

Article

No Expression Divergence despite Transcriptional Interference between Nested Protein-Coding Genes in Mammals

Raquel Assis

Department of Electrical Engineering and Computer Science, Institute for Human Health and Disease Intervention, Florida Atlantic University, Boca Raton, FL 33431, USA; rassis@fau.edu

Abstract: Nested protein-coding genes accumulated throughout metazoan evolution, with early analyses of human and *Drosophila* microarray data indicating that this phenomenon was simply due to the presence of large introns. However, a recent study employing RNA-seq data uncovered evidence of transcriptional interference driving rapid expression divergence between *Drosophila* nested genes, illustrating that accurate expression estimation of overlapping genes can enhance detection of their relationships. Hence, here I apply an analogous approach to strand-specific RNA-seq data from human and mouse to revisit the role of transcriptional interference in the evolution of mammalian nested genes. A genomic survey reveals that whereas mammalian nested genes indeed accrued over evolutionary time, they are retained at lower frequencies than in *Drosophila*. Though several properties of mammalian nested genes align with observations in *Drosophila* and with expectations under transcriptional interference, contrary to both, their expression divergence is not statistically different from that between unnested genes, and also does not increase after nesting. Together, these results support the hypothesis that lower selection efficiencies limit rates of gene expression evolution in mammals, leading to their reliance on immediate eradication of deleterious nested genes to avoid transcriptional interference.



Citation: Assis, R. No Expression Divergence despite Transcriptional Interference between Nested Protein-Coding Genes in Mammals. *Genes* **2021**, *12*, 1381. <https://doi.org/10.3390/genes12091381>

Academic Editor: Kumiko Ui-Tei

Received: 16 July 2021

Accepted: 24 August 2021

Published: 1 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: nested genes; overlapping genes; transcriptional interference; gene expression

1. Introduction

Surveys of eukaryotic genome architecture have uncovered high frequencies of nested protein-coding genes, in which one “internal” gene is located in an intron of a second “external” gene [1–6]. Internal genes are typically short and intronless, whereas external genes tend to be long and possess many large introns [2–4]. A comprehensive analysis across three metazoan lineages illustrated that internal genes often arise via gene duplication, and that nested genes are formed when the resulting young duplicate genes are inserted into introns of existing genes [3]. This study also revealed that nested genes accumulated over evolutionary time, as evidenced by the predominance of nesting relative to unnesting events in all three metazoan lineages [3].

The finding that frequencies of nested genes increased over evolutionary time [3] is surprising, as such structures are expected to be evolutionarily disfavored due to transcriptional interference between external and internal genes [7,8]. Indeed, most external and internal genes are transcribed from opposite strands [2,4–6]. Nevertheless, interrogations of early human [9] and *Drosophila melanogaster* [10] microarray data yielded positive correlations between expression profiles of nested genes [3,6]. Though smaller than positive correlations between expression profiles of adjacent genes, they were found to be no different than those between intra-chromosomal genes [3,6]. Thus, these results support the hypothesis that nested genes accumulated over time simply because of increased nesting opportunities provided by large metazoan introns [3].

Yet, the conclusions of these early studies [3,6] were clouded by their dependence on data from microarray experiments, which can yield inaccurate estimates of gene expression

levels for overlapping genes. Moreover, the usage of correlation coefficients to assess expression divergence is biased when measurement error is large [11]. With these limitations in mind, a recent study used RNA-seq data [12,13] and Euclidian distance estimates of gene expression divergence [11] to reexamine the hypothesis that transcriptional interference impacts nested gene evolution in *Drosophila* [6]. This analysis uncovered widespread expression divergence between nested genes that was greater than that between either intra- or inter-chromosomal genes, providing strong support for transcriptional interference between nested genes in *Drosophila* [6]. Further, both expression and sequence divergence were found to rapidly increase after nesting, indicating that natural selection plays an important role in avoidance of transcriptional interference between *Drosophila* nested genes [6].

These findings in *Drosophila* prompt the question of whether transcriptional interference drives the evolution of nested genes in other taxonomic groups. Thus, here I address this question in mammals, which also accumulated nested genes over evolutionary time [3]. To do so, I take advantage of high-quality genome sequence and annotation data, along with strand-specific RNA-seq data from the same seven tissues [14,15], in human and mouse. Following the approach taken in *Drosophila* [6], I investigate mammalian nested gene prevalence and evolutionary dynamics, genomic and transcriptomic properties, and expression divergence. Joint consideration of these findings allows me to assess whether and how transcriptional interference influences the evolution of nested genes in the mammalian lineage.

2. Materials and Methods

2.1. Identification of Nested and Unnested Gene Pairs

Genome annotation (gtf) files for human (*Homo sapiens*), mouse (*Mus musculus*), cow (*Bos taurus*), opossum (*Monodelphis domestica*), platypus (*Ornithorhynchus anatinus*), chicken (*Gallus gallus*), and zebrafish (*Danio rerio*) were retrieved from the Ensembl release 104 [16] FTP site at <ftp.ensembl.org> (accessed on 23 August 2021). There are 452 pairs of nested protein-coding genes annotated in human (4.4%), 484 in mouse (4.3%), 745 in cow (6.8%), 926 in opossum (8.7%), 521 in platypus (6.0%), 640 in chicken (7.6%), and 673 in zebrafish (5.3%). The number of human nested genes identified here is consistent with that obtained from an older genome assembly [3]. For this study, I focused on properties of nested genes in human (Table S1) and mouse (Table S2), which have high-quality genome assemblies and annotation data sets, similar proportions of annotated nested genes, and strand-specific RNA-seq data from the same seven tissues (see Section 2.3 below). For comparison, I also obtained all intra-chromosomal (8,033,791 in human and 11,247,158 in mouse) and inter-chromosomal (142,486,134 in human and 190,692,335 in mouse) protein-coding gene pairs from the 17,351 and 20,113 unnested and non-overlapping protein-coding genes annotated in human and mouse, respectively. The null model for all comparisons between nested and unnested genes in this study is that their properties are similar, as that is the expectation in the absence of transcriptional interference.

2.2. Inference of Gene Nesting and Unnesting Events

I obtained 1:1 orthologs for all protein-coding genes in human (*Homo sapiens*), mouse (*Mus musculus*), cow (*Bos taurus*), opossum (*Monodelphis domestica*), platypus (*Ornithorhynchus anatinus*), chicken (*Gallus gallus*), and zebrafish (*Danio rerio*) from Ensembl release 104 [16] via the BioMart database [17]. Nesting events that occurred before the divergence of human and mouse lineages were inferred based on their presence in both of these species. In contrast, nesting events that occurred after the divergence of human and mouse lineages were inferred based on their presence in only one of these species and their absence in all outgroups. Though it is possible that genes underwent nesting and unnesting multiple times throughout evolution, the stringent requirement that nesting be absent in all outgroups enabled conservative identification of human- or mouse-specific nesting events. Moreover, to ensure that incomplete genome assembly or annotation errors

did not bias the inference of such nesting events, I required that external and internal genes both have orthologs in human, mouse, and at least one outgroup. Thus, nesting events were not inferred when one or both genes are simply absent ancestrally.

2.3. Gene Expression Analyses

Tables of normalized strand-specific RNA-seq abundances in transcripts per million (TPM) from brain, lung, liver, spleen, kidney, colon, and testis tissues in human (E-MTAB-4344) [15] and mouse (E-MTAB-2801) [14] were downloaded from Expression Atlas [18] at <https://www.ebi.ac.uk/gxa/home/> (accessed on 23 August 2021). All data in Expression Atlas are obtained with the iRAP pipeline, averaged across technical replicates, and quantile normalized [19]. Though there are numerous gene expression data sets available for human and mouse, I chose these specifically because they contain seven of the same tissues and were obtained from strand-specific RNA-seq experiments, which enable more accurate expression quantification of overlapping genes [20]. To minimize noise, all genes with $\text{TPM} \geq 1$ in at least one of the seven tissues were retained for expression analyses, yielding 269 human (Table S1) and 265 mouse (Table S2) nested genes for which both external and internal genes met this threshold. The requirement that both external and internal genes be expressed was used to ensure that findings from expression analyses involving one and both genes are comparable, and also that transcriptional interference between the genes is possible. I estimated the expression breadth of each gene by computing the tissue specificity index τ , which ranges from 0 (broadly expressed) to 1 (tissue specific [21]), and the expression divergence between each pair of genes by computing the Euclidian distance between their relative TPM across tissues, which enables inter-species comparisons [11,22].

2.4. Statistical Analyses

All statistical analyses were performed in the R software environment [23]. Two-tailed binomial tests implemented with the `binom.test()` function in the `stats` package [23] were used to compare numbers of nested genes on the same vs. opposite strands, numbers of tissue-specific external vs. internal genes, and numbers of tissue-specific genes expressed in each tissue for nested vs. unnested genes. In each comparison of nested genes on the same vs. opposite strands, x was set to the number of nested genes on opposite strands, n to the total number of nested genes, and $p = 0.5$ to represent the expected frequency of opposite-strand nestings if orientation is random. In each comparison of numbers of tissue-specific external vs. internal genes, x was set to the number of tissue-specific external genes, n to the total number of tissue-specific external and internal genes, and $p = 0.5$ to represent the expected frequency of tissue-specific external genes if tissue specificity is random. In each comparison of numbers of tissue-specific genes expressed in each tissue for nested vs. unnested genes, x was set to the number of tissue-specific nested genes in the tissue of interest, n to the total number of tissue-specific nested genes, and p to the proportion of tissue-specific unnested genes in the tissue of interest. For these analyses, p -values were Bonferroni-corrected for the seven comparisons performed. Two-tailed Fisher's exact tests implemented with the `fisher.test()` function in the `stats` package [23] were used to compare numbers of nested genes on the same vs. opposite strands and numbers of tissue-specific external vs. internal genes between human and mouse. Two-tailed two-sample permutation tests implemented with the `permTS()` function in the `perm` package [24] were used for all pairwise comparisons between distributions. For comparisons involving intra- or inter-chromosomal gene pairs, the `permControl()` function was used to restrict the number of permutations to 1000.

3. Results

3.1. Prevalence and Evolutionary Dynamics of Nested Protein-Coding Genes in Mammals

Across the seven vertebrate species surveyed, which included five mammals, 4.3–8.7% of protein-coding genes were found in nested structures (see Materials and Methods for details). Human and mouse genomes sit at the lower end of this range, with 4.4%

and 4.3% of their genes nested, respectively. These proportions are roughly half of those observed across 12 *Drosophila* species [6], consistent with relative frequencies obtained in an earlier study of metazoan nested genes [3]. Thus, also taking into consideration that human and mouse have the highest quality and best annotated genomes among those of the species examined here, gene nesting appears to be much less common in mammals than in *Drosophila*. Because most nested genes arise from the insertion of young duplicate genes into the introns of existing genes [3], this difference may be attributed to either gene duplication or nesting. However, the higher gene duplication rates in mammals [25,26] and similar proportions of retained duplicate genes in mammalian and *Drosophila* genomes [27] are inconsistent with a difference due to either neutral or selective forces involved in gene duplication. A neutral scenario in which genomic composition impacts nesting probabilities is also unlikely, as intronic and intergenic regions display conserved 1:1 ratios across metazoans [28]. As a result, the lower frequency of nested genes in mammals may be best explained by stronger selection to eradicate such structures, which is likely the first mechanism of defense against transcriptional interference between external and internal genes.

Analysis of the evolutionary dynamics of human and mouse nested protein-coding genes (see Materials and Methods for details) uncovered 73 (Table S3) nesting events that occurred before the divergence of the two mammalian lineages, and 56 nesting events that occurred after their divergence—34 in the human lineage (Table S4) and 22 in the mouse lineage (Table S5). In contrast, I only identified four cases of unnesting events (three in human and one in mouse), mirroring previous findings of more frequent nesting than unnesting in human and mouse [3]. Because this earlier study also revealed a similar trend in *Drosophila* and *Caenorhabditis* lineages [3], the current analysis supports the hypothesis that gene nestings accumulate and contribute to the increased organizational complexity of mammalian and other metazoan genomes over evolutionary time [3]. This phenomenon was previously explained by the presence of large metazoan introns [3], which take up as much genomic space as intergenic regions [28] and offer ample opportunities for gene nesting. However, another contributing factor may be rapid gene duplication, as this mutational process creates most internal genes [3]. Indeed, experimental studies have shown that gene duplication occurs faster than all other types of spontaneous mutation in several metazoan species [26,29–32]. Hence, if this pattern holds in mammals, then it is possible that large and abundant mammalian introns provide much-needed homes for floods of newly generated young duplicate genes.

3.2. Genomic and Transcriptomic Properties of Nested Protein-Coding Genes in Mammals

Though large introns coupled with fast duplication rates may contribute to the rapid creation of mammalian nested protein-coding genes, it is curious how such structures persist in the presence of transcriptional interference between external and internal genes. As a first step in addressing this question, I examined relationships between external and internal genes. To facilitate direct 1:1 comparisons between external and internal genes, I restricted my analysis to the 296 human (Table S1) and 248 mouse (Table S2) simple nested protein-coding gene pairs, in which an external gene contains only one internal gene in its intron. Of these simple nested gene pairs, 220 in human (74.3%) and 193 in mouse (77.8%) contain external and internal genes on opposite strands (Table 1). Hence, there are similar opposite-strand biases in human and mouse, consistent with those observed in previous studies of human [2] and *Drosophila* [6] nested genes. These comparably strong biases point to a preference for opposite-strand nestings that crosses taxonomic boundaries, suggesting that purging of same-strand nestings by negative selection may serve as a global mechanism for reducing transcriptional interference between nested genes.

Table 1. Numbers of Simple Nested Protein-Coding Genes on the Same and Opposite Strands.

	Same	Opposite	Same vs. Opposite *
Human	76	220	$p = 2.13 \times 10^{-17}$
Mouse	55	193	$p = 3.67 \times 10^{-19}$
Human vs. Mouse **	$p = 0.37$		

* Binomial tests (see Materials and Methods for details). ** Fisher's exact test (see Materials and Methods for details).

Previous studies have shown that young duplicate genes in mammals and many other animals and plants tend to be expressed primarily in male reproductive tissues [33–39]. However, if transcriptional interference drives the evolution of nested genes, then one would expect external and internal genes to be expressed in different tissues. Consistent with this hypothesis, studies of *Drosophila* nested genes have shown that whereas internal genes are often testis specific, external genes tend to be broadly expressed across several tissues [5,6]. To determine whether this is also true in mammals, I first examined distributions of the tissue specificity index τ [21] across seven tissues in human and mouse unnested, external, and internal genes (Figure 1A; see Materials and Methods for details). In both mammals, internal genes tend to be more tissue specific than either external or unnested genes, mirroring results in *Drosophila* [5,6]. Also consistent with *Drosophila* [5,6], human external genes are significantly more broadly expressed than unnested genes. However, this is not the case for mouse external genes, which have similar expression breadths as unnested genes. Nevertheless, both mammals display elevated tissue specificities of their internal nested genes, consistent with observations of young duplicate genes [33–39], as well as clear differences between expression breadths of external and internal genes, supporting the expectation under transcriptional interference.

To further investigate expression breadths of nested protein-coding genes, I extracted all tissue-specific ($\tau > 0.9$) [21] genes. Application of this cutoff yielded 40 external and 81 internal tissue-specific genes in human, and 69 external and 101 internal tissue-specific genes in mouse (Table 2). Of this subset, there are 18 cases for which both external and internal genes in a pair are tissue specific in human (Table S1), and 78 such cases in mouse (Table S2). Thus, there are similar over-representations of internal tissue-specific genes in both mammals, consistent with previous findings in *Drosophila* [5,6]. For each tissue, I compared observed numbers of tissue-specific nested genes to expectations based on unnested genes (Figure 1B; see Materials and Methods for details). In human, no statistically significant trends were uncovered, perhaps due to a lack of power from small sample sizes (Table 2), though there may be a preference for testis specificity among internal genes ($p = 0.06$). In mouse, there are larger and statistically significant over-representations of testis-specific internal and brain-specific external genes. Thus, though external genes tend to be more tissue specific in mouse than in human, the primary tissue in which mouse external genes are expressed (brain) differs from the primary tissue in which their internal genes are expressed (testis). Therefore, the results in both mammals suggest that external and internal genes are typically expressed in different tissues, as one might expect under transcriptional interference.

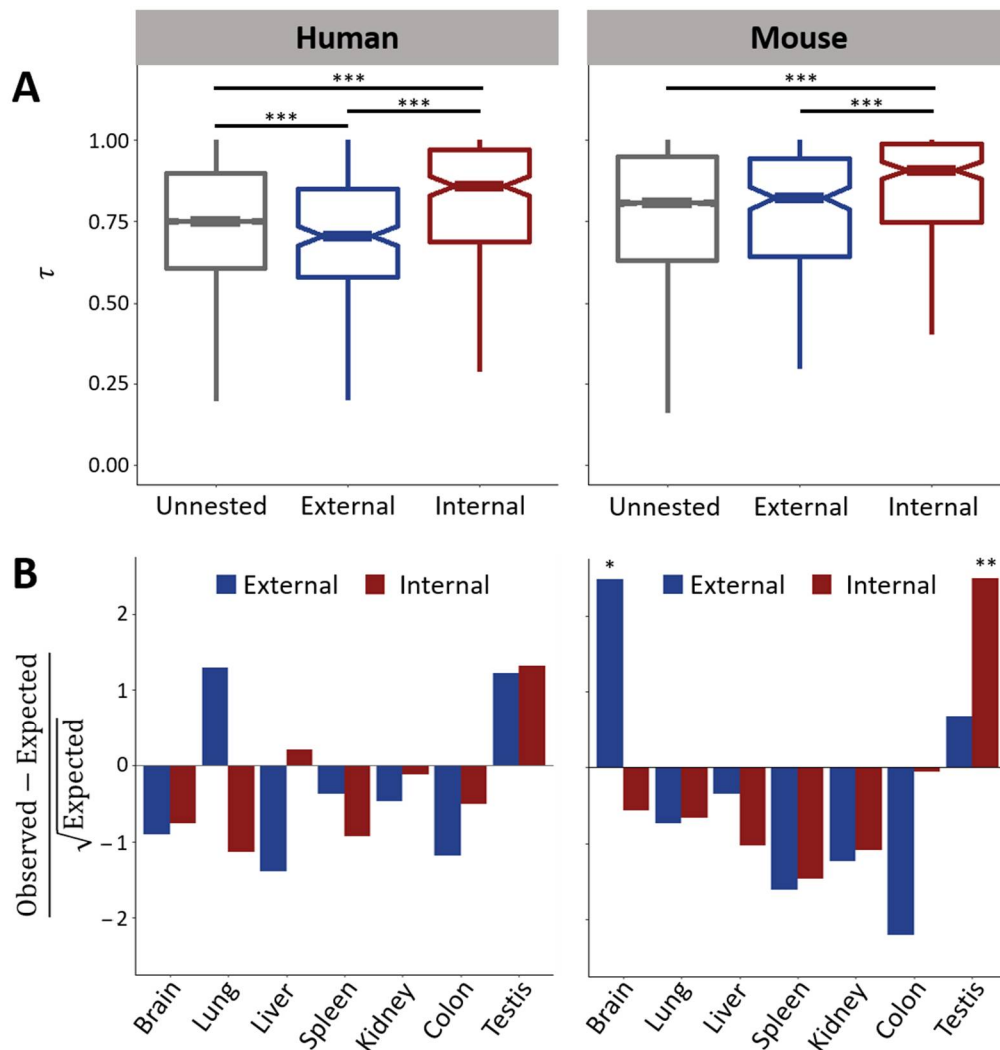


Figure 1. Expression breadths of external and internal protein-coding genes. **(A)** Distributions of tissue specificities (τ) across seven tissues in human (left) and mouse (right) unnested, external, and internal genes. Higher τ corresponds to greater tissue specificity. **(B)** Hanging chi-grams comparing observed numbers of human (left) and mouse (right) primary tissues of tissue-specific external and internal genes to expectations based on those of unnested genes. Positive and negative values indicate over-representations and under-representations, respectively. * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$ (after Bonferroni corrections for Figure 1B; see Materials and Methods for details).

Table 2. Numbers of Tissue-Specific External and Internal Protein-Coding Genes.

	External	Internal	External vs. Internal *
Human	40	81	$p = 2.13 \times 10^{-17}$
Mouse	69	101	$p = 3.67 \times 10^{-19}$
Human vs. Mouse **	$p = 0.22$		

* Binomial tests (see Materials and Methods for details). ** Fisher's exact test (see Materials and Methods for details).

3.3. Expression Divergence between Nested Protein-Coding Genes in Mammals

Last, I compared expression divergence across the seven tissues between pairs of nested, intra-chromosomal, and inter-chromosomal protein-coding genes (Figure 2; see Materials and Methods for details). In both mammals, expression divergence between nested genes is slightly elevated, but not significantly different from that between intra-chromosomal or inter-chromosomal genes. This result starkly contrasts the much higher

expression divergence observed between *Drosophila* nested genes [6]. Perhaps not surprisingly, there is also no support in mammals for the rapid increase in expression divergence after nesting that was observed in *Drosophila* [6]. In particular, though numbers of nesting events are small in both mammals (Tables S3 and S4), expression divergence between these derived nested genes and their ancestral unnested orthologs are similar to one another, as well as to that between nested genes conserved in both mammals (Table S5; $p > 0.05$ for both comparisons, permutation tests; see Materials and Methods for details). Hence, expression divergence does not appear to substantially increase either immediately or long after gene nesting has occurred in mammals. This lack of expression divergence is consistent with lower selection efficiencies in mammals than in *Drosophila* [40]. Further, perhaps the relative deficiency of nested genes in mammals is reflective of a preference for eradicating new nesting events in their avoidance of transcriptional interference, as their abilities to diverge and accommodate new nested gene structures are more limited than those of *Drosophila*.

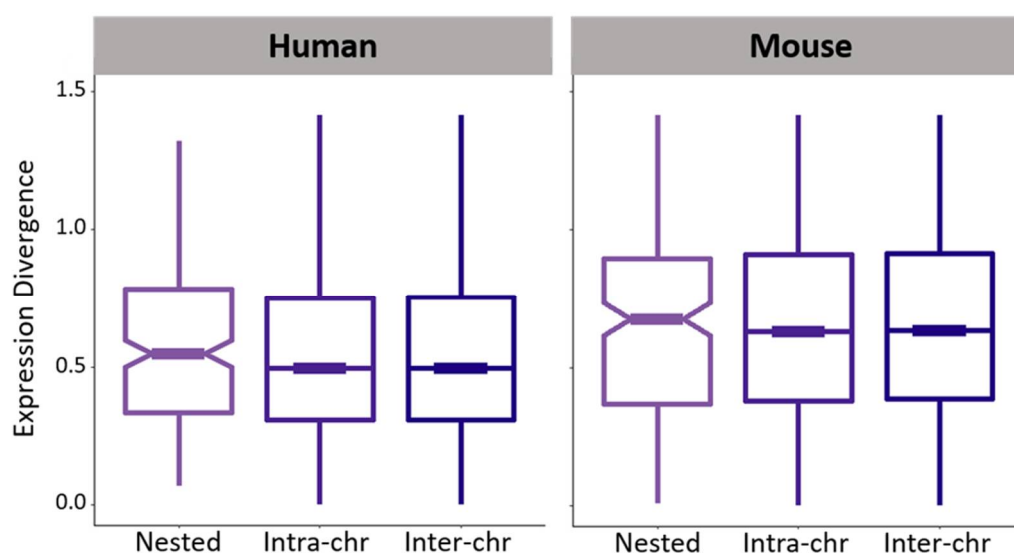


Figure 2. Expression divergence between nested, intra-chromosomal, and inter-chromosomal protein-coding genes. Distributions of Euclidian distances across seven tissues between gene pairs in human (left) and mouse (right). None of the pairwise differences between distributions are statistically significant (see Materials and Methods for details).

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/genes12091381/s1>, Table S1: Nested protein-coding genes in human, Table S2: Nested protein-coding genes in mouse, Table S3: Protein-coding gene nesting events that occurred before the divergence of human and mouse lineages, Table S4: Protein-coding gene nesting events that occurred in the human lineage, Table S5: Protein-coding gene nesting events that occurred in the mouse lineage.

Funding: This research was funded by NSF (grant no. DEB-2001059).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data produced in this study are provided as Supplementary Materials (Tables S1–S5).

Acknowledgments: I would like to thank the editor and two anonymous reviewers for their helpful comments.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Veeramachaneni, V.; Makalowski, W.; Galdzicki, M.; Sood, R.; Makalowska, I. Mammalian Overlapping Genes: The Comparative Perspective. *Genome Res.* **2004**, *14*, 280–286. [[CrossRef](#)]
2. Yu, P.; Ma, D.; Xu, M. Nested genes in the human genome. *Genomics* **2005**, *86*, 414–422. [[CrossRef](#)] [[PubMed](#)]
3. Assis, R.; Kondrashov, A.S.; Koonin, E.V.; Kondrashov, F.A. Nested genes and increasing organizational complexity of metazoan genomes. *Trends Genet.* **2008**, *24*, 475–478. [[CrossRef](#)] [[PubMed](#)]
4. Kumar, A. An Overview of Nested Genes in Eukaryotic Genomes. *Eukaryot. Cell* **2009**, *8*, 1321–1329. [[CrossRef](#)]
5. Lee, Y.C.G.; Chang, H.-H. The Evolution and Functional Significance of Nested Gene Structures in *Drosophila melanogaster*. *Genome Biol. Evol.* **2013**, *5*, 1978–1985. [[CrossRef](#)]
6. Assis, R. Transcriptional Interference Promotes Rapid Expression Divergence of *Drosophila* Nested Genes. *Genome Biol. Evol.* **2016**, *8*, 3149–3158. [[CrossRef](#)] [[PubMed](#)]
7. Shearwin, K.E.; Callen, B.P.; Egan, J.B. Transcriptional interference—A crash course. *Trends Genet.* **2005**, *21*, 339–345. [[CrossRef](#)]
8. Liao, B.Y.; Zhang, J. Coexpression of linked genes in mammalian genomes is generally disadvantageous. *Mol. Biol. Evol.* **2008**, *25*, 1555–1565. [[CrossRef](#)]
9. Su, A.I.; Wiltshire, T.; Batalov, S.; Lapp, H.; Ching, K.A.; Block, D.; Zhang, J.; Soden, R.; Hayakawa, M.; Kreiman, G.; et al. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 6062–6067. [[CrossRef](#)]
10. Chintapalli, V.R.; Wang, J.; Dow, J. Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. *Nat. Genet.* **2007**, *39*, 715–720. [[CrossRef](#)]
11. Pereira, V.; Waxman, D.; Eyre-Walker, A. A Problem with the Correlation Coefficient as a Measure of Gene Expression Divergence. *Genetics* **2009**, *183*, 1597–1600. [[CrossRef](#)]
12. Graveley, B.R.; Brooks, A.N.; Carlson, J.W.; Duff, M.O.; Landolin, J.M.; Yang, L.; Artieri, C.G.; van Baren, M.J.; Boley, N.; Booth, B.W.; et al. The developmental transcriptome of *Drosophila melanogaster*. *Nature* **2011**, *471*, 473–479. [[CrossRef](#)]
13. Kaiser, V.B.; Zhou, Q.; Bachtrog, D. Non-random gene loss from the *Drosophila miranda* neo-Y chromosome. *Genome Biol. Evol.* **2011**, *3*, 1329–1337. [[CrossRef](#)]
14. Merkin, J.; Russell, C.; Chen, P.; Burge, C.B. Evolutionary Dynamics of Gene and Isoform Regulation in Mammalian Tissues. *Science* **2012**, *338*, 1593–1599. [[CrossRef](#)] [[PubMed](#)]
15. Lin, S.; Lin, Y.; Nery, J.R.; Urich, M.A.; Breschi, A.; Davis, C.A.; Dobin, A.; Zaleski, C.; Beer, M.A.; Chapman, W.C.; et al. Comparison of the transcriptional landscapes between human and mouse tissues. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 17224–17229. [[CrossRef](#)] [[PubMed](#)]
16. Howe, K.L.; Achuthan, P.; Allen, J.; Allen, J.; Alvarez-Jarreta, J.; Amode, M.R.; Armean, I.M.; Azov, A.G.; Bennett, R.; Bhai, J.; et al. Ensembl 2021. *Nucleic Acids Res.* **2021**, *49*, 884–891. [[CrossRef](#)] [[PubMed](#)]
17. Smedley, D.; Haider, S.; Ballester, B.; Holland, R.; London, D.; Thorisson, G.; Kasprzyk, A. BioMart—Biological queries made easy. *BMC Genom.* **2009**, *10*, 22. [[CrossRef](#)]
18. Kapushesky, M.; Emam, I.; Holloway, E.; Kurnosov, P.; Zorin, A.; Malone, J.; Rustici, G.; Williams, E.; Parkinson, H.; Brazma, A. Gene Expression Atlas at the European Bioinformatics Institute. *Nucleic Acids Res.* **2010**, *38*, D690–D698. [[CrossRef](#)] [[PubMed](#)]
19. Papatheodorou, I.; Fonseca, N.A.; Keays, M.; Tang, Y.A.; Barrera, E.; Bazant, W.; Burke, M.; Fullgrave, A.; Fuentes, A.M.-P.; George, N.; et al. Expression Atlas: Gene and protein expression across multiple studies and organisms. *Nucleic Acids Res.* **2018**, *46*, D246–D251. [[CrossRef](#)] [[PubMed](#)]
20. Zhao, S.; Zhang, Y.; Gordon, W.; Quan, J.; Xi, H.; Du, S.; Von Schack, D.; Zhang, B. Comparison of stranded and non-stranded RNA-seq transcriptome profiling and investigation of gene overlap. *BMC Genom.* **2015**, *16*, 675. [[CrossRef](#)] [[PubMed](#)]
21. Yanai, I.; Benjamin, H.; Shmoish, M.; Chalifa-Caspi, V.; Shklar, M.; Ophir, R.; Bar-Even, A.; Horn-Saban, S.; Safran, M.; Domany, E.; et al. Genome-wide midrange transcription profiles reveal ex-expression level relationships in human tissue specification. *Bioinformatics* **2005**, *21*, 650–659. [[CrossRef](#)]
22. Liao, B.-Y.; Zhang, J. Evolutionary Conservation of Expression Profiles Between Human and Mouse Orthologous Genes. *Mol. Biol. Evol.* **2005**, *23*, 530–540. [[CrossRef](#)]
23. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020.
24. Fay, M.P.; Shaw, P.A. Exact and asymptotic weighted logrank tests for interval censored data: The interval R package. *J. Stat. Softw.* **2010**, *36*, 1–34. [[CrossRef](#)]
25. Itsara, A.; Wu, H.; Smith, J.D.; Nickerson, D.A.; Romieu, I.; London, S.J.; Eichler, E.E. De novo rates and selection of large copy number variation. *Genome Res.* **2010**, *20*, 1469–1481. [[CrossRef](#)]
26. Schrider, D.R.; Houle, D.; Lynch, M.; Hahn, M. Rates and Genomic Consequences of Spontaneous Mutational Events in *Drosophila melanogaster*. *Genetics* **2013**, *194*, 937–954. [[CrossRef](#)]
27. Zhang, J. Evolution by gene duplication: An update. *Trends Ecol. Evol.* **2003**, *18*, 292–298. [[CrossRef](#)]
28. Francis, W.R.; Wörheide, G. Similar Ratios of Introns to Intergenic Sequence across Animal Genomes. *Genome Biol. Evol.* **2017**, *9*, 1582–1598. [[CrossRef](#)] [[PubMed](#)]
29. Lynch, M.; Sung, W.; Morris, K.; Coffey, N.; Landry, C.R.; Dopman, E.B.; Dickinson, W.J.; Okamoto, K.; Kulkarni, S.; Hartl, D.L.; et al. A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 9272–9277. [[CrossRef](#)] [[PubMed](#)]

30. Lipinski, K.J.; Farslow, J.C.; Fitzpatrick, K.A.; Lynch, M.; Katju, V.; Bergthorsson, U. High Spontaneous Rate of Gene Duplication in *Caenorhabditis elegans*. *Curr. Biol.* **2011**, *21*, 306–310. [[CrossRef](#)]
31. Keith, N.; Tucker, A.E.; Jackson, C.E.; Sung, W.; Lledó, J.I.L.; Schrider, D.R.; Schaack, S.; Dudycha, J.L.; Ackerman, M.; Younge, A.J.; et al. High mutational rates of large-scale duplication and deletion in *Daphnia pulex*. *Genome Res.* **2015**, *26*, 60–69. [[CrossRef](#)] [[PubMed](#)]
32. Konrad, A.; Flibotte, S.; Taylor, J.; Waterston, R.H.; Moerman, D.G.; Bergthorsson, U.; Katju, V. Mutational and transcriptional landscape of spontaneous gene duplications and deletions in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 7386–7391. [[CrossRef](#)] [[PubMed](#)]
33. Betrán, E.; Thornton, K.; Long, M. Retroposed New Genes Out of the X in *Drosophila*. *Genome Res.* **2002**, *12*, 1854–1859. [[CrossRef](#)]
34. Vinckenbosch, N.; Dupanloup, I.; Kaessmann, H. Evolutionary fate of retroposed gene copies in the human genome. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 3220–3225. [[CrossRef](#)]
35. Kaessmann, H. Origins, evolution, and phenotypic impact of new genes. *Genome Res.* **2010**, *20*, 1313–1326. [[CrossRef](#)] [[PubMed](#)]
36. Assis, R.; Bachtrog, D. Neofunctionalization of young duplicate genes in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 17409–17414. [[CrossRef](#)]
37. Wu, N.-D.; Wang, X.; Li, Y.; Zeng, L.; Irwin, D.M.; Zhang, Y.-P. “Out of pollen” hypothesis for origin of new genes in flowering plants: Study from *Arabidopsis thaliana*. *Genome Biol. Evol.* **2014**, *6*, 2822–2829. [[CrossRef](#)]
38. Assis, R.; Bachtrog, D. Rapid divergence and diversification of mammalian duplicate gene functions. *BMC Evol. Biol.* **2015**, *15*, 138. [[CrossRef](#)] [[PubMed](#)]
39. Jiang, X.; Assis, R. Rapid functional divergence after small-scale gene duplication in grasses. *BMC Evol. Biol.* **2019**, *19*, 97. [[CrossRef](#)] [[PubMed](#)]
40. Charlesworth, B. Fundamental concepts in genetics: Effective population size and patterns of molecular evolution and variation. *Nat. Rev. Genet.* **2009**, *10*, 195–205. [[CrossRef](#)] [[PubMed](#)]