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

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]. Symptoms of AD range from memory loss and poor judgement in mild cases to the inability to communicate and seizures in severe cases [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 understanding the disease in 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]. As such, this paper explores one facet of AD pathogenesis, the long non-coding RNA (lncRNA) BC200. BC200 has been implicated in AD as it upregulates the expression of b-site APP-cleaving enzyme1 (BACE1) [4], [6]. The upregulation of BACE1 in turn leads to higher levels of beta-amyloid (Aβ) in the brain, thus, disrupting cell function [4], [6].

In hopes of better understanding how AD may affect other species in the hominoidea superfamily, specifically, 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 following 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?

Brain Cytoplasmic 200 lncRNA RNA (BC200) is a 200 nucleotide long RNA transcript which is found mostly in the brain [7]. As a non-coding RNA, 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 because of its role in regulating translation and inhibiting its initiation, as well as its impacts in the pathogenesis of Alzheimer's disease and cancer [4], [7]. These **[Derek: these being BC200?]** non-coding RNAs are involved in translation control, thus, they impact the synthesis of dendritic proteins which facilitates long-term plastic changes at the synapse [8].

B. The Relation Between AD and BC200

Alzheimer's disease (AD) is a neurodegenerative disease resulting from synaptic plasticity failure in neurons [8]. It is a complex disease, meaning that it involves multiple cell types and signaling pathways [4].

AD is thought to occur due to the accumulation of two proteins in the brain. One of them is beta-amyloid (Aβ) which accumulates in neurons, forms plaques, and disrupts cell functions. The other one is hyper-phosphorylated tau protein which in abnormal levels can form neurofibrillary tangles in neurons and block synaptic transmissions [4].

Aβ, a cleavage product of the amyloid precursor protein (APP), is generated by b-site APP-cleaving enzyme1 (BACE1) and γ-secretase complex, and it strongly influences the pathogenesis of AD. Inhibition of BACE1 activity and the subsequent reduction in Aβ levels may cure or prevent AD [4], [6].

BC200 facilitates AD pathogenesis by up-regulating Aβ production through the modulation 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 overexpression [4], [6].

Many researches have demonstrated the important role of BC200 in AD. El Mus *et al.* [8] show that there are steady decline in BC200 level from age 49 to 86, but, in AD brain its level was substantially higher. They also observe that BC200 expression is increased in brain areas that are involved in AD and it is parallel with severity of disease. Huanyen Li *et al.* [6] establish an AD cell model overexpressing Aβ1-42 to observe the effects of BC200 on the cell viability and apoptosis and to investigate the associated underlying mechanisms. They observe that BC200 and BACE1 were increased upon treatment with Aβ1-42, and inhibition of BC200 rescued this Aβ1-42-mediated dysfunction, as indicated by the interaction of BC200 directly targeting BACE1. Moreover, inhibition of BC200 increased AD cell growth and reduced cells 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.

III. Materials and Methods

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 [9] search. Specifically, Megablast [10] with default parameters was used as it is able to compare closely related sequences [11]. From the Blast results, the top six sequences which are known BC200 homologs as indicated by the inclusion of BC200 in their name were chosen. 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, AC183594.3) caused errors in the analysis software described in section III-B and section III-C. Table I outlines the sequence name including the organism and the accession number of each of the chosen sequences.

TABLE I: lncRNA Sequence Information

Organism Name	Sequence Accession Number
Homo sapiens	NR_001568.1
Pongo pygmaeus	AF067778.1
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 chosen in table I were used to build the phylogenetic tree. The phylogenetic tree was built with the Mega11 [12] bioinformatics software. The parameters of the analysis were kept to their default values. Namely, the statistical method used was maximum likelihood. In addition, the test of phylogeny performed was bootstrapping with 500 replications. Finally, as the substitution model used was the Tamura-Nei substitution model [13].

C. Structural Analysis

In order to find RNA secondary structure of BC200 and three of closest homologs of it, we used FORNA, a platform for drawing the secondary RNA structure [14]. We draw each homolog structure individually. We also draw BC200. First, we analyzed BC200 structure individually to get a better understanding of its function in body. Knowing that BC200 structure consist of three main domains, we also look at its homologs from this view points. Comparing each structure to BC200, we divided them to three main domains like BC200 structure and based on each part function we analyzed the structural differences.

IV. Results

A. Phylogenetic Tree of BC200

TODO

B. Structural difference between BC200 and its homologs

Brain cytoplasmic 200 long non-coding RNA (or BC200 lncRNA) is a 200 nucleotide RNA transcript that is found predominantly in the brain. It's primary function is regulating translation by inhibiting its initiation. It's role in AD is not fully understood, but research shows that lncRNA BC200 facilitates AD pathogenesis by upregulating AB through BACE1. Pathologically, AD is characterized by an imbalance in the production and clearance of amyloid-beta in the brain leading to plaque formation [6]. The BC200 structure consist of three main parts: A-rich domain, Alu domain and unique domain [15].

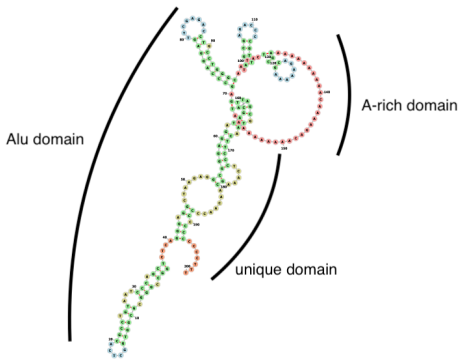

Fig. 1: BC200 RNA Secondary Structure

As we can see from the RNA secondary structures, all of the homologs have a nearly simialr A-rich domain, but on the other hand the the Alu domain is different between the homologs and BC200. The Alu domain of the mammalian

signal recognition particle (SRP) comprises the heterodimer of proteins SRP9 and SRP14 bound to the 5 and 3 terminal sequences of SRP RNA [16]. So, their function in brain should be very different from one another, even though there may be high similarity between their sequences. Figures 2, 3, and 4 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 Figure 5. 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.

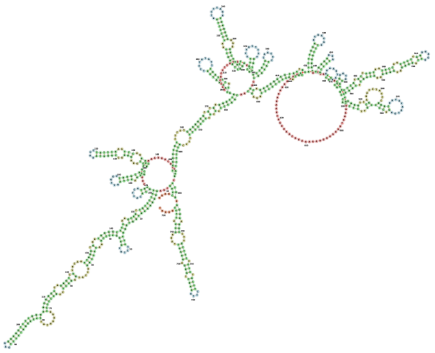

Fig. 2: BC200 RNA Secondary Structure in Gorilla

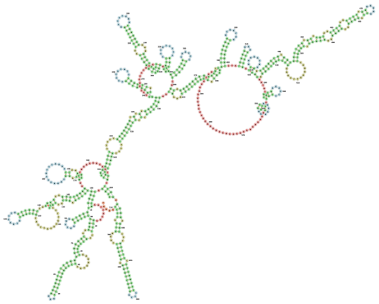

Fig. 3: BC200 RNA Secondary Structure in Pan

C. Comparing BC200 sequence with its homologs

One of the best ways for finding the conserved portions of a gene during evolution, is sequence alignment. Here, we aligned BC200 and the three most related homologs with each other to find the conserved parts of the sequence.

Here are the alignment results:

As it is obvious from the results, most parts of the sequence are conserved. Which shows that there should be a high similarity between BC200 RNA in human body and in other species. So, the following result does not support our hypothesis, and they

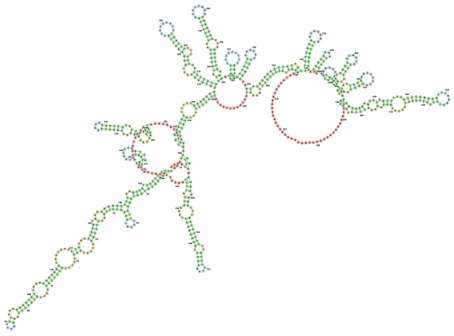

Fig. 4: BC200 RNA Secondary Structure in Pango

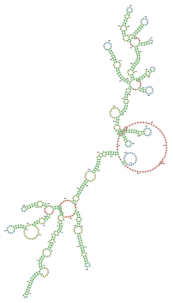

Fig. 5: BC200 RNA Secondary Structure in Hylobates lar

suggest that Alzheimers can be shared between human and great apes.

V. Discussion

TODO

```
Query: NC_000002.12:47335315-47335514 Homo sapiens chromosome 2,
GRCh38.p13 Primary Assembly Query ID: lcl|Query_43861 Length: 200


>AF067778.1 Pan paniscus BC200 alpha scRNA gene, complete sequence
Sequence ID: Query_43864 Length: 713
Range 1: 431 to 635

Score:340 bits(184), Expect:1e-97,
Identities:199/205(97%),  Gaps:5/205(2%), Strand: Plus/Plus

Query  1    GGCCGGGCGCGGTGGCTCACGCCTGTAATCCCAGCTCTCAGGGAGGCTAAGAGGCGGGAG
60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  431  GGCCGGGCGCGGTGGCTCACGCCTGTAATCCCAGCTCTCAGGGAGGCTAAGAGGCGGGAG
490

Query  61   GATAGCTTGAGCCCAGGAGTTCGAGACCTGCCTGGGCAATATAGCGAGACCCCGTTCTCC
120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  491  GATAGCTTGAGCCCAGGAGTTCGAGACCTGCCTGGGCAATATAGCGAGACCCCGTTCTCC
550

Query  121  AGaaaaaggaaaaaaaaaac-----aaaagacaaaaaaaaaaTAAGCGTAACTTCCCTC
175
            ||||||||||||||||||||     ||||||||||||||||||||||||||||||||||
Sbjct  551  AGAAAAAGGAAAAAAAAAAAAAAAAAAAAAGACAAAAAAAAAATAAGCGTAACTTCCCTC
610

Query  176  AAAGCAACAAccccccccccccTTT  200
            |||||||||||||||||||||||||
Sbjct  611  AAAGCAACAACCCCCCCCCCCCTTT  635
```

Fig. 6: BC200 Sequence Alignment with BC200 in pan

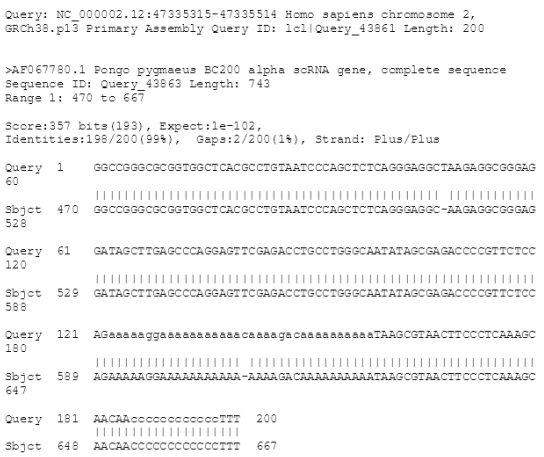

Fig. 7: BC200 Sequence Alignment with BC200 in Pango

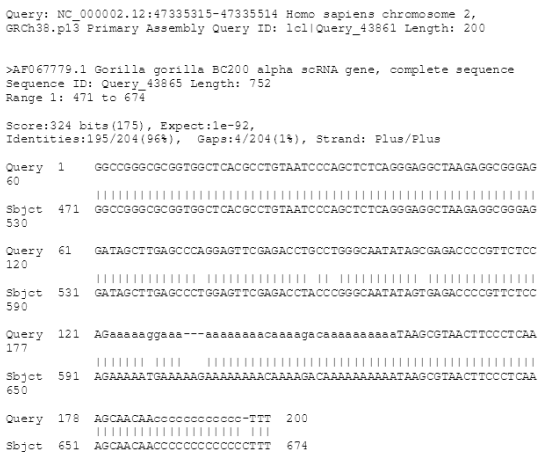

Fig. 8: BC200 Sequence Alignment with BC200 in Gorilla

VI. Conclusion

TODO

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] "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).

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] O. Weichenrieder, K. Wild, K. Strub, and S. Cusack, "Structure and assembly of the alu domain of the mammalian signal recognition particle," *Nature*, vol. 408, no. 6809, pp. 167–173, 2000.