For reprint orders, please contact: reprints@futuremedicine.com

Not so pseudo anymore: pseudogenes as therapeutic targets

Pseudogenes are junk DNA gene remnants generated by inactivating mutations or the loss of regulatory sequences, often following gene duplication or retrotransposition events. These pseudogenes have previously been considered to be molecular fossils derived from once-coding genes. In many cases, pseudogenes confer no observable selective advantage to the host organism and may be on a path towards removal from the genome. However, pseudogenes can also serve as raw material for the exaptation of novel functions, particularly in relation to the regulation of gene expression. Many pseudogenes are resurrected as noncoding RNA genes, which function in RNA-based gene regulatory circuits. As such, functional pseudogenes might simply be considered as 'genes'. Here, we discuss the role of these pseudogene-derived RNAs as regulators of gene expression in the context of human disease. In particular, we consider the manipulation of pseudogene transcripts through the use of antisense oligonucleotides, siRNAs, aptamers or classical gene therapy approaches as novel pharmacological strategies.

KEYWORDS: epigenetics ■ lncRNAs ■ long noncoding RNAs ■ microRNA ■ miRNA ■ pseudogenes ■ therapeutics

Thomas C Roberts^{1,2} & Kevin V Morris^{*1,3}

¹Department of Molecular & Experimental Medicine, The Scripps Research Institute, 10550 N Torrey Pines Road, La Jolla, CA 92037, USA

²Department of Physiology, Anatomy & Genetics, University of Oxford, South Parks Road, Oxford, OX1 3QX, UK

³School of Biotechnology & Biomedical Sciences, University of New South Wales, NSW 2052, Australia

*Author for correspondence: kmorris@scripps.edu

A pseudogene is a genomic DNA sequence that is closely related to a paralogous parent gene but is deficient with respect to the parent gene function [1,2]. Typically, these are gene copies that have lost their protein-coding potential. Pseudogenes were first identified in *Xenopus laevis* [3] and have since been found in a range of organisms, including bacteria, plants and animals [4]. The estimated number of pseudogenes in the human genome is comparable to that of protein-coding genes (~20,000) [5,6].

In a classic 1972 essay, Ohno proposed the concept of junk DNA to explain the dramatic differences in genome sizes between organisms of similar levels of complexity/occupying similar niches (the so-called C-value paradox) [7]. The sequencing of the human genome subsequently demonstrated that protein-coding genes constitute only approximately 1% of the genome [8,9], and the remainder was assumed to be noncoding junk. Junk DNA has often been considered nonfunctional 'trash' and, by extension, pseudogenes have been dismissed as failed copies of genes that contribute nothing to the survival of the organism. However, Ohno was prescient enough to propose that junk DNA could also serve as a raw material for evolutionary innovation (similar ideas were also proposed by Britten and Davidson at approximately the same time [10,11]).

Some pseudogenes are clearly nonfunctional 'dying' genes. For example, comparative genomic

studies in mammals have revealed a great diversity in the number of functional and pseudogenized olfactory receptor genes [12]. Indeed, all olfactory receptor genes in the dolphin (*Stenella coeruleoalba*) genome are pseudogenes [13]. Similarly, the ancient gene *ACYL3* has been lost from the human lineage, but retained as a pseudogene [14]. Conversely, other pseudogenes have been 'resurrected' as long noncoding RNA (lncRNA) genes that adopt new functions, as in the case of *XIST*, the noncoding RNA initiator of X chromosome inactivation [15]. In this article, we discuss those pseudogenes that have evolved functions related to human disease and discuss the therapeutic potential of manipulating their expression.

Pseudogene biogenesis

There are three principal types of pseudogene (FIGURE 1A). In the simplest case, mutations lead to loss of coding potential or loss of expression. Such mutations include promoter disruption/absence, loss of the translation start codon, introduction of premature termination codons, splice-site alterations and insertions/deletions leading to a shift in the translation reading frame. This class of decayed, formerly functional genes are referred to as 'unitary' pseudogenes, as the parent gene is lost during their evolution. Unitary pseudogenes are the rarest class of pseudogenes (~100 in humans) [16]. For example, the *GULO*P gene (which encodes gulonolactone oxidase, an

Figure 1. Biogenesis and transcription of pseudogenes. (A) There are three types of pseudogene. Unitary pseudogenes are formed when the parental gene is degraded by mutations (indicated in red). Duplicated pseudogenes are produced when a gene loci is copied, often to a site proximal to the parental gene. The gene copy is subsequently pseudogenized by inactivating mutations. Processed pseudogenes are generated when an mRNA is reverse transcribed and then integrated back into the genome. As a result, processed pseudogenes lack the intron–exon structure and upstream regulatory regions of the parental gene. **(B)** Some pseudogenes are transcribed. This can occur when transcription is driven by duplicated upstream regulatory regions (i.e., promoter or enhancer sequences). Conversely, processed pseudogenes may integrate in sites adjacent to regulatory elements. Pseudogene transcription can occur in both sense and antisense orientations. Please see color figure at www.futuremedicine.com/doi/pdf/10.2217/pgs.13.172.

enzyme required for synthesis of vitamin C) is a pseudogene in primates and guinea pigs, whereas all other mammals carry a functional copy [17].

Gene duplication and divergence is a mechanism that is responsible for the evolution of gene families [18]. The presence of multiple copies of a gene permits the generation of novel variants, as selection pressure is relieved on all but one copy of the gene (deleterious mutations occurring in the gene copies do not adversely affect the survival of the organism). In some circumstances, this gives rise to new genes with novel functions (e.g., the α -globin gene cluster [19]); in others, it leads to pseudogenization. Pseudogenes produced by tandem genomic DNA duplications or unequal crossing over are often located in the vicinity of their

paralogous parent gene and generally maintain the intron–exon structure of their ancestral variant.

In contrast with other types of pseudogenes, processed pseudogenes are generated via an RNA intermediate. Mature mRNAs transcribed from the protein-coding parent gene are retrotransposed back into the host genome [1]. These sequences may subsequently accumulate disabling mutations. Retrotransposition, leading to the formation of processed pseudogenes, is mediated by a long interspersed nuclear element-encoded reverse transcriptase, which generates a cDNA copy that subsequently integrates back into the genome [20]. As such, processed pseudogenes are also known as retropseudogenes. Processed pseudogenes are typically nonsynthetic, indicating that

they are not the result of tandem gene duplications. The lack of upstream regulatory regions (i.e., promoters) or introns and the presence of poly-A tails are hallmarks of their derivation from mature mRNA via a cDNA intermediate. Similarly, processed pseudogenes are often flanked by inverted repeats that are relics of the integration process [21,22]. Processed pseudogenes are the most abundant class in humans and are believed to have arisen due to increased retrotranspositional activity in ancestral primates approximately 40–50 million years ago [4]. (Indeed, ~80% of processed pseudogenes in the human genome are specific to primates [23] and only 5% have potential mouse orthologs). The majority of processed pseudogenes arise from highly expressed genes, such as *GAPDH* and *KRT18*, or the genes for ribosomal proteins (e.g., *RPL21*), suggesting that expression level may be a key determinant of retrotranspositional potential [4,24,25].

Pseudogenes are transcribed

The application of tiling arrays, next-generation sequencing (RNA-seq) and pseudogene-specific rapid amplification of cDNA ends (RACE) have demonstrated that many pseudogene sequences produce lncRNA transcripts (also called pseudomRNAs or ψ mRNAs) [23,26,27]. Whereas many pseudogenes accumulate mutations in their promoter regions or their upstream regulatory elements, or integrate in transcriptionally silent regions leading to a loss of transcriptional potential, others may retain their regulatory sequences or be inserted adjacent to active promoter regions (FIGURE 1B) [21,27]. Estimates of the number of transcribed pseudogenes range from approximately 6% [27] to approximately 20% [23], and pseudomRNAs were shown to constitute approximately 10% of mouse cDNAs in the FANTOM3 collection [26], with processed pseudogenes tending to be transcribed more often than nonprocessed pseudogenes [21]. Pseudogene loci have also been shown to produce transcripts in the antisense orientation [28–30]. Given that antisense pseudomRNAs have the potential to hybridize with their complementary paralogous mRNAs, they are a potential source of novel gene regulators (discussed below).

Transcription of pseudogenes is consistent with the emerging paradigm of pervasive transcription (specifically, that transcription is widespread in eukaryotic genomes) [31]. Non-protein-coding transcripts represent the vast majority of the transcriptional output of the cell and represent a previously unappreciated level of complexity in the control of gene expression and cellular function.

Pseudogene functions

Many pseudogene sequences have exapted functional roles in gene regulation, although the full extent to which pseudogene-derived RNAs are functional is currently unknown. Pseudogenes were originally defined specifically as having lost their functionality [1]. By this definition, the term ‘functional pseudogene’ is an oxymoron and pseudogenes that are found to be functional are simply ‘genes’ (we acknowledge that the definition of ‘the gene’ is somewhat nebulous [32,33]). This argument has been used recently in criticisms of the interpretations of data from the ENCODE project [34]. However, the nomenclature persists as it provides insight into the genetic origins of pseudogene-derived RNAs and the relationship between pseudogene and parent gene, and for historical reasons.

Although pseudogenes are generally considered to be evolving neutrally, many show evidence of evolutionary conservation [35–37]. Specifically, Svensson *et al.* identified 30 transcribed pseudogenes that were conserved between humans and mice [37]. Khachane and Harrison showed that whereas only approximately 3% of transcribed pseudogenes were conserved between humans and mice, the conservation between humans and rhesus monkeys was much higher (~50%) [35]. Similarly, transcribed pseudogenes show evidence of being under selection pressure in these monkeys, and 68 transcribed pseudogenes were found to be syntenic between humans and at least two other mammals [35]. Additionally, pseudogenes that show higher levels of evolutionary constraint show a greater tendency towards being transcribed [38].

Taken together, these studies suggest that many pseudogenes have evolved under positive selective pressure, consistent with biological functionality. The fact that many of these pseudogene sequences are conserved but restricted to certain mammals (or specific to the primate lineage) suggests that they may represent relatively recent evolutionary innovations. There are interesting parallels with studies that show that the amount of noncoding DNA (corrected for ploidy) scales with organismal complexity [39], whereas the number of protein-coding genes does not [40]. It is tempting to speculate that some of the novel gene-regulatory functions that result from exapted pseudogene sequences may account for the differences between organisms with similar coding gene repertoires.

The functionality of many lncRNA transcripts often manifests through the formation of secondary structure motifs that can act as

binding domains for proteins or other nucleic acids. Positive selection may be acting to conserve these structural features, but this would not be apparent at the primary sequence level [41]. As a result, sequence-based evolutionary comparisons and analysis could be misleading. As templates for lncRNAs, pseudogenes may be subject to this type of selection, although a structure-based view of conservation analysis would be required to detect this.

Transcribed pseudogenes show tissue-specific patterns of expression [23,38] and multiple studies have reported testis-specific pseudogene transcription [21,23,42,43]. In some cases, pseudogenes are expressed in tissues in which the parent gene is not. For example, the 5-HT₇ receptor (HTR7) has a pseudogene that is expressed in liver and kidney, which are tissues that do not express the parent *HTR7* gene [44,45]. Furthermore, expression of some pseudogenes is dynamic and responds to physiological signals, such as cell stress [43]. Pseudogene expression has also been implicated in the control of neuronal differentiation [46]. These observations are indicative of tightly regulated expression of pseudogene transcription, consistent with biological function. However, it could equally be argued that regulated pseudogene transcription is a consequence of differential transcription factor expression, leading to tissue-specific or stress-responsive 'leaky' transcription.

Of the few studies to date, several have identified a link between pseudogene misexpression and human disease (and cancer in particular). For example, transcription of the *MYLKPI* pseudogene [47] and pseudogenes of Oct-4 (*POUF51*) [48] is activated in cancers. Similarly, a pseudogene of Connexin 43 (*PsiCx43*) was found to be expressed in breast cancer cells, but not in normal cells. Inhibition of *PsiCx43* resulted in an increase in Connexin 43 expression and sensitized cells to cytotoxic chemotherapy [49]. Pseudogenes of BRAF were expressed in thyroid cancers, where they activate the MAPK pathway and promote oncogenic transformation [50]. Similarly, pseudogene-derived transcripts are upregulated in laryngeal cancer [51] and neuroblastoma [52]. Differential expression of pseudogene transcripts has also been implicated in the context of asthma [53] and HIV infection [54].

Gene regulation by pseudogene transcripts

Pseudogene transcripts can regulate expression of their parental genes if there is complementarity or homology between the two transcripts. The *HMGAI* gene regulates expression of the

insulin receptor and thereby contributes to insulin resistance in Type 2 diabetes. Overexpression of an *HMGAI* pseudogene (*HMGAI-p*) in diabetic patients resulted in destabilization of *HMGAI* mRNA through competition for the cytoplasmic RNA-stabilizing factor α CP1, which binds to a common site in the 3'-UTR of both transcripts, thereby contributing to insulin resistance (FIGURE 2A) [55]. Similarly, expression of a sense orientation, duplicated pseudogene *Makorin1-p1* stabilizes the parental *Makorin1* mRNA via a 5'-RNA decay element that is homologous between the two transcripts [56]. It is possible that the *Makorin1-p1* pseudotranscript also competes for a *trans*-acting destabilizing factor (although this notion has been challenged by another group [57]). Another example of pseudogene-mediated post-transcriptional gene regulation was observed in the CNS of the snail *Lymnaea stagnalis*. The gene encoding neuronal nitric oxide synthase (nNOS) has a pseudogene that produces an antisense pseudo-nNOS transcript and acts as a negative regulator of parental nNOS expression. The parental- and pseudo-nNOS transcripts hybridize, leading to suppression of nNOS translation (FIGURE 2B) [29].

Crosstalk between pseudogenes & the RNAi pathway

RNAi is an endogenous gene regulation mechanism mediated by small RNA-guided ribonucleoprotein complexes [58]. In humans, the primary role of the RNAi pathway is in the processing and function of miRNAs. miRNAs are 21–23 nucleotide ssRNA molecules that act primarily as post-transcriptional regulators of gene expression. miRNAs are initially expressed as long primary miRNA transcripts that contain one or more stem loop structures [59]. These stem loop hairpins are excised from the primary transcript by the enzyme Drosha [60] and the loop sequence is subsequently removed by Dicer in order to generate a small RNA duplex [61]. This duplex RNA is loaded into Argonaute-2 (the catalytic component of the RNA-induced silencing complex), whereby one strand is preferentially retained (i.e., the mature miRNA) and the other discarded [62]. The RNA-induced silencing complex is guided by the mature miRNA to complementary mRNA targets, which are subsequently silenced by translational repression and/or mRNA destabilization [63,64]. The duplex formed between an miRNA and its target mRNA is typically imperfect, requiring complementarity only in the 'seed' region (nucleotides 2–8). Importantly, miRNA targets have also been identified with little or no seed region

complementarity [65]. Conversely, complete complementarity between the small RNA guide and the target also result in gene silencing, although via a different mechanism. In this case, the target mRNA is cleaved by the slicer activity of Argonaute-2 and the resulting transcript fragments are destroyed by the cellular RNA degradation machinery. This activity has been widely exploited through the use siRNAs in the study of gene function and for the development of novel therapeutics. siRNAs are 21-nucleotide RNA duplexes with 2-nucleotide 3' overhangs. This structure is typical of the products of RNase III digestion (i.e., Droscha and Dicer), and so siRNAs mimic the later stages of miRNA processing. It has recently been shown that siRNAs are also produced in an endogenous context (endo-siRNAs or esiRNAs) by a non-miRNA processing pathway [66,67].

Endo-siRNA

Transcribed pseudogenes are sources of endo-siRNA clusters, which typically generate hundreds of unique small RNAs that can enter the RNAi pathway and induce gene silencing of complementary mRNA targets (a phenomenon that is especially apparent in oocytes) [68,69]. This can occur in a number of ways. Some pseudo-mRNA transcripts can adopt long hairpin structures that are processed by Dicer to produce endo-siRNAs. For example, long hairpin-processed endo-siRNAs derived from the *Au76* pseudogene are capable of regulating the parental gene, *Rangap1*, in *trans* (FIGURE 3A) [68]. Similarly, overlapping pseudo-mRNA and/or parental gene mRNA transcripts can hybridize and form substrates for Dicer processing, thereby generating clusters of endo-siRNAs, as exemplified by the *Pppr41* gene and its cognate pseudogene (FIGURE 3B). Conversely, in the case of the *Hdac1* pseudogene, endo-siRNAs are produced by a series of overlapping sense and antisense pseudogene-derived transcripts. The resulting endo-siRNAs regulate *Hdac1* expression but are not derived from *Hdac1* mRNA transcripts *per se* (FIGURE 3C) [69]. Pseudogene-derived endo-siRNAs have been shown to act as tumor suppressors, as in the case of the pseudogene ψ *PPMIK*, which produces endo-siRNAs that inhibit cell growth in the context of hepatocellular carcinoma [70].

PIWI-interacting RNA

Pseudogene transcripts also give rise to PIWI-interacting RNAs [68]. PIWI-interacting RNAs are approximately 26–31 nucleotide ssRNAs that act to silence transposable elements in the germline [71]. Transposon activity in the

Figure 2. Post-transcriptional pseudogene regulation. (A) Pseudogene transcripts can compete with their cognate parental mRNAs for RNA stability factors. For example, *HMGA1* mRNA and its pseudogene (*HMGA1-p*) contain a common 3' decay element that binds to the stability factor α CP1. **(B)** Pseudogene transcripts can also interact directly with the parental mRNA to influence its expression. For example, *nNOS* mRNA hybridizes with an antisense *nNOS* pseudogene transcript (*nNOS-p*), which results in suppression of *nNOS* translation.

germline can lead to infertility or the accumulation of deleterious mutations and so controlling transposable elements is of paramount importance to the survival of the species [72,73]. The testis-specific expression of some pseudogenes and the detection of endo-siRNAs in oocytes suggest that pseudogenes are involved in germline defense. The generation of endo-siRNAs in the germline may provide an explanation for why the slicer activity of Argonaute-2 has been conserved throughout evolution (given that slicer activity is dispensable for canonical miRNA functionality).

Competing endogenous RNA

Pseudogene transcripts can also act as miRNA sponges/decoys. An miRNA sponge is an RNA transcript containing multiple copies of binding sites for one or more miRNAs [74]. miRNA sponges compete with target transcripts for binding of miRNAs and, as such, have also been called competing endogenous RNAs (ceRNAs) [75]. Pseudogene-derived lncRNAs can act as ceRNAs that protect their parental paralogous mRNAs by sequestering miRNAs that would otherwise repress their expression (FIGURE 3D). Whereas exogenous miRNA sponges typically consist of multiple identical target sites designed to inhibit a specific miRNA, pseudogene-derived ceRNAs can be considered 'perfect decoys', as they sequester a set of miRNAs that converge on their parental gene [76]. Importantly, mRNAs can also act as ceRNAs [77–79]. The non-protein-coding nature of expressed pseudogenes makes them ideal ceRNAs, as it allows them to act as sinks for miRNAs independent of the protein-coding potential of the decoy. For example, *PTENPI*, the pseudogene of the tumor-suppressor gene *PTEN*, has lost its ability to encode the

Figure 3. Crosstalk between pseudogenes and the RNAi pathway. Pseudogene transcripts are sources of endo-siRNA clusters. **(A)** Pseudogene transcripts can fold into long hairpin structures that are processed by Dicer into approximately 21-bp endo-siRNA duplexes. Endo-siRNA-producing Dicer substrates are also generated when **(B)** the parental mRNA binds with an antisense pseudogene transcript or when **(C)** complementary sense and antisense transcripts derived from pseudogene loci hybridize. Endo-siRNAs produced by these mechanisms are complementary to the paralogous parental gene and act to regulate its expression at the post-transcriptional level. **(D)** Pseudogene transcripts can act as ceRNAs by competing with their parental mRNAs for miRNA binding. ceRNA: Competing endogenous RNA.

PTEN protein as a result of a missense mutation at the translation start codon [52], but otherwise shares close homology with the processed

PTEN mRNA (including miRNA binding sites for miR-17, miR-19, miR-21, miR-26 and miR-214). *PTENP1* can act as a sponge for these

miRNAs and thereby relieve the miRNA-mediated repression of *PTEN* expression [30,76]. As a result, *PTENP1* acts to finetune *PTEN* expression, which is of high importance given that subtle changes in PTEN protein levels can lead to tumor formation [80,81]. *PTENP1* is thus a tumor-suppressor gene in its own right. The importance of the *PTENP1* pseudogene is underlined by the observation that this locus is deleted in melanoma, leading to increased miRNA-mediated suppression of PTEN and promotion of tumor progression [82]. Other cancer-related genes also show signs of pseudogene-mediated ceRNA regulation. For example, the expression of *KRAS* and *KRASIP* is positively correlated in prostate cancer [76].

Pseudogene-mediated epigenetic control

Pseudogene-derived RNAs have also been shown to regulate epigenetic processes. The Oct4 gene (*POU5F1*), which acts as a driver of pluripotency, has six related pseudogenes [48]. Of these, *Oct4-pg5* produces an antisense transcript capable of mediating epigenetic silencing of the parental *Oct4* locus [28]. Similarly, *PTENP1* was also shown to produce antisense RNAs capable of regulating the *PTEN* locus. One antisense *PTNEP1* isoform was found to localize to the *PTEN* promoter and induce transcriptional silencing via recruitment of the H3K27 methyltransferase EZH2 and DNMT3A (FIGURE 4) [30]. Interestingly, this study also showed that a different *PTENP1* antisense isoform regulates the stability of the sense *PTENP1* transcript and thereby modulates its activity as a ceRNA. As a result, *PTENP1* pseudogene transcription regulates PTEN expression at multiple levels. Given that epigenetic control is a major function of lncRNAs [83–86], it is likely that other pseudogene transcripts may be involved in epigenetic gene regulation.

Therapeutic targeting of pseudogene RNAs

Inasmuch as pseudogene-derived transcripts are involved in the regulation of gene expression in a disease context, their expression can be manipulated for therapeutic benefit. A number of tools are available to alter the expression of noncoding RNA transcripts [87]. Antisense oligonucleotides (AOs) have been widely used to inhibit gene function and show promise in clinical trials [88]. Typically, one of two AO designs is used: gapmers and mixmers (when such technologies have been utilized to target

natural antisense transcripts, they have also been known as antagoNATs [89]). Gapmers are short oligonucleotide molecules that consist of RNA-based flanking sequences and an internal DNA ‘gap’ region (FIGURE 5A). Binding to a complementary RNA target results in the formation of an RNA–DNA heteroduplex, which is a substrate for cleavage by the enzyme RNase H [90]. RNase H is primarily located in the nucleus and so gapmer technology is ideal for targeting noncoding RNA transcripts with nuclear functions. Conversely, mixmers consist of alternating nucleic acid chemistries and are designed to sterically block association of the target transcript with other nucleic acids or ribonucleoproteins [91]. Mixmers are designed so that they do not contain strings of consecutive DNA nucleotides and are therefore not RNase H competent (FIGURE 5B). Mixmers could be used to block direct association between a pseudogene transcript and its cognate parental mRNA, or to inhibit the binding of proteins, such as epigenetic remodeling complexes. Alternatively, RNAi could also be utilized to knockdown expression of pseudogene-derived transcripts. In this case, exogenously introduced siRNAs (FIGURE 5C) or vector-expressed short hairpin RNAs could be used to target specific pseudogene transcripts, although their predilection for the RNAi pathway can result in some level of exclusion from nuclear processes.

A limitation of these approaches is that the high levels of homology between pseudogenes and their parental protein-coding genes mean that target sites for AOs and RNAi effectors

Figure 4. Pseudogene-mediated epigenetic control. Antisense pseudogene transcripts bind to epigenetic modifying factors (e.g., EZH2 and DNMT3A) and recruit them to the parental gene locus, leading to heterochromatinization (compact histones are indicated in turquoise) and DNA methylation (dark blue lollipops) of the promoter region. As a result, the parental gene is silenced at the transcriptional level.

Please see color figure at www.futuremedicine.com/doi/pdf/10.2217/pgs.13.172.

may be limited to nonhomologous regions or SNP-containing regions. As a result, potential off-target downregulation of the parental mRNA must be considered carefully. Moreover, many lncRNAs most likely form complex secondary structures, such structures could limit the access of the pseudogene RNA to oligonucleotide targeting [41]. The use of aptamers could provide a solution to this problem [92]. Aptamers are structured oligonucleotides that are developed by *in vitro*- or *in vivo*-directed evolution and bind to protein or nucleic acid targets with high affinity and specificity (FIGURE 5D) [93]. A number of nucleic acid chemical modifications have been developed in order to impart drug-like properties on AOs, siRNAs and aptamers (FIGURE 5E). For example, modifications to the oligonucleotide backbone or 2' ribose position alter the target binding affinity, reduce clearance, increase nuclease stability and improve pharmacokinetic properties [94,95].

Conversely, deficiency of a pseudogene transcript could be complemented by exogenous expression, mediated by a viral vector in a classical gene therapy approach. For example, in the case of *PTENP1* loss in melanoma, overexpression of a *PTENP1* transgene could restore tumor-suppressor activity and retard tumor growth.

Future perspective

For some genes, the loss of coding potential and pseudogenization leads to gene death, and ultimately removal from the genome. For others, inactive genes are resurrected as noncoding RNA genes and neo-functionalized as gene-regulatory elements. Pseudogene-derived transcripts can regulate gene expression by generating small

RNAs, regulating mRNA stability via direct binding or sequestration of *trans*-acting RNA decay proteins, sponging of endogenous miRNAs and promoter recruitment of epigenetic remodeling complexes. The extent to which pseudogenes are important in gene regulation and the pathogenesis of disease has so far been underappreciated for a number of reasons. First, only a relatively small number of functional pseudogenes have been extensively characterized. Second, functional pseudogenes exert their effects on gene expression via a plethora of mechanisms that are not always easily inferred by analysis of the primary nucleic acid sequence. Third, working with pseudogenes presents technical difficulties due to the close homology between the pseudo- and parental-mRNA sequences. Finally, the lineage specificity of pseudogenes and particularly the lack of conservation between humans and mice may prove to be an obstacle in the development of pseudogene-targeted therapeutics. Nevertheless, pseudogene-transcribed lncRNAs are emerging as both important regulators of gene expression and as promising novel targets for pharmacological intervention.

Information resource

- This study (published during the preparation of this manuscript) identified a pseudogene lncRNA at the centre of the regulation of inflammation: Rapicavoli NA, Qu K, Zhang J, Mikhail M, Laberge R-M, Chang HY. A mammalian pseudogene lncRNA at the interface of inflammation and anti-inflammatory therapeutics. *eLife* (Internet), 2 (2013). <http://elife.elifesciences.org/content/2/e00762>

Executive summary

Pseudogene biogenesis

- Pseudogenes are defective copies of paralogous genes.
- There are three types of pseudogene: unitary, duplicated and processed (retrotransposed).

Pseudogenes are transcribed

- Many pseudogenes produce RNA transcripts in sense and/or antisense orientations.
- Pseudogene transcripts show tissue-specific patterns of expression.

Pseudogene functions

- Some pseudogenes show signs of evolutionary conservation (e.g., positive selection and synteny, among others).
- Many pseudogenes are misexpressed in human disease.
- Pseudogene transcripts can exert gene-regulatory effects by regulating translation, competing for *trans*-acting stabilizing proteins, competing with parental mRNAs for miRNA binding, producing endo-siRNAs and directing epigenetic remodeling at specific promoters.

Therapeutic targeting of pseudogene RNAs

- The expression of disease-associated pseudogene transcripts can be manipulated using antisense oligonucleotides, siRNAs and aptamers.

Financial & competing interests disclosure

This project was supported by NIAID R56 AI096861-01 and PO1 AI099783-01 (to KV Morris) and CA151574-01 and RO1 CA153124-01 (to KV Morris). The authors have no other relevant affiliations or financial involvement with any

organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

References

Papers of special note have been highlighted as:
■ of interest

- Vanin EF. Processed pseudogenes: characteristics and evolution. *Annu. Rev. Genet.* 19, 253–272 (1985).
- Mighell AJ, Smith NR, Robinson PA, Markham AF. Vertebrate pseudogenes. *FEBS Lett.* 468(2–3), 109–114 (2000).
- Jacq C, Miller JR, Brownlee GG. A pseudogene structure in 5S DNA of *Xenopus laevis*. *Cell* 12(1), 109–120 (1977).
- Ohshima K, Hattori M, Yada T, Gojobori T, Sakaki Y, Okada N. Whole-genome screening indicates a possible burst of formation of processed pseudogenes and Alu repeats by particular L1 subfamilies in ancestral primates. *Genome Biol.* 4(11), R74 (2003).
- Harrison PM, Hegyi H, Balasubramanian S *et al.* Molecular fossils in the human genome: identification and analysis of the pseudogenes in chromosomes 21 and 22. *Genome Res.* 12(2), 272–280 (2002).
- Gonçalves I, Duret L, Mouchiroud D. Nature and structure of human genes that generate retropseudogenes. *Genome Res.* 10(5), 672–678 (2000).
- Ohno S. So much 'junk' DNA in our genome. *Brookhaven Symp. Biol.* 23, 366–370 (1972).
- Venter JC, Adams MD, Myers EW *et al.* The sequence of the human genome. *Science* 291(5507), 1304–1351 (2001).
- Lander ES, Linton LM, Birren B *et al.* Initial sequencing and analysis of the human genome. *Nature* 409(6822), 860–921 (2001).
- Britten RJ, Davidson EH. Gene regulation for higher cells: a theory. *Science* 165(3891), 349–357 (1969).
- Britten RJ, Davidson EH. Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. *Q. Rev. Biol.* 46(2), 111–138 (1971).
- Niimura Y, Nei M. Extensive gains and losses of olfactory receptor genes in mammalian evolution. *PLoS ONE* 2(8), e708 (2007).
- Freitag J, Ludwig G, Andreini I, Rössler P, Breer H. Olfactory receptors in aquatic and terrestrial vertebrates. *J. Comp. Physiol. A* 183(5), 635–650 (1998).
- Zhu J, Sanborn JZ, Diekhans M, Lowe CB, Pringle TH, Haussler D. Comparative genomics search for losses of long-established genes on the human lineage. *PLoS Comput. Biol.* 3(12), e247 (2007).
- Duret L, Chureau C, Samain S, Weissenbach J, Avner P. The *Xist* RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. *Science* 312(5780), 1653–1655 (2006).
- Zhang ZD, Frankish A, Hunt T, Harrow J, Gerstein M. Identification and analysis of unitary pseudogenes: historic and contemporary gene losses in humans and other primates. *Genome Biol.* 11(3), R26 (2010).
- Nishikimi M, Yagi K. Molecular basis for the deficiency in humans of gulonolactone oxidase, a key enzyme for ascorbic acid biosynthesis. *Am. J. Clin. Nutr.* 54(6 Suppl.), 1203S–1208S (1991).
- Soskine M, Tawfik DS. Mutational effects and the evolution of new protein functions. *Nat. Rev. Genet.* 11(8), 572–582 (2010).
- Hill AV, Nicholls RD, Thein SL, Higgs DR. Recombination within the human embryonic xi-globin locus: a common xi–xi chromosome produced by gene conversion of the psi xi gene. *Cell* 42(3), 809–819 (1985).
- Esnault C, Maestre J, Heidmann T. Human LINE retrotransposons generate processed pseudogenes. *Nat. Genet.* 24(4), 363–367 (2000).
- Shows the molecular mechanism through which retrotransposition generates processed pseudogenes.
- Vinckenbosch N, Dupanloup I, Kaessmann H. Evolutionary fate of retroposed gene copies in the human genome. *Proc. Natl Acad. Sci. USA* 103(9), 3220–3225 (2006).
- Jurka J. Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. *Proc. Natl Acad. Sci. USA* 94(5), 1872–1877 (1997).
- Zheng D, Frankish A, Baertsch R *et al.* Pseudogenes in the ENCODE regions: consensus annotation, analysis of transcription, and evolution. *Genome Res.* 17(6), 839–851 (2007).
- Zhang Z, Harrison P, Gerstein M. Identification and analysis of over 2000 ribosomal protein pseudogenes in the human genome. *Genome Res.* 12(10), 1466–1482 (2002).
- McDonnell L, Drouin G. The abundance of processed pseudogenes derived from glycolytic genes is correlated with their expression level. *Genome* 55(2), 147–151 (2012).
- Frith MC, Wilming LG, Forrest A *et al.* Pseudo-messenger RNA: phantoms of the transcriptome. *PLoS Genet.* 2(4), e23 (2006).
- Harrison PM, Zheng D, Zhang Z, Carriero N, Gerstein M. Transcribed processed pseudogenes in the human genome: an intermediate form of expressed retrosequence lacking protein-coding ability. *Nucleic Acids Res.* 33(8), 2374–2383 (2005).
- Hawkins PG, Morris KV. Transcriptional regulation of Oct4 by a long non-coding RNA antisense to Oct4-pseudogene 5. *Transcription* 1(3), 165–175 (2010).
- 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.* 19(18), 7711–7720 (1999).
- Demonstration of post-transcriptional gene regulation by a pseudogene-derived antisense transcript.
- Johnsson P, Ackley A, Vidarsdottir L *et al.* A pseudogene long-noncoding-RNA network regulates PTEN transcription and translation in human cells. *Nat. Struct. Mol. Biol.* 20(4), 440–446 (2013).
- First demonstration of epigenetic regulation by a pseudogene-derived lncRNA.
- Clark MB, Amaral PP, Schlesinger FJ *et al.* The reality of pervasive transcription. *PLoS Biol.* 9(7), e1000625; discussion e1001102 (2011).
- Gerstein MB, Bruce C, Rozowsky JS *et al.* What is a gene, post-ENCODE? history and updated definition. *Genome Res.* 17(6), 669–681 (2007).
- Djebali S, Davis CA, Merkel A *et al.* Landscape of transcription in human cells. *Nature* 489(7414), 101–108 (2012).
- Graur D, Zheng Y, Price N, Azevedo RBR, Zufall RA, Elhaik E. On the immortality of television sets: 'function' in the human genome according to the evolution-free gospel of ENCODE. *Genome Biol. Evol.* 5(3), 578–590 (2013).

- 35 Khachane AN, Harrison PM. Assessing the genomic evidence for conserved transcribed pseudogenes under selection. *BMC Genomics* 10, 435 (2009).
- 36 Podlaha O, Zhang J. Nonneutral evolution of the transcribed pseudogene *Makorin1-p1* in mice. *Mol. Biol. Evol.* 21(12), 2202–2209 (2004).
- 37 Svensson O, Arvestad L, Lagergren J. Genome-wide survey for biologically functional pseudogenes. *PLoS Comput. Biol.* 2(5), e46 (2006).
- 38 Pei B, Sisu C, Frankish A *et al.* The GENCODE pseudogene resource. *Genome Biol.* 13(9), R51 (2012).
- 39 Taft RJ, Mattick JS. Increasing biological complexity is positively correlated with the relative genome-wide expansion of non-protein-coding DNA sequences. *Genome Biol.* 5(1), P1 (2003).
- 40 Baltimore D. Our genome unveiled. *Nature* 409(6822), 814–816 (2001).
- 41 Novikova IV, Hennelly SP, Sanbonmatsu KY. Structural architecture of the human long non-coding RNA, steroid receptor RNA activator. *Nucleic Acids Res.* 40(11), 5034–5051 (2012).
- 42 Kleene KC, Mulligan E, Steiger D, Donohue K, Mastrangelo MA. The mouse gene encoding the testis-specific isoform of poly(A) binding protein (*Pabp2*) is an expressed retroposon: intimations that gene expression in spermatogenic cells facilitates the creation of new genes. *J. Mol. Evol.* 47(3), 275–281 (1998).
- 43 Kalyana-Sundaram S, Kumar-Sinha C, Shankar S *et al.* Expressed pseudogenes in the transcriptional landscape of human cancers. *Cell* 149(7), 1622–1634 (2012).
- 44 Olsen MA, Schechter LE. Cloning, mRNA localization and evolutionary conservation of a human 5-HT7 receptor pseudogene. *Gene* 227(1), 63–69 (1999).
- 45 Elliman SJ, Wu I, Kemp DM. Adult tissue-specific expression of a Dppa3-derived retrogene represents a postnatal transcript of pluripotent cell origin. *J. Biol. Chem.* 281(1), 16–19 (2006).
- 46 Lin M, Pedrosa E, Shah A *et al.* RNA-Seq of human neurons derived from iPS cells reveals candidate long non-coding RNAs involved in neurogenesis and neuropsychiatric disorders. *PLoS ONE* 6(9), e23356 (2011).
- 47 Han YJ, Ma SF, Yourek G, Park YD, Garcia JGN. A transcribed pseudogene of MYLK promotes cell proliferation. *FASEB J.* 25(7), 2305–2312 (2011).
- 48 Suo G, Han J, Wang X *et al.* *Oct4* pseudogenes are transcribed in cancers. *Biochem. Biophys. Res. Commun.* 337(4), 1047–1051 (2005).
- 49 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.* 8(4), 786–793 (2009).
- 50 Zou M, Baitei EY, Alzahrani AS *et al.* Oncogenic activation of MAP kinase by *BRAF* pseudogene in thyroid tumors. *Neoplasia* 11(1), 57–65 (2009).
- 51 Feenstra M, Bakema J, Verdaasdonk M *et al.* Detection of a putative *HLA-A*31012* processed (intronless) pseudogene in a laryngeal squamous cell carcinoma. *Genes Chromosomes Cancer* 27(1), 26–34 (2000).
- 52 Fujii GH, Morimoto AM, Berson AE, Bolen JB. Transcriptional analysis of the PTEN/MMAC1 pseudogene, *psiPTEN*. *Oncogene* 18(9), 1765–1769 (1999).
- 53 Tsitsiou E, Williams AE, Moschos SA *et al.* Transcriptome analysis shows activation of circulating CD8⁺ T cells in patients with severe asthma. *J. Allergy Clin. Immunol.* 129(1), 95–103 (2012).
- 54 Chang ST, Sova P, Peng X *et al.* Next-generation sequencing reveals HIV-1-mediated suppression of T cell activation and RNA processing and regulation of noncoding RNA expression in a CD4⁺ T cell line. *MBio* 2(5), pii: e00134-11 (2011).
- 55 Chiefari E, Iiritano S, Paonessa F *et al.* Pseudogene-mediated posttranscriptional silencing of HMGA1 can result in insulin resistance and Type 2 diabetes. *Nat. Commun.* 1, 40 (2010).
- 56 Hirotsune S, Yoshida N, Chen A *et al.* An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene. *Nature* 423(6935), 91–96 (2003).
- 57 Gray TA, Wilson A, Fortin PJ, Nicholls RD. The putatively functional *Mkrn1-p1* pseudogene is neither expressed nor imprinted, nor does it regulate its source gene in *trans*. *Proc. Natl Acad. Sci. USA* 103(32), 12039–12044 (2006).
- 58 Filipowicz W, Jaskiewicz L, Kolb FA, Pillai RS. Post-transcriptional gene silencing by siRNAs and miRNAs. *Curr. Opin. Struct. Biol.* 15(3), 331–341 (2005).
- 59 Lee Y, Kim M, Han J *et al.* MicroRNA genes are transcribed by RNA polymerase II. *EMBO J.* 23(20), 4051–4060 (2004).
- 60 Lee Y, Ahn C, Han J *et al.* The nuclear RNase III Drosha initiates microRNA processing. *Nature* 425(6956), 415–419 (2003).
- 61 Bernstein E, Caudy AA, Hammond SM, Hannon GJ. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* 409(6818), 363–366 (2001).
- 62 Liu J, Carmell MA, Rivas FV *et al.* Argonaute2 is the catalytic engine of mammalian RNAi. *Science* 305(5689), 1437–1441 (2004).
- 63 Zeng Y, Yi R, Cullen BR. MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms. *Proc. Natl Acad. Sci. USA* 100(17), 9779–9784 (2003).
- 64 Guo H, Ingolia NT, Weissman JS, Bartel DP. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* 466(7308), 835–840 (2010).
- 65 Helwak A, Kudla G, Dudnakova T, Tollervey D. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell* 153(3), 654–665 (2013).
- 66 Yang N, Kazazian HH. L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. *Nat. Struct. Mol. Biol.* 13(9), 763–771 (2006).
- 67 Kawaji H, Nakamura M, Takahashi Y *et al.* Hidden layers of human small RNAs. *BMC Genomics* 9, 157 (2008).
- 68 Watanabe T, Totoki Y, Toyoda A *et al.* Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* 453(7194), 539–543 (2008).
- 69 Tam OH, Aravin AA, Stein P *et al.* Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* 453(7194), 534–538 (2008).
- 70 Chan WL, Yuo CY, Yang WK *et al.* Transcribed pseudogene *ψPPM1K* generates endogenous siRNA to suppress oncogenic cell growth in hepatocellular carcinoma. *Nucleic Acids Res.* 41(6), 3734–3747 (2013).
- 71 Aravin AA, Sachidanandam R, Bourc'his D *et al.* A piRNA pathway primed by individual transposons is linked to *de novo* DNA methylation in mice. *Mol. Cell* 31(6), 785–799 (2008).
- 72 Aravin AA, Sachidanandam R, Girard A, Fejes-Toth K, Hannon GJ. Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science* 316(5825), 744–747 (2007).
- 73 Carmell MA, Girard A, van de Kant HJG *et al.* MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Dev. Cell* 12(4), 503–514 (2007).
- 74 Franco-Zorrilla JM, Valli A, Todesco M *et al.* Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat. Genet.* 39(8), 1033–1037 (2007).
- 75 Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis: the Rosetta stone of a hidden RNA language? *Cell* 146(3), 353–358 (2011).

- 76 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* 465(7301), 1033–1038 (2010).
- **Shows how a pseudogene transcript can compete with mRNAs for miRNA binding.**
- 77 Tay Y, Kats L, Salmena L *et al.* Coding-independent regulation of the tumor suppressor PTEN by competing endogenous mRNAs. *Cell* 147(2), 344–357 (2011).
- 78 Sumazin P, Yang X, Chiu HS *et al.* An extensive microRNA-mediated network of RNA–RNA interactions regulates established oncogenic pathways in glioblastoma. *Cell* 147(2), 370–381 (2011).
- 79 Karreth FA, Tay Y, Perna D *et al.* *In vivo* identification of tumor-suppressive *PTEN* ceRNAs in an oncogenic *BRAF*-induced mouse model of melanoma. *Cell* 147(2), 382–395 (2011).
- 80 Trotman LC, Niki M, Dotan ZA *et al.* Pten dose dictates cancer progression in the prostate. *PLoS Biol.* 1(3), E59 (2003).
- 81 Alimonti A, Carracedo A, Clohessy JG *et al.* Subtle variations in Pten dose determine cancer susceptibility. *Nat. Genet.* 42(5), 454–458 (2010).
- 82 Polisenio L, Haimovic A, Christos PJ *et al.* Deletion of *PTENP1* pseudogene in human melanoma. *J. Invest. Dermatol.* 131(12), 2497–2500 (2011).
- 83 Khalil AM, Guttman M, Huarte M *et al.* Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl Acad. Sci. USA* 106(28), 11667–11672 (2009).
- 84 Gupta RA, Shah N, Wang KC *et al.* Long non-coding RNA *HOTAIR* reprograms chromatin state to promote cancer metastasis. *Nature* 464(7291), 1071–1076 (2010).
- 85 Mayer C, Schmitz KM, Li J, Grummt I, Santoro R. Intergenic transcripts regulate the epigenetic state of rRNA genes. *Mol. Cell* 22(3), 351–361 (2006).
- 86 Yu W, Gius D, Onyango P *et al.* Epigenetic silencing of tumour suppressor gene *p15* by its antisense RNA. *Nature* 451(7175), 202–206 (2008).
- 87 Roberts TC, Wood MJA. Therapeutic targeting of non-coding RNAs. *Essays Biochem.* 54(1), 127–145 (2013).
- 88 Erba HP, Sayar H, Juckett M *et al.* Safety and pharmacokinetics of the antisense oligonucleotide (ASO) LY2181308 as a single-agent or in combination with idarubicin and cytarabine in patients with refractory or relapsed acute myeloid leukemia (AML). *Invest. New Drugs* 31(4), 1023–1034 (2013).
- 89 Modarresi F, Faghihi MA, Lopez-Toledano MA *et al.* Inhibition of natural antisense transcripts *in vivo* results in gene-specific transcriptional upregulation. *Nat. Biotechnol.* 30(5), 453–459 (2012).
- 90 Jepsen JS, Wengel J. LNA-antisense rivals siRNA for gene silencing. *Curr. Opin. Drug Discov. Devel.* 7(2), 188–194 (2004).
- 91 Ivanova G, Reigadas S, Ittig D *et al.* Tricyclo-DNA containing oligonucleotides as steric block inhibitors of human immunodeficiency virus type 1 tat-dependent trans-activation and HIV-1 infectivity. *Oligonucleotides* 17(1), 54–65 (2007).
- 92 Tsai MC, Spitale RC, Chang HY. Long intergenic noncoding RNAs: new links in cancer progression. *Cancer Res.* 71(1), 3–7 (2011).
- 93 Sundaram P, Kurniawan H, Byrne ME, Wower J. Therapeutic RNA aptamers in clinical trials. *Eur. J. Pharm. Sci.* 48(1–2), 259–271 (2013).
- 94 Rettig GR, Behlke MA. Progress toward *in vivo* use of siRNAs-II. *Mol. Ther.* 20(3), 483–512 (2012).
- 95 Choung S, Kim YJ, Kim S, Park HO, Choi YC. Chemical modification of siRNAs to improve serum stability without loss of efficacy. *Biochem. Biophys. Res. Commun.* 342(3), 919–927 (2006).