Predictive analysis of Tryptophan Hydroxylase 2 (TPH2) *missense* mutations in psychiatric disorders

Análise preditiva das mutações *missense* da Triptofano Hidroxilase 2 (TPH2)

DOI:10.34117/bjdv8n9-101

Recebimento dos originais: 08/08/2022
Aceitação para publicação: 08/09/2022

Gustavo Duarte Bocayuva Tavares[1]
Mestre em Ciências Biológicas
Institution: Universidade Federal do Estado do Rio de Janeiro (UNIRIO)
Address: Avenida Pasteur, 296, CEP: 22290-250, Urca, Rio de Janeiro - RJ
E-mail: gustavobocayuva@gmail.com

Gabriel Rodrigues Coutinho Pereira[1]
Mestre em Biologia Molecular e Celular
Institution: Universidade Federal do Estado do Rio de Janeiro (UNIRIO)
Address: Avenida Pasteur, 296, CEP: 22290-250, Urca, Rio de Janeiro - RJ
E-mail: gabrielkytz@hotmail.com

Gabriela Fontoura Borges
Bacharel em Biomedicina
Institution: Universidade Federal do Estado do Rio de Janeiro (UNIRIO)
Address: Avenida Pasteur, 296, CEP: 22290-250, Urca, Rio de Janeiro - RJ
E-mail: gfbfontoura@gmail.com

Joelma Freire de Mesquita
Doutorado em Bioquímica
Institution: Universidade Federal do Estado do Rio de Janeiro (UNIRIO)
Address: Avenida Pasteur, 296, CEP: 22290-250, Urca, Rio de Janeiro - RJ
E-mail: joelma.mesquita@unirio.br

ABSTRACT

Psychiatric disorders are syndromes characterized by cognitive disturbance and behavioral dysfunction, which affect over 800 million people worldwide. It is considered a major public health problem responsible for severe distress with significant impairment in social and working relationships. In the United States and Canada, psychiatric disorders are considered the main cause of disability in young individuals, in addition to being a key factor underlying suicide. Missense mutations in tryptophan hydroxylase 2 enzyme (TPH2) are associated with the development of psychiatric disorders. TPH2 catalyzes the first step of serotonin biosynthesis, a neurotransmitter that plays a central role in the regulation of emotional behavior and cognition. These mutations lead to TPH2 dysfunction with impaired enzymatic activity, which ultimately results in abnormally low levels of serotonin in the brain. Despite the importance of missense mutations in TPH2 to the development of psychiatric disorders, most of them have not yet been characterized,

[1] These authors contributed equally to this work.

so their effects are still unknown. In this study, we characterized the impact of missense mutations in TPH2 using prediction algorithms and evolutionary conservation analysis. We also used a penalty system to prioritize the most likely harmful mutations of TPH2 by combining the predictive analyses, evolutionary conservation, literature review, and alterations in physicochemical properties upon amino acid substitution. Three hundred and eighty-four missense mutations of TPH2 were compiled from the literature and databases. Our predictive analysis pointed to a high rate of deleterious and destabilizing predictions for the TPH2 mutations. These mutations mainly affect conserved and, possibly, functionally important residues. Among the uncharacterized mutations of TPH2, variants V295E, R441C T311P, Y281C, R441S, S383F, P308S, Y281H, and E363G received the highest penalties, thus, being the most likely deleterious and, consequently, important targets for future investigation. Our findings may guide the design of clinical and laboratory experiments, optimizing time and resources.

Keywords: psychiatric disorders, Tryptophan Hydroxylase 2, in silico analysis.

RESUMO

Os distúrbios psiquiátricos são síndromes caracterizadas por distúrbios cognitivos e disfunções comportamentais, que afetam mais de 800 milhões de pessoas em todo o mundo. É considerado um grande problema de saúde pública, responsável por graves problemas de saúde, com significativa deficiência nas relações sociais e de trabalho. Nos Estados Unidos e no Canadá, os distúrbios psiquiátricos são considerados a principal causa de incapacidade em indivíduos jovens, além de serem um fator-chave subjacente ao suicídio. As mutações de Missense na enzima triptofano hidroxilase 2 (TPH2) estão associadas ao desenvolvimento de distúrbios psiquiátricos. A TPH2 catalisa o primeiro passo da biossíntese da serotonina, um neurotransmissor que desempenha um papel central na regulação do comportamento emocional e da cognição. Estas mutações levam à disfunção do TPH2 com atividade enzimática prejudicada, o que resulta em níveis anormalmente baixos de serotonina no cérebro. Apesar da importância das mutações de falta de sentido no TPH2 para o desenvolvimento de distúrbios psiquiátricos, a maioria deles ainda não foi caracterizada, portanto seus efeitos ainda são desconhecidos. Neste estudo, caracterizamos o impacto das mutações de falha em TPH2 usando algoritmos de previsão e análise de conservação evolutiva. Também utilizamos um sistema de penalidades para priorizar as mutações mais provavelmente prejudiciais do TPH2, combinando as análises preditivas, conservação evolutiva, revisão de literatura e alterações nas propriedades físico-químicas sobre a substituição de aminoácidos. Foram compiladas trezentas e oitenta e quatro mutações de falta de sentido do TPH2 a partir da literatura e bancos de dados. Nossa análise preditiva apontou para uma alta taxa de previsões deletérias e desestabilizadoras para as mutações do TPH2. Estas mutações afetam principalmente resíduos conservados e, possivelmente, resíduos funcionalmente importantes. Dentre as mutações não caracterizadas do TPH2, as variantes V295E, R441C T311P, Y281C, R441S, S383F, P308S, Y281H, e E363G receberam as maiores penalidades, sendo, portanto, os alvos mais prováveis de deletérios e, conseqüentemente, importantes para investigações futuras. Nossas descobertas podem orientar o desenho de experimentos clínicos e laboratoriais, otimizando tempo e recursos.

Palavras-chave: distúrbios psiquiátricos, Tryptophan Hydroxylase 2, em análise silico.

1 INTRODUCTION

According to the DSM-5 (*Diagnostic and Statistical Manual of Mental Disorders 5th edition*), psychiatric disorders are syndromes characterized by clinical cognitive disturbance, in addition to significant behavioral and hormonal dysfunction. By definition, the following illnesses are included in the psychiatric disorders category: depression, anxiety, addiction, attention deficit hyperactivity disorder (ADHD), bipolar affective disorder (BPAD), eating disorders, and schizophrenia (American Psychiatric Association, 2013). Although there is no consensus on the cause of psychiatric disorders, most studies suggest that they are polygenic and multifactorial brain syndromes caused by neurotransmitter deregulation, genetic abnormalities, and defects in brain structures (OLIVEIRA *et al.*, 2019).

Psychiatric disorders are considered a major public health problem responsible for severe distress for the affected ones and their families, in addition to significant impairment in social and work relationships (TRAUTMANN; REHM; WITTCHEN, 2016). In 2016, 792 million people were affected by any type of mental disorder, mainly because of depression or anxiety disturbances (MUKHERJEE, 2016). The results of the National Comorbidity Survey Replication (USA) suggested that approximately half of the population will experience at least a mental disorder throughout life (KESSLER *et al.*, 2005). In the United States and Canada, psychiatric disorders are considered the main cause of disability in individuals from 15 to 44 years of age, corresponding to 3.5% of all cases (SANSONE; SANSONE, 2010). The cost of lost productivity in the workplace only due to depression and anxiety is estimated at one trillion dollars per year (WHO - Mental Health, 2017). The morbidity rate is also drastically increased in patients affected by psychiatric disorders, corresponding to two and a half times higher than that observed in the general population (STANLEY *et al.*, 2019). This rate is even higher for suicide, in which 90% of all deaths result from underlying mental health disorders, killing over 800.000 people every year (WHO - Mental Health, 2017).

Serotonin, or 5-hydroxytryptamine (5-HT), is a tryptophan-derived neurotransmitter that plays key roles in the regulation of several physiological processes in vertebrates, including appetite, homeostasis, sleep, emotional behavior, and cognition (ŠTRAC; PIVAC; MÜCK-ŠELER, 2016). Serotonin is mainly produced by neurons from the raphe nucleus of the midbrain. Serotonin biosynthesis occurs in two steps: i) the first and also the rate-limiting step consists of the hydroxylation of L-tryptophan, which is catalyzed by tryptophan hydroxylases (TPH) enzymes; ii) the second step consists of the

decarboxylation of 5-hydroxy-L-tryptophan by aromatic L-amino acid decarboxylase (AADC). Two isoforms of TPH enzymes are currently known: TPH1, almost exclusively expressed in peripheral tissues, and TPH2, mainly expressed in the central nervous system (WELFORD *et al.*, 2016).

Missense mutations in *TPH2* can lead to severe alterations in protein's stability and solubility (ZHANG; WANG, 2021), which results in inactive and dysfunctional forms of the enzyme, impairing serotonin synthesis and leading to abnormally low levels of this neurotransmitter in the brain. These mutations have been associated with the development of psychiatric disorders, including depression, BPAD, and ADHD (UNIPROT CONSORTIUM, 2021).

Millions of genetic variants have been discovered due to improvements in next-generation sequencing methods. Nonetheless, analyzing the effect of mutations by traditional wet-lab methods is an arduous and expensive task, so most of them have not yet been characterized (DE BAETS *et al.*, 2012). Bioinformatics, on the other hand, proved to be efficient in predicting the effect of genetic variants using only information obtained from the amino acid sequence. This approach has relatively high accuracy, *i.e.* around 80% (LÓPEZ-FERRANDO *et al.*, 2017), being successful in the study of several genetic and metabolic disorders. The *in silico* characterization of mutations saves time and resources thanks to its ability to screen for the most likely deleterious mutations so that they can be prioritized in future *in vitro* and *in vivo* studies (PEREIRA *et al.*, 2020).

Despite the relevance of TPH2 mutations in psychiatric disorders, the majority of them have not yet had their effects characterized (UNIPROT CONSORTIUM, 2021). Thus, this study aims to characterize these variants *in silico* following the methodology previously established by our group (PEREIRA *et al.*, 2020; PEREIRA *et al.* 2022; PEREIRA; ABRAHIM-VIEIRA; DE MESQUITA, 2021), which could guide the design of future experiments and provide relevant information in the field precision medicine (ROY CHOUDHURY *et al.*, 2017).

2 MATERIALS AND METHODS

2.1 DATA RETRIEVAL

The sequence of wild-type TPH2 was obtained from the UniProt database, which is stored under the identification code Q8IWU. Information on functional domains of TPH2 and relevant protein residues were obtained from the UniProt (NCBI, 2017) and a literature review on PubMed.

Missense mutations of TPH2 protein were compiled from the databases OMIM, dbSNP, and UniProt, in addition to a literature review (NCBI, 2017). Information on the effects of already characterized mutations was obtained from the same data sources.

The structure of wild-type TPH2 was obtained from the Protein Data Bank (PDB), which is stored under the identification code 4V06 (ROSE *et al.*, 2021).

2.2 PREDICTIVE ANALYSIS

The algorithms SNAP2, SNP&GO, PolyPhen-2, PMUT, PhD-SNP, MutPred2, Provean, SIFT, Panther, and Predict-SNP (PEREIRA *et al.*, 2022) were used to predict whether the mutation is neutral or deleterious.

The algorithm I-Mutant 3.0, FoldX (PEREIRA; ABRAHIM-VIEIRA; DE MESQUITA, 2021), and INPS were used to predict whether the mutations reduce, increase, or do not affect protein's stability. A consensus approach was used for the stability prediction analysis so that the final prediction was considered when more than half of all algorithms converge to the same response.

At last, the algorithms TANGO, WALTZ, and LIMBO of SNPEffect4.0 were used to predict whether the mutations increase, decrease, or do not affect the protein's aggregation, amyloid-fiber formation, or chaperone binding, respectively (DE BAETS *et al.*, 2012).

The predictive algorithms used in this study are based on different machine learning methods aiming at classifying the effects of missense mutations. These methods initially identify rules and patterns from large data sets (training), from which they build mathematical models able to predict new observations (PAIXÃO *et al.*, 2022) with relatively high precision, *i.e.* around 80% (LÓPEZ-FERRANDO *et al.*, 2017).

2.3 EVOLUTIONARY CONSERVATION ANALYSIS

The crystallographic structure of wild-type TPH2, *i.e.* 4V06, was submitted to ConSurf, which computed the evolutionary conservation degree of each amino acid of TPH2. ConSurf performs multiple sequence alignments to construct a phylogenetic tree, which is used to estimate the evolutionary conservation score of protein residues (ASHKENAZY *et al.*, 2016). The evolutionary conservation degree of each amino acid was then also projected on the protein's surface and colored according to ConSurf color-coding scale, which varies from 1 and cyan, representing variable positions to 9 and maroon, representing conserved positions.

The following parameters were selected for the ConSurf analysis: PDB ID:4V06; Chain identifier: A; homologous search algorithm: PSI-BLAST; number of iterations: 3; E-value cut-off: 0.0001; protein database: UniProt; reference sequence: closest; number of reference sequences selected: 150; maximum sequence identity: 95%; minimum identity for counterparts: 35%; alignment method: MAFFT-L-INS-i; calculation method: Bayesian; evolutionary substitution model: best model.

2.4 PENALTY ANALYSIS

We combined the predictive analysis with the evolutionary conservation degree, a literature review, and alterations in physicochemical properties to rank all uncharacterized mutations in TPH2 using a penalty system. The penalty system ranked the mutations according to their deleterious propensity so that mutations receiving the highest penalties were ranked first. Alterations in physicochemical properties upon amino acid substitution considered the following parameters: hydrophobicity, size, and charge, which were defined according to the classification described by Jean-Baptiste, Berthelot & Favre (2016). The penalties were attributed to alterations considered damaging for TPH2 in the literature consulted.

3 RESULTS AND DISCUSSION

3.1 DATA RETRIEVAL

The TPH2 sequence obtained from the UniProt has 490 amino acids, which corresponds to the complete protein length. As shown in Fig 1, TPH2 can be divided into three functional domains: regulatory domain (residues 1 – 150), catalytic domain (151 – 450), and tetramerization or oligomerization domain (residues 451 – 490) (CARKACI-SALLI *et al.*, 2006).

The regulatory domain contains two relevant regions: an initial flexible region, also known as the N-terminal tail (residues 1 – 46), and an ACT subdomain (residues 65 – 140). The first 46 residues of TPH2 are a highly mobile and relatively unstable region, which is believed to limit TPH2 purification and hamper protein crystallization. Truncation of the flexible region has been shown to impair TPH2 tetramerization and dimerization, suggesting that this region might also be involved in the protein's oligomerization (TIDEMAND *et al.*, 2016). The ACT subdomain receives this name from the first three enzymes in which this motif was recognized: aspartate kinase, chorismate mutase, and TyrA. ACT subdomain is a small structural motif associated with allosteric

regulatory function upon ligand binding, which is commonly observed in enzymes related to amino acid metabolism, including TPH2. The binding of L-Phe to the regulatory domain of phenylalanine hydroxylase (PAH), an enzyme that shares structural and functional similarities with TPH2, has been shown to stabilize the protein and activate PAH by inducing a conformational shift towards the exposure of its active site. A similar mechanism is believed to occur in TPH2 upon L-Phe or L-Trp binding (SKAWINSKA, 2020). Most of the regulatory domain has not yet been experimentally determined (ROSE *et al.*, 2021), as shown in Fig 1B, possibly due to the high mobility and instability (TIDEMAND *et al.*, 2016).

Two important phosphorylation sites are present in human TPH2 protein, all of them located within the regulatory domain of TPH2. Ser19 and Ser104 are phosphorylated by cyclic AMP-dependent protein kinase A (PKA), which acts as a post-translational mechanism that regulates TPH2 activity. Ser19 can be additionally phosphorylated by calmodulin-dependent protein kinase II (CaMKII) (CARKACI-SALLI *et al.*, 2014). TPH2 phosphorylation is responsible to increase enzymatic activity by up to 30%, possibly through a similar allosteric activation mechanism to that previously described by PAH upon L-Phe binding (SKAWINSKA, 2020).

The catalytic domain is the most evolutionarily conserved among the three TPH2 domains, being responsible for its enzymatic activity (PEREIRA *et al.*, 2020). TPH2 uses the Fe^{2+} co-factor and the co-substrates pterin and O_2 to hydroxylate the L-Trp substrate and generate 5-hydroxytryptophan (5HTP), which constitutes the rate-limiting reaction in the biosynthesis of serotonin neurotransmitters (TIDEMAND *et al.*, 2016). As shown in Fig 2, human TPH2 interacts with the L-Trp substrate through residues Tyr281, Arg303, Thr311, Pro314, Ser336, Phe359, Phe364, and Ile412. Among them, Tyr281, Phe359, and Ile412 are key residues to determine substrate specificity. Tyr281 is directly required for the correct placement of L-Trp on the active site of TPH2. The catalytic domain also contains all the TPH2 residues involved in pterin co-substrate binding, which are Gly280, Tyr281, Leu282, and Phe307 (SKAWINSKA, 2020). Fe^{2+} is the enzymatic co-factor of TPH2, thus being crucial for its catalytic mechanism. Residues His318, His323, and Glu363 are known to bind to Fe^{2+} (UNIPROT CONSORTIUM, 2021).

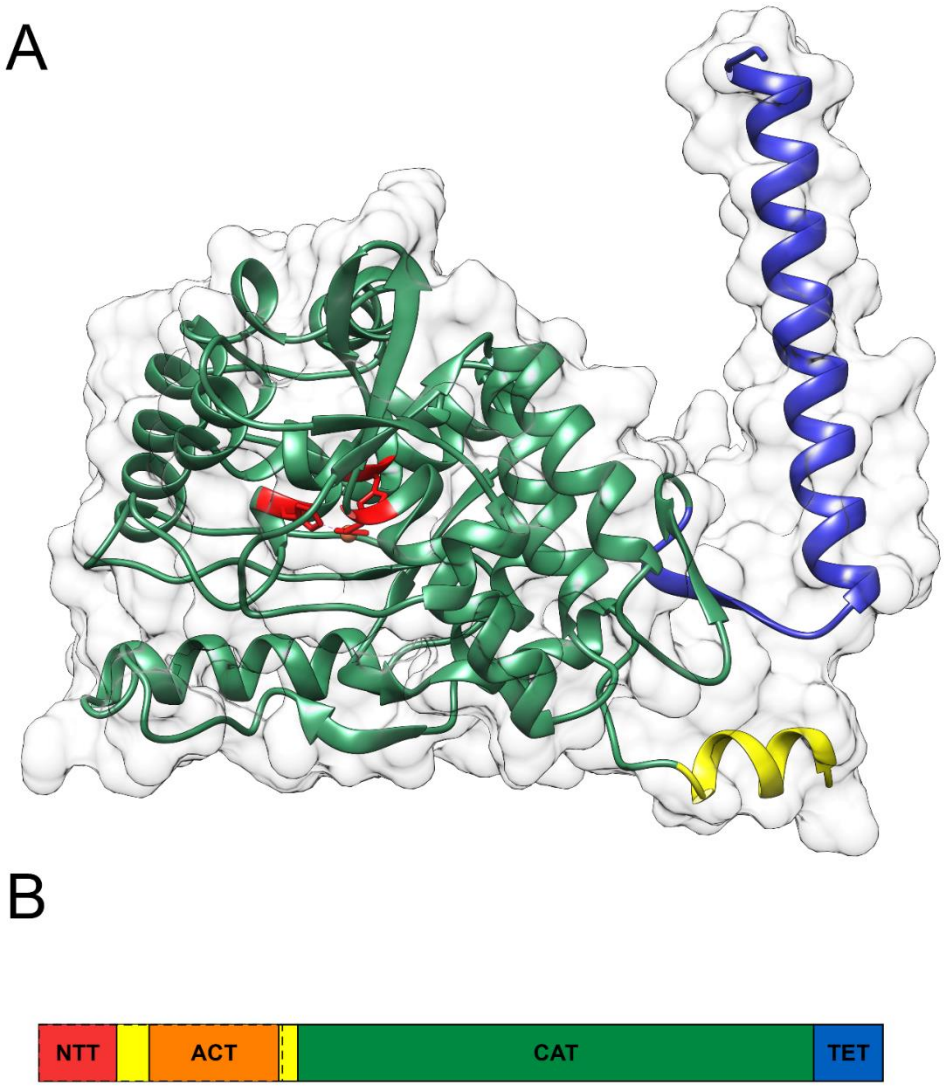
The oligomerization domain is composed of an alpha helix located at the C-terminus of TPH2. This helical structure is crucial for TPH2 tetramerization (PEREIRA *et al.*, 2020), which occurs through the formation of a coiled-coil motif. The coiled-coil motif formed consists of 4 amphipathic α-helices interacting through their hydrophobic

interfaces, which is responsible for maintaining the quaternary structure of TPH2 (SKAWINSKA, 2020). Truncation of this domain resulted in disrupted tetrameric assembly, reaffirming its role in TPH2 oligomerization (TIDEMAND *et al.*, 2016).

Three hundred and eighty-four missense mutations of TPH2 were compiled from the literature and databases consulted (S1_Table). Among them, only 12 mutations have already been characterized. Thus, the effects of most TPH2 mutations, i.e. 372, are still unknown. The mutations P206S, R303W, and R441H are known to be pathogenic, being previously associated with the development of BPAD, ADHD, and depressive disorder, respectively (UNIPROT CONSORTIUM, 2021). The study by CICHON *et a*l. (2008), which analyzed 2183 patients, suggested that the mutation P206S in TPH2 increases the risk of developing BPAD by 4.8 times (CICHON *et al.*, 2008). The study by McKinney *et al.* (2008) identified complete loss-of-function and deficient serotonin levels upon R303W mutations which is believed to make the mutation carriers susceptible to developing ADHD (MCKINNEY *et al.*, 2008). MCKINNEY *et al.* (2009) also studied the impact of nine missense mutations in TPH2 protein. The study reaffirms the complete loss of function for R303W and also indicated a severe loss of function for variant R441H, which has already been associated with the development of major depression (MCKINNEY *et al.*, 2009).

The mutations L36P, L36R, R88C, P138L, R471C, D473N, and N482D are known to impair serotonin synthesis (NCBI, 2017). Reduced serotonin levels in the brain have already been related to the development of psychiatric disorders, especially depressive disorders (YANG *et al.*, 2019). Therefore, these mutations may be potential targets for future investigation. At last, mutations L36V and S41Y are believed to be benign to TPH2. According to ClinVar, an extensive predisposition screen test performed in a healthy population together with literature evidence was sufficient to determine that the variants L36V and S41Y are unlikely to cause disease, thus being classified as likely benign (LANDRUM *et al.*, 2018). Furthermore, variant S41Y was previously reported to increase the enzymatic activity of TPH2, but no increase in serotonin levels was observed for this variant (CARKACI-SALLI *et al.*, 2014).

Fig 1. Crystalographic structure and schematic representation of human TPH2 protein. The regulatory domain is colored yellow, the catalytic domain (CAT) is colored green, and the tetramerization domain is colored blue (TET). (A) The structure of human TPH2 available at the Protein Data Bank (4V06) is shown as a three-dimensional representation displaying the secondary structure elements, where helices represent α-helices, arrows represent β-sheets, and thin lines represent coils. Iron binding residues are colored red. (B) Schematic representation of TPH2 sequence displaying the functional domains and relevant residues. The regulatory subdomains N-terminal tail (NTT) and ACT are highlighted in red and orange, respectively. The dashed region represents the absence of three-dimensional information for TPH2.

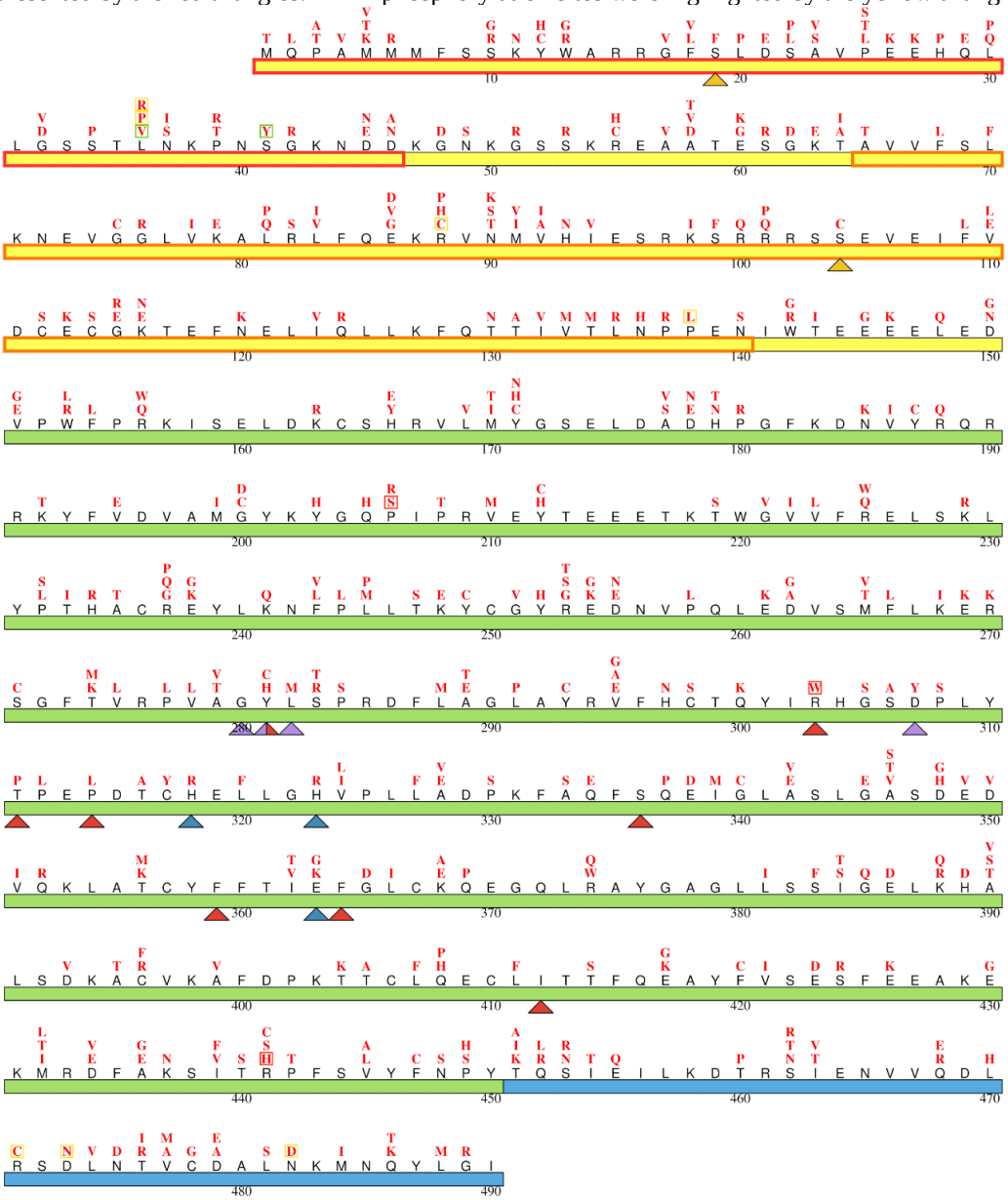


As shown in Fig 2, all mutations were displayed together with the functional domains and relevant residues of TPH2 for further comparison. The 384 mutations compiled are distributed all over the protein, similarly affecting the three functional domains of TPH2. Two TPH2 mutations affect known phosphorylation sites of the protein: S19F and S104C. Among the 20 essential amino acids, nine of them are currently known to be phosphorylated in human proteins: Ser, Thr, Tyr, His, Lys, Arg, Asp, Glu, and Cys. Approximately thirty percent of all phosphorylation occurs in Ser, Thr, and Tyr,

which also results in a relatively more stable form of this post-translational modification (ARDITO *et al.*, 2017; MAKWANA; MUIMO; JACKSON, 2018). S19F results in the substitution of a phosphorylable residue, Ser, for a non-phosphorylable one, *i.e*. Phe. Thus this mutation possibly impairs S19F phosphorylation, which is a key post-translational modification to modulate TPH2 activity that can increase enzymatic activity by up to 30% (SKAWINSKA, 2020). On the other hand, variant S104C results in the substitution of Ser to Cys, which is also phosphorylable. Nonetheless, Ser is known to be considerably more stable than Cys.

The mutations H318R, H323R, E363G, and E363K affect iron-binding residues, which is the cofactor required for the catalytic activity of TPH2 (UNIPROT CONSORTIUM, 2021). The study by Cao *et al.* (2017) analyzed the conservation of iron binding residues within protein sequences, which suggested that His, Glu, and Asp are by far the more frequent amino acids binding to iron. Cys, Tyr, and Asn are relatively less frequent than His, Glu, and Asp, but they present moderate to low frequency as iron-binding residues. The other 14 essential amino acids, in turn, are rarely bound to Fe^{2+} and Fe^{3+} ions (CAO *et al.*, 2017). The mutations H318R, H323R, E363G, and E363K result in the substitution of His and Glu, which are highly conserved amino acids binding to Fe^{2+} and Fe^{3+} ions for Arg, Gly, and Lys, which are rarely bound to iron. The substitutions of His to Arg upon H318R and H323R mutations result in small differences in amino acid properties, not affecting the amino acid charge. On the other hand, the substitutions of His to Gly upon E363G significantly affect amino acid properties, resulting in the change of a negatively charged and hydrophilic amino acid into a relatively small, hydrophobic, and non-charged amino acid. The substitutions of His to Lys upon E363K also result in substantial alteration in amino acid properties since His, which is positively charged is altered to a negatively charged amino acid. Thus, particularly E363G and E363K may hamper iron binding with consequent implications for TPH2 catalytic activity.

Fig 2. Schematic representation of TPH2 protein missense mutations, functional domains, and relevant residues. The regulatory domain is colored yellow, the catalytic domain is colored green, and the tetramerization domain is colored blue. The regulatory subdomains N-terminal tail (NTT) and ACT are highlighted by red and orange rectangles, respectively. Iron binding residues are represented by the blue triangles, pterin residues are represented by purple triangles, and substrate binding residues are represented by the red triangles. TPH2 phosphorylation sites were highlighted by the yellow triangles.



Variants Y281H, Y281C, L282M, and D307Y occur in TPH2 residues involved in pterin co-substrate binding. Interestingly, position 281 is also known to interact with the enzymatic substrate of TPH2, *i.e.* L-Trp. Tyr281 is a key residue to determine substrate specificity and is also required for the correct placement of L-Trp on the TPH2 activity site (SKAWINSKA, 2020). Thus, mutations Y281H and Y281C may affect TPH2's ability to bind to pterin and L-Trp, which are central interactions for TPH2 enzymatic activity, requiring further attention. Especially Y281C, which causes important

alterations in the physicochemical properties of Tyr, changing a hydrophobic and non-charged residue into a hydrophilic and positively charged amino acid.

In addition to Y281H and Y281C, previously discussed, the following mutations affect TPH2 residues involved in L-Trp binding: R303W, R303Q, T311P, and P313L. Among them, variant R303W is already known to be deleterious and associated with the development of ADHD. R303W induces complete loss of function in TPH2 with consequent deficient serotonin levels in the brain (MCKINNEY *et al.*, 2009). Interestingly, R303Q occurs at the same position as R303W and causes alterations in amino acid charge, thus, being an important target for future investigation.

3.2 PREDICTIVE ANALYSIS

Ten algorithms were used to predict the functional effects of TPH2 missense mutations, including SNAP2, SNPs&GO, PolyPhen-2, PMUT, PhD-SNP, MutPred2, Provean, SIFT, Panther, and Predict-SNP. As described in Table 1, these algorithms use particular training sets, predictor variables, and machine learning methods to classify whether the mutation is neutral or deleterious for the protein's function.

Overall, the algorithms presented high accuracy in detecting the deleterious or neutral potential of the five already classified mutations in TPH2. PANTHER, SIFT, Provean, PolyPhen-2, PHD-SNP, SNPs&GO, and PMUT were the algorithms with the highest accuracy rate, correctly classifying L36V, S41Y, P206S, R303W, and R441H mutations. On the other hand, MutPred2 and SNAP2 were the algorithms with the lowest accuracy rate, correctly classifying only 60% of these mutations. PredictSNP, in turn, presented an 80% accuracy rate.

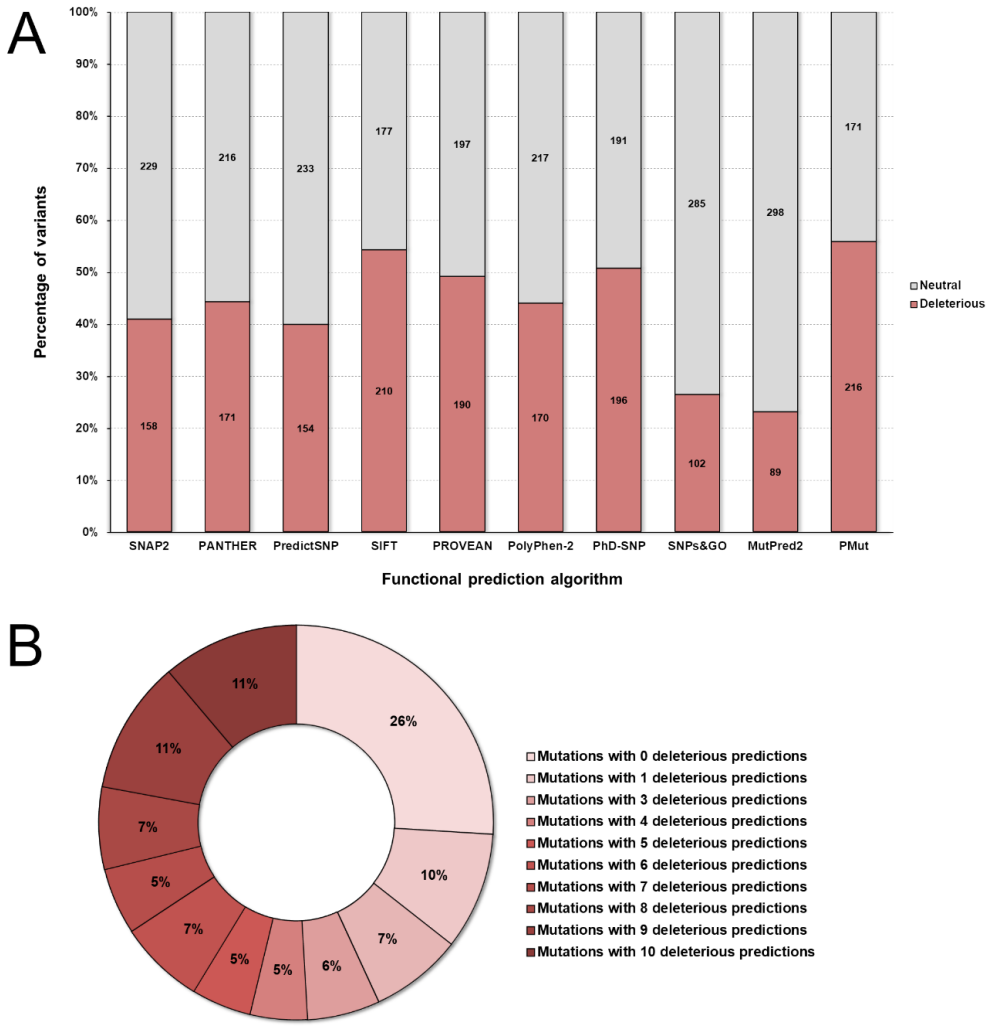
Table 1. Description of the functional prediction algorithms used.

Algorithm	**Training set**	**Predictor Variables**	**Learning method**	**Reference**
SNAP2	dataset derived from Protein Mutant Database	evolutionary conservation, solvent accessibility, secondary structure, flexibility, and information on functional domains	Neural networks	(BROMBERG; YACHDAV; ROST, 2008)
SNPs&GO	dataset derived from SwissVar	evolutionary conservation, secondary structure, and functional domains	Support Vector Machines	(CAPRIOTTI *et al.*, 2013)
PolyPhen-2	dataset derived from HumDiv and Humvar	structural information, evolutionary conservation, electrostatic interactions, hydrophobicity, and binding sites	Naïve Bayes	(ADZHUBEI *et al.*, 2010)

PMUT	dataset derived from the SwissProt database	physicochemical properties, protein interactome, and evolutionary conservation	Neural networks	(LÓPEZ-FERRANDO *et al.*, 2017)
PhD-SNP	dataset derived from the SwissProt database	evolutionary conservation	Support Vector Machines	(CAPRIOTTI; FARISELLI, 2017)
MutPred2	dataset derived from HGMD, SwissProt, and dbSNP	solvent accessible surface, secondary structure, metal binding, and post-translational modifications	Neural networks	(PEJAVER *et al.*, 2017)
Provean	dataset derived from UniProtKB and SwissProt	evolutionary conservation, physicochemical properties	Clustering	(CHOI; CHAN, 2015)
SIFT	dataset derived from dbSNP	evolutionary conservation, physicochemical properties	Naïve Bayes	(VASER *et al.*, 2015)
Predict-SNP	dataset derived from the UniProtKB and Protein Mutant Database	prediction scores of eight different algorithms (MAPP, nsSNPAnalyzer, PANTHER PhD-SNP, PolyPhen-1, PolyPhen-2, SIFT, and SNAP)	Consensus classifier	(BENDL *et al.*, 2014)
Panther	proprietary database of human mutations	evolutionary conservation and functional annotations	Hidden Markov Model	(MI *et al.*, 2017)

Individually, the functional prediction algorithms presented a moderate rate of deleterious predictions. PMut presented the highest rate of deleterious predictions, classifying 56% of the analyzed mutations as deleterious, while MutPred2 presented the lowest rate, with a 23% of deleterious rate (Fig 3A). Thus, it points to divergent predictions for TPH2 variants that can be attributed to unique aspects of the algorithm, including the training data, the set of predictor variables, and the machine learning method used, as described in Table 1. Since no currently gold standard method is currently available, it is indispensable to combine multiple algorithms to predict the functional effects of missense mutations (PEREIRA; ABRAHIM-VIEIRA; DE MESQUITA, 2021).

Fig 3. Functional prediction analysis of TPH2 missense mutations. (A) The number of deleterious (maroon), and neutral (gray) predictions per algorithm for the TPH2 missense mutations is displayed in a bar plot. (B) The number of mutations predicted as deleterious from zero (light maroon) to ten functional prediction algorithms (dark maroon) is displayed in a donut plot with the corresponding percentage values.



The state of art methods has individual accuracies ranging from 60% to 81% (LÓPEZ-FERRANDO *et al.*, 2017; PEJAVER *et al.*, 2017), but as previously shown by BENDL *et al.* (2014), combining multiple algorithms can significantly increase the overall predictive performance (BENDL *et al.*, 2014). By combining multiple algorithms in a consensus approach, *i.e.* considering the response of half plus one algorithm (≥ 6), our findings suggest that 41% of the analyzed TPH2 mutations were classified as deleterious (Fig 3B). Detailed information on the functional prediction analysis is shown in S2_Table.

As shown in Fig 3B, eleven percent of the TPH2 mutations were predicted as deleterious by the ten prediction algorithms used, including L81Q, L81P, V110E, W153L, W153R, F154L, Y171C, Y171N, H179T, Y187C, P206R, Y212H, Y212C,

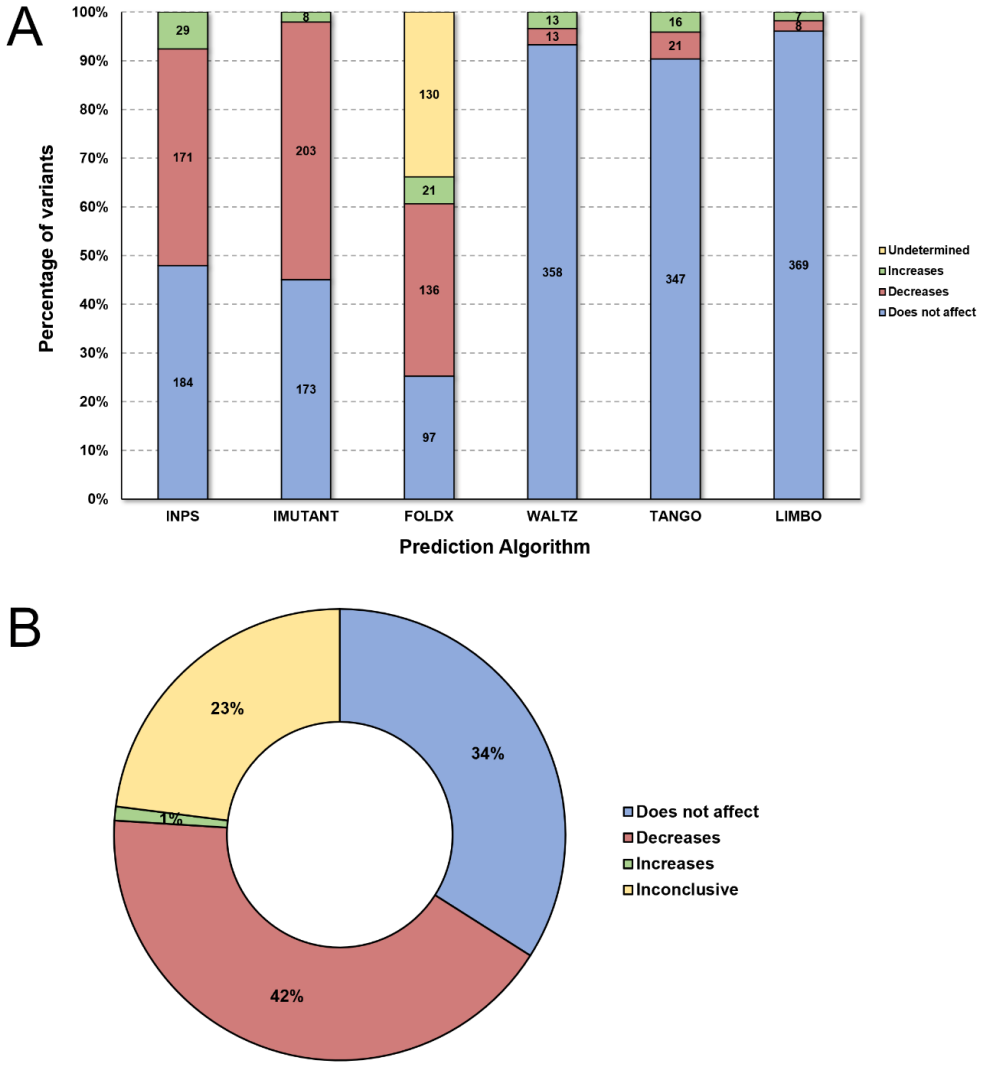
E238G, L245P, P258L, P277L, V278L, A279V, Y281H, Y281C, L291P, Y293C, V295E, V295G, Q300K, R303W, P308S, T311P, P314L, C317Y, H318R, Q337P, G430C, E363K, E363G, G365D, Q369P, S383F, Q408P, S424R, A436E, R441S, R441C, F447C, P449H, T451I, and L481S. Despite the diverse approaches, all methods converge on the same response, therefore, suggesting that these mutations may be potentially damaging to TPH2. Most of them (over 90%) occur in the catalytic domain, which is crucial for TPH2 enzymatic activity and, consequently, its function (SKAWINSKA, 2020). Interestingly, variants P206S, R303W, and R441H, which have already been characterized as pathogenic, also occur in the catalytic domain (NCBI, 2017).

We also performed stability predictions using I-Mutant3.0, INPS, and FoldX algorithms to further characterize the effects of TPH2 missense mutations, which are shown in Fig 4. FoldX did not return predictions for the mutations occurring between positions 1 and 148 of TPH2, given that this algorithm requires structural information obtained from the PDB (DE BAETS *et al.*, 2012), but no complete three-dimensional structure has been experimentally determined for TPH2 (ROSE *et al.*, 2021). Considering the individual prediction of the algorithms, most of the analyzed mutations reduce TPH2 stability. FoldX and I-Mutant were the algorithms with the highest number of destabilizing mutations, accounting for approximately 53% of their valid predictions, while INPS was the algorithm with the lowest number of predictions from this class, corresponding to 44% of the TPH2 mutations. Only a small fraction of variants were predicted to increase TPH2 stability, ranging from 2% in I-Mutant, to approximately 8% in FoldX and INPS.

As shown in S3_Table, distinct results were observed in the stability prediction analysis for the TPH2 variants, which can be attributed to the particular methods each algorithm used. I-Mutant3.0 is an algorithm trained on a dataset derived from the ProTerm, which contains thermodynamic information of experimentally determined protein structures. I-Mutant3.0 is based on Support Vector Machines (CAPRIOTTI; FARISELLI; CASADIO, 2005). FoldX, on the other hand, uses an empirical force field trained in a database of engineered proteins to estimate free-energy alterations upon amino acid substitution (DE BAETS *et al.*, 2012). At last, INPS is based on a support vector machine regressor with a radial basis function kernel (RBF) trained using a database derived from ProTherm. INPS extracts features from protein structure and sequence, including molecular weight, solvent accessibility, evolutionary information,

and hydrophobicity, which are used to predict the free-energy changes upon mutation (SAVOJARDO *et al.*, 2016).

Fig 4. Stability and SNPEffect4.0 analyses of TPH2 mutations. (A) The bar plot shows the number of stability predictions by INPS, IMutant, and FoldX, in addition to the number of predictions by WALTZ (amyloid propensity), TANGO (protein aggregation), and LIMBO (chaperone binding). Blue bars represent "do not affect" predictions, red bars represent "decrease" predictions", and green bars represent "increase" predictions. The yellow bars represent undetermined predictions. (B) The pie chart shows the consensus stability prediction analysis. The blue sector represents mutations predicted as not affecting stability, the red sector represents mutations predicted as destabilizing, the green sector represents mutations predicted as stabilizing, and the yellow sector represents inconclusive classifications.



The evaluations of FoldX, INPS, and I-Mutant3.0 were combined into a consensus approach, *i.e.* considering the response of half plus one algorithm (≥ 2). As shown in Fig 4B, the consensus analysis suggested that most TPH2 mutations are destabilizing (42%), which was followed by mutations that do not affect protein stability (34%). Interestingly, only 1% of all mutations analyzed were predicted to increase protein stability, which suggests that this phenotype is relatively rare among the TPH2 mutations. Twenty-three

percent of all mutations received inconclusive results in the consensus analysis, *i.e.* no prevalent classification, thus, not considered in our analysis.

Missense mutations can impact protein function by increasing or decreasing its stability. Changes in structural stability can prevent conformational changes necessary for protein function, constituting a frequent mechanism underlying numerous disease-related mutations (SANAVIA *et al.*, 2020). Deleterious mutations in TPH2, *i.e.* P206S, R303W, and R441H, are known to reduce protein stability and solubility, which ultimately leads to loss of enzymatic activity and diminished serotonin levels (UNIPROT CONSORTIUM, 2021). Thus, mutations predicted to alter TPH2 stability, particularly those classified as destabilizing, may lead to harmful effects on protein function.

Among the 384 variants analyzed in the TANGO algorithm, 21 of them were classified as increasing protein aggregation (Fig 4A), while 16 mutations were predicted to decrease this parameter. Protein aggregation is a central mechanism in the pathophysiology of neurodegenerative diseases and has also been linked to the development of psychiatric disorders. Nonetheless, mechanisms underlying protein aggregation in psychiatric disorders are still poorly understood. Korth (2012) analyzed the effects of the *DISC1* gene in the development of schizophrenia and other psychiatric disorders. According to the author, mutant forms of Disrupted-in-schizophrenia 1 (DISC1) protein were found in aggregates inside neurons of patients with schizophrenia. Moreover, a study by the Stanley Medical Research Institute's Consortium Collection, involving post-mortem investigation of patients with schizophrenia, major depression, and bipolar disorder, also found inclusions of DISC1 protein aggregates in neurons (KORTH, 2012). Kuhn *et al*. (2011) proposed a link between TPH2 aggregation upon oxidation with serotonin deficits and psychiatric symptoms. The authors observed that TPH2 oxidation cause misfolding with consequent loss of enzymatic activity and the formation of high molecular weight aggregates, which were associated with the development of neuropsychiatric conditions such as depression, sleep disorders, and anxiety (KUHN *et al.*, 2011). The study by Winge *et al*. (2006), in turn, suggested that defects in TPH2 catalytic function, usually related to the development of psychiatric disorders, possibly result from the increased aggregation and degradation propensity observed in mutant proteins (WINGE *et al.*, 2007).

The following mutations were predicted to increase TPH2 aggregation propensity: A4V, G62D, K63E, T64I, L70F, T134M, R225W, P244L, E261K, D262A, M265V, K268I, E269K, L291P, V295E, T356M, E417G, E417K, R441H, R441S, and R441C

(S3_Table). Interestingly, the following regions of TPH2 concentrated most of these predictions, including i) residues 62 – 70 in the regulatory domain; ii) residues 244 – 295 in the catalytic domain; iii) position 441, which contains three different mutations, all of them predicted to increase protein aggregation. Interestingly, this position is also affected by R441H mutation, which is deleterious. No mutation predicted to increase protein aggregation occurred in the oligomerization domain.

The WALTZ analysis indicated that 13 mutations were classified as increasing TPH2 amyloid propensity, while 13 mutations were predicted to reduce this parameter. Amyloid propensity is a central mechanism in the pathophysiology of neurodegenerative disorders like Alzheimer's (XU *et al.*, 2019). The review by Pandolfo *et al.* 2021 evidenced that altered amyloid metabolism is observed in major psychiatric disorders, although the mechanism underlying these changes is still poorly understood (PANDOLFO *et al.*, 2021). No studies on direct changes in the amyloid propensity of TPH2 protein were found in the literature consulted. Nonetheless, deficient serotonin signaling has been previously described as significantly increasing the formation of β-amyloid plaques in both humans and rats (XU *et al.*, 2019). The following mutations were predicted to increase TPH2 amyloid propensity: T64A, R82S, M91I, H93N, T130N, M170I, K268I, P308S, Q352R, R374W, S383F, T476I, and L481S (S3_Table).

In turn, the LIMBO analysis suggested that eight variants were predicted to increase the chaperone binding tendency of TPH2, while seven mutations were predicted to decrease this parameter (S3_Table). Chaperone binding is essential for the correct folding and functioning of proteins. Thus, impaired chaperone binding may have harmful implications for the protein (PEREIRA; VIEIRA; MESQUITA, 2021). The following variants were predicted to decrease the chaperone binding tendency of TPH2: H166Y, L169V, M170T, Y171C, R225W, R241Q, and F243V. Although no natural chaperones for TPH2 are currently known, *in vivo* studies indicated that using pharmacological chaperones to increase protein stability and promote the correct folding of TPH2 mutants has shown to be a promising strategy for the treatment of psychiatric disorders (KULIKOVA; KULIKOV, 2019).

3.3 EVOLUTIONARY CONSERVATION ANALYSIS

As shown in Fig 5, the evolutionary conservation degree of each amino acid of TPH2 was projected on the protein's surface and colored according to the ConSurf color-coding scale, which varies from 1 and cyan (highly variable) to 9 and maroon (highly

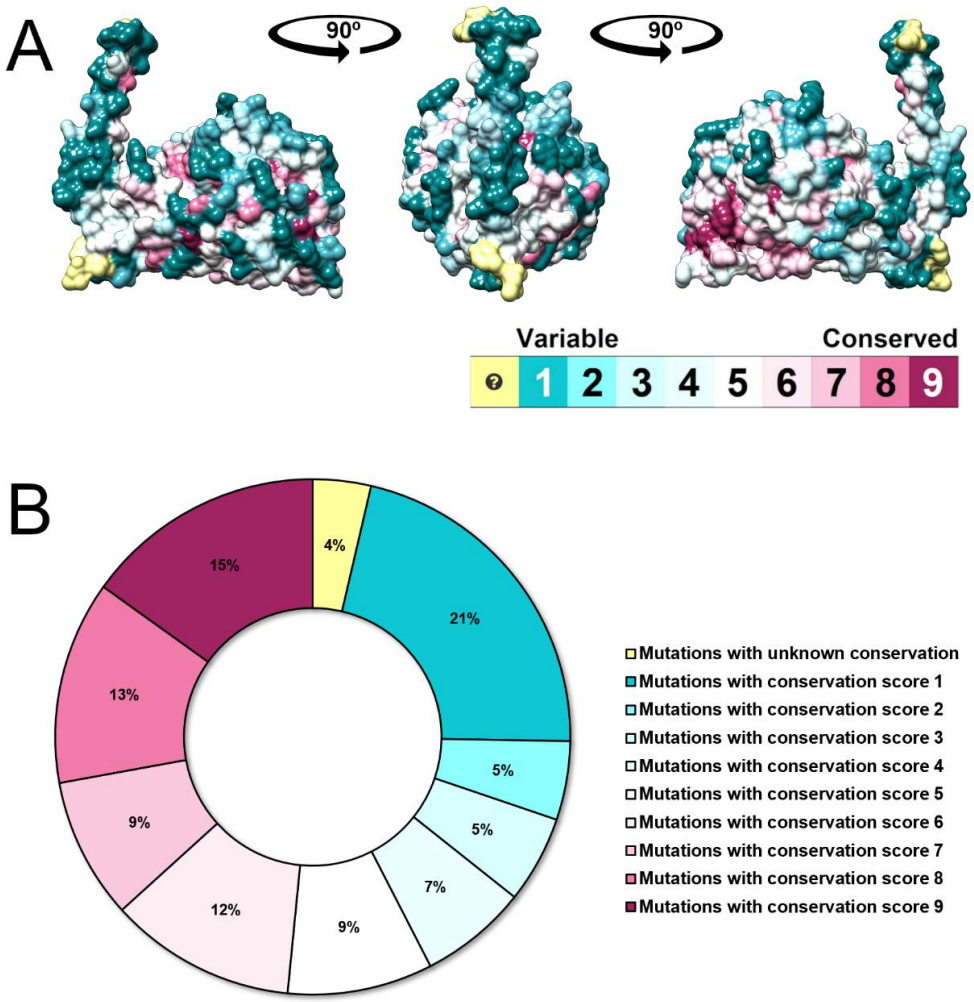
conserved). Biologically relevant amino acids, *i.e.* those with functional or structural importance, are often conserved throughout the evolution due to a higher selective pressure (CHOI *et al.*, 2012). Thus, highly conserved amino acids within a protein are usually functionally important (ASHKENAZY *et al.*, 2016), and mutations affecting these positions are those most likely to be harmful (PEREIRA *et al.*, 2022).

The catalytic domain is the most conserved among the three functional domains of TPH2, followed by the oligomerization domain, and then the evolutionary domain. Evolutionary data on the regulatory domain is not shown in Fig 5A since the three-dimensional structure of this region is still unknown (ROSE *et al.*, 2021). Our results are in agreement with those of SKAWINSKA, (2020), which analyzed the sequence conservation among the human aromatic amino acid hydroxylases (AAAH) protein family, including TPH2 (SKAWINSKA, 2020).

As shown in Fig 5B, the ConSurf analysis indicated that most TPH2 mutations (37%) occur in conserved positions, *i.e.* those with conservation score ≥ 7. It was followed by mutations occurring in variable positions (ConSurf-score ≤ 3), which correspond to 31%, and ultimately by mutations in average conserved positions (4 ≥ ConSurf-score ≤ 6), with 28%. Four percent of the TPH2 mutations did not receive a conservation score due to insufficient data.

Furthermore, a relatively high number of mutations, 15%, lie in highly conserved positions *i.e.* received the maximum conservation score (ConSurf-score = 9), thus, being possibly important for TPH2 structure and function (CHOI *et al.*, 2012). Most of these mutations (93%) occur in the catalytic domain (S4_Table). It could be related to the functional importance of this domain, which contains the catalytic site of TPH2 and amino acids involved in binding to substrate, co-factor, and co-substrate (Fig 2).

Fig 5. Evolutionary conservation analysis of human TPH2 protein. Each amino acid of TPH2 was colored according to its evolutionary conservation degree and projected on the protein's surface. The color-coding bar shows the evolutionary conservation scheme adopted by ConSurf, which ranges from cyan and variable to magenta and conserved. (A) TPH2 structure is represented as a space-filling model in three different angles rotating 90º from each other. (B) The number of mutations affecting amino acids with conservation degrees from 1 (cyan sector) to 9 (magenta sector) is shown in a pie chart.



Interestingly, among the missense mutations classified as deleterious by all the algorithms in the functional analysis (S2_Table), more than half also occur in positions that received maximum conservation scores (S4_Table). It may be related to the training strategy adopted by most of these algorithms, which mainly use evolutionary information extracted from the sequence to predict the effect of missense mutations on protein's function, as previously shown in Table 1.

Among the TPH2 variants associated with the development of psychiatric disorders, only P206S does not affect a conserved position. R303W affects a position with the maximum conservation degree (ConSurf-score = 9), while R441H affects an amino acid that received a conservation degree of eight. On the other hand, P206S affects a variable position of TPH2, with a conservation score of three. The high conservation of

positions 303 and 441, together with the known deleterious potential of R303W and R441H, suggest that these positions are likely to be biologically relevant for TPH2.

3.4 PENALTY ANALYSIS

From the 384 missense mutations compiled for TPH2 in the literature and databases, only five mutations have already been completely characterized. Additionally, seven TPH2 mutations have been partially characterized, so they are known to reduce serotonin synthesis but the functional consequence of this phenotype is still unknown (S1_Table). Thus, the effects of most TPH2 mutations, i.e. 379, are still unknown. In this scenario, aiming at facilitating the prioritization of most-likely deleterious mutations of TPH2, we adopted a penalty system by combining the predictive analysis with the evolutionary conservation degree, a literature review, and alterations in physicochemical properties. This approach was used to rank all uncharacterized mutations of TPH2 according to their harmful potential so that mutations receiving the highest penalties were ranked first, as shown in S5_Table.

The maximum penalty a mutation can receive is eleven, from which:

i) 0-2 points account for the number of deleterious predictions. Mutations consensus predicted as deleterious, *i.e.* equal or more than half plus one algorithm (≥ 6), received one point. An additional penalty was attributed to mutations predicted as deleterious by all the ten algorithms used (S2_Table).

ii) 0-2 points were attributed according to the evolutionary conservation degree. Mutations affecting conserved positions, *i.e.* ConSurf-score equal to or higher than 7, received a penalty. An additional point was attributed to mutations occurring in positions with the maximum conservation degree, *i.e.* 9. Biologically relevant residues are often conserved. Thus, mutations affecting these positions may have harmful consequences for protein function (S4_Table) (ASHKENAZY *et al.*, 2016).

iii) One point accounts for mutations consensus predicted as destabilizing (≥2 algorithms). Deleterious mutations in TPH2 are known to reduce protein stability and solubility, which ultimately leads to loss of enzymatic activity and diminished serotonin levels (S3_Table) (UNIPROT CONSORTIUM, 2021).

iv) 0-2 points were assigned to increased aggregation tendency and increased amyloid propensity. These alterations are known to be related to the development

of psychiatric disorders (S3_Table) (KUHN *et al.*, 2011; PANDOLFO *et al.*, 2021).

v) One point is attributed to whether the mutation occurs at a relevant residue for TPH2 function, which includes those binding to co-factors, co-substrate, and substrate, in addition to those involved in post-translational modifications (Fig 2).

vi) 0-1 point accounts for alterations in physicochemical properties upon amino acid substitution, including hydrophobicity, size, and charge. Two alterations received a penalty, while three alterations received 2 points. Changes in the size or biochemical properties of an amino acid can alter or prevent protein function, which is observed in most genetic disorders (KHAN; VIHINEN, 2007). Among them, radical amino acid substitutions, *i.e.* those involving amino acids with very distinct physicochemical properties, are more likely to be deleterious due to a stronger negative selective pressure (S5_Table) (WEBER; WHELAN, 2019).

vii) One point is attributed to whether a mutation occurs in the same position as a deleterious mutation of TPH2. Considering that positions 206, 303, and 441 are affected by mutations already known to be associated with psychiatric disorders (UNIPROT CONSORTIUM, 2021), these positions are likely to be biologically important for TPH2 (S1_Table).

viii) One point was attributed to TPH2 mutations already known to reduce serotonin synthesis (S1_Table). Altered serotonin levels in the brain have been related to the development of psychiatric disorders (FAROOK *et al.*, 2012).

Among the non-characterized variants of TPH2, V295E and R441C received the highest penalties, with 7 points, followed by variants T311P, Y281C, R441S, S383F, P308S, Y281H, and E363G, which received six penalty points (S5_Table). According to the criteria discussed earlier, these mutations are most likely to be deleterious for TPH2, thus, being important targets for future investigation.

4 CONCLUSIONS

In this work, we compiled 384 missense mutations of TPH2, which are distributed all over the protein, similarly affecting the three functional domains. Among them, 379 have not yet been characterized. Thus, their effects are still unknown. Our predictive analysis pointed to a high rate of deleterious and destabilizing predictions for the TPH2 mutations. These mutations mainly affect conserved and, possibly, functionally important

residues within the catalytic domain, responsible for protein's enzymatic activity. The predictive analysis also suggested that 16 mutations increase protein aggregation tendency, while 13 mutations increase amyloid fiber formation, which are alterations related to the development of psychiatric disorders. We also used a penalty system to prioritize the most likely harmful mutations of TPH2 by combining the predictive analyses, evolutionary conservation, literature review, and alterations in physicochemical properties of amino acid substitution. Among the uncharacterized mutations of TPH2, variants V295E, R441C T311P, Y281C, R441S, S383F, P308S, Y281H, and E363G received the highest penalties, thus, being the most likely deleterious and, consequently, important targets for future investigation. It may guide the design of clinical and laboratory experiments, optimizing time and resources.

FUNDING

This study was supported by Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Financiadora de Estudos e Projetos, Conselho Nacional de Desenvolvimento Científico e Tecnológico, Universidade Federal do Estado do Rio de Janeiro, and NVIDIA Corporation. The funders had no role in study design and manuscript preparation.

COMPETING INTERESTS

The material received as support from NVIDIA for this study does not alter our adherence to the journal policies on sharing data and materials.

REFERENCES

AMERICAN PSYCHIATRIC ADZHUBEI, IA *et al.* PolyPhen-2 : prediction of functional effects of human nsSNPs. **Curr Protoc Hum Genet**, 2010.

ARDITO, Fatima *et al.* The crucial role of protein phosphorylation in cell signalingand its use as targeted therapy (Review). **International Journal of Molecular Medicine**, v. 40, n. 2, p. 271–280, 2017.

ASHKENAZY, Haim *et al.* ConSurf 2016 : an improved methodology to estimate and visualize evolutionary conservation in macromolecules. **Nucleic Acids Research**, v. 44, n. May, p. 344–350, 2016.

ASSOCIATION, American Psychiatric. **Diagnostic and Statistical Manual of Mental Disorders 5**. 5th. ed. Arlington: American Psychiatric Publishing, 2013. *E-book*.

BENDL, Jaroslav *et al.* PredictSNP: Robust and Accurate Consensus Classifier for Prediction of Disease-Related Mutations. **PLoS Computational Biology**, v. 10, n. 1, p. 1–11, 2014.

BROMBERG, Yana; YACHDAV, Guy; ROST, Burkhard. SNAP predicts effect of mutations on protein function. **Bioinformatics**, v. 24, n. 20, p. 2397–2398, 2008.

CAO, Xiaoyong *et al.* Identification of metal ion binding sites based on amino acid sequences. **PLoS ONE**, *[S. l.]*, v. 12, n. 8, p. 1–16, 2017.

CAPRIOTTI, Emidio *et al.* WS-SNPs&GO: a web server for predicting the deleterious effect of human protein variants using functional annotation. **BMC genomics**, v. 14, n. Suppl 3, p. S6, 2013.

CAPRIOTTI, Emidio; FARISELLI, Piero. PhD-SNPg: A webserver and lightweight tool for scoring single nucleotide variants. **Nucleic Acids Research**, v. 45, n. W1, p. W247–W252, 2017.

CAPRIOTTI, Emidio; FARISELLI, Piero; CASADIO, Rita. I-Mutant2.0: Predicting stability changes upon mutation from the protein sequence or structure. **Nucleic Acids Research**, v. 33, n. Web Server, p. W306–W310, 2005.

CARKACI-SALLI, Nurgul *et al.* Functional domains of human tryptophan hydroxylase 2 (hTPH2). **Journal of Biological Chemistry**, v. 281, n. 38, p. 28105–28112, 2006.

CARKACI-SALLI, Nurgul *et al.* Functional characterization of the S41Y (C2755A) polymorphism of tryptophan hydroxylase 2. **Journal of neurochemistry**, v. 130, n. 6, p. 748–758, 2014.

CHOI, Yongwook *et al.* Predicting the Functional Effect of Amino Acid Substitutions and Indels. **PLoS ONE**, v. 7, n. 10, 2012.

CHOI, Yongwook; CHAN, Agnes P. PROVEAN web server: A tool to predict the functional effect of amino acid substitutions and indels. **Bioinformatics**, v. 31, n. 16, p. 2745–2747, 2015.

CICHON, Sven *et al.* Brain-specific tryptophan hydroxylase 2 (TPH2): A functional Pro206Ser substitution and variation in the 5′-region are associated with bipolar affective disorder. **Human Molecular Genetics**, v. 17, n. 1, p. 87–97, 2008.

DE BAETS, Greet *et al.* SNPeffect 4.0: On-line prediction of molecular and structural effects of protein-coding variants. **Nucleic Acids Research**, v. 40, n. Database, p. D935–D939, 2012.

FAROOK, M. Febin *et al.* Altered serotonin, dopamine and norepinepherine levels in 15q duplication and Angelman syndrome mouse models. **PLoS ONE**, v. 7, n. 8, p. 1–9, 2012.

JEAN-BAPTISTE, Lamy; BERTHELOT, Hélène; FAVRE, Madeleine. Rainbow boxes: A technique for visualizing overlapping sets and an application to the comparison of drugs properties. **Proceedings of the International Conference on Information Visualisation**, v. 2016-August, n. July, p. 253–260, 2016.

KESSLER, Ronald C. *et al.* Lifetime Prevalence and Age-of-Onset Distributions of. **Arch Gen Psychiatry**, v. 62, n. June, p. 593–602, 2005.

KHAN, Sofia; VIHINEN, Mauno. Spectrum of disease-causing mutations in protein secondary structures. **BMC Structural Biology**, v. 7, p. 1–18, 2007.

KORTH, Carsten. Aggregated proteins in schizophrenia and other chronic mental diseases: DISC1opathies. **Prion**, *[S. l.]*, v. 6, n. 2, p. 134–141, 2012.

KUHN, Donald M. *et al.* Tryptophan hydroxylase 2 aggregates through disulfide cross-linking upon oxidation: Possible link to serotonin deficits and non-motor symptoms in Parkinson's disease. **Journal of Neurochemistry**, *[S. l.]*, v. 116, n. 3, p. 426–437, 2011.

KULIKOVA, Elizabeth A.; KULIKOV, Alexander V. Tryptophan hydroxylase 2 as a therapeutic target for psychiatric disorders: focus on animal models. **Expert Opinion on Therapeutic Targets**, v. 23, n. 8, p. 655–667, 2019.

LANDRUM, Melissa J. *et al.* ClinVar : improving access to variant interpretations and supporting evidence. **Nucleic Acids Research**, v. 46, n. Database, p. D1062–D1067, 2018.

LÓPEZ-FERRANDO, Víctor *et al.* PMut: A web-based tool for the annotation of pathological variants on proteins, 2017 update. **Nucleic Acids Research**, v. 45, n. W1, p. W222–W228, 2017.

MAKWANA, Mehul V.; MUIMO, Richmond; JACKSON, Richard F. W. Advances in development of new tools for the study of phosphohistidine. **Laboratory Investigation**, v. 98, n. 3, p. 291–303, 2018.

MCKINNEY, J. *et al.* A loss-of-function mutation in tryptophan hydroxylase 2 segregating with attention-deficit/hyperactivity disorder. **Molecular Psychiatry**, v. 13, n. 4, p. 365–367, 2008.

MCKINNEY, Jeffrey A. *et al.* Functional properties of missense variants of human tryptophan hydroxylase 2. **Human Mutation**, *[S. l.]*, v. 30, n. 5, p. 787–794, 2009.

MI, Huaiyu *et al.* PANTHER version 11: Expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. **Nucleic Acids Research**, v. 45, n. D1, 2017.

MUKHERJEE, Joia S. Global Health and the Global Burden of Disease. **Oxford Scholarship Online**, n. December, 2016.

NCBI. Database Resources of the National Center for Biotechnology Information. **Nucleic Acids Research**, v. 45, n. D1, p. D12–D17, 2017.

OLIVEIRA, Clara Carolina Silva De *et al.* In silico analysis of the V66M variant of human BDNF in psychiatric disorders : An approach to precision medicine. **PLoS ONE**, v. 14, n. 4, p. e0215508, 2019.

PAIXÃO, Gabriela Miana de Mattos *et al.* Machine Learning in Medicine: Review and Applicability. **Arquivos Brasileiros de Cardiologia**, v. 118, n. 1, p. 95–102, 2022.

PANDOLFO, Gianluca *et al.* Mental illness and amyloid: A scoping review of scientific evidence over the last 10 years (2011 to 2021). **Brain Sciences**, v. 11, n. 10, 2021.

PEJAVER, Vikas *et al.* MutPred2: inferring the molecular and phenotypic impact of amino acid variants. **bioRxiv**, *[S. l.]*, p. 134981, 2017.

PEREIRA, Gabriel Rodrigues Coutinho *et al.* In silico analysis of the tryptophan hydroxylase 2 (TPH2) protein variants related to psychiatric disorders. **PLoS ONE**, v. 15, n. 3, p. 1–23, 2020.

PEREIRA, Gabriel Rodrigues Coutinho *et al.* Analysis of mutations in APP1 protein associated with development and protection against Alzheimer ' s disease - an In silico approach. **Brazilian Journal of Development**, *[S. l.]*, v. 8, n. 6, p. 46902–46924, 2022.

PEREIRA, Gabriel Rodrigues Coutinho; DE AZEVEDO ABRAHIM VIEIRA, Bárbara; DE MESQUITA, Joelma Freire. Comprehensive in silico analysis and molecular dynamics of the superoxide dismutase 1 (SOD1) variants related to amyotrophic lateral sclerosis. **PLoS ONE**, v. 16, n. 2 February, p. 1–27, 2021.

PEREIRA, Gabriel Rodrigues Coutinho; VIEIRA, Barbara de Azevedo Abrahim; MESQUITA, Joelma Freire De. Comprehensive in silico analysis and molecular dynamics of the superoxide dismutase 1 (SOD1) variants related to amyotrophic lateral sclerosis. **PLOS ONE**, v. 1, n. 0247841, p. 1–27, 2021.

ROSE, Yana *et al.* RCSB Protein Data Bank: Architectural Advances Towards Integrated Searching and Efficient Access to Macromolecular Structure Data from the PDB Archive. **Journal of Molecular Biology**, *[S. l.]*, v. 433, n. 11, p. 166704, 2021.

ROY CHOUDHURY, Amrita *et al.* Supporting precision medicine by data mining across multi-disciplines: An integrative approach for generating comprehensive linkages between single nucleotide variants (SNVs) and drug-binding sites. **Bioinformatics**, v. 33, n. 11, p. 1621–1629, 2017.

SANAVIA, Tiziana *et al.* Limitations and challenges in protein stability prediction upon genome variations: towards future applications in precision medicine. **Computational and Structural Biotechnology Journal**, *[S. l.]*, v. 18, p. 1968–1979, 2020.

SANSONE, Randy A.; SANSONE, Lori A. PSYCHIATRIC DISORDERS : A Global Look at Facts and Figures PREVALENCE OF PSYCHIATRIC. **Psychiatry**, v. 7, n. 12, p. 16–19, 2010.

SAVOJARDO, Castrense *et al.* INPS-MD: a web server to predict stability of protein variants from sequence and structure: Table 1. **Bioinformatics**, v. 32, n. 16, p. 2542–2544, 2016.

SKAWINSKA, Natalia. **Allosteric regulation of human tryptophan hydroxylase isoform 2 (hTPH2)**. 2020. - Technical University of Denmark, 2020.

STANLEY, Susanne *et al.* A model of integrated care to address the physical health of people with severe mental illness. **The Wellness Clinic**, n. June, p. 1–46, 2019.

ŠTRAC, Dubravka Švob; PIVAC, Nela; MÜCK-ŠELER, Dorotea. The serotonergic system and cognitive function. **Translational Neuroscience**, v. 7, n. 1, p. 35–49, 2016.

TIDEMAND, Kasper D. *et al.* Stabilization of tryptophan hydroxylase 2 by l-phenylalanine-induced dimerization. **FEBS Open Bio**, v. 6, n. 10, p. 987–999, 2016.

TRAUTMANN, Sebastian; REHM, Jürgen; WITTCHEN, Hans-Ulrich. The economic costs of mental disorders. **EMBO reports**, v. 17, n. 9, p. 1245–1249, 2016.

UNIPROT CONSORTIUM. UniProt: the universal protein knowledgebase in 2021. **Nucleic Acids Research**, *[S. l.]*, v. 49, n. D1, p. D480–D489, 2021.

VASER, Robert *et al.* SIFT missense predictions for genomes. **Nature Protocols**, v. 4, n. December 2015, p. 1073–1081, 2015.

WEBER, Claudia C.; WHELAN, Simon. Physicochemical amino acid properties better describe substitution rates in large populations. **Molecular Biology and Evolution**, v. 36, n. 4, p. 679–690, 2019.

WELFORD, Richard W. D. *et al.* Serotonin biosynthesis as a predictive marker of serotonin pharmacodynamics and disease-induced dysregulation. **Scientific Reports**, v. 6, n. June, p. 1–10, 2016.

WHO. **World Health Organization - Mental Health**. 2017.

WINGE, Ingeborg *et al.* Characterization of wild-type and mutant forms of human tryptophan hydroxylase 2. **Journal of Neurochemistry**, v. 100, n. 6, p. 1648–1657, 2007.

XU, Chao-Jin *et al.* Tph2 Genetic Ablation Contributes to Senile Plaque Load and Astrogliosis in APP/PS1 Mice. **Curr Alzheimer Res**, v. 16, n. 3, p. 219–232, 2019.

YANG, Jiarun *et al.* The interaction of TPH2 and 5-HT2A polymorphisms on major depressive disorder susceptibility in a Chinese Han population: A case-control study. **Frontiers in Psychiatry**, v. 10, n. APR, p. 1–7, 2019.

ZHANG, Xiaoyan; WANG, Yiming. TPH2: A Key Gene Risk Factor and Potential Therapy Target in Depression. **E3S Web of Conferences**, v. 271, p. 03070, 2021.

SUPPORTING INFORMATION

The supporting information is available at: https://doi.org/10.6084/m9.figshare.20388033