Prioritizing missense mutations in BDNF to predict variant pathogenicity in Alzheimer’s Disease (AD)

Srihas Rao *

February 20, 2021

Abstract

The number of elder people will double from 2000 to 2050. When older, people become more susceptible to neurodegenerative disorders. One gene that affects the phenotypes of neurodegenerative diseases is Brain-Derived Neurotrophic Factor, BDNF. BDNF is a protein with 5 different isoforms in the human Chromosome number 11. For this study, only missense mutations were analyzed. By limiting the analysis, we can develop strategies to predict potential pathogenic effects. These missense mutations could be one of the risk factors for developing neurodegenerative diseases. The bioinformatics tools MARRVEL, NCBI, Clustal Omega, STRING, and TMHMM 2.0 were used to analyze the mutations. We extracted missense mutations data, related parameters through Geno2MP, and added amino acid change, conserved up to, and an amino acid change position in the domain columns. In addition to this analysis, the structural analysis of the well known pathogenic mutations of BDNF were analyzed. With this data, the most potential damaging mutation was found by prioritizing characteristics of the mutations, and we determined it is V66M.

Keywords: Alzheimer’s Disease, BDNF, missense, pathogenicity, Geno2MP, MARRVEL

*Advised by: Zeynep Öztürk
1 Purpose:

In finding the most pathogenic mutation of the known missense mutations, the goal is to raise awareness for those with specific mutations to be more careful so they can decrease their chances of developing AD or AD-like symptoms. The BDNF Val66Met polymorphism should be considered as a target for the novel Alzheimer’s disease therapeutics.

2 Hypothesis:

If the val66met polymorphism occurs in the BDNF protein, then the protein will be most likely to cause Alzheimer’s related symptoms among the mutations that we studied. We predict this because there have been previous studies that show that the V66M mutation causes reduced BDNF protein in the brain, thus causing Alzheimer’s and Alzheimer’s related symptoms.

3 Introduction

What is BDNF?

Brain-derived neurotrophic factor (BDNF), a molecule known to regulate neuronal survival and plasticity, is widely expressed in the developing and adult mammalian brain [ZYC08]. Alzheimer’s is an illness that is a dynamic, unalterable mind issue that gradually demolishes memory and thinking aptitudes and, inevitably, the capacity to complete the most straightforward errands. In the vast majority with the sickness, side effects initially show up in their mid-60s [Fac19]. The disease is multifactorial, meaning that the disease mechanism has both genetics and epigenetics contribution [BDN20]. People’s lifestyle, diet, and environment are also involved [Fac19]. Thus, while there are multiple genes that can increase the risk to develop this disorder, it does not guarantee that one will get the disorder. Due to the complexity, the disease mechanism is not fully understood, but one of the potential candidates to understand it is BDNF, or Brain Derived Neurotrophic Factor. It is a potential factor because there have been many studies [JB15] showing that the BDNF gene has caused an increased risk of older people developing some of the symptoms like those of Alzheimer’s.

Effect of BDNF on Humans

The reduction of BDNF has shown to cause problems in elderly people [ZYC08]. These problems include short term memory loss, difficulty completing familiar tasks, and confusion with time or place, all common symptoms of
Alzheimer’s Disease. Some interventions like exercise or antidepressant administration, enhance the expression of BDNF in normal and pathological conditions. Thus, if one stays active, he or she can have a reduced risk of developing Alzheimer’s [ZYC08].

The BDNF gene encodes a protein called a brain-derived neurotrophic factor, found in the brain and spinal cord, and localized in the hippocampus [BDN20]. This protein advances the endurance of nerve cells by assuming a job in the development. In the brain, the BDNF protein capacities at the associations of neural connections, where cell-to-cell correspondence happens. The neural connections can change and adjust after some time considering understanding, a trademark called synaptic plasticity [JB15].

Effects of reduced BDNF:

Changes in BDNF articulation are related to both typical and neurotic maturing and mental sickness, in structures significant for memory procedures, for example, the hippocampus and parahippocampal regions. BDNF is urgent to learning and memory since it directs long term depression (LTD) and long-term potentiation (LTP), synaptic versatility, axonal growing, multiplication of dendritic arbour, and neuronal separation [JB15]. In addition, reduced BDNF messenger RNA and protein levels have been found in the hippocampus and other cortical areas in patients with AD. Thus, mutated BDNF proteins can lose functionality so there will be lowered protein levels and many long term ailments can arise.

Isoforms in BDNF:

In Brain Derived Neurotrophic Factor, there are five isoforms, with the longest being 247 amino acids long. Isoforms occur when a gene is transcripted from the same locus but are different in their transcription start sites. The sequences that are common among the five isoforms are the more important sequences because they are used every time. When modeling, we will look at the five isoforms and how they structurally change with the most influential mutations found from the bioinformatic analysis. This will help us understand which of the mutations found from the bioinformatic analysis are showing significant structural problems resulting in AD symptoms.

Mutations in BDNF:

To predict the effect of BDNF on AD, one approach could be analyzing missense alterations in the gene. In each human genome, there are polymorphisms, slight changes in genes that result in genetic variation, making each
human genome distinct from others. An example of mutations in BDNF impacting Alzheimer’s is the Val66Met polymorphism is implicated in synaptic excitation and neuronal integrity, and has previously been shown to moderate amyloid–related memory decline and hippocampal atrophy in preclinical sporadic Alzheimer's disease. From previous studies, the val66met polymorphism has influenced memory in people from ages 20-93 [ZYC08]. In studies of brain morphometry using structural magnetic resonance imaging (MRI) scans, Val/Met individuals have repeatedly been shown to have a smaller hippocampal volume relative to controls which are homozygous for Val allele [ZYC08]. In other studies, it is shown that Met66 carriers showed greater dysfunction in cognition, glucose metabolism and tau, with implications for clinical trial design [YYL16]. Finally, Val66Met also has shown an increased risk of developing AD in women, and Caucasian women, specifically [FMN].

To see how this polymorphism and others affect the phenotype of Alzheimer’s Disease, we used the bioinformatic tools MARRVEL (Geno2MP) and NCBI (protein database) to compare different alterations of BDNF gene to prioritize more deleterious missense mutations. With this information, we can sort and filter the different characteristics of each mutation to better understand which mutations are more harmful than the others. When filtering, the more important characteristics can be prioritized.

To see how this polymorphism and others affect the phenotype of Alzheimer’s Disease, we used the bioinformatic tools MARRVEL (Geno2MP) and NCBI (protein database) to compare different alterations of BDNF gene to prioritize more deleterious missense mutations. With this information, we can analyze the various scores that tell information about the mutations, like Grantham Score, PolyPhen2 Score, and Conserv Score. Grantham score predicts the effect of the polymorphisms based on the chemical properties, like polarity and molecular size, and PolyPhen2 Score is the probability that a mutation is harmful. A score below 50 for the Grantham score is considered more harmful and a score above .80 for PolyPhen2 is considered pathogenic. Also, Conserv Score tells how conserved a mutation is.

**Predicting effect of BDNF on Alzheimer’s with modeling:**

In this study, we analyzed the structural features of the BDNF and predicted the potential pathogenic or non-pathogenic alleles reported in databases. We used bioinformatic tools, such as: TMHMM which a server to predict transmembrane domain in BDNF protein, Clustal Omega which can compare the FASTA sequences of BDNF from different species, STRING to see proteins interacting with BDNF, and Geno2MP in MARRVEL to extract and compare the missense
mutations in BDNF. Once we found what was the most harmful mutation from the bioinformatic analysis, we used the Swiss model server in order to model the proteins to see the structural changes that occurred because of the mutations. In finding the most pathogenic mutation of the known missense mutations, the goal is to raise awareness for those with specific mutations to be more careful so they can decrease their chances of developing AD or AD-like symptoms.

4 Materials and Methods:

A Computer with high speed internet access and online Kinematics tools were required.

4.1 MARRVEL/Geno2MP

Individuals with a missense mutation in the BDNF/BDNF-AS gene can be found using Geno2MP (2020). With this data, an excel sheet was made and the columns, amino acid change, hydrophilic/hydrophobic change, significance, and conservation were added. Then, with the amino acid change, a hydrophilic/hydrophobic change can be found by seeing if it changed from a hydrophobic protein to a hydrophilic protein, vice versa, or stayed the same structure/chemistry/property. If the property stayed the same, the change is not significant. If the structure did change radically, then it is significant. - Find Individuals with a missense mutation in the BDNF/BDNF-AS gene using Geno2MP (2020)

- Create an excel sheet with the columns amino acid change, hydrophilic/hydrophobic change, significance, and conservation

- With the amino acid change, a hydrophilic/hydrophobic change can be found by seeing if it changed from a hydrophobic protein to a hydrophilic protein, vice versa, or stayed the same.

- Check the structure: If there is a change then it is significant otherwise it is not significant. Check for Grantham score, gerpscore, and PolyPhen2 score.

4.2 Missense mutation positioning in NGF Domain

Proteins have domains which are amino acids generating functional regions in the protein. If there is a mutation in a functional domain, it is expected that this mutation might affect the protein function. To interpret whether the amino acid changes negatively affect the protein function we need to know the positions of these changes. We used MARRVEL/DIOPT 7.1 interphase to know the NGF domain (amino acids 212-329) of the isoform ‘NP_001137282.1 [MMB19]. The data was used from Geno2MP is for the isoform sp|P23560.1|. NGF domain was found manually in the isoform sp|P23560.1| as it is the analyzed isoform in Geno2MP. NCBI was used to have FASTA amino acid sequences of these two isoforms and the NGF region was detected after having alignment from Clustal
Omega. Amino acids were counted, and changes were checked whether they are in the NGF domain or close to the domain.

- Download FASTA amino acid sequences of the isoform sp\{P23560.1\} from NCBI

- Find NGF domain manually in the isoform sp\{P23560.1\} using data collected from Geno2MP in step 4.1

4.3 Analysing missense alleles for conservation/position

To determine the severity of a mutation, criteria such as gerpscore, grantham score, conservation of amino acids and pathogenicity prediction from Geno2MP were used. Using Table 1, the several strategies were applied. For example, for strategy 1 we applied sorting/filtering in the order gerp score biggest to smallest, pathogenicity, important change II. To apply this strategy, we first sorted the gerpcors from highest to lowest. Then, Grantham scores are filtered as higher 100. By filtering the important traits by following different strategies, the most likely damaging mutation can be found. We tried different strategies to get the most possible damaging alterations and to test our approaches.

4.4 Analyzing secondary structure changes in Polymorphisms

1. Using www.Uniprot.org, collect the data for BDNF Natural variant P23560

2. DOPE scores were found for all of the known missense mutations

3. The mutations with the least DOPE scores were T2I, V66M, Q75H, M122T, R125M, and R127L.

4. Preparing file Multiple Sequence Alignment
   - Copy the FASTA format sequence of all 5 isoforms in text document
   - The fasta format of BDNF gene isoforms paste into the large text box in SOPMA and submit to get the results.

5. Homology modeling/ tertiary structure prediction
   - Use modeller software to modelle Structures
   - BLAST was performed and 3QB5 was selected as target
   - Structures were modelled by running Python script
   - Structures with least DOPE score were selected for superimposing
   - The reference structure and modelled structure were imported to PyMol and commands were performed to superimpose the two structure
5 Results

5.1 A. BDNF Missense mutations reported in Geno2MP

There are 25 missense mutations reported. The data comes with many characteristics like gene information (Chr:Pos, Alleles) and protein information (Protein change, amino acid change, and significance) as well as pathogenicity prediction. Here we tried to improve that pathogenicity prediction and apply an approach that we can filter missense variants more specific.

<table>
<thead>
<tr>
<th>Chr Pos</th>
<th>Allele</th>
<th>Gene</th>
<th>Choromosome</th>
<th>Protein Change</th>
<th>Amino Acid Change</th>
<th>Significance</th>
<th>Conservation</th>
<th>Pathogenicity</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:50000</td>
<td>C/C</td>
<td>BDNF</td>
<td>Chr1:50000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1:51000</td>
<td>C/C</td>
<td>BDNF</td>
<td>Chr1:51000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1:52000</td>
<td>C/C</td>
<td>BDNF</td>
<td>Chr1:52000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1:53000</td>
<td>C/C</td>
<td>BDNF</td>
<td>Chr1:53000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1:54000</td>
<td>C/C</td>
<td>BDNF</td>
<td>Chr1:54000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 1: Missense mutations of BDNF and BDNF-AS reported in the Geno2MP database. 25 missense mutations were found in the Geno2MP database and analyzed for their effect on protein nature. In this table we have information about chromosome position (Chr:Pos), Alleles, gene, annotations, protein change, hydrophilic/hydrophobic change, significance, conservation in other species, NGF domain, and Pathogenicity.
5.2 B. Prioritizing the most pathogenic variants with 2 different strategies

5.2.1 Strategy 1

In this strategy, we sorted and filtered the table based on gerpscore biggest to smallest, pathogenicity, important change II.

<table>
<thead>
<tr>
<th>Allele</th>
<th>r/s</th>
<th>Gene</th>
<th>Amino Acid Change</th>
<th>Hydropathy</th>
<th>Hydrophobic Change</th>
<th>哪家</th>
<th>将前 -&gt; 最后</th>
<th>将前 -&gt; 最后</th>
<th>将前 -&gt; 最后</th>
<th>将前 -&gt; 最后</th>
<th>将前 -&gt; 最后</th>
<th>将前 -&gt; 最后</th>
<th>将前 -&gt; 最后</th>
<th>将前 -&gt; 最后</th>
</tr>
</thead>
<tbody>
<tr>
<td>C/C</td>
<td>N/C</td>
<td>BDNF</td>
<td>L&lt;sup&gt;194&lt;/sup&gt;G</td>
<td>0.77</td>
<td>1.96</td>
<td>yes</td>
<td>be</td>
<td>1.77</td>
<td>1.96</td>
<td>18</td>
<td>probably damaging</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C/C</td>
<td>N/C</td>
<td>BDNF</td>
<td>L&lt;sup&gt;286&lt;/sup&gt;C</td>
<td>0.77</td>
<td>1.96</td>
<td>yes</td>
<td>be</td>
<td>1.77</td>
<td>1.96</td>
<td>18</td>
<td>probably damaging</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C/C</td>
<td>N/C</td>
<td>BDNF</td>
<td>L&lt;sup&gt;194&lt;/sup&gt;G</td>
<td>0.77</td>
<td>1.96</td>
<td>yes</td>
<td>be</td>
<td>1.77</td>
<td>1.96</td>
<td>18</td>
<td>probably damaging</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C/C</td>
<td>N/C</td>
<td>BDNF</td>
<td>L&lt;sup&gt;286&lt;/sup&gt;C</td>
<td>0.77</td>
<td>1.96</td>
<td>yes</td>
<td>be</td>
<td>1.77</td>
<td>1.96</td>
<td>18</td>
<td>probably damaging</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A/A</td>
<td>N/A</td>
<td>BDNF</td>
<td>L&lt;sup&gt;307&lt;/sup&gt;C</td>
<td>0.77</td>
<td>1.96</td>
<td>yes</td>
<td>be</td>
<td>1.77</td>
<td>1.96</td>
<td>18</td>
<td>probably damaging</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A/A</td>
<td>N/A</td>
<td>BDNF</td>
<td>L&lt;sup&gt;307&lt;/sup&gt;C</td>
<td>0.77</td>
<td>1.96</td>
<td>yes</td>
<td>be</td>
<td>1.77</td>
<td>1.96</td>
<td>18</td>
<td>probably damaging</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5.2.2 Strategy 2

In this strategy, we sorted and filtered the table based on polyPhen2 score → pathogenicity → Granthamscore.

5.3 5 missense alleles were found in NGF domain of BDNF

As shown in Figure 1, the 2 isoforms have the same domains, but the coordinates are different. From amino acids 83-329 in the long isoform, the sequence is repeated in the short isoform. In addition, the NGF domain can be found from the amino acids 212-329. This means that 5 of the mutations are in the NGF domain.

5.4 Conservation of mutated amino acids in Isoforms

Conservation is an important feature for an amino acid. Throughout evolution it is observed that essential amino acids are conserved among organisms for the proteins which have similar functions. We checked BDNF protein and tested conservation of amino acids which carry missense mutations. The Clustal Omega comparing the eight FASTA sequences of the isoforms shows which amino acids are conserved among different isoforms. The number of isoforms
that are conserved of a specific amino acid location can be used to determine if a mutation is conserved or not. The mutation that occurs at the 66th and 120th position, V66M and A120T, is conserved in isoforms 1, 7, 4, 5, and 6. The mutations that occur at the 144th position, and I144T. The amino acids that are conserved in all species are favored over others when analyzing the data.
5.5 Secondary and tertiary structural analysis

The Root Mean Square Deviation (RMSD) of two aligned structures indicates their divergence from one another. In Pymol RMSD will be printed as RMS and the units are Angstroms. Pymol shows the structure changes in the mutated protein as cyan and the wild type protein in green. T2I shows no significant structural changes, and as a RMS score of 0.233. V66M shows a significant structural change because of how the sticks differ where Figure 4 has the protein structure highlighted. Additionally, this mutation also had a RMS score 0.364, the highest among the mutations. Therefore, based on our structural analysis, the most pathogenic missense mutation in BDNF is the Val66Met mutation.

Homology modeling/ tertiary structure prediction

6 Data Analysis:

Alzheimer’s Disease is one of the common neurodegenerative diseases and is currently the sixth leading cause of death [Fac19]. It affects mostly elderly people and makes them dependent on caregivers. There is no cure and the disease mechanism is not completely understood. BDNF is one of the potential players in the disease mechanism [Fac19]. BDNF promotes the survival and differen-
tiation of selected neuronal populations of the peripheral and central nervous systems. Therefore, a decrease in BDNF protein in the brain can result in AD-like symptoms. There are several missense mutations reported in the Geno2MP database generated by the human genome project. Missense mutations may or may not affect protein function. If there is a missense mutation generating dysfunctional protein this could increase the risk of developing AD or cause some other detrimental symptoms. Here we showed that bioinformatic analysis could help us to understand and prioritize some of these mutations for future research. In doing this research, we have found that based on our analysis, V66M was the most pathogenic mutation. This was determined by focusing on the chemical structure and nature of the protein and finding which amino acid changes were significant, conserved and in the NGF domain, and thus the effect on the secondary and tertiary structure. We tried different strategies to have the best potential prediction. With these findings, we predict that those with a genetic history of these pathogenic mutations could be in the risk of developing AD or AD-like symptoms. There are no reported cases diagnosed with AD and carrying those mutations. Our prediction is based on having a dysfunctional protein which eventually could lead to neurodegeneration and AD mechanism or similar outcomes. However, we cannot be sure without experiments which test the effect of the mutation in vivo.

Once we can define the most pathogenic mutations, we will have a better
understanding of the disease mechanism. Since the amino acid changes cause alterations in the protein’s property, the cause of Alzheimer’s disease can be found on a molecular level and this could help us understand the heterogeneity of the disease.

7 Conclusion

AD is one of the mysterious diseases that we need to solve. Finding a cure for the disease is based on the knowledge we have about disease mechanisms. Here we tried to develop a strategy based on chemical changes in BDNF protein due to missense mutations in the gene. We extracted data from MARRVEL. The chromosome position (Chr:Pos), alleles, gene, annotations, protein change, hydrophilic/hydrophobic change, significance, conservation in other species, NGF domain, and pathogenicity was found by using the tools Geno2MP/MARRVEL, Clustal Omega, String, ClinVar, and NCBI. After gathering all the data (Table 1), the most pathogenic mutations were found by sorting and filtering for the strategies we defined in the Results (3.6). Results showed that like what is stated in the hypothesis, the most pathogenic mutations was V66M based on possibly having a dysfunctional protein which eventually could lead neurodegeneration and AD mechanism or similar symptoms. However, there are other pathogenic mutations and thus we can analyze missense variants for their pathogenicity to help us understand disease mechanisms. This can help us show what mutations have an effect on Alzheimer’s Disease or AD related symptoms (6, 7). Although having this mutation cannot be changed, individuals can still not develop Alzheimer’s Disease or AD-like symptoms. This could be either due to

Figure 7: Sopma results of missense mutation V66M with sequence length 247. The figure shows RMS of 0.364 with Superimpose of the two structures in Pymol and mutated protein as cyan and the wild type protein in green.
their environment, diet, life standards, and mental activities or the complexity of the AD mechanism. We are not claiming that carrying one pathogenic variant could lead AD alone. However, it would be beneficial to consider studying potential pathogenic variants to understand the disease mechanism.

8 Acknowledgement

In the successful completion of this project, I would like to thank my family for their help and guidance throughout the project.

References


[FMN] Critical issues in bdnf val66met genetic studies of neuropsychiatric disorders.

Figure 9: Sopma results of missense mutation M122T with sequence length 247. The figure shows RMS of 0.214 with Superimpose of the two structures in Pymol and mutated protein as cyan and the wild type protein in green.


Figure 10: Sopma results of missense mutation R125M with sequence length 247. The figure shows RMS of 0.199 with Superimpose of the two structures in Pymol and mutated protein as cyan and the wild type protein in green.

Figure 11: Sopma results of missense mutation R127L with sequence length 247. The figure shows RMS of 0.219 with Superimpose of the two structures in Pymol and mutated protein as cyan and the wild type protein in green.
Figure 12: Represents RMS score and Mutations with error bars.