Editorial comments:

1. Please take this opportunity to thoroughly proofread the manuscript to ensure that there are no spelling or grammar issues. The JoVE editor will not copy-edit your manuscript and any errors in the submitted revision may be present in the published version.

We proofread the manuscript as suggested.

2. Title: Please revise to avoid the use of punctuation (colon, dash, etc.).

We now changed the title into “Mapping Alzheimer’s disease variants to their target genes using computational analysis of chromatin configuration”

3. Please revise the Protocol to contain only action items that direct the reader to do something (e.g., “Do this,” “Ensure that,” etc.). The actions should be described in the imperative tense in complete sentences wherever possible. Avoid usage of phrases such as “could be,” “should be,” and “would be” throughout the Protocol. Any text that cannot be written in the imperative tense may be added as a “NOTE.” Please include all safety procedures and use of hoods, etc. However, notes should be used sparingly and actions should be described in the imperative tense wherever possible. Please move the discussion about the protocol to the Discussion.

We ensured that the protocol does not have any phrases such as “could be”, “should be”, and “would be.”

4. Please add more details to your protocol steps. There should be enough detail in each step to supplement the actions seen in the video so that viewers can easily replicate the protocol. Please ensure you answer the “how” question, i.e., how is the step performed?

This request is similar to request #5. Please see our response to request #5.

5. For actions involving software usage, please provide all specific details (e.g., button clicks, software commands, any user inputs, etc.) needed to execute the actions. Please include a step-wise description of software usage; mention what button is clicked on in the software, or which menu items need to be selected, and provide user input commands, etc.

We made the following changes in response to the request #4 and #5.

1. We directed the readers to install RStudio (Line 93: Install RStudio Desktop: https://www.rstudio.com/products/rstudio/download/), which will help them run the code provided.

2. We added a sentence “Type the following code into the console window in RStudio (e.g. Line 96)” in front of each section of code for added clarity.
3. We have revised the Download files section (section **1.3**) to include links to all files and explicit directions for downloading data.

6. Line 283: Figure 2C does not exist. Please revise.

Thanks for pointing this out. We changed it to **Figure 2B**. We also found out that Figures were not labeled properly within the manuscript, which is now corrected.

7. Please remove the embedded figure(s) from the manuscript.

We removed the figures from the manuscript, and submitted separately.

Reviewers' comments:

Reviewer #1:

The authors in this paper developed a computational pipeline for linking GWAS loci to genes using Hi-C data, and applied to Alzheimer's disease for discovering AD risk genes. The pipeline mapped credible SNPs from AD GWAS to various regions including enhancers, gene bodies and promoters, and then linked SNPs to genes if the mapped regions have potential interactions from Hi-C data. Overall, the paper was well organized and provided a complete set of R codes for pipeline implementation. Before recommending for publication, I have the following minor concerns that authors need to address:

We thank the reviewer for his/her positive and constructive comments.

1. Hi-C data description such as protocol, resolution & tissue/cell type is missing. Can the pipeline be scalable to the Hi-C datasets with different resolutions, which authors discussed at the end?

We thank the reviewer for this comment. We already described that Hi-C interaction profiles were generated from the adult brain. Thanks to the comment, we also added Line 125: “NOTE: In case other Hi-C datasets are used, this protocol requires Hi-C datasets processed at high resolution (5-20kb).”

2. Fig. 2A seems mix multiple regions together. Is there any particular developmental expression pattern for particular regions? Authors didn't introduce how to normalize gene expression either.

We used cortical expression data from brainspan. We now updated **Figure 2A legend** to describe the brain region: “AD risk genes are highly expressed in the postnatal cortex compared to the prenatal cortex”. We used the expression data provided by BrainSpan (https://www.brainspan.org/) and did not perform any additional normalization or processing. We now updated section **1.3.1** with detailed instructions for how to download the expression data file and process it.

3. Was the single cell data from healthy or AD brain? If healthy, the cell type specific expression might not represent AD cells. Will cell type Hi-C improve gene linking over tissue Hi-C? Also, details on gene expression normalization is missing.

Single cell data was from healthy (neurotypical) brain. We used the expression data provided by the original paper (Wang et al., Science *2018*) which has been already normalized. We also updated section **1.3.1** to describe that the single cell data comes from healthy brains: Line 127: “Single cell expression datasets from the PsychENCODE (Described as *singlecell.rda* below). This is from neurotypical control samples.”

Reviewer #2:

Manuscript Summary:

The manuscript describes a method to annotate non-coding variations to the candidate genes for GWAS SNPs. Furthermore, the candidate genes have been subjected to enrichment analysis and cross-cell expression comparisons. Addressing the following comments will improve the quality of the manuscript.

We thank the reviewer for his/her critical and insightful comments.

Major Concerns:

1. The explanation about figure 1B is completely missing in the manuscript. It's important to mention and explain what resource was being used to get the figure and what purpose does it serve as the results of the work done. Additionally, the information about the total number of genes involved in each of the GO terms should also be written. Do all the genes take part in each of these GO terms? I don't think so. Also, what does the -log10(FDR) mean? This needs explanation as well. Overall, the results have been represented in a very abstract way. They need to be explained in detail.

We are sorry that **Figure 1B** was labelled incorrectly. We corrected the figure captions for Figures 1 and 2: “AD risk genes were associated with amyloid precursor proteins, amyloid-beta formation, and immune response, reflecting the known biology of AD[15–18] (**Figure 1B**).”

Based on the reviewer’s suggestion, we

(1) Updated **Figure 1B-D** with the # of genes in each term and # of genes represented in our list.
(2) Updated **Figure 1B-D legend** as below:

(B-D) Enrichment of GO (**B**), KEGG (**C**), and Reactome (**D**) terms in AD risk genes was performed using HOMER as described in step **5**. The x-axis represents the FDR corrected -log10 (P-value). Enriched terms with FDR<0.1 were plotted. Grey vertical lines represent FDR=0.05. APP amyloid precursor protein. Numerator, the number of AD risk genes represented in each term; denominator, the number of genes in each term.

2. This is a follow up to the 1st comment: The authors have done a sort of enrichment analysis depicting top 10 GO terms. It's highly recommended to perform and include results from Gene Set Enrichment Analysis (GSEA) using Broad Institute's Molecular signature database. This will give a better overview of associated pathways from KEGG and Reactome for the genes the authors have shortlisted in their work. In addition to this, it is also recommended to use NeuroMMsig (https://neurommsig.scai.fraunhofer.de/) to find out what Alzheimer Disease (AD) mechanisms are represented by the list of genes. NeuroMMsig is a mechanism enrichment analysis platform developed for neurodegenerative diseases, especially for AD and Parkinson disease (PD).

Thanks to this insightful suggestion, we used HOMER to analyze KEGG and Reactome for AD risk genes and reported the result in Figure 1. We think this greatly improved the overall findings of the manuscript, and gave more detailed descriptions about the potential function of AD risk genes.

3. It's difficult to comprehend what do the different shapes of different cellular expression profiles refer to.

Sorry for the confusion. We now updated **Figure 2B legend** as below:

“(B) Violin plots depicting distributions of gene expression values (normalized expression) in different cell types from the cortex. These results show that AD risk genes are highly expressed in microglia, consistent with previous studies[14].”

4. The authors have reported about highly expressed genes in microglia (Figure 2B) as a whole. It is good to make overall comparison of gene expressions across different cell types but when it comes to gene expression analysis, the down-regulated genes are also of interests. Therefore, the author might have to report about individual genes that are either over-expressed or under-expressed to improve the quality of the manuscript.

While it is often of interest to investigate genes with different patterns of change, the intent of this plot is to show that the genes implicated by our analysis are expressed in cell types with a known role in AD. This offers support to the validity of our analyses.

While the reviewer raised an important point, we believe that the analysis of other genes that are differentially expressed across these cell types is outside the scope of this manuscript.

5. The R data files for devExpr.rda, ADgenes.rda and singlecell.rda should be made available.

Because *devExpr.rda* and *singlecell.rda* use datasets that have been made publicly available by the original paper, we do not feel comfortable providing these files. Instead, we updated section **1.3.1** in which we described the download procedure in detail so that readers can easily follow and retrieve the same data. We are providing ADgenes.rda with this paper.

Minor Concerns:

Line 60: There are double full stops.

Line 70: First mention of GWS, therefore needs to be written with the full name and the abbreviation.

Line 75: Double "the".

Line 326-328: Needs to be checked again.

We thank the reviewer for catching these minor errors. We now corrected all the points brought up by the reviewer.