

A Research Proposal for “Phylogenetic Analysis of lncRNAs Implicated in Alzheimer’s Disease”

Derek Robinson, Mina Emadi, Mazyar Ghezelji

University of Victoria

Victoria, Canada

{drobinson, minaemadi, mazyarghezelji}@uvic.ca

I. INTRODUCTION

There is no doubt that Alzheimer’s disease (AD) greatly affects the lives of those diagnosed and those who care for the diagnosed. In the United States, AD is currently 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]. While many humans suffer from AD, it is not well understood if other great apes (Hominidae) also suffer from AD. Finch and Austad argue that with our current understanding of AD, it is not possible to determine if AD is uniquely human [3]. As a group, we are interested in learning more about AD; specifically, the phylogeny of the long non-coding RNAs (lncRNAs) which have been implicated in AD [4]. We have chosen to work on lncRNA as a majority of the human genome is transcribed into non-coding RNA sequences.

Our main objective for this research project is to build the phylogenetic tree for one of the lncRNAs implicated in AD, as identified by Luo and Chen [4]. Our current target is brain cytoplasmic 200 (BC200) as we have already performed a preliminary BLAST search identifying several similar sequences to BC200 [5], [6]. Our secondary goal for this research project is to compare the structure of our lncRNA’s most closely related homologs to shine light on the possible differences that may lead to AD being uniquely human.

Thus, our research questions are as follows:

- RQ1. What does the phylogenetic tree of BC200 look like?
- RQ2. What structural differences exist between BC200 and its two most closely related homologs?

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 [8], [9]. This lncRNA has

recently been studied extensively because of its role in regulating translation and inhibiting its initiation, as well as its impacts in pathogenesis of Alzheimer’s disease and cancer [7], [8]. These 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 [9].

B. The Relation Between AD and BC200

Alzheimer’s disease (AD) is a Neurodegenerative disease resulting from synaptic plasticity failure in neurons [9]. It is a complex disease, meaning that it involves multiple cell types and signaling pathways [8].

AD occurs due to the accumulation of two proteins in the brain. One of them is beta-amyloid ($A\beta$) 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 [8].

$A\beta$, 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 $A\beta$ levels may cure or prevent AD [8], [10].

BC200 facilitates AD pathogenesis by up-regulating $A\beta$ 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 over-expression [8], [10].

Many researches have demonstrated the important role of BC200 in AD. El Mus *et al.* [9] 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.* [10] establish an AD cell model overexpressing $A\beta$ 1-42 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 A β 1-42, and inhibition of BC200 rescued this A β 1-42-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 A through BACE1.

III. MATERIALS AND METHODS

The methods we will be applying are those similar to Amirmahani and Goharrizi [11]. Specifically, we will be using the bioinformatics program NCBI-BLAST in order to identify homologs of our chosen lncRNA [5], [6]. Once we have identified several homologs of our chosen lncRNA, we will then perform the phylogenetic analysis using MEGA11 [12]. Our phylogenetic tree will be created via all available tree construction algorithms (Maximum Likelihood, Neighbour Joining, and Minimum-Evolution), and each of the results will be compared. Once we have built the phylogenetic tree, we will then compare the secondary structure of the three most closely related homologs using RNAz 2.0 [13].

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. Its primary function is regulating translation by inhibiting its initiation. Its role in AD is not fully understood, but research shows that Long noncoding RNA BC200 facilitates AD pathogenesis by upregulating AB through BACE1. [10] The BC200 structure consist of three main parts: A-rich domain, Alu domain and unique domain [14]. 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 [15]. So, their function in brain should be very different with eachother, eventhough there may be high similarity between their sequences.

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.

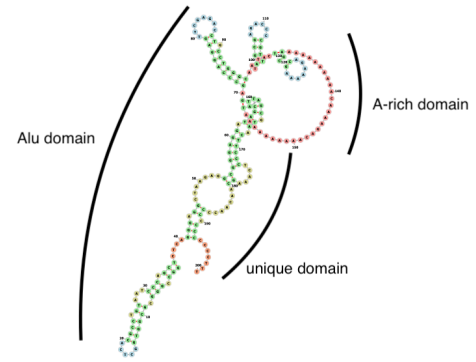


Fig. 1. BC200 RNA Secondary Structure

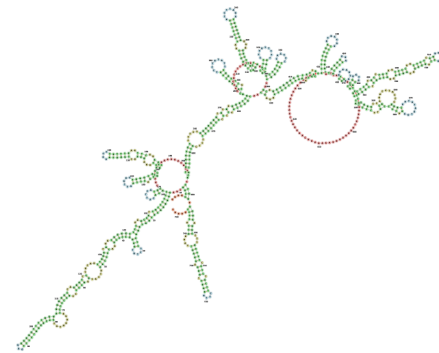


Fig. 2. BC200 RNA Secondary Structure in Gorilla

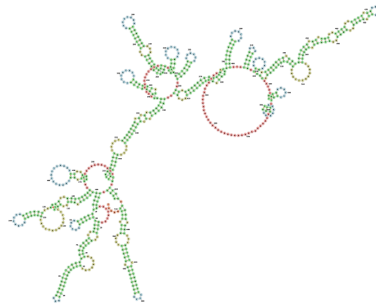


Fig. 3. BC200 RNA Secondary Structure in Pan

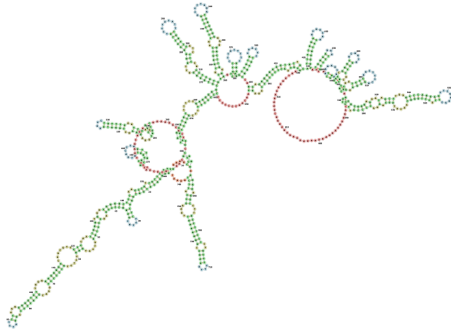


Fig. 4. BC200 RNA Secondary Structure in Pango

```

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

Query 1 99CCG99GCGCGT99CTCACGCTGTAAATCCAGCTTTCAGGAGGCTAAAGAGCGGAG
60
Sbjct 431 99CCG99GCGCGT99CTCACGCTGTAAATCCAGCTTTCAGGAGGCTAAAGAGCGGAG
490

Query 61 GATAGCTTGAGCCCAAGAGTTCAGAGCTGCTG99CAATATAGCAGACCCGTTCTTC
120
Sbjct 491 GATAGCTTGAGCCCAAGAGTTCAGAGCTGCTG99CAATATAGCAGACCCGTTCTTC
550

Query 121 AGaaaaaggaaaaaaaaaa-----aaagaacaaaaaaTAAGCTAACTTCCTC
175
Sbjct 551 AGAAAAAGGAAAAAAGAAAAAAGAAAAAGCAAAAAAATAAGCTAACTTCCTC
610

Query 176 AAAGCAACAAACcccccccccc-TTT 200
Sbjct 611 AAAGCAACAAACcccccccccc-TTT 635

```

Fig. 5. BC200 Sequence Alignment with BC200 in pan

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 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

>AF067780.1 Pongo pygmaeus BC200 alpha scRNA gene, complete sequence
Sequence ID: Query_43863 Length: 743
Range 1: 470 to 667

Score:157 bits (199), Expect:1e-102,
Identities:198/200 (99%), Gaps:2/200 (1%), Strand: Plus/Plus

Query 1 99CCG99GCGCGT99CTCACGCTGTAAATCCAGCTTTCAGGAGGCTAAAGAGCGGAG
60
Sbjct 470 99CCG99GCGCGT99CTCACGCTGTAAATCCAGCTTTCAGGAGGCTAAAGAGCGGAG
528

Query 61 GATAGCTTGAGCCCAAGAGTTCAGAGCTGCTG99CAATATAGCAGACCCGTTCTTC
120
Sbjct 529 GATAGCTTGAGCCCAAGAGTTCAGAGCTGCTG99CAATATAGCAGACCCGTTCTTC
588

Query 121 AGaaaaaggaaaaaaaaaa-----aaagaacaaaaaaTAAGCTAACTTCCTC
180
Sbjct 589 AGAAAAAGGAAAAAAGAAAAAAGAAAAAGCAAAAAAATAAGCTAACTTCCTC
647

Query 181 AAACAACcccccccccc-TTT 200
Sbjct 648 AAACAACcccccccccc-TTT 667

```

Fig. 6. BC200 Sequence Alignment with BC200 in Pango

```

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

>AF067779.1 Gorilla gorilla BC200 alpha scRNA gene, complete sequence
Sequence ID: Query_43865 Length: 752
Range 1: 471 to 674

Score:124 bits (175), Expect:1e-50,
Identities:195/204 (96%), Gaps:9/204 (4%), Strand: Plus/Plus

Query 1 99CCG99GCGCGT99CTCACGCTGTAAATCCAGCTTTCAGGAGGCTAAAGAGCGGAG
60
Sbjct 471 99CCG99GCGCGT99CTCACGCTGTAAATCCAGCTTTCAGGAGGCTAAAGAGCGGAG
530

Query 61 GATAGCTTGAGCCCAAGAGTTCAGAGCTGCTG99CAATATAGCAGACCCGTTCTTC
120
Sbjct 531 GATAGCTTGAGCCCTGAGTTCAGAGCTACCCG99CAATATAGCAGACCCGTTCTTC
590

Query 121 AGaaaaagga-----aaaaaaacaaagaacaaaaaaTAAGCTAACTTCCTC
177
Sbjct 591 AGAAAAAGGAAAAAAGAAAAAAGAAAAAGCAAAAAAATAAGCTAACTTCCTC
650

Query 178 AGCAACAAACcccccccccc-TTT 200
Sbjct 651 AGCAACAAACcccccccccc-TTT 674

```

Fig. 7. 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.>, (Accessed on 10/19/2021).
- [2] "Alzheimer's and dementia in canada," <https://www.alz.org/ca/dementia-alzheimers-canada.asp>, (Accessed on 10/19/2021).
- [3] 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.
- [4] Q. Luo and Y. Chen, "Long noncoding rnas and alzheimers disease," *Clinical interventions in aging*, vol. 11, p. 867, 2016.
- [5] T. Madden, "Chapter 16: The blast sequence analysis tool, the ncbi handbook (2002)," 2012.
- [6] "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).
- [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] 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.
- [9] 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.
- [10] 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.
- [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] A. R. Gruber, S. Findeiß, S. Washietl, I. L. Hofacker, and P. F. Stadler, "Rnaz 2.0: improved noncoding rna detection," in *Biocomputing 2010*. World Scientific, 2010, pp. 69–79.
- [14] 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.
- [15] 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.