Potential AD Pathogenic Protein Domains in PSEN1 Isoforms

Aanya Kamath (student)[1]

1. Evergreen Valley High School, 3300 Quimby Rd, San Jose, CA, 95135, USA

aanya.kamath@gmail.com

Abstract

This study identifies a potential AD pathogenic protein domain in a certain isoform generated from alternative splicing of the PSEN1 gene. Alternative splicing of the PSEN1 gene creates several functionally and structurally different isoforms of the PSEN1 protein. Certain isoforms possess AD pathogenic protein domains that other isoforms do not have. Using the STAR RNA-seq alignment algorithm, this study identifies alternative splicing events of PSEN1 transcripts that could potentially result in AD pathogenic protein domains in PSEN1 isoforms.

Keywords

Biology; Genetics; Alzheimer's Disease; Genetic Risk Factors; Presenilin 1

Introduction

Several genetic variants of the gene Presenilin 1 (PSEN1) characterized by Single Nucleotide Polymorphisms (SNPs) have been identified. Some of these genetic variants have been classified as genetic risk factors for Early Onset Alzheimer's Disease (EOAD). Possession of these genetic risk factors in addition to the alternative splicing of the PSEN1 genes affects an individual's risk of Developing Alzheimer's Disease (AD). AD is a neurodegenerative disease characterized by a decline in cognitive abilities such as reasoning and speaking. Symptoms of AD include memory loss, difficulty solving simple problems and completing daily tasks, mood changes, and increased anxiety[1]. In 2024, approximately 6.9 million Americans are living with Alzheimer's Disease. Early Onset AD is an onset of Alzheimer's Disease (AD) before the age of 65.

PSEN1 in Alzheimer's Disease

Genetic Variants of the gene Presenilin 1 (PSEN1), characterized by single nucleotide polymorphisms (SNPs) have been identified as risk for early onset AD (EOAD)[2]. However, only a small percentage of Americans living with Alzheimer's Disease have Early Onset AD. According to the Alzheimer's Association's 2024 Facts and Figures,

approximately 200,000 Americans are living with Early Onset AD. The remaining Americans living with AD have Late Onset AD, which is an onset of AD after the age of 65[3].

Alternative Splicing and Isoforms

Isoforms are versions of proteins that originate from the same gene, but have differences in their amino acid sequences that account for their different structure and function[4]. The formation of isoforms can result from alternative splicing, the process by which exons from a single gene are arranged in different ways, resulting in several mRNA transcripts generated from the same gene. These mRNA transcripts are translated into isoforms[5]. This study focuses on the isoforms of variant PSEN1 proteins. The protein PSEN1 is a component of the protein complex γ-secretase. γ-secretase cleaves amyloid precursor protein (APP) to form amyloid-beta (AB) peptides. An abnormal accumulation of AB in the form of plaques in the brain causes AD[2]. Alternative splicing of variant PSEN1 and PSEN2 genes can result in different protein isoforms that affect AD pathology by increasing or decreasing AB production. An abundance of a certain isoform of the variant PSEN1 and PSEN2 proteins can affect a certain individual's risk for AD. Some isoforms may have a certain protein domain, and the presence of the protein domain could determine whether the isoform increases or decreases risk for AD.

Protein Domains in Isoforms of APP

The Amyloid-Precursor Protein (APP) is the precursor protein of Amyloid-beta (AB) and its production is correlated to AD pathology. Ho et. al[6] identified 3 alternatively spliced isoforms of APP: APP751, APP770, and APP695. The study found out that the APP751 and APP770 isoforms contained a Kunitz-Protease Inhibitor (KPI) domain but the isoform APP695 did not. The study also found out that the KPI-free isoform APP695 was more abundant in the brain than the other APP isoforms, suggesting that the KPI domain has a role in AD pathology. The goal of this study is to identify a protein domain found in PSEN1 and PSEN2 isoforms that plays the same role as the KPI domain in APP isoforms.

Results

mRNA Transcripts

This study aligned 10 different PSEN1 mRNA transcripts to the chromosome 14 DNA transcript. All mRNA transcript sequences were obtained from the NCBI Entrez Database. This study used two PSEN1 transcripts with accession IDs starting with 'NM' and 'XM'. Accession IDs starting with 'NM' indicate mRNA transcripts from the International Nucleotide Sequence Database Collection (ISNDC) submissions or RefSeq records. Accession IDs starting with 'XM' are predicted mRNA transcripts generated from NCBI's genome annotation pipeline or computationally annotated submissions to the ISNDC.

CIGAR Analysis

Each alignment is contained in a BAM file generated by the STAR algorithm. The alignment is represented by a CIGAR (Compact Idiosynctratic Gapped Alignment Report) in the BAM file. The CIGARs of the provided mRNA sequences indicate how different PSEN1 mRNA transcripts align to the PSEN1 reference gene. Each CIGAR shows what parts of the mRNA transcript match the reference transcript and indicate the presence or absence of specific exons within the mRNA transcripts. (how does STAR produce the CIGAR alignments)

Transcript ID	CIGAR
NM_000021.4	77M11211N82M90N140M22690N251M2518N99M
NM_007318.3	77M11211N82M90N128M22702N251M2518N111M
XM_047431602.1	32M11346N82M90N128M22702N251M2518N142M13145N14M
XM_054376420.1	32M11346N82M90N128M22702N251M2518N142M13145N14M
XM_054376417.1	145M11215N82M90N128M22702N251M2518N43M
XM_054376413.1	145M11215N82M90N140M22690N251M2518N31M
XM_054376415.1	29M11346N82M90N140M22690N251M2518N144M
XM_047431601.1	29M11346N82M90N140M22690N251M2518N144M
XM_005267866.3	73M11215N82M90N128M22702N251M2518N115M
XM_005267864.4	73M11215N82M90N140M22690N251M2518N103M

Transcript	Exon 1	Intron Following Exon 1	Exon 2	Intron Following Exon 2	Exon 3	Intron Following Exon 3	Exon 4	Intron Following Exon 4	Exon 5	Intron Following Exon 5	Exon 6
XM_054376417.1	145 bp	11,215 bp	82 bp	90 bp	128 bp	22,702 bp	251 bp	2,518 bp	43 bp		
XM_054376413.1	145 bp	11,215 bp	82 bp	90 bp	140 bp	22,690 bp	251 bp	2,518 bp	31 bp		
NM_000021.4	77 bp	11,211 bp	82 bp	90 bp	140 bp	22,690 bp	251 bp	2518 bp	99 bp		

NM_007318.3	77 bp	11,211 bp	82 bp	90 bp	128 bp	22,702 bp	251 bp	2,518 bp	111 bp		
XM_047431602.1	32 bp	11,346 bp	82 bp	90 bp	128 bp	22,702 bp	251 bp	2,518 bp	142 bp	13145 bp	14 bp
XM_054376420.1	32 bp	11,346 bp	82 bp	90 bp	128 bp	22,702 bp	251 bp	2,518 bp	142 bp	13145 bp	14 bp
XM_054376415.1	29 bp	11,346 bp	82 bp	90 bp	140 bp	22,690 bp	251 bp	2,518 bp	144 bp		
XM_047431601.1	29 bp	11,346 bp	82 bp	90 bp	140 bp	22,690 bp	251 bp	2,518 bp	144 bp		

The 1st and 5th expressed exons in all transcripts have different splicing patterns. For the 3rd expressed intron there are two splicing patterns, one that results in a length of 128 bp and one that results in a length of 140 bp. Following the 3rd exon, there is an intron in the PSEN1 gene that is unexpressed in all of the transcripts. This intron has two main splicing patterns, one that results in 22,702 bp being spliced out of the PSEN1 gene and another that results in 22,690 bp being spliced out of the PSEN1 gene. The two transcripts just discussed both have the 22,702 bp spliced out. The final segment of the alignment “13145N14M” is not present in any of the other mRNA transcripts. The last 14 bp is an extra exon expressed in the two mRNA transcripts. The second to last exon of the two transcripts has a length of 142. This exon aligns with the last expressed exon in the other mRNA transcripts. This exon also has different splicing patterns in different transcripts.

Splicing Patterns of Transcripts XM_054376417.1 and XM_054376413.1

- 3rd Exon Splicing Patterns
  - The third expressed exon in XM_054376417.1 has 12 bp less than in XM_054376413.1 (128 bp vs. 140 bp).
  - Across all transcript variants, the 3rd expressed exon has either 128 bp or 140 bp, indicating two distinct splicing patterns.
- 3rd Intron Splicing Patterns
  - In XM_054376417.1, 22,702 bp following the 3rd exon in the reference are unexpressed in the transcript
  - In XM_054376413.1, 22,690 bp following the 3rd exon in the reference are unexpressed in the transcript.
  - These differences (22,702 vs. 22,690 bp) are consistent across all mRNA transcripts, reflecting two distinct splicing patterns at the 3rd intron
- Final Exon Splicing Patterns
  - The final exon in XM_054376417.1 is 43 bp
  - The final exon in XM_054376413.1 is 31 bp
  - This represents a third difference in the splicing of the two transcripts.

Splicing Patterns of Transcripts NM_000021.4 and NM_007318.3

- 3rd Exon Splicing Patterns
    - Length: 140M in NM_000021.4 and NM_007318.3, compared to 128M in XM_054376417.1 and XM_054376413.1.
- 3rd Intron Splicing Patterns:
    - Length in NM_000021.4: 22,690 bp.
    - Length in NM_007318.3: 22,702 bp.
- Final Exon Splicing Patterns:
    - NM_000021.4 final exon: 99 bp.
    - NM_007318.3 final exon: 111 bp.

Summary

The CIGARs generated from aligning mRNA transcripts to the DNA reference demonstrated 3 main splicing patterns:

1. All Spliced transcripts showed either 128 bp. or 140 bp in their third expressed exon.
2. The third unexpressed intron in spliced transcripts had a length of either 22,690 bp or 22,702 bp.
3. Spliced transcripts XM_047431602.1 and XM_054376420.1 contained an extra expressed exon of 14 bp, not found in other transcripts.
    a. The presence of this extra exon, combined with their consistent length of 50,450 bp, which is longer than the average PSEN1 mRNA transcript, indicates a distinct alternative splicing pattern. In the context of PSEN1 isoform diversity, these results imply that XM_047431602.1 and XM_054376420.1 are examples of isoforms with specific roles. Their unique structure and extended transcript length may influence PSEN1's molecular functions or its interaction within cellular pathways.

Discussion

These findings suggest certain implications of PSEN1 splicing and its potential contribution to diseases. The presence of extra exons in PSEN1 transcripts XM_047431602.1 and XM_054376420.1 indicates the potential for certain splicing patterns to contribute to Alzheimer's disease (AD) pathogenesis. The PSEN1 gene encodes a core component of the γ-secretase complex, which plays a critical role in cleaving amyloid precursor protein (APP) into amyloid-beta (Aβ) peptides. Aberrant splicing of the PSEN1 gene could lead to isoforms with altered sequences, affecting the structure or function of the PSEN1 protein. This may disrupt γ-secretase activity, altering the balance of Aβ42/Aβ40 peptides, a characteristic of AD pathology. The consistency of these splicing patterns across transcripts suggests a regulated but potentially aberrant splicing event, possibly triggered by mutations or environmental factors associated with AD.

Methods

This study utilizes the DNA sequence of the PSEN1 gene and various mRNA transcripts of the PSEN1 gene from the NCBI database. Several alternatively spliced mRNA sequences of PSEN1 were aligned to the PSEN1 DNA sequence using the Spliced Transcripts Alignment to a Reference (STAR) algorithm. The STAR alignment identifies the matches between each mRNA transcript and the reference as well as the regions of the reference that are missing from the transcript. Because a protein domain is often coded by a single distinct exon, protein domains that are present or missing in certain mRNA isoforms can be identified based on which exons are present or missing in the transcripts. An mRNA transcript with an exon(s) present or missing may code for an AD pathogenic PSEN1 isoform. The procedure used to find missing or extra exons in PSEN1 mRNA transcripts consists of 4 steps:

1. Retrieval: mRNA transcripts and the DNA transcript for chromosome 14, which the PSEN1 gene lies on, are retrieved from the NCBI database using Biopython. Biopython retrieves PSEN1 mRNA and DNA transcripts in FASTA format.
2. Alignment: Each mRNA transcript are aligned to the DNA transcript using STAR. STAR is an RNA-seq aligner that can align RNA-seq reads to a reference genome. STAR produces alignments that can identify differences in the exons expressed in different mRNA transcripts due to certain splicing patterns.
3. Visualization: The STAR alignments are visualized in the Interactive Genome Viewer (IGV) application.
4. Analysis: The differences in the alignments are analyzed to find the different splicing patterns of the PSEN1 gene and certain exons that are present in certain isoforms but absent in others, which could code for potentially AD pathogenic protein domains.

Detailed Description of Methods

Retrieval

This study used the Bio.Entrez module in biopython to obtain genetic sequences from the National Center for Biotechnology Information's (NCBI) Entrez Molecular Sequence Database System. The efetch command in the Bio.Entrez module retrieves sequences from the NCBI Entrez Database when given an ID and a requested format. The following procedure is used for retrieval:

- mRNA transcripts of the PSEN1 gene and a reference DNA transcript of the chromosome are retrieved using Biopython

- mRNA sequences are retrieved using ID starting with ‘NM’, which indicates that the sequence is an mRNA transcript
- The DNA sequence of chromosome 14 is retrieved using an ID starting with ‘NC’, which indicates that the sequence represents a chromosome.

Alignment and Visualization

Once biopython has been used to obtain an mRNA transcript for either PSEN1 or PSEN2 and a DNA reference transcript of either Chromosome 14 or Chromosome 1, STAR can align the two. STAR takes in the two FASTA files containing the sequences as input and produces a BAM file containing the alignment as output.

The procedure for alignment and visualization is:

1. The STAR algorithm was run with each mRNA transcript and the DNA transcript as input to produce the alignment.
2. Visualization: the BAM file is visualized in IGV (Interactive Genome Viewer).

All alignments discussed in this study are primary alignments. Primary alignments represent the location in the reference genome that matches the mRNA read the best. All other alignments are secondary alignments. STAR aligns each PSEN1 mRNA sequence to the Chromosome 14 reference genome and produces a BAM file. The BAM file contains data on the alignment. It contains the boundaries of the alignment and the specific length of the alignment in bp (base pairs). It states whether the alignment is Primary or Secondary. The BAM file contains:

- Whether the alignment is primary or secondary
- the Alignment Score (AS) that indicates how well the mRNA transcript aligns the reference transcript. The score is numerical and calculated from the gaps in the alignment and mismatches
- a Compact Idiosyncratic Gapped Alignment Report (CIGAR), which identifies the matches between each mRNA read and the reference as well as the regions of the reference that are skipped in the read. An example CIGAR string, from an alignment between a PSEN1 mRNA read and the Chromosome 14 reference genome, “77M11211N82M90N140M22690N251M2518N99M”.

Interactive Genome Viewer (IGV) application was used to view each BAM file.

Analysis

In a CIGAR string, ‘M’ refers to the bases that match between the read and reference, whereas ‘N’ refers to the bases that are skipped in the read, indicating introns in the reference. The CIGAR string leads from left to right of the alignment. “77M11211N” means that starting from the left of the alignment, 77 bases match between the read and reference and then 11,211 bases are skipped in the read. The 11,211 bases skipped in the read indicate an intron in the reference that is not expressed in the read as a result of splicing.

- The CIGAR strings can be analyzed to identify the different splicing patterns of the PSEN1 gene
    - In a CIGAR string, ‘M’ refers to the bases that match between the read and reference
    - ‘N’ refers to the bases that are skipped in the read, indicating introns in the reference transcript.

Conclusion

This study identified certain isoforms of the protein PSEN1 that could potentially correlate with increased production of pathogenic Aβ42 peptides or other disruptions in γ-secretase activity. Further research is needed to determine whether these splicing events are directly linked to the pathology of AD and whether they could serve as therapeutic targets for regulating PSEN1 activity to restore normal Aβ production.

Acknowledgements

I would like to thank Dr. Anil Kumar Challa and Purushotham Kamath for their assistance and advice.

References

1. F. Bature, B. Guinn, D. Pang, Y. Pappas. Signs and symptoms preceding the diagnosis of Alzheimer’s disease: a systematic scoping review of literature from 1937 to 2016. *BMJ Open*, **7(8)**, e015746, (2017) https://doi.org/10.1136%2Fbmjopen-2016-015746
2. J. Bagaria, E. Bagyinszky, & S. S. A. An. Genetics, Functions, and Clinical Impact of Presenilin-1 (PSEN1) Gene. *International Journal of Molecular Sciences*, **23(18)**, 10970, (2022) https://doi.org/10.3390/ijms231810970
3. 2024 Alzheimer’s Disease Facts and Figures [Review of 2024 Alzheimer’s Disease Facts and Figures] *Alzheimer’s Association*. (2024) https://www.alz.org/media/documents/alzheimers-facts-and-figures.pdf
4. M. Stastna, & J. E.Van Eyk Analysis of protein isoforms: can we do it better?. *Proteomics*, **12(19-20)**, 2937–2948, (2012). https://doi.org/10.1002/pmic.201200161
5. Q. Liu, L. Fang, & C. Wu. Alternative Splicing and Isoforms: From Mechanisms to Diseases. *Genes*, **13(3)**, 401, (2022). https://doi.org/10.3390/genes13030401
6. L. Ho, K. Fukuchi, & S.G. Younkin. The Alternatively Spliced Kunitz Protease Inhibitor Domain Alters Amyloid b Protein Precursor Processing and Amyloid b Protein Production in Cultured Cells [Review of The Alternatively Spliced Kunitz Protease Inhibitor Domain Alters Amyloid b Protein Precursor Processing and Amyloid b Protein Production in Cultured Cells]. *Journal of Biological Chemistry.* (1996) https://www.jbc.org/article/S0021-9258(19)79152-2/pdf