer cells. *Biochem Biophys Res Commun* 2012;418:199–204.
- 8 Jeter CR, Badeaux M, Choy G, Chandra D, Patrawala L, Liu C, Calhoun-Davis T, Zaehres H, Daley GQ, Tang DG. Functional evidence that the self-renewal gene NANOG regulates human tumor development. *Stem Cells* 2009;27:993–1005.
- 9 Sun C, Orozco O, Olson DL, Choi E, Garber E, Tizard R, Szak S, Sanicola M, Carulli JP. CRIPTO3. A presumed pseudogene, is expressed in cancer. *Biochem Biophys Res Commun* 2008;377:215–20.
- 10 Takahashi K, Mitsui K, Yamanaka S. Role of ERas in promoting tumour-like properties in mouse embryonic stem cells. *Nature* 2003;423:541–5.
- 11 Zou M, Baitei EY, Alzahrani AS, Al-Mohanna F, Farid NR, Meyer B, Shi Y. Oncogenic activation of MAP kinase by BRAF pseudogene in thyroid tumors. *Neoplasia* 2009;11:57–65.
- 12 Karro JE, Yan Y, Zheng D, Zhang Z, Carriero N, Cayting P, Harrison P, Gerstein M. Pseudogene.org: a comprehensive database and comparison platform for pseudogene annotation. *Nucleic Acids Res* 2007;35:D55–60.
- 13 Zhang Z, Gerstein M. Large-scale analysis of pseudogenes in the human genome. *Curr Opin Genet Dev* 2004;14:328–35.
- 14 Cooke SL, Shlien A, Marshall J, Pipinikas CP, Martincorena I, Tubio JM, Li Y, Menzies A, Mudie L, Ramakrishna M, Yates L, Davies H, Bolli N, Bignell GR, Tarpey PS, Behjati S, Nik-Zainal S, Papaemmanuil E, Teixeira VH, Raine K, O'Meara S, Dodoran MS, Teague JW, Butler AP, Iacobuzio-Donahue C, Santarius T, Grundy RG, Malkin D, Greaves M, Munshi N, Flanagan AM, Bowtell D, Martin S, Larsimont D, Reis-Filho JS, Boussioutas A, Taylor JA, Hayes ND, Janes SM, Futreal PA, Stratton MR, McDermott U, Campbell PJ; ICGC Breast Cancer Group. Processed pseudogenes acquired somatically during cancer development. *Nat Commun* 2014;5:3644.
- 15 Kalyana-Sundaram S, Kumar-Sinha C, Shankar S, Robinson DR, Wu YM, Cao X, Asangani IA, Kothari V, Prensner JR, Lonigro RJ, Iyer MK, Barrette T, Shanmugam A, Dhanasekaran SM, Palanisamy N, Chinnaiyan AM. Expressed Pseudogenes in the Transcriptional Landscape of Human Cancers. *Cell* 2012;149:1622–34.
- 16 Han L, Yuan Y, Zheng S, Yang Y, Li J, Edgerton ME, Diao L, Xu Y, Verhaak RG, Liang H. The Pan-Cancer analysis of pseudogene expression reveals biologically and clinically relevant tumour subtypes. *Nat Commun* 2014;5:3963.
- 17 Zhang Z, Carriero N, Zheng D, Karro J, Harrison PM, Gerstein M. PseudoPipe: an automated pseudogene identification pipeline. *Bioinformatics* 2006;22:1437–9.
- 18 Solovyev V, Kosarev P, Seledsov I, Vorobyev D. Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biol* 2006;7(Suppl 1):S10.1–12.
- 19 Zheng D, Gerstein MB. A computational approach for identifying pseudogenes in the ENCODE regions. *Genome Biol* 2006;7(Suppl 1):S13.1–10.
- 20 Zheng D, Frankish A, Baertsch R, Kapranov P, Reymond A, Choo SW, Lu Y, Denoeud F, Antonarakis SE, Snyder M, Ruan Y, Wei CL, Gingeras TR, Guigó R, Harrow J, Gerstein MB. Pseudogenes in the ENCODE regions: consensus annotation, analysis of transcription, and evolution. *Genome Res* 2007;17:839–51.
- 21 Pei B, Sisic C, Frankish A, Howald C, Habegger L, Mu XJ, Harte R, Balasubramanian S, Tanzer A, Diekhans M, Reymond A, Hubbard TJ, Harrow J, Gerstein MB. The GENCODE pseudogene resource. *Genome Biol* 2012;13:R51.
- 22 Ruan Y, Ooi HS, Choo SW, Chiu KP, Zhao XD, Srinivasan KG, Yao F, Choo CY, Liu J, Ariyaratne P, Bin WG, Kuznetsov VA, Shahab A, Sung WK, Bourque G, Palanisamy N, Wei CL. Fusion transcripts and transcribed retrotransposed loci discovered through comprehensive transcriptome analysis using Paired-End diTags (PETs). *Genome Res* 2007;17:828–38.
- 23 Shen L, Du M, Wang C, Gu D, Wang M, Zhang Q, Zhao T, Zhang X, Tan Y, Huo X, Gong W, Xu Z, Chen J, Zhang Z. Clinical Significance of POU5F1P1 rs10505477 Polymorphism in Chinese Gastric Cancer Patients Receiving Cisplatin-Based Chemotherapy after Surgical Resection. *Int J Mol Sci* 2014;15:12764–77.
- 24 Pan Y, Sun C, Huang M, Liu Y, Qi F, Liu L, Wen J, Liu J, Xie K, Ma H, Hu Z, Shen H. A genetic variant in pseudogene EZFP1 contributes to prognosis of hepatocellular carcinoma. *J Biomed Res*. 2014;3:194–200.
- 25 Hayashi H, Arao T, Togashi Y, Kato H, Fujita Y, De Velasco MA, Kimura H, Matsumoto K, Tanaka K, Okamoto I, Ito A, Yamada Y, Nakagawa K, Nishio K. The OCT4 pseudogene POU5F1B is amplified and promotes an aggressive phenotype in gastric cancer. *Oncogene* 2013. doi: 10.1038/onc.2013.547.
- 26 Khachane AN, Harrison PM. Assessing the genomic evidence for conserved transcribed pseudogenes under selection. *BMC Genomics* 2009;10:435.
- 27 Podlaha O, Zhang J. Nonneutral evolution of the transcribed pseudogene Makorin1-p1 in mice. *Mol Biol Evol* 2004;12:2202–9.
- 28 Svensson O, Arvestad L, Lagergren J. Genome-wide survey for biologically functional pseudogenes. *PLoS Comput Biol* 2006;2:e46.
- 29 Hirotsune S, Yoshida N, Chen A, Garrett L, Sugiyama F, Takahashi S, Yagami K, Wynshaw-Boris A, Yoshiki A. An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene. *Nature* 2003;423:91–6.
- 30 Kandouz M, Bier A, Carystinos GD, Alaoui-Jamali MA, Batist G. Connexin43 pseudogene is expressed in tumor cells and inhibits growth. *Oncogene* 2004;23:4763–70.
- 31 Bischof JM, Chiang AP, Scheetz TE, Stone EM, Casavant TL, Sheffield VC, Braun TA. Genome-wide identification of pseudogenes capable of disease-causing gene conversion. *Hum Mutat* 2006;27:545–52.
- 32 Kehrer-Sawatzki H, Kluwe L, Sandig C, Kohn M, Wimmer K, Krammer U, Peyrl A, Jenne DE, Hansmann I, Mautner VF. High frequency of mosaicism among patients with neurofibromatosis type 1 (NF1) with microdeletions caused by somatic recombination of the JAZ1 gene. *Am J Hum Genet* 2004;3:410–23.
- 33 Ganster C, Wernstedt A, Kehrer-Sawatzki H, Messiaen L, Schmidt K, Rahner N, Heinimann K, Fonatsch C, Zschocke J, Wimmer K. Functional PMS2 hybrid alleles containing a pseudogene specific missense variant trace back to a single ancient intrachromosomal recombination event. *Hum Mutat* 2010;5:552–60.
- 34 Wang J, Pitarque M, Ingelman-Sundberg M. 3'-UTR polymorphism in the human CYP2A6 gene affects mRNA stability and enzyme expression. *Biochem Biophys Res Commun* 2006;340:491–7.
- 35 Puget N, Gad S, Perrin-Vidoz L, Sinilnikova OM, Stoppa-Lyonnet D, Lenoir GM, Mazoyer S. Distinct BRCA1 rearrangements involving the BRCA1 pseudogene suggest the existence of a recombination hot spot. *Am J Hum Genet* 2002;70:858–65.
- 36 Lu W, Zhou D, Glusman G, Utleg AG, White JT, Nelson PS, Vasicek TJ, Hood L, Lin B. KLK31P is a novel androgen regulated and transcribed pseudogene of kallikreins that is expressed at lower levels in prostate cancer cells than in normal prostate cells. *Prostate* 2006;66:936–44.
- 37 Korneev SA, Park JH, O'Shea M. Neuronal expression of neural nitric oxide synthase (nNOS) protein is suppressed by an antisense RNA transcribed from an NOS pseudogene. *J Neurosci* 1999;19:7711–20.
- 38 Okamura K, Lai EC. Endogenous small interfering RNAs in animals. *Nat Rev Mol Cell Biol* 2008;9:673–8.
- 39 Watanabe T, Totoki Y, Toyoda A, Kaneda M, Kuramochi-Miyagawa S, Obata Y, Chiba H, Kohara Y, Kono T, Nakano T, Surani MA, Sakaki Y, Sasaki H. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* 2008;453:539–43.
- 40 Tam OH, Aravin AA, Stein P, Girard A, Murchison EP, Cheloufi S, Hodges E, Anger M, Sachidanandam R, Schultz RM, Hannon GJ. Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* 2008;453:534–8.
- 41 Chan WL, Yuo CY, Yang WK, Hung SY, Chang YS, Chiu CC, Yeh KT, Huang HD, Chang JG. Transcribed pseudogene yPPM1K generates endogenous siRNA to suppress oncogenic cell growth in hepatocellular carcinoma. *Nucleic Acids Res* 2013;41:3734–47.
- 42 Tay Y, Kats L, Salmena L, Weiss D, Tan SM, Ala U, Karreth F, Polisenio L, Provero P, Di Cunto F, Lieberman J, Rigoutsos I, Pandolfi PP. Coding-independent regulation of the tumor suppressor PTEN by competing endogenous mRNAs. *Cell* 2011;147:344–57.
- 43 Kumar MS, Armenteros-Monterroso E, East P, Chakravorty P, Matthews N, Winslow MM, Downward J. HMG2A functions as a competing endogenous RNA to promote lung cancer progression. *Nature* 2014;505:212–17.
- 44 Karreth FA, Tay Y, Perna D, Ala U, Tan SM, Rust AG, DeNicola G, Webster KA, Weiss D, Perez-Mancera PA, Krauthammer M, Halaban R, Provero P, Adams DJ, Tuveson DA, Pandolfi PP. In vivo identification of tumor-suppressive PTEN ceRNAs in an oncogenic BRAF-induced mouse model of melanoma. *Cell* 2011;147:382–95.
- 45 Sumazin P, Yang X, Chiu HS, Chung WJ, Iyer A, Llobet-Navas D, Rajbhandari P, Bansal M, Guarnieri P, Silva J, Califano A. An extensive microRNA-mediated network of RNA-RNA interactions regulates established oncogenic pathways in glioblastoma. *Cell* 2011;147:370–81.
- 46 Wang L, Guo ZY, Zhang R, Xin B, Chen R, Zhao J, Wang T, Wen WH, Jia LT, Yao LB, Yang AG. Pseudogene OCT4-pg4 functions as a natural micro RNA sponge to regulate OCT4 expression by competing for miR-145 in hepatocellular carcinoma. *Carcinogenesis* 2013;34:1773–81.
- 47 Han YJ, Ma SF, Yourek G, Park YD, Garcia JG. A transcribed pseudogene of MYLK promotes cell proliferation. *FASEB J* 2011;25:2305–12.
- 48 Bier A, Oviedo-Landaverde I, Zhao J, Mamane Y, Kandouz M, Batist G. Connexin43 pseudogene in breast cancer cells offers a novel therapeutic target. *Mol Cancer Ther* 2009;8:786–93.
- 49 Lau CC, Sun T, Ching AK, He M, Li JW, Wong AM, Co NN, Chan AW, Li PS, Lung RW, Tong JH, Lai PB, Chan HL, To KF, Chan TF, Wong N. Viral-human chimeric transcript predisposes risk to liver cancer development and progression. *Cancer Cell* 2014;25:335–49.

Review

- 50 Ren G, Zhang Y, Mao X, Liu X, Mercer E, Marzec J, Ding D, Jiao Y, Qiu Q, Sun Y, Zhang B, Yeste-Velasco M, Chelala C, Berney D, Lu YJ. Transcription-Mediated Chimeric RNAs in Prostate Cancer: Time to Revisit Old Hypothesis? *OMICS* 2014;18:615–24.
- 51 Varley KE, Gertz J, Roberts BS, Davis NS, Bowling KM, Kirby MK, Nesmith AS, Oliver PG, Grizzle WE, Forero A, Buchsbaum DJ, LoBuglio AF, Myers RM. Recurrent read-through fusion transcripts in breast cancer. *Breast Cancer Res Treat* 2014;146:287–97.
- 52 Lai J, Lehman ML, Dinger ME, Hendy SC, Mercer TR, Seim I, Lawrence MG, Mattick JS, Clements JA, Nelson CC. A variant of the KLK4 gene is expressed as a cis sense-antisense chimeric transcript in prostate cancer cells. *RNA* 2010;6:1156–66.
- 53 Betran E, Wang W, Jin L, Long M. Evolution of the phosphoglyceratemutase processed gene in human and chimpanzee revealing the origin of a new primate gene. *Mol Biol Evol* 2002;19:654e663.
- 54 Zhang J, Wang X, Li M, Han J, Chen B, Wang B, Dai J. NANOGP8 is a retrogene expressed in cancers. *FEBS J* 2006;273:1723e1730.
- 55 Jeter CR, Badeaux M, Choy G, Chandra D, Patrawala L, Liu C, Calhoun-Davis T, Zaehres H, Daley GQ, Tang DG. Functional evidence that the self-renewal gene NANOG regulates human tumor development. *Stem Cells* 2009;27:993–1005.
- 56 Ishiguro T, Sato A, Ohata H, Sakai H, Nakagama H, Okamoto K. Differential expression of nanog1 and nanogp8 in colon cancer cells. *Biochem Biophys Biochem Biophys Res Commun* 2012;418:199–204.
- 57 Zou M, Baitei EY, Alzahrani AS, Al-Mohanna F, Farid NR, Meyer B, Shi Y. Oncogenic activation of MAP kinase by BRAF pseudogene in thyroid tumors. *Neoplasia* 2009;11:57–65.
- 58 Moreau-Aubry A, Le Guiner S, Labarrière N, Gesnel MC, Jotereau F, Breathnach R. A processed pseudogene codes for a new antigen recognized by a CD8(+) T cell clone on melanoma. *J Exp Med* 2000;191:1617–24.
- 59 Hendrickson RC, Cicinnati VR, Albers A, Dworacki G, Gambotto A, Pagliano O, Tüting T, Mayordomo JI, Visus C, Appella E, Shabanowitz J, Hunt DF, DeLeo AB. Identification of a 17betahydroxysteroid dehydrogenase type 12 pseudogene as the source of a highly restricted BALB/c Meth A tumor rejection peptide. *Cancer Immunol Immunother* 2010;59:113–24.
- 60 Hofree M, Shen JP, Carter H, Gross A, Ideker T. Network-based stratification of tumor mutations. *Nat Methods* 2013;10:1108–15.
- 61 Hoadley KA, Yau C, Wolf DM, Cherniack AD, Tamborero D, Ng S, Leiserson MD, Niu B, McLellan MD, Uzunangelov V, Zhang J, Kandoth C, Akbani R, Shen H, Omberg L, Chu A, Margolin AA, Van't Veer LJ, Lopez-Bigas N, Laird PW, Raphael BJ, Ding L, Robertson AG, Byers LA, Mills GB, Weinstein JN, Van Waes C, Chen Z, Collisson EA; Cancer Genome Atlas Research Network, Benz CC, Perou CM, Stuart JM. Multiplatform Analysis of 12 Cancer Types Reveals Molecular Classification within and across Tissues of Origin. *Cell* 2014;158:929–44.



Pseudogene in cancer: real functions and promising signature

Lu Xiao-Jie, Gao Ai-Mei, Ji Li-Juan and Xu Jiang

J Med Genet published online November 12, 2014

Updated information and services can be found at:

<http://jmg.bmj.com/content/early/2014/11/12/jmedgenet-2014-102785>

These include:

References

This article cites 60 articles, 11 of which you can access for free at:

<http://jmg.bmj.com/content/early/2014/11/12/jmedgenet-2014-102785#BIBL>

Email alerting service

Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Topic Collections

Articles on similar topics can be found in the following collections

[Editor's choice](#) (109)

[Molecular genetics](#) (1234)

Notes

To request permissions go to:

<http://group.bmj.com/group/rights-licensing/permissions>

To order reprints go to:

<http://journals.bmj.com/cgi/reprintform>

To subscribe to BMJ go to:

<http://group.bmj.com/subscribe/>