nature portfolio

Peer Review File

Proximity Analysis of Native Proteomes Reveals Phenotypic Modifiers in a Mouse Model of Autism and Related Neurodevelopmental Conditions

Editorial note: This manuscript has been previously reviewed at another journal that is not operating a transparent peer review scheme. This document only contains reviewer comments and rebuttal letters for versions considered at *Nature Communications*.

REVIEWER COMMENTS

Reviewer #2 (Remarks to the Author):

Unfortunately I still have the same opinion regarding the findings of the manuscript. The authors added some useful discussions for the reader, in regards to the limitations of the methods, but the general tone is the same. I don't agree that the methods "provided a significant advance over prior studies"

As the authors acknowledge:

We agree with the Reviewer that for any new method, it is important to discuss the Potential al limitations. In the revised manuscript, we highlighted 3 potential limitations to the

approach on lines 430-445: (1) quantitative comparisons to negative control do not mean a protein is not present; (2) low abundant proteins may be more difficult to biotinylate and enrich; (3) we cannot guarantee that the tag does not influence some binding partners. We are

sorry if we gave the false impression to the Reviewer that we claimed to have discovered all (or

exact) PPIs, but we provided an honest discussion of limitations.

This is in answer to the fact that the assays didn't detect many PPIs that have been validated over the years (I already discussed this). Thus, while there are some advantage, there are obviously clearly severe limitations. Again, I repeat that I'm not against the use of these assays, but they are not a "significant improvement", they manuscript doesn't show that "HiUGE-iBioID captures relevant data better than existing methods". What is this better relevant data? Surely is not capturing very well studied Dlgap1, Dlgap4 in Shank2 and Shank3 PPIs (among others). All of them developmental disease genes, ASD genes, epilepsy, general development, core components of the networks described in the manuscript, etc…

I repeat that my main issue is that there is no “new paradigm” or a new “biology” proposed. Therfore, the tone of he manuscript. You can say that after the use of these assays, the conclusions are the same as discussed over many years and countless manuscripts. Also, the Syngap-Anks1b PPI used to highlight the assays have been already described by other methods that re supposed to be of inferior quality. Thus I don’t support the statement that the methods offer a way to “reveal unexpected and highly significant interactions with other lower-confidence autism risk gene products, positing new avenues to prioritize genetic risk”. And I don’t want to enter the discussion of what is an ASD risk gene.

Reviewer #4 (Remarks to Author):

A. Methodological considerations

A major aspect of this paper is the promotion of a method for “a scalable genome editing mediated approach to target…genes…for proximity-based endogenous proteomics”. Paragraphs 2 and 3 of the discussion present the case that their method (termed HiUGE-iBioID) has “four key benefits”. Benefits 1 and 2 refer to the fact that the tag is on an endogenous protein in the intact animal. Benefit 3 is a restatement of the endogenous in vivo issue. Benefit 4 refers to the fact that the approach is not limited by antibody availability and can be used against more targets. The authors go on to say that HiUGE-iBioID may be an “optimal approach”.

Unfortunately, the authors do not present the major limitations of their approach, which I will outline below.

Limitation 1. Somatic mosaicism.

Their method relies on the use of an AAV that is injected into the brain of mouse pups. This method will infect a subpopulation of cells and some of these will have their genome edited, which both contribute to somatic mosaicism. Evidence for somatic mosaicism is apparent in Supplementary Figure S1 and S2. For example, in the leftmost panel of Fig S1 there is heterogenous expression between each hemisphere, and in other panels where individual

cells are visible it is apparent that subpopulations of cells are labelled and some regions show no labelling (e.g. dentate gyrus). In the higher magnification images in Fig S2 it is clear that the endogenous protein labelled with antibodies reveals more puncta than the genetically tagged puncta and that the cell bodies.

Somatic mosaicism presents a number of important problems. First, it is not possible to argue which cell type the proteomic data has been obtained from. Second, variability in the efficacy of infection or injection between individual experiments will likely produce variability in the proteomic data. Thirdly, it is not possible to easily determine if there is a mutant phenotype.

Limitation 2. Control of zygosity.
CRISPR-Cas9 engineering is remarkably efficient and is well known to cause homozygous or heterozygous mutations in cells. Inserting a tag on a protein always risks making a mutation that changes the function or localization of the protein and the physiological impact of this will be determined by the zygosity of the mutation. It is therefore crucial to control the zygosity, or at the very least, know what the zygosity is of the engineered cells. It is quite possible, if not highly likely, that in a mouse that has been engineered with HiUGE-iBioID that some cells will carry heterozygous mutations and other carry homozygous mutations.

The two limitations raised above make it particularly difficult for the investigator to rule out a functional mutation. Consider the situation where subsets of neurons carry either a homozygous or heterozygous mutation. What kinds of physiological, behavioral or other functional assays can be performed that will demonstrate there are no phenotypes? I am disappointed that the authors have not attempted to address the issues of mosaicism and zygosity nor raised them in their discussion. I note that Reviewers 1-3 did not raise these critical issues either.

Moreover, the authors do not present a balanced discussion of their methods compared to the established methods of tagging proteins using germline engineering. The issues of mosaicism and zygosity are not limitations of the germline tagging approach. Furthermore, with germline modification one can control the cell-types that are tagged using proven Cre

lines or viruses with specific promotor/enhancers for cell type discrimination. Presently the HiUGE-iBioID does not address cell-type specificity (discussed further below).

What might be the impact of mosaicism and zygosity in BioID proteomic experiments? There are several ways that one could envision that it would cause the proteomic data to be skewed and misrepresent the ground truth profile of proximal proteins. For example, the tagged protein could mislocalize in some neurons (and not others) and its localization could be influenced by zygosity. The authors make a point of comparing their HiUGE-iBioID data with immunoprecipitation data (for Anks1b and Scn2a) and come up with overlapping, but distinct profiles. What are we to think about this difference? Could it arise for the reasons I mention above?

While this cannot be excluded by the authors, it is important to point out another limitation of the authors attempt to make this comparison: they are comparing immunoprecipitation (IP) with proximity labelling. IP measures the proteins found in complexes; proximity labelling measures proteins that are near to each other (and can be in distinct and separate complexes). Thus, we cannot expect the two datasets to be identical. It is wrong for the authors to say at line 398 that their approach “outperforms” immunoprecipitations.

Limitation 3. Cell-type specificity

In addition to the comments above, I note that Reviewer 1 at point 2 raises questions on the issue of cell type. The authors response that “The data is cell-type specific to the extent the bait’s endogenous expression is.” is misleading because the tropism of the virus influences this as well. AAVs do not infect all brain cells equally.

Limitation 4. Off-target effects

The authors make no mention about the concern of off-target effects and the problems that their method imposes as far as detecting these effects or the advantages that germline tagging approaches have for dealing with this issue.

B. General comments

Description of protein datasets

Proteins assemble into complexes and these complexes are packed together in compartments within cells. It is possible to identify the proteins in complexes by purifying the complexes and then performing proteomic experiments. It is also possible to identify proteins that are nearby to each other (which can be in the same and/or in different complexes) using the proximity labelling (BioID) approach.

It is important to distinguish these two methods, but unfortunately the authors conflate the two methods by referring to “interactomes” in the context of proximity labelling. Proteins that are in different nearby complexes don’t necessarily interact whereas proteins that are in complexes do, by definition, interact. It would be more accurate for the author to describe their datasets as “proximity proteomes” and not “interactomes”. By being more accurate the authors could nicely compare real interactome data with their proximity data.

Moreover, because their data are “proximity proteomes” and the method they use produces somatic mosaicism, the author should address in what way their data is a reflection of different cell types (and synapse types). They could have had a very interesting discussion about the spatial nature of their data at the level of “within cells” and “between cells”. I suspect they may not want to have raised this too much because it might reflect on the limitations of their approach.

Novelty of methods

Regarding the novelty of the HiUGE-iBioID and its context within tagging of endogenous brain proteins and their proteomic analysis. Somatic tagging of brain proteins using CRISPR has been done by other groups (including Harald MacGillavry’s group). People have used Cas9 mice before in conjunction with AAVs. So there is very little technical novelty in the HiUGE-iBioID approach here. Having said this, there is a place for it because it can be used to fairly rapidly screen proteins in vivo.

Novelty of biological findings

The authors reported a set of “interactomes” for 14 autism risk genes. Others have isolated proteins interacting with autism proteins and found large sets containing other autism

genes, so this is not a major conceptual advance. We have known for a long time that synapse proteomes and other cellular compartments are enriched for autism genes. The datasets could be useful to others but did not reveal any major new insights. I concur with Reviewer 2 in this regard.

Response to Reviewers

Reviewer #2 (Remarks to the Author):

Unfortunately I still have the same opinion regarding the findings of the manuscript. The authors added some useful discussions for the reader, in regards to the limitations of the methods, but the general tone is the same. I don't agree that the methods "provided a significant advance over prior studies"
As the authors acknowledge:
We agree with the Reviewer that for any new method, it is important to discuss the Potential al limitations. In the revised manuscript, we highlighted 3 potential limitations to the approach on lines 430-445: (1) quantitative comparisons to negative control do not mean a protein is not present; (2) low abundant proteins may be more difficult to biotinylate and enrich; (3) we cannot guarantee that the tag does not influence some binding partners. We are sorry if we gave the false impression to the Reviewer that we claimed to have discovered all (or exact) PPIs, but we provided an honest discussion of limitations.

This is in answer to the fact that the assays didn't detect many PPIs that have been validated over the years (I already discussed this). Thus, while there are some advantage, there are obviously clearly severe limitations. Again, I repeat that I'm not against the use of these assays, but they are not a "significant improvement", they manuscript doesn't show that "HiUGE-iBioID captures relevant data better than existing methods". What is this better relevant data? Surely is not capturing very well studied Dlgap1, Dlgap4 in Shank2 and Shank3 PPIs (among others). All of them developmental disease genes, ASD genes, epilepsy, general development, core components of the networks described in the manuscript, etc…

Response:
We have now removed the statement "provided a significant advance over prior studies", and we have included a statement on lines 426-428 in the Discussion that, not all previously reported interactions were found be enriched in our experiments, including Dlgap1 and Dlgap4 for the Shank2 and Shank3 baits. We do wish the reviewer would have commented on the new additional analysis we provided comparing our method to other methods for the same baits in Supplemental Figs 9-10.

I repeat that my main issue is that there is no "new paradigm" or a new "biology" proposed. Therfore, the tone of he manuscript. You can say that after the use of these assays, the conclusions are the same as discussed over many years and countless manuscripts. Also, the Syngap-Anks1b PPI used to highlight the assays have been already described by other methods that re supposed to be of inferior quality. Thus I don't support the statement that the methods offer a way to "reveal unexpected and highly significant interactions with other lower-confidence autism risk gene products, positing new avenues to prioritize genetic risk". And I don't want to enter the discussion of what is an ASD risk gene.

Response:
We have edited the statement that our method could “reveal unexpected and highly significant interactions with other lower-confidence autism risk gene products, positing new avenues to prioritize genetic risk”. Instead, this sentence now reads on lines 36-38, “reveal proximity interactions between proteins from high-confidence risk genes with those of lower-confidence that may provide new avenues to prioritize genetic risk.”

Reviewer #4 (Remarks to Author):

A. Methodological considerations

A major aspect of this paper is the promotion of a method for “a scalable genome editing mediated approach to target…genes…for proximity-based endogenous proteomics”. Paragraphs 2 and 3 of the discussion present the case that their method (termed HiUGE-iBioID) has “four key benefits”. Benefits 1 and 2 refer to the fact that the tag is on an endogenous protein in the intact animal. Benefit 3 is a restatement of the endogenous in vivo issue. Benefit 4 refers to the fact that the approach is not limited by antibody availability and can be used against more targets. The authors go on to say that HiUGE-iBioID may be an “optimal approach”.

Unfortunately, the authors do not present the major limitations of their approach, which I will outline below.

Limitation 1. Somatic mosaicism.
Their method relies on the use of an AAV that is injected into the brain of mouse pups. This method will infect a subpopulation of cells and some of these will have their genome edited, which both contribute to somatic mosaicism. Evidence for somatic mosaicism is apparent in Supplementary Figure S1 and S2. For example, in the leftmost panel of Fig S1 there is heterogenous expression between each hemisphere, and in other panels where individual cells are visible it is apparent that subpopulations of cells are labelled and some regions show no labelling (e.g. dentate gyrus). In the higher magnification images in Fig S2 it is clear that the endogenous protein labelled with antibodies reveals more puncta than the genetically tagged puncta and that the cell bodies.

Somatic mosaicism presents a number of important problems. First, it is not possible to argue which cell type the proteomic data has been obtained from. Second, variability in the efficacy of infection or injection between individual experiments will likely produce variability in the proteomic data. Thirdly, it is not possible to easily determine if there is a mutant phenotype.

Response:
We thank the reviewer for highlighting the somatic mosaicism. We have added a statement to the Discussion (lines 439-442) stating that “HiUGE-iBioID relies on sparse editing of cells within brain tissue. While we have demonstrated that the efficiency is

sufficient for biochemical approaches such as proximity proteomics, one should keep in mind that some cell types may edit better than others, presenting potential bias in the data." In addition, we have removed the statement "optimal approach" from the Discussion.

Limitation 2. Control of zygosity.
CRISPR-Cas9 engineering is remarkably efficient and is well known to cause homozygous or heterozygous mutations in cells. Inserting a tag on a protein always risks making a mutation that changes the function or localization of the protein and the physiological impact of this will be determined by the zygosity of the mutation. It is therefore crucial to control the zygosity, or at the very least, know what the zygosity is of the engineered cells. It is quite possible, if not highly likely, that in a mouse that has been engineered with HiUGE-iBioID that some cells will carry heterozygous mutations and other carry homozygous mutations.

The two limitations raised above make it particularly difficult for the investigator to rule out a functional mutation. Consider the situation where subsets of neurons carry either a homozygous or heterozygous mutation. What kinds of physiological, behavioral or other functional assays can be performed that will demonstrate there are no phenotypes? I am disappointed that the authors have not attempted to address the issues of mosaicism and zygosity nor raised them in their discussion. I note that Reviewers 1-3 did not raise these critical issues either.

Moreover, the authors do not present a balanced discussion of their methods compared to the established methods of tagging proteins using germline engineering. The issues of mosaicism and zygosity are not limitations of the germline tagging approach. Furthermore, with germline modification one can control the cell-types that are tagged using proven Cre lines or viruses with specific promotor/enhancers for cell type discrimination. Presently the HiUGE-iBioID does not address cell-type specificity (discussed further below).

Response:
Regarding zygosity, we have added a statement to the Discussion (lines 443-447) stating that "Unlike transgenic mouse production, HiUGE editing can result in cells that are either heterozygous or homozygous tagged in the same tissue, potentially contributing to functional heterogeneity in the edited cells. Furthermore, germline modifications can leverage Cre lines to specify cell types, while the currently presented HiUGE-iBioID approach does not. We note, however, that cell-type specificity for HiUGE-iBioID could be achieved by using Cre-dependent expression of Cas9."

What might be the impact of mosaicism and zygosity in BioID proteomic experiments? There are several ways that one could envision that it would cause the proteomic data to be skewed and misrepresent the ground truth profile of proximal proteins. For

example, the tagged protein could mislocalize in some neurons (and not others) and its localization could be influenced by zygosity. The authors make a point of comparing their HiUGE-iBioID data with immunoprecipitation data (for Anks1b and Scn2a) and come up with overlapping, but distinct profiles. What are we to think about this difference? Could it arise for the reasons I mention above?

While this cannot be excluded by the authors, it is important to point out another limitation of the authors attempt to make this comparison: they are comparing immunoprecipitation (IP) with proximity labelling. IP measures the proteins found in complexes; proximity labelling measures proteins that are near to each other (and can be in distinct and separate complexes). Thus, we cannot expect the two datasets to be identical. It is wrong for the authors to say at line 398 that their approach "outperforms" immunoprecipitations.

Response:
We have removed the statement "outperforms immunoprecipitations" from the Discussion, the sentence now reads: "While further comparisons are needed to confirm the above observations, the available data suggest that endogenous proximity proteomics using HiUGE-iBioID is a valuable addition to the existing methods." (lines 392-394).

To further tone down our interpretation of the observed differences between different methods, we have removed the following statements: "direct comparisons of HiUGE-iBioID to the prior "gold standard" of immunoprecipitation…" from the Introduction, "these data demonstrate that HiUGE-iBioID excels in specificity for detecting biologically-relevant proteomic interactions" from the Result Section, and "employing HiUGE-iBioID for the in vivo study of endogenous protein complexes appears to be an advantageous method overall", and "with higher confidence interactomes than immunoprecipitation or BioID over-expression…" from the Discussion.

Finally, it is not our intention to criticize studies using immunoaffinity-based or recombinant BioID approaches, as they do have their own merits. These direct comparison analyses were added during the first revision cycle, as suggested by the reviewers.

Limitation 3. Cell-type specificity
In addition to the comments above, I note that Reviewer 1 at point 2 raises questions on the issue of cell type. The authors response that "The data is cell-type specific to the extent the bait's endogenous expression is." is misleading because the tropism of the virus influences this as well. AAVs do not infect all brain cells equally.

Response:
As mentioned above, we have added a statement to the discussion stating that "HiUGE-BioID relies on sparse editing of cells within brain tissue. While we have demonstrated that the efficiency is sufficient for biochemical approaches such as proximity proteomics,

one should keep in mind that some cell types may edit better than others, presenting potential bias in the data." (lines 439-442).

Limitation 4. Off-target effects
The authors make no mention about the concern of off-target effects and the problems that their method imposes as far as detecting these effects or the advantages that germline tagging approaches have for dealing with this issue.

Response:
We have now added to the Discussion the statement, "Off-target integrations can occur in cells using CRISPR approaches, which is why it is important to verify the bait protein is detected in the resulting proteomics analysis. In our prior publication on the HiUGE approach we experimentally determined the off-target rate is low and largely results in integration in non-coding regions. As the integration of BioID into each cell in non-dividing neurons is an independent effect, it is likely that rare off-target integrations that occurred in-frame with a coding region would make limited contribution to the total detected proximity proteome" (lines 433-439).

B. General comments

Description of protein datasets
Proteins assemble into complexes and these complexes are packed together in compartments within cells. It is possible to identify the proteins in complexes by purifying the complexes and then performing proteomic experiments. It is also possible to identify proteins that are nearby to each other (which can be in the same and/or in different complexes) using the proximity labelling (BioID) approach.

It is important to distinguish these two methods, but unfortunately the authors conflate the two methods by referring to "interactomes" in the context of proximity labelling. Proteins that are in different nearby complexes don't necessarily interact whereas proteins that are in complexes do, by definition, interact. It would be more accurate for the author to describe their datasets as "proximity proteomes" and not "interactomes". By being more accurate the authors could nicely compare real interactome data with their proximity data.

Response:
We have adopted the term "proximity proteomes" throughout.

Moreover, because their data are "proximity proteomes" and the method they use produces somatic mosaicism, the author should address in what way their data is a reflection of different cell types (and synapse types). They could have had a very interesting discussion about the spatial nature of their data at the level of "within cells" and "between cells". I suspect they may not want to have raised this too much because it might reflect on the limitations of their approach.

Response:
We are not sure what the reviewer meant by discussing data at the level of “within cells” and “between cells”, but it does appear he/she is asking for a change here.

Novelty of methods
Regarding the novelty of the HiUGE-iBioID and its context within tagging of endogenous brain proteins and their proteomic analysis. Somatic tagging of brain proteins using CRISPR has been done by other groups (including Harald MacGillavry’s group). People have used Cas9 mice before in conjunction with AAVs. So there is very little technical novelty in the HiUGE-iBioID approach here. Having said this, there is a place for it because it can be used to fairly rapidly screen proteins in vivo.

Response:
Yes, other labs have published CRISPR approaches to tagging proteins similar to HiUGE. Indeed, Harald MacGillavry’s publication on his method termed “ORANGE” (PMID: 32275651, 2020) was published ***after*** our publication describing HiUGE. Importantly, the novelty of the methods presented here is the combined use of HiUGE and iBioID for the identification of functionally significant proximity proteomes *in vivo*, not the somatic tagging of brain proteins using HiUGE (published in 2019).

Novelty of biological findings
The authors reported a set of “interactomes” for 14 autism risk genes. Others have isolated proteins interacting with autism proteins and found large sets containing other autism genes, so this is not a major conceptual advance. We have known for a long time that synapse proteomes and other cellular compartments are enriched for autism genes. The datasets could be useful to others but did not reveal any major new insights. I concur with Reviewer 2 in this regard.

Response:
We respectfully disagree with this statement. Yes, many of the proximity interactions we detected were previously known, in addition to many novel ones. Yet, as discussed in prior rebuttals, neither Reviewer 2 nor Reviewer 4 commented on the novelty of the Scn2a phenotypic rescue based on the proximity proteomics results we presented.