

An integrative systems-biology approach defines mechanisms of Alzheimer's disease neurodegeneration

Corresponding Author: Dr Ernest Fraenkel

This manuscript has been previously reviewed at another journal. This document only contains information relating to versions considered at Nature Communications.

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #2

(Remarks to the Author)

The revised manuscript has improved in clarity since its initial submission, and the emphasis on linking the genetic screen to aging has been toned down. While the integration of the multi-omics datasets is still not as clear, the study makes a good effort to advance findings from the genetic screen through functional experiments, providing insights into underlying mechanisms. The results largely confirm previously established pathways, such as DNA damage and cell cycle re-entry, but reinforce the importance of these pathways in neurodegeneration.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)

I appreciate your extensive revisions to the article in response to my many pointers. You clearly have the technical skills to perform these difficult but necessary checks on robustness, which I think sets this work in its updated state apart from 90% of systems biology work, and will ultimately serve to help the field narrow down to effective targets more quickly. Moreover, it's great to see a team, such as the one on this paper, making insightful findings about the means through which cells carry out their molecular functions.

There's a key point in your intro (in terms of your global motivation) where you say: "Despite the wealth of human pathological and genetic data, the pathways through which both single gene mutations and QTL-associated molecular changes impact neurodegenerative disease pathogenesis remain incompletely defined". I think the "incompletely defined" is itself too vague at this point - pretty much all of science could fall under that label, and there must be 100+ papers now trying to find overarching/systems level info from human GWAS, so you probably have to be more specific about what they're failing to do.

In figure one where you have "hypothesis generation" that reminds me of the Sydney Harris comic strip where for a key step "then a miracle occurs" - so perhaps you could be more descriptive on this stage in the figure?

You've done a great job of strengthening the robustness of results in several ways, but one seems untouched: "We do not believe that adding additional transcriptional data will substantially improve the manuscript." Any particular data-based, demonstrated reason for that belief? There was one marker gene you pointed to, but the overall similarity and any impact on downstream results doesn't seem completely determined by that. Furthermore, majority of human cell type specific data is coming out of single cell RNAseq, those papers are highly cited, so it seems logical to use some of those signatures to prove robustness of results to the AD definition and explore if particular subtypes of neurons are more or less relevant. In fact if you use the ROSMAP samples for this you could segment out cognition, tau and amyloid genes, which might help address some of the minimal overlap discussed later.

"the lack of overlap" [between degeneration and tau screens] "is a phenomenon that has been noted previously in other comparisons of genetic and expression data". It's been noticed before, so apparently it's robust - why do you think that's the case? Think this issue bears discussion as integration is a central function of the paper. Relatedly," Furthermore, we observe that more neurodegeneration screen hits show proteomic and phosphoproteomic changes in A β 1-42 transgenic flies than in tauR406W transgenic flies." - I don't see a reference, figure or table cited for this claim.

Both reviewer two and I found regarding the if claims of detecting age-related genes are fully justified without interim measurements. The core of your response appears to be: "the very substantial scale of our screen (5,261 genes) made systematic analysis at multiple time points for the entire cohort of genes tested infeasible. have documented progressive neurodegeneration over time, and using additional markers of neurodegeneration, on selected candidates emerging from the screen, including fly APP and presenilin" There appear to be avenues to directly test this and to avoid possibly cherry-picked cases - you could either pick a random subset of genes that you predict to show aging progression through experiments, or you can perform some sort of structured literature review for all genes with some sort of aging ramp and show that your set is enriched in those. Just saying you've done a lot (true) and pointing to a few genes that have the expected temporal trajectory (also true) just doesn't address the larger issue of the generalizability of claims that this extends to the full set, unless you did something along this lines of what I've proposed, or some other approach you come up with that answers the issue robustly.

With regard to the discussion across several points of selection of specific targets for validation, and obtaining false positive rates, what would be very convincing is to take the highest ranked gene with absolutely no prior evidence for Alzheimer's disease and show you get convincing phenotypes out of that. I see the point of your response that you're looking for pathways rather than prioritizing molecules, but I think the #1 thing any experimental lab or pharma group is going to ask given the pathway is what they should be targeting in light of this, so then we're back to the same issue, and you could really display the transformative power of your approaches by showing genes with no prior evidence can produce disease phenotypes.

(Remarks on code availability)

The code is of above average quality in terms of organization and helpful in determining exactly what was done.

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

I appreciate your extensive revisions to the article in response to my many pointers. You clearly have the technical skills to perform these difficult but necessary checks on robustness, which I think sets this work in it's updated state apart from 90% of systems biology work, and will ultimately serve to help the field narrow down to effective targets more quickly. Moreover, it's great to see a team, such as the one on this paper, making insightful findings about the means through which cells carry out their molecular functions.

We thank the reviewer for recognizing the efforts we have made to improve the manuscript.

There's a key point in your intro (in terms of your global motivation) where you say: "Despite the wealth of human pathological and genetic data, the pathways through which both single gene mutations and QTL-associated molecular changes impact neurodegenerative disease pathogenesis remain incompletely defined". I think the "incompletely defined" is itself too vague at this point - pretty much all of science could fall under that label, and there must be 100+ papers now trying to find overarching/systems level info from human GWAS, so you probably have to be more specific about what they're failing to do.

We thank the reviewer for the suggestion and have revised the text as follows.

"Despite the wealth of human pathological and genetic data, the pathways through which both single gene mutations and QTL-associated molecular changes impact neuronal cell death remain incompletely defined (Page 2)."

In figure one where you have "hypothesis generation" that reminds me of the Sydney Harris comic strip where for a key step "then a miracle occurs" - so perhaps you could be more descriptive on this stage in the figure?

We have revised the figure to be more descriptive of our process for choosing nodes and subnetworks of interest.

You've done a great job of strengthening the robustness of results in several ways, but one seems untouched: "We do not believe that adding additional transcriptional data will substantially improve the manuscript." Any particular data-based, demonstrated reason for that belief? There was one marker gene you pointed to, but the overall similarity and any impact on downstream results doesn't seem completely determined by that. Furthermore, majority of human cell type specific data is coming out of single cell RNAseq, those papers are highly cited, so it seems logical to use some of those signatures to prove robustness of results to the AD definition and explore if particular subtypes of neurons are more or less relevant. In fact if you use the ROSMAP samples for this you could segment out cognition, tau and amyloid genes, which might help address some of the minimal overlap discussed later.

We thank the reviewer for encouraging us to revisit this question. We performed differential expression analyses in excitatory neurons from previous a previously published RNA-sequencing study. We identified differentially expressed genes with respect to cognitive decline, Braak stage, amyloid pathology and Alzheimer's disease status¹. The results of this analysis have been included as a table in Supplementary Table 7 and in the revised manuscript in the tab "Disease_phenotype_enrichments".

We found that the subnetworks associated with postsynaptic activity and insulin signaling were enriched for genes that were differentially expressed with respect to Alzheimer's disease status, amyloid pathology, and Braak staging. The subnetworks associated with NOTCH signaling and condensation of prometaphase chromosomes were enriched for genes differentially expressed with respect to Braak staging and Alzheimer's disease status. We note that the genes we tested experimentally (*HNRNPA2B1*, *NOTCH1*, *CSNK2A1*, *MEPCE*) all belong to a cluster enriched for differentially expressed genes associated with Alzheimer's disease pathology. This analysis strengthens the relevance of our findings to Alzheimer's disease.

"the lack of overlap" [between degeneration and tau screens] "is a phenomenon that has been noted previously in other comparisons of genetic and expression data". It's been noticed before, so apparently it's robust - why do you think that's the case? Think this issue bears discussion as integration is a central function of the paper.

The reviewer suggests we comment on the basis for relative lack of overlap of genetic and expression approaches. As we suggested in prior work, possible explanations include robustness of redundant pathways to knockdown of individual proteins that nonetheless change in expression, and lack of differential expression in proteins regulated by posttranslational modification². As requested by the reviewer, we have included these possibilities in our manuscript. We agree with the reviewer that integration of different data sources with sensitivity for diverse types of signals is a strength of our study.

Relatedly," Furthermore, we observe that more neurodegeneration screen hits show proteomic and phosphoproteomic changes in A β 1-42 transgenic flies than in tauR406W transgenic flies."- I don't see a reference, figure or table cited for this claim.

We apologize for this oversight and have added a citation to figures 3d and 3e after sentence indicated above.

Both reviewer two and I found regarding the if claims of detecting age-related genes are fully justified without interim measurements. The core of your response appears to be: "the very substantial scale of our screen (5,261 genes) made systematic analysis at multiple time points for the entire cohort of genes tested infeasible. have documented progressive neurodegeneration over time,

and using additional markers of neurodegeneration, on selected candidates emerging from the screen, including fly APP and presenilin" There appear to be avenues to directly test this and to avoid possibly cherry-picked cases - you could either pick a random subset of genes that you predict to show aging progression through experiments, or you can perform some sort of structured literature review for all genes with some sort of aging ramp and show that your set is enriched in those. Just saying you've done a lot (true) and pointing to a few genes that have the expected temporal trajectory (also true) just doesn't address the larger issue of the generalizability of claims that this extends to the full set, unless you did something along this lines of what I've proposed, or some other approach you come up with that answers the issue robustly.

We apologize that we did not more clearly state our results and appreciate the suggestion that we expand our study by including a structured literature review. As suggested, we performed a structured literature review to identify *Drosophila* genes with significant changes in gene expression and proteomic abundance in fly brains and neurons³⁻⁷. In our literature review, we used the search terms "age-associated genes in *Drosophila* brain" in google scholar and included all papers that measured aging-related changes in proteomics or RNA-seq data. We supplemented this list with papers found using the following search terms in a PubMed advanced search: "transcriptomics and drosophila and brain and neuron and age", "RNA-seq and drosophila and brain and neuron and age", "proteomics and drosophila and brain and neuron and age", and "mass spectrometry and drosophila and brain and neuron and age". We note that the preprint of our manuscript was the first result when using the search terms "proteomics and drosophila and brain and neuron and age" after sorting the results by "best match". We included studies where the results of a differential expression analysis with respect to age or a regression analysis of expression with respect to age in *Drosophila* were reported. All included studies were performed in *Drosophila* whole brains or *Drosophila* neurons in either brains or photoreceptors. We included photoreceptor neurons since hits from the neurodegeneration screen included genes with evidence of retinal neurodegeneration. We excluded studies that performed limited analysis of candidate genes with respect to aging. We also excluded studies that did not use sequencing-based methods for transcript profiling such as microarrays and excluded proteomic studies that did not use mass spectrometry for protein quantification. One study was excluded because the table that reported differential proteomic abundance was unavailable⁸. Other differential expression studies were excluded if they used only a fold-change cutoff without including an FDR-adjusted p-value cutoff^{9,10}. We extracted the tables provided from the studies that met our search criteria and removed duplicate gene identifiers to construct our list of genes with an aging ramp. We found that this set of genes was enriched for neurodegeneration screen hits (hypergeometric test p-value=2.87*10⁻³). We have added this result to the revised manuscript. We further have revised our manuscript to include the full complement of genes, in addition to APP and presenilin, recovered in our screen that have been previously associated with age-dependent neurodegeneration in *Drosophila*.

We also show in the manuscript that the expression of the human orthologs of the screen hits are associated with chronological age. We performed a Gene Set Enrichment analysis where genes were ranked by the correlation between their expression and age. We found a negative association between the expression of screen hits and age (Fig. 2b). This result and our systematic literature review support our claim that the hits from the neurodegeneration screen are age-associated.

With regard to the discussion across several points of selection of specific targets for validation, and obtaining false positive rates, what would be very convincing is to take the highest ranked gene with absolutely no prior evidence for Alzheimer's disease and show you get convincing phenotypes out of that. I see the point of your response that you're looking for pathways rather than prioritizing molecules, but I think the #1 thing any experimental lab or pharma group is going to ask given the pathway is what they should be targeting in light of this, so then we're back to the same issue, and you could really display the transformative power of your approaches by showing genes with no prior evidence can produce disease phenotypes.

The reviewer highlights two related, but distinct ideas: pathways discovery and gene discovery. We could have focused on “the highest ranked gene with absolutely no prior evidence for Alzheimer's disease” if our study was intent on gene discovery. However, as our title indicates, we were interested in defining mechanisms of neurodegeneration, which is just as important for a gene with prior association with AD as it is for a completely novel gene. For example, by showing that *CSNK2A1* knockdown leads to neurodegeneration, increased DNA damage and inappropriate cell cycle activation, we infer that reduced *CSNK2A1* expression contributes to neurodegeneration through inappropriate cell cycle activation. Our work suggests that application of cell cycle inhibitors could be neuroprotective in DNA damage-burdened neurons. Through this example, we argue that our work can be used to identify molecules and targets for therapeutic intervention in neurodegenerative disease.

Reviewer #3 (Remarks on code availability):

The code is of above average quality in terms of organization and helpful in determining exactly what was done.

We thank the reviewer for recognizing our efforts to share our code and make our results reproducible.

References

1. Mathys, H. *et al.* Single- Cell Transcriptomic Analysis of Alzheimer's disease. *Nature* **570**, 332–337 (2019).

2. Yeager-Lotem, E. *et al.* Bridging high-throughput genetic and transcriptional data reveals cellular responses to alpha-synuclein toxicity. *Nat Genet* **41**, 316–323 (2009).
3. Mangleburg, C. G. *et al.* Integrated analysis of the aging brain transcriptome and proteome in tauopathy. *Mol Neurodegener* **15**, 56 (2020).
4. Pacifico, R., MacMullen, C. M., Walkinshaw, E., Zhang, X. & Davis, R. L. Brain transcriptome changes in the aging *Drosophila melanogaster* accompany olfactory memory performance deficits. *PLOS ONE* **13**, e0209405 (2018).
5. Yang, L. *et al.* Multidimensional Proteomics Identifies Declines in Protein Homeostasis and Mitochondria as Early Signals for Normal Aging and Age-associated Disease in *Drosophila**[S]. *Molecular & Cellular Proteomics* **18**, 2078–2088 (2019).
6. Hall, H. *et al.* Transcriptome profiling of aging *Drosophila* photoreceptors reveals gene expression trends that correlate with visual senescence. *BMC Genomics* **18**, 894 (2017).
7. Islam, M. S. *et al.* Human R1441C LRRK2 regulates the synaptic vesicle proteome and phosphoproteome in a *Drosophila* model of Parkinson's disease. *Hum Mol Genet* **25**, 5365–5382 (2016).
8. Scholes, H. M. *et al.* Dynamic changes in the brain protein interaction network correlates with progression of A β 42 pathology in *Drosophila*. *Sci Rep* **10**, 18517 (2020).
9. Wang, M., Peng, I.-F., Li, S. & Hu, X. Dysregulation of antimicrobial peptide expression distinguishes Alzheimer's disease from normal aging. *Aging (Albany NY)* **12**, 690 (2020).

10. Zhang, S. *et al.* Aging and Intermittent Fasting Impact on Transcriptional Regulation and Physiological Responses of Adult *Drosophila* Neuronal and Muscle Tissues. *Int J Mol Sci* **19**, 1140 (2018).