Opinion TRENDS in Genetics Vol.23 No.5

ELSEVIER
Full text provided by www.sciencedirect.com
ScienceDirect

The ambiguous boundary between genes and pseudogenes: the dead rise up, or do they?

Deyou Zheng[1] and Mark B. Gerstein[1,2,3]

[1] Department of Molecular Biophysics and Biochemistry, Yale University, 266 Whitney Avenue, New Haven, CT 06520, USA
[2] Department of Computer Science, Yale University, 51 Prospect Street, New Haven, CT 06520, USA
[3] Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT 06520, USA

Pseudogenes have long been considered to be 'dead', nonfunctional by-products of genome evolution. However, several lines of evidence now show that some pseudogenes are transcriptionally 'alive', and a few might even have biochemical roles. Therefore, the boundary between genes (often considered to be 'living') and pseudogenes (often considered to be 'dead') might be ambiguous and difficult to define. Here, we examine the evidence for and against pseudogene functionality, and we argue that the time is ripe for revising the definition of a pseudogene. Furthermore, we suggest a classification system to accommodate pseudogenes with various levels of functionality.

Pseudogenes and the common assumption of nonfunctionality

In the late 1970s, researchers were mapping the chromosomal locations of several genes (e.g. the genes encoding 5S rRNA [1] and globins [2]), and they stumbled on DNA sequences that looked similar to functional genes but contained genetic lesions such as truncations and premature stop codons. The term pseudogene was coined to describe these gene lookalikes [1]. Since then, this terminology has been used to encompass a group of genomic sequences with the following two characteristics: first, sequence similarity to a functional gene; and, second, genetic defects that preclude the generation of functional products [3–5]. The first feature can be characterized precisely by aligning the sequence of a pseudogene with that of its paralogous gene(s). The paralog is often referred to as the parental gene; however, mechanistically, it is the ancestor of the parental gene from which the pseudogene arose. For pseudogenes that have resulted from gene 'death' and do not have a functional paralog, the homologous gene(s) in a different species can be used to define sequence similarity. The second feature, however, is rather elusive but has conventionally been interpreted as mutations that disrupt the capability of a sequence to encode a protein or a structural RNA (e.g. rRNA and tRNA), because these have traditionally been considered to be the functional products of genes. This characteristic implies that pseudogenes are nonfunctional, a view that has been accepted into the neutral theory of molecular evolution [6].

It was, therefore, considered provocative by many when a study by Hirotsune *et al.* in 2003 described a functional pseudogene in mice: makorin 1 pseudogene 1 (*Mkrn1-ps1*; also known as makorin1-p1 and *Mkrn1-p1*; GenBank Accession Number AF494488) [7]. But can pseudogenes be functional? And how should the concept of 'nonfunctional' be interpreted in defining pseudogenes? How could this finding be amalgamated with the established evolutionary theory, which often uses pseudogenes as nonfunctional and neutrally evolving DNAs for estimating various parameters in evolution? The scientific community, especially those dealing with molecular evolution and gene or pseudogene annotation, began to ponder these questions.

Although a recent study [8] has shown that the function assigned to *Mkrn1-ps1* is highly questionable, several reports – including that proposing a function for *Mkrn1-ps1*, an earlier report of a functional pseudogene in snails [9], and the discovery of numerous transcriptionally 'alive' pseudogenes in humans [10,11], mice [12], plants [13] and yeast [14] – have raised concern about the conventional definition of pseudogenes. Therefore, we propose a modified definition that is based more on the sequence features of pseudogenes than on their functional status. We focus our discussion on pseudogenes that are derived from protein-coding genes, but the main points also apply to those derived from RNA-coding genes.

Nonfunctionality can be difficult to define

For the pseudogenes that were discovered in the 1970s, and for the many others that were defined later using biochemical experiments, it was clear that the sequences had defects such as a lack of promoters, the presence of frameshift or nonsense mutations, or a loss of splice sites [3–5]. These changes are usually deleterious to protein-coding genes, so the nonfunctionality is relatively transparent in that there is no (functional) product. However, those experimentally defined pseudogenes constitute only a tiny proportion of the large number of pseudogenes present in a variety of genomes [15]. For example, both the human genome [16–18] and the mouse genome [19,20] were estimated to contain ~20 000 pseudogenes. Even in

Corresponding authors: Zheng, D. (deyou.zheng@yale.edu); Gerstein, M.B. (mark.gerstein@yale.edu).
Available online 26 March 2007.

www.sciencedirect.com 0168-9525/$ – see front matter © 2007 Elsevier Ltd. All rights reserved. doi:10.1016/j.tig.2007.03.003

small and compact bacterial genomes, up to 5% of all gene-like sequences could be pseudogenes [21–23]: for example, the *Escherichia coli* genome probably contains hundreds of pseudogenes [22,24,25]. It is unlikely that the functional status of most pseudogenes will be verified using biochemical assays. And, although comparative genomic methods could help to assess the nonfunctionality of a sequence, these methods have various degrees of uncertainty [17,18,21,26]. Fundamentally, the property of being nonfunctional is difficult to establish, and the rule of there being 'no functional product' does not hold for all pseudogenes, as discussed here.

Discovery of a putative functional NOS pseudogene in snails

The ambiguity about nonfunctionality and pseudogenes first emerged in 1999, when Michael O'Shea's research group investigated the gene encoding nitric oxide synthase (NOS; GenBank Accession Number AF012531) in the snail *Lymnaea stagnalis* [9]. They found that a NOS pseudogene (GenBank Accession Number AF165914) was transcribed in certain neurons in the central nervous system of *L. stagnalis* and that the pseudo-NOS RNA could form an RNA–RNA duplex with the mRNA transcribed from a *NOS* gene *in vivo*. In this case, the transcript from the NOS pseudogene is a natural antisense RNA. The formation of this duplex was, in turn, sufficient to suppress the synthesis of NOS proteins in neurons and has a role in memory formation [9,27]. The unusual function of the NOS pseudogene seems to have evolved through a gene duplication and inversion [28]. Clearly, this result is inconsistent with the established concept of pseudogenes being nonfunctional. Nevertheless, this NOS pseudogene does have defects, including multiple premature stop codons in all three reading frames of its transcript [9]; therefore, unlike its parental gene, it cannot encode a protein.

Controversy over a putative 'functional' pseudogene in mice: *Mkrn1-ps1*

Even molecular biological experiments can fall short of revealing the functional status of a sequence. In 2003, a mouse pseudogene that seemed to have a regulatory role in development was reported [7]. While characterizing transgenic mice carrying the *Drosophila* gene *Sex lethal*, Hirotsune *et al.* obtained one line of transgenic mice that died shortly after birth as a result of multiple organ failure. Subsequent biochemical and genetic investigations found that transcription of the pseudogene *Mkrn1-ps1* was disrupted in these mice, owing to the insertion of *Sex lethal*. In wild-type mice, *Mkrn1-ps1* pseudo-RNA was suggested to be essential for the stability of one particular mRNA isoform produced from the parental gene makorin 1 (*Mkrn1*; GenBank Accession Number NM_018810). The proposed function of *Mkrn1-ps1* was further supported by the authors' finding that the abnormal development of the transgenic mice could be rescued by overexpression of either *Mkrn1-ps1* or *Mkrn1* [7]. The evidence was so compelling that a functional mammalian pseudogene seemed to have been identified.

There is, however, one problem with this work, and this has proven to be costly. In a follow-up study [8], Gray *et al.* demonstrated that *Mkrn1-ps1* was not transcribed, so the 'legend' of *Mkrn1-ps1* seems to be fading away. The new study contended that the transcript that was originally attributed to *Mkrn1-ps1* was an overlooked mRNA isoform from the *Mkrn1* locus. Furthermore, the 5′-regions of both *Mkrn1-ps1* alleles were shown to be fully methylated, so this pseudogene is unlikely to be transcribed. These contradictory results underscore the difficulty in evaluating the functional status of a pseudogene.

Pseudogene transcription

Pseudogene transcription has been observed in both small-scale gene-centered studies [3] and genome-scale unbiased mapping of transcriptionally active regions in the human genome and the mouse genome. Several surveys have provided a conservative estimate that 5–20% of human pseudogenes can be transcriptionally active [10,11,29,30]. However, in a report from the Functional Annotation of Mouse (FANTOM) project (http://fantom.gsc.riken.go.jp), 9278 of the ~100 000 full-length mouse cDNA sequences were probably from mouse pseudogenes, suggesting ~50% of mouse pseudogenes can be transcribed [12]. The ENCyclopedia Of DNA Elements (ENCODE) project (http://www.genome.gov/ENCODE) will provide more-systematic evidence for the scale of pseudogene transcription in humans.

Because many pseudogenes have high sequence similarities to functional genes and because some are almost identical to their parental genes, obvious concerns in these analyses include cross-hybridization and erroneously assigning mRNAs of parental or paralogous genes to pseudogenes. In practice, these issues are often addressed computationally by assigning an RNA (or the transcriptional evidence associated with the RNA) to a pseudogene only if the RNA shows a higher sequence identity to the pseudogene locus than to other genomic region(s) [11,30]. Such an approach is markedly conservative and does not entirely resolve ambiguous cases. Nevertheless, it is preferred in the analysis of pseudogene transcription, especially if transcription is to be coupled to a biochemical function.

Pseudogene transcription in the context of genome transcription

Traditionally, pseudogenes have been thought to be transcriptionally silent, so a strong skepticism might exist concerning the phenomenon of pseudogene transcription. Is it actually possible? Studies of transcribed regions in the human, mouse and other genomes have revealed that the transcriptome is more complex than was expected [31,32]. Not only is most of the mammalian genome transcribed but also more than half of the transcribed regions are mapped outside known genes. This picture has now emerged from numerous studies, which used several distinct high-throughput techniques, including tiling microarrays, various tag sequencing methods and full-length cDNA analysis (reviewed in Refs [31] and [32]). Given this finding, perhaps we should not be entirely surprised that pseudogenes, which seem to be as abundant as protein-coding genes in the human genome, contribute to the complex pool of the human transcriptome.

It is worth emphasizing that the evidence for pseudogene transcription in the large-scale studies mentioned earlier [10,11,29,30] was largely derived from processed RNAs. In addition, carefully designed conventional biochemical experiments have established evidence for the transcription of many pseudogenes, including the β-globin pseudogene (*HBBP1*; GenBank Accession Number X02133) [33], an interferon pseudogene (*IFNAP22*; GenBank Accession Number NG_005640) [34], a DNA topoisomerase I pseudogene [35] and a pseudogene derived from a gene encoding a heat-shock transcription factor (HSFY) [10,36].

Pseudogene transcription and function

Does pseudogene transcription provide an indication of a stochastic cellular process or an intrinsic biological function? The answer is, perhaps both. There is no doubt that (duplicated) sequences that have recently become pseudogenes can have some residual transcriptional activity, but the biological implication of this is unclear. Even if such transcribed pseudogenes influence the expression of their parental or paralogous genes – for example, owing to competition for cellular RNA degradation machinery [7] – their 'functions' might simply be a consequence of their initial temporarily high sequence identities to their parental genes and, therefore, these functions could eventually be lost after these pseudogenes accumulate more mutations. Retrotransposed sequences (also known as retroposed sequences) can also be transcribed if they land in the neighborhood of an active promoter; in fact, transcribed processed (i.e. retroposed) pseudogenes are markedly enriched near both the 5′-ends and the 3′-ends of genes [11]. Several hundred new protein-coding genes in primates might also have risen through this mechanism [37–39]. Some of these retroposed genes presumably have functions different from those of their parental genes. If evolution can produce novel protein-coding genes through retroposition, then how can we exclude the possibility that novel noncoding RNA (ncRNA) genes can arise from the same process?

More generally, given that the evolutionary 'bricolage' can take advantage of any opportunity to evolve new functions from existing sequences, and given that pseudogenes derive from functional genes, pseudogenes are certainly predisposed to be exapted (i.e. to acquire novel functions that differ markedly from the initial function of these sequences). Indeed, it has long been speculated that pseudogenes could be a source of regulatory RNA because of their sequence similarity to their parental genes [40]. The finding that the NOS pseudogene functions as an antisense RNA [9] clearly adds support to this hypothesis.

How many ways can a pseudogene be functional?

The evolutionary process undergone by the NOS pseudogene [9] is also proposed to be the mechanism for the evolution of some genes encoding microRNAs: for example, *Arabidopsis thaliana MIR161* and *MIR163* evolved by inverted duplication of their target protein-coding genes [41]. Sometimes, pseudogenes can also contribute indirectly to cellular functions by hosting novel functional elements. The human *XIST* ncRNA gene (GenBank Accession Number NR_001564), the key initiator of

Table 1. Classification of genes and pseudogenes

	Living gene	Ghost pseudogene (some intermediate functionality): similar to a functional gene, transcribed and/or functional			Dying pseudogene: dying but retains some transcriptional activity			Dead pseudogene (no indication of functionality): retains fading similarity to a functional gene, not transcribed, evolves neutrally	
		Exapted pseudogene: gain of a defined, new biological function		Piggy-back pseudogene: hosts unrelated, novel functional sequences				Non-disabled pseudogene	Disabled pseudogene
		Parent associated	Not parent associated		Chimeric transcript	Truncated transcript	Non-ORF transcript		
Key features[a]	Expressed, and exons specify a protein with a defined biological function	Regulates the function of its parental gene	Regulates the function of a gene other than its parent	As above	Can form a chimeric transcript with RNA transcribed from its neighboring gene	Generates a truncated, functionless transcript with an ORF	Retains transcriptional activity, transcript contains no ORF	Without obvious genetic defects, 'normal' ORF	With obvious genetic defects, disrupted ORF
Example	β-Globin-encoding gene [33]	A NOS pseudogene [9]	*XIST* [42]	Host gene of the snoRNA U19 [46]	An RPL23A pseudogene fused with an HLA-F-encoding gene [48]	A TCP10 pseudogene [51]	*HBBP1* [33]	An RPL21 pseudogene [18]	A GAPDH pseudogene [18]
Illustration	Figure 1a	Figure 1c	Figure 1d	Figure 1d	Figure 1e	Figure 1f	Not shown	Figure 1g	Figure 1b

[a]Abbreviation: ORF, open reading frame.

X chromosome inactivation, evolved from a pseudogene: two of its eight exons are relics of a disabled protein-coding gene [42]. Similarly, the host genes of some small nucleolar RNAs (snoRNAs) – for example, human U19 and U22 and mouse U22 – have been proposed to be derived from protein-coding genes that have lost their protein-coding potential [43–46], although in these cases the functional relatives of the putative pseudogenes have not been identified. Alternatively, pseudogenes might have functioned as DNA reservoirs for increasing antibody diversity in humans, chickens and other vertebrates, through gene conversion [3,47]. The flip side of this is that pseudogenes can swap their 'bad' DNA into their functional paralogs and, consequently, cause diseases in humans [16].

The diverse ways that a genomic sequence can realize its biological function have made it difficult to define a nonfunctional sequence, as elaborated in Table 1 and illustrated in Figure 1. It has also been found that parts of the transcripts of several pseudogenes can fuse with mRNAs from adjacent functional genes to form chimeric RNAs [48]. Do we consider these pseudogenes to be functional? Furthermore, there are pseudogenes that are

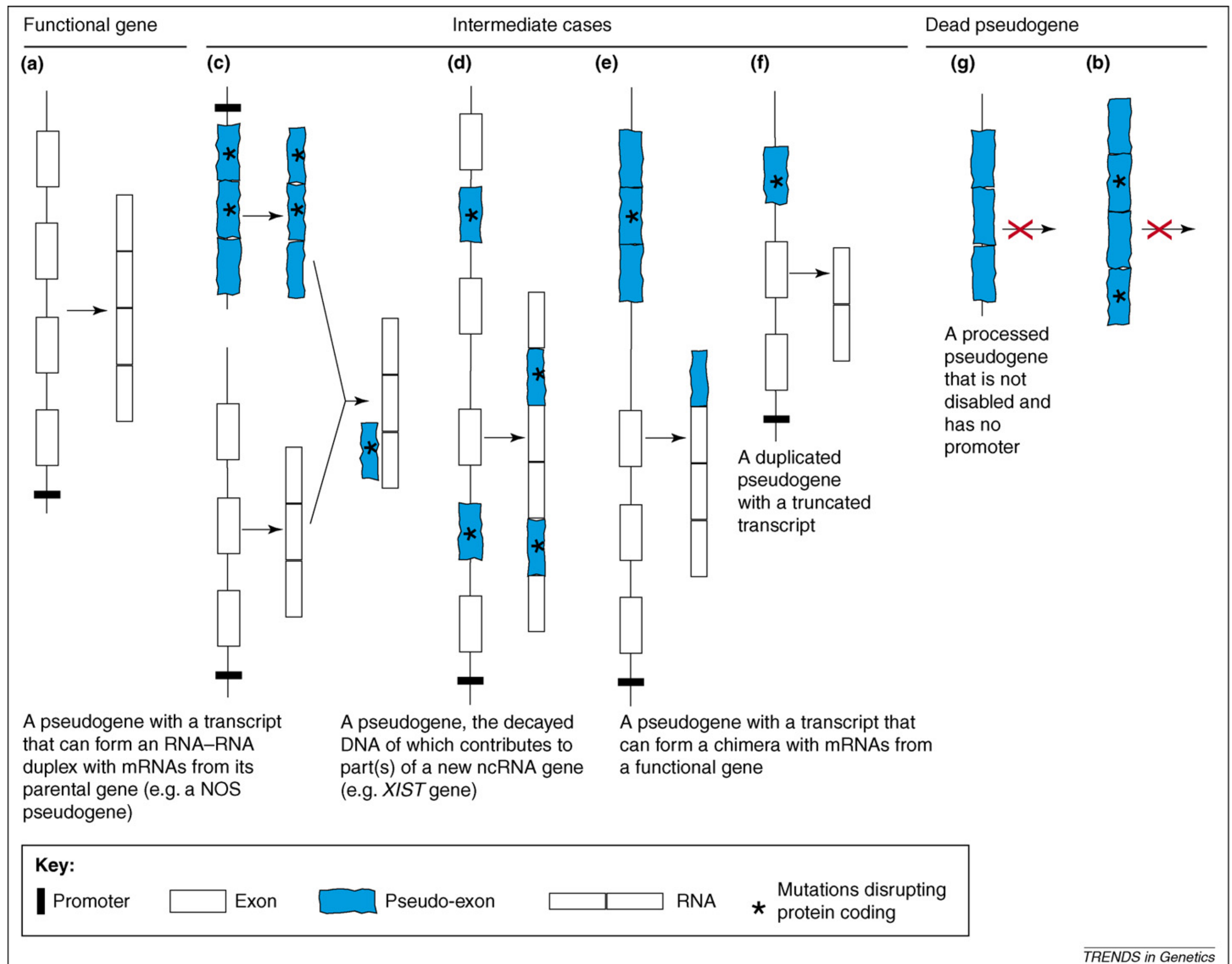

Figure 1. The distinction between genes and pseudogenes can be vague. In most cases, the separation of genes and pseudogenes is clear. A gene **(a)** can be expressed, and its exons are joined to specify a protein sequence with a defined biochemical function. By contrast, a completely dead pseudogene **(b)** does not encode a functional product owing to lack of a promoter or other sequence defects that disrupt protein coding (such as frameshift or nonsense mutations). However, such distinctions can become elusive in many scenarios. First, even though a pseudogene cannot encode a protein, its RNA transcript might have regulatory roles **(c)**: for example, a snail NOS pseudogene [9] is implicated in regulating the function of its parental gene. This occurs through the formation of an RNA–RNA duplex that is mediated by an antisense region in the pseudo-NOS RNA, which is generated from an inversion mutation after gene duplication [28]. Second, the remnant of a dead gene can be combined with new DNA to generate a new gene **(d)**: for example, the human *XIST* ncRNA gene [42]. Cases c and d are more appropriately called functional pseudogenes (or exapted pseudogenes). Third, the transcripts of certain pseudogenes, or at least some of their sequences, can be found in gene–pseudogene chimeric transcripts [48] **(e)**, although the functional implications of such transcripts have not been established. Pseudogene assignment for the examples illustrated in **(f)** and **(g)** can also be debatable. Even though a pseudogene has disrupted exon(s), a modified transcript with a new open reading frame (ORF) might be produced if the disrupted exon(s) are skipped (f). Some retroposed sequences have an intact ORF but do not have an easily detectable promoter (g), so it can be difficult to determine whether such sequences are retroposed functional genes or processed pseudogenes. Some sequences in category (f) or (g) might be young pseudogenes; the ambiguity in categorizing these sequences results from pseudogenization not necessarily being an instantaneous process (i.e. a gene copy accumulates disabling mutations randomly and progressively after it has lost its function). If we assume that the functionality of these young pseudogenes can be approximated by their protein-coding potential, then analysis of K_a/K_s (i.e. the ratio of nonsynonymous substitutions to synonymous substitutions), the pattern of insertions and deletions, or the evolutionary rate periodicity [52] that is observed in protein-coding sequences should distinguish these sequences.

under evolutionary constraint, and some of these are transcribed [10,11,29]. But are they good candidates for functional DNAs? One particular example is the pseudogene Ψ*Est-6* in *Drosophila,* which seems indispensable for the function of the *β-esterase* (*Est*) gene cluster, but the exact molecular mechanism underlying this is unclear [49].

Suggested revision of current pseudogene definition

The dilemma is what to call the sequences that seem to be derived from functional genes and have lost their protein-coding potential but might have biological functions. Considering that the definition of genes is increasingly blurring, as it is about to include heritable units of DNA sequence with various functions other than encoding proteins [32,50], we suggest that such sequences could still be called pseudogenes. As such, we define pseudogenes as genomic sequences that arise from functional genes but that cannot encode the same type of functional product (i.e. protein, tRNA or rRNA) as the original genes. Under this definition, retroposed genes would not be considered to be pseudogenes, because they encode the same type of product (protein) as their parental genes. This definition, however, does not exclude the possibility that a pseudogene can be transcribed or even functional in another context, although we consider that functional pseudogenes are rare instances of evolution. Certainly, we can restrict our definition to sequences that are nonfunctional. Such a definition is conceptually superior but, in practice, is extremely difficult to apply, because exhaustive investigation is required to define nonfunctionality. Some researchers might even argue that none of the known pseudogenes meets this criterion. We further suggest that several scenarios should be considered in defining functional or nonfunctional pseudogenes (Table 1).

Our proposed definition resolves the intrinsic irony of functional pseudogenes. To make the terminology clearer, it might be useful to refer to functional pseudogenes as exapted pseudogenes: this would indicate that they have been recruited to have a new function, and it would distinguish them from completely dead pseudogenes (i.e. functionless pseudogenes) and recently generated pseudogenes (i.e. those on their way to becoming completely dead pseudogenes). We also would like to point out that pseudogenes have previously been proposed to be 'potogenes' because of their potential to acquire new function and thus become genes [3,49].

Concluding remarks

It seems that, so far, there are only a few cases of functional pseudogenes in eukaryotes. Considering the diversity of how these pseudogenes carry out their function and the prevalence of pseudogenes in mammalian genomes, we anticipate that many more functional pseudogenes (or exapted pseudogenes) remain to be discovered. Eventually, we would like to see more studies directly targeted at pseudogene loci to prove convincingly the existence of a functional pseudogene. Before then, perhaps we should prepare ourselves by updating the definition of pseudogenes.

Acknowledgements

We thank the referees for their valuable suggestions for improving this article. This work was supported by the National Institutes of Health (grant numbers P50 HG02357, 1U01HG003156).

References

1 Jacq, C. *et al.* (1977) A pseudogene structure in 5S DNA of *Xenopus laevis*. *Cell* 12, 109–120
2 Little, P.F. (1982) Globin pseudogenes. *Cell* 28, 683–684
3 Balakirev, E.S. and Ayala, F.J. (2003) Pseudogenes: are they 'junk' or functional DNA? *Annu. Rev. Genet.* 37, 123–151
4 Vanin, E.F. (1985) Processed pseudogenes: characteristics and evolution. *Annu. Rev. Genet.* 19, 253–272
5 Mighell, A.J. *et al.* (2000) Vertebrate pseudogenes. *FEBS Lett.* 468, 109–114
6 Kimura, M. (1983) *The Neutral Theory of Molecular Evolution,* Cambridge University Press
7 Hirotsune, S. *et al.* (2003) An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene. *Nature* 423, 91–96
8 Gray, T.A. *et al.* (2006) The putatively functional *Mkrn1-p1* pseudogene is neither expressed nor imprinted, nor does it regulate its source gene *in trans*. *Proc. Natl. Acad. Sci. U. S. A.* 103, 12039–12044
9 Korneev, S.A. *et al.* (1999) Neuronal expression of neural nitric oxide synthase (nNOS) protein is suppressed by an antisense RNA transcribed from an NOS pseudogene. *J. Neurosci.* 19, 7711–7720
10 Zheng, D. *et al.* (2005) Integrated pseudogene annotation for human chromosome 22: evidence for transcription. *J. Mol. Biol.* 349, 27–45
11 Harrison, P.M. *et al.* (2005) Transcribed processed pseudogenes in the human genome: an intermediate form of expressed retrosequence lacking protein-coding ability. *Nucleic Acids Res.* 33, 2374–2383
12 Frith, M.C. *et al.* (2006) Pseudo-messenger RNA: phantoms of the transcriptome. *PLoS Genet.*,2, e23 DOI: 10.1371/journal.pgen.0020023 (genetics.plosjournals.org)
13 Yamada, K. *et al.* (2003) Empirical analysis of transcriptional activity in the *Arabidopsis* genome. *Science* 302, 842–846
14 Harrison, P. *et al.* (2002) A small reservoir of disabled ORFs in the yeast genome and its implications for the dynamics of proteome evolution. *J. Mol. Biol.* 316, 409–419
15 Karro, J.E. *et al.* (2007) Pseudogene.org: a comprehensive database and comparison platform for pseudogene annotation. *Nucleic Acids Res.* 35, D55–D60
16 Bischof, J.M. *et al.* (2006) Genome-wide identification of pseudogenes capable of disease-causing gene conversion. *Hum. Mutat.* 27, 545–552
17 Torrents, D. *et al.* (2003) A genome-wide survey of human pseudogenes. *Genome Res.* 13, 2559–2567
18 Zhang, Z. *et al.* (2003) Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome. *Genome Res.* 13, 2541–2558
19 Zhang, Z. *et al.* (2004) Comparative analysis of processed pseudogenes in the mouse and human genomes. *Trends Genet.* 20, 62–67
20 Khelifi, A. *et al.* (2005) HOPPSIGEN: a database of human and mouse processed pseudogenes. *Nucleic Acids Res.* 33, D59–D66
21 Lerat, E. and Ochman, H. (2004) Psi-Phi: exploring the outer limits of bacterial pseudogenes. *Genome Res.* 14, 2273–2278
22 Lerat, E. and Ochman, H. (2005) Recognizing the pseudogenes in bacterial genomes. *Nucleic Acids Res.* 33, 3125–3132
23 Liu, Y. *et al.* (2004) Comprehensive analysis of pseudogenes in prokaryotes: widespread gene decay and failure of putative horizontally transferred genes. *Genome Biol.*,5, R64 DOI: 10.1186/gb-2004-5-9-r64 (genomebiology.com)
24 Ochman, H. and Davalos, L.M. (2006) The nature and dynamics of bacterial genomes. *Science* 311, 1730–1733
25 Harrison, P.M. and Gerstein, M. (2002) Studying genomes through the aeons: protein families, pseudogenes and proteome evolution. *J. Mol. Biol.* 318, 1155–1174
26 Harrison, P.M. *et al.* (2002) Molecular fossils in the human genome: identification and analysis of the pseudogenes in chromosomes 21 and 22. *Genome Res.* 12, 272–280
27 Korneev, S.A. *et al.* (2005) Timed and targeted differential regulation of nitric oxide synthase (NOS) and anti-NOS genes by reward conditioning leading to long-term memory formation. *J. Neurosci.* 25, 1188–1192

28 Korneev, S. and O'Shea, M. (2002) Evolution of nitric oxide synthase regulatory genes by DNA inversion. *Mol. Biol. Evol.* 19, 1228–1233
29 Svensson, O. *et al.* (2006) Genome-wide survey for biologically functional pseudogenes. *PLoS Comput. Biol.*,2, e46 DOI: 10.1371/journal.pcbi.0020046 (compbiol.plosjournals.org)
30 Yano, Y. *et al.* (2004) A new role for expressed pseudogenes as ncRNA: regulation of mRNA stability of its homologous coding gene. *J. Mol. Med.* 82, 414–422
31 Willingham, A.T. and Gingeras, T.R. (2006) TUF love for 'junk' DNA. *Cell* 125, 1215–1220
32 Carninci, P. (2006) Tagging mammalian transcription complexity. *Trends Genet.* 22, 501–510
33 Chang, L.Y. and Slighton, J.L. (1984) Isolation and nucleotide sequence analysis of the β-type globin pseudogene from human, gorilla and chimpanzee. *J. Mol. Biol.* 180, 767–784
34 Goeddel, D.V. *et al.* (1981) The structure of eight distinct cloned human leukocyte interferon cDNAs. *Nature* 290, 20–26
35 Zhou, B.S. *et al.* (1992) Identification of antisense RNA transcripts from a human DNA topoisomerase I pseudogene. *Cancer Res.* 52, 4280–4285
36 Footz, T.K. *et al.* (2001) Analysis of the cat eye syndrome critical region in humans and the region of conserved synteny in mice: a search for candidate genes at or near the human chromosome 22 pericentromere. *Genome Res.* 11, 1053–1070
37 Long, M. *et al.* (2003) The origin of new genes: glimpses from the young and old. *Nat. Rev. Genet.* 4, 865–875
38 Brosius, J. (1991) Retroposons – seeds of evolution. *Science* 251, 753
39 Marques, A.C. *et al.* (2005) Emergence of young human genes after a burst of retroposition in primates. *PLoS Biol.*,3, e357 DOI: 10.1371/journal.pbio.0030357 (biology.plosjournals.org)
40 McCarrey, J.R. and Riggs, A.D. (1986) Determinator–inhibitor pairs as a mechanism for threshold setting in development: a possible function for pseudogenes. *Proc. Natl. Acad. Sci. U. S. A.* 83, 679–683
41 Allen, E. *et al.* (2004) Evolution of microRNA genes by inverted duplication of target gene sequences in *Arabidopsis thaliana*. *Nat. Genet.* 36, 1282–1290
42 Duret, L. *et al.* (2006) The *Xist* RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. *Science* 312, 1653–1655
43 Tycowski, K.T. *et al.* (1996) A mammalian gene with introns instead of exons generating stable RNA products. *Nature* 379, 464–466
44 Tycowski, K.T. *et al.* (1994) Requirement for intron-encoded U22 small nucleolar RNA in 18S ribosomal RNA maturation. *Science* 266, 1558–1561
45 Moore, M.J. (1996) Gene expression. When the junk isn't junk. *Nature* 379, 402–403
46 Bortolin, M.L. and Kiss, T. (1998) Human U19 intron-encoded snoRNA is processed from a long primary transcript that possesses little potential for protein coding. *RNA* 4, 445–454
47 Ota, T. and Nei, M. (1995) Evolution of immunoglobulin V_H pseudogenes in chickens. *Mol. Biol. Evol.* 12, 94–102
48 Vinckenbosch, N. *et al.* (2006) Evolutionary fate of retroposed gene copies in the human genome. *Proc. Natl. Acad. Sci. U. S. A.* 103, 3220–3225
49 Balakirev, E.S. and Ayala, F.J. (2004) The β-esterase gene cluster of *Drosophila melanogaster*: is ψ*Est-6* a pseudogene, a functional gene, or both? *Genetica* 121, 165–179
50 Pearson, H. (2006) Genetics: what is a gene? *Nature* 441, 398–401
51 Zheng, D. and Gerstein, M.B. (2006) A computational approach for identifying pseudogenes in the ENCODE regions. *Genome Biol.*,7 (Suppl. 1), S13 DOI: 10.1186/gb-2006-7-s1-s13 (genomebiology.com)
52 David, L. *et al.* (2006) A high-resolution map of transcription in the yeast genome. *Proc. Natl. Acad. Sci. U. S. A.* 103, 5320–5325

Endeavour

The quarterly magazine for the history and philosophy of science.

You can access *Endeavour* online on ScienceDirect, where you'll find book reviews, editorial comment and a collection of beautifully illustrated articles on the history of science.

Featuring:

Information revolution: William Chambers, the publishing pioneer by A. Fyfe
Does history count? by K. Anderson
Waking up to shell shock: psychiatry in the US military during World War II by H. Pols
Deserts on the sea floor: Edward Forbes and his azoic hypothesis for a lifeless deep ocean by T.R. Anderson and T. Rice
'Higher, always higher': technology, the military and aviation medicine during the age of the two world wars by C. Kehrt
Bully for *Apatosaurus* by P. Brinkman

Coming soon:

Environmentalism out of the Industrial Revolution by C. Macleod
Pandemic in print: the spread of influenza in the Fin de Siècle by J. Mussell
Earthquake theories in the early modern period by F. Willmoth
Science in fiction - attempts to make a science out of literary criticism by J. Adams
The birth of botanical *Drosophila* by S. Leonelli

And much, much more...

***Endeavour* is available on ScienceDirect, www.sciencedirect.com**