<u>1st Round of Reviews</u> (submitted to J Neurosci Nov 17, 2021; reviews Dec 17, 2021)

Thank you for the reviews. We have re-analyzed and/or changed graph formats to improve readability, we have included more statistical details (effect sizes, statistics summary table, full dataset), and we have made numerous smaller changes all of which are highlighted in the article file that shows the tracked changes. Please see our response to each concern and suggestion below, in bold.

Reviewer #1 (Rationale for Significance Rating for Authors):

This is an interesting article presenting data from studies designed to explore potential sex differences in the function of adult-generated neurons in spatial navigation learning in rats. The authors show that adult-generated neurons improve learning under conditions of stress in males and have the opposite effect in females.

Reviewer #1 :

This is an interesting article presenting data from studies designed to explore potential sex differences in the function of adult-generated neurons in spatial navigation learning in rats. Using transgenic rats that permit the inhibition of neurogenesis in adulthood with valganciclovir treatment, the authors showed different effects in males and females but only when spatial memory was assessed under aversive conditions (cold versus warm water). Females show improvements in spatial learning under stress in the absence of new neurons, while males show impairments. The authors also present evidence showing sex differences in immediate early gene activation in the dentate gyrus, as well as sex differences in the influence of learning under aversive conditions on dendritic morphology of new neurons. Overall, the findings are novel and raise interesting questions about sex differences in the role of new neurons in stress regulation. My comments for improving the manuscript are listed below:

1) I do not think that a graph showing that few studies have addressed these issues is a strong way to start this paper. I think stating that the literature lacks these types of studies is sufficient. Figure 1 should be removed.

We do agree it could be sufficient to simply state that little is known about sex differences in the function of neurogenesis, but do not understand if/why this detracts from the manuscript. Papers often begin with general statements that are meant to justify the work, such as "a small number of newborn neurons" or "little is known about topic X". But these claims are often subjective and even the opposite claims could be made depending on perspective. And so here we performed a quantitative analysis in order to provide objective data that could definitively justify our research (and could also facilitate future research). From an empirical, data driven standpoint we therefore feel that this figure only strengthens the manuscript. But if there are reasons why this figure detracts from the manuscript we are certainly open to them. In the revision we have continued to include these data.

2) The figures are organized to highlight comparisons between TK and WT rats treated with valganciclovir (Figures 3 and 4) which raises questions about whether the effects are related to the genotype or the genotype + valganciclovir. The legends of these figures don't mention valganciclovir, which is confusing. Furthermore, data from TK and WT rats not treated with ganciclovir is presented in a different set of figures, which makes comparison very difficult. It

would be better to organize the data so that WT and TK with and without valganciclovir were presented on the same figure. Key comparisons could be made across sexes as well.

These are all good points and, in principle, we agree that it would be ideal to have data from WT and TK rats, over days, both with and without valganciclovir present in the same graphs and analyses. Of course it would also be ideal to include both sexes. However, our valganciclovir-treated and untreated rats were not tested at the same time. Since temporal gaps can lead to baseline fluctuations in performance, we did not include valganciclovir treatment as a 4th factor in our ANOVA, and are wary of making direct comparisons between untreated and valganciclovir-treated rats, since this would increase the likelihood of making a type 2 error. Since we did not observe any genotype differences in rats that were not treated with valganciclovir, we hope this provides reasonably strong support for the interpretation that behavioral differences are not due to nonspecific genotype effects.

To make the group treatments more clear, in the revision we now state in the figure caption and figure legends where rats were untreated or treated with valganciclovir (eg "TK male$_{VAL}$" and "WT female$_{UNTREATED}$").

3) Related to the comment in number 2, the methods section should be clarified to make it clear that within each sex, each genotype was divided into two groups - one treated with ganciclovir and the other not treated with it.

We have adjusted to methods to highlight the fact that separate groups of males and females were untreated vs treated with valganciclovir, and have explicitly linked these groups to their respective figures in the methods.

4) This transgenic model also inhibits neurogenesis in other regions. This should at least be mentioned in the manuscript.

We have included this point in the opening of the discussion.

5) The inset image showing a fos positive cell in Figure 6 is not convincing. A better example should be provided.

We agree that it would be nice to include a GAD+ cell with stronger Fos labelling (and could possibly find one). However, to be honest, this is representative of what we see. We are able to obtain very strong Fos labelling in excitatory neurons of the DG and CA3 (and elsewhere), but inhibitory neurons did not express Fos at such high levels (this may fit with others' findings, such as those by Vazdarjanova JCN 2006 who found that another IEG, Arc, is not expressed by inhibitory interneurons at all). Given the weaker expression, we quantified Fos staining intensity in GAD+ neurons and only counted cells that had levels twice background, in order to objectively quantify more modest expression. We note these points in the revision.

6) The wrong figure is mentioned in the figure legend of Figure 4 - it should be Fig 3, not Fig 2.

Thanks – this has been fixed.

7) I am a little confused about the corticosterone results because they don't seem to support the

claim that cold water is aversive to the rats. A better explanation of those results should be provided.

Yes, the interpretation would be easier if there were clear differences HPA activation between rats trained at the 2 temperatures. We have expanded this portion of the discussion to consider why this is the case, and speculate on other stress-related changes that could be responsible (eg. catecholamines)

8) Latency to reach the platform is highlighted in many of the figures as evidence for learning. Because of the potential confounds of swimming capabilities, it would be useful to highlight path length as well.

Thanks. Yes to address this we also analyzed path error (which is not confounded by differences in speed and distance traveled) but primarily focussed on the latency data as a measure of learning. In the revision we present a more balanced assessment of both latency and path error as evidence for learning.

Reviewer #2 (Rationale for Significance Rating for Authors):

Overall, the interaction between sex, stress, spatial learning, and neurogenesis is an interesting topic, although the specific directions of interactions that the authors are studying here is not clearly articulated. Of more concern is the difficulty in following the precise methods used and analyses conducted, some important control groups missing, and overall an over interpretation of the findings (particularly as it relates to sex).

Reviewer #2 :

Many of the figures are very difficult to follow. For example: in Figure 1, the % of studies seems to add up to more than 100%;

Yes – this was because studies could fall under multiple categories. For example, studies that examined both males and females were counted under both “male” and “male & female”. Upon reflection we agree that it is more intuitive to categorize studies as “male only” and “female only” and so we have reanalyzed and regraphed these data (such that male only + female only + male&female + unspecified = 100%). We have also adjusted the methods, figure legend and figure caption clarify how the data were collected and analyzed, and what the bars represent.

In figure 5, if there is a ~17% increase in "random search" in TK males, surely there should be a compensatory decrease in other strategies?

Yes! For some reason we decided upon an unnecessarily complex difference score, where the %TK-WT change was normalized to WT levels and also weighted according to the proportion of trials where a given strategy was used. And so the x-axis units were not % points, and increases in one strategy were not visually/obviously balanced by an equal reduction in other strategies. Clearly we overthought this and so in the revision we have adjusted the graphs to show the suggested difference scores (ie %TK trials -%WT trials). The take home message is the same, but the new graphs are clearer and more intuitive.

In figures 3 and 4, please show males and females on the same graph so that the data can be compared, since the analysis includes sex as a variable. In figures 3c and F and 4C and F, why

are categorial variables (sex) displayed as line graphs, and why are the Y-axes truncated? These make the data difficult to follow and understand.

As discussed above, we do agree that males and females could/should be presented in the same graphs. At least to some extent since we presume readers will be interested in direct comparisons between the sexes. However, many previous studies have compared spatial learning between males and females and so direct comparisons between the sexes was not our main objective. Rather we were interested in effects of neurogenesis ablation. For the latency and path error data over days, putting both sexes and both genotypes on the same graphs actually makes the data more difficult to follow, because the lines become crowded. For this reason we presented the same data 2 ways, one to show the genotype differences (performance over the 3 days of training) and another to show sex differences and the interaction between sex and genotype (where we graphed averaged data to minimize visual clutter associated with training day). Also, for the graphs of sex x genotype (now Fig 3C,F,J,M) we used lines to visually connect groups of the same genotype and generate slopes that can provide an intuitive sense of sex x genotype interactions (or the lack thereof). The y-axes in these graphs are truncated to focus on relative group differences (but of course the axes are labelled and readers can extract absolute values as well).

In the revision we have clarified that our main objective was to examine effects of neurogenesis ablation within the sexes. We also have provided our full dataset as supplementary material should others wish to examine sex differences in more detail, include these data in meta analyses, etc.

The large n in studies throughout the paper is laudable, but please also include effect size - the size of effects seem small, and this is important information to include in any analyses.

This is a good suggestion. We (and most in the field, for that matter) have not been accustomed to using effect size measures but we appreciate their value. In the revised manuscript we have included partial eta squared as an effect size for our ANOVAs, and have included Hedge's g as an effect size for post-hoc comparisons of 2 groups. Given the debate about optimal effect size measures (eg Lakens 2013) we have included our full dataset as supplementary material should others wish to examine the matter in additional detail.

Did exposure to the 16 degree water trigger HPA axis activation in any mice? There is reference in the discussion to corticosterone data but none that I could find.

Yes this is in Fig 9. Previously it was in the extended data (Fig 1-7).

The main "sex differences" seem to be in Figure 3 - time in target zone. And yet, this is a 2 second difference, which though significant, is unlikely to be meaningful, especially with proportion of time in target vs other quadrants similar in males and females **[actually, the greater time spent in the target zone does reflect a true spatial bias, and is not an artefact of males simply spending more time in both the correct and incorrect zones (effect of sex on Target–Other difference score, P = 0.03)**]; and indeed, total time shown on the graph is lower in females compared with males. It is also not clear why is time in the target zone/time in probe test so low? It seems that males spend more time in the maze (6+3+3+2 =14 seconds; vs 4+2+1+2 = 8 seconds)? Even if this is in a maze region more narrowly defined than an entire quadrant, the interpretation of this "difference" should be more nuanced - especially as this

effect does not replicate in the experimental data in figure 4 - where there is no significant difference between WT males and females in time in target area.

Where we observed the greatest/main sex differences is an interesting question, and we agree that differences are not necessarily meaningful just because they reach a statistical threshold. Generally, we didn't focus extensively on sex differences in traditional measures of water maze performance, since these have already been studied extensively by others (for example, we didn't mention these 25C probe data outside of the results). We did find that males showed greater spatial learning according to traditional metrics, in several but not all experiments. This is in line with the pattern observed in the literature, the fact that some degree of replication failure (eg 20%) is to be expected even in well-powered experiments, and suggests that there is a real sex difference in spatial behavior.

The use of a zone that is smaller than a quadrant is fairly standard practice (eg Morris, 1982, Nature; Kee, 2007, Nat Neurosci) as it allows for a more precise measure of spatial behavior. Is a 2 second difference in 25C probe trial performance meaningful? Given that it is comparable to chance performance (2.4 seconds) females searched at ~twice chance levels and males search at ~three times chance levels, which is arguably a meaningful standardized difference (alternatively, sex explained 9% of the variance in the 25C probe trial). Looking deeper, we observed that estrus stage dictated probe performance at 25C (where we observed a sex difference) but not 16C (where we observed no significant sex difference), and so perhaps these data may be useful for future investigations into temperature-dependent interactions between hormones and behavior. More pertinent to our study, we found that neurogenesis differentially regulated strategy choice depending on sex. Thus, differences in strategy, which may only result in modest differences in acquisition latency & probe search times (because of the myriad ways in which the water maze can be solved), could result in very meaningful differences in tasks that are highly dependent on which strategy is employed (eg navigational or choice tasks that have fixed response options). In sum, we tried to perform a comprehensive analysis of the behavior to understand the reason for the sex differences that are observed with traditional metrics. We hope that with continued investigation we can come to a better conclusion of how meaningful these differences are.

In Figure 7 - is the effect on dendritic spines training or temperature exposure - I could not see reference to a group exposed to 16 degree water, but no spatial training (a cued group would be an ideal control here), and since these changes were observed only in the 16 but not 25 degree training groups, the obvious alternative hypothesis is that these changes are stress, not changing, related.

We agree and think that multiple additional experiments need to be performed to settle this. We have therefore minimized claims about it being learning-specific, but instead propose that this plasticity may promote learning under stress (even if it is triggered solely by the stress of cold water).

The method for the literature review needs additional detail. If I do a pubmed search for "neurogenesis" and "dentate gyrus" - the search terms described in the methods - I get 4398 results in the 2001 -2020 time range (with about 540 review articles). How were the 76-112 studies per 5 year bin selected? What were exclusion/inclusion criteria?

We have expanded the methods section to include these details (in short, ~20 primary research articles on mammalian adult neurogenesis were randomly selected per year). We have also included the full list of studies in the dataset file.

Some minor but important notes on terminology:
Sex (the noun) does not modulate hippocampal plasticity, although sex (the verb) might. There are, however, sex differences in hippocampal plasticity.

Good point – we have fixed this.

When talking about humans and human disorders, the term "women" is more appropriate than "female" as it takes gender into account too

In the introduction we have changed “female” to “women” and “male” to “men”.

<u>2nd round of reviews</u> (resubmitted to eNeuro Feb 3, 2022; reviews March 7, 2022)

Please see our responses below, in bold.

- Extended Data should be labeled as Figure 1-1, Figure 1-2, Table 1-1, etc., so they indicate which figure they are supporting (i.e. Extended Data table supporting Figure 5 labeled as Figure 5-1).

- Extended Data figure/table should be referenced in the legend for the figure/table it is supporting. Please add a reference to the Extended Data figure/table in the corresponding main article figure/table.

In "Preparing a manuscript" [https://www.eneuro.org/content/preparing-manuscript] it states "Extended data that supports more than one figure and/or table should be labeled as supporting the figure or table referred to first in the text." It sounds like this applies to our situation, since our extended data file contains the data that supports all of the figures. We therefore refer to it as Extended Data Fig. 1-1. We now state at the beginning of the results, and in the legend for the first figure, that all of the underlying data and statistical analyses for all figures can be found in this file. Is this an acceptable approach? This is also how we provided, and referred to, the underlying data in Cole et al., 2020, J Neurosci. Alternatively, we could break the data file up into discrete files that only support individual figures but this would result in 11 different data files and would make accessing the data more cumbersome.

Synthesis Statement for Author (Required):
This manuscript is a revised version of a paper reviewed at the Journal of Neuroscience and transferred to eNeuro. The reviewers of the revised paper felt it provides interesting new data regarding potential sex differences in the role of new neurons in stress-modulated spatial learning and morphology of newborn cells in rats. Although it was felt that conclusions are limited by the model used, the data set was thought to be useful and appropriately interpreted. Overall, the reviewers thought that this is a much improved, much clearer version than the previous submission, and many of the major critiques were addressed. However, they felt that several issues still remain, as detailed below.

Before describing those issues, however, I would like to comment on the unusual way in which statistical analyses are presented and ask the Results be edited to provide a more conventional presentation of significant results. It appears that all statistical results are written into the Figure Legends and contained in the Extended Data file, leaving few in the Results section. As a reader, I find it frustrating and challenging to go searching for the statistical results on which the author's conclusions rest, and prefer not to have to dig through figure legends of at least half a page. Typically, significant stats are located within the Results section as a way to support the author's conclusions about the data. Thus, I request that you find a way to incorporate significant findings into the Results section if at all possible. If there is some compelling reason why this is not possible, then please make that case in your rebuttal.

We don't have a perfect answer or solution for this since spreading results over the text, figure and figure legend will always require one to navigate between these 3 items (often on different pages) to fully absorb the data. We appreciate that putting all of the statistics in the results text makes for a coherent package of descriptive and statistical evidence. But visual inspection of the data itself is also important for readers to assess conclusions that are present in the results text. Typically, our preference (in papers

we've published at SFN journals and elsewhere) has been to keep the figures and analyses together, so one can directly relate patterns that are visible in the data with their corresponding statistical analyses. In some cases we have found that extensive statistical reporting in the results text makes for challenging reading. And so this keeps the results text easy to read. We don't know which style is more common (and likely there isn't a one-size-fits-all approach here) but, in skimming our recently-read articles, we often find statistical analyses in the figure legends (a couple of examples are very recently published articles on sex differences that we have now cited in the revised manuscript: Waters, Gould et al, 2022, Neurobiol Stress; Le, Lynch et al, 2022, Nat Neurosci). And so, since making this change wouldn't solve the problem navigating between text, figure and legend, and since this style preference doesn't affect the validity of our analyses or interpretations (but it would involve a fair bit of work), we would prefer to keep many of the statistical analyses in the legend alongside the figures.

If you include the Extended Data file in your revision, then you must clearly indicate in the Results section which data are included in this file and make sure to reference the file throughout the Results. In addition, although I appreciate the transparency of the Extended Data file, it appears that at least some of its contents are already presented in graphical or text form in the figures and figure legends. If so, it is necessary to also include that information in the Extended Data file?

We have put all of the underlying data in the Extended Data file and so in a sense this duplicates what is present in the figures, except the actual data is more precise. We feel it is important to include the actual data for reasons that are aligned with eNeuro's mission to promote statistical rigor and reproducibility. In our original reviews we were asked to provide effect size measures. In researching this practice we learned that there are many different types of effect sizes, that some are better (eg less biased) than others, but that tools and methods for calculating some effect sizes for certain experimental designs are not yet readily available (eg see Lakens, 2013, Frontiers in Psychology). We therefore opted to report partial eta squared because, while it is not perfect, it is widely-used and easy to calculate. However, by providing the underlying data, readers can compute their own effect size measurement as tools become available and norms change. This may be particularly important for future meta-analytic studies of neurogenesis and sex differences, since effects are variable across studies and theory is best assessed from a body of work.

In addition to the underlying data, the Extended Data file also contains the full statistical analyses for all experiments. Some of these are duplicated in the results text or figure legends. However, many are not duplicates and so for consistency's sake, and to provide a resource that allows readers to easily access any of the statistical results (but minimally increases file size), we opted to simply put all of the statistical results in the Extended Data file.

Additional reviewer comments are as follows:

1. "Critically, these disorders affect women to a greater extent than men, suggesting that 62 neurogenesis functions in stress may vary depending on sex and gender (Kessler et al., 2012)" This is not logically consistent. Together with the data in rodents? Maybe?

Agreed, "together with the data from rodents" was what we originally intended but this may have been lost as the rodent literature was discussed in the previous paragraph.

We have now changed it to "Stress-related disorders such as anxiety, PTSD and depression impact a substantial fraction of the population and these disorders affect women to a greater extent than men. Together with the data from rodents, this suggests that neurogenesis functions in stress may vary depending on sex and gender (Kessler et al., 2012)."

2. One reviewer felt that Figure 1 should be deleted and the findings incorporated into the text of the manuscript. They thought that showing that information as a display item is more appropriate for a position paper than to provide rationale for a research article. On this point, I agree. If you do decide to publish this figure elsewhere, the other reviewer had the following questions/comments: "The sum of Male only, Female only, Male & Female, and Sex unspecified should add to 100%, correct? It would then make sense if "reported data by sex" was reported separately (or at least after the other measures) since this is a separate but related issue. In addition, when you say "reported data by sex" do you mean use sex as a variable, or do you mean disaggregated by noting male and females on individual data points/examining means/ranges for males and females separately? This would be an interesting statistics to add. The graph with the additional methods now makes sense. It is extremely frustrating to see that the proportion of studies using males only is roughly the same while the proportion studying "females only" decreases as studies including both sexes increasing. "

We removed this figure from the manuscript and have published it on Figshare (https://figshare.com/articles/figure/Adult_neurogenesis_studies_primarily_use_males/19319849) and now only refer to it in the text. For the record, we liked the idea of putting the "reported data by sex" after the other measures.

3. Figure 2 images do not reflect the quantitative data shown on the graphs. A very large difference in DCX cells is shown in the graphs and the images between groups look almost identical.

Good catch. In making the high resolution figures somehow the WT image was duplicated in place of the TK image. This has been fixed.

4. In Figure 3, noting on the figure which are cold and which are warm water tests would be helpful (or a key noting the meaning of the blue and red color coding)

This has been added to the figure (large text on top of each group of graphs that states "16C" or "25C").

5. Graphs showing sex differences would be more accurate if they were not shown with lines connecting the male and female data. This gives the impression of some continuity between the groups and is misleading. These data should be shown as bar graphs.

This has been changed.

6. Individual data points should be shown for behavioral graphs.

This has been changed (except for data over the 3 days of training, since the graphs become too muddled).

7. The species should be mentioned in the title.

Done.

8. Throughout the paper, the authors refer to "neurogenesis" modulating effects. It seems overly general to say that a process is modulating an effect when what it seems they mean is that immature neurons are modulating the effect. The authors should consider changing this wording.

Done.

9. Line 578 - please temper the interpretation of fos-activity reflecting overall activity (here and elsewhere). You can say that decreased neurogenesis did not alter activation of fos, but not all neurons/activity increases fos. It's suggestive, but not deterministic. Especially, as you note in your response to reviewer 1, that IEGs have different thresholds and likelihood for activation states.

Agreed, and we have added another statement to this effect, on the fact that there are other forms of neuronal activity that could have changed/could be investigated.

<u>3rd round of reviews</u> (resubmitted March 16, 2022; reviews April 12, 2022)

Please see our responses below, in bold.

Thank you for your thoughtful response to the reviews. I now have a better understanding of your presentation of the results and find the revised text more clear in describing the location of the statistical analyses. The way in which you have graphed the Extended Data will work--no need to replicate multiple times.

Although the reviewers felt you had addressed most of their concerns, one reviewer has asked that you modify your language about neurogenesis in the title and abstract as per their original comment. The specific comment on the revision was as follows: "Although the reviewers agreed to change the word "neurogenesis" to something more specific (e.g., immature neurons, new neurons, adult-generated neurons) when the goal is to describe an actual mechanism rather than a process, there are still several important places where it has not been changed, including in the title and the abstract.". Thus, I would ask you to modify the title and abstract to reflect the reviewer's concern.

In the revised version we have changed the title, abstract and introduction to use terms such as "adult-born neurons" instead of "adult neurogenesis". The changes can be seen in the marked up version of the manuscript file.

Resubmitted April 12, 2022 and accepted April 15, 2022. Yay!