In silico identification of the rare-coding pathogenic mutations and structural modeling of human NNAT gene associated with Anorexia Nervosa

Muhammad Bilal Azmi (bilal.azmi@duhs.edu.pk)
Dow University of Health Sciences https://orcid.org/0000-0001-8320-4479
Unaiza Naeem
Dow Medical College
Arisha Saleem
Dow Medical College
Areesha Jawed
Dow Medical College
Haroon Usman
University of Karachi Department of Biochemistry
Shamim Akhter Qureshi
University of Karachi Department of Biochemistry
Muhammad Kamran Azim
MAJU: Mohammad Ali Jinnah University

Research Article

Keywords: Anorexia nervosa, acetylation, in silico, mutations, neuronatin

Posted Date: April 11th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1521353/v1

Abstract

Purpose Increase susceptibility towards *Anorexia nervosa* (AN) was reported with reduced levels of NNAT gene. We sought to investigate the most pathogenic rare-coding missense mutations (nsSNPs) of NNAT and their potential damaging impact on protein function through transcript level sequence and structure based *in silico* approaches.

Methods Gene sequence, SNPs of NNAT was retrieved from public databases and the putative post-translational modification (PTM) sites were analyzed. Distinctive *in silico* algorithms were recruited for transcript level SNPs analyses and to characterized high risk rare-coding nsSNPs along with their impact on protein stability function. *Ab initio* 3D-modeling of wild-type, alternate model prediction for most deleterious nsSNP, validation and recognition of druggable binding pockets were also performed. AN 3D therapeutic compounds followed rule of drug-likeness were docked with most pathogenic variant of NNAT to estimate the drugs' binding free energies.

Results Conclusively, 10 transcripts (201–205) based nsSNPs from 3 rare-coding missense variants i.e., *rs539681368*, *rs542858994*, *rs560845323* out of 840 exonic SNPs were identified. Transcript based functional impact analyses predicted *rs539681368* (C30Y) from NNAT-204 as the high risk rare-coding pathogenic nsSNP, deviating protein functions. The 3D-modeling analysis of AN drugs' binding energies indicated lowest binding free energy (ΔG) and significant inhibition constant (K_i) with mutant models C30Y.

Conclusions Mutant model (C30Y) exhibiting significant drug binding affinity and the commonest interaction observed at the acetylation site K59. Thus, based on these findings, we concluded that the identified nsSNP may serve as potential targets for various studies, diagnosis and therapeutic interventions.

Level of evidence No level of evidence

What Is Already Know On The Subject?

Lombardi *et al.,* 2019 and Ceccarrini *et al* in 2021 reported the role of confirmation of the involvement of the NNAT gene in the pathogenesis of AN. Further, its role in energy metabolic homeostasis has been clearly established through a vast spectrum of literature.

What your study adds?

Transcript level predictions of rare-coding high risk pathogenic human NNAT gene's SNPs by different *in silico* ML algorithms and their functional impact on protein stability. Characterizing the structural inference of rare-coding deleterious missense mutations and computing the molecular binding energies of most pathogenic 3D model of NNAT variant with therapeutic compounds of AN.

Introduction

Eating disorders (ED) are heterogeneous psychiatric condition that manifest as abnormal dietary habits which can affect the physical, psychological and social functions, leading to a reduced quality of life [1]. EDs includes Anorexia Nervosa (AN), Bulimia Nervosa (BN), Binge Eating Disorders (BEDs), and Eating Disorders Not Otherwise Specified (EDNOS) [1]. *Wu* and co-workers (2020) reported the Age Standardized Rates (ASRs) of prevalence for EDs have increased by yearly average of 0.65 while the DALYS (Disability Adjusted Life Years) of EDs have also constantly increased globally by yearly average of 0.66 [1]. AN a perplexing heritable disorder, recognized as the most incapacitating disease [2] characterized by an abnormally low weight perception with an intense fear of gaining body weight with a compulsive desire to be thin (Ref). AN is well-known to be triggered at puberty [3] and diagnosis of the disease is confirmed with the presence of subsequent symptoms [2].

Globally, women with the age between 12–35 years have been found to be more affected with EDs and currently there is a rising upsurge of EDs specifically in Asian countries [4]. Similarly, East Asia has had high increment in ASRs for AN followed by South Asia during last two decades [1]. South Asian women generally have been reported to have a statistically high risk of developing an ED [5] and the mean age of presentation of AN in South Asian female is 15.07 years [6]. In Pakistan, the ASRs of AN prevalence is observed to have an increment of 30.60% per 100,000 population while studies with small female sample population showed 21.67% [7] and 42% [8] respectively reported to be having AN in 2002 and 2015.

Genome wide searches have been unsuccessful in clearly identifying specific genes that predispose an individual to this disorder. Hence, a proper mechanism based association of AN and its etiological basis has not been established [2]. A study undertaken by *Lomabrdi* and his co-researchers in 2019 sequenced some families (with AN) and found significant association of NNAT gene variants (single nucleotide polymorphism - SNP) in identified population with AN [2]. The possible mechanism of AN development is through reduced level of the NNAT-α isoform in brain [2]. This classifiably relates with brain specific developmental gene involved in neuronal differentiation and other associated metabolic functions [9]. They correlated the dysregulation of NNAT gene (α isoform) expression could be linked with the propensity of having AN [2]. Ceccarrini *et al* in 2021 reported the role of confirmation of the involvement of the NNAT gene in the pathogenesis of AN [10]. In coherence with this posit we have attempted to further examine NNAT's changing expression in the brain as a pivotal finding and relate its role in development of AN. Previously pivotal impact of nsSNPs on NNAT protein regarding AN pathogenesis is not well understood. Thus, in present effort, we explored the transcript level prediction of the rare-coding deleterious variants of human NNAT gene through *in silico* sequence and structural machine learning (ML) algorithms and their impact of mutation on protein stability. We also attempted to determine the structural inference of rare-coding missense SNPs of NNAT gene and to estimate the *in silico* molecular binding energies of 3D homology model of most pathogenic variant with therapeutic compounds of AN.

Methodology

Extraction of AN genomic information

Present scheme summarizing the important steps of present study are mentioned in Fig. 1. The genomic information was extracted through database associated protein search and the strategy for the retrieval of AN proteins was associated with AN key terms (synonyms) *"Anorexia nervosa"*, *"Homo sapiens"* from the NCBI (the National Center for Biotechnology Information) databases (https://www.ncbi.nlm.nih.gov/genes). Similarly, the AN genes along with their corresponding proteins were extracted from the period of January 2010 till December 2020 (date of retrieval: 8th March 2021). All the essential genomic information was further retrieved from HGNC (HUGO Gene Nomenclature Committee), NCBI https://www.ncbi.nlm.nih.gov/gene, ENSEMBL http://asia.ensembl.org/Homo_sapiens/Gene/Summary; HGNC (HUGO Gene Nomenclature Committee) https://www.genenames.org/; and GeneCards (https://www.genecards.org/) database resources.

Retrieving human tissue-specific expression and gene-gene interactions

HPA RNA-sequence normal tissues data available through https://www.ncbi.nlm.nih.gov/bioproject/PRJEB4337/ were used to identify the tissue-specificity of all AN protein-coding genes [11]. Subsequently, the highest expressed gene in brain tissues was selected to further explore its expression pattern through GTEx (Genotype-Tissue Expression portal [12]. The gene-gene interactions network was studied with GeneMANIA [13] and STRING [14] (https://string-db.org/cgi/) web-servers (Ref).

Extraction of rare-coding exonic SNPs

SNPs of selected gene were extracted from ENSEMBL (https://asia.ensembl.org/index.html) database (Human Genome Assembly, GRCh38.p13 build 38). Exonic variants were extracted by the application of the filter in VEP (Variant Effect Predictor) module of ENSEMBL [15]. The rare-coding variants were selected by the global minor allele frequency (MAF ≤ 0.001) [16] and 'missense variants' mutation consequence type, led to the extraction of rare-coding missense variants (nsSNP). This were further filtered to extract the rare-coding pathogenic variants through Storing Intolerant from Tolerant (SIFT) [17], Polymorphism Phenotyping (PolyPhen2) [18], Combined Annotation Dependent Depletion (CADD) [19], Rare Exome Variant Ensemble Learner (REVEL) [20], MetaLR and Mutation Assessor [21, 22].

Transcript based functional analyses and estimation of expected accuracy through consensus classifier for predicting rare-coding nsSNPs

Rare-coding nsSNPs were computationally investigated by the transcript based functional analyses through various *in silico* ML algorithms, to assess whether the respective variants are pathogenic or not. These *in silico* ML approaches include; **PROVEAN** - a sequence homology-based tool, developed through web server on http://provean.jcvi.org/index.php. The cutoff score of PROVEAN was found to be -2.5 and mutations with scores over − 2.5 are predicted to be pathogenic [23]. PhD-SNP - pathogenicity-predicting tool based on support vector machines through https://snps.biofold.org/phd-snp/phd-snp.html [24]. SNAP2 **-** a neural network-based classifier which predicts the impact (effect) of single amino acid substitutions on protein function available through https://rostlab.org/services/snap2web/. A high score of > 50 indicates a strong signal for effect [25]. SNPs&GO **–** this tool predicts the variation of a protein sequence and whether the variation is disease related or not, available on https://snps.biofold.org/snps-and-go/snps-and-go.html [26]. PMut **-** a web-based tool for annotation of pathological variants on proteins, was trained and tested manually, available through http://mmb.irbbarcelona.org/PMut/. The prediction scores range from 0 to 1, and the cutoff value is set at 0.5 (neutral, 0 to 0.5; pathological, 0.5 to 1) [27]. MutPred2 - a web-server application developed to categorize amino acid substitutions on functional proteins as pathogenic or benign in human, available through http://mutpred.mutdb.org/. The scores ranging from 0.5 to 1.0 are classified as pathogenic whereas, values less than 0.5 are benign [28]. The combined expected accuracy of nsSNP tools, were estimated with PredictSNP1.0 (http://loschmidt.chemi.muni.cz/predictsnp1/) web server as the consensus classifier predictor of the SNP effect on protein function. This consensus classifier jointly used nine best performing prediction tools: SIFT, PolyPhen-1, PolyPhen-2, MAPP, PhD-SNP, SNAP, PANTHER, PredictSNP, and nsSNPAnalyzer [29].

Transcript based computation of changes in free energy of mutation and structural inference of rare-coding pathogenic nsSNPs

The computations of changes in free energy in transcript level nsSNPs on the basis of transcript sequences were estimated through **MUpro** ML programs, which predict how single-site amino acid mutation affects protein stability on the basis of two ML methods *i.e.*, Support Vector Machines (SVM) and Neural Networks (NN)., available at http://mupro.proteomics.ics.uci.edu/ [30]. **I-Mutant -** v.3.0 is a predictor of protein stability changes upon mutations, tool for the automatic prediction of protein stability changes upon single point mutations at pH 7 and 25°C temperature [31]. **iPTREE-STAB -** iStable v.2.0 is an interpretable decision tree-based method for discriminating the stability of proteins (predicting their stability changes, ΔΔG) upon single amino acid substitutions from amino acid sequence, http://ncblab.nchu.edu.tw/iStable2/seqsubmit.html [32]. **INPS-MD** (Impact of Non-synonymous mutations on Protein Stability-Multi Dimension), https://inpsmd.biocomp.unibo.it [33].

For structural inference of rare-coding nsSNPs, the SOPMA (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html) - a neural network method, jointly used to predict the secondary structure (*i.e., α* helix, *β* turn and coil) of amino acids [34]. The tool HOPE (Have (y) Our Protein Explained), available on http://www.cmbi.ru.nl/hope/input/, was used to predict the effect(s) of amino acid variation due to nsSNPs on physical and chemical properties, hydrophobicity, spatial structure, and function of proteins, [35].

Sequence alignment, ab initio **3D structure modeling, validation, and alternate homology model prediction for pathogenic (**rs539681368:C30Y**) mutation of NNAT gene**

The pathogenic variant (*rs539681368:C30Y*, NNAT-204) was observed with the highest pathogenic scores, as compared to others and it has the gold standard type - high quality, reviewed transcript sequence, compared to rest others. NNAT-204 transcript sequence has been represented as protein coding canonical sequence / NNAT α-isoform, retrieved through UNIPROT database (accession code: Q16517). The wild type (WT) sequence of NNAT-204 transcript was downloaded and the multiple sequence alignment (MSA) of NNAT (α-isoform) gene was performed through Predict Protein web server

(https://predictprotein.org/) [36]. InterPro tool was used for domain classification and functional analysis (https://www.ebi.ac.uk/interpro/) [37]. The protein post-translational modification (PTM) site prediction was performed by MusiteDeep (https://www.musite.net) deep-learning web-tool [38].

As yet, no previous 3D structural coverage was reported from PDB (Protein Data Bank) [39], hence, QUARK *ab initio* protein structural prediction algorithm from Zhang Lab [40] was used to construct the 3D model of canonical sequence of NNAT (α-isoform) gene. The reason for selecting of QUARK web-server was based on its suitability for protein 3D structural modeling for specific protein sequences that do not have homologous templates in the PDB library. The validation of 3D homology model was carried out by PROCHECK stereo-chemical assessment through quality factor, Ramachandran plot and residual properties of constructed model. The dihedral angles φ against ψ of possible conformations of amino acids in protein structure had also been studied in Ramachandran plot [41]. SAVES (Structure Validation Server: https://saves.mbi.ucla.edu/) web server was used to know the probable structural errors and z-score [41]. Secondary structure properties of NNAT protein was investigated through the α-helices, identification of trans-membrane region and low complexity region [34]. Alternate model prediction for *NNAT* pathogenic *(rs539681368:C30Y)* variant was done through residue specific mutation approach by applying the UCSF Chimera visualization program [42].

To investigate the 3D structural variation associated with the location of the mutated residue in protein structure (PDB file) environment the CUPSAT web-server (http://cupsat.tu-bs.de/) was used [43], which estimated the difference in specific atom potentials and torsion angle potentials (ΔΔG = kcal/mol), and the difference in free energy of unfolding between WT and mutant protein.

AN therapeutic compounds, drug-likeness prediction and identification of druggable protein cavities

3D structures of AN drugs were explored from the publicly accessible DrugBank database [44], the search strategy included key terms and synonyms, modified to retrieve AN-related drugs and therapeutics were "*Anorexia Nervosa* (AN)". Overall, six 3D compounds were filtered. Lipinski's rule of five was applied to all 3D compounds, and those were selected which had PASS in all five physicochemical properties. The drug-likeness prediction were further computed and validated with the aid of OSIRIS Property Explorer (https://www.organic-chemistry.org/prog/peo/) which uses chemical structures for drug property prediction in terms of Topological polar surface area (TPSA), c log *P* calculation, log *S* calculation, molecular weight, fragment based drug-likeness, and drug score.

To identify the druggable protein binding cavities, CB (Cavity-detection guided Blind Docking approach) dock was used to explore the potential binding cavities/residues [45]. CB-Dock is an automatic protein-ligand docking approach which identifies the binding cavities, compares and ranked them through a process called 'CurPocket' using the benchmark set of COACH server [45], as prediction methods. This method was carefully optimized and achieved ~ 70% success rate for the top-ranking poses whose root mean square deviation (RMSD) was within 2 Å from the X-ray pose [45]. The 3D model of NNAT (WT) was CB docked with all 3D compounds of AN and binding affinities were noted. The consensus binding residues (CBR) were examined with 2D plot and selected according to the common appearance in highest grading cavities.

Molecular docking based energy estimation of NNAT pathogenic variant (C30Y)

The 3D model of pathogenic (*rs539681368*, C30Y) model of NNAT was docked with selected AN 3D compounds by AutoDock Vina tool (version 4.2) to determine the effect of pathogenic mutation over the drug-binding affinity of NNAT [46]. Initially, polar hydrogen was added and partial charges were assigned to the standard residue using Gasteiger partial charge, which assumes all hydrogen atoms were represented explicitly. The obtained interacting residues as a result of drugagble cavity binding pattern (CB-dock) with lowest free binding energy and RMSD value were further used to explore the interacting targets with NNAT pathogenic variant. It combines Genetic Algorithm that provides fast prediction of bound conformations with predicted free energies of association. Prior to docking, the grid parameter was defined to cover the full CBR of NNAT modeled structure using a grid-box spanning the whole protein centered at the coordinates x: -3.306Å, y: -3.861Å and z: 1.917Å.

Results

Genes and proteins of AN

In humans, database based search strategy resulted in 4 genes *i.e.*, Neuronatin, Membrane bound O-acyltransferase domain containing 4, Spermatogenesis associated 17 and Zinc finger protein 804B coding 6 proteins associated with AN. Neuronatin (*NNAT or Peg5* – location: 20q11.23) gene (ID-4826) with 3 exon codes for NNAT proteins. Membrane bound O-acyltransferase domain containing 4 (*FKSG89/ MBOAT4/ OACT4/ GOAT* – location: 8p12) gene (ID- 619373) with 4 exon codes for protein ghrelin O-acyltransferase. Spermatogenesis associated 17 (*CFAP305/ FAP305/ MOT17/ MSRG11/ MSRG-11/ IQCH/ SPATA17* – location: 1q41) gene (ID- 128153) with 12 exon codes spermatogenesis-associated protein 17. Zinc finger protein 804B (*ZNF804B* – location: 7q21.13) gene (ID: 219578) with 4 exon codes for zinc finger protein 804B.

Gene tissue expression

The retrieved data showed highest expression of NNAT gene in brain tissue with a value of 49.77 ± 24.26 RPKM. The genes SPATA17, ZNF804B and MBOAT4, were expressed with a value of 0.20 ± 0.09, 0.12 ± 0.003 and 0.08 ± 0.03 RPKM in brain, respectively. NNAT gene was highly expressed in the placenta (54.26 ± 25.45 RPKM). NNAT's expression was further explored in brain tissues through GTEx portal which showed the highest median expression especially in *Nucleus accumbens* (Basal ganglia) region i.e., 943 TPM and *Caudate* (Basal ganglia) i.e., 300.4 TPM. The gene-gene interaction profile of NNAT showed its physical interaction with NHLRC1 (NHL Repeat Containing E3 Ubiquitin Protein Ligase 1) and FKBP1B (FKBP prolyl isomerase 1B). The overall genetic interactions, co-expression, pathway, shared protein domains function are mentioned in Fig. 2.

Mutational estimation of NNAT

Total available SNPs obtained from ENSEMBL genome browser of NNAT gene (retrieved on August 2021) were 4647. The exonic variants were 840 and through filtration strategy rare-coding SNPs were 19. The filtration through functional downstream effects resulted in 10 rare pathogenic missense / nsSNP (see Fig. 3).

In silico **functional analyses and estimation of free energy change in rare-coding nsSNPs**

Overall, we extracted 3 rare-coding nsSNPs i.e., *rs539681368*, *rs542858994*, *rs560845323* distributed with 5 transcripts of NNAT (201–205) (see Table 1). The most pathogenic scores computed through various *in silico* algorithms were observed in nsSNP '*rs539681368'* (C30Y) of transcript sequence NNAT-204 and the lowest scores were computed in *rs542858994* (Q52E) of transcript NNAT-203 (see Table 1). Consensus classifier prediction of the rare-coding nsSNPs also categorized *rs539681368:C30Y* (NNAT-204) as the most deleterious nsSNP of the NNAT gene (see Table 2). The free energy changes estimated through mentioned *in silico* tools showed a decrement in the stability of protein upon these three mutations *i.e., rs539681368*, *rs542858994*, *rs560845323* (see Table 3). Mutation *rs539681368:C30Y* (NNAT-204) also showed a change in the protein stability.

The mutation *rs539681368:C30Y* (NNAT-204) was also validated from dbSNP (https://www.ncbi.nlm.nih.gov/snp/) with aggregate (ALFA) allele frequency which showed G = 1.00000 (reference allele), T = 0.00000 (alternate allele) in global sample size of 10680 population. The South Asian sample size of 94 showed G = 1.00 (reference allele) and T = 0.00 (alternate allele).

The nsSNPs *rs539681368*, *rs542858994* and *rs539681368* were predicted in α-helix, random coil and extended strand (see Table 4). The most pathogenic nsSNP i.e., *rs539681368:C30Y* (NNAT-204) identified through the application of above *in silico* tools have 53% buried residues with 60% α-helix proportion. Owing to the significant difference between wild type (cysteine) and the mutant (tyrosine) amino acid, this mutation leads to decrement in the hydrophobic character with bigger tyrosine residues. Hence hydrophobic interactions, either in the core of the protein or on the surface, will be lost (see Table 4).

MSA, homology modeling, structural validations, for WT and pathogenic (rs539681368:C30Y) **variant of human NNAT gene**

MSA of NNAT gene resulted in comparison of human NNAT gene with 8 other organisms. *Microtus ochrogaster* and *Rhinopithecus bieti* were the genes that are most similar to the NNAT WT (identified gap = 2.5%). However, *Microtus ochrogaster* is a low quality protein and in *Eptesicus fuscus* the largest gap of 58% was observed (see Fig. 4).

The transcript ID (NNAT-204: *ENST00000649451.1*) of rare pathogenic variant of NNAT (C30Y) was validated from ENSEMBL database which showed gold standard (reviewed and high standard) transcript. The transcript NNAT-204 comprised 81 amino acids (1259 base pairs), and *RefSeq* match as NM_005386.4. The QUARK algorithm has been effectively applied to model entire sequence (81 residues WT), and has performed well in Critical Assessment of protein Structure Prediction (CASP) blind experiments with five constructed models. Through 3D sterochemical analysis and validation of all predicted structures, model rank 2 of NNAT had the highest quality factor *i.e.*, 100% and ~ 93% of residues lied in the most favored region. Only 7.1% residues were in additional allowed region whereas no residues were found in the generously allowed or disallowed region of Ramachandran plot. The secondary structure prediction of NNAT protein identified 3 α- helices. The non-glycine and non-proline residues were 70. The residual properties showed that maximum deviation observed was 13.6, the bond length/angle = 4.7, the bond contact was 1. The G-factor analyses showed that dihedrals score was − 0.39, covalent were 0.02 with overall score of -0.20. Planar group of 3D model showed that 87.1% were within the limit and 12.9% were highlighted ones. The trans-membrane region of NNAT ranged from Ile13 to Gly35, which means that our targeted mutation C30Y resides in the trans-membrane region. The low complexity region (LCR) is starting from Ala2 and ending on Ala9 (see Fig. 4).

The 3D NNAT homology model (WT) obtained through *ab initio* modeling was done with residue specific substitution of cysteine to tyrosine at 30th position. The overall quality factor for variant (C30Y) model was 90.411, of which 92.9% residues were in the most favorable region, 7.1% were in additional allowed region and no residue was found in generously and disallowed region of Ramachandran plot (see Fig. 5). The residual properties showed that maximum deviation bond length/angle remains same meanwhile, the bond contact was increased. The G-factor and Planar group analysis also remains the same as that of WT 3D model. The mutation C30Y of NNAT 3D protein structure resulted in an overall destabilization of protein structure environment (see Table 5). The torsion angles were unfavorable and the predicted with the aid of free energy difference of unfolding between WT and mutant 3D protein (-2.77 kcal/mol).

AN compounds, drug-likeness assessment and in silico **exploration of NNAT's druggable protein cavities**

The six 3D compounds used to manage AN manifesting Amitriptyline (PUBCHEM ID: 2160), Citalopram (PUBCHEM ID: 2771), Desipramine (PUBCHEM ID: 2995), Fluoxetine (PUBCHEM ID: 3386), Lysine (PUBCHEM ID: 5962) and Mirtazapine (PUBCHEM ID: 4205). Their 3D structure was obtained from PUBCHEM database. Application of rule for drug likeness resulted 'Lysine' 3D compound that was only excluded. The computed clog*P* value, drug-likeness and drug score for lysine was much lowers compared to rest of the five drugs (see Table 6).

Due to the lack of the availability of the active site / interacting residues information, CB-docking procedure was applied on NNAT WT model; the rationale was to investigate the best possible residue-interaction. Binding energies according to highest cavity grading was obtained as; Amitriptyline = -5.7 kcal/mol, Citalopram = -4.9 kcal/mol, Desipramine = -5.1 kcal/mol, Flouxetine = -5. kcal/mol and Mirtazipine = -4.9 kcal/mol (see Fig. 6). Identification of residues with commonest appearance having lowest binding scores as well as lowest RMSD values for interacted targets were considered for CBR.

Molecular docking of 3D homology model of pathogenic variant (rs539681368:C30Y **)**

The molecular docking of pathogenic variant (*rs539681368:C30Y*) of NNAT with all 5 compounds of AN exhibited lower free energy (ΔG) values encompassing Amitriptyline = -4.46 kcal/mol, Citalopram = -4.1 kcal/mol, Desipramine = -4.82 kcal/mol, Flouxetine = -4.54 kcal/mol and Mirtazipine = -4.39

kcal/mol (see Fig. 7, and Table 7). The inhibition constant (K_i) values and lower intermolecular binding energies further elucidated a higher *in silico* affinity of (NNAT- rs539681368:C30Y) the variant's binding with selected AN therapeutic 3D compounds (see Table 7).

Discussion

The gene NNAT in humans consists of two introns and three exons. Its genomic structure (mRNA sequence) specifies its expression as two mRNA species (generated by alternative splicing) *i.e.*, one isoform is α, the other one is β [47]. The α-isoform which we used in our current work contains all three exons, and its mRNA encodes a protein of 81 amino acids by the aid of all three exons, whereas, the β-isoform of NNAT mRNA encodes a protein consisting of 54 amino acids through only first and third exons [47]. NNAT is a paternally expressed gene, and its absence correlates with postnatal growth restriction which is suggestive NNAT's role as a developmental protein [48]. Therefore, the reported high expression of NNAT in human brain tissues is associated with its role as a specific integral protein for fetal neural cell development, terminal brain differentiation, and pituitary development [2].

Exploring the tissue specific gene expression mechanism(s) provides the elaborated basis to understand how tissues are distinguished by gene expression patterns and implying their significant regulatory role. The preferred function and expression of particular gene in one or more several tissues (or cells) types set better understanding of genes functionality or tissue - gene relationship, its etiology and discovery of novel tissue-specific drug targets that are especially studied to manage the gene-associated diseases [49].

The main reason for NNAT gene selection and its further investigation through our *in silico* study was based on its higher gene expression pattern in brain tissue, in comparison to other tissues [11, 12]. Predominant expression of NNAT α-isoform during an initial stage of brain development was well reported, in comparison to NNAT β-isoform [2] while the primary adipocytes express only the α-isoform of NNAT gene [2]. Additionally, NNAT has also been expressed in various other tissues which are particularly involved in metabolic energetic homeostasis of body [2, 50]. Especially in human brain, its mRNA is present in the hypothalamus and pituitary regions, whereas, peripheral tissue regions to the brain, NNAT has its expression including in tissues of thyroid and pancreas, with coordinating role in energetic homeostasis [2, 47, 48, 50]. Studies also reported that NNAT is normally expressed in islet of pancreas β - cells [2, 50] and its abnormal expression (NNAT β-isoform) is related with the pancreatic β-cells destruction [2, 47, 48, 50]. The accumulation of great amount of NNAT β isoform was reported as misfolded ubiquitin-positive aggregate protein structures which contributes to the damage of pancreas β - cells and possibly implicated in diabetes mellitus [2, 47]. On the contrary, loss of NNAT caused a decrease in the basal and insulin-stimulated glucose uptake and glycogen synthesis processes [2]. Hence, in AN, it was also suggested that the ratio of NNAT β-isoform to NNAT α-isoform was affected, explaining the unexpected resistance to hypoglycemia in AN patients [2].

Researchers reported that AN runs in family in relation with genetic factors. *Lombardi et al* in 2019 and Ceccarrini *et al.*, 2021 sequenced some families (with AN) and found significant association of NNAT (neuronatin) gene variants (single nucleotide polymorphism - SNP) in an identified population with AN [2, 10]. *Lombardi et al.*, suggested that the constitutional reduced level of NNAT-α isoform in brain and in adipocytes from inclusive study subjects may predispose to AN [2]. Previously published findings in humans and in other organisms strengthen the notion that NNAT expression changes or its gene variants may be associated with having susceptibility to EDs like AN [10]. Past researches suggested that NNAT could be considered as a potential therapeutic target for the treatment of human obesity and our data extend this suggestion to AN [2]. The data from tissue specific expression and earlier findings suggested that gene NNAT was involved in adiposity, metabolic homoeostasis and pancreatic mechanism which might dispose to AN [2, 10, 47, 50]. Therefore, our study's rationale was to investigate transcript level prediction of the rare-coding deleterious nsSNPs of NNAT gene and mutational impact on protein stability. Furthermore, determination of the structural inference of rare-coding nsSNPs of NNAT gene and to estimating the *in silico* molecular binding energies of 3D homology model of most pathogenic variant(s) with therapeutic compounds of AN was also intended to propose its role in AN.

SNPs are the most common and simplest type to represent genomic variation among individual DNAs, responsible for the diversity in genome evolution and serve as the basis to understand devastating role of disease that have governing impact from gene to phenotype [51]. Many GWAS (Genome wide association studies) have recognized number of SNPs that increase the risk of cancer like breast cancer, colorectal cancer, leukemia among others [52]. Most of the pathogenic SNPs have been reported to modify the secondary structure, influence promoter activity, influentially impact the expression levels, and subcellular localization of mRNAs / affecting the protein regulatory function, thereby contributing to the progression of disease [51, 53]. Generally, mutations (SNPs) are neutral, but few of them have pathogenic predisposition by altering the regular function of protein [53]. In different human diseases, nsSNPs constitute about half of all genetic variation which can impact either neutral or deleterious [54]. To investigate the disease-causing mutation(s) as well as studying their potential role as genetic markers is a popular science and researchers believe that finding these genomic variants may affect the response to treatment [51–54]. Due to increase frequency of SNPs and non-redundant databases, recently it has become challenging to determine each SNPs and its substantial contribution in disease development. From a huge set of mutations, the role of computational analysis to mark off or prioritize pathogenic SNPs for genetic disease screening has its own significance [22–29]. *In silico* analyses of SNPs is always considered as a cost-effective and feasible option to investigate or to predict the status in terms of pathogenicity or relating its high risk impact as non-synonymous mutation through the available various bioinformatics tools [22].

Therefore, we aimed to investigate all rare-coding nsSNPs in the human NNAT gene and predict their transcript level effects on the basis of functional impact analyses, computation of changes in free energies due to nsSNPs (see Table 1–3), PTMs and their structural inference in relation to function, stability and regulation of its respective protein (see Fig. 4, and Table 4). Initially we have 840 exonic mutations from record; the filtration strategies reduced this count to 19 which were only rare-coding mutations. This count was reduced to 10 when segregated on the basis of missense mutation type (see Fig. 3). The distribution of these ten nsSNPs was probed which showed three nsSNPs *i.e., rs539681368*, *rs542858994*, *rs560845323* as rare-coding nsSNP, distributed in five NNAT transcripts sequences (NNAT201-205), resulting total count of 10 transcript level nsSNP (see Table 1). The transcript based functional impact analyses of nsSNPs by the aid of various *in silico* mentioned tools such as PROVEAN, PhD-SNP, SNAP2, SNP&GO, PMut, MutPred2 and using consensus classifier for predicting high risk pathogenic mutations computed the nsSNP '*rs539681368*' (C30Y: NNAT-204) as high risk pathogenic mutation (see Table 1–

2). The computation (through MUpro, I-Mutant, iPTree, INPS) of mutation '*rs539681368'* (C30Y: NNAT-204) in terms of their impact on protein stability showed the change in protein function (see Table 3).

We did *ab initio* structure prediction of NNAT α-isoform (this isoform sequence is the NNAT-204 transcript sequence) and afterwards we generated an alternate 3D model for *rs539681368'* (C30Y: NNAT-204) mutation (see Fig. 5). The identification of active sites or functional interacting residues in protein structure actually described the critical and conserved functional areas of proteins, which can serve as a decisive step in predicting protein activity. When no information of crystal 3D structure of protein is available the selection and identification of active consensus residues for further investigating the role of protein in term of ligand / drug binding, is an important way to proceed. To resolve this, we used CB dock approach which computationally detects the binding sites, and determines the center and size of the druggable cavities in the NNAT protein (see Fig. 6).

In structural biology and computer-assisted drug design, molecular docking is an important tool to identify the best fit orientation of drug / ligand binding with the particular protein targets [55]. The purpose of ligand-protein docking is to anticipate a ligand's most common binding mode(s) with a protein of a known 3D structure. Virtual screening, binding affinity, free energy binding estimates, as well as drawing out and visualizing various forms of bonds and non-bonded interactions between the ligand and amino acid residues of a protein, are all done using molecular docking [55]. To examine our second objective we selected and screened five drugs commonly used for the managing AN (see Table 6). The tricyclic antidepressant 'amitriptyline' is used to treat depression, both endogenous and psychotic, as well as anxiety linked with depression. The actual mechanism is not completely understood. The proposed role of amitriptyline is thought to prevent the membrane pump mechanism responsible for the re-uptake of amines like norepinephrine and serotonin, increasing their concentration in synaptic clefts. One of the oldest hypotheses in depression is that the deficiency of serotonin (5-HT) and/or norepinephrine (NE) neurotransmission in the brain causes depressive symptoms hence, amitriptyline inhibits these pathways, which may be the mechanism by which it alleviates depressed symptoms especially in AN [56]. Another tricyclic antidepressant 'Desipramine' used to treat depression, as it is referred as a choice of medication that selectively inhibits / blocks the re-uptake of norepinephrine (noradrenaline) neural synapse and inhibits serotonin reuptake [57]. In addition, a tetracyclic antidepressant, 'Mirtazapine' is used for treating of major depressive disorder (MDD) and is sometimes prescribed as off-label drug for appetite stimulation. The suggested mechanism of action may be explained by its rapid onset of action, a high degree of responsiveness, a lesser side-effect profile and it also has effects on central adrenergic and serotonergic activity that distinguish it from other antidepressants [58]. The antidepressant 'Citalopram' used in AN, belong to the class of drugs which is referred as SSRI (selective serotonin reuptake inhibitor) usually prescribed to treat depression. Citalopram's mechanism of action is based on the suppression of CNS neuronal re-uptake of serotonin (5-HT), and prevents serotonin absorption in the synaptic cleft by blocking the serotonin transporter (solute carrier family 6 member 4, *SLC6A4*) [59]. Fluoxetine is another SSRI used in managing severe depressive disorder. It decreases the presynaptic re-uptake of the 'serotonin', resulting in the rise of the level of 5-hydroxytryptamine (5-HT) levels in different regions of the brain [60]. The binding of these five drugs with the deleterious variant (*rs539681368:C30Y*) of NNAT gene were estimated in terms of lower free energy values (see Fig. 7, and Table 7). Additionally, in protein-ligand complexes, when hydrogen atom connected to a strongly electronegative atom is in the neighborhood of another electronegative atom with a lone pair of electrons, it creates a specific dipole-dipole attraction called a hydrogen bond [55, 61]. In proteins, hydrogen bonding in secondary structure and in tertiary conformation was observed as a characteristics structural entity which serves as a stabilizing force. Interactions (mostly hydrogen bonds) between surrounding polypeptide backbones that contain Nitrogen-Hydrogen bonded pairs and oxygen atoms make up a protein's secondary structure [61]. Due to the strong electronegative nature of both N and O, hydrogen atoms bound to nitrogen in one polypeptide backbone can hydrogen bond to oxygen atoms in another chain, and vice versa [61]. Docking guided analysis of AN drugs' binding energies indicated with lowest binding free energy (ΔG), inhibition constant (K_i) and lower intermolecular energies with mutant model, strengthened that NNAT variants (*rs539681368:C30Y*) significantly interact with AN drugs (see Fig. 7). The all five drugs showed commonest interaction at K59, which was also characterized as the putative PTM site for acetylation; existed in the cytoplasmic domain of gene (see Fig. 4). Therefore, there is a probability that these drugs binds with the identified variant of NNAT significantly which results interference in the binding of these AN therapeutic compounds with their respective target proteins, leads to unmanaged status of AN. However, in vivo studies are still needed to support this propose mechanism.

Strength And Limits

We evaluated human NNAT gene in terms of the identification of the most pathogenic rare-coding nsSNP and their impact on protein stability. In a total of 840 exonic SNPs, three (*rs539681368*, *rs542858994*, *rs560845323*) were rare-coding nsSNP, which were further distributed with five transcripts sequences (constituting a total count of 10 nsSNPs). Our computational findings categorized '*rs539681368* (C30Y: NNAT-204)' variants as the most pathogenic mutation so far. The 3D structural analysis of *rs539681368:C30Y* (NNAT-204) revealed decrease in stability, quality factor, with unfavorable torsion angles as well as the predicted free energy difference of unfolding between WT and mutant proteins. The molecular docking based energetic estimates of 3D deleterious variant (*rs539681368:C30Y*, NNAT-204) of NNAT protein with AN drugs showed lower energy values. Thus, we concluded that this variant may possibly interfere in the binding of mentioned AN therapeutic 3D compounds with their respective target proteins, resulting in the unmanaged status of AN. The identified pathogenic nsSNP could be an important candidate, underlying various pathological processes associated with NNAT pathophysiology. The main strength of this work is to presents a rich array of opportunities to study the gene structural variant relationship which will assist in expediting the discovery of new genomic targets, and ultimately lead to novel compounds with predicted biological activity to manage AN challenge.

The recent unavailability of crystal structures, protein flexibility, and molecular confirmation for NNAT could serve as the possible study limitation and has considered an obstacle in predicting the reliable mechanism for AN pathogenesis.

Declarations

Author contributions MBA, UN, SAQ and MKA: conceived the study. UN, AS, AJ and HU: performed the literature search. UN, AS, AJ, HU and MBA: analyzed data. All authors contributed to interpretation of data, as well as drafting and approval of the manuscript.

Funding None

Data availability The data used in the article are given with the information from where the data were taken.

Declarations

Conflict of interest None.

Ethical approval Not applicable

Consent to participate Not applicable.

Consent for publication Not applicable.

References

1. Wu J, Liu J, Li S, Ma H, Wang Y (2020) Trends in the prevalence and disability-adjusted life years of eating disorders from 1990 to 2017: results from the Global Burden of Disease Study 2017. Epidemiol Psychiatr Sci 29:e191. doi: 10.1017/S2045796020001055
2. Lombardi L, Blanchet C, Poirier K, Lebrun N, Ramoz N, Moro MR, Gorwood P, Bienvenu T (2019) Anorexia nervosa is associated with Neuronatin variants. Psychiatr Genet 29:103–110. doi: 10.1097/YPG.0000000000000224
3. Batista M, Žigić Antić L, Žaja O, Jakovina T, Begovac I (2018) Predictors of eating disorder risk in anorexia nervosa adolescents. Acta Clin Croat 57:399–410. doi: 10.20471/acc.2018.57.03.01
4. Kwok C, Kwok V, Lee HY, Tan SM (2020) Clinical and socio-demographic features in childhood vs adolescent-onset anorexia nervosa in an Asian population. Eat Weight Disord 25:821–826. doi: 10.1007/s40519-019-00694-9
5. Mumford DB, Whitehouse AM, Platts M (1991) Sociocultural correlates of eating disorders among Asian schoolgirls in Bradford. Br J Psychiatry 158:222–228. doi: 10.1192/bjp.158.2.222.. doi: 10.1192/bjp.158.2.222
6. Tareen A, Hodes M, Rangel L (2005) Non-fat–phobic anorexia nervosa in British South Asian adolescents. Int J Eat Disord 37:161–165. doi: 10.1002/eat.20080
7. Babar N, Alam M, Ali SS, Ansari A, Atiq M, Awais A, Amin F, Israr SM (2002) Anorexic behaviour and attitudes among female medical and nursing students at a private university hospital. J Pak Med Assoc 52:272–276
8. Hisam A, Rahman MU, Mashhadi SF (2015) Anorexia nervosa among teenage girls: Emerging or prevalent? Pak J Med Sci 31:1290–1294. doi: 10.12669/pjms.316.7617
9. Vrang N, Meyre D, Froguel P, Jelsing J, Tang-Christensen M, Vatin V, Mikkelsen JD, Thirstrup K, Larsen LK, Cullberg KB, Fahrenkrug J (2010) The imprinted gene neuronatin is regulated by metabolic status and associated with obesity. Obesity 18:1289–1296. doi: 10.1038/oby.2009.361
10. Ceccarini MR, Precone V, Manara E, Paolacci S, Maltese PE, Benfatti V, Dhuli K, Donato K, Guerri G, Marceddu G, Chiurazzi P (2021) A next generation sequencing gene panel for use in the diagnosis of anorexia nervosa. Eat Weight Disord 25:1–12. doi: 10.1007/s40519-021-01331-0
11. Fagerberg L, Hallström BM, Oksvold P, Kampf C, Djureinovic D, Odeberg J, Habuka M, Tahmasebpoor S, Danielsson A, Edlund K, Asplund A (2014) Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. Mol Cell Proteomics 13:397–406. doi: 10.1074/mcp.M113.035600
12. GTEx Consortium (2020) The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science 369:1318–1330. doi: 10.1126/science.aaz1776
13. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, Franz M, Grouios C, Kazi F, Lopes CT, Maitland A (2010) The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res 38:W214–W220. doi: 10.1093/nar/gkq537
14. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ (2019) STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res 47:D607–D613
15. Hunt SE, Moore B, Amode RM, Armean IM, Lemos D, Mushtaq A, Parton A, Schuilenburg H, Szpak M, Thormann A, Perry E (2021) Annotating and prioritizing genomic variants using the Ensembl Variant Effect Predictor—A tutorial. Hum Mutat 1–12. doi: 10.1002/humu.24298
16. Linck E, Battey CJ (2019) Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. Mol Ecol Resour 19:639–647. doi: 10.1111/1755-0998.12995
17. Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC (2016) SIFT missense predictions for genomes. Nat Protoc 11:1–9. doi: 10.1038/nprot.2015.123
18. Adzhubei I, Jordan DM, Sunyaev SR (2013) Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet 76:7–20. doi: 10.1002/0471142905.hg0720s76
19. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M (2019) CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res 47:D886–894. doi: 10.1093/nar/gky1016
20. Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, Musolf A, Li Q, Holzinger E, Karyadi D, Cannon-Albright LA (2016) REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. Am J Hum Genet 99:877–885. doi: 10.1016/j.ajhg.2016.08.016
21. Cunningham F, Achuthan P, Akanni W, Allen J, Amode MR, Armean IM, Bennett R, Bhai J, Billis K, Boddu S, Cummins C 2019 Ensembl 2019.Nucleic Acids Res47:D745-D751. doi: 10.1093/nar/gky1113

22. Frousios K, Iliopoulos CS, Schlitt T, Simpson MA (2013) Predicting the functional consequences of non-synonymous DNA sequence variants—evaluation of bioinformatics tools and development of a consensus strategy. Genomics 102:223–228. doi: 10.1016/j.ygeno.2013.06.005
23. Choi Y, Chan AP (2015) PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics 31:2745–2747. doi: 10.1093/bioinformatics/btv195
24. Capriotti E, Fariselli P (2017) PhD-SNPg: a webserver and lightweight tool for scoring single nucleotide variants. Nucleic Acids Res 45:W247–W252. doi: 10.1093/nar/gkx369
25. Hecht M, Bromberg Y, Rost B (2015) Better prediction of functional effects for sequence variants. BMC Genomics 16:1–2. doi: 10.1186/1471-2164-16-S8-S1
26. Capriotti E, Martelli PL, Fariselli P, Casadio R (2017) Blind prediction of deleterious amino acid variations with SNPs&GO. Hum Mutat 38:1064–1071. doi: 10.1002/humu.23179
27. López-Ferrando V, Gazzo A, De La Cruz X, Orozco M, Gelpí JL (2017) PMut: a web-based tool for the annotation of pathological variants on proteins, 2017. update Nucleic Acids Res 45:W222–W228. doi: 10.1093/nar/gkx313
28. Pejaver V, Urresti J, Lugo-Martinez J, Pagel KA, Lin GN, Nam HJ, Mort M, Cooper DN, Sebat J, Iakoucheva LM, Mooney SD (2020) Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. Nat Commun 11:1–13. doi: 10.1038/s41467-020-19669-x
29. Bendl J, Stourac J, Salanda O, Pavelka A, Wieben ED, Zendulka J, Brezovsky J, Damborsky J (2014) PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations. PLoS Comput Biol 10:e1003440. doi: 10.1371/journal.pcbi.1003440
30. Khan S, Vihinen M (2010) Performance of protein stability predictors. Hum Mutat 31:675–684. doi: 10.1002/humu.21242
31. Capriotti E, Calabrese R, Casadio R (2006) Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. Bioinformatics 22:2729–2734. doi: 10.1093/bioinformatics/btl423
32. Chen CW, Lin MH, Liao CC, Chang HP, Chu YW (2020) iStable 2.0: predicting protein thermal stability changes by integrating various characteristic modules. Comput Struct Biotechnol J 18:622–630. doi: 10.1016/j.csbj.2020.02.021
33. Savojardo C, Fariselli P, Martelli PL, Casadio R (2016) INPS-MD: a web server to predict stability of protein variants from sequence and structure. Bioinformatics 32:2542–2544. doi: 10.1093/bioinformatics/btw192
34. Angamuthu K, Piramanayagam S (2017) Evaluation of in silico protein secondary structure prediction methods by employing statistical techniques. Biomed Biotech Res J 1:29–36. doi: 10.4103/bbrj.bbrj_28_17
35. Dunlavy DM, O'leary DP, Klimov D, Thirumalai D (2005) HOPE: A homotopy optimization method for protein structure prediction. J Comput Biol 12:1275–1288. doi: 10.1089/cmb.2005.12.1275
36. Yachdav G, Kloppmann E, Kajan L, Hecht M, Goldberg T, Hamp T, Hönigschmid P, Schafferhans A, Roos M, Bernhofer M, Richter L (2014) PredictProtein—an open resource for online prediction of protein structural and functional features. Nucleic Acids Res 42:W337–W343. doi: 10.1093/nar/gku366
37. Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, Chang HY, Dosztányi Z, El-Gebali S, Fraser M, Gough J (2017) InterPro in 2017—beyond protein family and domain annotations. Nucleic Acids Res 45:D190–D199. doi: 10.1093/nar/gkw1107
38. Wang D, Liu D, Yuchi J, He F, Jiang Y, Cai S, Li J, Xu D (2020) MusiteDeep: a deep-learning based webserver for protein post-translational modification site prediction and visualization. Nucleic Acids Res 48:W140–W146. doi: 10.1093/nar/gkaa275
39. Velankar S, Burley SK, Kurisu G, Hoch JC, Markley JL (2021) The protein data bank archive. Methods Mol Biol 2305:3–21. doi: 10.1007/978-1-0716-1406-8_1
40. Mortuza SM, Zheng W, Zhang C, Li Y, Pearce R, Zhang Y (2021) Improving fragment-based ab initio protein structure assembly using low-accuracy contact-map predictions. Nat Commun 12:1–12. doi: 10.1038/s41467-021-25316-w
41. Nooranian S, Kazemi Oskuee R, Jalili A (2022) Characterization and Evaluation of Cell-Penetrating Activity of Brevinin-2R: An Amphibian Skin Antimicrobial Peptide. Mol Biotechnol 11:1–14. doi: 10.1007/s12033-021-00433-5
42. Huang CC, Meng EC, Morris JH, Pettersen EF, Ferrin TE (2014) Enhancing UCSF Chimera through web services. Nucleic Acids Res 42:W478–W484. doi: 10.1093/nar/gku377
43. Parthiban V, Gromiha MM, Schomburg D (2006) CUPSAT: prediction of protein stability upon point mutations. Nucleic Acids Res 34:W239–W242. doi: 10.1093/nar/gkl190
44. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, Sajed T, Johnson D, Li C, Sayeeda Z, Assempour N (2018) DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res 46:D1074–D1082. doi: 10.1093/nar/gkx1037
45. Liu Y, Grimm M, Dai WT, Hou MC, Xiao ZX, Cao Y (2020) CB-Dock: a web server for cavity detection-guided protein–ligand blind docking. Acta Pharmacol Sin 41:138–144. doi: 10.1038/s41401-019-0228-6
46. Nguyen NT, Nguyen TH, Pham TN, Huy NT, Bay MV, Pham MQ, Nam PC, Vu VV, Ngo ST (2020) Autodock vina adopts more accurate binding poses but autodock4 forms better binding affinity. J Chem Inf Model 60:204–211. doi: 10.1021/acs.jcim.9b00778
47. Joseph RM (2014) Neuronatin gene: Imprinted and misfolded: Studies in Lafora disease, diabetes and cancer may implicate NNAT-aggregates as a common downstream participant in neuronal loss. Genomics 103:183–188. doi: 10.1016/j.ygeno.2013.12.001
48. Millership SJ, Tunster SJ, Van de Pette M, Choudhury AI, Irvine EE, Christian M, Fisher AG, John RM, Scott J, Withers DJ (2018) Neuronatin deletion causes postnatal growth restriction and adult obesity in 129S2/Sv mice. Mol Metab 18:97–106. doi: 10.1016/j.molmet.2018.09.001
49. Somepalli G, Sahoo S, Singh A, Hannenhalli S (2021) Prioritizing and characterizing functionally relevant genes across human tissues. PLoS Comput Biol 17(7):e1009194. doi: 10.1371/journal.pcbi.1009194

50. Millership SJ, Xavier GD, Choudhury AI, Bertazzo S, Chabosseau P, Pedroni SM, Irvine EE, Montoya A, Faull P, Taylor WR, Kerr-Conte J (2018) Neuronatin regulates pancreatic β cell insulin content and secretion. J Cin Invest 128:3369–3381. doi: 10.1172/JCI120115
51. Guo X, Song Y, Liu S, Gao M, Qi Y, Shang X (2021) Linking genotype to phenotype in multi-omics data of small sample. BMC Genomics 22:1–11. doi: 10.1186/s12864-021-07867-w
52. Sud A, Kinnersley B, Houlston RS (2017) Genome-wide association studies of cancer: current insights and future perspectives. Nat Rev Cancer 17:692–704. doi: 10.1038/nrc.2017.82
53. Studer RA, Dessailly BH, Orengo CA (2013) Residue mutations and their impact on protein structure and function: detecting beneficial and pathogenic changes. Biochem J 449:581–594. doi: 10.1042/BJ20121221
54. Ancien F, Pucci F, Godfroid M, Rooman M (2018) Prediction and interpretation of deleterious coding variants in terms of protein structural stability. Sci Rep 8:1–11. doi: 10.1038/s41598-018-22531-2
55. Pinzi L, Rastelli G (2019) Molecular docking: shifting paradigms in drug discovery. Int J Mol Sci 20:4331. doi: 10.3390/ijms20184331
56. Gupta SK, Shah JC, Hwang SS (1999) Pharmacokinetic and pharmacodynamic characterization of OROS and immediate-release amitriptyline. Br J Clin Pharmacol 48:71–78. doi: 10.1046/j.1365-2125.1999.00973.x
57. Souery D, Serretti A, Calati R, Oswald P, Massat I, Konstantinidis A, Linotte S, Kasper S, Montgomery S, Zohar J, Mendlewicz J (2011) Citalopram versus desipramine in treatment resistant depression: Effect of continuation or switching strategies. A randomized open study. World J Biol Psychiatry 12:364–375. doi: 10.3109/15622975.2011.590225
58. San L, Arranz B (2006) Mirtazapine: only for depression? Acta Neuropsychiatr 18:130–143. doi: 10.1111/j.1601-5215.2006.00143.x
59. Sangkuhl K, Klein TE, Altman RB (2011) PharmGKB summary: citalopram pharmacokinetics pathway. Pharmacogenet Genomics 21:769–772. doi: 10.1097/FPC.0b013e328346063f
60. Sommi RW, Crismon ML, Bowden CL (1987) Fluoxetine: a serotonin-specific, second-generation antidepressant. Pharmacotherapy 7:1–15. doi: 10.1002/j.1875-9114.1987.tb03496.x
61. Azmi MB, Sultana S, Naeem S, Qureshi SA (2021) In silico investigation on alkaloids of Rauwolfia serpentina as potential inhibitors of 3-hydroxy-3-methyl-glutaryl-CoA reductase. Saudi J Biol Sci 28:731–737. doi: 10.1016/j.sjbs.2020.10.066

Tables

Table 1 Transcript based functional impact analyses of rare-coding missesne nsSNPs of human NNAT gene

Transcripts	Variant:AA	Chromosome: base pair	PROVEAN score	PhD-SNP	SNAP2	SNPs&GO	PMut	MutPred2
NNAT-201	rs542858994:F26L	20:37522672	-5.345	0.901(disease)	78 (85%)	0.482 (neutral)	0.33	0.356
NNAT-202	rs539681368:C30Y	20:37522374	-7.15	0.971 (disease)	81 (91%)	0.859 (disease)	0.61	0.361
NNAT-202	rs542858994:F35L	20:37522672	-3.573	0.915 (disease)	50 (75%)	0.528 (disease)	0.52	0.238
NNAT-203	rs539681368:C30Y	20:37522374	-7.5	0.971 (disease)	80 (91%)	0.750 (disease)	0.54	0.215
NNAT-203	rs542858994:Q52E	20:37522672	-0.036	0.480 (neutral)	44 (71%)	0.065 (neutral)	0.42	0.128
NNAT-203	rs560845323:R99Q	20:37522814		0.724 (disease)	37 (66%)	0.131 (neutral)	0.24	0.1
NNAT-203	rs560845323:R99P	20:37522814		0.606 (disease)	45 (71%)	0.081 (neutral)	0.31	0.341
NNAT-204	rs539681368:C30Y	20:37522374	-7.15	0.971 (disease)	85 (91%)	0.875 (disease)	0.53	0.884
NNAT-204	rs542858994:F53L	20:37522672	-3.823	0.705 (disease)	52 (75%)	0.266 (neutral)	0.16	0.321
NNAT-205	rs542858994:S35L	20:37522672	-3.083	0.506 (disease)	43 (71%)	0.025 (neutral)	0.52	0.166

Table 2 Transcript based percent expected accuracy in silicoalgorithms through consensus classifier prediction of the rare-coding nsSNPs of NNAT gene

Transcripts	Variant:AA	Chromosome: base pair	PredictSNP	PhD-SNP	PolyPhen-1	PolyPhen-2	SIFT	SNAP	PANTHER	MAPP
NNAT-201	rs542858994:F26L	20:37522672	61*	86*	-	-	68^	56*	64^	-
NNAT-202	rs539681368:C30Y	20:37522374	87*	88*	59*	45*	79*	89*	48^	-
NNAT-202	rs542858994:F35L	20:37522672	60^	82*	67^	87^	71^	81*	64^	-
NNAT-203	rs539681368:C30Y	20:37522374	55*	88*	59*	40*	76^	85*	48^	71^
NNAT-203	rs542858994:Q52E	20:37522672	75^	78^	67^	61^	76^	55^	-	84*
NNAT-203	rs560845323:R99Q	20:37522814	60^	83^	67^	43*	74^	85*	-	66*
NNAT-203	rs560845323:R99P	20:37522814	61*	72^	59*	47*	68^	85*	-	86*
NNAT-204	rs539681368:C30Y	20:37522374	87*	88*	74*	54*	79*	87*	48^	77*
NNAT-204	rs542858994:F53L	20:37522672	60^	58*	67^	87^	68^	72*	64^	57*
NNAT-205	rs542858994:S35L	20:37522672	60^	58^	67^	41*	79^	62*	-	77*

* indicates the deletrious prediction of mutation, ^ indicates the neutral prediction of mutation

Table 3 Transcript based computation of changes in free energy of amino acids variations of rare-coding nsSNPs of human NNAT gene

Transcripts	Variant:AA	MUpro ΔΔG (kcal/mol)	MUpro_SVM ΔΔG (kcal/mol)	MUpro_NN ΔΔG (kcal/mol)	I-Mutant ΔΔG (kcal/mol)	SVM2 : SVM3	iPTREE-STAB ΔΔG (kcal/mol)	INPS ΔΔG (kcal/mol)
NNAT-201	rs542858994:F26L	-0.64878677	0.389	0.786	-1.37	Decrease : Large Decrease	-2.713	-1.56971
NNAT-202	rs539681368:C30Y	-0.83286733	0.1576	-0.6122	-0.24	Decrease : Decrease	1.9	-1.91537
NNAT-202	rs542858994:F35L	-0.53540105	-0.1783	0.7028	-1.23	Decrease : Large Decrease	-2.7133	-1.99448
NNAT-203	rs539681368:C30Y	-0.83286733	0.1576	-0.6122	-0.17	Decrease : Decrease	1.9	-1.68446
NNAT-203	rs542858994:Q52E	-0.3898909	-0.1501	0.5879	-0.04	Decrease : Neutral	1.3467	-0.645429
NNAT-203	rs560845323:R99Q	-0.5479955	-0.838	-0.755	-0.95	Decrease : Large Decrease	0.74	-0.529108
NNAT-203	rs560845323:R99P	-0.88545867	-0.3753	0.562	-0.57	Decrease : Large Decrease	0.74	-0.781829
NNAT-204	rs539681368:C30Y	-0.83286733	0.1576	-0.6122	-0.17	Decrease : Decrease	1.9	-1.86067
NNAT-204	rs542858994:F53L	-0.38258542	0.098	0.5923	-1.06	Decrease : Large Decrease	-2.7133	-1.91724
NNAT-205	rs542858994:S35L	-0.01027647	0.534	0.861	-0.21	Decrease : Decrease	-1.5486	-0.716182

SVM = Support Vector Machine. SVM3= Large Decrease (values less than -0.5); SVM2 = Decrease (Values in negative)

Table 4 Structural inference of rare-coding nsSNPs of human NNAT gene

Transcripts	Variant:AA	SOPMA Predicting Secondary Structure	Secondary Structural Information	Surface Accessibility	Change of charge	Hydrophobicity	HOPE Inference
NNAT-201	rs542858994:F26L	Alpha-helix	Alpha-helix = 72.2%, Coil = 25.9% & Beta-strand = 1.9%	Burried = 51.9% & Exposed = 48.1%			The mutant residue is smaller, this might lead to loss of interactions. The mutant residue is located near a highly conserved position.
NNAT-202	rs539681368:C30Y	Random coil	Alpha-helix = 79.4% & Coil = 20.6%	Burried = 55.6% & Exposed = 44.4%		Decreases	The mutant residue is bigger than the wild-type residue and located near a highly conserved position, this might lead to bumps. The wild-type residue is more hydrophobic than the mutant residue. The residue is located in a region annotated in the UniProt database as a transmembrane domain. This size difference can affect the contacts with the lipid-membrane. This differences in hydrophobicity can affect the hydrophobic interactions with the membrane lipids.
NNAT-202	rs542858994:F35L	Alpha-helix					The mutant residue is smaller than the wild-type residue. The residue is located in a region annotated in the UniProt database as a transmembrane domain. The mutant residue is smaller than the wild-type residue this might lead to loss of interactions.. This size difference can affect the contacts with the lipid-membrane.
NNAT-203	rs539681368:C30Y	Extended-strand	Alpha-helix = 27.2%, Coil = 65.6% & Beta-strand = 7.2%	Burried = 39.2% & Exposed = 60.8%		Decreases	The mutant residue is bigger than the wild-type residue, this might lead to bumps. The wild-type residue is more hydrophobic than the mutant residue. Hydrophobic interactions, either in the core of the protein or on the surface, will be lost.
NNAT-203	rs542858994:Q52E	Extended-strand			Nuetral > Negative	Decreases	The mutation introduces a charge, this can cause repulsion of ligands or other residues with the same charge. Mutant residue is located near a highly conserved position and it's more likely that the mutation is damaging to the protein.
NNAT-203	rs560845323:R99Q	Random coil			Positive > Neutral	Increases	The charge of the wild-type residue will be lost, this can cause loss of interactions with other molecules or residues. The mutant residue is smaller, this might lead to loss of interactions.
NNAT-203	rs560845323:R99P				Positive > Neutral	Increases	The mutant residue is smaller than the wild-type residue this might leads to loss of interactions. The mutant residue is more hydrophobic than the wild-type residue, this can result in loss of hydrogen bonds and/or disturb correct folding.
NNAT-204	rs539681368:C30Y	Random coil	Alpha-helix = 60.5% & Coil = 39.5%	Burried = 53.1% & Exposed = 46.9%		Decreases	The wild-type and mutant amino acids differ in size. The mutant residue is bigger, this might lead to bumps. The hydrophobicity of the wild-type and mutant residue differs. Hydrophobic interactions, either in the core of the protein or on the surface, will be lost.
NNAT-204	rs542858994:F53L	Alpha-helix					The wild-type residue is very conserved. The wild-type and mutant amino acids differ in size. The mutant residue is smaller, this might lead to loss of interactions.
NNAT-205	rs542858994:S35L	Random coil	Alpha-helix = 21.4%, Coil = 67.9% &	Burried = 42% & Exposed = 58%		Increases	The mutant residue is bigger, this might lead to bumps. The mutant residue is more hydrophobic than the wild-type residue. This can result in

	Beta-strand = 10.7%		loss of hydrogen bonds and/or disturb correct folding.

Table 5 Transcript based computation of surface accessibility, secondary structure, disorder, and dihedral torsion angles of rare-coding nsSNPs of NNAT gene

Transcript	Variant ID : AA	NNATwild type					NNAT mutant type				
		RSA (%)	ASA (Å)	Phi (ϕ)	Psi (ψ)	Pdisorder (%)	RSA (%)	ASA (Å)	Phi (ϕ)	Psi (ψ)	Pdisorder (%)
NNAT-201	rs542858994 : F26L	60	120	-71	-34	0	55	100	-69	-36	0
NNAT-202	rs539681368 : C30Y	47	66	-89	37	3	70	150	-76	-26	1
NNAT-202	rs542858994 : F35L	51	102	-65	-41	0	52	95	-68	-41	0
NNAT-203	rs539681368 : C30Y	36	51	-95	94	8	55	118	-102	95	5
NNAT-203	rs542858994 : Q52E	62	111	-108	135	8	64	112	-114	136	3
NNAT-203	rs560845323 : R99Q	76	173	-95	136	54	73	131	-96	136	47
NNAT-203	rs560845323 : R99P						74	105	-68	149	42
NNAT-204	rs539681368 : C30Y	37	52	-86	30	8	60	129	-76	-32	2
NNAT-204	rs542858994 : F53L	50	100	-69	-41	2	37	68	-68	-42	0
NNAT-205	rs542858994 : S35L	77	90	-73	121	33	72	131	-68	126	29

RSA = Relative Solvent Accessibility; ASA = accessible surface area (ASA)

Table 6 Drug-likeness assessment of the selected therapeutic compounds of Aneroxia nervosa

Compound PubChem ID	Name	Chemical Formula	Mass (Dalton)	H-bond Donor	H-bond Acceptor	logP	Molar Refractivity	clogP	Solubility	TPSA	Druglikeness	D S
2160	Amitriptyline	$C_{20}H_{23}N$	277	0	1	3.703199	88.675964	4.41	-3.67	3.24	2.89	0
2771	Citalopram	$C_{20}H_{21}FN_2O$	324	0	3	3.812979	90.913971	2.79	-3.97	36.26	0.27	0
2995	Desipramine	$C_{18}H_{22}N_2$	266	1	2	3.532799	85.841675	3.62	-4.28	15.27	3.48	0
3386	Fluoxetine	$C_{17}H_{18}F_3NO$	309	1	2	4.114699	80.766685	3.62	-3.56	21.26	0.05	0
5962	Lysine	$C_6H_{14}N_2O_2$	146	5	4	-0.4727	38.51659	-4.68	-0.78	89.34	-18.21	0
4205	Mirtazapine	$C_{17}H_{19}N_3$	265	0	3	2.1807	81.877983	2.66	-2.67	19.37	5.74	0

Table 7 Docking-guided molecular mechanics and estimation of AN drugs binding energies with deleterious variant (rs539681368:C30Y) of human NNAT gene

Drugs	ΔG_{bind} (kcal/mol)	Ligand Efficiency	Inhibition Constant (K_i cal)	ΔE_{vdw}	ΔE_{ele}	ΔE_{MM}	$\Delta E_{(unbound)}$	$\Delta E_{(torsional)}$	ΔE (Total Internal)	$\Delta E_{(intermolecular)}$
Amitriptyline	-4.46	-0.21	534.78 uM	-4.93	-0.36	-5.29	-0.63	0.82	-0.63	-5.29
Citalopram	-4.1	-0.17	988.07 uM	-5.14	-0.34	-5.47	-1.12	1.37	-1.12	-5.47
Desipramine	-4.82	-0.24	294.37 uM	-4.77	-1.14	-5.91	-0.65	1.1	-0.65	-5.91
Fluoxetine	-4.54	-0.21	472.96 uM	-5.42	-1.03	-6.46	-0.92	1.92	-0.92	-6.46
Mirtazapine	-4.39	-0.22	604.9 uM	-4.71	0.32	-4.39	-	-	-	-4.39

Figures

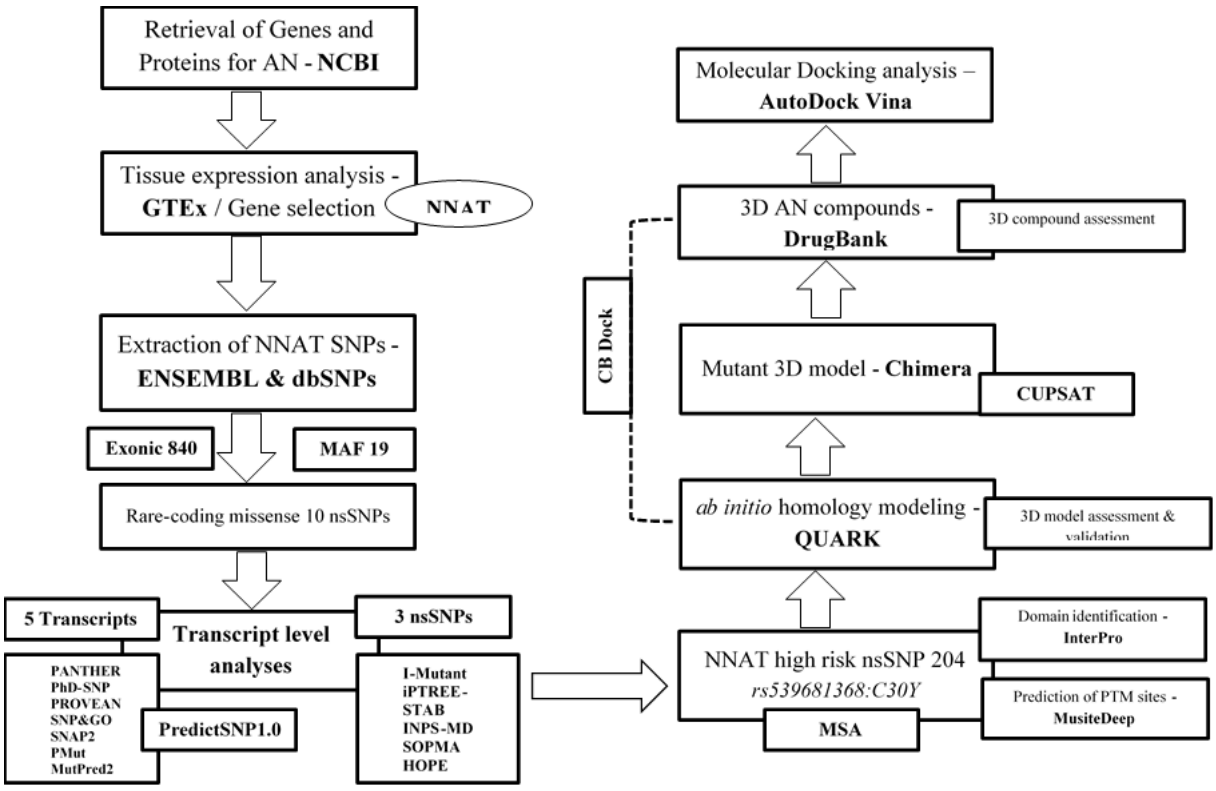

Figure 1

Scheme summarizing the important steps of present study

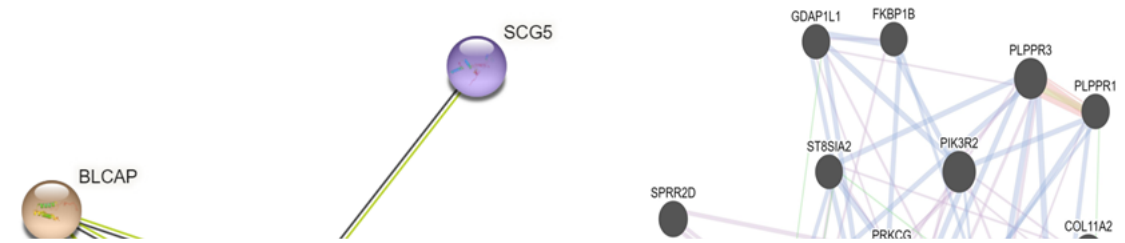

Figure 2

NNAT gene-gene interaction

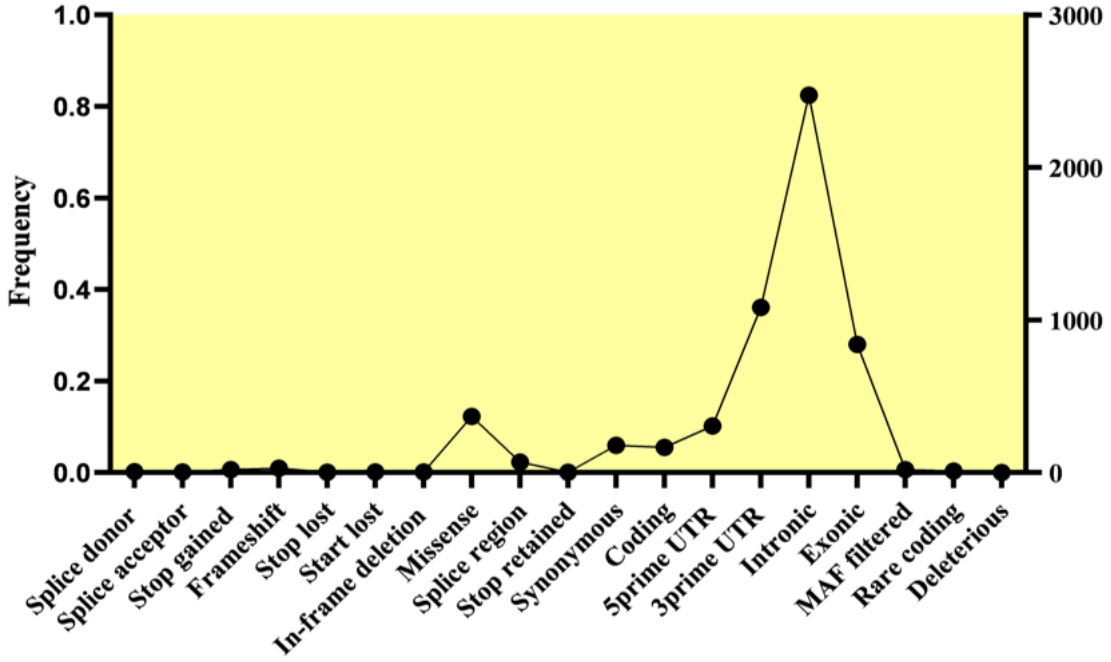

Figure 3

Distribution of identified mutation of human NNAT gene

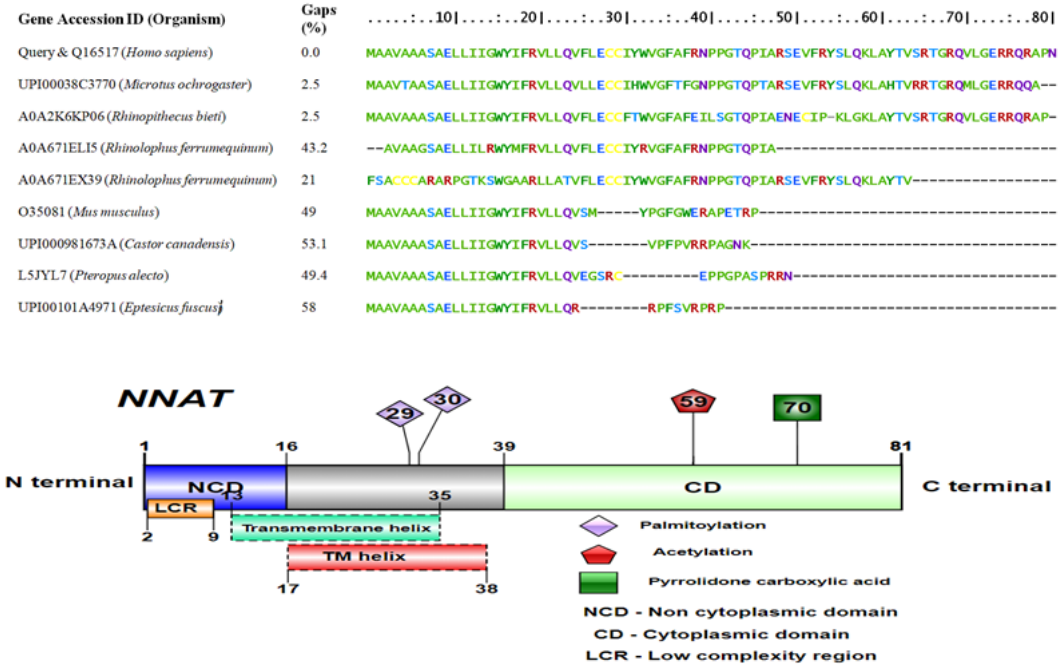

Figure 4

MSA analysis (NNAT-204) and the predicted PTM sites by *MusiteDeep* of the human NNAT α-isoform (using IBS software)

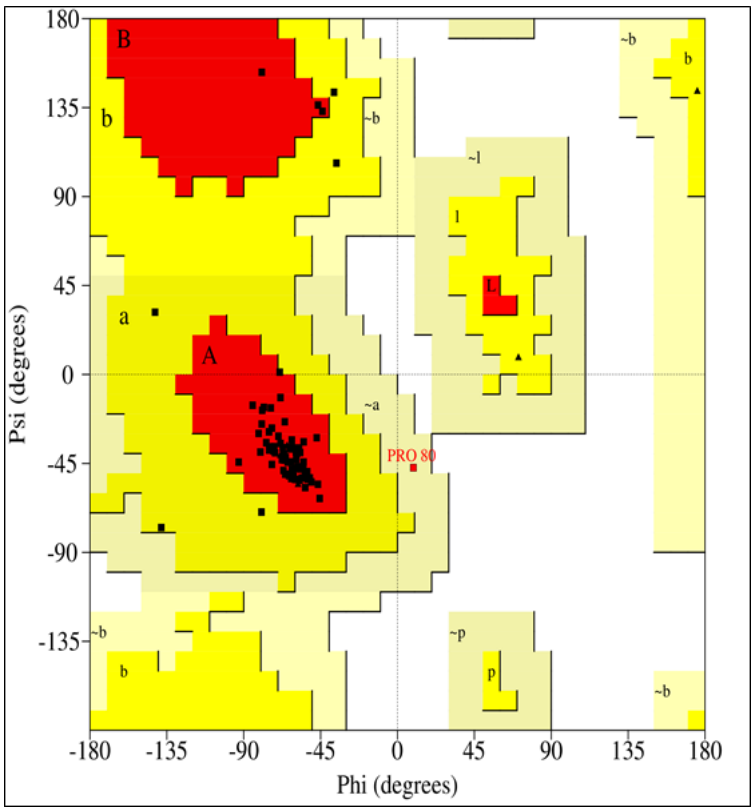

Figure 5

Ramachandran plot of NNAT pathogenic (rs539681368:C30Y) variant

Figure 6

2D plots showing blind docked cavity binding interaction of NNAT homology model (wild type) with selected AN drugs. Coloring scheme showed green color = conventional H-bonds; light green = van der Waals; Dark purple = Pi-Pi stacked; light purple = Pi-alkyl; red = unfavorable positive-positive; electric blue = halogen

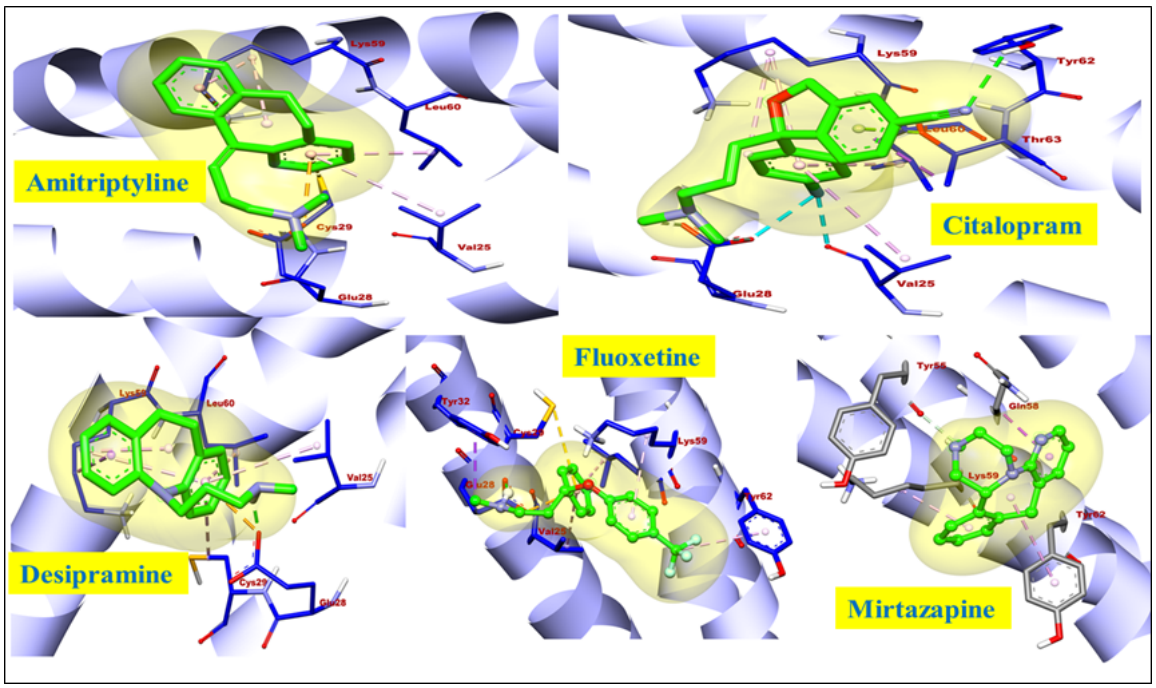

Figure 7

Binding of AN drugs with the homology model of pathogenic variants (*rs539681368*: C30Y) of human NNAT gene