Response to reviewers' Comments for "Pseudogenes in the mouse lineage: transcriptional activity and strain-specific history "

Response Letter

-- Ref1.1: Introductory comments --

Reviewer Comment	In this paper the authors provided a comprehensive and updated annotation of pseudogenes in a list of 16 mouse strains, encompassing evolutionary time of 6 million years. This effort complements and completes the genic annotations of these species, and provides a unique perspective on the evolution of genomes. Perhaps the biggest surprise is the large number of unitary pseudogenes reported from each species, which will no doubt shed light on the function of these genes and why the loss of which were tolerated in each species.
Author Response	We sincerely thank the reviewer for these kind words and their constructive comments, which we believe made our paper stronger. We respond to the reviewer's comments below.

-- Ref1.2: Reference --

Reviewer Comment	Reference 17 (Line 779) has no author names. Troublesome variability in mouse studies. Nat Neurosci 12, 1075 (2009).
Author Response	We thank the reviewer for pointing out this reference. However this citation relates to an Editorial article from Nature Neuroscience and it does not specify any authors names. We updated the reference to highlight this.
Excerpt From Revised Manuscript	17. Editorial, Troublesome variability in mouse studies. *Nat Neurosci* **12**, 1075 (2009).

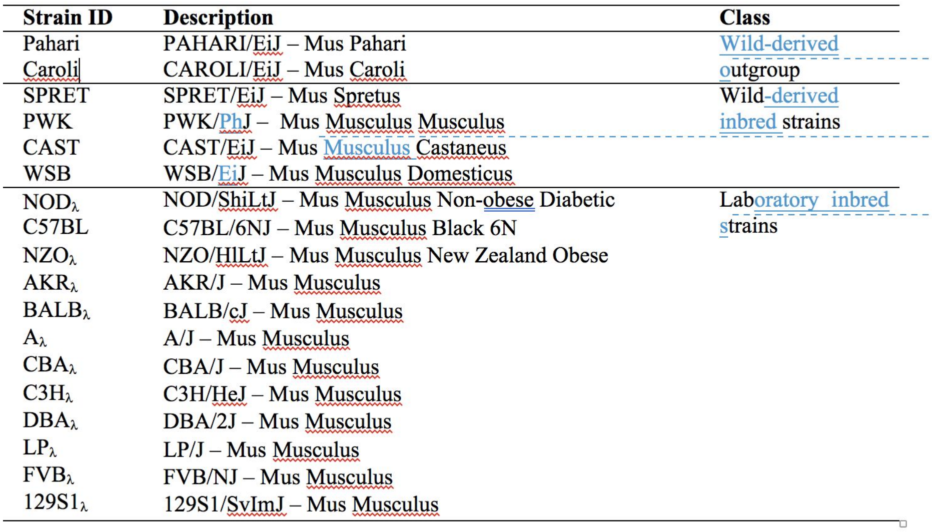
-- Ref1.3: Unitary pseudogene --

Reviewer Comment	I find the definition of unitary pseudogenes (or lack of it) rather ambivalent. These are loci that have become pseudogenes in one organism but maintain as functional gene in other organisms. Please define them properly.
Author Response	We updated the manuscript to include a complete definition of unitary pseudogenes that is in accord with that of the reviewer.

1.3 ORTHO

Excerpt From Revised Manuscript	There is also a third class of pseudogenes, called unitary. These pseudogenes are formed when a functional gene acquires disabling mutations resulting in the inactivation of the original coding locus. Unitary pseudogenes are also characterized by the presence of a functional gene on the same locus in other species.

-- Ref1.4: Table 2 Typo --

<table>
<tr><td>Reviewer Comment</td><td>"Must Castaneus" should be "Mus Castaneus" I am also wondering why CAST is not listed as "Mus musculus castaneus" as PWK and WSB. SPRET is generally recognized as a separate species, so that explains why it is listed as Mus Spretus.</td></tr>
<tr><td>Author Response</td><td>We corrected the mistake in Table 2 and revised the nomenclature of mouse strains both in the text as well as in the table.</td></tr>
<tr><td>Excerpt From Revised Manuscript</td><td>The strains are broadly organized into 3 classes (<b>Table 2</b>): the outgroup wild-derived inbred strains – formed by two independent mouse species, <i>Mus Caroli</i> and <i>Mus Pahari</i>; wild-derived inbred strains – covering the subspecies <i>Mus Spretus,</i> and three musculus strains (<i>Mus Musculus Castaneus, Mus Musculus Musculus,</i> and <i>Mus Musculus Domesticus</i>), and a set of 12 laboratory inbred strains. A detailed summary of the genome composition for each strain is presented in [36].

<table>
<tr><th>Strain ID</th><th>Description</th><th>Class</th></tr>
<tr><td>Pahari</td><td>PAHARI/EiJ – Mus Pahari</td><td>Wild-derived outgroup</td></tr>
<tr><td>Caroli</td><td>CAROLI/EiJ – Mus Caroli</td><td></td></tr>
<tr><td>SPRET</td><td>SPRET/EiJ – Mus Spretus</td><td>Wild-derived inbred strains</td></tr>
<tr><td>PWK</td><td>PWK/PhJ – Mus Musculus Musculus</td><td></td></tr>
<tr><td>CAST</td><td>CAST/EiJ – Mus Musculus Castaneus</td><td></td></tr>
<tr><td>WSB</td><td>WSB/EiJ – Mus Musculus Domesticus</td><td></td></tr>
<tr><td>NOD$_\lambda$</td><td>NOD/ShiLtJ – Mus Musculus Non-obese Diabetic</td><td>Laboratory inbred strains</td></tr>
<tr><td>C57BL</td><td>C57BL/6NJ – Mus Musculus Black 6N</td><td></td></tr>
<tr><td>NZO$_\lambda$</td><td>NZO/HlLtJ – Mus Musculus New Zealand Obese</td><td></td></tr>
<tr><td>AKR$_\lambda$</td><td>AKR/J – Mus Musculus</td><td></td></tr>
<tr><td>BALB$_\lambda$</td><td>BALB/cJ – Mus Musculus</td><td></td></tr>
<tr><td>A$_\lambda$</td><td>A/J – Mus Musculus</td><td></td></tr>
<tr><td>CBA$_\lambda$</td><td>CBA/J – Mus Musculus</td><td></td></tr>
<tr><td>C3H$_\lambda$</td><td>C3H/HeJ – Mus Musculus</td><td></td></tr>
<tr><td>DBA$_\lambda$</td><td>DBA/2J – Mus Musculus</td><td></td></tr>
<tr><td>LP$_\lambda$</td><td>LP/J – Mus Musculus</td><td></td></tr>
<tr><td>FVB$_\lambda$</td><td>FVB/NJ – Mus Musculus</td><td></td></tr>
<tr><td>129S1$_\lambda$</td><td>129S1/SvImJ – Mus Musculus</td><td></td></tr>
</table>
</td></tr>
</table>

-- Ref1.5: Gene loss rate is similar in human and mouse --

Reviewer Comment	I have doubts on the statement that "the gene loss rate is similar in both mouse and primate lineage" (ref, 39 and 40). If this is indeed true, then it conveys that gene loss is a evolutionarily neutral process and the rate of gene loss is proportional to the time of divergence. However, in the discussions before this part in the manuscript, the

1.5 JULIE

	authors seem to be hinting that the creation of unitary pseudogenes is likely the result of changes in selective pressure or has adaptive benefits.
Author Response	The gene loss process has been shown to be subjected to either neutral or adaptive evolution. In the case of pseudogenes as a whole we do work on the premise that in most cases a loss of function event will be under neutral evolution. However there are a number of examples that highlight the causative relationship between the LOF event and adaptive evolution. To Add examples from: PMID: 27087500, 26438339,19411603
Excerpt From Revised Manuscript	

-- Ref1.6: Figure 3A --

Reviewer Comment	I have trouble understanding Figure 3A. What do the columns and different colors represent?
Author Response	Figure 3A is a set diagram representing the distribution of pseudogenes across the strains. Each column is a particular subset of the strains. The strains included in each subset are identified by the filled-in circles below the column. The column colors correspond to the types of strains included within each subset. Blue represents the outgroup strains, red represents the wild strains, yellow represents the lab strains, green represents the reference strain, and black columns indicate a subset that contains strains from multiple of the groups listed above. We have updated the figure to include a color legend and added a description the significance of each colour to the figure caption in order to improve clarity.
Excerpt From Revised Manuscript	A – Summary of pseudogene distribution in the pangenome mouse strain dataset. The different group of mouse strains are highlighted by colours: blue relates to outgroup mice (Mus Pahari and Mus Caroli), red corresponds to wild-derived mice (Mus Spretus, Mus Musculus Castaneus, Mus Musculus Musculus, and Mus Musculus Domesticus), yellow indicates the laboratory inbred strains as listed in Table 2, and green highlights the laboratory inbred "reference" strains C57BL/6NJ that is the closest related strain to the mouse reference genome C57BL/6J.

1.6 REP

-- Ref1.7: Phylogeny --

Reviewer Comment	How were these phylogenetic trees generated? When constructing a phylogenetic tree from a group of sequences, one can either concatenate the sequences into a super gene and build a tree from the supergene, or one can build trees from individual gene or pseudogene and then derive the consensus tree form these individual trees. Which approach did the author take? The authors also need to provide bootstrapping values for each branching point. Also please confirm that the tree for protein coding genes are aligned on amino acid sequences, while the trees on pseudogenes are aligned on nucleotide sequences.
Author Response	We expanded our methods section to give more detail on the phylogenetic analysis conducted. The trees have been redrawn to include the bootstrapping values.
Excerpt From Revised Manuscript	For each of the 18 mouse genomes, the extracted sequences were concatenated into strain-specific contig (supergene). The order of the pseudogene sequences was kept the same in all 18 contigs. The 18 supergenes were subjected to a multi-sequence alignment using MUSCLE aligner [65] under standard conditions. Similarly, the sequences of parent protein coding genes of the 1,460 pseudogenes were assembled into a strain specific sequence and aligned using MUSCLE. The tree was generated using Tamura-Nei genetic distance model and neighbouring-joining tree build method with Pahari as outgroup using GENEIOUS 10.2 software package [66]. The phylogenetic trees exemplifying 4 pfam families, were constructed using conserved pseudogenes across the 18 strains that belong to each of the families. The list of the selected pseudogenes is give in Table SNEW-Phylo.

1.7
CONFIRM
?

1.7b
BOOT
IN TEXT
(MORE)

-- Ref1.8: Figure 4A --

Reviewer Comment	The duplicated pseudogenes are dominated by olfactory factors, and the processed pseudogenes are dominated by 70-80 ribosomal proteins. I wonder how the plots look like with only these two families or with these two families removed. Also I am not sure I agree with the statements listed in line 288-293, regarding expression level of single member of large protein families. Please elaborate with examples.
Author Response	-- to add plots but I am pretty sure that the plots would look rather similar, the correlation would be rather poor too. -- look for examples!!!! **TODO**
Excerpt From Revised Manuscript	

TODO
1.8

1.9 EXPL

-- Ref1.9: Figure 4C --

Reviewer Comment	The graph shows that there is even a reduction in mouse in the number of processed pseudogenes negated in the recent history (right side of the curve). Does this mean that retrotransposition process also slowed down in mouse ?
Author Response	The reviewer raises a very good and interesting question. We point out that we use sequence similarity to parents as a proxy for pseudogene age. While this method is representative for the majority of pseudogenes, it is not accurate for pseudogenes that have been preserved under positive selection. Moreover, the right side of the curve in figure 4C covers pseudogenes that have high sequence similarity to their functional homologs. In order to prevent false positive, we use very stringent criteria in calling pseudogenes, thus our pipeline might overlook potential candidates with high sequence similarity to their parent genes. Therefore, based on the current data we would like to refrain on making any such speculation on the rate of the retrotransposition in mouse based solely on the number of pseudogenes. We have added a paragraph in the text discussing this issue.
Excerpt From Revised Manuscript	Moreover, a close examination of the young pseudogene density in suggests a reduction in the number of new pseudogenes being created. However, this observation is most likely a consequence of the stringent criteria used in calling pseudogenes at high sequence similarity to parents. Thus the results are indicative of a high quality annotation process.

-- Ref1.10: Calibrating the age of processed pseudogenes in mouse --

Reviewer Comment	There is actually another way to calibrate the age of processed pseudogenes in mice, by correlating with the presence and absence of a pseudogene in the synteny region in various mouse strains and species (Figure 1).
Author Response	-- ok --maybe add a sentence about that and look at the synteny and count the conserved pseudogenes…. **TODO** **Suggested response:** The reviewer raises a good point that processed pseudogene age could also be evaluated based on the presence/absence of the pseudogene in syntenic regions across the mouse strains

	and species. However, in practice we must use similarity and overlap cutoffs in order to identify pseudogenes conserved across syntenic regions. For this work we have employed a strict 90% reciprocal overlap requirement in order to generate a high confidence set of shared pseudogenes. These cutoffs can influence the number of strains in which a pseudogene is identified as conserved within a syntenic region. Consequently we feel that pseudogene similarity to the parent gene provides a better method for estimating the age of the pseudogene.
Excerpt From Revised Manuscript	

-- Ref1.11: Figure 5C and Section 3.3 --

Reviewer Comment	It may make more sense to discuss and show the processed pseudogenes and duplicated pseudogenes separately since they were created from different mechanisms.
Author Response	Figure created and added
Excerpt From Revised Manuscript	

-- Ref1.12: Unitary pseudogenes --

Reviewer Comment	It is interesting that the authors discovered additional unitary pseudogenes in mice and human. I wonder whether more analysis can be presented on this group. For example, in addition to GO enrichment, do the functional counterpart of these unitary pseudogenes tend to be highly expressed? tend to be non-essential genes ? Also for the mouse unitary pseudogenes, are the null deletion of the counterparts in human more likely to be tolerated ?
Author Response	hmmm maybe --- the reality is that when we look at two species that diverged a considerable time ago, having more knowledge on their annotation by comparison we can identify potentially new unitary pseudogenes in both species and there is plenty of non-coding genome there. The only problem that arises is the correct identification of syntenic regions, otherwise the pseudogenes might not necessarily be unitary but rather duplicated ….

1.12
NO ANSW

	TODO
Excerpt From Revised Manuscript	

-- Ref2.1: Introductory comments --

Reviewer Comment	This paper is both interesting and timely. Many of the raw results are useful and if the manuscript is edited in depth it would be of general interest. The paper is too long and there are too many figures that are not essential. More importantly, there are several major issues with this paper that detract from its potential interest and its usefulness for the research community. The paper is essentially descriptive and focuses on the origin and evolution of pseudogenes in the mouse "lineage" by comparing the results in 18 mouse inbred strains with similar data in humans. Although much of the raw data is certainly useful, the evolutionary analyses are compromised by the lack of proper context. My comments below center in some of the major issues.
Author Response	We sincerely thank the reviewer for kind words and the constructive comments, which we believe made our paper stronger. We respond to the reviewer's comments below.

-- Ref2.2: C57BL/6J vs C57BL/6NJ --

Reviewer Comment	In contrast with the human genomes, the mouse reference genome is mostly the result shotgun sequencing of BACs from a single mouse inbred strain, C57BL/6J (not C57BL/6NJ as shown in Figure 3A. The later distinction is a fairly minor but indicative of the lack of rigor in the use of terms and designations as far as mouse genetics is concerned)
Author Response	We thank the reviewer for stressing the importance of distinguishing between the two strains and the differences between the human and mouse reference genomes. We do make distinction in the text, and following the reviewer's comments we expanded the corresponding section in the text.
Excerpt From Revised Manuscript	Mouse reference genome is based on the Mus Musculus strain C57BL/6J strain. The mouse reference annotation is based on GENCODE vM12/Ensembl 87. The human reference genome annotation is based on GENCODE v25/Ensembl 87.

	The 16 laboratory and wild strains (Table 2) assemblies and strain specific annotations were obtained from the Mouse Genome Project [36] (http://www.sanger.ac.uk/science/data/mouse-genomes-project, last accessed on 21.08.2017). The laboratory strain C57BL/6NJ is a subline of the reference strain [15]. There is high sequence and evolutionary similarity the reference genome single inbred strain C57BL/6J and the laboratory inbred mouse strain C57BL/6NJ. For the purpose of this study and in order to facilitate a reliable comparison across all the studied mouse genomes, we used the laboratory inbred strain C57BL/6NJ as a reference point.

-- Ref2.3: Comparing human and mouse is not fair --

Reviewer Comment	Therefore, comparing content between human and mouse reference genomes is not completely fair as segregating pseudogenes may be absent (or functionally different) in the single chromosome chosen to represent a biological species (M. musculus) that has a much greater sequence diversity and effective population size than humans.
Author Response	fair enough but this argument can also be used against using mouse as a model organism! -- need to address it
Excerpt From Revised Manuscript	

TODO 2.3

-- Ref2.4: Mosaic genomes -- [[EVOLUTIONARY]]

Reviewer Comment	The genome of the C57BL/6J inbred strain is a mosaic of haplotypes with different taxomical origins (M. m. domesticus, M. m. musculus and M .m castaneus; in order of frequency). Although specific overall contributions and the exact genomic locations of these contributions is still under some debate, the consensus view is pretty much settled. The mosaic origin applies to all standard laboratory strains (including the 12 analyzed here). Thus global phylogenies and local phylogenies maybe discordant in many places. The impact of this on Figure 3C is not discussed at all.
Author Response	We agree with the reviewer that on a large scale the mosaicism exhibited by the reference genome and the 12 analysed mouse strains can potentially impact a phylogenetic study. However in our evolutionary analysis we by-passed this potential confounding factor by creating contigs (super genes) from pseudogenes and correspondingly protein coding genes that are conserved across all the 18 mouse genomes. Moreover, the sequences of the conserved elements were concatenated

2.4 AVG. OVER ALL 2.

	together in the same order in all the strain. As such the phylogenetic difference observed would result only from differences in the actual gene sequence and will not be altered by the differences in the relative location of the genes in each of the strains. We comment upon this both in the text as well as in the Methods section detailing on the creation of the phylogenetic trees.
Excerpt From Revised Manuscript	Main text: Next, we took advantage of pseudogenes' ability to evolve with little or no selective constraints [41], and compared mutational processes across the mouse strains. To this end, we built a phylogenetic tree based on sequences from selected from the 3,000 pseudogenes that are conserved across all strains (**Figure 3C**). This pseudogene-based tree follows closely the tree constructed from protein coding genes and correctly identifies and clusters the mice into three classes: outgroup, wild-derived, and laboratory strains. In constructed the phylogenetic trees we concatenated the gene sequences in the same order in all the strains, thus overriding any potential biased induced by the laboratory strain mosaicism, and focusing only on the sequence alterations. Methods: Sequences of the 1,460 pseudogenes were randomly selected out of the total of 2925 conserved pseudogenes in the 18 mouse strains accounting for approximately 50% of the total number of conserved pseudogenes. For each of the 18 mouse genomes, the extracted sequences were concatenated into strain-specific contig (supergene). The order of the pseudogene sequences was kept the same in all 18 contigs. Preserving the same order of pseudogenes or protein coding genes across all strains eliminates any potential bias resulting from the laboratory strain mosaicism, as the relative location of a gene is not considered when creating the trees. Thus the resulting phylogeny is depended only on the sequence evolution. The 18 supergenes were subjected to a multi-sequence alignment using MUSCLE aligner [65] under standard conditions. Similarly, the sequences of parent protein coding genes of the 1,460 pseudogenes were assembled into a strain specific sequence and aligned using MUSCLE. The tree was generated using Tamura-Nei genetic distance model and neighbouring-joining tree build method with Pahari as outgroup using GENEIOUS 10.2 software package [66].

-

-- Ref2.5: Contamination --

Reviewer Comment	All mice used in this study are laboratory strains, including the representatives of the two distantly related species, Mus caroli and M pahari, the representative of the more closely related species M spretus and the so called

	"wild" mice CAST/EiJ, PWK/PhJ and WSB/EiJ. All these mice were bred for many generations in the lab until complete indreeding and at least two of them (CAST/EiJ and PWK/PhJ) are "contaminated" in the sense that they have haplotypes present in other subspecies including haplotypes from standard lab mice.
Author Response	--- add a sentence about how these contamination can interfere with the correctly annotating species specific pseudogenes ...
Excerpt From Revised Manuscript	TODO

-- Ref2.6: Taxonomy and nomenclature --

Reviewer Comment	The entire first paragraph of the Mouse strain section needs rewriting after careful consideration of the taxonomy and nomenclature.
Author Response	We thank the reviewer for pointing out this weakness. We have updated the manuscript using the correct nomenclature for each of the mouse strains, as well as updating Table 2 to reflect the correct nomenclature and terminology.
Excerpt From Revised Manuscript	The Mouse Genome Project has sequenced and assembled genomes for 12 laboratory, and 4 wild mice, and developed a draft annotation of each organisms' protein coding genes [36]. Another two distant Mus species, *Mus Caroli* and *Mus Pahari*, were also sequenced and assembled [37]. Collectively the 18 strains provide a unique overview of mouse evolution. The strains are broadly organized into 3 classes (**Table 2**): the outgroup wild-derived inbred strains – formed by two independent mouse species, *Mus Caroli* and *Mus Pahari*; wild-derived inbred strains – covering the subspecies *Mus Spretus,* and three musculus strains (*Mus Musculus Castaneus, Mus Musculus Musculus,* and *Mus Musculus Domesticus*), and a set of 12 laboratory inbred strains. A detailed summary of the genome composition for each strain is presented in [36].

-- Ref2.7: Distribution of pseudogenes --

Reviewer Comment	Given of of these considerations, the distribution of pseudogenes makes no sense. There should be many more in the wild-derived strains suggesting that the approach of combining a inbred reference genome and the sequence annotation in the more distant strains leads to massive undercounts.
Author Response	We agree with the reviewer that using only the conserved protein coding genes between the reference genome and the more distant strains will result in under-annotating the pseudogenes in those strains. We added a section in the text highlighting this point. However based on the similarity between the mouse reference strain C57BL/6J and the laboratory reference strain C57BL/6NJ, and under the hypothesis that the rate of pseudogene generation is expected to be the same for all the 18 mouse strain, we are able to extrapolate the total number of pseudogene in each strain. The results show that all the strains have on average between 17000 and 20000 pseudogenes. The difference between the number of annotated pseudogenes and the estimate total is the result of a trade-off between the methods specificity and sensitivity. Thus the strict annotation criteria used decrease the number of false positives results at the same time in a decrease of sensitivity. Moreover, the induced reduction in the number of pseudogenes in distant strains does not impact the annotation accuracy, as we observed from Figure 3A that more distant species are enriched in pseudogenes with no direct orthologs in the other strains.
Excerpt From Revised Manuscript	The observed reduction in the number of pseudogenes in the distant species is correlated to the decrease in the number of conserved protein coding genes (between the analysed and the reference mouse genome) used as input in the annotation workflow (**Figure SF-New1C**). However, based on close relationship between the mouse reference strain C57BL/6J and its related laboratory inbred strain counterpart C57BL/6NJ, we are able to estimate the total number of pseudogenes in each of the 18 mouse genomes (**Table S-NEW1C**). The results suggest that all of the studied strains have pseudogene complements of similar size. The difference between the number of annotated pseudogenes and the expected total can be overcome by improving the protein coding annotation in each of the studied strains.

2.7
LOW FP
LIST
V
ACC.
EST.

-- Ref2.8: Using "population" to describe the mouse strains --

Reviewer Comment	P 66 states that there is a large range of divergence in the mouse "population". There several things at play here, but most biologists will disagree that mice from different species (and there are four at play here) or different subspecies (three represented here) are a population in any legitimate sense of the word.

2.8 TO DISC

Author Response	We have rectified the incorrect phrasing of "mouse population" with "mouse lineage" which correctly conveys the meaning of the sentence. NOTE: The main mouse paper has a sentence stating "Inbred laboratory strains are broadly organised into two groups; classical and wild-derived strains, a phenotypically and genetically diverse cohort capturing high allelic diversity, that can be used to model the variation observed in human populations [17317875, 26839397]." We might be able to use these references to respond to reviewer 2's comments about viewing variation across the mouse lineage as analogous to that present in the human population.
Excerpt From Revised Manuscript	While it is hard to make a direct comparison between the two species, there is a large range of divergence in the mouse lineage, with some approaching human-chimp divergence levels in terms of the number of intervening generations.

-- Ref2.9: Figure 7 --

Reviewer Comment	Figure 7 is unreadable
Author Response	We have updated the figure with an improved resolution picture.
Excerpt From Revised Manuscript	

-- Ref2.10: Typos and rewording --

Reviewer Comment	There are several statements that are poorly worded and incorrect on their own. For example in line 146 "there us a considerable overlap, of over 83% between manual and automatic annotation sets" can not be reconciled with number on Table 1. 83% is the overlap with the manual annotation.
Author Response	We thank the reviewer for pointing out these mistakes. The manuscript has been edited to correct them.
Excerpt From Revised Manuscript	e.g. "[33, 34] there is a considerable overlap, of over 83%, between the manual and automatic annotation sets."

DISC

-- Ref2.11: Pseudogene genesis in female germline --

Reviewer Comment	The section on the pseudogenome genesis only considers the female germline and not the male. Given the fact that retrotranspositions is particularly notable among genes highly expressed during spermatogesisis (GAPDH, ALDOA, etc) this seems like a major limitation to reach conclusions.
Author Response	
Excerpt From Revised Manuscript	

-- Ref3.1: Introductory comments --

Reviewer Comment	This paper describes the annotation and analysis of pseudogenes in the genomes/transcriptomes of 18 mouse strains. Comparisons are made with the human genome and between specific mouse strains. The transcription of pseudogenes and parent genes is examined in different contexts using phase 3 ENCODE transcriptional data. There is a lot of data analysis in this paper, and the most interesting aspects get somewhat lost or are not highlighted enough. I think that the key observations are the large numbers of strain-specific pseudogenes linked to specific biological processes, the greater strain specificity of pseudogene expression compared to protein-coding gene expression. The greater expression of retropseudogene parent genes at later stages of embryonic development is also potentially very interesting, but I think further work may have to be done on that point.
Author Response	We sincerely thank the reviewer for kind words and the constructive comments, which we believe made our paper stronger. We respond to reviewer's comments below.
Excerpt From Revised Manuscript	

-- Ref3.2: Originality and significance --

Reviewer Comment	This is certainly an original analysis. The key observations are of general significance to researchers who use the mouse as a model organism. Obviously, the annotations themselves are a significant resource for investigating strain-specific responses to diet, disease, etc. However, the authors should be more careful in identifying what are confirmatory observations of results that have been previously published, e.g., the most common families, the derivation of processed pseudogenes from highly transcribed genes, the amount of pseudogenes that are transcribed.
Author Response	WG Make sure all the previous analysis are cited and we highlight what is new. 3,2 WORD BETTER
Excerpt From Revised Manuscript	

-- Ref3.3: Data and Methodology --

Reviewer Comment	The approaches are valid, and the annotations seem of sufficient quality. The figure presentation is of a high standard, except that: (i) the resolution of Figures 5A and 6B are much too low. (ii) I cannot follow what Figure 7C is representing. There are a lot of very faint grey lines, and there is a red oval around an area of white space labelled 'strain-specific transcribed pseudogenes'; it is not clear to me what this means. Could the authors make this figure clearer and explain it in the legend?
Author Response	Improve resolution on the figure and give a better explanation of the circos plots
Excerpt From Revised Manuscript	

-- Ref3.4: Pseudogene genesis --

Reviewer Comment	Presumably in section 3.1 'Pseudogene genesis', they are referring to Figure SF4 when they talk about parent gene expression during development. Also, this figure does not have any fitted lines or correlation coefficients,

	the authors just say in the text '...the correlation improves...".
Author Response	Add correlation coefficients to the sup figure.
Excerpt From Revised Manuscript	

-- Ref3.5: Pseudogene genesis during early development --

Reviewer Comment	Data for earlier stages of oocyte and sperm development are missing from this analysis. Is this data available? It would make a more interesting discussion to include it. Furthermore, couldn't the correlation that the data is converging to be identified? Then, interesting outliers relative to this correlation at earlier developmental stages could be picked out.
Author Response	Get more data …..PM.
Excerpt From Revised Manuscript	

-- Ref3.6: Abstract --

Reviewer Comment	The abstract does not really do justice to the most interesting observations in the paper, and there are several confirmatory observations in the abstract that are not identified as such (e.g., the most common families, the amount of pseudogenes that are transcribed, the derivation of processed pseudogenes from highly transcribed genes).
Author Response	Rewrite abstract ….
Excerpt From Revised Manuscript	

-- Ref3.3: Pseudogene as marker of genome remodeling --

Reviewer Comment	Pseudogenes as 'ideal markers of genome remodeling...'; I am not sure what the authors mean by this. Is this really

LIKE PROT COD.
BUT NO SEL.

	borne out by the analysis in the paper? How exactly do pseudogenes mark genome remodelling events?
Author Response	Clarify
Excerpt From Revised Manuscript	