Phylogenetic and Structural Analysis of BC200 and *Hominoidea* Homologs

Derek Robinson, Mina Emadi, Mazyar Ghezelji
University of Victoria
Victoria, Canada
{drobinson, minaemadi, mazyarghezelji}@uvic.ca

Abstract—**Alzheimers disease (AD) is a neurodegenerative disorder resulting from synaptic plasticity failure in neurons, which greatly affects the cognitive ability of patients. AD is linked to the accumulation of beta-amyloid protein and tau protein, which in abnormal levels can disrupt cell functions and potentially lead to diseases such as AD. Previous studies have suggested that AD is a uniquely human disease. In this paper, we investigate this claim by analyzing genetic similarities between humans and other species of the *Hominoidea* superfamily. We break down our problem into two crucial parts; performing phylogenetic analysis of a long non-coding RNA associated with AD named Brain Cytoplasmic 200 lncRNA (BC200), and investigating the structural differences between BC200 and its four most closely related *Hominoidea* homologs. We start by first performing a NCBI Blast search to determine similar sequences, followed by building a phylogenetic tree to determine closely related homologs. We then observe the most similar sequences and build their secondary structures, which are then used to compare functions of BC200 and its homologs. Although these sequences are similar, secondary structure comparison revealed structural differences between human BC200 and its homologs, meaning that they may exert different functions in the body. As a result, our findings neither support nor refute that AD is a uniquely human disease.**

I. INTRODUCTION

Alzheimer's disease (AD) is a neurodegenerative disease which greatly affects the lives of those who are diagnosed and those who care for the diagnosed. In the United States, 6.2 million patients are estimated to be living with AD and it is the sixth leading cause of death for American adults [1]. In Canada, over 747,000 patients are living with AD, or another form of dementia [2]. In mild cases, patients with AD may exhibit memory loss and poor judgement [3]. The symptoms of higher severity cases may include the the inability to communicate, seizures, and difficulty swallowing [3].

AD involves multiple cell types and signaling pathways [4]. As such, the collective knowledge of AD is spread across many different domains. This spread of knowledge means that fully understanding AD in humans is difficult, let alone attempting to understand if the disease affects other species of the *Hominoidea* superfamily. Finch and Austad argue that with our current understanding of AD, it is not possible to determine if AD is uniquely human [5]. Thus, this paper explores one facet of AD pathogenesis, the long non-coding RNA (lncRNA) BC200. BC200 has been implicated in AD bu upregulating the expression of b-site APP-cleaving enzyme1 (BACE1) [4], [6]. The upregulation of BACE1 in turn leads to higher levels of beta-amyloid ($A\beta$) in the brain, which can forms plaques, accumulate in neurons, and disrupt cell function. [4], [6].

In hopes of better understanding the role that *Hominoidea* homologs of BC200 play in AD, we answer the following research questions:

RQ1. What does the phylogenetic tree of BC200 look like?

RQ2. What structural differences exist between BC200 and its four most closely related *Hominoidea* homologs?

The remainder of this paper is structured as follows: section II gives background into BC200 and the role it plays in AD, section III lays out the methods used for selecting BC200 homologs and performing both the phylogenetic structural analysis, section IV presents the results of our analysis, section V discusses the relevancy of the results, and section VI concludes the paper.

II. Background

A. What is BC200?

BC200, sometimes referred to as BCYRN1, is a 200 nucleotide long RNA transcript which is found predominantly in the brain [7]. As a lncRNA, BC200 is not translated into protein, but can be used as a potential therapeutic target and biomarker due to its regulatory role in biological processes involved in disease development [4], [8]. This lncRNA has recently been studied extensively due to its regulatory role in translation, as well as its pathogenic impact in AD [4], [7]. Because of the crucial role BC200 plays in translation control, it impacts the synthesis of dendritic proteins which facilitates long-term plastic changes at the synapse [8].

B. The Relation Between AD and BC200

AD is a neurodegenerative disease resulting from synaptic plasticity failure in neurons [8]. One protein implicated in the onset of AD is beta-amyloid. Aβ, a cleavage product of the amyloid precursor protein (APP), is generated by b-site APP-cleaving enzyme1 (BACE1) and γ-secretase complex. Beta-amyloid strongly influences the pathogenesis of AD due to the formation of plaques, accumulation in neurons, and disruption of cell functions. BC200 facilitates AD pathogenesis by increasing Aβ production through the upregulation of BACE1 expression. The inhibition of BC200 significantly suppresses BACE1 expression, increases cell viability and reduces cell apoptosis in an AD model, and these effects are reversed by BC200 over-expression [4], [6]. Inhibition of BACE1 activity subsequently leads to a reduction in Aβ levels and may cure or prevent AD [4], [6].

Many researches have demonstrated the important role of BC200 in AD. Mus *et al.* show that there is a steady decline in BC200 levels from ages 49 to 86, but a substantially higher amount was detected in AD patients [8]. They also observe that BC200 expression is increased in brain areas that are implicated in AD and it is correlated with disease severity. Li *et al.* establish an AD cell model which over-expresses Aβ 1-42 to observe the effects of BC200 on cell viability on cell viability, apoptosis, and to investigate it's mechanistic role on these processes [6]. They observe that knockdown of BC200 led to an increase in cell viability and a decrease in cell apoptosis via the mediation of BACE1. Inhibition of BC200 rescued this Aβ 1-42 mediated dysfunction, as indicated by the interaction of BC200 directly targeting

BACE1. Moreover, inhibition of BC200 increased AD cell growth and reduced cell apoptosis. They demonstrate that BC200 is a potent positive regulator of BACE1 in AD cells and in conclusion, lncRNA BC200 facilitates AD pathogenesis by up-regulating Aβ through BACE1.

C. What are Alu domains and what is their role in body?

Alu domains (or elements) are sections of RNA which are responsible for for the regulation of tissue-specific genes. Additionally, they are involved in the transcription of nearby genes and may change the expression of said gene. In some cases, Alu domains can be detrimental and result in inherited disorders such as AD, Lung cancer, and Gastric cancer [9].

D. Sequence Conservation

Over many generations, random mutations and deletions can change nucleic acid sequences in the genome of an evolutionary lineage. Chromosomal rearrangement may also recombine or delete sequences. Conserved sequences are sequences which persist in the genome despite such forces, and have slower rates of mutation than the background mutation rate [10]. Non-coding sequences important for gene regulation, such as the binding or recognition sites of ribosomes and transcription factors, may be conserved within a genome. For example, the promoter of a conserved gene or operon may also be conserved.

III. Materials and Methods

The methodology workflow we followed is presented in Fig. 1.

A. Selection of lncRNAs

The lncRNA BC200 was selected for phylogenetic and structural analysis due to the role it plays in AD as discussed in section II. The homologs of BC200 were selected as a result of an NCBI Blast [11] search. Specifically, Megablast [12] with default parameters was used as it is able to compare closely related sequences [13]. From the Blast results, the top six sequences of known BC200 homologs were chosen , identified by the inclusion of BC200 in their name. Two of the top hits in the Blast results were complete bacterial artificial chromosome sequences for *Pan troglodytes*. Inclusion of these two sequences (accession numbers: AC185986.3,
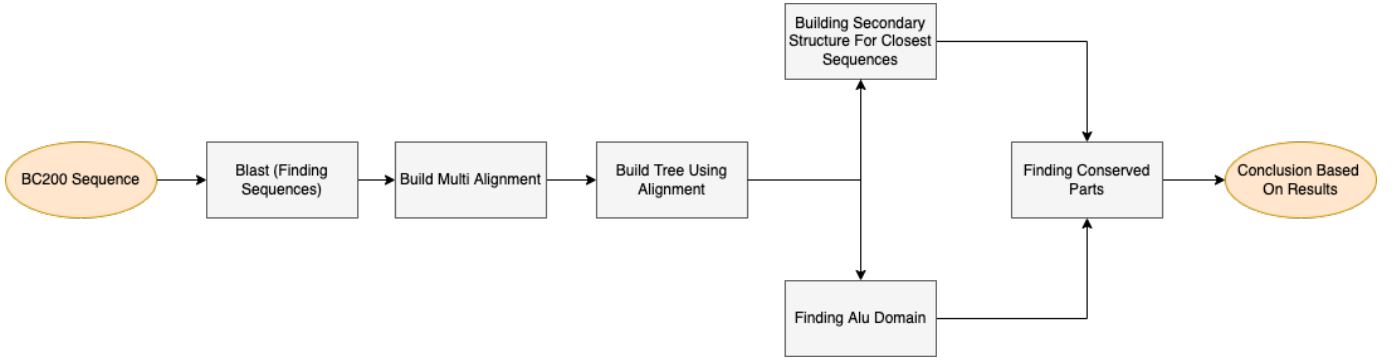
Fig. 1: Methodology Workflow

AC183594.3) caused errors in the analysis software described in section III-B and section III-C, as such we did not consider these full chromosome sequences in our analysis. Table I outlines the organism names and accession numbers for each of the chosen sequences.

TABLE I: lncRNA Sequence Information

Organism Name	Sequence Accession Number
Homo sapiens	NR_001568.1
Pongo pygmaeus	AF067780.1
Pan paniscus	AF067778.1
Gorilla gorilla	AF067779.1
Macaca mulatta	AF067784.1
Hylobates lar	AF067781.1
Papio hamadryas	AF067782.1

B. Phylogenetic Analysis

The lncRNA sequences in Table I were subject to phylogenetic analysis. The phylogenetic tree was built using the MEGA11 [14] bioinformatics software. The sequences outlined in Table I were aligned using the MEGA11 alignment module. The multiple sequence alignment was performed using the MUSCLE algorithm [15] with default parameters in the MEGA11 alignment module. Once the multiple sequence alignment had been performed, it was then possible to build the phylogenetic tree.

The phylogenetic tree was built using the maximum likelihood statistical method and Tamura-Nei substitution model [16], [17]. The bootstrap method was used as a test of phylogeny with 500 replications. In order to further support the phylogenetic tree, estimates of evolutionary divergence were also calculated by MEGA11 [14]. Evolutionary divergence estimates were calculated using Maximum Composite Likelihood and the Tamura-

Nei model [16], [17]. All ambiguous positions were removed for each sequence pair (pairwise deletion option).

C. Structural Analysis

In order to approximate the RNA secondary structure of BC200 and four of its closest homologs, we used FORNA, a platform which estimates RNA secondary structure [18]. FORNA is able to draw the RNA secondary structure from the dot bracket notation provided by the user or approximate the minimum free energy (MFE) structure as calculated by RNAFold [19]. As this analysis aims at computationally determining a secondary structure, the latter option of using the MFE structure was employed. Using structural similarity as a metric for functional similarity, we compared the structure of BC200 to its homologs. The approximated BC200 structure consists of three main domains, the Alu, A-rich, and unique domain. Comparing each homolog structure into BC200, we divided them to three main domains as seen in the BC200 structure and analyzed structural differences in each domain.

D. Finding Alu Domains

In order to find the Alu domains present in our sequences, we used Dfam [20]. The Dfam database is an open collection of Transposable Element DNA sequence alignments, hidden Markov Models (HMMs), consensus sequences, and genome annotations. By using the sequence search we identified the Alu domains of the BC200 and its homologs. We performed a sequence search for each of the target lncRNAs and compared the first row names of the diagram for each sequence search. Comparing the Alu domain structure gives us a suitable result for the RNA function.

E. Finding Conserved Parts of BC200

One of the best ways for visualize conserved sequences during evolution, is multiple sequence alignment. Here, we aligned BC200 and the four most related homologs with each other to find the conserved parts of the sequence. We used Clustal Omega for aligning the sequences [21]. The CLUSTAL format includes a plaintext key to annotate conserved columns of the alignment, denoting conserved sequence (*), conservative mutations (:), semi-conservative mutations (.), and non-conservative mutations ( ). The guide tree for the multiple sequence alignment is shown in Fig. 2. The conservation percentage of BC200 was calculated by comparing the number of conserved nucleotides to number of total nucleotides in the sequence.
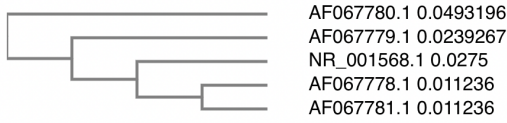
Fig. 2: Multiple Sequence Alignment Guide Tree

IV. RESULTS

A. Phylogenetic Tree of BC200

The results of the phylogenetic analysis can be found in Fig. 3. The tree shows that the *Homo sapiens* BC200 lncRNA and the *Gorilla gorilla* BC200 lncRNA are likely to share a common ancestor. Similarly, the BC200 homologs from *Pan paniscus* and *Hylobates lar* are likely to share a common ancestor. This is also the case for the BC200 homologs from *Macaca mulatta*, *Papio hamadryas*, and *Pongo pygmaeus*.

Additionally, Table II shows the pairwise distances between sequences which estimate the evolutionary divergence between sequences. The evolutionary divergence estimates show that the four most closely related homologs of BC200 found in other *Hominoidea* come from *Pan paniscus* (AF067778.1), *Pongo pygmaeus* (AF067780.1), *Hylobates lar* (AF067781.1), and *Gorilla gorilla* (AF067779.1), in order of evolutionary distances. Thus, the aforementioned BC200 homologs (AF067778.1, AF067780.1, AF067781.1, AF067779.1) are the closest relatives of BC200 and have their secondary structure analyzed in section IV-B.
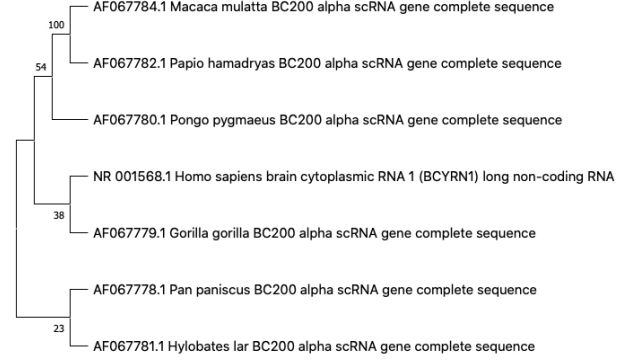
Fig. 3: Phylogenetic tree of BC200 and *Hominoidea* homologs. The numbers at each node are the bootstrap support values obtained by maximum likelihood.

B. Structural difference between BC200 and its homologs

The BC200 structure consist of three main parts: the Alu, A-rich, and unique domain [22] as shown in Fig. 4.
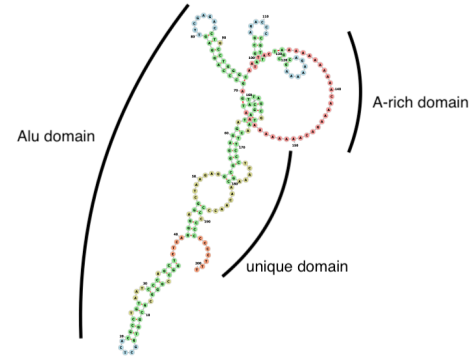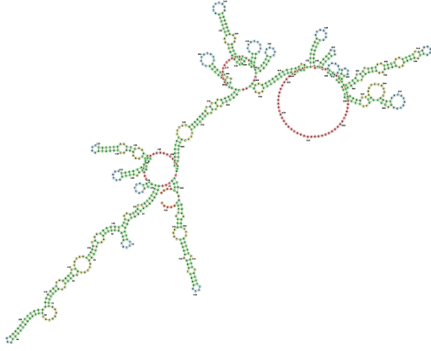
Fig. 4: BC200 RNA Secondary Structure

By comparing the Alu domain of BC200 and its homologs (see Appendix A), we concluded that the Alu domain between human BC200 and its selected homologs are nearly identical. From this conclusion, we can infer that the functions of each of the homologs may be the same as those of BC200. Fig. 5a, 5b, and 5c show the RNA secondary structure of the three most closely related BC200 homologs from great apes (*Hominidae*). Additionally, we have chosen to depict the RNA secondary structure of the *Hylobates lar* BC200 homolog in Fig. 5d. While *Hylobates lar* is not a great ape, it is still part of the *Hominoidea* superfamily, and Phylogenetic analysis revealed that it was closely related to BC200.

TABLE II: Estimates of Evolutionary Divergence between Sequences

Accession Number						
NR_001568.1						
AF067780.1	0.01023					
AF067778.1	0.00503	0.02169				
AF067779.1	0.03070	0.02890	0.01709			
AF067784.1	0.03088	0.04573	0.03802	0.05235		
AF067781.1	0.02561	0.02763	0.01562	0.02439	0.04353	
AF067782.1	0.03120	0.04478	0.03947	0.04731	0.00852	0.04259

(a) BC200 RNA Secondary Structure of *Gorilla gorilla*
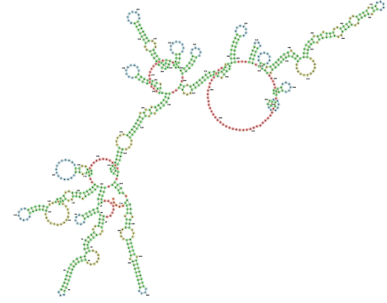
(b) BC200 RNA Secondary Structure of *Pan paniscus*
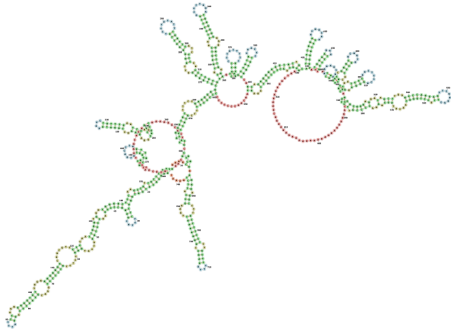
(c) BC200 RNA Secondary Structure of *Pongo pygmaeus*
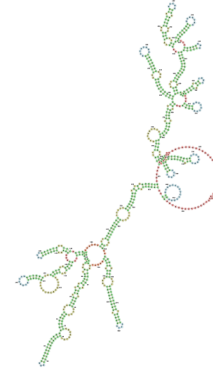
(d) BC200 RNA Secondary Structure of *Hylobates lar*

Fig. 5: RNA Secondary Structures of *Hominoidea* BC200 Homologs, generated by FORNA [18] using default parameters

C. Finding Conserved Parts of BC200

Using multiple sequence alignment we aligned BC200 with its four mostly related homologs to find conserved parts of BC200. The result of the multiple sequence alignment can be found in Appendix B

From the multiple sequence alignment, we can conclude that a majority (88%) of the BC200 homolog sequences are conserved. Sequence conservation in non-coding RNAs is generally poor compared to protein-coding sequences, and base pairs that contribute to structure or function are often conserved instead [23]. The result shows that there should be a high similarity between the function of BC200 in *Homo sapiens* and closely related species. So, the following result does not support

or refuse our hypothesis, the homologs that we found didn't have the same secondary structure, which import scoffolding function to RNA. But they had the same Alu elements, so we cannot state that whether their function is different in differenct species or not.

V. Discussion

In this study we first began by gathering known homologs of *Homo sapiens* BC200 via NCBI Blast [11], [24]. Using these homologs, we built the phylogenetic tree in order to determine which homologs were the most evolutionarily related. Once we had determined that the BC200 homologs from *Pan paniscus*, *Pongo pygmaeus*, *Hylobates lar*, and *Gorilla gorilla* were the most closely related to *Homo sapiens* BC200, we used FORNA to determine their secondary structure [19]. We found that the secondary structures of *Homo sapiens* BC200, shown in Fig. 4, differs from that of of *Gorilla gorilla*, *Pan paniscus*, and *Pongo pygmaeus* BC200, shown in Fig. 5a, 5b, 5c. However, by comparing the Alu domains present in *Homo sapiens* BC200 with the Alu domains of *Gorilla gorilla*, *Pan paniscus*, and *Pongo pygmaeus* BC200 we can say that their gene regulation functions may be similar.

Here we showed that several *Hominoidea* homologs of BC200 share evolutionary ancestors as shown in Fig. 3. Additionally, we have shown that the four most closely related *Hominoidea* homologs of *Homo sapiens* BC200 have differing RNA secondary structures. Although many papers suggest that beta-amyloid may cause AD, there is biological evidence that contradicts this hypothesis [25]. As mentioned above, overexpression of BC200 will increase expression of BACE1, which in turn generates beta-amyloid. Despite the high similarity between BC200 lncRNA in *Homo sapiens* and in *Pan paniscus* and *Pongo pygmaeus*, this does not definitively answer the question of whether or not AD is a uniquely human disease. This indicates that further research in this area is needed to draw a definitive conclusion. Alzheimers disease is extremely complex and this work is only aimed determining one facet of how AD may proceed in *Hominoidea* other than *Homo sapiens*. In a recent review paper, Drummond *et al.* discuss the many models of AD pathology including transgenic mice/rats, invertebrate animals, and *in vitro* human cell culture models [26]. After reviewing the many models of AD pathology, they conclude that AD is a uniquely human disease [26]. Although, the conclusion

of Drummond *et al.* is not supported by our observations, it does not necessarily contradict it. Our results show that further research involving laboratory methods are necessary to understand the role of BC200 in other *Hominoidea*.

A. Limitations

Our analysis began by selecting similar sequences via NCBI Blast. As the sequences we chose were selected based on being known BC200 homologs, we may have missed sequences which are potential homologs. An example is the exclusion of the bacterial artificial chromosome sequences for *Pan troglodytes*. If we had included these sequences and were able to isolate the section of the chromosome which encodes for BC200 our Phylogenetic tree would have contained an extra data point. After performing phylogenetic analysis, we built the RNA secondary structure of BC200 and its homologs. These structures have not been experimentally confirmed and are based on the minimum free energy. Consequently, the structures are approximations are should be thought of as such.

VI. Conclusion

In this paper, we set out to determine if there was bioinformatic evidence that supports the hypothesis that AD is uniquely human. In order to determine if such evidence exists we first performed phylogenetic analysis to determine candidate species for RNA secondary structure analysis. Once we had determined the RNA secondary structure for the four most closely related *Hominoidea* homologs of BC200, we compared their secondary structures to the secondary structure of *Homo sapiens* BC200. We determined that the Alu domain of *Homo sapiens*, *Pan paniscus*, and *Pongo pygmaeus*, *Gorilla gorilla* and *Hylobates lar* BC200 are structurally similar. Further analysis determined that 88% of *Homo sapiens* BC200 sequence is conserved during evolution, as shown in Appendix B. While much of the *Homo sapiens* BC200 sequence is conserved, the BC200 homologs were much longer than that of *Homo sapiens* BC200. The result of a different RNA secondary structure may suggest that BC200 has a different regulatory function each host organism. Other results, including having the same Alu domains, the connection of each homolog through the phylogenetic tree, and the existence of a large portion of conserved nucleotides in BC200, suggests that their function may be similar. In conclusion, our analysis

of BC200 homologs in *Hominoidea* suggest that they may have conserved function in the body, but that more research on the topic is required to determine if AD is uniquely human.

REFERENCES

[1] "Alzheimer's disease — disease or condition of the week — cdc," https://www.cdc.gov/dotw/alzheimers/index.html#:~:text=Alzheimer's%20disease%20is%20the%20most,of%20death%20for%20all%20adults., June 2021, (Accessed on 10/19/2021).

[2] "Alzheimer's and dementia in canada," https://www.alz.org/ca/dementia-alzheimers-canada.asp, 2021, (Accessed on 10/19/2021).

[3] "What are the signs of alzheimer's disease? — national institute on aging," https://www.nia.nih.gov/health/what-are-signs-alzheimers-disease, May 2017, (Accessed on 11/21/2021).

[4] Y. Zhang, Y. Zhao, X. Ao, W. Yu, L. Zhang, Y. Wang, and W. Chang, "The role of non-coding rnas in alzheimers disease: From regulated mechanism to therapeutic targets and diagnostic biomarkers," *Frontiers in Aging Neuroscience*, p. 384, 2021.

[5] C. E. Finch and S. N. Austad, "Commentary: is alzheimer's disease uniquely human?" *Neurobiology of aging*, vol. 36, no. 2, pp. 553–555, 2015.

[6] H. Li, L. Zheng, A. Jiang, Y. Mo, and Q. Gong, "Identification of the biological affection of long noncoding rna bc200 in alzheimers disease," *Neuroreport*, vol. 29, no. 13, pp. 1061–1067, 2018.

[7] H. Tiedge, W. Chen, and J. Brosius, "Primary structure, neural-specific expression, and dendritic location of human bc200 rna," *Journal of Neuroscience*, vol. 13, no. 6, pp. 2382–2390, 1993.

[8] E. Mus, P. R. Hof, and H. Tiedge, "Dendritic bc200 rna in aging and in alzheimer's disease," *Proceedings of the National Academy of Sciences*, vol. 104, no. 25, pp. 10 679–10 684, 2007.

[9] Y.-M. Tseng, S.-M. Tsai, C.-C. Lin, Y.-R. Jin, W.-H. Yeh, J.-K. Hsiao, C.-F. Chen, W.-H. Lan, and L.-Y. Tsai, "Oxidative stress-related enzyme polymorphisms associated with the immunological biomarkers levels in heavy drinkers in taiwan," *Journal of clinical laboratory analysis*, vol. 27, no. 6, pp. 494–503, 2013.

[10] M. Kimura and T. Ohta, "On some principles governing molecular evolution," *Proceedings of the National Academy of Sciences*, vol. 71, no. 7, pp. 2848–2852, 1974.

[11] "Nucleotide blast: Search nucleotide databases using a nucleotide query," https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome, (Accessed on 10/20/2021).

[12] A. Morgulis, G. Coulouris, Y. Raytselis, T. L. Madden, R. Agarwala, and A. A. Schäffer, "Database indexing for production megablast searches," *Bioinformatics*, vol. 24, no. 16, pp. 1757–1764, 2008.

[13] F. Amirmahani and K. Jamshidi Goharrizi, "Phylogenetic analysis of three long non-coding rna genes: Ak082072, ak043754 and ak082467," *Journal of Genetic Resources*, vol. 4, no. 1, pp. 56–64, 2018.

[14] K. Tamura, G. Stecher, and S. Kumar, "Mega11: molecular evolutionary genetics analysis version 11," *Molecular Biology and Evolution*, vol. 38, no. 7, pp. 3022–3027, 2021.

[15] R. C. Edgar, "Muscle: a multiple sequence alignment method with reduced time and space complexity," *BMC bioinformatics*, vol. 5, no. 1, pp. 1–19, 2004.

[16] K. Tamura and M. Nei, "Estimation of the number of nucleotide substitutions in the control region of mitochondrial dna in humans and chimpanzees." *Molecular biology and evolution*, vol. 10, no. 3, pp. 512–526, 1993.

[17] K. Tamura, M. Nei, and S. Kumar, "Prospects for inferring very large phylogenies by using the neighbor-joining method," *Proceedings of the National Academy of Sciences*, vol. 101, no. 30, pp. 11 030–11 035, 2004.

[18] P. Kerpedjiev, S. Hammer, and I. L. Hofacker, "Forna (force-directed rna): Simple and effective online rna secondary structure diagrams," *Bioinformatics*, vol. 31, no. 20, pp. 3377–3379, 2015.

[19] R. Lorenz, S. H. Bernhart, C. Höner zu Siederdissen, H. Tafer, C. Flamm, P. F. Stadler, and I. L. Hofacker, "Viennarna package 2.0," *Algorithms for molecular biology*, vol. 6, no. 1, pp. 1–14, 2011.

[20] J. Storer, R. Hubley, J. Rosen, T. J. Wheeler, and A. F. Smit, "The dfam community resource of transposable element families, sequence models, and genome annotations," *Mobile DNA*, vol. 12, no. 1, pp. 1–14, 2021.

[21] F. Madeira, Y. M. Park, J. Lee, N. Buso, T. Gur, N. Madhusoodanan, P. Basutkar, A. R. Tivey, S. C. Potter, R. D. Finn *et al.*, "The embl-ebi search and sequence analysis tools apis in 2019," *Nucleic acids research*, vol. 47, no. W1, pp. W636–W641, 2019.

[22] E. Jung, J. Lee, H. J. Hong, I. Park, and Y. Lee, "Rna recognition by a human antibody against brain cytoplasmic 200 rna," *Rna*, vol. 20, no. 6, pp. 805–814, 2014.

[23] P. Johnsson, L. Lipovich, D. Grandér, and K. V. Morris, "Evolutionary conservation of long non-coding rnas; sequence, structure, function," *Biochimica et Biophysica Acta (BBA)-General Subjects*, vol. 1840, no. 3, pp. 1063–1071, 2014.

[24] T. Madden, "Chapter 16: The blast sequence analysis tool, the ncbi handbook (2002)," 2012.

[25] D. J. Selkoe and J. Hardy, "The amyloid hypothesis of alzheimer's disease at 25 years," *EMBO molecular medicine*, vol. 8, no. 6, pp. 595–608, 2016.

[26] E. Drummond and T. Wisniewski, "Alzheimers disease: experimental models and reality," *Acta neuropathologica*, vol. 133, no. 2, pp. 155–175, 2017.

APPENDIX A
DFAM ANALYSIS OF ALU DOMAINS

A. *Homo sapiens*

B. Gorilla gorilla

C. Pan paniscus

D. Pongo pygmaeus

CLUSTALW MULTIPLE SEQUENCE ALIGNMENT OF BC200 AND *Hominoidea* HOMOLOGS

```
CLUSTAL O(1.2.4) multiple sequence alignment


AF067780.1      GAAGCCAGCAAGAAGAACATGGTGGACTCTCTGTGAGGCCCCAAACCCCAAATCATGAGC
AF067779.1      GAAGCCAGCAAGAAGAACATGGTGGACTCTCTGTGAGGCCCCAAACCCCAAATCATGAGC
NR_001568.1     ------------------------------------------------------------
AF067778.1      -------------------------------------------CCAAACCCCAAATCATGAGC
AF067781.1      -------------------------------------------CCAAACCCCAAATCATGAGC


AF067780.1      CTCTGCCCCCTGAACGTATTCCAGGATCTCCAAGGGAG-TATTGGGAGGGGGCCAACGTC
AF067779.1      CTCCGCCCCCTGAAAGTATTCCAGGATCTCCGAGGGAGGTTTTGGGAGGGGGCCAAGGTC
NR_001568.1     ------------------------------------------------------------
AF067778.1      CTCCGCCCCCTGAAAGTATTCCAGGATCTCTGAGGGAGGTTTTGGGAGGGGGCCAAGGTC
AF067781.1      CTCCGCCCCCTGAAAGTATTCCAGGATCTCCGAGGGAGGTTTTGGGAGGGGGCCAAGTTC


AF067780.1      CTCCAATTGGGTGGGTGGGGCTCCAAGCCATTGTTCTTTGCACATTCATGCCAATGTGAC
AF067779.1      CTGCAATTGGGTGGGTGGGGCTCCAAGCCATTGTTCTTTGCACATTCATGCCAATGTGAC
NR_001568.1     ------------------------------------------------------------
AF067778.1      CTGCAATTGGGTGGGTGGGGCTCCAAGCCATTGTTCTTTGCACATTCATGCCAATGTGAC
AF067781.1      CTGCAATTGGGTGGGTGGGGCTCCAAGCCATTGTTCTTTGCACATTCTTGCCAATGTGAC


AF067780.1      CTAATTGGCTAAGCACAGGCTGGAGAAGCTGGATAAACCTATTTAAATTCAGAGCCAAAA
AF067779.1      TTAATTGGCTAAGCACAGGCTGGAGAAGCTAGATAAACCTATTTAAATTCAGAGCCAAAA
NR_001568.1     ------------------------------------------------------------
AF067778.1      CTAATTGGCTAAGCACAGGCTGGAGAAGCTAGATAAACCTATTTAAATTCAGAGCCAAAA
AF067781.1      CTAATTGGCTAAGCACAGGCTGGAGAAGCTAGATAAACCTATTTAAATTCAGAGCCAAAA


AF067780.1      AACACTCAAGGATGTTTTCTGAGGGGGCGTGAGGGTTGGAAAATTATCCTATATTTGATT
AF067779.1      GACACTCAAGGATGTTTTCTGAGGGGGTGTGAGGGTTGGAAAATCATCCTATATTTGATT
NR_001568.1     ------------------------------------------------------------
AF067778.1      GACACTCAAGGATGTTTTCTGAGGGGGTGTGAGGGTTGGAAAATCATCCTATATTTGATT
AF067781.1      GACACTCAAGGATGTTTTCTGAGGGGGTGTGAGGGTTGGAAAATCATCCTATATTTAATT


AF067780.1      TATCTCCCACCTCATGTCGTGAGTCAATTCCTTAGTCTTAGATGAGGGAGAGGACTTGGG
AF067779.1      TATCTCCCACCTCATGTCCTGAGTCAATTCCTTAGTCTTAGATGAGGGAGAGGACTTGGG
NR_001568.1     ------------------------------------------------------------
AF067778.1      TATCTCCCACCTCATGTCCTGAGTCAATTCCTTAGTCTTAGATGAGGGAGAGGACTTGGG
AF067781.1      TATCTCCCACCTCATGTCCTGAGTCAATTCCTTAGTCTTAGATGAGGGAGAGGACTTGGG


AF067780.1      AGTCATCCTTGTTTTTGATGAGCTATACAACCCTATGGCCAGCAGAGGGAAGTACTGCAT
AF067779.1      AGTCATCCTTGTTTTTGATGAGCTATACAACCCTATGGCCAGCAGAGGGAAGTACTGCAT
NR_001568.1     ------------------------------------------------------------
AF067778.1      AGTCATCCTTGTTTTTGATGAGCTATACAACCCTATGGCCAGCAGAGGGAAGTACTGCAT
AF067781.1      AGTCATCCTTGTTTTTGATGAGCTATACAACCCTATGGCCAGCAGAGGGAAGTACTGCAT


AF067780.1      TTCAGAGCGACAATTTGAGATCTATGAAAGAATTTCAATCGAGAATAAGAGGCCGGGCGC
AF067779.1      TTCAGAGCGACAATTTGAGATCTATGAAAGAATTTCAATCGAAAATAAGAGGCCGGGCGC
NR_001568.1     ---------------------------------------------------GGCCGGGCGC
AF067778.1      TTCAGAGCGACAATTTGAGATCTATGAAAGAATTTCAATCGAGAATAAGAGGCCGGGCGC
AF067781.1      TTCAGAGCGACAATTTGAGATCTATGAAAGAATTTCAATCGAGAATAAGAGGCCGGGCGC
                                                                   **********

AF067780.1      GGTGGCTCACGCCTGTAATCCCAGCTCTCAGGGAGGC-AAGAGGCGGGAGGATAGCTTGA
AF067779.1      GGTGGCTCACGCCTGTAATCCCAGCTCTCAGGGAGGCTAAGAGGCGGGAGGATAGCTTGA
NR_001568.1     GGTGGCTCACGCCTGTAATCCCAGCTCTCAGGGAGGCTAAGAGGCGGGAGGATAGCTTGA
AF067778.1      GGTGGCTCACGCCTGTAATCCCAGCTCTCAGGGAGGCTAAGAGGCGGGAGGATAGCTTGA
AF067781.1      GGTGGCTCACGCCTGTAATCCCAGCTCTCAGGGAGGC-AAGAGGCGGGAGGATAGCTTGA
                *************************************** ********************

AF067780.1      GCCCAGGAGTTCGAGACCTGCCTGGGCAATATAGCGAGACCCCGTTCTCCAGAAAAAGGA
AF067779.1      GCCCTGGAGTTCGAGACCTACCCGGGCAATATAGTGAGACCCCGTTCTCCAGAAAAATGA
NR_001568.1     GCCCAGGAGTTCGAGACCTGCCTGGGCAATATAGCGAGACCCCGTTCTCCAGAAAAAGGA
AF067778.1      GCCCAGGAGTTCGAGACCTGCCTGGGCAATATAGCGAGACCCCGTTCTCCAGAAAAAGGA
AF067781.1      GCCCAGGAGTTCGAGACCTGCCTGGGCAATATAGCGAGACCCCGTTCTCCCGAAAAAGAA
                **** ************** ** ********** *************** ******   *

AF067780.1      AAAA------AAAAAAAAAGACAAAAAAAAAAATAAGCGTAACTTCCCTCAAAGCAACAA
AF067779.1      AAAAGAAA--AAAAACAAAAGACAAAAAAAAAAATAAGCGTAACTTCCCTCAAAGCAACAA
NR_001568.1     AAAAAAA-----AAACAAAAGACAAAAAAAAAAATAAGCGTAACTTCCCTCAAAGCAACAA
AF067778.1      AAAAAAAAAAAAAAAAAAAGACAAAAAAAAAAATAAGCGTAACTTCCCTCAAAGCAACAA
AF067781.1      AAAAAAAAAAAAAAAAAAAGACAAAAAAAAAAATAAGCGTAACTCCCCTCAAAGCAACAA
                ****         *** ****************************** ************

AF067780.1      CCCCCCCCC--CCCTTTTCATATTCTTGCCTGGAAGAAAGGCCTTGCTTTCCTCTCGCTTC
AF067779.1      CCCCCCCCCCCCCCTTTTTCATATTCTTGCCTGGAAGAAAGGCCTTGCTTTCCTCTCGCTTC
NR_001568.1     CCCCCC-CCCCCCTTT-------------------------------------------
AF067778.1      CCCCCC-CCCCCCTTTTTCATATTCTTGCTTGGAAGAAAGGCCTTGCTTTCCTCAGCTTC
AF067781.1      CCCCCC-CCCCCTTTTTTCATATTCTTGCCTGGAAGAAAGGCCTTGCTTTTCTCTGCTTC
                ****** **   *   **

AF067780.1      TTAAAGCTGGGA-AAGTTAAAGCCATTCTGAAAT
AF067779.1      TTAAAGCTGGGAAAAGTTAAAGCCATTCTGAAAT
NR_001568.1     ----------------------------------
AF067778.1      TTAAAGCTGGGAAAAGTTAAAGCCATTCTGAAAT
AF067781.1      TTAAAGCTGGGAAAAGTTAAAGCCATTCTGAAAT
```