Prospects & Overviews

Processed pseudogenes: A substrate for evolutionary innovation

Retrotransposition contributes to genome evolution by propagating pseudogene sequences with rich regulatory potential throughout the genome

Robin-Lee Troskie1 | Geoffrey J. Faulkner1,2 | Seth W. Cheetham1

1 Mater Research Institute, University of Queensland, Woolloongabba, Australia
2 Queensland Brain Institute, University of Queensland, Brisbane, Australia

Abstract

Processed pseudogenes may serve as a genetic reservoir for evolutionary innovation. Here, we argue that through the activity of long interspersed element-1 retrotransposons, processed pseudogenes disperse coding and noncoding sequences rich with regulatory potential throughout the human genome. While these sequences may appear to be non-functional, a lack of contemporary function does not prohibit future development of biological activity. Here, we discuss the dynamic evolution of certain processed pseudogenes into coding and noncoding genes and regulatory elements, and their implication in wide-ranging biological and pathological processes. Also see the video abstract here: https://youtu.be/iUY_mteVoPI

KEYWORDS

evolution, lncRNA, neofunctionalisation, pseudogene, retrotransposition

INTRODUCTION

The human genome is an evolutionary playground for repurposing ancestral coding sequences, facilitating the birth of novel genes and regulatory elements. The diversification of ancestral genes by accumulated mutations is a well-studied mechanism of genetic evolution that can drive speciation. Adaptive evolutionary innovations can also arise through the emergence of novel genes leading to species-specific phenotypic changes. Most novel genes arise through DNA- and RNA-based gene duplication of ancestral sequences.

RNA-based gene duplication is mediated by retrotransposition. In humans, long interspersed element-1 (LINE-1 or L1) retrotransposons can ‘copy and paste’ genes into new genomic locations via an mRNA intermediate. These retroposed gene copies are often termed processed pseudogenes and are largely presumed to be ‘dead on arrival’ due the loss of flanking regulatory elements and introns, or the rapid accumulation of disruptive mutations. However, a current state of non-functionality does not imply that a pseudogene never was nor will be biologically relevant. Evolution is a dynamic continuum that constantly re-arranges the building blocks of life, leaving space for novel combinations to come together and break apart. The birth and death of genes plays a clear role in this process, yet the contribution of pseudogenes to evolution remains a controversial topic. Do these ‘defective’ gene copies merely drift through evolutionary time, accumulating mutations until they are unrecognisable? Or do some represent an opportunity for evolutionary innovations to take place by providing the raw material to engineer novel genes or regulatory regions under positive selective pressure?

The implication of pseudogenes in biological processes including neurogenesis, inflammatory responses and cancer necessitates revisitation of the notion that pseudogenes are evolutionary ‘junk’. However, the extent of pseudogene activity remains poorly investigated, in part perhaps due the bias inherent to the term ‘pseudogene’ which presumes non-functionality. Furthermore, technical shortcomings have impeded unambiguous distinction of pseudogene activity from their near identical parental counterparts. Here, we propose that evolutionary innovations are
facilitated by retrotransposons that in rare instances propagate gene copies throughout the genome. Some of these gene copies, known as processed pseudogenes, may act as a genetic repository of raw material for evolution, giving rise to novel genes and regulatory elements that shape the human genome.

RETROTRANSPOSITION DRIVES GENOME EVOLUTION

The notion that novel genes emerge through the repurposing of ancestral genes dates to the early 1930s[15,16] and persisted through the genomics revolution, influenced by Susumu Ohno’s essay that proposed gene duplication as the major driving force of evolution.[17] There are several key features of Ohno’s model: firstly, gene duplication produces redundant gene copies that accumulate mutations that act as raw material for evolutionary innovation (neofunctionalisation), while ancestral genes continue to fulfil original functions. Secondly, duplicated genes serve as a genetic surplus for ancestral genes that are silenced during dosage compensation. Thirdly, the most probable fate of a duplicated gene is degeneracy.[13]

Duplicated genes can arise through segmental duplication or through the activity of retrotransposons.[2] Retrotransposons are a class of transposable elements that contribute to the instability and evolution of the human genome.[18] Once referred to as ‘selfish entities’, retrotransposons copy genetic information through an mRNA intermediate by evading host genome defences.[19] Over one third of the human genome is composed of retrotransposons,[20] raising the question as to whether they contribute to the functional evolution of the human genome.

L1 retrotransposition is capable of genome-wide mutagenesis, through cis-preferential reverse transcription of L1 mRNA into single-stranded DNA, followed by second strand synthesis (double-stranded DNA) and integration into a new genomic location.[21,22] Full-length L1 transcripts span ~6 kb and consist of a 5′ untranslated region (UTR) with an internal RNA polymerase II promoter,[23] two non-overlapping open reading frames (ORF1 and ORF2) and a polyadenylated (polyA) 3′ UTR.[24] The ORFs encode the molecular machinery necessary for mobilisation and integration of an L1 transcript into a new genomic location.[21] Over 500,000 L1 copies are annotated in the human reference genome[20] although the majority are inactivated by truncations or disruptive mutations, precluding the ability to mobilise.[20,25] Nonetheless, approximately 100 L1s retain intact ORFs and are retrotransposition-competent.[26,27] These active elements are engaged in an evolutionary arms race against host defences to infiltrate the human genome.[28,29]

Past and ongoing L1 activity has generated profound changes to the human genome (reviewed in ref. [18]). For example, L1 retrotransposition can directly alter the genomic landscape by generating instability via insertional mutagenesis, producing heritable insertions when mobilised in the committed germline or during early embryonic development. High L1 copy number and sequence similarities can also indirectly spur genomic rearrangements through recombination-associated deletions. In contrast to L1s, higher processed pseudogene density is inversely correlated with recombination rate.[30] The impact of L1 on human genome evolution is not limited to mobilisation of the L1 RNA; L1s can trans-mobilise other polyA RNAs in rare cases, including those produced by the nonautonomous retrotransposons Alu[31] and SVA.[32] L1 retrotransposons can also mobilise processed protein-coding mRNAs[5] (Figure 1). Transcripts in the cellular vicinity of the L1-encoded molecular machinery are ‘hijacked’, reverse transcribed and re-integrated into the genome, dispersing intronless gene copies sometimes stripped of parental cis-regulatory elements. These stripped-down gene copies are called ‘processed pseudogenes’[8] and are considered constituents of ‘junk DNA’.

Seven years after Ohno’s seminal publication on gene birth through duplication, the term ‘pseudogene’ was officially coined by Jacq and co-workers to describe transcriptionally silent, tandemly repeated truncated copies of the 5S ribosomal RNA gene in Xenopus laevis.[33] Indeed, segmental duplication or erroneous non-equal cross-over results in widespread formation of duplicated pseudogenes that retain parent

![Figure 1 L1-mediated generation of processed pseudogenes.](image-url)
gene intron-exon structures and regulatory elements.[7] A rarer class of unitary pseudogenes are generated de novo by accumulation of disabling mutations in a protein-coding gene rendering it transcriptionally and translationally silent.[34] Following Ohno’s assumption that most duplicated genes are destined for degeneracy, Jacq and co-workers concluded that the identified 5S ribosomal RNA ‘pseudo’ genes were artefacts of evolution. These remarks provided the foundation for a framework that categorises apparently defective sequences with similarity to another gene as pseudogenes. The genomics revolution subsequently saw regions of the genome with pseudogene hallmarks labelled functionless en mass. Although pseudogenes are almost as numerous as protein-coding genes (14,767, of which, 72% are processed, 24% duplicated, 1.6% duplicated, 2.4% other, and 19,957, respectively) (Gencode v38[35]), comparatively little is known about the contribution of pseudogenes to the evolution of the human genome.

RESURRECTED PSEUODGENES EMERGE THROUGH ACQUISITION OF NOVEL PROMOTERS AND REGULATORY ELEMENTS

The prevailing view of pseudogenes has long been that of defective gene copies undergoing evolutionary decay. Indeed, most processed pseudogenes are incapable of producing mRNA due to loss of parental promoters and regulatory elements as well as acquisition of disabling mutations.[8] Nonetheless, evidence of pseudogene transcription in tissue- and cancer-specific patterns emerged following the advent of high throughput, short and long-read RNA-sequencing (RNA-seq).[36–39] How do pseudogenes devoid of the regulatory machinery necessary for transcription become expressed? Various studies have identified sources of novel regulatory elements that facilitate the birth of functional pseudogenes (retrogenes).[40–43]

Processed pseudogenes may preferentially integrate into genomic regions with open-chromatin and actively expressed genes[44,45] and are transcribed by ‘hitchhiking’ on pre-existing regulatory machinery.[45,46] For example, pseudogenes may integrate within an intron or exon and become transcribed from a ‘host gene’s’ core promoter, generating a fusion transcript of pseudogene and host gene exons[45,47–51] (Figure 2A). Alternatively, bidirectional promoters can facilitate the transcription of nearby pseudogenes[42,45] (Figure 2B). Open chromatin can permit interactions between distal enhancers and regulatory elements associated with neighbouring genes, strengthening pseudogene transcription[45] (Figure 2C). Furthermore, promoters from retrotransposons immediately upstream of pseudogene integration sites can regulate pseudogene expression[39,42,45] (Figure 2D). For example, PDCL3P4, a long noncoding pseudogene identified by Pacific Biosciences (PacBio) long-read complementary DNA (cDNA) sequencing, is expressed from the promoter of an upstream human endogenous retrovirus-K[49] in an acute myeloid leukaemia cell line. In addition to benefiting from pre-existing regulatory elements, pseudogenes can acquire novel promoters from GpG-rich regions containing proto-promoter sequences with the inherent potential to facilitate transcription[42,52,53] (Figure 2E). Certain retrocopy promoters may have evolved through the repurposing of regulatory elements,[54] and promoters and enhancers that share similar architectures and functional features[55,56] can lead to enhancer or enhancer-like elements being co-opted as promoters[46] (Figure 2F). Indeed, putative mouse enhancers were found to be orthologous to the promoters of several rat-specific retrocopies suggesting that regulatory element conversion is a mechanism through which pseudogenes can become transcriptionally active.[46]

The recruitment of novel regulatory elements enables pseudogenes to evolve distinct expression patterns and functions compared to their parent genes. Notably, many pseudogenes show highly specific expression patterns, suggestive of coordinated regulation.[36,38,39] For example, RNA-seq transcriptome profiling encompassing 293 human samples (13 tissue types, of which, 248 samples were cancerous and 45 benign) showed strong evidence of widespread pseudogene transcription with the identification of 2082 distinct pseudogenes exhibiting lineage-specific, cancer-specific and ubiquitous expression patterns.[36,38,39] Of note, pseudogene transcription is not indicative of function; only a minority of pseudogenes have identified functions (discussed below) while the majority are not currently functional or have functions yet to be elucidated.

PROCESSED PSEUDOGENES CAN EXPAND THE GENETIC REPERTOIRE THROUGH NOVEL GENE FUSIONS AND EXON GAIN

New genes can form when pseudogenes integrate into unrelated host genes. These insertions result in fusion transcripts containing sequences derived from both the pseudogene and host gene that may have functions distinct from those of the original host or parent gene. In rare instances that these novel sequences provide an adaptive evolutionary benefit, they can be selectively preserved. A classic example and the first discovered ‘young’ chimeric pseudogene fusion formed in a common ancestor of two African Drosophila species around 2.5 M years ago.[47] The chimeric gene jgw arose through a series of evolutionary events where a retroposed copy of Adh inserted into an intron downstream of the 5′ regulatory region and several exons of a segmental gene duplicate, ynd, producing a novel coding fusion transcript. Under the influence of positive selection, the novel testis-expressed gene evolved a functional role in hormone and pheromone metabolism.[57] Other retrogene fusions have been identified including Sphinxa,[48] a Drosophila noncoding RNA-gene fusion implicated in male-male courtship behaviour, as well as young coding fusions that emerged in the primate lineage.[45] Notably, retrotransposition of CypA into the coding sequence of TRIMZ, a restriction factor, produced the functionally-important gene fusion TRIM-CypA that confers retroviral resistance.[58–60] Strikingly, the chimeric gene arose independently in both New and Old World monkeys through convergent evolution. Recently, 71 human-specific transcripts containing exapted sequences of 56 retrocopies were identified.[61] Retrocopy insertion events can have profound structural implications by generating novel proteins from alternative splicing.[39] For example, a splice variant of BRCA1
FIGURE 2 Recruitment of regulatory elements. Pseudogenes can become expressed through various mechanisms including (A) retrotransposing into a host gene’s intron and being incorporated into a fusion transcript, (B) inserting near a gene with a bi-directional promoter, (C) inserting into open, actively transcribed chromatin domains that may facilitate interactions of distant enhancers and regulatory elements of neighbouring genes with nearby pseudogenes, (D) upstream promoters embedded in retrotransposons, (E) sequences containing CpG-rich regions with proto-promoter properties or single nucleotide substitutions under intense positive selection can evolve into novel promoters and transcribe nearby pseudogenes and (F) enhancers and promoters share similar sequences architectures that may enable the conversion of one to another during evolution.

contains a portion of DPH3P1 that encodes a peptide that may modify the trans-activation domain 1 and interaction domain. Fusion gene formation through retrocopy insertions enables shuffling of conserved protein-coding domains between distant genes, providing a mechanism for evolutionary innovation.

Exon gain is associated with evolution of pseudogenes into functional genes. While some processed pseudogenes become multi-exonic through intronisation (acquisition of splice sites within a parent-derived exon), the majority recruit novel exons from upstream or downstream flanking sequences. Notably, multi-exonic pseudogenes with novel 5’ exons are highly overrepresented in several vertebrate species, presumably providing distal promoters. Over 80% of ancient retrogenes have accumulated complex gene structures that have resulted in broad expression patterns in contrast to younger mono-exonic pseudogenes that tend to be testes-specific. The importance of exon gain is exemplified in the mouse retrocopy Rps23r1 that emerged through retroposition of the ribosomal protein S23. Rps23r1 is transcribed from the complementary strand relative to the parent gene and incorporates additional coding sequence from sites flanking the insertion. The structurally distinct yet functional protein confers heightened resistance to the formation of amyloid plaques associated with Alzheimer’s disease, demonstrating the recycling potential of pseudogene sequences to fulfil functions unrelated to their parent genes. Mono- and multi-exonic pseudogenes can also generate transcript isoforms from alternative splice sites and termination start/stop sites as exemplified by the retrogene HNRNP1 that encodes one broadly expressed and one testes-expressed isoform in several species. Thus, alternative splicing of mono- and multi-exonic pseudogenes can enable expansion of genetic repertoires by producing functionally distinct transcript isoforms that may display organ-specific expression patterns.

EXPRESSION PROCESSED PSEUDOGENES CAN ACT THROUGH NONCODING MECHANISMS

Since early anecdotal evidence of functional pseudogenes, it is now clear that many processed pseudogenes have evolved into bona fide
FIGURE 3 Pseudogenes contribute to the regulatory landscape. Processed pseudogenes can function through several RNA- and protein-based mechanisms. (A) A retrotransposed transcript is integrated in reverse orientation relative to the parent gene and expressed by a proximal promoter to generate an antisense pseudogene transcript. Alternatively, a bi-directional promoter in close proximity can generate an antisense pseudogene transcript. (B) Pseudogene asRNA can hybridise with parent gene sRNA forming an RNA-RNA duplex, inhibiting the translation of the parent gene. (C) Pseudogene antisense-mediated gene regulation was first identified in neurons from the Lymnaea stagnalis snail. High sequence similarity is the regulation of parent gene expression due to high sequence similarity. For example, integration of processed pseudogenes into the genome in a reverse orientation relative to an adjacent promoter produces antisense RNA (asRNA) (Figure 3A). Highly complementary regions of pseudogene asRNA can hybridise with parental sense RNA and form a stable double-stranded RNA (dsRNA) duplex, inhibiting protein synthesis from the parent gene (Figure 3B). Pseudogene antisense-mediated gene regulation was first identified in neurons from the Lymnaea stagnalis snail. Co-expression of RNA from nNOS and antisense mRNA from the corresponding pseudogene leads to the formation of a stable dsRNA duplex, suppressing nNOS translation. Similarly, in humans, the pseudogene FLT1P1 produces sense and antisense transcripts that reduce the levels of the parent gene, VEGFR1, decreasing human colorectal tumour cell proliferation and xenograft tumour growth. Pseudogene asRNA transcripts can also regulate epigenetic processes (Figure 3C). For example, PTEN is regulated by the presence of asRNA from its pseudogene, PTENP1 that localises to the PTEN promoter and induces transcriptional silencing via recruitment of the H3K27 methyl-transferase EZH2 and DNA methylase DNMT3A. Similarly, OCT4, a master regulator of pluripotency has six related pseudogenes, of which, OCT4-pg5 generates an asRNA capable of binding to the OCT4 locus and inducing epigenetic silencing.

RNA-RNA duplexes consisting of pseudogene-pseudogene or pseudogene-parent gene transcripts can process into endogenous
small interfering RNAs in mouse oocytes and downregulate parent gene expression levels through RNA interference (RNAi)\(^{[73,74]}\) (Figure 3D). RNAi is a crucial biological process responsible for the suppression of gene expression by post-transcriptional targeting of mRNA with complementary siRNA molecules that direct targeted cleavage.\(^{[75]}\) Pairing of sense and antisense pseudogene transcripts of Hdac1 induces cleavage by Dicer,\(^{[76]}\) an enzyme that recognises and cleaves dsRNA, generating siRNAs complementary to the parent gene transcript. siRNAs incorporated into an RNA-induced silencing complex (RISC) localise to complementary regions on parent gene mRNA and induce cleavage by argonaute 2, the catalytic component of the RISC.\(^{[72]}\) Similarly, the pseudogene Ppp4r1 generates an anti-sense transcript that hybridises to its parent gene’s complementary transcript and downregulates its expression by RNAi.\(^{[73,74]}\) Processing of internal secondary structures or hairpin loops formed by pairing of homologous regions within a single pseudogene transcript can also elicit an inhibitory effect on parent genes including the Au76 pseudogene and its parent gene, Rangap1.\(^{[74]}\)

Pseudogene-derived siRNAs can also act as tumour-suppressors. For example, a pseudogene of PPM11k, \(\psi\)PPM1K, produces siRNAs that inhibit cell growth in hepatocellular carcinoma (HCC).\(^{[77]}\) Inverted repeats within the pseudogene RNA form a hairpin that is cleaved by Dicer, generating two siRNAs that bind and downregulate both parent gene and NEK8, a target gene of PPM1K that promotes cellular proliferation and is overexpressed in HCC.

Pseudogene transcripts can also increase expression of their parent genes (Figure 3E). PTENP1, KRASP1 and BRAFP1 enhance expression of their parent genes by acting as molecular sponges that sequester microRNAs (miRNAs) via shared binding sites.\(^{[78,79]}\) The concept of competitive endogenous RNAs (ceRNAs) explains upregulation of PTEN, an intensively studied tumour suppressor gene that is frequently mutated in cancer, when its highly homologous pseudogene, PTENP1, is expressed.\(^{[78]}\) In a pseudogene context, the ceRNA hypothesis postulates that shared miRNA response elements (MREs) within the high homology region of the 3′UTR of both pseudogene and parent gene transcripts results in competition for parent gene-targeting miRNAs. Increasing PTENP1 levels depletes the pool of PTEN-targeting miRNA molecules, relieving PTEN inhibition. PTENP1 acts as tumour suppressor in cancer by increasing PTEN levels and consequently, the PTENP1 locus is frequently deleted in cancers. Oncogenic KRAS was similarly upregulated when the MRE-containing 3′UTR of its pseudogene, KRASP1, was overexpressed in vitro.\(^{[76]}\)

The first causal role of pseudogene ceRNA in cancer formation was identified in vivo when Braf-rs-1 overexpression in transgenic mice induced an aggressive malignancy resembling human diffuse large B cell lymphoma.\(^{[79]}\) Braf-rs-1 and its human ortholog, BRAFP1, elicit oncogenic activity as ceRNAs by elevating BRAF expression and MAPK activation, promoting cell proliferation, differentiation and migration. However, there is increasing scepticism about the impact of pseudogene-miRNA interactions as most evidence for the ceRNA hypothesis relies on non-physiological levels of pseudogene expression.\(^{[80]}\) Indeed, normal physiological changes in usually low pseudogene transcript levels do not sufficiently diminish the pool of parent gene-targeting miRNA to impact the often higher parent gene expression levels.\(^{[81]}\) ceRNA levels must approach target abundance of parent gene-targeting miRNA to have a de-suppressive effect, making pseudogenes unlikely candidates as ceRNAs as they generally have considerably lower expression levels than their parent genes. Caution must be exercised when generalised theories of gene-regulatory networks are applied to large, heterogeneous bodies of noncoding RNAs that are not yet functionally characterised to simplify complex genomic interactions.

**EXPRESSION PROCESSED PSEUODGENES CAN IMPACT THE GENOME THROUGH CODING-DEPENDENT MECHANISMS**

While the definition of pseudogene is not compatible with protein-coding capacity, some processed pseudogenes have undisrupted ORFs that are translated (Figure 3F). PGK2 was the first reported retro-transposed gene-copy capable of producing a biologically significant protein.\(^{[65]}\) PGK2 exhibits hallmarks of a processed pseudogene (no introns, genomic polyA tract) yet contains a complete ORF and is translated in the human testes to compensate for inactivation of its X-linked counterpart, PGK1. Mass-spectrometry-based analyses of the human proteome revealed 140 pseudogenes that are capable of producing over 200 peptides\(^{[82]}\) and, more recently, over 700 retrogenes were found to produce uniquely matching peptides.\(^{[161]}\) Furthermore, estimates based on ribosomal profiling that utilises a percentage of maximum entropy (PME) approach to measure read distribution uniformity and distinguish coding from noncoding RNAs, suggests 40% of annotated pseudogenes are translated.\(^{[183]}\) Long-read cDNA sequencing of mixed human tissues and cell lines identified 160/318 full-length pseudogene transcripts (50%) that encode ORFs over 100 amino acids in length.\(^{[39]}\) Notably, 53 pseudogenes contained ORFs over 90% the length of their parent genes, highlighting the potential of translated pseudogenes to produce intact proteins.

Recently, a systemic analysis of CRISPR-Cas9-induced frameshift mutations in protein-coding genes revealed that protein production was often not completely ablated.\(^{[84]}\) The disruption of 136 distinct genes by frame-shift mutations resulted in residual protein expression for one third of gene targets (ranging from low to wild-type levels). Two causal mechanisms explain continued protein production: translational re-initiation downstream of a mutated exon producing N-terminally truncated proteins and skipping of the mutated exon during splicing. The same phenomena may apply to pseudogenes with disrupted ORFs, revealing why numerous pseudogenes retain the ability to translate into protein. While the functions of most translated pseudogenes remains to be elucidated, several sporadically characterised pseudogenes encode functional proteins that play important roles in tumorigenesis\(^{[12,85,86]}\) and have been re-annotated as novel protein-coding genes.\(^{[87]}\)

Newly established protein-coding pseudogenes can evolve novel functions through the relocation of encoded proteins to novel cellular niches and perform compartment-specific functions under the
influence of natural selection. The process, known as subcellular adaptation, is reflected in the functional adaptation of a hominoid-specific retrogene, CDC14Bretro. The retrogene emerged as a splice variant of CDC14B and encodes a protein that became expressed in the brain and testes, playing a role in stabilising microtubules. After a period of intense positive selection, CDC14Bretro completely relocalised and began to function within the endoplasmic reticulum. The functional diversification of GLUD2, a retrogene derived from GLUD1, occurred when it underwent subcellular relocalisation in a common ancestor of humans and apes 18–25 million years ago. Under positive selection, two key amino acid substitutions induced biochemical alterations in the encoded protein causing sublocalisation from both the cytoplasm and mitochondrial compartments to just the mitochondria where it began to target the neurotransmitter glutamate in the brain for degradation. Thus, numerous processed pseudogenes have contributed to genome evolution, serving as ncRNAs or producing functional proteins with biologically important roles.

PSEUDOGENES CAN EVOLVE PROTO-PROMOTER ACTIVITY AND INFLUENCE PROXIMAL GENES

Processed pseudogenes are a rich source of genetic information that can provide raw material for the evolution of novel regulatory regions. We argue that retroposed transcripts originating from protein-coding genes contain both high GC content and various transcription factor-binding motifs (TFBMs) that provide a favourable environment to evolve enhancer-like or proto-promoter properties. For example, the presence of TFBMs in young processed pseudogenes could increase their propensity to evolve into enhancer-like elements and impact neighbouring genes. In the absence of selective pressure for their retention, TFBMs degenerate over evolutionary time. Under selection, processed pseudogenes could evolve promoter activity and profoundly impact the coding sequences of nearby genes or host genes. For example, some pseudogene transcripts are embedded with alternative downstream transcriptional start sites inherited from parental transcripts that can evolve promoter activity.

These pseudogenes can serve as alternative promoters for adjacent or host genes, contributing coding sequence and generating novel splice isoforms. For example, the imprinted tumour suppressor gene RB1 harbours PPP1R26P1, a 5′-truncated retrocopy of PPP1R26. The retrotransposition event occurred before the split of New and Old World monkeys and resulted in the integration of PPP1R26P1 into intron 2 of RB1, in reverse orientation relative to the host gene. The region of PPP1R26P1 derived from exon 4 of the parent gene contains a differentially methylated CpG island that evolved promotor activity. Furthermore, the CpG island harbours an exon that is spliced into exon 3 of RB1 generating an imprinted pseudogene transcript. The contribution of pseudogene sequences to protein-coding transcripts is widespread. PacBio long-read cDNA sequencing of mixed human tissues reveals that the retrocopy constitutes the majority of the 5′ exon of RB1 and can contribute 179 codons to the fusion sequence. cDNA sequencing also identifies 93 protein-coding genes that contain coding sequence derived from pseudogenes.

Semi-processed pseudogenes present an interesting mechanism through which pseudogenes could impact proximal genes. Retrotransposition of partially spliced transcripts can generate semi-processed pseudogenes that retain one or more parent gene introns. Parentally-derived enhancer-like or regulatory elements embedded within unspliced introns could provide an avenue for semi-processed pseudogenes to become transcribed or influence nearby genes. Thus, some processed pseudogenes that integrate within exons of host genes can evolve promoter properties that can influence the expression of novel fusion transcripts, or provide alternative regulatory elements embedded within unspliced introns, highlighting the potential of pseudogenes to disperse information-rich sequences throughout the genome and impact surrounding genes.

PROCESSED PSEUDOGENES ARE A GENETIC RESERVOIR FOR EVOLUTIONARY INNOVATION

While most pseudogenes appear to pose no contemporary biological benefit, a current state of non-functionality does not necessarily imply that a pseudogene never was nor never will be biologically relevant. The tens of thousands of processed pseudogenes in the human genome that appear to be ‘inert’ may be undergoing subtle changes that slowly impact the functional landscape of the genome. Indeed, pseudogenes are not strongly conserved across non-primate mammals suggesting that many pseudogenes are non-adaptive and rather exaptive. In other words, a pseudogene may not initially arise as a consequence of adaptive selection but rather exists as a substrate for selection to act upon. This is a well-known evolutionary mechanism for the birth and death of genes that generates substrates for adaptive selection and evolution to occur. This is exemplified in the recent discovery of a potent inhibitor of enveloped virus budding, retroCHMP3. The retrogene arose through independent retrotransposition events in New World monkeys and mice, producing variants that evolved the same antiviral function. retroCHMP originated from the retroduplication of CHMP3, an ESCRT-III protein involved in cellular membrane remodelling events. The ESCRT pathway is frequently exploited by enveloped viruses to enable budding from the cellular membrane. While inhibition of the ESCRT pathway prevents viral replication, loss-of-function induces cytotoxicity due to the essential nature of the pathway. Remarkably, retroCHMP3 evolved an exquisitely balanced functional role as a potent antiviral factor by inhibiting the ESCRT pathway during infection while causing minimal cytotoxicity. Interestingly, the evolutionary pathways of retroCHMP3 emergence display species- and lineage-specific diversity producing a variety of full-length, truncated and degraded copies that present differing levels of antiviral activity and cytotoxicity. Full-length copies subjected to C-terminus-truncating mutations display enhanced inhibitory activity of the ESCRT pathway with minimal cytotoxicity as seen in squirrel monkeys and mice. Many primates surveyed contained full-length retroCHMP3, concomitant with recent duplication.
or long-term selection, indicating that the retrocopy is one truncating mutation away from becoming a potent antiviral defence mechanism. This demonstrates the ability for multiple evolutionary trajectories to lead to the repeated emergence of retrogenes that act as a genetic reservoir for the evolution of common immune defences across multiple species. The existence of functional truncated pseudogenes also highlights the lack of empirical basis for computationally distinguishing inert pseudogenes from functional genes.

With numerous reported cases of functional pseudogenes, the distinction between a gene and pseudogene is blurred. What do functional pseudogenes lack that precludes them being perceived as a conventional gene? Furthermore, what constitutes a gene? The precise definition is highly contested and has continued to evolve since the outdated and impractical ‘one gene – one protein’ convention. As our understanding of the molecular basis of our biology has expanded so has our view of what constitutes a gene (for review, see ref. [101]). For example, we now know that a large fraction of genes are noncoding and govern vital biological processes through RNA. Additionally, a single gene locus may contain differing transcriptional start and stop sites and undergo alternative splicing, producing various transcripts that give rise to proteins with profoundly different structures and functions. Furthermore, many genes overlap on the same or opposite strands or reside within introns and produce fusion transcripts that increase the repertoire of genetic diversity. Nonetheless, a gene-centric view of molecular biology may be detrimental as it acts as an oversimplification for complex biological processes and blurs potential lines of enquiry. This generalisation is mirrored in pseudogenes that are understudied and underappreciated perhaps not only due to the technical challenge of distinguishing pseudogenes from parent genes, but also due to the bias implied by the term ‘pseudogene’, which immediately assumes non-functionality.

As a result, the growing body of pseudogenes with demonstrated functions are cast aside from those of conventional genes and continue to be thought of as biologically inconsequential artefacts. Further investigation of coding and noncoding pseudogenes will likely greatly expand the pseudogene functional repertoire, highlight their contribution to the evolution of the human genome and alleviate bias carried by the term ‘pseudogene’.

CONCLUSION

Here, we have argued that processed pseudogenes contribute functional elements to genomes through diverse mechanisms, including as gene-regulatory elements, novel protein-coding genes and noncoding RNAs. Thus, whilst retrotransposition has generally been regarded as deleterious to the host genome, its role in mobilising mRNAs intoRNAs. Thus, whilst retrotransposition has generally been regarded as deleterious to the host genome, its role in mobilising mRNAs ingRNAs. Thus, whilst retrotransposition has generally been regarded as deleterious to the host genome, its role in mobilising mRNAs intoRNAs. Thus, whilst retrotransposition has generally been regarded as deleterious to the host genome, its role in mobilising mRNAs into.

ACKNOWLEDGEMENTS

The authors thank N. Jansz and N. Smits for helpful discussions. Funding for this study was provided by the Australian National Health and Medical Research Council (NHMRC) (GNT1173711 to G.J.F. and GNT1161832 to S.W.C.), a CSL Centenary Fellowship to G.J.F., an Australian Government Research Training Program (RTP) Scholarship and Mater Research Frank Clair Scholarship to R-L.T. and by the Mater Foundation (Equity Trustees/AE Hingeley Trust).

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

AUTHOR CONTRIBUTIONS

Robin-Lee Troskie and Seth W. Cheetham contributed to all aspects of the article. Geoffrey J. Faulkner revised the manuscript.

DATA AVAILABILITY STATEMENT

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

REFERENCES

1. Necsulea, A., & Kaessmann, H. (2014). Evolutionary dynamics of coding and non-coding transcriptomes. Nature Reviews Genetics, 15(11), 734–748.
2. Long, M., Betrán, E., Thornton, K., & Wang, W. (2003). The origin of new genes: Glimpses from the young and old. Nature Reviews Genetics, 4, 865–875.
3. Kaessmann, H. (2010). Origins, evolution, and phenotypic impact of new genes. Genome Research, 20(10), 1313–1326.
4. Kaessmann, H., Vinckenbosch, N., & Long, M. (2009). RNA-based gene duplication: Mechanistic and evolutionary insights. Nature Reviews Genetics, 10(1), 19–31.
5. Esnault, C., Maestre, J., & Heidmann, T. (2000). Human LINE retrotransposons generate processed pseudogenes. Nature Genetics, 24(4), 363–367.
6. Wei, W., Gilbert, N., Ooi, S. L., Lawler, J. F., Ostertag, E. M., Kazazian, H. H., Boeke, J. D., & Moran, J. V. (2001). Human L1 retrotransposition: Cis preference versus trans complementation. Molecular and Cellular Biology, 21(4), 1429–1439.
7. Mighell, A. J., Smith, N. R., Robinson, P. A., & Markham, A. F. (2000). Vertebrate pseudogenes. FEBS Letters, 468, 109–114.
8. Vanin, E. F. (1984). Processed pseudogenes. BBA-Gene. Struct. Expr., 782(3), 231–241.
9. Suzuki, I. K., Gacquer, D., Van Heurck, R., Kumar, D., Wojno, M., Bilheu, A., Herpoel, A., Lambert, N., Cheron, J., Polleux, F., Detours, V., & Vanderhaeghen, P. (2018). Human-specific NOTCH2NL genes expand cortical neurogenesis through delta/notch regulation. Cell, 173(6), 1370–1384.e16.
10. Fiddes, I. T., Lodewijk, G. A., Mooring, M., Bosworth, C. M., Ewing, A. D., Mantalas, G. L., Novak, A. M., van den Bout, A., Bishara, A., Rosenkranz, J. L., Lorig-Roach, R., Field, A. R., Haeussler, M., Russo, L., Bhaduri, A., Nowakowski, T. J., Pollen, A. A., Dougherty, M. L., Nuttle, X.,... Haussler, D. (2018). Human-specific NOTCH2NL genes affect notch signaling and cortical neurogenesis. Cell, 173(6), 1356–1369.e22.
11. Rapicavoli, N. A., Qu, K., Zhang, J., Mikhail, M., Laberge, R. M., & Chang, H. Y. (2013). A mammalian pseudogene IncRNA at the interface of inflammation and antiinflammatory therapeutics. eLife, 2013(2), 1–16.
12. Hayashi, H., Arao, T., Togashi, Y., Kato, H., Fujita, Y., De Velasco, M. A., Kimura, H., Matsumoto, K., Tanaka, K., Okamoto, I., Ito, A., Yamada, Y,
Nakagawa, K., & Nishio, K. (2015). The OCT4 pseudogene POUSF18 is amplified and promotes an aggressive phenotype in gastric cancer. Oncogene, 34(2), 199–208.

13. Ohno, S. (1972). So much “junk” DNA in our genome. Brookhaven Symposia in Biology, 23, 366–370.

14. Cheetham, S. W., Faulkner, G. J., & Dinger, M. E. (2020). Overcoming challenges and dogmas to understand the functions of pseudogenes. Nature Reviews Genetics, 21, 191–201.

15. Haldane, J. B. S. (1933). The part played by recurrent mutation in evolution. American Naturalist, 67, 708.

16. Muller, H. J. (1935). The origination of chromatin deficiencies as minute deletions subject to insertion elsewhere. Genetica, 17, 237–252.

17. Ohno, S. (1970). Evolution by gene duplication. Berlin: Springer.

18. Cordaux, R., & Batzer, M. A. (2009). The impact of retrotransposons on human genome evolution. Nature Reviews Genetics, 10, 691–703.

19. Goodier, J. L., & Kazazian, H. H., Jr. (2008). Retrotransposons revisited: The restraint and rehabilitation of parasites. Cell, 135(1), 23–35.

20. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., Levine, R., McEwan, P., … Morgan, M. J. (2001). Initial sequencing and analysis of the human genome. Nature, 409(6822), 860–921.

21. Luan, D. D., Korman, M. H., Jakubczak, J. L., & Eickbush, T. H. (1993). Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: A mechanism for non-LTR retrotransposition. Cell, 72(4), 595–605.

22. Kazazian, H. H., Wong, C., Youssoufian, H., Scott, A. F., Phillips, D. G., & Antonarakis, S. E. (1988). Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. Nature, 332, 164–166.

23. Swergold, G. D. (1990). Identification, characterization, and cell specificity of a human LINE-1 promoter. Molecular and Cellular Biology, 10(12), 6718–6729.

24. Babushok, D. V., & Kazazian, H. H. (2007). Progress in understanding the biology of the human mutagen LINE-1. Human Mutation, 28, 527–539.

25. Szak, S. T., Pickeral, O. K., Makalowski, W., Boguski, M. S., Landsman, D., & Boeke, J. D. (2002). Molecular archeology of L1 insertions in the human genome. Genome Biology, 3, research0052.

26. Brouha, B., Schustak, J., Badge, R. M., Lutz-Pfiege, S., Farley, A. H., Morant, J. V., & Kazazian, H. H. (2003). Hot L1s account for the bulk of retrotransposition in the human population. Proceedings of the National Academy of Sciences of the United States of America, 100(9), 5280–5285.

27. Beck, C. R., Collier, P., Macfarlane, C., Malig, M., Kidd, J. M., Eichler, E. E., Badge, R. M., & Moran, J. V. (2010). LINE-1 retrotransposition activity in human genomes. Cell, 141(7), 1159–1170.

28. Imbeault, M., Helleboid, P. Y., & Trono, D. (2017). KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. Nature, 543, 550–554.

29. Jacobs, F. M. J., Greenberg, D., Nguyen, N., Haeussler, M., Ewing, A. D., Katzman, S., Paten, B., Salama, S. R., & Haussler, D. (2014). An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. Nature, 516, 242–245.

30. Liu, G., Li, H., & Cai, L. (2010). Processed pseudogenes are located preferentially in regions of low recombination rates in the human genome. Journal of Evolutionary Biology, 23(5), 1107–1115.

31. Dewannieux, M., Esnault, C., & Heidmann, T. (2003). LINE-mediated retrotransposition of marked Alu sequences. Nature Genetics, 35, 41–48.

32. Ostertag, E. M., Goodler, J. L., Zhang, Y., & Kazazian, H. H. (2003). SVA elements are nonautonomous retrotransposons that cause disease in humans. American Journal of Human Genetics, 73(6), 1444–1451.

33. Jacq, C., Miller, J. R., & Brownlee, G. G. (1977). A pseudogene structure in SS DNA of Xenopus laevis. Cell, 12(1), 109–120.

34. Zhang, Z. D., Frankish, A., Hunt, T., Harrow, J., & Gerstein, M. (2010). Identification and analysis of unitary pseudogenes: Historic and contemporary gene losses in humans and other primates. Genome Biology, 11(3), R26.

35. Frankish, A., Diekhans, M., Ferreira, A.-M., Johnson, R., Jungreis, I., Loveland, J., Mugue, J. M., Sisu, C., Wright, J., Armstrong, J., Barnes, I., Berry, A., Bignell, A., Carbonell Sala, S., Chrast, J., Cunningham, F., Di Domenico, T., Donaldson, S., Fiddes, I. T., … Flicek, P. (2019). GENCODE reference annotation for the human and mouse genomes. Nucleic acids research, 47(D1), D766–D773.

36. Kalyana-Sundaram, S., Kumar-Sinha, C., Shankar, S., Robinson, D. R., Wu, Y. M., Cao, X., Asangani, I. A., Kothari, V., Prensner, J. R., Lonigro, R. J., Iyer, M. K., Barrette, T., Shanmugam, A., Dhanasekaran, S. M., Palanisamy, N., & Chinnaiyan, A. M. (2012). Expressed pseudogenes in the transcriptional landscape of human cancers, Cell, 149(7), 1622–1634.

37. Pei, B., Sisu, C., Frankish, A., Howald, C., Habegger, L., Lu, X. J., Harte, R., Balasubramanian, S., Tanzer, A., Diekhans, M., Raymond, A., Hubbard, T. J., Harrow, J., & Gerstein, M. B. (2012). The GENCODE pseudogene resource. Genome Biology, 13(9), R51.

38. Guo, X., Lin, M., Rockowitz, S., Lachman, H. M., & Zheng, D. (2014). Characterization of human pseudogene-derived non-coding RNAs for functional potential. PLoS One, 9(4), e93972.

39. Troskie, R.-L., Jafary, I., & Cheetham, S. W. (2021). Long-read cDNA sequencing identifies functional pseudogenes in the human transcriptome. Genome Biology, 22, 146.

40. Bai, Y., Casola, C., Feschotte, C., & Betrán, E. (2007). Comparative genomics reveals a constant rate of origin and convergent acquisition of functional retrogenes in Drosophila. Genome Biology, 8, R11.

41. Okamura, K., & Nakai, K. (2008). Retrotransposition as a source of new promoters. Molecular Biology and Evolution, 25(6), 1231–1238.

42. Fablet, M., Bueno, M., Potrzebowski, L., & Kaessmann, H. (2009). Evolutionary origin and functions of retrogene introns. Molecular Biology and Evolution, 26(9), 2147–2156.

43. Sorourian, M., Kunte, M. M., Domingues, S., Gallach, M., Özdíl, F., Rio, J., & Betrán, E. (2014). Relocation facilitates the acquisition of short cis-regulatory regions that drive the expression of retrogenes during spermatogenesis in Drosophila. Molecular Biology and Evolution, 31(8), 2170–2180.

44. Zhang, Y., Li, S., Abyzov, A., & Gerstein, M. B. (2017). Landscape and variation of novel retro duplications in 26 human populations. PLoS Computational Biology, 13(6), e1005567.

45. Vincenboshc, N., Dopianloup, I., & Kaessmann, H. (2006). Evolutionary fate of retroaped gene copies in the human genome. Proceedings of the National Academy of Sciences of the United States of America, 103(9), 3220–3225.

46. Carelli, F. N., Hayakawa, T., Go, Y., Imai, H., Warnefors, M., & Chinnaiyan, A. M. (2012). Expressed pseudogenes and Evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. Nature, 516, 242–245.

47. Liu, G., Li, H., & Cai, L. (2010). Processed pseudogenes are located preferentially in regions of low recombination rates in the human genome. Journal of Evolutionary Biology, 23(5), 1107–1115.

48. Dewannieux, M., Esnault, C., & Heidmann, T. (2003). LINE-mediated retrotransposition of marked Alu sequences. Nature Genetics, 35, 41–48.
Stricker, H., Hwang, C., Chitale, D. A., Williamson, S. R., Gupta, N. S., Navone, N. M., Rogers, C., Menon, M., Varambally, S., & Palanisamy, N. (2019). Pseudogene associated recurrent gene fusion in prostate cancer. Neoplasia, 21(10), 989–1002.

Lai, J., Lehman, M. L., Dinger, M. E., Hendy, S. C., Mercer, T. R., Seim, I., Lawrence, M. G., Mattick, J. S., Clements, J. A., & Nelson, C. S. (2010). A variant of the KLK4 gene is expressed as a cis sense-antisense chimeric transcript in prostate cancer cells. RNA, 16, 1156–1166.

Kundu, T. K., & Rao, M. R. S. (1999). CpG islands in chromatin organization and gene expression. Journal of Biochemistry, 125, 217–222.

Illingworth, R. S., Gruenewald-Schneider, U., Webb, S., Kerr, A. R. W., James, K. D., Turner, D. J., Smith, C., Harrison, D. J., Andrews, R., & Bird, A. P. (2010). Orphan CpG islands identify numerous conserved promoters in the mammalian genome. PLoS Genetics, 6, e1001134.

Carelli, F. N., Liechti, A., Halbert, J., Warnefors, M., & Kaessmann, H. (2018). Repurposing of promoters and enhancers during mammalian evolution. Nature Communications, 9, 4066.

Core, L. J., Martins, A. L., Danko, C. G., Watters, C. T., Siepel, A., & Lis, J. T. (2014). Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. Nature Genetics, 46, 1311–1320.

Andersson, R., Sandelin, A., & Danko, C. G. (2015). A unified architecture of transcriptional regulatory elements. Trends in Genetics, 31, 426–433.

Zhang, Z., Carriero, N., & Gerstein, M. (2004). Comparative analysis of processed pseudogenes in the mouse and human genomes. Trends in Genetics, 20(2), 62–67.

Sayah, D. M., Sokolsskaia, E., Berthoux, L., & Luban, J. (2004). Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1. Nature, 430, 569–573.

Virgen, C. A., Kratovac, Z., Bieniasz, P. D., & Hatzioioannou, T. (2008). Independent genesis of chimeric TRIM5-cyclophilin proteins in two primate species. Proceedings of the National Academy of Sciences of the United States of America, 105, 3563–3568.

Brennan, G., Kozyrev, Y., & Hu, S. L. (2008). TRIMCyp expression in Old World primates Macaca nemestrina and Macaca fascicularis. Proceedings of the National Academy of Sciences of the United States of America, 105, 3569–3574.

Kubiak, M. R., Szczesniak, M. W., & Makalowska, I. (2020). Complex analysis of retrotransposed genes’ contribution to human genome, proteome and transcriptome, Genes, 11, 542.

Zhang, Y. W., Liu, S., Zhang, X., Li, W. B., Zhang, X., Liu, C., Zheng, X., Calin, G. A., Wang, J., Lu, X., & Ellis, L. M. (2015). VEGFR-1 pseudogene expression and regulatory function in human colorectal cancer cells. Molecular Cancer Research, 13(9), 1274–1282.

Johnsson, P., Ackley, A., Vidarsdottir, L., Lui, W. O., Corcoran, M., Grandér, D., & Morris, K. V. (2013). A pseudogene long-noncoding-RNA network regulates PTEN transcription and translation in human cells. Nature Structural & Molecular Biology, 20(4), 440–446.

Hawkins, P. G., & Morris, K. V. (2010). Transcriptional regulation of Oct4 by a long non-coding RNA antisense to Oct4-pseudogene 5. Nature, 3(1), 165–175.

Tam, O. H., Aravin, A. A., Stein, P., Giraud, A., Murchison, E. P., Cheloufi, S., Hodges, E., Anger, M., Sachidanandam, R., Schultz, R. M., & Hanlon, G. J. (2008). Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. Nature, 453(7194), 534–538.

Watanabe, T., Totoki, Y., Toyota, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., Chiba, H., Kohara, Y., Kono, T., Nakano, T., Surani, M. A., Sakaki, Y., & Sasaki, H. (2008). Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. Nature, 453(7194), 539–543.

Fire, A., Xu, S., Montgomery, M. K., Costas, S. A., Driver, S. E., & Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature, 391(6669), 806–811.

Bernstein, E., Caudy, A. A., Hammond, S. M., & Hanlon, G. J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature, 409(6818), 363–366.

Chan, W. L., Yuo, C. Y., Yang, W. K., Hung, S. Y., Chang, Y. S., Chiu, C. C., Yeh, K. T., Huang, H. D., & Chang, J. G. (2013). Transcribed pseudogene ψRpPM1K generates endogenous siRNA to suppress oncogenic cell growth in hepatocellular carcinoma. Nucleic Acids Research, 41(6), 3734–3747.

Poliseno, L., Salmena, L., Zhang, J., Carver, B., Haveman, W. J., & Pandolfi, P. P. (2010). A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature, 465(7301), 1033–1038.

Karreth, F. A., Reschke, M., Ruocco, A. G., Ng, C., Chapuy, B., Léopold, V., Sjoberg, M., Keane, T. M., Verma, A., Alia, U., Tay, Y., Wu, D., Seitzer, N., Velasco-Herrera, M. D. C., Bothmer, A., Fung, J., Langellotto, F., Rodig, S. J., Elemento, O., ... Pandolfi, P. P. (2015). The BRAF pseudogene functions as a competitive endogenous RNA and induces lymphoma in vivo. Cell, 161(2), 319–332.

Thomson, D. W., & Dinger, M. E. (2016). Endogenous microRNA sponges: Evidence and controversy. Nature Reviews Genetics, 17(5), 272–283.

Denzler, R., Agarwal, V., Stefano, J., Bartel, D. P., & Stoffel, M. (2014). Assessing the cRNA hypothesis with quantitative measurements of miRNA and target abundance. Molecular Cell, 54(5), 766–776.

Kim, S. M., Pinto, S. M., Getnet, D., Nirujogi, R. S., Manda, S. S., Chaerkady, R., Madugundu, A. K., Kelkar, D. S., Isserlin, R., Jain, S., Thomas, J. K., Muthusamy, B., Leal-Rojas, P., Kumar, P., Sahasrabudhe, N. A., Balakrishnan, A., Advani, J., George, B., Renuse, S., ... Pandey, A. (2014). A draft map of the human proteome. Nature, 509(7502), 575–581.

Ji, Z., Song, R., Regev, A., & Struhl, K. (2015). Many IncRNAs, 5’UTRs, and pseudogenes are translated and some are likely to express functional proteins, eLife, 4, e08890.

Smith, A. H., Ziebell, F., Joberty, G., Zinn, N., Mueller, W. F., Clauder-Münster, S., Eberhard, D., Fälth Savitski, M., Grandi, P., Jakob, P., ...
Michon, A.-M., Sun, H., Tessmer, K., Bürckstümmer, T., Bantscheff, M., Steinmetz, L. M., Drewes, G., & Huber, W. (2019). Biological plasticity rescues target activity in CRISPR knockouts, *Nature Methods*, 16(11), 1087–1093.

85. Kandouz, M., Bier, A., Carystinos, G. D., Alaoui-Jamali, M. A., & Batist, G. (2004). Connexin43 pseudogene is expressed in tumor cells and inhibits growth. *Oncogene*, 23(27), 4763–4770.

86. Zhang, J., Wang, X., Li, M., Han, J., Chen, B., Wang, B., & Dai, J. (2006). NANOGP8 is a retrogene expressed in cancers. *The FEBS Journal*, 273(8), 1723–1730.

87. Wright, J. C., Mudge, J., Weisser, H., Barzine, M. P., Gonzalez, J. M., Brazma, A., Choudhary, J. S., & Harrow, J. (2016). Improving GENCODE reference gene annotation using a high-stringency proteogenomics workflow. *Nature Communications*, 7, 11788.

88. Rosso, L., Marques, A. C., Weier, M., Lambert, N., Lambot, M.-A., Vanderhaeghen, P., & Kaessmann, H. (2008). Birth and rapid subcellular adaptation of a hominoid-specific CDC14 protein. *PLoS Biology*, 6(6), e140.

89. Burki, F., & Kaessmann, H. (2004). Birth and adaptive evolution of a hominoid gene that supports high neurotransmitter flux. *Nature Genetics*, 36, 1061–1063.

90. Rosso, L., Marques, A. C., Reichert, A. S., & Kaessmann, H. (2008). Mitochondrial targeting adaptation of the hominoid-specific glutamate dehydrogenase driven by positive darwinian selection. *PLoS Genetics*, 4, e1000150.

91. Liu, G., Cui, X., Li, H., & Cai, L. (2016). Evolutionary direction of processed pseudogenes. *Science China. Life Sciences*, 59(8), 839–849.

92. Sandelin, A., Carninci, P., Lenhard, B., Ponjavic, J., Hayashizaki, Y., & Hume, D. A. (2007). Mammalian RNA polymerase II core promoters: Insights from genome-wide studies. *Nature Reviews Genetics*, 8, 424–436.

93. Kanber, D., Buiting, K., Roos, C., Gromoll, J., Kaya, S., Horsthemke, B., & Lohmann, D. (2013). The origin of the RB1 imprint. *PLoS One*, 8(11), e81502.

94. Soares, M. B., Schon, E., Henderson, A., Karathanasis, S. K., Cate, R., Zeitlin, S., Chirgwin, J., & Efstratiadis, A. (1985). RNA-mediated gene duplication: The rat preproinsulin I gene is a functional retroposon. *Molecular and Cellular Biology*, 5(8), 2090–2103.

95. Kazazian, H. H., Jr. (2014). Processed pseudogene insertions in somatic cells. *Mobile DNA*, 5(1), 20.

96. Khachane, A. N., & Harrison, P. M. (2009). Assessing the genomic evidence for conserved transcribed pseudogenes under selection. *BMC Genomics [Electronic Resource]*, 10, 435.

97. Gilbert, W., De Souza, S. J., & Long, M. (1997). Origin of genes. *Proceedings of the National Academy of Sciences of the United States of America*, 94(15), 7698–7703.

98. Rheinemann, L., Downhour, D. M., Bredbenner, K., Mercenne, G., Davenport, K. A., Schmitt, P. T., Necessary, C. R., McCullough, J., Schmitt, A. P., Simon, S. M., Sundquist, W. I., & Elde, N. C. (2020). RetroCHMP3 Blocks Budding of Enveloped Viruses Without Blocking Cytokinesis, bioRxiv, https://doi.org/10.1101/2020.08.30.273656.

99. Rheinemann, L., Downhour, D. M., Davenport, K. A., McKeown, A. N., Sundquist, W. I., & Elde, N. C. (2021). Recurrent Emergence of an Antiviral Defense through Repeated Retrotransposition and Truncation of CHMP3, bioRxiv, https://doi.org/10.1101/2021.04.27.441704.

100. Beadle, G. W., & Tatum, E. L. (1941). Genetic control of biochemical reactions in Neurospora. *Proceedings of the National Academy of Sciences of the United States of America*, 27, 499–506.

101. Pesole, G. (2008). What is a gene? An updated operational definition. *Gene*, 417(1-2), 1–4.