Pseudogenes and their composers: delving in the 'debris' of human genome

Kamalika Sen and Tapash Chandra Ghosh

Advance Access publication date 29 July 2013

Abstract

Pseudogenes, the nonfunctional homologs of functional genes and thus exemplified as 'genomic fossils' provide intriguing snapshots of the evolutionary history of human genome. These defunct copies generally arise by retrotransposition or duplication followed by various genetic disablements. In this study, focusing on human pseudogenes and their functional homologues we describe their characteristic features and relevance to protein sequence evolution. We recapitulate that pseudogenes harbor disease-causing degenerative sequence variations in conjunction with the immense disease gene association of their progenitors. Furthermore, we also discuss the issue of functional resurrection and the potentiality observed in some pseudogenes to regulate their functional counterparts.

Keywords: *human genome; pseudogenes; disease genes; evolutionary rate*

INTRODUCTION

Pseudogenes are now regarded as an important resource in evolutionary and comparative genomics and are considered assets for the studies of evolutionary relatedness and protein evolution [1]. It was in 1977, that a genomic region with a homologous structure of the gene encoding the oocyte-type 5S RNA of *Xenopus laevis* was first reported as a pseudogene owing to its 5′-end truncation and 14-bp mismatches compared with its functional counterpart [2]. The terms 'pseudogene', 'relics of evolution' and 'genomic fossil' were introduced due to their analogy with parental genes and acquired nonfunctionality due to loss-of-function mutations or removal of gene regulatory regions [3, 4]. On the other hand, other groups of scientists coined a new term 'potogenes' (for 'junk DNAs') as the DNA sequence within them bear the potential to evolve into novel genes [5, 6]. The occurrence of the faulty replicates of normal genes in a genome is still a confounding matter. The significant overlapping of the Ka/Ks values (ratio of nonsynonymous substitutions per nonsynonymous sites to synonymous substitutions per synonymous sites) among genes and pseudogenes suggests their functional roles and in support of that some pseudogenes are even observed to be transcribed to play regulatory roles in gene expression [7–11]. Besides, several works on human genes (associated with diseases) designated pseudogenes as the potential regulator of disease progression [11] capable of controlling tumor suppressors and oncogenes by acting as microRNA decoys [12, 13]. Apart from all the investigations on pseudogenes, analyses of the structural and functional attributes of their composers are also shedding light on the formation and evolution of the naturally occurring mutant genes [14, 15]. In this review, we recapture the story of functional resurrection and disease

Downloaded from https://academic.oup.com/bfg/article/12/6/536/238174 by guest on 20 August 2022

Corresponding author. Tapash Chandra Ghosh, Bioinformatics Centre, Bose Institute, P 1/12, C.I.T. Scheme VII M, Kolkata 700 054, India. Tel.: +91 33 2355 6626; Fax: +91 33 2355 3886; E-mail: tapash@jcbose.ac.in

Kamalika Sen is a senior research fellow at Bioinformatics Centre in Bose Institute, A.J.C. Bose Centenary Building, Kolkata, India, where she works on the evolutionary genomics of human pseudogene composers.

Tapash Chandra Ghosh is a professor at Bioinformatics Centre in Bose Institute, A.J.C. Bose Centenary Building, Kolkata, India. His research work is focused on the identification and understanding of the different evolutionary forces by analyzing the genomic and functional data of various organisms.

Bose Institute was founded by Acharya Jagadish Chandra Bose, F.R.S., in Calcutta on 30th November 1917. The institute is one of the earliest, perhaps the first modern research institute in India. J. C. Bose, generally acknowledged as the father of modern scientific research in India, dedicated the institute to the Nation.

© The Author 2013. Published by Oxford University Press. All rights reserved. For permissions, please email: journals.permissions@oup.com

association of the so-called 'nonfunctional' or 'junk' DNA and their significance in the evolution of human genome.

HUMAN PSEUDOGENE CLASSES

Pseudogenes, previously depicted as 'skeletons' of long-dead genes residing in our genetic closet [16] were once functional but are gradually acquiring deadly injuries to their structures (debilitating mutations) over time and are in the process of decomposition. On the basis of origin and characteristic features, they are categorized as (i) duplicated or nonprocessed pseudogenes, (ii) processed or retrotransposed pseudogenes and (iii) unitary pseudogenes (Figure 1).

The duplicated pseudogenes arise due to unequal crossing over between two homologous chromosomes (during the process of DNA replication) followed by nondeleterious mutations. In spite of having original promoter or other regulatory regions, intron and exon sequences intact [1], the erroneous recombination and subsequent mutations (misplaced stop codons, insertions, deletions) steer them to the path of nonfunctionalization. Duplicated pseudogenes are frequently found in clusters of analogous functional sequences on the same chromosome or adjacent to their paralogous functional genes and even can be inserted into a different chromosome [17]. Most of the human immunoglobulin Vλ light chain pseudogenes are clustered along with their paralogous functional genes on chromosome 22q11.2. Another two 'orphan' human Vλ light chain pseudogenes which arose by a single duplication and translocation event (which occurred before the divergence of humans and gorillas) are seen to be present on chromosome 8q11.2. [18], while human olfactory receptor pseudogenes are dispersed in most of the human chromosomes by duplication of genomic DNA [17, 19, 20]. A well-studied example of duplicated pseudogenes is human olfactory receptor (OR) pseudogene repertoire whose proportion in that gene family was reported to be significantly higher in human than in other apes and also significantly larger in apes than in the mouse. As an explanation it was suggested that the acquirement of full trichromatic vision relaxed the need for a sensitive sense of smell, albeit it did not render olfaction so fundamental and consequently some OR genes accumulated coding region disruptions while others are still evolving under evolutionary constraints [21]. All together, through a homology-based approach, almost 3000 duplicated pseudogenes were identified in human genome [4].

On the other hand, processed pseudogenes arising by reverse transcription and reinsertion (back to the genome) process can also provide molecular records on the dynamics and evolution of genomes. In this case, the processed transcript of a functional gene is reverse transcribed and incorporated into a staggered chromosome break, followed by DNA synthesis and repair [22]. The process of reverse transcription and insertion are guided by the enzymatic machinery of LINE1 non-LTR retrotransposons [23]. The processed ones are derived from mature mRNA products, lack the upstream promoters and are often entitled as 'dead on arrival' [22] because of the acquired nonfunctionality or loss of function [24] immediately upon the reinsertion process. Their structural feature shows total lack of both 5′-promoter sequence and introns besides the presence of small flanking direct repeats and polyadenylation at the 3′-end [1, 17]. Approximately 40% of the processed pseudogenes, known as ancestral pseudogenes (were formed before human and mouse diverged), stay in a syntenic region preserved in both human and mouse [25]. The processed or retropseudogenes [26] earlier speculated as 'fossilized footprints' of their parental gene expression [27] have become of increasing interest in the field of pseudogene evolution and comparative genomics since a burst of processed pseudogene genesis was observed early in the primate evolution [28]. Sequencing projects revealed that mammalian genomes show a preponderance of processed pseudogenes in the pseudogene population [25]. Based on the supposition of nearly 75 000 genes in human, 'presence of truncation' criterion (identification of pseudogenes is performed in the intergenic regions containing loss-of-function mutations in their potential coding regions, i.e. regions homologous to known protein sequences) [7] led to an estimate of 23 000–33 000 processed pseudogenes [29], whereas a lower estimation of 9000–11 000 processed pseudogenes appeared considering the number of human genes as 30 000–35 000 (International Human Genome Sequencing Consortium, 2001). Later on, after the completion of annotating euchromatic sequence of the human genome, it seems to encode only 20 000–25 000 protein-coding genes [30] and the most recent estimation of processed pseudogenes in human genome is ~8780 [31]. Chromosome 19 shows the highest

Downloaded from https://academic.oup.com/bfg/article/12/6/536/238174 by guest on 20 August 2022

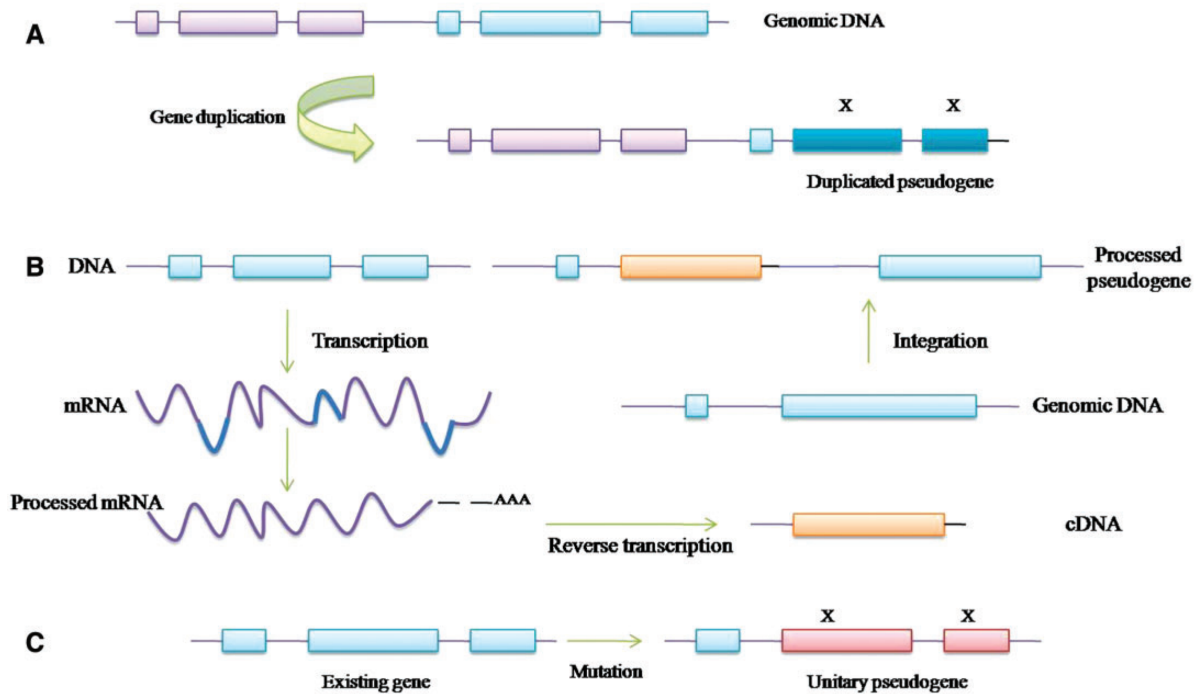

Figure 1: Origin of different kinds of pseudogenes. (**A**) Duplicated pseudogene: functional genes in a stretch of DNA are being duplicated and one of them after acquiring loss-of-function mutations at some positions (labeled as X) becomes a duplicated pseudogene. (**B**) Processed pseudogene. (**C**) Unitary pseudogene.

gene–pseudogene ratio [7] and chromosome 22 alone contains 110 retropseudogenes [32]. Apart from the ribosomal protein genes which process the maximum number of that kind of pseudogenes in human genome, there are some other genes identified to produce a large number of processed pseudogenes (Table 1) [4].

Unitary pseudogenes are like 'vestigial DNA sequences' or genetic relics of genes whose functions were important in ancestral species but became unnecessary in modern species. These pseudogenes are a natural consequence of mutations that failed to be eliminated by negative natural selection, because the functions of their products become unnecessary in present species. The deactivated singleton genes became fixed in the population by random genetic drift as unitary pseudogenes [33, 34]. Human L-gulono-γ-lactone oxidase gene (GULO) is frequently used as an instance of this category. The nonfunctionalization of GULO gene (in Human and guinea pigs), which is an enzyme that produces the precursor of vitamin C [35], has happened as a result of the relaxation of functional constraints that resulted in less severe selective pressure. It was hypothesized that the guinea pig and human ancestors survived on a naturally ascorbic acid-rich diet and hence, the loss of the enzyme did not constitute a disadvantage [33] in them. A few other examples of human pseudogenes in this category are urate oxidase (an enzyme catalyzing the oxidation of uric acid to allantoin) [36], Farnesoid x receptor beta (a nuclear receptor for lanosterol) [37], cardiotrophin-2 (CTF2), hyaluronoglucosaminidase 6 (HYAL6) [33]. As a whole, 76 unitary pseudogenes were reported by Zhang *et al*. [33] identifying the sequence signature left by genic losses in human and having functional counterparts in the mouse genome.

EVOLUTIONARY FACETS AND GENOMIC ATTRIBUTES OF PSEUDOGENES AND THEIR COMPOSERS IN HUMAN

The theory of positive Darwinian selection asserts that functionally unimportant genomic regions have lower rates of nucleotide substitution as mutations in those places do not offer any significant selective advantage [38, 39]. On the contrary, neutral mutation hypothesis claims higher substitution rate for the genes, or gene regions, with less functional importance than the essential regions as the latter ones undergo stronger purifying (negative) selection

Downloaded from https://academic.oup.com/bfg/article/12/6/536/238174 by guest on 20 August 2022

Table 1: Examples of human genes with large number of processed pseudogenes

Gene name	Number of processed pseudogenes
Cyclophilin A	63
Glyceraldehyde-3-phosphate dehydrogenase	52
Nucleophosmin	34
Cytochrome c	31
Histone H3.3	25
Prohibitin	20
Ubiquitin-like protein SMT3B	19
Small nuclear ribonucleoprotein G	18
NADH-ubiquinone oxidoreductase MLRQ subunit	16
Hsc70-interacting protein	14

[40–43]. This latter interpretation of course assumes that natural selection only plays a significant role eliminating deleterious variants from populations. The higher rate of nucleotide substitution at the third codon positions than the first two ascertains the neutral mutation hypothesis as most of the third position substitutions are synonymous and do not change the encoded amino acids [44–46]. According to the theory of neutral evolution [40], pseudogenes remain unconstrained by selection, randomly accumulate mutations (insertions, deletions and substitutions) over time and this neutral nature of all pseudogenic regions renders them suitable to determine different forms and rates of neutral sequence evolution among different regions in the genome and even among different organisms [7]. In case the pseudogenes have a functional counterpart, the presence of one copy of the gene suffices for the requirements of the organism. Then, the pseudogenic mutations, whether disruptive or not, will not undergo purifying selection and will have an equal probability of becoming fixed in the population [6, 47–49]. On account of the rapid accumulation of frequent mutations, the pseudogenes usually degenerate and melt into the background of the surrounding DNA which has been detected in prokaryotes [6, 50, 51]. In contrast, the pseudogenic regions in eukaryotic genomes avoid full degeneration owing to less selective pressure for deletion [17, 52]. Again, the loci of pseudogene inclusion also affect their evolution as the incorporation of some pseudogenes is deleterious for other functional genes and thus will undergo purifying selection and are lost. In general, established pseudogenes will evolve and undergo genetic drift due to the absence of apparent selection pressures to prevent random mutations which is supported by the faster evolutionary rate of processed pseudogenes than their corresponding functional paralogs [17, 53, 54].

Besides the pseudogenes, human genes configuring them are also now coming forth as significant resources in the study of human protein evolution. Functionally more important genes are prone to encounter stronger selective pressure than the genes having less functional importance [42]. So, it is interesting to delve the evolutionary features of those genes which are functionally important but give rise to nonfunctional duplicated and processed pseudogenes. Human genes configuring duplicated pseudogenes act like an essential group of genes as they were observed to harbor significantly higher number of highly expressed and hub protein encoding genes (incidentally, hubs are essential for the maintenance of the network structure and the genes encoding them are widely expressed and essential for the survival of the organism) [14, 55, 56]. On the other hand, the progenitors of duplicated pseudogenes were seen to recombine more frequently, be longer genes and yield functionally redundant duplicates corroborating with their highly evolving nature compared to the other duplicated genes casting functional genes [14]. As an explanation for their higher rate of evolution, it was speculated that the essential-like nature of the genes casting duplicated pseudogenes signifies their urge to increase gene recombination rate in order to elevate the paralog number. It was also supported by the previous proposition that in mammals, the hub protein encoding genes display higher gene duplicability (paralog number) by virtue of their need to be produced in a high dosage [57] and they may intend to reserve back-up copy for future defense. Even though the functional similarity of the duplicate genes can offer a back-up for gene loss through mutations [58], the redundant copies are not protected against deleterious mutations and thus are evolutionarily unstable [59]. Moreover, the progenitors while increasing the duplicates with redundant functions may exceed the optimum necessity of the cell. The redundant copies may result into the dosage imbalance which is deleterious for the cellular integrity according to the balance hypothesis [60] of proteins in interaction network. In such a scenario, when the unstable copies encounter mutations (mutations are loss of function in nature but the functions are compensated by their functional copies), the duplicated

Downloaded from https://academic.oup.com/bfg/article/12/6/536/238174 by guest on 20 August 2022

genes will escape the filtration process of natural selection and will be restored in the genome as duplicated pseudogenes [61].

During the last several years, processed or retropseudogenes are being cataloged and characterized in many completely sequenced genomes including human. However, there remains a dearth of reports on the structural and functional characterization of the human genes configuring this kind of pseudogenes. The pioneering work of Goncalves *et al.* [29] focused on 181 human functional genes casting 249 retropseudogenes. Their study delved out a highly expressed and evolutionarily conserved character of the retropseudogene ancestors. The genes (forming processed pseudogenes) were also reported to be short and GC-poor (GC level: the molar ratio of guanine+cytosine in DNA), indicating their high efficiency for retrotransposition. As an explanation of the aforementioned observations, it was argued that the mechanism of retropseudogene insertion involves Long Interspersed Nuclear Element (LINE) reverse transcriptase [23, 62] which is actively transcribed in the nucleus, transported to the cytoplasm for getting translated and the ribonucleoprotein particle (including L1 reverse transcriptase) formed there, is again transported back to the nucleus for target-primed reverse transcription to carry out retrotransposition [63]. As mammalian LINE elements were reported to be GC-poor [64], the low GC content of genes forming retropseudogenes signifies their high efficiency of reverse transcription using LINE reverse transcriptase. Moreover, the shorter coding sequence length (bears a positive correlation with the number of retropseudogenes per ancestor gene) also corroborates with the ease of a short sequence to get transferred between cytoplasm and nucleus, to be reverse transcribed and integrated within the genome with less deleterious effects [29]. Besides, the high expressivity of the genes configuring processed pseudogenes (which were reported to be evolutionarily conserved) was elucidated by their high interactivity along with a predominance of hub protein encoding genes [15]. Again, the higher disorderness of the genes, although being incompatible with their expression level [65], was demonstrated as prerequisite of intense network involvement as the unstructured regions of the translated forms facilitate the network connectivity [66]. The progenitor genes were also observed to retain the translated form in a stable configuration (higher protein stability) which may be due to the fact that the regions with structural disorder keep up the stability of proteins *in vivo* through the attachment with the corresponding target molecules [67]. Besides being enriched with disordered residues, hub proteins were seen to harbor sequence repeats to enlarge the available surface area predisposing them for functioning via protein–protein interactome [68], which is also consistent with the abundance of repeat sequence containing genes in the genes with retropseudogenes supporting the presence of protein disordered regions in them. In addition, a positive alliance between the mRNA abundance and the propensity of repeat sequence containing genes supports the fact that tandem repeats in human genes (configuring retropseudogenes) can positively regulate the level of transcription [69]. On the other hand, proteins encoded by the progenitor genes were observed to configure large ribosomal subunits engaged in transcription-associated jobs which supports the idea that proteins carrying disordered regions are able to perform some essential functions directly linked to their structural disorder [70]. Again, it was argued that intrinsically unstructured regions of a polypeptide segment offer sites for alternative splicing as the disordered regions can tolerate functional or regulatory diversity without any disturbance in the protein sequence [71] and as the alternative splicing event accounts for the quantitative and qualitative regulation of gene expression [72], it was hypothesized that progenitor genes go through an extensive alternative splicing event to form a number of spliced isoforms elevating the mRNA abundance level which, together with the higher extent of endurance (the genes executed a lower mRNA decay rate), may contribute to enhance the level of gene expression. The higher mRNA abundance contributes for an elevated reverse transcription process which in turn increases the chance for retro-pseudogenization [15].

A number of genes belonging to the third category of human pseudogene series systematically identified by Zhang *et al.* (2010) [33] are especially interesting as they are themselves the functional genes encountering several disruptive mutations hindering any successful transcription or translation. In their work, Zhang *et al.* (2010) [33] reported that before being pseudogenized, human unitary pseudogenes were involved in many different biological processes (e.g. integrin-mediated signaling pathway) and their translated forms executed miscellaneous molecular functions (like endopeptidase activity,

Downloaded from https://academic.oup.com/bfg/article/12/6/536/238174 by guest on 20 August 2022

serine-type peptidase activity, serine-type endopeptidase activity, metalloendopeptidase activity and serine hydrolase activity) at various cellular locations. Again, analysis on protein domain revealed that two Pfam domains—reprolysin family propeptide and reprolysin (M12B) family zinc metalloprotease—are affluent in the human unitary pseudogenes. Moreover, it was observed that in the course of evolution (when compared with mouse), the following five testis-specific genes were pseudogenized in humans: testicular cell adhesion molecule 1(TCAM1), testis expressed gene 16 (TEX16), testis expressed gene 21 (TEX21), testis-specific serine kinase 5(TSSK5) and cytochrome c testis (CYCT) [73]. Analyses of primate evolution also show that gene loss by virtue of unitary pseudogenization events occurred at every stage, i.e. from the human lineage alone to the last common ancestor of the great apes, the old world monkeys, the new world monkeys and the tarsiers [33].

FUNCTIONAL EVIDENCES OF PSEUDOGENES

A number of intergenic areas exhibiting transcriptional activity overlap with annotated human pseudogenes, which implies that some of the 'nonfunctional', 'junk' pseudogenes may have life left in them [16]. The high abundance of pseudogenes in various species and their evolutionary conservation signifies their involvement in important biological processes [74]. Recent evidences indicate that some pseudogenes are transcribed into noncoding RNAs and post-transcriptionally modulate their parental genes by three distinct mechanisms: (i) gene expression suppression by natural antisense RNA; (ii) RNA interference by producing short interfering RNAs (siRNAs) and (iii) act as microRNA decoys (miRNA) [74] (Figure 2). In 1986, McCarrey and Riggs [75] were the first to state the possible significant role of pseudogenes in morphogenesis as a source of intracellular inhibitor molecules. They speculated that nonprocessed pseudogenes (pseudogenes with disabling mutations that have been accumulated during the event of gene duplication) can have the possibility to be transcribed from the opposite strand of their corresponding parental genes and subsequently suppress the translation of the functional counterparts by forming sense–antisense double-stranded RNAs [74, 75]. Later in 1992, a human pseudogene TOP1 (Human DNA Topoisomerase I Pseudogene) was observed to produce a naturally occurring antisense transcript and was the first report of unlinked antisense transcription (antisense transcripts and their targets are from different genes) in eukaryotes [76]. Pseudogene acting as 'antigene' was then demonstrated in a fresh water snail *Lymnaea stagnalis*, where neural nitric oxide synthase (nNOS) gene was observed to be post-transcriptionally regulated by an antisense transcript encoded by its own pseudogene (pseudo-NOS) [10]. Another stimulating study by Hirotsune *et al.* [8] on the mouse genome showed that Makorin1-p1, a processed pseudogene of Makorin1, shows transcriptional activity and regulates the functional gene. But, the work was argued against in 2006 as it was discovered that both Makorin1-p1 alleles are methylated and thus the pseudogene is an 'unexpressed or silent pseudogene', which re-established the 'evolutionary relic' nature of mammalian pseudogenes [77, 78]. The idea that pseudogenes can evolve into novel genes was proposed by Proudfoot [3] and later, Brosius *et al.* [5] coined the terms 'potonuons' or 'potogenes' for pseudogenes as they (pseudogenes) showed the potentiality to be the vast repertoire of sequences with the ability to shape the evolution of an organism rather than considered as merely dispensable genomic noise. This possibility to contribute sequences for future use or eventually acquire distinctive function [6] can almost resolve the disparity arising because of the dual face of pseudogenes (capability to regulate their functional genes or staying as junk DNA). Oct4P1 can serve the example of 'potonuons' as it is seen to preserve some aspect of the parent Oct4 function in stem cells in spite of its coding sequence truncation [79]. The interdependence of a gene and its pseudogene is observed during the knockdown of ABCC6P1 expression, which is a co-expressed pseudogene of the pseudoxanthoma elasticum gene ABCC6 and it results in a decrease of the expression of the later one signifying a regulatory interrelation between the parental gene and its pseudogene [80]. Again, in context of the potential of human pseudogenes to introduce pathogenic mutations into functional genes by pseudogene-mediated gene conversion, it was also speculated that pseudogenes might have served as templates of multiple, potentially advantageous changes in their single-copy functional parental genes which eventually became fixed on the course of evolution [81]. The speculation was based on the comparative analysis of Hayakawa *et al.* in 2005 on human sialic acid

Downloaded from https://academic.oup.com/bfg/article/12/6/536/238174 by guest on 20 August 2022

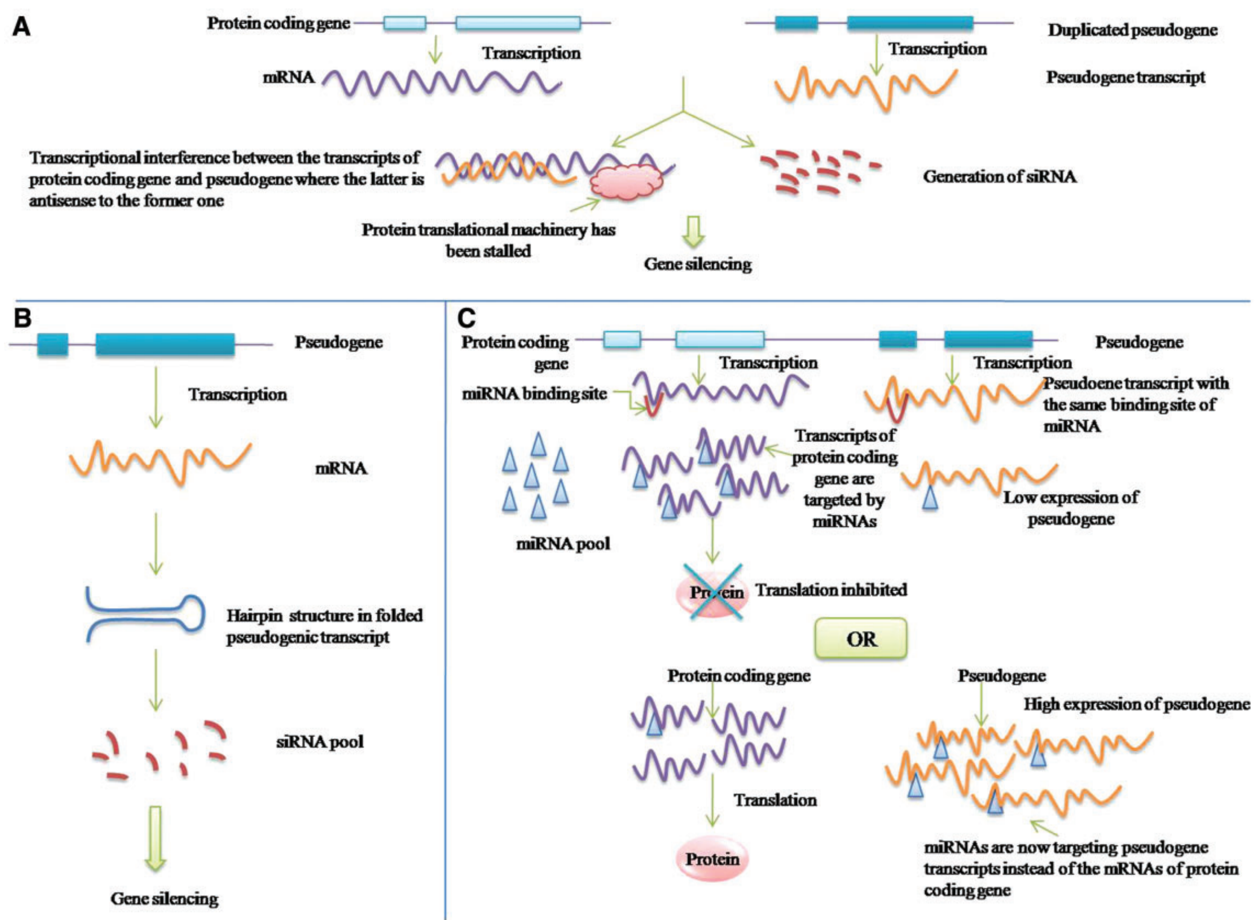

Figure 2: Different mechanisms of pseudogene function at post-transcriptional level. (**A**) Transcriptional interference between a sense strand mRNA of parental gene and antisense transcript of its duplicated pseudogene leading to translational inhibition or siRNA generation resulting in gene silencing. (**B**) siRNAs are generated from hairpin/stem–loop structures formed in folded pseudogenic transcripts. (**C**) Homologous pseudogenes with binding site (for miRNA) same as that of the parent gene can act as decoy of miRNA. When the pseudogene is lowly expressed, miRNA pool target parental gene inhibiting its translation. A high level of expression of the pseudogene sequesters miRNA from the parental gene results in the latter being translated into protein. (**C**) is modified from the Figure I in the original article by Muro *et al*. [78].

binding Ig-like lectin 11 (SIGLEC11) gene and its pseudogene with their homologs in the chimpanzee, bonobo, gorilla and orangutan where they observed human SIGLEC11 gene to have converted its 5′ upstream region and exons encoding sialic acid recognition domain (~2 kb) by the closely flanking SIGLECP16 pseudogene. Thus, in comparison with other primates, human SIGLEC11 shows altered substrate binding capacity suggesting an adaptive change that could have borne an imperative role in the evolution of the genus Homo [82].

A recent work by Muro and Andrade-Navarro [83] revealed the functional relevance of mammalian pseudogenes in antisense transcription as they illustrated the formation of duplicated pseudogenes as the mechanism of purposeful *trans*-NATs (*trans*-acting naturally occurring antisense transcripts) generation providing potential regulation of the parental gene. In support of this, they observed an elevated selection pressure to preserve the similarity between the duplicated pseudogene and the parental gene in the region corresponding to the *trans*-NAT, suggesting a functional association between the *trans*-NATs and the parental gene. Besides, pseudogene can even function unrelated to its ancestral gene, which is evidenced for human Xist noncoding RNA gene which evolved in eutherians by pseudogenization of the protein coding gene Lnx3 and initiates X chromosome inactivation [80, 84]. In a broad perspective, pseudogenes encoding endo-siRNAs

Downloaded from https://academic.oup.com/bfg/article/12/6/536/238174 by guest on 20 August 2022

endow with a linkage to the evolution of miRNA-mediated regulation which again may provide insights in the identification of new therapeutic targets in cancer [13, 85].

DISEASE ASSOCIATION OF PSEUDOGENES AND THEIR PARENTAL GENES

The fascinating work of Poliseno *et al.* [86] demonstrated gene expression regulatory interaction between a gene and its pseudogene and correlated the presence/absence of that pseudogene with onset of diseases. They observed that PTENP1, a biologically active pseudogene of Phosphatase and tensin homolog (PTEN) tumor suppressor gene, has the ability to regulate the cellular levels of the latter by acting as a perfect 'miRNA decoy' for its parental gene. The pseudogene retains miRNA binding sites, competes for the binding of many miRNA molecules at once and consequently rescues PTEN from miRNA-mediated repression. Thus, underexpression of the pseudogene results in decreased level of PTEN transcripts leading an abnormal cell proliferation in prostate cancer cells [86]. The de-repression regulatory ability of PTENP1 was also seen to be abrogated in DICER-null (Dicer: an endoribonuclease in the RNase III family) colon carcinoma cells [13, 86]. It was thus suggested that PTENP1 acts as a bona fide tumor suppressor gene, hence the locus subjected to copy number loss (correlates with a decrease in PTEN) during tumorigenesis. A similar relationship was also observed between the pseudogene KRAS1P and its parental oncogene KRAS [86]. A high expression of BRAF pseudogene was also seen to be correlated with the initiation of goiter formation and then in the progression of papillary thyroid carcinoma [87]. In this case, it was speculated that the pseudogene may escape X-inactivation and remains active in some thyroid tumors. Again, pseudogenes which are co-localized with their parental genes were seen to be the potential candidates of gene conversion event that may produce intriguing candidate disease genes [11]. The close proximity of pseudogenes and the parental genes increases the likelihood of recombination between them. Hence, the sequence variations accumulated in pseudogenes turn out to be disease causing mutations when they are transferred to other genes by gene conversion (Table 2) [11, 88–100]. Gupta *et al.* [88] reported the recombination events between the functional gene and its pseudogene on chromosome 22 as a common cause of von Willebrand disease in humans. In another study, Bischof *et al.* [11] identified 1945 duplicated pseudogenes of that kind and evaluated their probable function in gene conversion and disease. This includes: (i) retinitis pigmentosa 9 (RP9) pseudogene carrying a c.509A4G mutation which gives rise to p.Asp170Gly substitution, which is again associated with the RP9 form of autosomal dominant retinitis pigmentosa (adRP); (ii) inosine monophosphate dehydrogenase 1 (IMPDH1) pseudogene encounters a c.676G4A mutation that forms a p.Asp226Asn substitution resulting in another type of retinitis pigmentosa (RP10); and (iii) phosphoglycerate kinase 1 (PGK1) pseudogene (PGK1P1) carrying a c.837T4C mutation which produces a p.Ile252Thr substitution associated with a phosphoglycerate kinase deficiency.

Table 2: Examples of human candidate genes for gene conversion (with their corresponding pseudogenes) mediated disease

Gene name	Disease
IGLLI	B cell deficiency
ABCC6	Pseudoxanthoma elasticum
CRYBB2	Autosomal dominant cataract
CYP2IA2	Congenital adrenal hyperplasia
FOLRI	Neural tube defects
GBA	Type 2 Gaucher disease
IDS	Hunter syndrome
NCFI	Chronic granulomatous disease
PKDI	Autosomal dominant polycystic kidney disease
SBDS	Shwachman–Bodian–Diamond syndrome
VWF	Type 3 vonWillebrand disease

Though pseudogenes were observed to harbor disease causing sequence discrepancies, the involvement of their parental genes with disease still needs to be explored thoroughly. Recently, we unveiled a strong association of diseases with the genes casting pseudogenes in human [101]. We provided evidence by finding a higher abundance of genes targeted by disease-associated miRNAs, genes with polymorphisms on miRNA target sites, presence of 'disease gene-specific' network properties, affluence of dosage sensitive genes and genes having disease causing nonsynonymous mutations.

DISCUSSION

Pseudogenes, the dysfunctional genomic copies which remain in the genome for millions of years

Downloaded from https://academic.oup.com/bfg/article/12/6/536/238174 by guest on 20 August 2022

are often posited as evolutionary relics. They arose as functional pieces of genomes but in the course of evolution, switched off that mode and set down as 'junk DNA'. Certainly, pseudogenes are considered among the most persuasive pieces of support for biological evolution as they appear as left over or plagiarized slip-ups from evolutionary predecessors. But now, instead of being seen as silent relics, many of them are observed to be transcribed into RNA, some exhibit tissue-specific patterns of activation and some function as a source of information for producing genetic diversity [12]. However, in the context of pseudogene functionality, it was reported that, transcription from those regions as such does not necessarily imply a functional role, because transcriptional activation at a particular genomic locus may have a ripple effect on the neighboring loci [102] which again can be a possible reason of the pseudogenic transcriptional activity [103]. Nevertheless, several lines of evidences are now redefining the ambiguous margin between the 'living' functional genes and 'dead' pseudogenes as some of them were observed to be transcriptionally 'alive' and even possessing biochemical roles [104]. In their very sensible piece of opinion, Zheng and Gerstein [104] proposed a new definition, rather classification system, to resolve the intrinsic paradox of functional pseudogenes. They named the pseudogenes with intermediate functionality as 'Ghost pseudogenes' and subdivided them as: (i) Exapted (with new biological function), (ii) Piggy-back (with novel functions but unrelated to the hosts) and (iii) Dying pseudogenes (dying though retaining some transcriptional activity). Rest of the pseudogenes with no sign of functionality were termed as 'Dead'. Evidently, pseudogenes exhibit dissimilarity in their functional potentiality as a number of them are transcriptionally silent but others are active, raising the question of whether they are spurious execution of cellular energy or instead harnessed by the cell to control coding sequences [6]. The issue of retaining some specific functions, acquiring novel ones or total resurrection of their original functions [6] is now a matter of interest to the geneticists. Possessing a pseudogene with such functional potentiality to constructively regulate the parental genes is also expected to be evolutionarily conserved [12, 86]. The reports demonstrating the deregulation of human pseudogenes in disease progression and the immense association of the progenitor genes of pseudogenes with known disease genes are just the tip of the iceberg as so many questions are yet to be addressed regarding the pseudogene deregulation in human genetic diseases.

Future perspectives

The moniker of 'genomic nuisance' had so long restrained the pseudogenes from being considered as a regulatory element in the field of understanding the biology of health and disease and consequently, the pseudogene probes were often absent from the commercially available microarray chips [12] which obviously to some extent retarded the advancement of medical genomics. Accordingly, apposite experimental designs using next-generation sequencing and other related transcriptomic studies on pseudogene-derived transcripts and proper interpretation is truly required to get a vivid view of their role in the field of molecular genetics. Obviously, 'not a nuisance', rather the pseudogenes are now 'making more sense' in medical genomics. With the advent of the researches on the human pseudogene composers, whether it is duplicated or processed or unitary, exposing the genomic imperatives constraining their evolution as well as a new facet of their physical and functional attributes, we will indubitably be able to trace the probable course of pseudogenization of human genes and relevance in proteome evolution. Besides, research on the involvement of human pseudogenes in the disease domain will unveil a new stratum of complexity in their functional roles. Novel findings in this field may introduce more 'dead' genes giving rise to a new dawn of the dead.

Key Points

- Pseudogenes provide glimpses of the evolutionary chronicles of human genome.
- The neutral characteristic of all pseudogenic regions renders them relevant to determine the nature of neutral sequence evolution among different regions in the genome and even among different organisms.
- Some pseudogenes were observed to post-transcriptionally modulate their parental genes by three distinct mechanisms: (i) gene expression suppression by natural antisense RNA; (ii) RNA interference by producing siRNAs and (iii) act as miRNA decoys.
- Pseudogenes can partially retain or totally resurrect their original functions.
- Possessing a pseudogene with functional potentiality to beneficially regulate their parental genes is evolutionarily conserved.
- Pseudogenes were observed to harbor disease causing sequence discrepancies over their entirety and recently an immense association of disease genes was unrevealed in the repertoire of the human genes casting pseudogenes.

Downloaded from https://academic.oup.com/bfg/article/12/6/536/238174 by guest on 20 August 2022

ACKNOWLEDGEMENTS

We thank Mr Jyotirmoy Das for his technical help. We are grateful to Professor Fernando Alvarez of Seccion Biomatematica Facultad de Ciencias, Montevideo, Uruguay and Dr Pamela Mukhopadhyay of University of Queensland, Australia for their critical reading of the manuscript. We are also thankful to two anonymous referees for their helpful suggestions in improving the manuscript.

FUNDING

The Department of Biotechnology, Government of India, Bioinformatics Centre (Centre Of Excellence), BT/BI/03/002/2003; Date: 17-01/2003.

References

1. Zhang ZL, Gerstein M. Large-scale analysis of pseudogenes in the human genome. *Curr Opin Genet Dev* 2004;**14**: 328–35.
2. Jacq C, Miller JR, Brownlee GG. A pseudogene structure in 5S DNA of Xenopus laevis. *Cell* 1977;**12**:109–20.
3. Proudfoot N. Pseudogenes. *Nature* 1980;**286**:840–1.
4. Zhang ZL, Harrison PM, Liu Y, *et al*. Millions of years of evolution preserved: A comprehensive catalog of the processed pseudogenes in the human genome. *Genome Res* 2003;**13**:2541–58.
5. Brosius J, Gould SJ. On "genomenclature": a comprehensive (and respectful) taxonomy for pseudogenes and other "junk DNA". *Proc Natl Acad Sci USA* 1992;**89**: 10706–10.
6. Balakirev ES, Ayala FJ. Pseudogenes: are they "junk" or functional DNA? *Annu Rev Genet* 2003;**37**:123–51.
7. Torrents D, Suyama M, Zdobnov E, *et al*. A genome-wide survey of human pseudogenes. *Genome Res* 2003;**13**: 2559–67.
8. Hirotsune S, Yoshida N, Chen A, *et al*. An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene. *Nature* 2003;**423**:91–6.
9. Kandouz M, Bier A, Carystinos GD, *et al*. Connexin43 pseudogene is expressed in tumor cells and inhibits growth. *Oncogene* 2004;**23**:4763–70.
10. Korneev SA, Park JH, O'Shea M. Neuronal expression of neural nitric oxide synthase (nNOS) protein is suppressed by an antisense RNA transcribed from an NOS pseudogene. *J Neurosci* 1999;**19**:7711–20.
11. Bischof J, Chiang A, Scheetz T, *et al*. Genome-wide identification of pseudogenes capable of disease-causing gene conversion. *Hum Mutat* 2006;**27**:545–52.
12. Pink RC, Wicks K, Caley DP, *et al*. Pseudogenes: pseudo-functional or key regulators in health and disease? *RNA* 2011;**17**:792–8.
13. Swami M. Small RNAs: Pseudogenes act as microRNA decoys. *Nat Rev Cancer* 2010;**10**:535.
14. Sen K, Podder S, Ghosh TC. Insights into the genomic features and evolutionary impact of the genes configuring duplicated pseudogenes in human. *Febs Lett* 2010;**584**:4015–8.
15. Sen K, Podder S, Ghosh TC. On the quest for selective constraints shaping the expressivity of the genes casting retropseudogenes in human. *BMC Genomics* 2011;**12**:401.
16. Gerstein M, Zheng D. The real life of pseudogenes. *Sci Am* 2006;**295**:48–55.
17. Mighell AJ, Smith NR, Robinson PA, *et al*. Vertebrate pseudogenes. *Febs Lett* 2000;**468**:109–14.
18. Frippiat JP, Dard P, Marsh S, *et al*. Immunoglobulin lambda light chain orphons on human chromosome 8q11.2. *Eur J Immunol* 1997;**27**:1260–5.
19. Rouquier S, Taviaux S, Trask BJ, *et al*. Distribution of olfactory receptor genes in the human genome. *Nat Genet* 1998;**18**:243–50.
20. Brand-Arpon V, Rouquier S, Massa H, *et al*. A genomic region encompassing a cluster of olfactory receptor genes and a myosin light chain kinase (MYLK) gene is duplicated on human chromosome regions 3q13-q21 and 3p13. *Genomics* 1999;**56**:98–110.
21. Gilad Y, Wiebe V, Przeworski M, *et al*. Loss of olfactory receptor genes coincides with the acquisition of full trichromatic vision in primates. *PLoS Biol* 2004;**2**:E5.
22. Vanin EF. Processed pseudogenes - Characteristics and evolution. *Annu Rev Genet* 1985;**19**:253–72.
23. Esnault C, Maestre J, Heidmann T. Human LINE retrotransposons generate processed pseudogenes. *Nat Genet* 2000;**24**:363–7.
24. Graur Dan LW-H. *Fundamentals of Molecular Evolution*. Sunderland, MA: Sinauer Associates Inc., 2000.
25. Zhang ZL, Carriero N, Gerstein M. Comparative analysis of processed pseudogenes in the mouse and human genomes. *Trends Genet* 2004;**20**:62–7.
26. Csuros M, Miklos I. Statistical alignment of retropseudogenes and their functional paralogs. *Mol Biol Evol* 2005;**22**: 2457–71.
27. Podlaha O, Zhang JZ. Processed pseudogenes: the 'fossilized footprints' of past gene expression. *Trends Genet* 2009; **25**:429–34.
28. Ohshima K, Hattori M, Yada T, *et al*. Whole-genome screening indicates a possible burst of formation of processed pseudogenes and Alu repeats by particular L1 subfamilies in ancestral primates. *Genome Biol* 2003;**4**:R74.
29. Goncalves I, Duret L, Mouchiroud D. Nature and structure of human genes that generate retropseudogenes. *Genome Res* 2000;**10**:672–8.
30. Collins F, Lander E, Rogers J, *et al*. Finishing the euchromatic sequence of the human genome. *Nature* 2004;**431**: 931–45.
31. Karro JE, Yan Y, Zheng D, *et al*. Pseudogene.org: a comprehensive database and comparison platform for pseudogene annotation. *Nucleic Acids Res* 2007;**35**:D55–60.
32. Dunham I, Shimizu N, Roe BA, *et al*. The DNA sequence of human chromosome 22. *Nature* 1999;**402**:489–95.
33. Zhang ZD, Frankish A, Hunt T, *et al*. Identification and analysis of unitary pseudogenes: historic and contemporary gene losses in humans and other primates. *Genome Biol* 2010; **11**:R26.
34. Zhu J, Sanborn JZ, Diekhans M, *et al*. Comparative genomics search for losses of long-established genes on the human lineage. *PLoS Comput Biol* 2007;**3**:e247.
35. Nishikimi M, Fukuyama R, Minoshima S, *et al*. Cloning and chromosomal mapping of the human nonfunctional gene for L-gulono-gamma-lactone oxidase, the enzyme for l-ascorbic-acid biosynthesis missing in man. *J Biol Chem* 1994;**269**:13685–8.

Downloaded from https://academic.oup.com/bfg/article/12/6/536/238174 by guest on 20 August 2022

36. Wu X, Lee C, Muzny D, *et al*. Urate oxidase - primary structure and evolutionary implications. *Proc Natl Acad Sci USA* 1989;**86**:9412–6.
37. Zhang Z, Cayting P, Weinstock G, *et al*. Analysis of nuclear receptor pseudogenes in vertebrates: How the silent tell their stories. *Mol Biol Evol* 2008;**25**:131–43.
38. Clarke B. Selective constraints on amino-acid substitutions during evolution of proteins. *Nature* 1970;**228**:159.
39. Milkman R. Molecular evolution – selection is major determinant. *Trends Biochem Sci* 1976;**1**:N152–4.
40. Kimura M. Evolutionary rate at molecular level. *Nature* 1968;**217**:624–6.
41. King J, Jukes T. Non-Darwinian evolution. *Science* 1969; **164**:788–98.
42. Kimura M, Ohta T. Some principles governing molecular evolution. *Proc Natl Acad Sci USA* 1974;**71**: 2848–52.
43. Dickerson RE. The structures of cytochrome c and the rates of molecular evolution. *J Mol Evol* 1971;**1**:26–45.
44. Kimura M. Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature* 1977;**267**:275–6.
45. Jukes TH. Neutral changes during divergent evolution of hemoglobins. *J Mol Evol* 1978;**11**:267–9.
46. Li WH, Gojobori T, Nei M. Pseudogenes as a paradigm of neutral evolution. *Nature* 1981;**292**:237–9.
47. Gojobori T, Li WH, Graur D. Patterns of nucleotide substitution in pseudogenes and functional genes. *J Mol Evol* 1982;**18**:360–9.
48. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 1980;**16**:111–120.
49. Li WH, Gojobori T. Rapid evolution of goat and sheep globin genes following gene duplication. *Mol Biol Evol* 1983; **1**:94–108.
50. Andersson JO, Andersson SG. Pseudogenes, junk DNA, and the dynamics of Rickettsia genomes. *Mol Biol Evol* 2001;**18**:829–39.
51. Kuo CH, Ochman H. The extinction dynamics of bacterial pseudogenes. *PLoS Genet* 2010;**6**:e1001050.
52. Harrison PM, Gerstein M. Studying genomes through the aeons: protein families, pseudogenes and proteome evolution. *J Mol Biol* 2002;**318**:1155–74.
53. Blake RD, Hess ST, Nicholson-Tuell J. The influence of nearest neighbors on the rate and pattern of spontaneous point mutations. *J Mol Evol* 1992;**34**:189–200.
54. Ophir R, Itoh T, Graur D, *et al*. A simple method for estimating the intensity of purifying selection in protein-coding genes. *Mol Biol Evol* 1999;**16**:49–53.
55. He X, Zhang J. Why do hubs tend to be essential in protein networks? *PLoS Genet* 2006;**2**:826–34.
56. Goh KI, Cusick ME, Valle D, *et al*. The human disease network. *Proc Natl Acad Sci USA* 2007;**104**:8685–90.
57. Liang H, Li WH. Gene essentiality, gene duplicability and protein connectivity in human and mouse. *Trends Genet* 2007;**23**:375–8.
58. Wagner A. Robustness against mutations in genetic networks of yeast. *Nat Genet* 2000;**24**:355–61.
59. Nowak MA, Boerlijst MC, Cooke J, *et al*. Evolution of genetic redundancy. *Nature* 1997;**388**:167–71.
60. Papp B, Pal C, Hurst LD. Dosage sensitivity and the evolution of gene families in yeast. *Nature* 2003;**424**:194–7.
61. Zhang JZ. Evolution by gene duplication: an update. *Trends Ecol Evol* 2003;**18**:292–8.
62. Dhellin O, Maestre J, Heidmann T. Functional differences between the human LINE retrotransposon and retroviral reverse transcriptases for in vivo mRNA reverse transcription. *EMBO J* 1997;**16**:6590–602.
63. Kazazian HH, Moran JV. The impact of L1 retrotransposons on the human genome. *Nat Genet* 1998;**19**:19–24.
64. Korenberg JR, Rykowski MC. Human genome organization: Alu, lines, and the molecular structure of metaphase chromosome bands. *Cell* 1988;**53**:391–400.
65. Singh GP, Dash D. How expression level influences the disorderness of proteins. *Biochem Biophys Res Commun* 2008; **371**:401–4.
66. Haynes C, Oldfield CJ, Ji F, *et al*. Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. *PLoS Comput Biol* 2006;**2**:890–901.
67. Yen HCS, Xu QK, Chou DM, *et al*. Global protein stability profiling in mammalian cells. *Science* 2008;**322**:918–23.
68. Dosztányi Z, Chen J, Dunker AK, *et al*. Disorder and sequence repeats in hub proteins and their implications for network evolution. *J Proteome Res* 2006;**5**:2985–95.
69. Gemayel R, Vinces MD, Legendre M, *et al*. Variable tandem repeats accelerate evolution of coding and regulatory sequences. *Annu Rev Genet* 2010;**44**:445–77.
70. Hazy E, Tompa P. Limitations of induced folding in molecular recognition by intrinsically disordered proteins. *Chemphyschem* 2009;**10**:1415–9.
71. Romero PR, Zaidi S, Fang YY, *et al*. Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. *Proc Natl Acad Sci USA* 2006;**103**:8390–5.
72. Pavlicek A, Gentles AJ, Paces J, *et al*. Retroposition of processed pseudogenes: the impact of RNA stability and translational control. *Trends Genet* 2006;**22**:69–73.
73. Zhang Z, Gerstein M. The human genome has 49 cytochrome c pseudogenes, including a relic of a primordial gene that still functions in mouse. *Gene* 2003;**312**:61–72.
74. Wen YZ, Zheng LL, Qu LH, *et al*. Pseudogenes are not pseudo any more. *RNA Biol* 2012;**9**:27–32.
75. McCarrey J, Riggs A. Determinator-inhibitor pairs as a mechanism for threshold setting in development - a possible function for pseudogenes. *Proc Natl Acad Sci USA* 1986;**83**:679–83.
76. Zhou BS, Beidler DR, Cheng YC. Identification of antisense RNA transcripts from a human DNA topoisomerase I pseudogene. *Cancer Res* 1992;**52**:4280–5.
77. Gray TA, Wilson A, Fortin PJ, *et al*. 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* 2006;**103**:12039–44.
78. Muro EM, Mah N, Andrade-Navarro MA. Functional evidence of post-transcriptional regulation by pseudogenes. *Biochimie* 2011;**93**:1916–21.
79. Lin H, Shabbir A, Molnar M, *et al*. Stem cell regulatory function mediated by expression of a novel mouse Oct4 pseudogene. *Biochem Biophys Res Commun* 2007;**355**:111–6.
80. Piehler AP, Hellum M, Wenzel JJ, *et al*. The human ABC transporter pseudogene family: Evidence for transcription and gene-pseudogene interference. *BMC Genomics* 2008;**9**:165.

Downloaded from https://academic.oup.com/bfg/article/12/6/536/238174 by guest on 20 August 2022

81. Chen JM, Cooper DN, Chuzhanova N, *et al*. Gene conversion: mechanisms, evolution and human disease. *Nat Rev Genet* 2007;**8**:762–75.
82. Hayakawa T, Angata T, Lewis AL, *et al*. A human-specific gene in microglia. *Science* 2005;**309**:1693.
83. Muro EM, Andrade-Navarro MA. Pseudogenes as an alternative source of natural antisense transcripts. *BMC Evol Biol* 2010;**10**:338.
84. Duret L, Chureau C, Samain S, *et al*. The Xist RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. *Science* 2006;**312**:1653–5.
85. Sasidharan R, Gerstein M. Genomics: protein fossils live on as RNA. *Nature* 2008;**453**:729–31.
86. Poliseno L, Salmena L, Zhang J, *et al*. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* 2010;**465**:1033–8.
87. Zou M, Baitei EY, Alzahrani AS, *et al*. Oncogenic activation of MAP kinase by BRAF pseudogene in thyroid tumors. *Neoplasia* 2009;**11**:57–65.
88. Gupta PK, Adamtziki E, Budde U, *et al*. Gene conversions are a common cause of von Willebrand disease. *Br J Haematol* 2005;**130**:752–8.
89. Bunge S, Rathmann M, Steglich C, *et al*. Homologous nonallelic recombinations between the iduronate-sulfatase gene and pseudogene cause various intragenic deletions and inversions in patients with mucopolysaccharidosis type II. *Eur J Hum Genet* 1998;**6**: 492–500.
90. Cai L, Lumsden A, Guenther UP, *et al*. A novel Q378X mutation exists in the transmembrane transporter protein ABCC6 and its pseudogene: implications for mutation analysis in pseudoxanthoma elasticum. *J Mol Med* 2001;**79**: 536–46.
91. De Marco P, Moroni A, Merello E, *et al*. Folate pathway gene alterations in patients with neural tube defects. *Am J Med Genet* 2000;**95**:216–23.
92. Lee H, Niu D, Lin R, *et al*. Structural analysis of the chimeric CYP21P/CYP21 gene in steroid 21-hydroxylase deficiency. *J Hum Genet* 2002;**47**:517–22.
93. Martinez-Arias R, Calafell F, Mateu E, *et al*. Sequence variability of a human pseudogene. *Genome Res* 2001;**11**:1071–85.
94. Minegishi Y, Coustan-Smith E, Wang Y, *et al*. Mutations in the human lambda 5/14.1 gene result in B cell deficiency and agammaglobulinemia. *J Exper Med* 1998;**187**: 71–7.
95. Nakashima E, Mabuchi A, Makita Y, *et al*. Novel SBDS mutations caused by gene conversion in Japanese patients with Shwachman-Diamond syndrome. *Hum Genet* 2004; **114**:345–8.
96. Roesler J, Curnutte JT, Rae J, *et al*. Recombination events between the p47-phox gene and its highly homologous pseudogenes are the main cause of autosomal recessive chronic granulomatous disease. *Blood* 2000;**95**:2150–6.
97. Stone DL, Tayebi N, Orvisky E, *et al*. Glucocerebrosidase gene mutations in patients with type 2 Gaucher disease. *Hum Mutat* 2000;**15**:181–8.
98. Surdhar GK, Enayat MS, Lawson S, *et al*. Homozygous gene conversion in von Willebrand factor gene as a cause of type 3 von Willebrand disease and predisposition to inhibitor development. *Blood* 2001;**98**:248–50.
99. Vanita, Sarhadi V, Reis A, *et al*. A unique form of autosomal dominant cataract explained by gene conversion between beta-crystallin B2 and its pseudogene. *J Med Genet* 2001;**38**:392–6.
100. Watnick T, Gandolph M, Weber H, *et al*. Gene conversion is a likely cause of mutation in PKD1. *Hum Mol Genet* 1998;**7**:1239–43.
101. Sen K, Ghosh TC. Evolutionary conservation and disease gene association of the human genes composing pseudogenes. *Gene* 2012;**501**:164–70.
102. Ebisuya M, Yamamoto T, Nakajima M, *et al*. Ripples from neighbouring transcription. *Nat Cell Biol* 2008;**10**:1106–13.
103. Khachane AN, Harrison PM. Assessing the genomic evidence for conserved transcribed pseudogenes under selection. *BMC Genomics* 2009;**10**:435.
104. Zheng D, Gerstein M. The ambiguous boundary between genes and pseudogenes: the dead rise up, or do they? *Trends Genet* 2007;**23**:219–24.

Downloaded from https://academic.oup.com/bfg/article/12/6/536/238174 by guest on 20 August 2022