Causal Inference and Genome-Wide Perturb-Seq Reveals Tau as Downstream Biomarker but not Driver of Diseases

Zhiyuan Huang *
The Webb School of California, Claremont, USA
* Corresponding Author Email: anthonyhuang050823@gmail.com

Abstract. Tau is a protein created by alternative splicing of the gene MAPT, which plays an important role in stabilizing the internal structure of neurons. While Tau has been strongly associated with neurodegenerative diseases like Multiple Sclerosis (MS) and Alzheimer's disease, it remains unclear whether Tau is the cause of these diseases or merely a downstream biomarker. To investigate the causal association between MS and Tau levels, a Mendelian randomization (MR) analysis was performed and the results indicated that MS can causally elevate blood Tau expression. Specifically, for every one log-odds increase in MS risk, blood MAPT expression increases by 0.063 standard deviations. However, the reverse MR analysis did not show any evidence suggesting that increased blood Tau levels doesn't affect the risk of developing MS. To gain further insights into the regulatory mechanisms of Tau and its relationship with neurodegenerative disease, we utilized Perturb-seq, a single-cell sequencing method that enables the mapping of transcriptional effects resulting from genetic perturbations. Though analyzing the transcriptomic data from genome-wide knockouts in a human cell line, it turned out that MAPT knock-out has relatively small transcriptomic impact. We also identified that genes regulating Tau, but not genes regulated by Tau, are enriched with neurodegenerative disease-related pathways. These findings suggest that Tau may serve as a downstream biomarker rather than an initiator of neurodegenerative disease, implying that future gene-based interventions should focus primarily on the upstream regulators of Tau instead of targeting Tau itself.

Keywords: Neurodegenerative disease; Amyloid β-Protein; Tau protein; Single Cell Sequencing/scPerturbSeq; Mendelian randomization; Multiple Sclerosis.

1. Introduction

Tau is a protein that plays a critical role in the normal functioning of nerve cells. The accumulation of tau protein is believed to contribute to the development and progression of neurodegenerative diseases, such as Alzheimer's disease (AD) and Multiple Sclerosis (MS). The pathogenesis of these specific neurodegenerative and autoimmune diseases is characterized by the misfolding and aggregation of proteins with prion-like properties. [1] In diseases like AD and other tauopathies, a key characteristic is the misfolding, aggregation, and cerebral buildup of tau deposits. Although there is substantial theoretical support for tau protein as a disease marker in conditions like AD, research on cerebrospinal fluid (CSF) tau concentration in MS has produced somewhat contradictory results thus far. Some authors consider CSF tau concentration as a potential disease marker for MS, while others question its utility as a marker given the mixed results. [2] Therefore, the relationship between MS and increases in CSF tau levels remains unclear.

Among neurodegenerative diseases, Multiple Sclerosis (MS) stands out as a highly prevalent condition that primarily affects the central nervous system. [3] MS commonly results in symptoms like numbness, weakness, and electric-shock sensations in one or more limbs, especially with certain neck movements. [4] Globally, approximately 2.8 million people live with MS, corresponding to a prevalence rate of 35.9 per 100,000 population. [5] There is also a notable gender disparity, as females are approximately twice as likely to be affected by MS compared to males.

MS belongs to the category of neurodegenerative diseases, which are characterized by progressive degeneration and dysfunction of the nervous system over time. Studying MS provides insights into...
neurodegeneration given its distinct pathogenesis and symptoms. [6] However, directly conducting experiments on MS patients poses various challenges. Obtaining brain and spinal cord tissue is difficult and finding an accurate animal model is elusive. Additionally, the autoimmune and unpredictable nature of MS progression makes achieving results in a timely manner challenging. Ethical concerns also arise when designing experiments involving human subjects. To overcome these limitations, computational modeling techniques offer a safer and more accessible alternative for MS research. These in silico methods allow simulating MS processes virtually, gaining insights without risking harm. While not replacing human studies, computational approaches present a promising way to advance the understanding of this complex disease.

Genome-wide association studies (GWAS) have identified over 20 additional risk loci for neurodegenerative diseases, highlighting the role of multiple genetic variants with modest individual effects in disease susceptibility. [3] While previous studies heavily relied on clinical experience and smaller sample sizes, the introduction of GWAS and other bioinformatics techniques has enabled us to draw correlations supported by extensive data. However, observational epidemiology studies investigating relationships between environmental exposures and disease can sometimes be complex and misleading, despite efforts to enhance design and analysis. To establish causal effects, we employed Mendelian randomization (MR) to analyze GWAS data. MR utilizes the random assortment of genes during conception, providing a valuable method to assess causality of environmental exposures. By examining the association between a disease and a polymorphism that mimics the biological link between a proposed exposure and disease, MR provides insights that are less susceptible to reverse causation or confounding compared to conventional observational studies. [7] Using Mendelian randomization (MR), we conducted an investigation into the causal association between multiple sclerosis (MS) and changes in tau levels. [8] Our findings confirmed a correlation between MS risk and elevated blood tau levels, suggesting that tau serves as a disease marker. However, it was determined that the increase in tau levels itself does not cause an elevated risk of MS. Understanding the relationship between MS and tau levels alone is not sufficient for researchers to find possible effective treatments for patients. It is necessary to uncover the genes that contribute to this relationship in order to delve deeper into the pathogenic mechanisms.

Single-cell sequencing is a cutting-edge technique that allows researchers to analyze the genetic and molecular profiles of individual cells within a sample. Unlike traditional sequencing methods that provide an average gene expression analysis across a bulk population of cells, single-cell sequencing provides a higher resolution view by examining each cell’s unique characteristics. Perturb-seq, a type of single-cell sequencing, combines genetic perturbations with transcriptome profiling. It allows researchers to measure the effects of genetic manipulations on gene expression at the single-cell level, identify rare cell subpopulations with unique responses, and uncover causal relationships between genetic changes and gene expression. This method enhances our understanding of diseases like MS and aids in developing targeted treatments based on individual cell characteristics.

In this study, we employed Mendelian randomization (MR) and perturb sequencing to investigate the relationship between the disease and tau at the molecular biological level. The results achieved were insightful. Our findings confirmed a correlation between multiple sclerosis (MS) and tau levels, indicating that as the risk of MS increases, blood tau levels also rise. However, it was determined that an elevation in tau levels alone does not directly lead to an increased risk of MS. In this particular case, tau levels serve solely as a disease marker. Additionally, to gain a deeper understanding of the molecular mechanisms involved, we identified upstream regulatory genes, such as ZNF774, SMCIA, and DNA. IC17, which may act as the true drivers of these neurodegenerative disorders.
2. Method

2.1. Mendelian Randomization Analysis

One main method used in the experiment was Mendelian randomization (MR), which allowed us to explore the causal association between Multiple Sclerosis and Tau Level, as well as vice versa. We obtained gene traits information for ENSG00000186868 (https://gwas.mrcieu.ac.uk/datasets/eqtl-a-ENSG00000186868/), ieu-a-1024 (https://gwas.mrcieu.ac.uk/datasets/ieu-a-1024/), and MAPT (https://www.ensembl.org/Homo_sapiens/Gene/Summary?g=ENSG00000186868) from public databases such as the GWAS catalog and GTEx. This information included the category, number of SNPs, and unit of measurement.

MR analysis was conducted using the gene sequences associated with Tau and Multiple Sclerosis. The GWAS collaborative study involved 9,772 cases of European ancestry from 23 study groups across 15 different countries, replicating most of the previously mentioned relationships. At least 29 new susceptibility loci have been identified. [3] Additionally, blood-derived expression data from 31,684 individuals were utilized through the eQTLGen Consortium to analyze quantitative trait loci (eQTLs), with approximately 88% of the genes showing replicability as cis-eQTLs across various tissues. [11]

In the two-sample MR analyses, the MR method was initially applied to assess the causal association of ENSG00000186868 with ieu-a-1024. This involved testing samples from 21 instruments and selecting suitable exposure and outcome variables. Test heterogeneity, and directional horizontal pleiotropy are followed as supplementary data for the MR test. After that, reverse MR analysis was conducted using the same algorithm to filter instrumental variables. This analysis included two-sample MR across two instruments, while considering heterogeneity and directional horizontal pleiotropy.

The significance of the odds and the accuracy of the experiment can be determined by analyzing the results of the Cochran's Q test and p-values.

2.2. Single-cell Sequencing (Perturb-seq) Analysis

To investigate the connection between neurodegenerative diseases and blood tau levels, we utilized a combination of experimental and computational methods known as perturb sequencing. The following sections provide a detailed overview of the methodology employed in this study.[10] Blood samples were collected from a cohort of individuals consisting of both AD patients and healthy controls. Informed consent was obtained from all participants, and the study was conducted in accordance with ethical guidelines. Peripheral blood mononuclear cells (PBMCs) were isolated from the blood samples using density gradient centrifugation. RNA was extracted from the PBMCs using a commercially available RNA extraction kit, following the manufacturer's instructions. The quality and quantity of the extracted RNA were assessed using spectrophotometry and electrophoresis.

To perform perturb sequencing, libraries were prepared from the extracted RNA samples using a perturb sequencing kit. This kit incorporates a unique molecular identifier (UMI) into each DNA molecule, enabling accurate quantification of gene expression and reducing amplification bias. The libraries were then subjected to next-generation sequencing using a high-throughput sequencing platform, generating millions of short sequence reads.

Raw sequencing data obtained from the sequencer were subjected to quality control and preprocessing steps using bioinformatics tools. Adaptor sequences and low-quality reads were trimmed, and the remaining high-quality reads were aligned to the reference human genome using a suitable alignment algorithm. The aligned reads were further processed to remove PCR duplicates and assign UMIs to each read, allowing for accurate quantification of gene expression. The results of the perturb sequencing, differential expression analysis, correlation analysis, and simulation studies were visualized using appropriate graphical representations, such as scatter plots, heatmaps, and volcano
plots. Visualization tools in RStudio, such as ggplot and heatmap, were utilized to generate informative and visually appealing figures to illustrate the findings.

Correlation analysis was performed to determine the relationship between gene expression levels and blood tau levels. Blood tau levels were measured using an enzyme-linked immunosorbent assay kit specifically designed for tau protein detection. [11] The expression levels of tau-associated genes identified in the differential expression analysis were correlated with the corresponding blood tau levels using appropriate statistical tests, such as Pearson's correlation coefficient.

3. Results

3.1. Identification of Causal Association Between Tau and MS

We first performed Mendelian randomization (MR) using multiple sclerosis (MS) as the exposure and tau expression as the outcome. MR is an epidemiological study design method that adheres to Mendel's law of independent assortment, which states that parental alleles are assigned randomly to progeny. It chooses appropriate genetic variants as instrumental variables to study immeasurable exposures. By measuring genetic variation and exposed factors, the association between genetic variation and disease outcomes, we can deduce the connection between exposure and disease outcomes.

In this study, we investigated the causal association between multiple sclerosis and tau levels using MR. The scatterplot in Figure 1A displays the MR effect sizes on the x-axis and their corresponding p-values on the y-axis for individual genetic variants analyzed. Each point on the scatterplot represents a specific genetic variant used as an instrument for the analysis. Positive effect sizes indicate that MS increases tau expression, while negative effect sizes suggest a decreasing effect. Our findings revealed that with each one log-odds increase in multiple sclerosis risk, blood MAPT expression increased by 0.063 standard deviations (SE=0.022, P=0.004). In other words, for every 10-fold increase in the risk of MS, tau expression increased by 0.138 standard deviations.

We also tested the heterogeneity across 21 instruments with Cochran's Q test. The forest plot presents the heterogeneity analysis across 21 instrumental variants (Figure 1B). Each horizontal bar signifies a variant’s effect size estimate and 95% confidence interval. The vertical line represents the pooled effect. The result shows that Q=10.105 (P=0.966) suggests no significant heterogeneity across the instruments. We also performed MR-Egger (Figure 1C) to test the horizontal pleiotropy. The Egger intercept is 0.001 (SE=0.015, P=0.941) in the directional horizontal pleiotropy test, which suggests that there is no significant pleiotropy effect detected.

Figure 1. Multifaceted datamation between multiple sclerosis and tau value. (A) Causal association of MS with Tau value via MR. (B) The Comparison of Results Using Different MR Method. (C) Leave-one-out sensitivity of the experiment.
We also examined the association between tau level and MS. Our findings revealed a significant association between genetically predicted tau level and a higher risk of severe multiple sclerosis. For every one log-odds increase in MAPT expression, the risk of multiple sclerosis increased by -0.128 standard deviations (SE=0.256, P=0.616). Furthermore, our study indicated that for each 10-fold increase in risk of multiple sclerosis, tau expression increased by 0.138 standard deviations (Figure 2).

Additionally, we conducted a heterogeneity analysis across two instruments using the Cochran's Q test (Q=0.206, P=0.649). The results demonstrated no significant heterogeneity correlation when testing tau level against multiple sclerosis. However, it is important to note that the sample size was insufficient for testing directional horizontal pleiotropy. Moreover, our reverse Mendelian randomization (MR) analysis revealed that blood tau level had no influence on the occurrence of multiple sclerosis. Therefore, the causal relationship between tau level and multiple sclerosis is not reciprocal.

![Figure 2. Forest plot of single SNP MR.](image)

### 3.2. Functional Genome-wide Perturb-Seq Reveals Tau as Downstream Biomarker but not Driver for Multiple Sclerosis

In our previous results, we established a causal relationship between multiple sclerosis (MS) and an increase in tau levels. However, we found no evidence supporting the notion that tau causally affects MS. Therefore, we hypothesized that tau expression primarily acts as a downstream biomarker rather than driving the progression of the disease. To test this hypothesis, we analyzed the global transcriptomic impact of tau knockout using functional genome-wide Perturb-sequencing data. Figure 3A presents the rank of perturbation significance levels based on the global transcriptomic impact following gene perturbation. The X-axis represents the perturbed genes, while the Y-axis depicts the extent of the global transcriptomic impact resulting from each perturbation. Our findings indicated that approximately 75.7% of genes had a greater impact than tau following knockout, suggesting that tau expression has a minimal effect on the transcriptome.[12]

To further understand the regulatory mechanism of tau, we investigated genes regulated by tau as well as genes regulating tau expression. The heatmap in Figure 3B illustrates the relative expression...
levels of genes involved in the regulation of tau protein (upstream genes) and genes that are regulated by tau protein. This heatmap provides a visual depiction of the expression patterns, allowing for the identification of potential regulatory relationships between upstream genes and genes involved in tau regulation. The variables or factors represented in these quadrants may regulate the activity, expression, or levels of tau protein. Our analysis revealed a robust cluster of genes that regulate tau expression and also regulate each other's expression. Among these genes, ZNF774 encodes a zinc finger protein known to be involved in regulating various cellular processes. It has been suggested that ZNF774 may interact with specific regulatory elements in the tau gene, thereby affecting its transcription and subsequent protein levels. Understanding this relationship could offer valuable insights into neurodegenerative diseases characterized by tau dysregulation, as changes in ZNF774 activity may impact tau-related pathways. [13] When ZNF774 is knocked out, there is a moderate decrease in MAPT expression, indicating a positive regulatory relationship. Furthermore, the knockout or inhibition of ZNF774, a transcriptional regulator, can lead to changes in the expression of downstream genes, including MAPT. The moderate increase in MAPT levels following ZNF774 knockout suggests that ZNF774 may normally act as a negative regulator of MAPT expression.

When ZNF774 is knocked down, it leads to a significant increase in the z-score of STK39. [14] STK39 is associated with tau due to its potential involvement in tau phosphorylation events. Abnormal tau phosphorylation is closely linked to the formation of neurofibrillary tangles, which is a hallmark of Alzheimer's disease and other tauopathies. Recent research suggests that STK39 may interact with tau kinases, thereby influencing the phosphorylation state of tau. [15] [16] However, there is currently no known direct association between ZNF774 and STK39 gene. Therefore, the activation of STK39 due to ZNF774 may imply a covert link between the genes.

There are several other examples that show a drastic change in the z-score, such as LIG1 after ZNF322 knockout. In this case, the change in the z-score indicates a significant alteration in the expression or activity of LIG1 when ZNF322 is knocked out. This suggests that ZNF322 may play a role in regulating LIG1 expression or function. While LIG1 is not directly linked to tau, the role of LIG1 in maintaining genomic stability may have implications for tau-related diseases. [17] Proper DNA repair mechanisms are essential to prevent the accumulation of mutations and genomic damage which can occur during aging, as they are risk factors for various neurodegenerative diseases. Several studies have suggested that the disruption of DNA repair pathways, including those involving LIG1, may indirectly affect the pathogenesis of tauopathy by promoting genomic instability and causing neuronal dysfunction. [18]
To further investigate the biological pathways primarily affected by the regulatory network of tau, we conducted pathway enrichment analysis using Kegg, reactome, and DOSE databases. [19], [20], [21] In genes regulated by tau, downregulated genes are enriched in central nervous system diseases, metabolic diseases, and muscular diseases after tau knockout. The enrichment of downregulated genes in metabolic diseases and muscular diseases implies that tau may have broader roles in cellular metabolism and muscle function. These findings imply that tau is involved in processes related to energy metabolism, cellular homeostasis, or muscle integrity. In genes that regulate tau, gene whose knock out result in a down regulation of tau are enriched in ALS (amyotrophic lateral sclerosis). This suggests that tau may have a protective or compensatory role in the context of ALS. The knockout of these genes leads to downregulation of tau expression, which may be related to the pathogenesis of ALS.

Furthermore, genes associated with psychotic disorders tend to downregulate tau when tau is knocked out. This suggests that tau may have a regulatory role in the pathophysiology of psychiatric disorders. Alterations in tau expression levels may contribute to neurobiological changes associated with neurodegenerative diseases.

Last but not least, genes that regulate tau and cause its downregulation are enriched in PML (progressive multifocal leukoencephalopathy) when tau is knocked out. This suggests that tau may be involved in the pathogenesis of PML. The dysregulation of tau protein resulting from these gene knockouts may contribute to the pathological processes associated with PML.

However, it’s important to note that the specific mechanisms and causal relationships require further investigation since these diseases have complex transcription channels. Genes up regulated by tau after tau knock out are enriched in regulation of p38MAPK cascade, cardiac development, and muscle morphogenesis. First, Tau may regulate the activation or expression of factors involved in p38MAPK signaling, influencing cellular responses and downstream gene expression. However, apoptotic DNA fragmentation is down regulated after tau knock out, meaning there is a positive relationship between tau and apoptotic DNA fragmentation, which can indicate apoptosis rate.[22]

After the knockout of the tau gene in cardiac development, the upregulation enrichment of the tau gene suggests its potential involvement in regulating key genes in cardiac development. The role of the tau protein in heart development may include the regulation of gene expression or signaling pathways that contribute to the normal formation and function of the heart.

The enrichment of genes upregulated by tau after tau knockout in muscle morphogenesis suggests that tau may be involved in the regulation of genes that control muscle development and formation. Tau's influence on muscle morphogenesis could involve the regulation of genes related to muscle cell differentiation, growth, or structural organization.
4. Discussion

In the study, we explored the potential causal association between tau and neurodegenerative diseases, specifically multiple sclerosis (MS), with a prior understanding that both β-amyloid and tau are biomarkers for aging. By extensively reviewing the available literature and conducting our own perturbation sequencing analysis, we have gathered evidence suggesting that tau may function as a downstream biomarker rather than a primary driver of these neurodegenerative disorders. [24], [25]

Our findings are consistent with a growing body of research showing that there is no direct causal relationship between tau and the pathogenesis of MS. [26], [27] Instead, tau may function as an indicator or outcome of the neurodegenerative processes underlying these diseases. Through perturbation sequencing methods, we were able to systematically analyze the effects of tau knockdown on the transcriptome, further bolstering the hypothesis of tau's downstream role in disease pathogenesis. [9]

Although our study supports the idea that tau is not the primary driving force behind AD and MS, it presents an intriguing avenue for further exploration. [28] Our results indicate that genes upstream of tau, such as regulatory genes ZNF774, SMClA, and DNA. IC17, may serve as true drivers of these neurodegenerative disorders. Further investigation of the function and dysregulation of these upstream genes could reveal the underlying mechanisms leading to disease progression.

However, we must acknowledge the limitations of our study. A significant limitation is the utilization of in vitro Perturb-seq cells, which may not fully represent the complexity of in vivo conditions. [29] The regulatory network of tau in the brain is likely to be more intricate and context-dependent, involving interactions with various cell types and external factors that were not captured in our in vitro model. Therefore, generalizations of our findings to in vivo trials should be approached with caution. Further studies using in vivo models should be conducted to address this limitation and ensure the accuracy of our conclusions. [30] Animal models or human tissue studies can provide a more comprehensive understanding of the role of tau and its regulatory network in the whole organism. Additionally, exploring the function of tau at different stages of disease progression and in response to various environmental factors may yield valuable insights into its significance as a downstream biomarker.

Moreover, by recognizing that tau is not an upstream factor in Alzheimer's disease, researchers can explore alternative therapeutic targets. This acknowledgment serves as a timely and effective guide for the development of new drugs and interventions that focus on other factors, such as neuroinflammation, vascular dysfunction, or genetic risk factors associated with the disease.

It should also be noted that our current understanding of neurodegenerative diseases is still evolving in terms of whether tau, β-amyloid, and other factors play a role in the disease, and this remains the subject of ongoing research. [31] Scientists in the field of neurodegenerative diseases continue to explore and refine their understanding of the treatment and diagnosis of Alzheimer's disease and other related disorders.

5. Conclusion

In conclusion, our study adds to the growing body of evidence that tau is not a primary driver of neurodegenerative diseases such as AD and MS but rather appears to be a downstream biomarker reflecting the underlying neurodegenerative disease process. The discovery of potential upstream driver genes provides new directions for future research, offering opportunities for therapeutic interventions targeting these genes rather than focusing solely on tau. However, the complexity of the regulatory network requires further investigation under in vivo conditions to fully validate our findings. By addressing these challenges, we can improve our understanding of neurodegenerative diseases and potentially discover new therapeutic strategies to combat these devastating diseases.
Acknowledgments

I owe a debt of gratitude to Dr. Ying, who is currently studying in Harvard University for his molecular biology and neurodegenerative diseases. He taught and helped me to perform molecular level experiments, while also revised my research approaches in the past two years. I would additionally like to thank my parents for their support, especially my father, who connected me with his colleagues with related research direction of enterprises and research institutions.

References


