UNCONVENTIONAL GENE BEHAVIOR AND ITS RELATIONSHIP TO PSEUDOGENES

JOHN WOODMORAPPE, MA, BA
6505 N. NASHVILLE #301
CHICAGO, IL 60631, USA

Proving that a gene unit is totally nonfunctional, and is therefore definitely a pseudogene, is impossible.

Mounsey et. al. [49, p. 772]

KEYWORDS :
Junk DNA, Teleology, Dysteleological Arguments, suboptimal design, recoding, readthrough

ABSTRACT
Traditionally, pseudogenes have been regarded as “dead” gene copies as a result of features such as the absence of promoters and the existence of premature stop codons. However, the recognition of a truly disabled gene is not as straightforward as once believed. It is now known that promoters may be cryptic. Genomic recoding processes can allow for the synthesis of a peptide despite the present of premature stop codons. Alternative splicing can allow for the omission of exons that contain premature stop codons. Finally, negative evidence for pseudogene expression, for the relatively few pseudogenes for which it is available, must be interpreted with caution. This is in view of the fact that many genes express themselves only under very restricted conditions.

INTRODUCTION

Each component of the eukaryotic gene plays a role in its expression. The DNA sequence of the gene that is transcribed into mRNA and then translated into a protein (in the case of protein-coding genes) is called the open reading frame. The promoter indicates the approximate location where transcription is to begin while the polyadenylation signal terminates the transcription. Some genes have both exons and introns, and these are marked off by splice sites. The primary RNA transcript includes both exons and introns, but the latter are spliced out in the mature mRNA. A translation initiation codon signals the location where the mature mRNA is to be translated into a peptide, while the stop codon terminates the translation process.

In many genes, features such as the promoter and initiation codons can be recognized solely by inspection of the gene’s DNA sequence, and gene-search computer algorithms have been designed and used to exploit this fact. However, it is commonly acknowledged that an unknown fraction of genes is missed by such algorithms as a result of unconventional genes whose sequences deviate in some important-but presently unknown-way from canonical genes. Such genes can only be discovered through traditional genetics or biochemistry.

Three decades ago, Susumu Ohno, a prominent Japanese geneticist, suggested that noncoding DNA lacks function, and originates from the remnants of dysfunctional genes. In fact, by analogy to the fossils found in the Earth’s crust, he called them fossilized genes [51, 37]. In time, “dead” gene copies came to be called pseudogenes. They clearly resemble specific genes but seem to be disabled as a result of inferred mutational lesions:

> Almost any imaginable defect can be found in pseudogenes, including abnormal initiation and termination codons, altered normally invariant codons, deletons, insertions, and nonsense mutations, as well as defects in promoter, RNA splicing, and polyadenylation sequences. [33, p. 253]

Pseudogenes are often more dissimilar from their certainly functional paralogs than the latter are from each other. This sequence diversity is interpreted as indicative of the fact that purifying selection is not acting to eliminate mutations from these putatively untranscribed gene-like sequences. Finally, pseudogenes at first were not observed to be transcribed nor translated in standard cell biology assays. All of this led to the belief that they are simply disabled copies of genes.

This thinking was perhaps reasonable some twenty years ago. However, advances in genomic knowledge now complicate the identification of pseudogenes. Features such as genomic recoding, alternative splicing, and cryptic promoters provide evidence that some apparent lesions, as occur in pseudogenes, can be over-ridden. Moreover, this over-riding can be regulatory as well as reparative in nature.

Pseudogenes have been exploited by opponents of special creation. The existence of seemingly disabled gene copies, notably in human genomes, have been taken as evidence against the design of genomes by a Designer. In addition, some of the apparent disablements found in pseudogenes have been found to be shared by humans and apes, prompting the argument that only shared common descent could account for these “shared mistakes” [43]. There is no shortage of anti-creationists who have repeated this assertion. However, the “shared mistake” argument is a dubious one. It ignores or belittles the fact that pseudogenes can acquire identical “lesions” independently [70]. If it turns out that pseudogenes are not disabled genes after all, the “shared mistake” argument would become all but moot.

Owing to the breadth of this topic, a series of papers is planned which address pseudogene-related phenomena. There are two main types of pseudogenes: classical and processed (the latter also called retropseudogenes). Interpersed repeats, considered by some to be retropseudogenes, are not discussed in this work. This paper emphasizes classical pseudogenes, although most of the discussion is also applicable to retropseudogenes. Consideration of pseudogene features themselves is limited to absent or disabled promoters, and to premature stop codons. Throughout this work, the emphasis is on eukaryotic genomes, and on the similarities which they share. In fact, in this regard, Resnick and Cox [55], in all seriousness, call yeast an honorary mammal.

DIVERSITY OF PROMOTERS PREVENTS READY IDENTIFICATION

Early studies of genes had indicated that certain conserved sequences are responsible for transcription. Because such regulatory sequences can be experimentally inactivated by artificial mutation, it is commonly assumed that inferred natural mutations of regulatory sequences will have an identical effect, transcriptionally silencing an otherwise active gene. However, even an obvious alteration of normally conserved gene transcription elements does not necessarily imply their inactivation. The γ-globin tarsier pseudogene was once “pronounced dead” solely because the distal CCAAT box was missing, and the second one had an ostensibly inactivating mutation (CcgAT). Now the γ-globin pseudogene is believed to be functional based on the absence of nonsynonymous substitutions in the coding regions [44], although the role of this sequence needs to be clarified.

The foregoing discussion is especially applicable to promoters, which are at the very core of the transcription process. A functional promoter is required for transcription of a gene to begin. Inferences are commonly made about the putative mutational inactivation of promoters in classical pseudogenes and the absence of promoters in retropseudogenes. However, determining whether or not some kind of functional promoter exists is no longer straightforward.

The most commonly occurring promoter complex includes the TATA box, which is usually situated about 30 nucleotides upstream of the translation start site (though it will still function as close as 15 nucleotides to the same: [52]). The TATA motif itself is an idealization, as it actually subsumes a large diversity of AT-rich short sequences [5]. Moreover, the TATA box can turn up in unexpected locations, even within the gene’s coding region [50].

We now realize that a variety of non-TATA sequences can function as promoters from the same 30 nucleotide distance [69]. Furthermore, in compositional contrast to the AT-rich TATA-type promoters, GC-rich promoters also exist for a significant number of genes. They can be found embedded within

CpG islands, occurring at variable distances from genes and sometimes even serving two genes simultaneously [39]. It is significant that an increasing number of short nucleotide sequences are being discovered that exhibit previously unsuspected promoter activity [59, 64]. These aptly named cryptic promoters are more prevalent than previously supposed [75, p. 7373]. And, contrary to earlier beliefs, neither the TBP nor the TFIID transcription factors need be present to bind to whatever promoter sites there are in existence, as other $TAF_{II}S$ can perform this role [68]. This further complicates our understanding of promoter sequences.

Functional promoters can be difficult to identify because some of them regulate genes from a considerable distance [63]. Moreover, some promoter sequences display little activity unless they work together with enhancers. The latter, in turn, are especially difficult to identify because they are quite variable in sequence and may be located at considerable distances from the promoter [28]. In fact, enhancers and other gene regulatory elements that are situated hundreds of kilobases from the promoter may be unexpectedly common [9].

It is a fairly common practice to compare pseudogenes with their gene paralogs, and then suspect that the pseudogene is transcriptionally inert owing to the fact that its promoter region differs significantly from that of its gene paralog, and the promoter is therefore presumably nonexistent or disabled. However, it is certainly possible that an apparently retrotranscribed gene could be under the control of an uncharacterized promoter. Such is the case for the human and rodent *Supt4h* and *Supt4h2* genes, prompting this comment:

> This illustrates another potential complexity of the mammalian genome, i. e., the use of a processed gene under the control of a different promoter region than the unspliced gene. [7, p. 4960]

It is commonly supposed that, keeping in mind the usual loss of the promoter sequence during reverse transcription, it is very unlikely that a retroposited gene sequence would fortuitously land near a suitable pre-existing promoter. However, it is sobering to realize that, at least in yeast, 1% of randomly chosen 16bp DNA sequences can function as promoters at an equal or better rate than the canonical TATA box [60]! And this does not consider weaker potential promoter sequences, at least some of which could serve as promoters if they happen to be paired with suitable enhancers, which, as noted earlier, may influence the promoter from a considerable distance.

Judgments about absent or damaged promoters must be re-examined in the light of the fact that unexpectedly many genes have more than one promoter [1]. As an example, the experimental removal of the entire 250 bp proximal promoter region (including the TATA box) of the human K18 gene fails to silence the gene. It only causes the activation of a second ("Lazarus") promoter [57]. The implication of plural promoters for the avoidance of gene inactivation is recognized:

> Alternatively, initiation of transcription at several sites, generating transcripts differing only in their 5'-untranslated region, might render expression of the gene less vulnerable to mutations in promoter sites. [1, p. 456]

The N-*myc2* retrosequence in the woodchuck [16] constitutes a retropseudogene that is transcribed thanks to a secondary promoter that had been unmasked by the reverse transcription process that produced this retroposon. The transcriptional activity of the secondary promoter in N-*myc2* is verified experimentally. N-*myc2* is transcribed in pathological tissue (liver tumors) as well as healthy tissue (the brain). Owing to the fact that the N-*myc2* sequence itself appears to be conserved, N-*myc2* is suspected of playing a biological role.

The *Adh* retrosequences in *Drosophila* provide an instructive example of the fact that the absence of a functional promoter in a pseudogene is all but impossible to prove by examining the sequence:

> We have gathered conflicting evidence about the putative functionality of these retrosequences…Inspection of the 5' region flanking the *Adh* retrosequences of the *obscura* species did not show any remnant similar to the original promoter sequences or structural homologies with an already-identified coding sequence. Nevertheless, we cannot discard the idea that these retrosequences might have captured a new promoter and might be transcribed. [41, p. 1323]

Since very few "promoter damaged" or "promoterless" pseudogenes have been experimentally investigated for cryptic promotjjjer or secondary promoter activity, we cannot assume that they are transcriptionally inert. Finally, it must be recognized that a significant number of pseudogenes have been found to be transcriptionally active after all. However, "The functional relevance of pseudogene transcripts remains unclear" [46, p. 112].

INFERRING TRANSLATION-PREVENTING MUTATIONS

If a pseudogene can be transcribed, the next question concerns its capability for translation. Various transcribed pseudogenes seem untranslatable because an initiator codon (usually AUG) is putatively absent. However, some investigators [34] recognize the fact that the pseudogenes under study may in fact be able to undergo translation initiation, and therefore potentially code a functional product, because some alternative, downstream AUG codon can serve as the initiator. Moreover, analyses of alternative splicing, a genomic process now known to be common [40], demonstrate that alternative AUG start codons occur at a previously unsuspected high frequency in genes [47].

One form of splicing takes place in the bacterium *Anaplasma phagocytophila*. The *p44-18* pseudogene lacks an AUG start codon and is out of frame with its upstream overlapping paralogous gene, *p44-1*. Thanks, however, to a splicing mechanism that excises the intervening sequence between *p44-1* and *p44-18*, the apparent impediments to pseudogene expression are overcome [76].

Assuming that translation initiation is possible, it is not straightforward to determine which alterations in the ORF would necessarily prevent the completion of the translation process :

> …it is likely that a substantial fraction of degenerative mutations (perhaps the majority) do not lead to complete loss-of-function (subfunctionalization), …[42, p. 470]

As an example of this, Fletcer et. al. [15] are unsure whether they are dealing with a gene or a pseudogene owing to the fact that, in spite of containing many inferred missense mutations, it may still be capable of being translated into a variant but functional peptide. Similarly, Mezzina et . al. [45, p. 111] raise the following caution about a transcribed rat pseudogene:

> Nevertheless, we cannot rigorously exclude that the *ΨmtTFA* might have evolved to fulfill some alternative function, in which case, the sequence drift could be ascribed to selection for this alternative function rather to a lack of functionality.

This fact can be generalized. In his comprehensive survey of functional retrosequences (including retropseudogenes), Brosius [4, p. 223] alludes to the ambiguity of mutational inactivation:

> There are probably *numerous* additional retrogenes whose ORFs are not severely compromised or could yield a truncated polypeptide, partially in a different reading frame [Refs.]. However, transcription and/or translation are not documented. (emphasis added).

Even according to evolutionary long-age assumptions, a (pseudo)gene can be undergoing an inferred rate of nonsynonymous to synonymous substitutions at a ratio approaching 1.0 without necessarily being in a “mutational freefall”. For example, some erstwhile *Drosophila* pseudogenes were subsequently promoted to at least potentially active gene status. The relatively large inferred rates of nonsynonymous substitutions were reinterpreted as being the outcome of selection pressure instead of the absence of functional constraints [53].

Many proteins contain certain domains or motifs deemed indispensable for the performance of their biological roles. However, we now know that these may differ between paralogous genes, or even between different splice variants from the same gene. Consider, for instance, the Bcl-rambo gene, involved in proapoptotic activities, and which contains the BH1, BH2 and BH3 motifs. The Bc1-rambo beta splice variant of the preceding, by contrast, lacks the foregoing domains and translates into a BH4-only protein. An exonized Alu insertion, containing a premature stop codon, is responsible for the synthesis of the domain-deprived shortened peptide [73]. Clearly, no pseudogene should be reckoned “dead” just because its deduced peptide sequence lacks one or more functional domains found in its gene paralogs. This fact can be extended to instances where a missing-domain deduced peptide sequence fails to closely correspond to currently known functional peptides:

> If the transcript from *11Ψh-mtTFA* was to be translated, it would produce a protein of 89 amino acids having 67.44% similarity with human mtTFA protein from amino acids 52-142, thus lacking the second HGM box and subsequently the ability to bind DNA. A Blast search against the SwissProt database has not revealed any other significant similarity, suggesting that this putative protein, if present, could have a different role from h-mtTFA [56. p. 231]

TRANSLATION IN THE CASE OF PARTIAL OPEN READING FRAMES

In genes exhibiting conventional behavior, the stop codon’s sole function is to give a signal for translation of a peptide product to terminate. A nucleotide substitution may convert an amino acid coding codon into a premature stop codon, or else an insertion or deletion may shift the open reading frame, causing a premature stop codon to come into phase with the remaining ORF. This is commonly believed

to indicate that any peptide, if synthesized at all, would be inert. Premature stop codons are characteristic of pseudogenes [11, 19], so much so that they are employed to help identify them in the genome [26].

There are several reasons, however, for no longer accepting the premature stop codon as an ipso facto indicator of a pseudogene's "dead" status. Numerous instances are now known where a truncated peptide, far from being inert, exhibits much the same biological activity as its full-length isoform [42, 35, 36, 62, 66]. A premature stop codon may simply create two shorter ORFs, at least one of which may be translated into a viable peptide, as is the case with a shortened *Drosophila* kelch protein that appears sufficient for the insect's development [58]. In their survey of the nematode worm's (*C. elegans*) genome, Harrison et al. [23] point out that some of their so-designated pseudogenes may actually be functional genes that make shortened peptides. Other investigators of pseudogenes [54] have voiced similar concerns.

PREMATURE STOP CODONS SEEN IN A NEW LIGHT

Since, as noted earlier, the stop codon acts to terminate peptide synthesis, it seems self evident that a premature stop codon necessarily indicates abnormal truncated translation. In striking opposition to this, various cellular processes, collectively known as recoding, are now known (see Baranov et al. [2] for review). Because recoding processes are known from taxonomically diverse groups, such as bacteria and mammals, such processes may occur in many, if not most, groups. Without recoding, most of the genes in question would not function properly, if at all, and would, by conventional definition, be pseudogenes.

One notable recoding process is codon redefinition, which allows a certain codon to possess an alternative mode of expression. Pointedly, codon redefinition allows for the premature stop codon to be read-through so that the (pseud)ogene can complete the translation of a full-length peptide. Some forms of readthrough are able to "over-rule" all three possible premature stop codons [62]. Those ORFs in yeast genes that are subject to readthrough have explicitly been referred to as containing pseudogenic features by Harrison et al. [25]. In certain yeast genes, readthrough of premature stop codons is said to be fairly common [35]. Certain prions, responding to environmental stress, can activate a seemingly-disabled yeast (pseudo)gene by readthough of its premature stop codon:

> Clearly, the lethal consequences of a nonsense mutation in an essential gene would be overcome in an allosuppressor background. *[PSI+]* is an omnipotent allosuppressor determinant, in that it enhances nonsense suppression of all three termination codons… [10, p. 1978]

The foregoing example is recognized by Harrison and Gerstein [24] as a process for "resurrecting" disabled genes. Other investigators [31, 32] are unsure whether they are dealing with a gene or a pseudogene because of the recently appreciated ambiguity of the effects of premature stop codons.

In their investigation of pseudogenes in *E. coli*, Homma et al. [30] acknowledge the possibility of recoding processes, but suggest that their (apparent) rarity makes them "…unlikely to be operative in many of the pseudogene candidates:" However, sweeping generalizations about the rarity of recoding processes may be unwarranted. Hammell et. al. [22] have found that 2.54% of yeast genes have the distinctive heptamer sequence that is known to serve as a –1 ribosomal frameshift signal. Furthermore, the apparent overall infrequency of recoding processes owes at least partly to research biases favoring highly expressed, well-studied, canonically behaved genes [18]. Finally, we do not know the extent of recoding processes in sporadically expressed genes, let alone supposed pseudogenes.

Some forms of translational readthrough are rather surprising. The premature stop codon is now known to function as a "wild card" codon for specifying noncanonical amino acids, namely selenocysteine [29] and possibly L-pyrrolysine [61]. The amino acid "alphabet" turns out to be larger than conventionally believed. No one knows how many other rarely occurring amino acids are, or were, encoded by otherwise pseudogenic premature stop codons.

Finally, even if all of the above-discussed factors are not applicable to a pseudogene with a premature stop codon, this still does not guarantee its "dead" status, thanks to alternative splicing. Morgan et. al. [48] have recently pointed out that this process has not been explored in detail even in certain highly-expressed genes. Yousef et al. [74] caution that there may be functional splice variants of the human siglec pseudogenes that are expressed only in certain tissues, certain developmental stages, or pathological situations, or which can encode a truncated, yet functional peptide. This fact can now be generalized. Thus, Harrison et al. [27] point out that it is now actually difficult to distinguish a functional

gene from a nonprocessed pseudogene by inspection of the DNA sequence, all because an exon containing a perceived disablement may be spliced out and the remaining transcript then translated into a functional product.

ASSESSING THE EXPRESSION OF GENES AND PSEUDOGENES

Up to this point, discussion has been limited to the inference of potential pseudogene expression from DNA sequence alone. It is clear that there are numerous difficulties in determining whether or not a gene-like structure is biologically active. Consider, first of all, human genes in general:

> Even with genome sequences in hand, our ability to identify genes is largely limited to relatively large, evolutionarily conserved, moderately to highly expressed protein coding genes. We know there are exceptions that fly below our radar—tiny genes, rapidly evolving genes, genes expressed in only a few cells at special times—but we had hoped, with some justification, that they aren't too important or numerous [12, p. 137]

To the contrary, such rarely expressed genes may be numerous [14], and expressed only at low levels [72, p. 811]. Large-scale surveys of mRNA sequences (EST libraries) have their own biases [40]. For instance, they typically contain very few transcripts that have a low copy number or short half-life. Pointedly, comparable characteristics may hold for most products of pseudogene expression. This would especially include those situations where only a small part of the pseudogene's sequence ever becomes translated. Gene-search algorithms are particularly weak in this regard:

> There exist two basic problems in gene recognition: recognition of protein coding regions and recognition of the functional sites of genes. They are not yet satisfactorily solved, especially in recognizing shorter coding regions of human genes…When sequences are shortened, the difference of statistical characteristics between coding and noncoding sequences tend to be small. [67, p. 207, 211]

As for pseudogene sequences themselves, there is no foolproof correlation between the number of "disablements" in a pseudogene, and whether or not, at minimum, it turns out to be transcribed. Harrison et al. [25] have found expressed ORFs in yeast genes in spite of these having at least five "lesions." The D2 pseudogene is transcribed in the ovary although it has seven lesions, excluding missense mutations [8].

Having seen the inadequacy of inferential judgments of pseudogene viability, let us now consider actual in vitro and in vivo experiments regarding the same. It turns out that many pseudogenes, inert according to initial experiments, turn out to be transcribed after all [3]. Some are eventually found expressed only in a few tissues [31, 32], including particular tissues in which other investigations had failed to detect any expression [8, 13]. Various researchers [21, 46] caution that silence of a tested pseudogene may only indicate that the minimal conditions for expression have not been met during the test. In like manner, negative evidence of pseudogene translation can hardly be called conclusive. Note that any synthesized peptide may degrade too rapidly for detection [20] and/or be translated in cell lines, tissues, or developmental stages missed by experimental analyses [17].

It is an intriguing fact that, at least in the nematode *C. elegans*, genes involved in early development have significantly more paralogous pseudogenes than those expressed in the later stages of the worm's life [6]. Might this imply at least a one-time expression of pseudogenes in the early development of an organism? Pointedly, Fourel et al. [16] caution that their failure to observe the N-*myc2* (pseudo)gene expressing itself in fetal tissue does not exclude its potential expression in stem cell tissue. Owing to the fact that only a vanishing number of pseudogenes have been tested for expression in stem cell tissue, this takes on further significance.

CONCLUSIONS

Genes are much more complex than believed until recently, and we know much less about them than we thought we did. Many genes do not behave in a straightforward, canonical manner. It is obvious that the conventional gene/pseudogene demarcation is becoming more and more difficult to draw. More and more (pseudo)genes with apparent lesions turn out to be transcribed and even translated. Sweeping assertions about the nonfunctionality of pseudogenes [43] appear to be more and more dubious.

When pseudogenes are studied as a whole, it is almost always from the assumption of biological inactivity and neutral molecular evolution [23]. Personal discussions with genomic specialists have revealed the sobering fact that no one has examined pseudogenes for expression in a comprehensive and systematic manner. All research is paradigm driven, and there is a disinclination to study objects that are deemed relatively unimportant within the scope of the ruling paradigm. Add to this the fact that the biotech industry prefers to allocate funding to the study of interesting genes in preference to the

decoding of junk DNA [37, p. 1125]. It is time for pseudogenes to be systematically tested for biological expression.

No attempt has been made in this paper to examine all pseudogene-related phenomena, nor to present a scientific creationist hypothesis for their origin. However, the fact that some genes have apparent lesions that actually serve regulatory purposes (as in recoding) allows one to consider at least some pseudogenes as a type of highly-regulated noncanonical gene. These could have been specially created with the pseudo-lesions serving as regulators of transcription and/or translation. Further creationist research on pseudogenes should not only include systematic investigation of their expression, in *all* tissues, but also of the relationship of pseudogenes to recoding processes. Studies should also be undertaken on the potential function of pseudogenic transcripts and truncated proteins, both of which have already been demonstrated to be functional in the case for two snail pseudogenes [71].

REFERENCES

Standard PubMed Journal Abbreviations Are in Use.

[1] Ayoubi, T. A. Y, and W. J. M. van de Ven. **Regulation of gene expression by alternative promoters**. FASEB J 10, (1996) 453-460.

[2] Baranov, P. V., et al. **Recoding: translational bifurcations in gene expression**. Gene 286, (2002) 187-201.

[3] Betran, E., et al**. Evolution of the Phosphoglycerate mutase processed gene in human and chimpanzee revealing the origin of a new primate gene.** Mol Biol Evol 19(2002)654-663.

[4] Brosius, J. **Genomes were forged by massive bombardments with retroelements and retrosequences**. Genetica 107, (1999) 209-238.

[5] Burke, T. W. et al. **The DPE, a conserved downstream core promoter element that is functionally analogous to the TATA box**. Cold Spring Harb Symp Quant Biol 63, (1998) 75-82.

[6] Castillo-Davis, C. I., and D. L. Hartl. **Genome evolution and developmental constraint in *Caenorhabdities elegans***. Mol Biol Evol 19, (2002) 728-735.

[7] Chiang, P.-W. et al. **Comparison of murine *Supt4h* and a nearly identical expressed, processed gene**. Nucleic Acids Res 26, (1998) 4960-4964.

[8] Choi, D. et al. **The expression of pseudogene cyclin D2 mRNA**. J Assist Reprod Genet 18, (2001) 110-113.

[9] DiLeone, R. J., et al. **An extensive 3' regulatory region controls expression of *Bmp5* in specific anatomical structures of the mouse embryo**. Genetics 148, (1998) 401-408.

[10] Eaglestone, S. S. **Translation termination efficiency can be regulated by *Saccharomyces cerevisiae***. Embo J 19, (1999) 1974-1981.

[11] Echols, N., et al. **Comprehensive analysis of amino acid and nucleotide composition of eukaryotic genomes, comparing genes and pseudogenes**. Nucleic Acids Res 30, (2002), 2515-2523.

[12] Eddy, S. R. **Computational genomics of noncoding RNA genes**. Cell 109:137-140.

[13] Endrizzi, K., et al**. Discriminative quantification of cytochrome P4502D6 and 2D7/8 pseudogene expression by TaqMan real-time reverse transcriptase polymerase chain reaction.** Anal Biochem 300, (2002) 121-131.

[14] Ewing, B, and P. Green. **Analysis of expressed sequence tags indicates 35,000 human genes**. nat genet 25, (2000) 232-240.

[15] Fletcher, B. H. et. al. **The murine chaperonin 10 gene family contains an intronless, putative gene for early pregnancy factor, *Cpn10-rs1***. Mamm Genome 12, (2001) 123-140.

[16] Fourel, G., et al. **Expression of the woodchuck N-*myc2* retroposon in brain and in liver tumors is driven by a cryptic *N-myc2* promoter**. Mol Cell Biol 12, (1992) 5336-5344.

[17] Fujii, G. H. **Transcriptional analysis of the PTEN/MMAC1 pseudogene, ΨPTEN**. Oncogene 18, (1999)1765-1769.

[18] Gesteland, R. F., and J. F. Atkins. **RECODING.** Annu Rev Biochem 65, (1996) 741-768.

[19] Goncalves, I. et al. **Nature and structure of human genes that generate retropseudogenes.** Genome Res 10, (2000) 672-678.

[20] Görlach, A. **A p47-*phox* pseudogene**. J Clin Invest 100, (1997) 1907-1918.

[21] Gregor, P., et al. **Molecular characterization of a second mouse pancreatic polypeptide**. J Biol Chem 271, (1996) 27776-27781.

[22] Hammell, A. B. et. al**. Identification of putative programmed –1 ribosomal frameshift signals in large DNA databases**, Genome Res 9, (1999) 417-427.

[23] Harrison, P. M. et al. **Digging for dead genes**. Nucleic Acids Res 29, (2001) 818-830.

[24] Harrison, P. M. and M. Gerstein. **Studying genomes through the aeons**. J Mol Biol 318, (2002) 1155-1174.

[25] Harrison, P. M. et al. **A small reservoir of disabled ORFs in the yeast genome and its implications**. J Mol Biol 316, (2002a) 409-419.

[26] Harrison, P. M. et al. **Molecular fossils in the human genome**. Genome Res 12, (2002b) 272-280.

[27] Harrison, P. M. et al. **A question of size: the eukaryotic proteome and the problems in defining it**. Nucleic Acids Res 30, (2002c) 1083-1090.

[28] Helden, J. van, et al. **Extracting regulatory sequences from the upstream region of yeast genes by computational analysis of oligonucleotide frequencies**. J Mol Biol 281, (1998) 827-842.

[29] Hill, K. E. et al. **The cDNA for rat selenoprotein P contains 10 TGA codons in the open reading frame.** J Biol Chem 266, (1991) 10050-10053.

[30] Homma K. et al. **A systematic investigation identifies a significant number of probable pseudogenes in the *Escherichia coli* genome**. Gene 294, (2002) 25-33.

[31] Hsu, L. C., and W.-C. Chang. **Sequencing and expression of human *ALDH8***. Gene 174, (1996) 319-322.

[32] Hsu, L. C., et al. **Human aldehyde dehydrogenase genes.** Gene 189, (1997) 89-94.

[33] Jeffreys, A. J., and S. Harris. **Pseudogenes**. Bioessays 1 (1984) 253-257.

[34] John, T. R. et al. **A phospholipase A_2-like pseudogene retaining the highly conserved introns of Mojave toxin**. DNA Cell Biol 15, (1996) 661-668.

[35] Kopczynski, J. B. **Translational readthrough at nonsense mutations in the HSF1 gene**. Mol Gen Genet 234, (1992) 369-378.

[36] Kroiher, M. et al. **A gene whose major transcript encodes only the substrate-binding domain of a protein-tyrosine kinase**. Gene 241, (2000) 317-324.

[37] Kuska, B. **Should scientists scrap the notion of junk DNA?** J Natl Cancer Inst 90, (1998a) 1032-1033.

[38] Kuska, B. **The semantics of junk DNA**. J Natl Cancer Inst 90, (1998b) 2215-1127.

[39] Larsen, F. et al. **CpG islands as gene markers in the human genome**. Genomics 13, (1992) 1095-1107.

[40] Lewis, B. P. et al. **Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans**. Proc Natl Acad Sci U S A 100, (2003) 189-192.

[41] Luque, T. et al. **Characterization and molecular analysis of *Adh* retrosequences in species of the *Drosophila* obscura group**. Mol Biol Evol 14, (1997) 1316-1325.

[42] Lynch, M., and A. Force. **The probability of duplicate gene preservation by subfunctionalization**. Genetics 154, (2000) 459-473.

[42] Mango, S. F., et al. **Carboxy-terminal truncation activates glp-1 protein**. Nature 352, (1991) 811-815.

[43] Max, E. E. **Plagiarized errors and molecular genetics.** http://www.talkorigins.org/faqs/molgen/. March 19, 2002.

[44] Meireles, C. M. **The tarsius γ-globin gene: pseudogene or active gene?** Mol Phylogenet Evol 13, (1999) 434-439.

[45] Mezzina, M. et al. **Characterization of the mtTFA gene and identification of a processed pseudogene in rat.** Gene 286 (2002) 111.

[46] Mighell, A. J., et al. **Vertebrate pseudogenes.** FEBS Lett 468, (2000) 109-114.

[47] Modrek, B., et al**. Genome-wide detection of alternative splicing in expressed sequences of human genes**. Nucleic Acids Res 29, (2001) 2850-2859.

[48] Morgan, K., et. al. **A transcriptionally active human type II gonadotropiin-releasing hormore receptor gene homolog**. Endocrinology 144, (2003) 423-436.

[49] Mounsey, A. et al. **Evidence suggesting that a fifth of annotated *Caenorhabditis elegans* genes may be pseudogenes**. Genome Res 12, (2002) 770-775.

[50} Mullick, J., et al. **Localization of a transcription promoter within the second exon of the cytochrome P-450c27/25 gene**. Biochemistry 34, (1995) 13729-13742.

[51] Ohno, S. **So much "junk" in our genome**. Brookhaven Symp Biol 23 (1972) 366-370.

[52] Pugh, B. F., and R. Tijan. **Diverse transcriptional functions of the multisubunit eukaryotic TFIID complex**. J Biol Chem 267, (1992) 679-682.

[53] Ramos-Onsins, S., and M. Aguade. **Molecular evolution of the *Cecropin* multigene family in Drosophila**. Genetics 150, (1998) 157-171.

[54] Reisman, D,. **A novel transcript encoded within the 10-kb first intron of the human p53 tumor suppressor gene**. Genomics 38, (1996) 364-370.

[55] Resnick, M. A., and B. S. Cox. **Yeast as an honorary mammal**. Mutat Res 451, (2000) 1-11.

[56] Reyes, A., et al. **Human mitochrondrial transcription factor A (mtTFA): gene structure and characterization of related pseudogenes**. Gene 291 (2002) 223-232.

[57] Rhodes, K., and R. G. Oshima. **A regulatory element of the human keratin 18 gene**. J Biol Chem 273, (1998), 26534-26542.

[58] Robinson, D. N., and L. Cooley**. Examination of the function of two kelch proteins generated by stop codon suppression**. Development 124, (1997) 1405-1417.

[59] Seroussi, E., **Uniquely conserved non-translated regions are involved in the generation of the two major transcripts of protein phosphatase 2Cβ .** J Mol Biol 312, (2001) 439-451.

[60] Singer, V. L., et al. **A wide variety of DNA sequences can functionally replace a yeast TATA element for transcriptional activation**. Genes Dev 4, (1990) 636-645.

[61] Srinivasan, G. et al. **Pyrrolysine encoded by UAG in Archaea**. Science 296, (2002) 1459-1462.

[62] Steneberg, P., and C. Samakovlis. **A novel stop codon readthrough mechanism produces functional Headcase protein in *Drosophila* trachea**. EMBO Rep 2, (2001) 593-597.

[63] Szeverenyi, I., et al. **Vector for IS element entrapment and functional characterization based on turning on expression of distal promoterless genes**. Gene 174, (1996) 103-110.

[64] Terrinoni, A., et al. **Cyclin D1 gene contains a cryptic promoter**. Genes, Chromosomes, and Cancer 31, (2001) 209-220.

[65] True, H. L., and S. L. Lindquist. **A yeast prion provides a mechanism for genetic variation and phenotypic diversity**. Nature 407, (2000) 477-483.

[66] Viklund, L., et al. **Expression and characterization of minican, a recombinant syndecan-1 with extensively truncated core protein**. Biochem Biophys Res Commun 290, (2002) 146-152.

[67] Wang, Y., Zhang, C.-T., and P. Dong. **Recognizing shorter coding regions of human genes based on the statistics of stop codons**. Biopolymers 63, (2002) 207-216.

[68] Wieczorek, E. et al. **Function of TAF_{II}-containing complex without TBP in transcription by RNA polymerase II.** Nature 393, (1998) 187-191.

[69] Wiley, S., R. et al. **Functional binding of the "TATA" box binding component of transcription factor TFIID to the –30 region of TATA-less promoters**. Proc Natl Acad Sci U S A 89, (1992) 5814-5818.

[70] Woodmorappe, J. **Are pseudogenes "shared mistakes'" between primate genomes?** Creation Ex Nihilo Technical Journal 14, No. 3 (2000) 55-71.

[71] Woodmorappe, J. **Pseudogene function: Regulation of gene expression**. Creation Ex Nihilo Technical Journal 21, (2003) in press.

[72] Yeh, R.F., Lim, L. P., and C. B. Burge. **Computational inference of homologous gene structures in the human genome**. Genome Res, 11 (2001) 803-816.

[73] Yi, P. et al. **Bcl-rambo beta, a special splicing variant with an insertion of an Alu-like cassette**. FEBS Lett 534, (2003) 61-68.

[74] Yousef, G. M. et al. **Genomic organization of the siglec gene locus on chromosome 19q13.4 and cloning of two new siglec pseudogenes**. Gene 286, (2002) 259-270.

[75] Zhang, B., and J.-T. Zhang. **Regulation of gene expression by internal ribosome entry sites or cryptic promoters**. Mol Cell Biol, 22 (2002) 7372-7384.

[76] Zhi, N. et. al. **Activation of a *p44* pseudogene in *Anaplasma phagocytophila* by bacterial RNA splicing**. Mol Microbiol 46, (2002) 135-145.