



Article SNP and Structural Study of the Notch Superfamily Provides Insights and Novel Pharmacological Targets against the CADASIL Syndrome and Neurodegenerative Diseases

Louis Papageorgiou ^{1,2}, Lefteria Papa ¹, Eleni Papakonstantinou ^{1,3}, Antonia Mataragka ¹, Konstantina Dragoumani ¹, Dimitrios Chaniotis ², Apostolos Beloukas ², Costas Iliopoulos ⁴, Erik Bongcam-Rudloff ⁵, George P. Chrousos ³, Sofia Kossida ⁶, Elias Eliopoulos ¹, and Dimitrios Vlachakis ^{1,3,4,*}

- ¹ Laboratory of Genetics, Department of Biotechnology, School of Applied Biology and Biotechnology, Agricultural University of Athens, 75 Iera Odos, 11855 Athens, Greece; louis.papageorgiou@aua.gr (L.P.); lefteria.papa@aua.gr (L.P.); eleni@aua.gr (E.P.); antonia.mataragka@gmail.com (A.M.); kwstantina@gmail.com (K.D.); eliop@aua.gr (E.E.)
- ² Department of Biomedical Sciences, School of Health and Care Sciences, University of West Attica, Agioy Spyridonos, 12243 Egaleo, Greece; dchaniotis@uniwa.gr (D.C.); abeloukas@uniwa.gr (A.B.)
- ³ University Research Institute of Maternal and Child Health & Precision Medicine, National and Kapodistrian University of Athens, "Aghia Sophia" Children's Hospital, 11527 Athens, Greece; chrousos@gmail.com
- ⁴ School of Informatics, Faculty of Natural & Mathematical Sciences, King's College London, Bush House, Strand, London WC2R 2LS, UK; csi@kcl.ac.uk
- ⁵ Department of Animal Biosciences, Swedish University of Agricultural Sciences, 756 51 Uppsala, Sweden; erik.bongcam@slu.se
- ⁶ IMGT[®], The International ImMunoGenetics Information System[®], Laboratoire d'ImmunoGénétique Moléculaire LIGM, Institut de Génétique Humaine, (IGH), Centre National de la Recherche Scientifique (CNRS), Université de Montpellier (UM), 34000 Montpellier, France; sofia.kossida@igh.cnrs.fr
 * Correspondence: dimvl@aua.gr

Abstract: The evolutionary conserved Notch signaling pathway functions as a mediator of direct cell-cell communication between neighboring cells during development. Notch plays a crucial role in various fundamental biological processes in a wide range of tissues. Accordingly, the aberrant signaling of this pathway underlies multiple genetic pathologies such as developmental syndromes, congenital disorders, neurodegenerative diseases, and cancer. Over the last two decades, significant data have shown that the Notch signaling pathway displays a significant function in the mature brains of vertebrates and invertebrates beyond neuronal development and specification during embryonic development. Neuronal connection, synaptic plasticity, learning, and memory appear to be regulated by this pathway. Specific mutations in human Notch family proteins have been linked to several neurodegenerative diseases including Alzheimer's disease, CADASIL, and ischemic injury. Neurodegenerative diseases are incurable disorders of the central nervous system that cause the progressive degeneration and/or death of brain nerve cells, affecting both mental function and movement (ataxia). There is currently a lot of study being conducted to better understand the molecular mechanisms by which Notch plays an essential role in the mature brain. In this study, an in silico analysis of polymorphisms and mutations in human Notch family members that lead to neurodegenerative diseases was performed in order to investigate the correlations among Notch family proteins and neurodegenerative diseases. Particular emphasis was placed on the study of mutations in the Notch3 protein and the structure analysis of the mutant Notch3 protein that leads to the manifestation of the CADASIL syndrome in order to spot possible conserved mutations and interpret the effect of these mutations in the Notch3 protein structure. Conserved mutations of cysteine residues may be candidate pharmacological targets for the potential therapy of CADASIL syndrome.

Keywords: Notch family members; neurodegenerative diseases; CADASIL; genetics; polymorphism analysis; mutation analysis; EGF; cysteine; protein structure analysis



Citation: Papageorgiou, L.; Papa, L.; Papakonstantinou, E.; Mataragka, A.; Dragoumani, K.; Chaniotis, D.; Beloukas, A.; Iliopoulos, C.; Bongcam-Rudloff, E.; Chrousos, G.P.; et al. SNP and Structural Study of the Notch Superfamily Provides Insights and Novel Pharmacological Targets against the CADASIL Syndrome and Neurodegenerative Diseases. *Genes* **2024**, *15*, 529. https://doi.org/ 10.3390/genes15050529

Academic Editor: Xingguang Luo

Received: 15 March 2024 Revised: 17 April 2024 Accepted: 22 April 2024 Published: 23 April 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/).

1. Introduction

Research on *Drosophila melanogaster* with notched wings led to the discovery of the Notch gene in 1914 [1]. To date, it seems that the evolutionary history of the Notch family is closely related to the biological tree of life. The Notch protein and its homologs, Notch1, Notch2, Notch3, Notch4, LIN-12, and GPL-1, have been detected in the genomes of all kingdoms, demonstrating the evolutionary development of the Notch family [2]. Members of the Notch family were discovered to have a comparable structure across several kingdoms, extending from bacteria to chordates [3]. Only one Notch receptor is found in *D. melanogaster*. The Notch receptors LIN-12 and GLP-14 in *Caenorhabditis elegans* are redundant [4]. Mammals have four Notch paralogs, Notch1, Notch2, Notch3, and Notch4, displaying both redundant and distinct activities [5].

The Notch receptors (Notch1–Notch4), found in mammalian cells, are four different transmembrane proteins expressed on the cell's surface as heterodimers not covalently bonded [6]. Notch proteins have an extracellular domain (NECD) that operates as the signal receiver and a transmembrane-intracellular domain (NICD) that operates as the signal transducer. The Notch1–Notch4 ECDs contain 36, 35, 34, and 29 epidermal growth factor-like repeats (EGF-like domain), respectively. Also, the ECD of Notch receptors has three cysteine-rich Lin12-Notch repeats (LNRs) and a heterodimerization domain (HD). The Notch ICD has an RBPJk-associated molecule domain (RAM) and nuclear localization sequences (NLSs) on both sides of the ANK domains. Also, Notch ICD consists of five to six ankyrin repeats (ANK), a transcriptional activation domain (TAD), and a C-terminal domain (PEST) rich in proline, glutamic acid, serine, and threonine. Notch family proteins function as cell surface receptors and direct regulators of gene transcription, constituting a particular signal transduction pathway that enables cells to affect the gene expression of their neighboring cells [7]. Notch signaling is activated upon cell-to-cell contact due to interactions between four transmembrane receptors encoded by Notch genes (Notch1-4) and five Notch ligands encoded by JAG1, JAG2, and DLL1, DLL3, and DLL4 [8]. Notch ECD contains EGF-like repeats that condense ligand–receptor binding [9].

The human (*Homo sapiens*) Notch1 gene is found at locus 9q34.3 on chromosome 9. Loss of function of the Notch1 protein is linked to abnormalities in angiogenesis, cardiogenesis, and somitogenesis, which could lead to the death of an embryo. This gene is involved in forming the first definitive adult hematopoietic stem cells (HSCs) [10]. Moreover, the development of B and T cells is regulated by Notch1 signaling. Mutations in the signaling and transcriptional regulator Notch1 result in various developmental aortic valve abnormalities, severe valve calcification, and T-cell acute lymphoblastic leukemia [10,11]. The Notch2 gene is located on chromosome 1p12. Notch2 has specific functional activity in determining cell fate and in the development of kidney, ovary, smooth muscle, T, and B cells [12]. Postnatal signaling regulates homeostasis, bone regeneration, and immune system function [13,14]. Mutations resulting in excessive Notch2 activity may lead in systemic issues typical of Alagille and Hajdu-Cheney syndromes such as heart abnormalities, chronic cholestasis, osteoporosis, polycystic kidneys, skeletal deformities, and neurological disorders [15,16]. The Notch3 gene is found between locations 13.2 and 13.1 on the short arm (p) of chromosome 19. This large type I transmembrane receptor, mostly expressed in pericytes and vascular smooth muscle cells adjacent to the local blood arteries, takes part in maintaining and renewing tissues as well as in important developmental functions [17]. Overexpression and aberrant activation of the Notch3 gene are linked to cancer, particularly breast and ovarian cancer. Mutations in Notch3 have been directly linked to the CADASIL syndrome [18]. The Notch4 gene is found at locus 6p21.32 on chromosome 6. It has been observed that both the overexpression and mutations of the Notch4 gene are related to cancer [19]. Notch4 is considered a new biomarker of cancer stem cells (CSCs) [20].

Notch genes are involved in various critical biological processes including somitogenesis, angiogenesis, vasculogenesis, cardiac development and function, neuronal development, and the specification and maintenance of neural stem cells (NSCs) [21]. All four mammalian Notch receptor paralogs and several pathway components (ligands and targets) are expressed with different cell type specificities in both the adult mouse and human brain [22]. There is evidence of Notch receptor expression in neurons (Notch1 and Notch2), neural stem cells (Notch1 and Notch2), vascular smooth muscle cells and pericytes (Notch3), endothelial cells (Notch1 and Notch4), and astrocytes (Notch1 and Notch2) [22]. Notch has been linked to maintaining NSCs in an undifferentiated state, preventing neuronal development, and even causing terminal differentiation inside the astrocyte lineage [23,24]. Notch signaling is essential for neural stem cell maintenance and neurogenesis in both the embryonic and adult brains [6]. The elderly's brain function, cell differentiation, and neurite formation are all impacted by Notch signaling, which is crucial for the nervous system's regular operation [5,25]. Accordingly, multiple mutations in Notch proteins have been linked to neurodegenerative conditions [21].

Quantitative data on this pathway's structural, biochemical, and biophysical features have emerged during the last few years [26]. Various loss-of-function mutations in the embryo and adult highlight the critical role of Notch signaling. Numerous studies on neurogenesis have used *Drosophila melanogaster*, zebrafish, and mice as model species [6]. The conditional loss of Notch signaling in the embryo causes the precocious differentiation of NSCs and neurodevelopmental defects such as impaired survival and the aberrant migration of progenitor cells [6]. In the adult brain, NSCs are predominantly quiescent and rarely divide. However, it is likely that quiescent NSCs enter the cell cycle and transform into active NSCs before quitting the cell cycle again and reentering the quiescent state [27]. In adults, Notch signaling pathway mutations are involved in many neurodegenerative diseases and brain disorders [6].

Neurodegenerative diseases are incurable disorders of the central nervous system that present clinically and pathologically in various ways and damage particular neuronal subsets and anatomical functioning systems [28]. There are currently no therapies that target the underlying cause of neurodegenerative diseases. Therefore, it is not feasible to prevent or stop the progression of these disorders [29]. The involvement of Notch receptor genes and proteins in aging, cerebrovascular disorder, and Alzheimer's disease is significant. Notch signaling may be a fundamental overlap between age-related vascular and Alzheimer's pathogenesis that contributes to their comorbidity and combined impact on cognitive decline and dementia. Numerous results from genetics, cell culture model studies, and neuropathology all point to a connection between aberrant Notch signaling and the pathogenesis of Alzheimer's disease [21]. In addition, it is generally established that the Notch3 protein plays a significant role in the development of CADASIL [30–32].

Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy is a hereditary dominant rare disease caused by mutations in the Notch3 protein, affecting adults beyond middle age and resulting in dementia and disability [33]. CADASIL is a fatal late-onset disease that primarily appears as a degenerative disorder of the central nervous system, and it is defined by specific clinical, neuroradiological, and pathological characteristics [30,34]. Over the last two decades, extensive efforts have been directed toward research on Notch3, identifying more than 280 mutations [35]. Some of these mutations cause a phenotype whereas others remain silent. Extensive analysis for categorizing, organizing, and mapping these mutations is required for a simple genotypephenotype linkage [33]. Numerous pathogenic mutations in the Notch3 gene change the number of cysteine residues in the receptor's extracellular domain, leading to protein misfolding and receptor aggregation [33]. Each EGF-like repeat contains six cysteines, which combine to create three disulfide bonds and provide the EGF repeat its three-dimensional structure [36]. However, non-Cys mutations have also been reported in recent years. These mutations do not match the disease's typical pattern and pathology [30]. Even though most of the mutations in Notch3 are point mutations, it has been established that each one has a major impact on the three-dimensional structure of the Notch3 protein [30].

The study of the Notch family has increased, significantly, the availability of biological data on polymorphisms and mutations that are related with neurodegenerative diseases [15,21,37–39]. The initial purpose of this work was to gather all the cases and link them between nucleotides and protein sequences. Today, the scientific research seems to be focused on understanding the way the EGF region functions due to the significant mutagenesis it presents through a series of scientific publications [39–43]. The ultimate goal of this study is the holistic study of all mutations occurring in *Notch3*, with an emphasis on the EGF region and the CADASIL syndrome, in order to identify specific patterns of mutagenesis in the EGF-repeats that may be related to the clinical phenotype, sex, and age data of the various patients [38,39,44–46]. The study and analysis of all mutations can additionally open new horizons thus contributing to the identification of new pharmacological targets as well as contributing to the identification of a candidate treatment against CADASIL syndrome and, generally, neurodegenerative diseases. The outline of the integrated bioinformatic method is presented in Figure 1.



Figure 1. Flow chart presentation of the bioinformatic method, presented in five steps.

2. Materials and Methods

2.1. Dataset Collection and Filtering

Data were collected from polymorphism databases, disease-specific mutation databases, and publications. Specifically, single-nucleotide polymorphisms (SNPs) on Notch1–Notch4 genes associated with neurodegenerative diseases were extracted from available online databases such as GWAS-Catalog, dbSNP, LitVar, and ClinVar. Likewise, a second search was carried out in the online database PubMed (https://pubmed.ncbi.nlm.nih.gov/, accessed on 18 March 2024) for publications that contained the key terms "Neurodegenerative diseases", "Cognitive Disorders", "Alzheimer's disease", "CADASIL" AND "NOTCH1", "NOTCH2", "NOTCH3", and "NOTCH4" with no date restriction. The collected SNPs from all databases were extracted, filtered, and annotated using Matlab bioinformatics toolbox for data mining and semantic techniques. All SNPs causing mutations on the protein level and directly related to neurodegenerative diseases were stored the final dataset. The Human Gene Mutation Database (HGMD[®]) (https://www.hgmd.cf.ac.uk/ac/index.php, accessed on 18 March 2024) was searched for missense mutations on Notch1-Notch4 proteins. HGMD^{®®} attempts to aggregate all known (published) human gene mutations responsible for human diseases. The mutations associated with neurodegenerative diseases have been collected. For each mutation, the access number, codon change, mutation position, and the phenotype it induces were recorded.

2.2. Gene and Protein Mapping

In this step, mapping of human Notch family genes and proteins was accomplished. The terms "Notch1", "Notch2", "Notch3", and "Notch4" were searched on the NCBI database (https://www.ncbi.nlm.nih.gov, accessed on 18 March 2024) while the filter "Gene" was previously selected for extracting the nucleotide sequences of these genes. Furthermore, additional information like gene location, chromosome, nucleotide sequence length, access number, and alternative gene names was extracted from the NCBI database. A second search was carried out on NCBI database with the "protein" filter for extracting the amino acid sequences of human Notch family proteins. These terms were also searched on available online protein databases such as UniProt and InterPro. Information on the amino acid sequence lengths of proteins and disease involvement was obtained. Domains of each protein were also recorded. Protein domain data were also extracted from publications in the PubMed database.

2.3. Data Integration

The data collected from polymorphism databases and mutation databases were merged and annotated. SNPs and the mutations in the human Notch family associated with neurodegenerative diseases were integrated to correlate polymorphisms and mutations. The integrated data are presented in a table, providing information for SNP ID, nucleotide change, codon change, mutation position, domain in which the mutation is located, and the phenotype it causes. The finding of correlations among polymorphisms and mutations on Notch1–Notch4 was necessary to find out which ones and how many mutations are associated with a known polymorphism, which neurodegenerative disease is most often caused by Notch3 mutations, which Notch3 domain contains the majority of these mutations, and which amino acid appears to be most frequently mutated. To comment on these queries, specific diagrams have been created where the data analysis is presented.

2.4. Mutation Analysis

The majority of polymorphisms/mutations associated with neurodegeneration were located in Notch3 and specifically in the EGF region. For this reason, our work then focused on this domain (Table 7). In order to analyze the mutations identified in the EGF-like repeats of the Notch3 protein, a FASTA file with the amino acid sequences of 34 EGF-like repeats was created. This file was analyzed using the Jalview platform (https://www.jalview.org/, accessed on 18 March 2024). JalView computes and visualizes a large number of sequences with high performance. The main advantage of this method is that it allows for the identification of conserved motifs with a quick overview of alignment. In this step, a multiple alignment of 34 EGF-like amino acid sequences was performed in order to find the conserved amino acids in the EGF-like repeats. Jalview's comments section, which displays amino acid conservation with logos and histograms, was also examined to discover novel motifs. The sequencing results were further elaborated. A histogram was created, showing the percentage of mutations at each amino acid position in the EGF-like sequences according to multiple-sequence alignment numbering. Finally, mutations in conserved amino acids in the EGF-like repeats were studied to identify conserved amino acid changes. A chart presenting the results of this analysis was also constructed.

2.5. Structural Analysis

Using the MOE (Molecular Operating Environment) platform, the mutant EGF-like repeat structure was analyzed to figure out the consequences for the Notch3 protein when including amino acid changes. MOE 2019.01 is an integrated life sciences software that supports drug design through molecular simulation, protein structure analysis, small molecule processing data, protein binding, and small molecule design. A homology modeling of the Notch3 approach was used for the structural analysis since no structure has been determined for this protein. The homology modeling of Notch3 protein structure was extracted from the AlphaFold Protein structure Database (https://alphafold.ebi.ac.uk/,

accessed on 18 March 2024). Structural analysis of the EGF-like 2 repeat was performed while introducing conserved mutations associated with CADASIL syndrome.

3. Results

3.1. Dataset

As a consequence of systematic data mining, SNPs of human Notch family members correlated with neurodegenerative diseases were identified (Table 1). These polymorphisms have been derived both from the data mined from the biological database GWAS CATALOG as well as from the publications that contained the ontologies of interest based on PubMed searches. In total, 1887 relevant publications were extracted, of which 188 described polymorphisms. Only single-nucleotide polymorphisms that cause mutations at the protein level and are directly related to neurodegenerative diseases were used. In the Notch1 gene, 57 polymorphisms were collected from online polymorphism databases and publications, of which 41 missense variants were screened. Only one variant was identified in this gene related to Alzheimer's disease (Table 4). Twenty-six polymorphisms in the Notch2 gene were retrieved. Among these, 15 missense SNPs were examined, and only one was found to be associated with autism multiplex disorder (Table 5). A total of 59 polymorphisms in the Notch3 gene were extracted from online databases and publications. Twenty-eight missense SNPs were screened, and all were found to be involved in the manifestation of neurodegenerative diseases (Table 7). Most SNPs are associated with CADASIL disease. Finally, from the Notch4 gene were extracted only two SNPs (missense variants), and neither was found to be associated with neurodegenerative disease (Table 6). Unlike SNPs in Notch1, Notch2, and Notch4, missense SNPs in the Notch3 gene seem more strongly related to neurodegenerative disorders. Additionally, through filtering the results of the HGMD database, all mutations in Notch1–Notch4 associated with neurodegenerative diseases were reported and imported. More specifically, 1, 1, 312, and 4 mutations in Notch1, Notch2, Notch3, and Notch4, respectively, were identified to be correlated with neurodegenerative diseases.

Table 1. Dataset of collected SNPs and mutations from online databases.

	SNPs	Mutations	Neurodegenerative Diseases
Notch1	1	1	Mental retardation, autosomal dominant, Alzheimer's disease
Notch2	1	1	Autism multiplex
			CADASIL, white matter lesions, Alzheimer's disease, ischemic
Notch3	28	312	stroke, cerebral small-vessel disease, arteriopathy and
			cavitating leukoencephalopathy, autism, migraine
Notch4	0	4	Schizophrenia, multiple sclerosis, migraine

3.2. Gene and Protein Mapping

Data on Notch1–Notch4 genes and proteins were retrieved from the National Center for Biotechnology Information (NCBI), UniProt, and InterPro. NCBI-Gene results for Notch1–Notch4 are shown in Table 2. The search results for Notch1–Notch4 proteins are shown in Table 3. Notch1 has the greatest amino acid sequence length of the four Notch proteins (2555 aa), followed by Notch2 (2471 aa), Notch3 (2321 aa), and Notch4 (2003 aa). Mutations in these genes are linked to the manifestation of several diseases. Mutations in the Notch1 gene are associated with diseases such as aortic valve disease, type 1 (AOVD1), Adams–Oliver syndrome type 5 (AOS5), T-cell acute lymphoblastic leukemia (T-ALL), chronic lymphocytic leukemia, and squamous cell carcinoma of the head and neck. Mutations in Notch2 are linked to Hajdu–Cheney syndrome, Alagille syndrome 2 (ALGS2), and cancer. Mutations in Notch3 have been identified as the cause of diseases such as CADASIL, infantile myofibromatosis, early-onset arteriopathy with cavitating leukodystrophy, lateral meningocele syndrome, and cancer. Finally, mutations in the Notch4 gene may be associated with schizophrenia.

Gene	Notch1	Notch2	Notch3	Notch4
Locus	9q34.3	1p12	19p13.12	6p21.32
DNA linear	58343 bp	165110 bp	48349 bp	36225 bp
Exon count	34	34	33	31
Accession number	NG_007458	NG_008163	NG_009819	NG_028190
Organism	Homo sapiens	Homo sapiens	Homo sapiens	Homo sapiens
Also known as	hN1; AOS5; TAN1; AOVD1	hN2; AGS2; HJCYS	IMF2; LMNS; CASIL; CADASIL; CADASIL1	INT3

Table 2. NCBI data for *Notch1–4* genes.

Table 3. NCBI data for Notch1–Notch4 proteins.

Gene	Notch1	Notch2	Notch3	Notch4
Protein length	2555 aa	2471 aa	2321 aa	2003 aa
Accession number	NP_060087	NP_077719 XP_946791 XP_950472	NP_000426	NP_004548
Organism	Homo sapiens	Homo sapiens	Homo sapiens	Homo sapiens
Disease	T-cell acute lymphoblastic leukemia, Adams–Oliver syndrome, aortic valve disease, cancer	Hajdu–Cheney syndrome, Alagille syndrome, cancer	CADASIL, infantile myofibromatosis, early-onset arteriopathy with cavitating leukodystrophy, lateral meningocele syndrome, cancer	Schizophrenia

By mining information from databases and publications, Notch1–Notch4 proteins were mapped. The Notch1–4 proteins consist of EGF, LNR, NOD, NODP, TM, RAM, NLS, ANK, TAD, and PEST domains [47]. In mammals, the TAD region is present in Notch1 and 2 but not in Notch3 and 4 [48]. The EGF domain is made up of EGF-like repeats. In human Notch1, Notch2, Notch3, and Notch4, the EGF domains consist of 36 EGF-like, 35 EGF-like, 34 EGF-like, and 29 EGF-like repeats, respectively. Each EGF-like repeat comprises 30–40 amino acids and contains six cysteine residues (C). The LNR sector consists of three micro-domains, LNR1, LNR2, and LNR3. The ANK domain in Notch1–2 proteins is made of six ankyrin repeats while in Notch3-4, proteins are made of five ankyrin repeats. Notch1–Notch4 protein domains are demonstrated with specific colors in protein sequences in Figures 2–5.

3.3. Data Integration

The integrations of polymorphism and mutation datasets of Notch1–Notch4 are shown in Tables 4–7. Since Notch3 is associated with the manifestation of neurodegenerative diseases to a considerably more significant degree than other human Notch family members, the annotation of Notch3 data followed. In Table 6, the consolidated data from the recording of Notch3 polymorphisms and mutations are presented in order to correlate them. Information on nucleotide change, amino acid change, protein domain on which the mutation is located, and the phenotype it induces were obtained for the recorded SNPs. According to the association between polymorphisms and mutations reported, 23 polymorphisms are related to mutations not identified in the HGMD data source. A blank cell in the "Accession number" column of the table indicates the particular polymorphisms. Although there are 43 mutations associated with known SNPs, there are 292 mutations unrelated to any identified SNP (Figure 6).

EGF-like 1 L 20 10 30 40 50 CLNGGKCEAA MPPLLAPLLC LALLPALAAR GPRCSQP STEACV EGF-like 2 11 1 70 80 90 100 60 QDP NPCLSTPCKN AC DRR GVADYACSCA LO CHV ١ EGF-like 3 11 110 130 140 120 150 KCRCP PGWSGKSCQQ ADPCASNPCA PLDNACLTNP CRNGGTCDLL EGF-like 4 EGF-like 5 11 170 180 190 200 160 HCPPSF HGPTCRQDVN **LPFEA** GLCR NEVG EGF-like 6 11 210 220 240 250 230 RATH TGPNCERPYV PCSPSP HECACLPGFT CONG GDVT EGF-like 7 11 11 260 270 280 290 300 EENIDD CPGNNCKNGG A /NTYN PPEWTGQ YCT<mark>EDVDECQ</mark> EGF-like 8 11 350 330 310 320 340 ONGG GGYN C WTGE SENI DDCA SAA GATC EGF-like 9 11 EGF-like 10 370 380 400 360 390 GRTGLLC HLNDACISNP (DRVASFYCE CDTN /NGKAICTC EGF-like 11 1 1 1 410 420 430 440 450 DECSLGA NPCEHA GPACS Q KCI PRCE ECQC EGF-like 12 I 11 480 490 500 460 470 ICMP GYEGVHCE<mark>VN TDECASSPCL</mark> VSNP ONDATCLDQ EGF-like 13 EGF-like 14 11 510 520 530 540 550 PTGF TGHLCQYDVD KNG PNTY DKIN 11 EGF-like 15 570 560 580 590 600 GYTG THCEVDIDEC DPDF HYGSC GHHC FTCL CRPGY EGF-like 16 11 1 1 640 610 620 630 650 NCEI NI TCQD FCL SSPC EGF-like 17 11 EGF-like 18 670 700 660 680 690 TGSMCNINID E TCLDKID ACEPGY CHNG GINGF EGF-like 19 11 710 720 730 740 750 RCPEGYHD PTCLSEVNEC NSNE HGAC GYKCD GTNC EGF-like 20 11 1 1 790 800 760 770 780 EGFSGPNCQT NINECA DINNNECESN PCVNGG ICKD TCR SNPC EGF-like 21 11 EGF-like 22 820 830 850 840 810 CLLP YTGATCE<mark>VVL APCAPSPCRN</mark> LNOGTCIDDV ROSEDY EGF-like 23 11 870 860 880 890 900 EV<mark>DI NE</mark> PCRH HGG AGYS PTG 1.1 EGF-like 24 11 910 920 940 950 930 GFRGT TDIDD TAF (NECA HNGG EGF-like 25 EGF-like 26 11 960 970 980 990 1000 GTHC ENN GANC YTCT TES NGGTCVD EGF-like 27 1 1 1020 1010 1030 1040 1050 GFTGSYCQH CQDGC TCPQG CLCP SQPC I I 1100 EGF-like 28 1 1 1060 1070 1080 1090 QNLV F PCKN THTQ PSGWT CDV<mark>PSVS</mark> EGF-like 29 11 1130 1140 1150 1110 1120 ORQGV DVARLCQHGG AGYTGS YCEDLVDECS GNTHH C EGF-like 30 EGF-like 31 11 1170 1190 1200 1160 1180 PSPCQNGATC TDYLGGYSCK CVAGYHGVNC SEEII ECLSH CQNGGTCLD 1 1 EGF-like 32 1210 1220 1230 1240 1250

LPNTYKCSCP RGTQGVHCEI NVDDCNPPVD PVSRSPKCFN NGTCVDQVGG

	1.1	EGE	-like 33	
1260	1270	1280	1290	1300
YSCTCPPGFV	GERCEGDVNE	CLSNPCDARG	TQNCVQRVND	FHCECRAGHT
1310	1320	GF-like 34	1340	1350
GRRCESVING	CKGKPCKNGG	TCAVASNTAR	GFICKCPAGF	EGATCENDAR
1	EGF-like 35		1	
1360	1370	1380	DECOEPASSP	1400
EGF-like	36		r Ecgr Phoor	
1410	1420	1430	1440	1450
GTCEPTSESP	FYRCLCPAKF	NGLLCHILDY	SFGGGAGRDI	PPPLIEEA <mark>CE</mark>
1460	1470	1480	1490	1500
LPECQEDAGN	KVCSLQCNNH	ACGWDGGDCS	LNFNDPWKNC	TQSLQCWKYF
LNR	2	1530	11 LI 1540	NR 3
SDGHCDSOCN	SAGCLFDGFD	CORAEGOCNP	LYDOYCKDHF	SDGHCDOGCN
	11		NOD	
1560	1570	1580	1590	1600
SAECEMDGED	CALIVPERLA	AGITAAATU	PPEQLENSSE	UL PKEPSKAP
1610	1620	1630	1640	1650
HTNVVFKRDA	HGQQMIFPYY	GREEELRKHP	IKRAAEGWAA	PDALLGQVKA
1660	1670	1680	1690	1700
SLLPGGSEGG	RRRRELDPMD	VRGSIVYLEI	DNRQCVQASS	QCFQSATDVA
	I		TM	
1710	SINTPYKTEA	1730	1740 PAOLHEMYVA	1750
HI DOUDADDO	ODWITINIDA		LAQUITINI VA	
1760	1770	1780	1790	1800
GCGVLLSRKR	RRQHGQLWFP	EGFKVSEASK	KKRREPLGED	SVGLKPLKNA
1810	1820	1830	1840	1850
SDGALMDDNQ	NEWGDEDLET	KKFRFEEPVV	LPDLDDQTDH	RQWTQQHLDA
1960	1970	1000	1900	1000
ADLRMSAMAP	TPPOGEVDAD	CMDVNVRGPD	GFTPLMIASC	SGGGLETGNS
1	NLS	1 1	AN	K 1
1910	1920	1930	1940	1950
EEEEDA <mark>PAVI</mark>	ANK	2	ALHLAAKISK	SDAAKKLLEA
1960	1970	1980	1990	2000
SADAN I QDN <mark>M</mark>	GRTPLHAAVS	ADAQGVFQIL	IRNRATDLDA	RMH <mark>DGTTPLI</mark>
2010	NK 3 2020	2030	2040	4 2050
LAARLAVEGM	LEDLINSHAD	VNAVDDLGKS	ALHWAAAVNN	VDAAVVLLKN
	A	NK 5		
CANKDMONNR	EETPLELAAR	EGSYETAKVI.	LDHFANRDTT	DHMDRLPRDT
ANK 6	DD11 D1 D1 D1		NLS	
2110	2120	2130	2140	2150
AQERMHHDIV	RLLDEYNLVR	SPQLHGAPLG	GTPTLSPPLC	SPNGYLGSLK
2160	2170	2180	2190	2200
PGVQGKKVRK	PSSKGLACGS	KEAKDLKARR	KKSQDGKGCL	LDSSGMLSPV
2210	2220	2230	2240	2250
DSLESPHGYL	SDVASPPLLP	SPFQQSPSVP	LNHLPGMPDT	HLGIGHLNVA
		TAD		
2260	2270	2280	2290	2300
AKPEMAALGG	GGRLAFEIGP	PKT2HTLAY	GISIATG222	GGALNE IVGG
2310	2320	2330	2340	2350
STSLNGQCEW	LSRLQSGMVP	NQYNPLRGSV	APGPLSTQAP	SLQHGMVGPL
2360	2370	2380	2390	2400
HSSLAASALS	QMMSYQGLPS	TRLATQPHLV	QTQQVQPQN	QMQQQNLQPA
2410	2420	2430	2440	2450
NIQQQQSLQP	PPPPPQPHLG	PEST	RSFLSGEPSQ	ADVQPLGPSS
2460	2470	2480	2490	2500
LAVHTILPQE	SPALPTSLPS	SLVPPVTAAQ	FLTPPSQHSY	SSPVDNTPSH
2510	2520	2520	2540	2550
OLOVPEHPFL	TPSPESPDOW	SSSSPHSNVS	DWSEGVSSPP	TSMOSOIARI
I				
D A DW				

Figure 2. Notch1 protein domains. Colors represent protein domains.

,								
		POP 14h	. 1				FOF	-1iko 33
10	20	EGF-IIK	e 1 40	50	1260	1270	1280	-11ke 33
PALRPALLW	ALLALWLCCA	APAHALOCRD	GYEPCVNEGM	CVTYHNGTGY	RCLPGFAGER	CEGDINECLS	NPCSSEGSLD	CIQLTNDY
	11	EGF-11	ke 2		1.1	I	EGF-like 34	
60	70	80	90	100	1310	1320	1330	13
KCPEGFLGE	YCQHRDPCEK	NRCQNGGTCV	AQAMLGKATC	RCASGFTGED	CE <mark>TFVDVCPQ</mark>	MPCLNGGTCA	VASNMPDGFI	CRCPPGFS
1 1	EGF-	like 3		11		1	EGF-	like 35
110	120	130	140	150	1360	1370	1380	13
QYS <mark>TSHPCF</mark>	VSRPCLNGGT	CHMLSRDTYE	CTCQVGFTGK	ECQWTDACLS	CRKGEQCVHT	ASGPRCFCPS	PRDCESGCAS	SPCQHGGS
EGF-lik	(e 4		EGF-	-like 5	1410	1420	1420	14
		180		200				ADKARDCU
PCANGSICI	TVANQESCRC	LIGITGUNCE	-like 6	HCQHGGTCLN	QUAPPESGSK		PPAICLOQIC	LND 2
210	220	230	-11Ke 0	250	1460	1470	1480	14
PGSYOCOCP	OGETGOYCDS	LYVPCAPSPC	VNGGTCROTG	DETEECNCLP	DGGDCSLTME	NPWANCSSPL	PCWDYINNOC	DELCNTVE
	201102102	EGF-lik	e 7	1 1	11	LI	NR 3	
260	270	280	290	300	1510	1520	1530	15
FEGSTCERN	IDDCPNHRCQ	NGGVCVDGVN	TYNCRCPPQW	TGQFCTEDVD	KTCKYDKYCA	DHFKDNHCDQ	GCNSEECGWD	GLDCAADQ
	EGF-like	8	11			NOD		1
310	320	330	340	350	1560	1570	1580	15
CLLQPNACQ	NGGTCANRNG	GYGCVCVNGW	SGDDCSENID	DCAFASCTPG	VLMPPEQLLQ	DARSFLRALG	TLLHTNLRIK	RDSQGELM
EGF-1	Like 9	11	EGF-lik	(e 10	1.61.0	1	1.600	NODP
360	370	380	390	400	1610	1620	1630	16
TCIDRVASE	SCMCPEGRAG	LLCHLDDACI	SNPCHKGALC	DINPLNGQII	KKQRMIKKSL	PGEQEQEVAG	SKVELEIDNK	QCVQDSDR
410	420	430	440	450	1660	1670	1680	16
TCPOGYKGA	DCTEDVDECA	MANSNPCEHA	GKCVNTDGAF	HCECLKGYAG	ASHATOGTLS	YPLVSVVSES	LTPERTOLLY	I.T.AVAVVI
11	EG	F-like 12		1 1		TM	Ditt goo	1
460	470	480	490	500	1710	1720	1730	17
RCEMDINEC	HSDPCQNDAT	CLDKIGGFTC	LCMPGFKGVH	CELEINECQS	KRKRKHGSLW	LPEGFTLRRD	ASNHKRREPV	GQDAVGLK
EGF-	like 13	1	EGF-lik	e 14	F	RAM		
510	520	530	540	550	1760	1770	1780	17
PCVNNGQCV	DKVNRFQCLC	PPGFTGPVCQ	IDIDDCSSTP	CLNGAKCIDH	GTGTSEHWVD	DEGPQPKKVK	AEDEALLSEE	DDPIDRRF
	11	EGF-	like 15		NI	S	1 1	
560	570	580	590	600	1810	1820	1830	18
NGTECQCAT	GFTGVLCEEN	CE-like 16	HGQCQDGIDS	TTCICNPGYM	RTPSLALTPP	QAEQEVDVLD	VNVRGP <mark>DGCT</mark>	PLMLASLE
610	620	630	640	650	1060	1970	1990	ANK 2
ATCSDOTDE	CYSSPCLNDG	RCIDLVNGYO	CNCOPGTSGV	NCEINEDDCA	AEDSSANITT	DLVXOGASLO	AOTDRTGEMA	I.HI.AARYS
TTTO PETER	or oper oper of	nor philipping	01102101000	HOL IN DOCK	ALLOUANTIT	PTAT COUPLY	AT DIVE OPPIN	Turner(10

EGF-like 17 1 1 EGF-like 18 1 1 1910 660 670 680 690 700 SNPCIHGICM DGINRYSCVC SPGFTGORCN IDIDEC ASNP CRKGATCING QDN<mark>MG</mark> R. | | 720 EGF-like 19 710 740 750 730 1960 VNGFRCICPE GPHHPSCYSQ VNECI LSNPCI GLSG YKCLCDAGWV EGF-like 20 11 11 I 760 770 800 780 790 2010 GINCEVDKNE CLSNP FKGY NCOVNIDECA ONGG IGYR MQDNKE EGF-like 21 EGF-like 22 11 830 810 820 840 850 2060 SNPCLNQGTC IGKNC OTV CENAAVCKE SGYTCH SPN EGF-like 23 11 870 890 860 880 900 2110 SPNFESYTCL GWQGQRC TIDIDECISK QGSYMCECP LCHN EGF-like 24 1 1 11 910 930 940 950 920 2160 PGFSGMDCEE IDDCLANPC QN CLPG FTGDKCQTDM MDGV EGF-like 25 11 980 960 970 990 1000 2210 NECLSEPCKN G ENNINE VNS QAGFD FNGG 6 | | 1020 1030 GFTGS FCL<mark>HEINECS</mark> EGF-like 27 EGF-like 26 1010 1030 1040 1050 2260 TYRCS INSES EGTC 11 EGF-like 28 1060 1070 1080 1090 1100 2310 CPLGYTGKNC LCSRS QCLCP <mark>aycd</mark>v EGF-like 29 11 I 1130 1110 1120 1140 1150 2360 PNVSCDIAAS RRGVLVEHLC CINAG HYCOCPLG YI (CE<mark>EQL</mark> EGF-like 30 11 EGF-like 31 1160 1170 1180 1190 1200 DECASNPCQH OFIGG YR VPGYQ EY<mark>EVDE</mark> CQNGG

EGF-like 32

1210 1220 1230 1240 1250 TCIDLVNHFK CSCPPGTRGL LCEENIDDCA RGPHCLNGGQ CMDRIGGYSC

11



Figure 3. Notch2 protein domains. Colors represent protein domains.

1250

1290

1340

1390

GSCH

LNR 1

CRCPPGFSGA

TNDYLC

1300

1400

AFTGRH

Т 1350

COSSCGOVK

QRQPPYYSC

1260	EGF-like 1270	32 1280	 1290	130
ESQPCQHGGQ	CRPSPGPGGG	LTFTCHCAQP	FWGPRCERVA	RSCRELQCP
EGF-1	ike 33	1	EGF-1:	ike 34
1310	1320	1330	1340	135
GVPCQQTPRG	PRCACPPGLS	GPSCRSFPGS	PPGASNASCA	AAPCLHGGS
1260	1270	1200	LNK 1	140
	CACAOCHTCP	DCEADAADE	UCEEDD CDDA	140
KPAPLAPT F K	CACAQGWIGP	I.NP	2	ACQARREDQ
1410	1420	1430	1440	145
CDRECNSPGC	GWDGGDCSLS	VGDPWROCEA	LOCWRLENNS	RCDPACSSP
		TODI III QULI	LNR	3
1460	1470	1480	1490	150
CLYDNFDCHA	GGRERTCNPV	YEKYCADHFA	DGRCDOGCNT	EECGWDGLD
			NOD	
1510	1520	1530	1540	155
ASEVPALLAR	GVLVLTVLLP	PEELLRSSAD	FLQRLSAILR	TSLRFRLDA
1		1	NO	DP
1560	1570	1580	1590	160
GQAMVFPYHR	PSPGSEPRAR	RELAPEV <mark>IGS</mark>	VVMLEIDNRL	CLQSPENDH
	1	T	м	
1610	1620	1630	1640	165
FPDAQSAADY	LGALSAVERL	DFPYPLRDVR	GEPLEPP EPS	VPLLPLLVA
1.000			1.000	170
1660	1670	1680	1690	170
AAPPPAATPAP	GVMVARRE	HST LWF PEGF	SLHKUVASGH	KORKEBAQŬ
1710	1720	1720	1740	175
	ECI MORUARD	LINDER OPENK	1740	175
ALGENNEARG	LOLMGEVAID	MMDIECPEAR	KERVEEPGHG	ALEAVDURU
1760	1770	1780	1790	180
TOHHLVAADI	RVAPAMALTP	POGDADADGM	DVNVRGPDGF	TPLMLASEC
2	NLS	S I	ANK 1	11211211010
1810	1820	1830	1840	185
GALEPMPTEE	DEADDTSASI	ISDLICQGAQ	LGAR TDR TGE	TALHLAARY.
	1 1		ANK 2	
1860	1870	1880	1890	190
RADAAKRLLD	AGADTNA QDH	SGRTPLHTAV	TADAQGVFQI	LIRNRSTDL
1 1	ANF	3	1 1	ANK 4
1910	1920	1930	1940	195
ARMADGSTAL	ILAARLAVEG	MVEELIASHA	DVNAVDELGK	SALHWAAAV
1	1	ANK 5		1
1960	1970	1980	1990	200
NVEATLALLK	NGANKDMQDS	NTC	REGSTEAAKL	LUHFANRE
2010	2020	2020	2040	205
TDHLDPL PPD	VAOERLHODT	VRLLDOPSCP	RSPPGPHGLG	PLLCPPGAF
1 Shabhar AD	CHARLEN IN COL	SGF SGF		. DDOLLOUL
2060	2070	2080	2090	210
PGLKAAOSGS	KKSRRPPGKA	GLGPQGPRGR	GKKLTLACPG	PLADSSVTL
2110	2120	2130	2140	215
PVDSLDSPRP	FGGPPASPGG	FPLEGPYAAA	TATAVSLAQL	GGPGRAGLG
		1		
2160	2170	2180	2190	220
QPPGGCVLSL	GLLNPVAVPL	DWARLPPPAP	PGPSFLLPLA	PGPQLLNPG
2210	2220	2230	2240	225
	YLAVPGHGEE	YPAAGAHSSP	PKARFLRVPS	EHPYLTPSP
		PEST	0000	000
0000	~~~~	2280	2290	230
2260	2270	DODAMAMONIA		
2260 SPEHWASPSP	2270 PSLSDWSEST	PSPATATGAM	ATTTGALPAQ	LTLT2AL22
2260 SPEHWASPSP 2310	2270 PSLSDWSEST	PSPATATGAM	ATTTGALPAQ	LPLPSAL22
2260 SPEHWASPSP 2310	2270 PSLSDWSEST 2320	PSPATATGAM	ATTTGALPAQ	EPEPSAL22

10	20	30	40	50
MGPGARGRRR	RRRPMSPPPP	PPPVRALPLL	LLLAGPGAAA	PPCLDGSPCA
EGF	-like 1	11	EGF-lik	e 2
60	70	80	90	100
NGGRCIQLES		EGF-lik	e 3	GVCQ33VVAG
110	120	130	140	150
TARFSCRCPR	GFRGPDCSLP	DPCLSSPCAH	GARCSVGPDG	RFLCSCPPGY
	170	EGF-like 4	100	1 1
OGRSCRSDVD	ECRYGEPCRH	GGTCLNTPGS	FRCOCPACYT	GPLCENPAVP
EGF-	like 5	1	EGF-lil	ke 6
210	220	230	240	250
CAPSPCRNGG	TCROSGDLTY	DCACLPGFEG	QNCEVNVDDC	PGHRCLNGGT
260	270	280	1Ke / 290	300
CVDGVNTYNC	QCPPEWTGQF	CTEDVDECQL	QPNACHNGGT	CFNTLGGHSC
1.1	EGF-1	like 8		1
310	320	330	340	350
VCVNGWTGES	CSQNIDDCAT	AVCFHGATCH	DRVASFYCAC	PMGKTGLLCH
360	370	380	390	400
LDDACVSNPC	HEDAICDTNP	VNGRAICTCP	PGFTGGACD Q	DVDE CS IGAN
EGF	-like 10	1 1	EGF-like	11
410	420	430	440	450
PCEHLGRCVN	TQGSFLCQCG	RGTTGPRCET EGE-	DVNECLSGPC	RNQATCLDRI
460	470	480	490	500
GQFTCICMAG	FTGT YC EVDI	DECQSSPCVN	GGV C KDRVNG	FSCTCPSGFS
	EGF	-like 13		
COTCOL DUDE	CASTROPACA	530	CRONECEECT	10000000
EGF	-like 14	I I	EGF-like	15
560	570	580	590	600
PDPCHHG R CV	DGIASFSCAC	APGYTGTRCE	S <mark>QVDECRSQP</mark>	CRHGGKCLDL
(10		EGF-1	ike 16	(50
VDKYLCRCPS	GTTGVNCEVN	TDDCASNPCT	FGVCRDGINR	YDCVCOPGET
1.1	EG	F-like 17		11
660				
000	670	680	690	700
GPLCNVEINE	670 CASSPCGEGG	680 SCVDGENGFR	690 CLCPPGSLPP	700 LCLPPSHPCA
GPLCNVEINE EGF-1 710	670 CASSPCGEGG ike 18 720	680 SCVDGENGFR I 730	690 CLCPPGSLPP EGF-113 740	700 LCLPPSHPCA ke 19 750
GPLCNVEINE EGF-1 710 HEPCSHGICY	670 CASSPCGEGG ike 18 720 DAPGGFRCVC	680 SCVDGENGFR I 730 EPGWSGPRCS	690 CLCPPGSLPP EGF-1i) 740 QSLARDACES	700 LCLPPSHPCA ke 19 750 QPCRAGGTCS
GPLCNVEINE EGF-1 710 HEPCSHGICY	670 CASSPCGEGG ike 18 720 DAPGGFRCVC 	680 SCVDGENGFR I 730 EPGWSGPRCS EGF-Li	690 CLCPPGSLPP EGF-li) 740 QSL <mark>ARDACES ke 20</mark>	700 LCLPPSHPCA ce 19 750 QPCRAGGTCS
GPLCNVEINE EGF-1 710 HEPCSHGICY 760	670 CASSPCGEGG ike 18 720 DAPGGFRCVC I I 770	680 SCVDGENGFR I 730 EPGWSGPRCS EGF-Li 780	690 CLCPPGSLPP EGF-111 740 QSLARDACES ke 20 790	700 LCLPPSHPCA ce 19 750 QPCRAGGTCS 800
GPLCNVEINE EGF-1 710 HEPCSHGICY 760 SDGMGFHCTC	670 CASSPCGEGG ike 18 720 DAPGGFRCVC I I 770 PPGVQGRQCE	680 SCVDGENGFR I 730 EPGWSGPRCS EGF-1i 780 LLSPCTPNPC EGF-1ike 2	690 CLCPPGSLPP EGF-1iJ 740 QSLARDACES ke 20 790 EHGGRCESAP 1	700 LCLPPSHPCA ce 19 750 QPCRAGGTCS 800 GQLPVCSCPQ
GPLCNVEINE EGF-1 710 HEPCSHGICY 760 SDGMGFHCTC I I 810	670 CASSPCGEGG ike 18 720 DAPGGFRCVC I I 770 PPGVQGRQCE 820	680 SCVDGENGFR I 730 EPGWSGPRCS EGF-li 780 LLSPCTPNPC EGF-like 2 830	690 CLCPPGSLPP EGF-1iJ 740 QSLARDACES ke 20 790 EHGGRCESAP 1 840	700 LCLPPSHPCA ce 19 750 QPCRAGGTCS 800 GQLPVCSCPQ I I 850
GPLCNVEINE EGF-1 710 HEPCSHGICY 760 SDGMGFHCTC I I 810 GWQGPRCQQD	670 CASSPCGEGG ike 18 720 DAPGGFRCVC I I 770 PPGVQGRQCE 820 VDECAGPAPC	680 SCVDGENGFR I 730 EPGWSGPRCS EGF-1i EGF-1ike 2 830 GPHGICTNLA	690 CLCPPGSLPP EGF-1il 740 QSLARDACES ke 20 790 EHGGRCESAP 1 840 GSFSCTCHGG	700 LCLPPSHPCA (e 19 750 QPCRAGGTCS 800 GQLPVCSCPQ I I 850 YTGPSCDQDI
GPLCNVEINE EGF-1 710 HEPCSHGICY 760 SDGMGFHCTC I I 810 GWQGPRCQQD EGF-1	670 CASSPCGEGG ite 18 720 DAPGGFRCVC I 1 770 PPGVQGRQCE 820 VDECAGPAPC ite 22	680 SCVDGENGFR I 730 EPGWSGFRCS EGF-1i 780 LLSPCTPNPC EGF-1ike 2 830 GPHGICTNLA	690 CLCPPGSLPP EGF-1i1 740 QSLARDACES ke 20 790 EHGGRCESAP 1 840 GSFSCTCHGG EGF-1i1	700 LCL PSHPCA cc 19 750 QPCRAGGTCS 800 GQLPVCSCPQ I I 850 YTGPSCDQDI cc 23
GPLCNVEINE EGF-1 710 HEPCSHGICY 760 SDGMGFHCTC I I 810 GWQGPRCQQD EGF-1 860 NDCDPNCLN	670 CASSPCGEGG ike 18 720 DAPGGFRCVC I I 770 PPGVQGRQCE 820 VDECAGPAPC ike 22 870 CGSCODEVGS	680 SCVDGENGFR I 730 EPGWSCFRCS EGF-1i 780 LLSPCTPNPC EGF-1ike 2 830 GPHGICTNLA 880 FSCSCLPCFA	690 CLCPPGSLPP EGF-1i1 740 QSLARDACES ke 20 790 EHGGRCESAP 1 840 GSFSCTCHGG EGF-1i1 890 GPBC&P DVDE	700 LCL PSHPCA cc 19 750 QPCRAGGTCS 800 GQLPVCSCPQ I I 850 YTGPSCDQDI cc 23 900 CLENPCCBCT
GPLCNVEINE EGF-1 710 HEPCSHGICY 760 SDGMGFHCTC I I 810 GWQGPRCQQD EGF-1 860 NDCDPNPCLN	670 CASSPCGEGG ite 18 720 DAPGGFRCVC I 1 770 PPGVQGRQCE 820 VDECAGPAPC ite 22 870 GGSCQDGVGS I 1	680 SCVDGENGFR I 730 EPGWSCFRCS EGF-1i 80 LLSPCTPNPC EGF-1ike 2 830 GPHGICTNLA 880 FSCSCLPGFA EGF-1i	690 CLCPPGSLPP EGF-1i1 740 QSLARDACES ke 20 790 EHGGRCESAP 1 840 GSFSCTCHGG EGF-1i1 890 GPRCARDVDE ke 24	700 LCL PSHPCA cc 19 750 QPCRAGGTCS 800 GQLPVCSCPQ I I 850 YTGPSCDQDI cc 23 900 CLSNPCGPGT
GPLCNVEINE EGF-1 710 HEPCSHGICY 760 SDGMGFHCTC I I 810 GWQGPRCQD EGF-1 860 NDCDPNPCLN 910	670 CASSPCGEGG ite 18 720 DAPGGFRCVC I 1 770 PPGVQGRQCE 820 VDECAGPAPC ite 22 870 GGSCQDGVGS I 1 920	680 SCVDGENGFR I 730 EPGWSCFRCS EGF-1ik 830 GPHGICTNLA 880 FSCSCLPGFA EGF-1i 930	690 CLCPPGSLPP EGF-1i1 740 QSLARDACES ke 20 790 EHGGRCESAP 1 840 GSFSCTCHGG EGF-1i1 890 GPRCARDVDE ke 24 940	700 LCL PSHPCA cc 19 750 QPCRAGGTCS 800 GQLPVCSCPQ I I 850 YTGPSCDQDI cc 23 900 CLSNPCGPGT 950
GPLCNVEINE EGF-1 710 HEPCSHGICY 760 SDGMGFHCTC I I 810 GWQGPRCQQD EGF-1 860 NDCDPNPCLN 910 CTDHVASFTC	670 CASSPCGEGG ite 18 720 DAPGGFRCVC I 1 770 PPGVQGRQCE 820 VDECAGPAPC ite 22 870 GGSCQDGVGS I 1 920 TCPPGYGGFH	680 SCVDGENGFR I 730 EPGWSCFRCS EGF-1ik 830 GPHGICTNLA 880 FSCSCLPGFA EGF-1i 930 CEQDLPDCSP	690 CLCPPGSLPP EGF-1i1 740 QSLARDACES ke 20 790 EHGGRCESAP 1 840 GSFSCTCHGG EGF-1i1 890 GPRCARDVE ke 24 940 SSCFNGGTCV	700 LCL PSHPCA ce 19 750 QPCRAGGTCS 800 GQLPVCSCPQ I I 850 YTGPSCDQDI ce 23 900 CLSNPCGPGT 950 DGVNSFSCLC
GPLCNVEINE EGF-1 710 HEPCSHGICY 760 SDGMGFHCTC I I 810 GWQGPRCQQD EGF-1 860 NDCDPNPCLN 910 CTDHVASFTC I I 960	670 CASSPCGEGG ike 18 720 DAPGGFRCVC I 1 770 PPGVQGRQCE 820 VDECAGPAPC ike 22 870 GGSCQDGVGS I 1 920 TCPPGYGGFH	680 SCVDGENGFR I 730 EPGWSCFRCS EGF-1ik 780 LLSPCTPNPC EGF-1ike 2 830 GPHGICTNLA 880 FSCSCLPGFA EGF-1i 930 CEQDLPDCSP EGF-1ike 2 980	690 CLCPPGSLPP EGF-1i1 740 QSLARDACES ke 20 790 EHGGRCESAP 1 840 GSFSCTCHGG EGF-1i1 890 GPRCARDVDE ke 24 940 SSCFNGGTCV 5 990	700 LCL PSHPCA ce 19 750 QPCRAGGTCS 800 GQLPVCSCPQ I I 850 YTGPSCDQDI ce 23 900 CLSNPCGPGT 950 DGVNSFSCLC I I 1000
GPLCNVEINE EGF-1 710 HEPCSHGICY 5DGMGFHCTC I I 810 GWQGPRCQQD EGF-1 860 NDCDPNPCLN 910 CTDHVASFTC I I 960 RPGYTGAHCQ	670 CASSPCGEGG ike 18 720 DAPGGFRCVC I 1 770 PPGVQGRQCE 820 VDECAGPAPC ike 22 870 GGSCQDGVGS I 1 920 TCPPGYGGFH 970 HEADPCLSRP	680 SCVDGENGFR I 730 EPGWSCFRCS EGF-1ik 780 LLSPCTPNPC EGF-1ike 2 830 GPHGICTNLA 880 FSCSCLPGFA EGF-1i 930 CEQDLPDCSP EGF-1ike 2 980 CLHGGVCSAA	690 CLCPPGSLPP EGF-1i1 740 QSLARDACES ke 20 790 EHGGRCESAP 1 890 GSFSCTCHGG EGF-1i1 890 GPRCARDVDE ke 24 940 SSCFNGGTCV 5 990 HPGFRCTCLE	700 LCL PSHPCA ce 19 750 QPCRAGGTCS 800 GQLPVCSCPQ I I 850 YTGPSCDQDI ce 23 900 CLSNPCGPGT 950 DGVNSFSCLC I I 1000 SFTGPQCQTL
GPLCNVEINE EGF-1 710 HEPCSHGICY 5DGMGFHCTC I I 810 GWQGPRCQQD EGF-1 860 NDCDPNPCLN 910 CTDHVASFTC I I 960 RPGYTGAHCQ EGF-	670 CASSPCGEGG ike 18 720 DAPGGFRCVC I 1 770 PPGVQGRQCE 820 VDECAGPAPC ike 22 870 GGSCQDGVGS I 1 920 TCPPGYGGFH 970 HEADPCLSRP -1ike 26	680 SCVDGENGFR 1 730 EPGWSCFRCS EGF-1it 780 LLSPCTPNPC EGF-1ike 2 880 FSCSCLPGFA EGF-1ike 2 980 CLHGGVCSAA	690 CLCPPGSLPP EGF-1i1 740 QSLARDACES ke 20 790 EHGGRCESAP 1 890 GFFSCTCHGG EGF-1i1 890 GPRCARDVDE ke 24 940 SSCFNGGTCV 5 990 HPGFRCTCLE EGF-1ike	700 LCL PSHPCA ce 19 750 QPCRAGGTCS 800 GQLPVCSCPQ I I 850 YTGPSCDQDI ce 23 900 CLSNPCGPGT 950 DGVNSFSCLC I I 1000 SFTGPQCQTL 27
GPLCNVEINE EGF-1 710 HEPCSHGICY 5DGMGFHCTC I I 810 GWQGPRCQQD EGF-1 860 NDCDPNPCLN 910 CTDHVASFTC I I 960 RPGYTGAHCQ EGF- 1010	670 CASSPCGEGG ike 18 720 DAPGGFRCVC I 1 770 PPGVQGRQCE 820 VDECAGPAPC ike 22 870 GGSCQDGVGS I 1 920 TCPPGYGGFH 970 HEADPCLSRP -1ike 26 1020	680 SCVDGENGFR 1 730 EPGWSCFRCS EGF-1it 780 LLSPCTPNPC EGF-1ike 2 830 GPHGICTNLA 880 FSCSCLPGFA EGF-1ik 930 CEQDLPDCSP EGF-1ike 2 980 CLHGGVCSAA 1030 CLHGGVCSAA	690 CLCPPGSLPP EGF-1i1 740 QSLARDACES ke 20 790 EHGGRCESAP 1 840 GSFSCTCHGG EGF-1i1 890 GPRCARDVE ke 24 940 SSCFNGGTCV 5 990 HPGFRCTCLE EGF-1ike 1040	700 LCL PSHPCA cc 19 750 QPCRAGGTCS 800 GQLPVCSCPQ I I 850 YTGPSCDQDI cc 23 900 CLSNPCGPGT 950 DGVNSFSCLC I I 1000 SFTGPQCQTL 27
GPLCNVEINE EGF-1 710 HEPCSHGICY SDGMGFHCTC I I 810 GWQGPRCQQD EGF-1 860 NDCDPNPCLN 910 CTDHVASFTC I I 960 RPGYTGAHCQ EGF- 1010 VDWCSRQPCQ	670 CASSPCGEGG ike 18 720 DAPGGFRCVC I 1 770 PPGVQGRQCE 820 VDECAGPAPC ike 22 870 GGSC0DGVGS I 1 920 TCPPGYGGFH 970 HEADPCLSRP -1ike 26 1020	680 SCVDGENGFR 1 730 EPGWSCFRCS EGF-1it 780 LLSPCTPNPC EGF-1ike 2 830 GPHGICTNLA 880 FSCSCLPGFA EGF-1ike 2 930 CEQDLPDCSP EGF-1ike 2 980 CLHGGVCSAA 1030 YCLCPPGWSG EGF-1ike 2	690 CLCPPGSLPP EGF-1i1 740 QSIARDACES ke 20 790 EHGGRCESAP 1 890 GSFSCTCHGG EGF-1i1 890 GPRCARDVDE ke 24 940 SSCFNGGTCV 5 990 HPGFRCTCLE EGF-1ike 1040 RLCDIRSLPC 28	700 LCL PSHPCA ce 19 750 QPCRAGGTCS 800 GQLPVCSCPQ I I 850 YTGPSCDQDI ce 23 900 CLSNPCGPGT 950 DGVNSFSCLC I I 1000 SFTGPQCQTL 27 1050 REAAAQIGVR
GPLCNVEINE EGF-1 710 HEPCSHGICY SDGMGFHCTC I I 800 GWQGPRCQQD EGF-1 860 NDCDPNPCLN 910 CTDHVASFTC I I 960 RPGYTGAHCQ EGF- 1010 VDWCSRQPCQ 1060	670 CASSPCGEGG ike 18 720 DAPGGFRCVC I 1 770 PPGVQGRQCE 820 VDECAGPAPC ike 22 870 GGSCQDGVGS I 1 920 TCPPGYGGFH 970 HEADPCLSRP -1ike 26 1020 NGGRCVQTGA I 1070	680 SCVDGENGFR 1 730 EPGWSCFRCS EGF-1it 780 LLSPCTPNPC EGF-1ike 2 830 GPHGICTNLA 880 FSCSCLPGFA EGF-1ik 930 CLCPCFGWSG 1030 YCLCPPCWSG 1080	690 CLCPPGSLPP EGF-1i1 740 QSLARDACES ke 20 790 EHGGRCESAP 1 840 GSFSCTCHGG EGF-1i1 890 GPRCARDVDE ke 24 940 SSCFNGGTCV 5 990 HPGFRCTCLE EGF-1ike 1040 RLCDIRSLPC 28 1090	700 LCL PPSHPCA (ce 19 750 QPCRAGGTCS 800 GQLPVCSCPQ I I 850 YTGPSCDQDI (ce 23 900 CLSNPCGPGT 950 DGVNSFSCLC I 1 1000 SFTGPQCQTL 27 1050 REAAAQIGVR 1100
GPLCNVEINE GPLCNVEINE EGF-1 710 HEPCSHGICY 760 SDGMGFHCTC I I 810 GWQGPRCQQD EGF-1 860 NDCDPNPCLN 910 CTDHVASFTC I I 960 RPGYTGAHCQ EGF- 1010 VDWCSRQPCQ 1060 LEQLCQAGGQ	670 CASSPCGEGG ike 18 720 DAPGGFRCVC I 1 770 PPGVQGRQCE 820 VDECAGPAPC ike 22 870 GGSCQDGVGS I 1 920 TCPPGYGGFH 970 HEADPCLSRP -like 26 1020 NGGRCVQTGA I 1070 CVDEDSSHYC	680 SCVDGENGFR 1 730 EPGWSCFRCS EGF-1it 780 LLSPCTPNPC EGF-1ike 2 830 GPHGICTNLA 880 FSCSCLPGFA EGF-1ik 930 CEQDLPDCSP EGF-1ike 2 980 CLHGGVCSAA 1030 YCLCPPGWSG EGF-1ike 2 1080 VCPEGRTGSH	690 CLCPPGSLPP EGF-1i1 740 QSLARDACES ke 20 790 EHGGRCESAP 1 840 GSFSCTCHGG EGF-1i1 890 GPRCARDVDE ke 24 940 SSCFNGGTCV 5 990 HPGFRCTCLE EGF-1ike 1040 RLCDIRSLPC 28 1090 CEQEVDPCLA	700 LCL PPSHPCA (ce 19 750 QPCRAGGTCS 800 GQLPVCSCPQ I 850 YTGPSCDQDI (ce 23 900 CLSNPCGPGT 950 DGVNSFSCLC I 1 1000 SFTGPQCQTL 27 1050 REAAAQIGVR 1100 QPCQHGGTCR
GPLCNVEINE GPLCNVEINE EGF-1 710 HEPCSHGICY 760 SDGMGFHCTC I I 810 GWQGPRCQQD EGF-1 860 NDCDPNPCLN 910 CTDHVASFTC I I 960 RPGYTGAHCQ EGF- 1010 VDWCSRQPCQ 1060 LEQLCQAGGQ	670 CASSPCGEGG ike 18 720 DAPGGFRCVC I I 770 PPGVQGRQCE 820 VDECAGPAPC ike 22 870 GGSCQDGVGS I I 920 TCPPGYGGFH 970 HEADPCLSRP -like 26 1020 NGGRCVQTGA I 070 CVDEDSSHYC I 1 1070	680 SCVDGENGFR 1 730 EPGWSCFRCS EGF-1ik 780 LLSPCTPNPC EGF-1ike 2 830 GPHGICTNLA 880 FSCSCLPGFA EGF-1ike 2 930 CLHGGVCSAA 1030 VCLCPPGWSG EGF-1ike 2 1080 VCPEGRTGSH EGF-1	690 CLCPPGSLPP EGF-1i1 740 QSIARDACES ke 20 790 EHGGRCESAP 1 840 GSFSCTCHGG EGF-1i1 890 GPRCARDVDE ke 24 940 SSCFNGGTCV 5 990 HPGFRCTCLE EGF-1ike 1040 RLCDIRSLPC 28 1090 CE0EVDPCLA ike 29	700 LCL PSHPCA ce 19 750 QPCRAGGTCS 800 GQLPVCSCPQ I I 850 YTGPSCDQDI ce 23 900 CLSNPCGPGT 950 DGVNSFSCLC I I 1000 SFTGPQCQTL 27 1050 REAAAQIGVR 1100 QPCQHGGTCR
GPLCNVEINE GPLCNVEINE EGF-1 710 HEPCSHGICY 760 SDGMGFHCTC I I 810 GWQGPRCQQD EGF-1 860 NDCDPNPCLN 910 CTDHVASFTC I I 960 RPGYTGAHCQ EGF- 1010 VDWCSRQPCQ 1060 LEQLCQAGGQ 1110	670 CASSPCGEGG ike 18 720 DAPGGFRCVC I I 770 PPGVQGRQCE 820 VDECAGPAPC ike 22 870 GGSCQDGVGS I I 920 TCPPGYGGFH 970 HEADPCLSRP -like 26 1020 NGGRCVQTGA I 1070 CVDEDSSHYC I 1 1120	680 SCVDGENGFR 1 730 EPGWSCFRCS EGF-1ik 22 830 GPHGICTNLA 880 FSCSCLPGFA EGF-1ik 22 930 CEQDLPDCSP EGF-1ike 2 980 CLHGGVCSAA 1030 VCLCPPGWSG EGF-1ike 1 1080 VCPEGRTGSH EGF-1 1130	690 CLCPPGSLPP EGF-1i1 740 QSLARDACES ke 20 790 EHGGRCESAP 1 840 GSFSCTCHGG EGF-1i1 890 GPRCARDVDE ke 24 940 SSCFNGGTCV 5 990 HPGFRCTCLE EGF-1ike 1040 RLCDIRSLPC 28 1090 CEQEVDPCLA ike 29 1140	700 LCL PSHPCA ce 19 750 QPCRAGGTCS 800 GQLPVCSCPQ I I 850 YTGPSCDQDI ce 23 900 CLSNPCGPGT 950 DGVNSFSCLC I I 1000 SFTGPQCQTL 27 1050 REAAAQIGVR 1100 QPCQHGGTCR 1150
GPLCNVEINE EGF-1 710 HEPCSHGICY SDGMGFHCTC I I 810 GWQGPRCQQD EGF-1 860 NDCDPNPCLN 910 CTDHVASFTC I I 960 RPGYTGAHCQ EGF 1010 VDWCSRQPCQ 1060 LEQLCQAGGQ 1110	670 CASSPCGEGG ike 18 720 DAPGGFRCVC I 1 770 PPGVQGRQCE 820 VDECAGPAPC ike 22 870 GGSCQDGVGS I 1 920 TCPPGYGGFH 970 HEADPCLSRP -like 26 1020 NGGRCVQTGA I 1070 CVDEDSSHYC I 1 1120 LPGYNGDNCE I 1	680 SCVDGENGFR 1 730 EPGWSCFRCS EGF-1ik 780 LLSPCTPNPC EGF-1ike 2 880 FSCSCLPGFA EGF-1ike 2 980 CLHGGVCSAA 1030 VCLCPPGWSG I EGF-1ike 1 1080 VCPEGRTGSH EGF-1 1130	690 CLCPPGSLPP EGF-1i1 740 QSIARDACES ke 20 790 EHGGRCESAP 1 890 GFSCTCHGG EGF-1i1 890 GFSCTCHGG EGF-1i8 990 HPGFRCTCLE EGF-1ike 1040 RLCDIRSLPC 28 1090 CEQEVDPCLA ike 29 1140	700 LCL PSHPCA ce 19 750 QPCRAGGTCS 800 GQLPVCSCPQ I I 850 YTGPSCDQDI ce 23 900 CLSNPCGPGT 950 DGVNSFSCLC I I 1000 SFTGPQCQTL 27 1050 REAAAQIGVR 1100 QPCQHGGTCR 1150 VARYLCSCPP
GPLCNVEINE EGF-1 710 HEPCSHGICY SDGMGFHCTC I I 800 GWQGPRCQQD EGF-1 860 NDCDPNPCLN 910 CTDHVASFTC I I 960 RPGYTGAHCQ EGF 1010 VDWCSRQPCQ 1060 LEQLCQAGGQ 1110 GYMGGYMCEC 1160	670 CASSPCGEGG ike 18 720 DAPGGFRCVC I 1 770 PPGVQGRQCE 820 VDECAGPAPC ike 22 870 GGSCQDGVGS I 1 920 TCPPGYGGFH 970 HEADPCLSRP -like 26 1020 NGGRCVQTGA I 1070 CVDEDSSHYC I 1 1120	680 SCVDGENGFR 1 730 EPGWSCFRCS EGF-1ik 22 830 GPHGICTNLA 880 FSCSCLPGFA EGF-1ik 23 930 CEQDLPDCSP EGF-1ike 2 980 CLHGGVCSAA 1030 VCLCPPGWSG I GGF-1ike 1 1080 VCPEGRTGSH EGF-1 1130	690 CLCPPGSLPP EGF-1i1 740 QSIARDACES ke 20 790 EHGGRCESAP 1 840 GSFSCTCHGG EGF-1i1 890 GPRCARDVDE ke 24 940 SSCFNGGTCV 5 990 HPGFRCTCLE EGF-1ike 1040 RLCDIRSLPC 28 1090 CE0EVDPCLA ike 29 1140	700 LCL PPSHPCA ce 19 750 QPCRAGGTCS 800 GQLPVCSCPQ I I 850 YTGPSCDQDI ce 23 900 CLSNPCGPGT 950 DGVNSFSCLC I I 1000 SFTGPQCQTL 27 1050 REAAAQIGVR 1100 QPCQHGGTCR 1150 VARYLCSCPP 1200
GPLCNVEINE GPLCNVEINE EGF-1 710 HEPCSHGICY 760 SDGMGFHCTC I I 810 GWQGPRCQQD EGF-1 860 NDCDPNPCLN 910 CTDHVASFTC I I 960 RPGYTGAHCQ EGF- 1010 VDWCSRQPCQ 1060 LEQLCQAGGQ 1110 GYMGGYMCEC 1160 GTLGVLCEIN	670 CASSPCGEGG ike 18 720 DAPGGFRCVC I 1 770 PPGVQGRQCE 820 VDECAGPAPC ike 22 870 GGSCQDGVGS I 1 920 TCPPGYGGFH 970 HEADPCLSRP -like 26 1020 NGGRCVQTGA I 1070 CVDEDSSHYC I 1 1070 CVDEDSSHYC I 1 1170 EDDCGPCPEL	680 SCVDGENGFR 1 730 EPGWSCFRCS EGF-1ik 780 LLSPCTPNPC EGF-1ike 2 880 FSCSCLPGFA EGF-1ike 2 980 CLHGGVCSAA 1030 VCLCPPGWSG I EGF-1ike 1 1080 VCPEGRTGSH EGF-1 1180 DSGFRCLHNG	690 CLCPPGSLPP EGF-1i1 740 QSIARDACES ke 20 790 EHGGRCESAP 1 840 GSFSCTCHGG EGF-1i1 890 GPRCARDVE ke 24 940 SSCFNGGTCV 5 990 HPGFRCTCLE EGF-1ike 1040 RLCDIRSLPC 28 1090 CEQEVDPCLA ike 29 1140 CQHGGSCIDL ike 30 1190 TCVDLVGGFR	700 LCL PSHPCA ce 19 750 QPCRAGGTCS 800 GQLPVCSCPQ I I 850 YTGPSCDQDI ce 23 900 CLSNPCGPGT 950 DGVNSFSCLC I I 1000 SFTGPQCQTL 27 1050 REAAAQIGVR 1100 QPCQHGGTCR 1150 VARYLCSCPP 1200 CTCPPGYTGL
GPLCNVEINE GPLCNVEINE EGF-1 710 HEPCSHGICY 760 SDGMGFHCTC I I 810 GWQGPRCQQD EGF-1 860 NDCDPNPCLN 910 CTDHVASFTC I I 960 RPGYTGAHCQ EGF- 1010 VDWCSRQPCQ 1060 LEQLCQAGGQ 1110 GYMGGYMCEC 1160 GTLGVLCEIN I I 1210	670 CASSPCGEGG ike 18 720 DAPGGFRCVC I 1 770 PPGVQGRQCE 820 VDECAGPAPC ike 22 870 GGSC0DGVGS I 1 920 TCPPGYGGFH 970 HEADPCLSRP -like 26 1020 NGGRCVQTGA I 1070 CVDEDSSHYC I 1 1170 EDDCGPCGPEL 1220	680 SCVDGENGFR 1 730 EPGWSCFRCS EGF-1ik 2 830 GPHGICTNLA 880 FSCSCLPGFA EGF-1ik 2 930 CEQDLPDCSP EGF-1ike 2 980 CLHGGVCSAA 1030 VCLCPPGWSG I EGF-1 1130 DVDECASQP EGF-1 1130 DSGPRCLHNG 1230	690 CLCPPGSLPP EGF-1i1 740 QSIARDACES ke 20 790 EHGGRCESAP 1 840 GSFSCTCHGG EGF-1i1 890 GPRCARDVDE ke 24 940 SSCFNGGTCV 5 990 HPGFRCTCLE EGF-1ike 1040 RLCDIRSLPC 28 1090 CEQEVDPCLA ike 29 1140 CQHGGSCIDL ike 30 1190 TCVDLVGGFR	700 LCL PPSHPCA ce 19 750 QPCRAGGTCS 800 GQLPVCSCPQ I 850 YTGPSCDQDI ce 23 900 CLSNPCGPGT 950 DGVNSFSCLC I 1 1000 SFTGPQCQTL 27 1050 REAAAQIGVR 1100 QPCQHGGTCR 1150 VARYLCSCPP 1200 CTCPPGYTGL I 1 1250

Figure 4. Notch3 protein domains. Amino acids marked with bold red represent the CADASIL mutations. Colors represent protein domains.

	NOD		Ĩ	
1310	1320	1330	1340	1350
VVLSPPALDQ	QLFALARVLS	LTLRVGLWVR	KDRDGRDMVY	PYPGARAEEK
	1	N	ODP	
1360	1370	1380	1390	1400
LGGTRDPTYQ	ERAAPQTQPL	GKETDSLSAG	FVVVMGVDLS	RCGPDHPASR
1410	1 1400	1420	1440	1450
	1420	1430	144U	DANOT PHOUL
TM	I LAAPAAVGA	PELPERGEPP	AVHPHAGIAP	PANQLEWEVL
1460	1470	1480	1490	1500
CSPVAGVILL	ALGALLVLOL	IRRRREHGA	LWLPPGFTRR	PRTOSAPHRR
		1		
1510	1520	1530	1540	1550
RPPLGEDSIG	LKALKPKAEV	DEDGVVMCSG	PEEGEEVGQA	EETGPPSTCQ
1560	1570	1580	1590	1600
LWSLSGGCGA	LPQAAMLTPP	QESEMEAPDL	DTRGPDGVTP	LMSAVCCGEV
1	NLS	1	ANK	1
1610	1620	1630	1640	1650
QSGTFQGAWL	GCPEPWEPLL	DGGACPQAHT	VGTGETPLHL	AARFSRPTAA
11		ANK 2		11
1000	1 6 7 0	1 (0 0	1	1700
1660	1670	1680	1690	1700
1660 RRLLEAGANP	1670 NOPDRAGRTP	1680 LHAAVAADAR	1690 EVCQLLLRSR	1700 QTAVDART <mark>E</mark> D
1660 RRLLEAGANP	1670 NQPDRAGRTP ANK 3 1720	1680 LHAAVAADAR 1730	1690 EVCQLLLRSR ANK 4 1740	1700 OTAVDARTED 1750
1660 RRLLEAGANP 1710 GTTPLMLAAR	1670 NQPDRAGRTP ANK 3 1720 LAVEDLVEEL	1680 LHAAVAADAR II 1730 LAAOADVGAR	1690 EVCQLLLRSR ANK 4 1740 DKWGKTALHW	1700 QTAVDARTED 1750 AAAVNNARAA
1660 RRLLEAGANP 1710 GTTPLMLAAR	1670 NOPDRAGRTP ANK 3 1720 LAVEDLVEEL	1680 LHAAVAADAR II 1730 IAAQADVGAR ANK 5	1690 EVCQLLLRSR ANK 4 1740 DKWGKTALHW	1700 OTAVDARTED 1750 AAAVNNARAA
1660 RRLLEAGANP 1710 GTTPLMLAAR I 1760	1670 NOPDRAGRTP ANK 3 1720 LAVEDLVEEL 1770	1680 LHAAVAADAR II 1730 IAAQADVGAR ANK 5 1780	1690 EVCQLLLRSR ANK 4 1740 DKWGKTALHW 1790	1700 QTAVDARTED 1750 AAAVNNARAA I 1800
1660 RRLLEAGANP 1710 GTTPLMLAAR I 1760 RSLLQAGADK	1670 NOPDRAGRTP ANK 3 1720 LAVEDLVEEL 1770 DAODNREOTP	1680 LHAAVAADAR 11 1730 IAAQADVGAR ANK 5 1780 LFLAAREGAV	1690 EVCQLLLRSR ANK 4 1740 DKWGKTALHW 1790 EVAQLLLGLG	1700 QTAVDARTED 1750 AAAVNNARAA I 1800 AARELRDQAG
1660 RELLEAGANP 1710 GTTPLMLAAR I 1760 RSLLOAGADK	1670 NOPDRAGRTP ANK 3 1720 LAVEDLVEEL 1770 DAQDNREQTP I	1680 LHAAVAADAR 11 1730 IAAQADVGAR ANK 5 1780 LFLAAREGAV NLS	1690 EVCQLLLRSR ANK 4 1740 DKWGKTALHW 1790 EVAQLLLGLG I	1700 QTAVDARTED 1750 AAAVNNARAA I 1800 AARELRDOAG
1660 RELLEAGANP 1710 GTTPLMLAAR II 1760 RSLLOAGADK 1810	1670 NQPDRAGRTP ANK 3 1720 LAVEDLVEEL 1770 DAQDNREQTP I 1820	1680 LHAAVAADAR I 1730 IAAQADVGAR ANK 5 1780 LFLAAREGAV NLS 1830	1690 EVCOLLLRSR ANK 4 1740 DKWGKTALHW 1790 EVAOLLLGLG I 1840	1700 OTAVDARTED 1750 AAAVNNARAA I 1800 AARELRDOAG 1850
1660 RELLEAGANP 1710 GTTPLMLAAR II 1760 RSLLQAGADK 1810	1670 NOPDRAGRTP ANK 3 1720 LAVEDLVEEL 1770 DAQDNREQTP I 1820 NHWDILTUL	1680 LHAAVAADAR I 1730 IAAQADVGAR ANK 5 1780 LFLAAREGAV NLS 1830 GAGPPEARHK	1690 EVCOLLESE ANK 4 1740 DKWGKTALHW 1790 EVAQLELGE I 1840 ATPGREAGPF	1700 OTAVDARTED 1750 AAAVNNARAA I 1800 AARELRDOAG 1850 PRARTVSVSV
1660 RRLLEAGANP 1710 GTTPLMLAAR 1760 RSLLQAGADK 1810 LAPADVAHQR	1670 NOPDRAGRTP ANK 3 1720 LAVEDLVEEL 1770 DAODNREQTP I 1820 NHWDLLTLLE	1680 LHAAVAADAR I 1730 IAAQADYGAR ANK 5 1780 LFLAAREGAV NLS 1830 GAGPPEARHK I	1690 EVCOLLERSE ANK 4 1740 EKWGKTALHW 1790 EVAQLLEG I 1840 ATPGREAGPF	1700 CTAVDARTE 1750 AAAVNNARAA 1 1800 AARELRDO AG 1850 PRARTVSVSV
1660 RRLLEAGANP 1710 GTTPLMLAAR 1760 RSLLQAGADK 1810 LAPADVAHOR 1860	1670 NOPDRAGRTP ANK 3 1720 LAVEDLVEEL 1770 DAQDNREQTP 1 1820 NHNDLLTLIE 1870	1680 LHAAVAADAR I 1730 IAAQADVGAR ANK 5 1780 LFLAAREGAV NLS 1830 GAGPPEARHK I 1880	1690 EVCOLLERSE ANK 4 1740 DKWGKTALHW 1790 EVAOLLIGIG 1 1840 ATPGREAGPF 1890	1700 OTAVDARTE 1750 AAAVNNARAA I 1800 AARELRD AG 1850 PRARTVSVSV 1900
1660 RRLLEAGANP 1710 GTTPLMLAAR 1760 RSLLQAGADK 1810 LAPADVAFOR 1860 PPHGGGALPR	1670 NOPDRAGRTP ANK 3 1720 LAVEDLVEEL 1770 DAQDNREQTP 1 1820 NHWDLLTLL 1870 CRTLSAGAGP	1680 LHAAVAADAR II 1730 IAAQADVGAR ANK 5 1780 LFIAAREGAV NLS 1830 GAGPPEARHK I 1880 RGGGACLQAR	1690 EVCOLLERSE ANK 4 1740 DKWGKTALHW 1790 EVAQULLIGIG I 1840 ATPGREAGPF 1890 TWSVDLAARG	1700 OTAVDARTE 1750 AAAVNNARAA I 1800 AARELRDO AG 1850 PRARTVSVSV 1900 GGAYSHCRS
1660 RRLLEAGANP 1710 GTTPLMLAAR 1760 RSLLQAGADK 1810 LAPADVAFOR 1860 PPHGGGALPR	1670 NOPDRAGRTP ANK 3 1720 LAVEDLVEEL 1770 DAQDNREQTP 1 1820 NHWDLLTLL 1870 CRTLSAGAGP	1680 LHAAVAADAR II 1730 IAAQADVGAR ANK 5 1780 LFIAAREGAV NLS 1830 GAGPPEARHK I 1880 RGGGACLQAR PEST	1690 EVCOLLERSE ANK 4 1740 DKWGKTALHW 1790 EVAQLILIGIG I 1840 ATPGREAGPF 1890 TWSVDLAARG	1700 OTAVDARTE 1750 AAAVNNARAA I 1800 AARELRDO AG 1850 PRARTVSVSV 1900 GGAYSHCRS
1660 RRLLEAGANP 1710 GTTPLMLAAR 1760 RSLLQAGADK 1810 APADVAHQR 1860 PPHGGGALPR 1910	1670 NOPDRAGRTP ANK 3 1720 LAVEDLVEEL 1770 DAQDNREQTP I 1820 NHWDILTIL 1870 CRTLSAGAGP 1920	1680 LHAAVAADAR II 1730 IAAQADVGAR ANK 5 1780 LFLAAREGAV NLS 1830 GAGPPEARHK I 1880 RGGGACLQAR PEST 1930	1690 EVCOLLERSE ANK 4 1740 DKWGKTALHW 1790 EVAQLLLGLG I 1840 ATPGREAGPF 1890 TWSVDLAARG 1940	1700 OTAVDARTE 1750 AAAVNNARAA I 1800 AARELRDO AG 1850 PRARTVSVSV 1900 GGAYSHCRS 1950
1660 RRLLEAGANP GTTPLMLAAR II 1760 RSLLQAGADK 1810 LAPADVAHOR 1860 PPHGGGALPR 1910 SGVGAGGGGPT	1670 NOPDRAGRTP ANK 3 1720 LAVEDLVEEL 1770 DAQDNREQTP I 1820 NHWDLLTLL 1870 CRTLSAGAGP 1920 PRGRRESAGM	1680 LHAAVAADAR I 1730 IAAQADVGAR ANK 5 1780 LFLAAREGAV NLS 1830 GAGPPEARHK I 1880 RGGGACLQAR PEST 1930 RGPRPNPAIM	1690 EVCOLLERSE ANK 4 1740 DKWGKTALHW 1790 EVAQLLLGLG I 1840 ATPGREAGPF 1890 TWSVDLAARG 1940 RGRYGWAAGR	1700 OTAVDARTED 1750 AAAVNNARAA I 1800 AARELRDO AG PRARTVSVSV 1900 GGAYSHCRS 1950 GGRVSTDDWH
1660 RRLLEAGANP 1710 GTTPLMLAAR 1760 RSLLQAGADK 1810 APADVAHOR 1860 PPHGGGALPR 1910 SGVGAGGGPT	1670 NOPDRAGRTP ANK 3 1720 LAVEDLVEEL 1770 DAQDNREQTP I 1820 NHWDLLTLLE 1870 CRTLSAGAGP 1920 PRGRRFSAGM	1680 LHAAVAADAR II 1730 IAAQADVGAR ANK 5 1780 LFLAAREGAV NLS 1830 GAGPPEARHK I 1880 RGGGACLQAR PEST 1930 RGPRPNAAIM	1690 EVCOLLERSE ANK 4 1740 DKWGKTALHW 1790 EVAQLLLGLG I 1840 ATPGREAGPF 1890 TWSVDLAARG 1940 RGRYGVAAGR	1700 OTAVDARTE 1750 AAAVNNARAA 1 1800 AARELRDO AG 1850 PRARTVSVSV 1900 GGAYSHCRS 1950 GGRVSTDDWH 2000
1660 RRLLEAGANP 1710 GTTPLMLAAR II 1760 RSLLQAGADK 1810 APADVAHOR 1860 PPHGGGALPR 1910 SGVGAGGGPT 1960	1670 NOPDRAGRTP ANK 3 1720 LAVEDLVEEL 1770 DAQDNREQTP 1820 NHWDLITLIE 1870 CRTLSAGAGP 1920 PRGRRESAGM 1970 SASNE PIEPEP	1680 LHAAVAADAR II 1730 IAAQADVGAR ANK 5 1780 LFLAAREGAV NLS 1830 GAGPPEARHK I 1880 RGGGACLQAR PEST 1930 RGPRPNPAIM 1980 CLEPSPERS	1690 EVCOLLERSE ANK 4 1740 DKWGKTALHW 1790 EVAQLLLGLG I 1840 ATPGREAGPF 1890 TWSVDLAARG 1940 RGRYGVAAGE 1990 POLDCCOPPA	1700 OTAVDARTE 1750 AAAVNNARAA 1 1800 AARELRDO AG 1850 PRARTVSVSV 1900 GGAYSHCRS 1950 GGRVSTDDWH 2000 OEMPINOOT
1660 RRLLEAGANP 1710 GTTPLMLAAR 1760 RSLLQAGADK 1810 DAPADVAHOR 1860 PPHGGGALPR 1910 SGVGAGGGPT 1960 DDWALGAGG	1670 NOPDRAGRTP ANK 3 1720 LAVEDLVEEL 1770 DAODNREOTP 1 1820 NHWDLITLLE 1870 CRTLSAGAGP 1920 PRGRRESAGM 1970 SASNI PIPPP	1680 LHAAVAADAR I 1730 IAAQADYGAR ANK 5 1780 IELAAREGAV NLS 1830 GAGPPEARHK I 1880 RGGGACLQAR PEST 1930 RGPRPNPAIM 1980 CLTPSPERGS	1690 EVCOLLERSR ANK 4 1740 EKWGKTALHW 1790 EVAOLLEGEG 1840 ATPGREAGPF 1890 TWSVDLAARG 1940 RGRYGVAAGR 1990 EOLDOGEPAL	1700 CTAVDARTE 1750 AAAVNNARAA 1 1800 AARELRDCAG 1850 PRARTVSVSV 1900 GGAYSHCRS 1950 GGRVSTDDHH 2000 OEMPINQGGE
1660 RRLLEAGANP 1710 GTTPLMLAAR 1760 RSLLQAGADK 1810 CAPADVAHOR 1860 PPHGGGALPR 1910 SGVGAGGGPT 1960 CDWVALGACG	1670 NOPDRAGRTP ANK 3 1720 LAVEDLVEEL 1770 DAODNREQTP 1820 NHWDLLTLL 1870 CRTLSAGAGP 1920 PRGRRESAGM 1970 SASNI PIPPP	1680 LHAAVAADAR I 1730 IAAQADYGAR ANK 5 1780 LFLAAREGAV NLS 1830 GAGPPEARHK I 1830 GAGPPEARHK I 1880 RGGGACLQAR PEST 1930 RGBRPNPAIM 1980 CLTPSPERGS	1690 EVCOLLERSR ANK 4 1740 EKWGKTALHW 1790 EVAQLLGGG 1 1840 ATPGREAGPF 1890 TWSVDLAARG 1940 RGRYGVAAGR 1990 POLDCGPPAL	1700 CTAVDARTE 1750 AAAVNNARAA 1 1800 AARELROCAG PRARTVSVSV 1850 PRARTVSVSV 1900 GGAYSHCRS 1950 GGRVSTDDWH 2000 QEMPINQGGH

10	20	EGF-li	ke 1	50
MQPPSLLLLL	LLLLLLCVSV	VRPRGLLCGS	FPEPCANGGT	CLSLSLGQGT
60	70	EGF 80	90 P-like 2	100
CQCAPGFLGE	TCQFPDPCQN	AQLCQNGGSC (QALLPAPLGL	PSSPSPLTPS
110	120	130	140	150
FLCTCLPGFT	GERCQAKLED	EGF-like	GRCHIQASGR 4	PQCSCMPGWT
160	170	180	190	200
EGF-	like 5		EGF-11	ke 6
DPGPCPKGTS	220 CHNTLGSEOC	230	240	250
		EGF-lik	e 7	
260 LMPEKDSTFH	LCLCPPGFIG	280 PDCEVNPDNC	290 VSHQCQNGGT	300 CQDGLDTYTC
31.0	11	230	F-like 8	350
LCPETWTGWD	CSEDVDECET	QGPPHCRNGG	TCQNSAGSFH	CVCVSGWGGT
I I 360	370	EGF-like 9 380	390	400
SCEENLDDCI	AATCAPGSTC	IDRVGSFSCL (CPPGRTGLLC	HLEDMCLSOP
EGF 410	-11ke 10 420	430	EGF-like 440	450
CHGDAQCSTN	PLTGSTLCLC	QPGYSGPTCH	QDLDECLMAQ	QGPSPCEHGG
460	470	480	490	500
SCLNTPGSFN	CLCPPGYTGS	RCEADHNECL : EGF-like 13	SQPCHPGSTC	LDLLATFHCL
510	520	530	540	550
CPPGLEGQLC	EVETNECASA	CLNHADCHD GF-like 14	LINGFQCICL	PGFSGTRCEE
560	570	580	590	600
EGF	-like 15	II	EGF-like	16
GASCLDLPGA	620	630	640	650
1	1	EGF-like	17	1
	4 10 4			
660 CPDGSPGCAP	670 PEDNCTCHHG	680 HCQRSSCVCD	690 VGWTGPECEA	700 ELGGCISAPC
660 CPDGSPGCAP EGF- 710	670 PEDNCTCHHG -like 18	680 HCQRSSCVCD	690 VGWTGPECEA EGF-like 740	700 ELGGCISAPC 19
660 CPDGSPGCAP EGF- 710 AHGGTCYPQP	670 PEDNCTCHHG -like 18 720 SGYNCTCPTG	680 HCQRSSCVCD II 730 YTGPTCSEEM	690 VGWTGPECEA EGF-like 740 TACHSGPCLN	700 ELGGCISAPC 19 750 GGSCNPSPGG
660 CPDGSPGCAP EGF- 710 AHGGTCYPOP 760	670 PEDNCTCHHG -like 18 720 SGYNCTCPTG I I 770	680 HCQRSSCVCD I 730 YTGPTCSEEM EGF-1 780	690 VGWTGPECEA EGF-like 740 TACHSGPCLN ike 20 790	700 ELGGCISAPC 19 750 GGSCNPSPGG 800
660 CPDGSPGCAP EGF- 710 AHGGTCYPQP 760 YYCTCPPSHT	670 PEDNCTCHHG -like 18 720 SGYNCTCPTG I I 770 GPQCQTSTDY	680 HCQRSSCVCD I I 730 YTGPTCSEEM EGF-1 780 CVSAPCENGG	690 VGWTGPECEA EGF-like 740 TACHSGPCLN iike 20 790 TCVNRPGTFS	700 ELGGCISAPC 19 750 GGSCNPSPGG 800 CLCAMGFQGP
660 CPDGSPGCAP EGF- 710 AHGGTCYPQP 760 YYCTCPPSHT I I 810	670 PEDNCTCHHG -like 18 720 SGYNCTCPTG I I 770 GPQCQTSTDY 820	680 HCQRSSCVCD 1 1 1 730 YTGPTCSEEM EGF-1 780 CVSAPCFNGG EGF-1ike 21 830	690 VGWTGPECEA EGF-like 740 TACHSGPCLN Like 20 790 TCVNRPGTFS 840	700 ELGGCISAPC 19 750 GGSCNPSPGG 800 CLCAMGFQGP I I 850
660 CPDGSPGCAP EGF- 710 AHGGTCYPQP 760 YYCTCPPSHT I I 810 RCEGKLRPSC	670 PEDNCTCHHG -like 18 720 SGYNCTCPTG I I 770 GPQCQTSTDY 820 ADSPCRNRAT	680 RCQRSSCVCD 7 1 1 730 YTGPTCS EM EGF-1 780 CVSAPCFNGG 7 EGF-11ke 21 830 CQDSPQGPRC 1	690 VGWTGPECEA EGF-like 740 TACHSGPCLN Like 20 790 TCVNRPGTES 840 LCPTGYTGGS	700 ELGGCISAPC 19 750 GGSCNPSPGG 800 CLCAMGFQGP I I 850 CQTLMDLCAQ
660 CPDGSPGCAP EGF 710 AHGGTCYPQP 760 YYCTCPPSHT I I 810 RCEGKLRSC EGF 860	670 PEDNCTCHHG -like 18 720 SGYNCCPTG I 1 770 GPQCQTSTDY 820 ADSPCRNRAT -like 22 870	680 HCQRSSCVCD 730 YTGPTCSEEM EGF-1 780 CVSAPCFNGG EGF-1ike 21 830 CQDSPQGPRC I I 880	690 VGWTGPECEA EGF-like 740 TACHSGPCLN 1ike 20 790 TCVNRPGTFS 840 LCPTGVTGGS EGF-like 890	700 ELGGCISAPC 19 750 GGSCNPSPGG 800 CLCAMGFQGP I I 850 CQTLMDLCAQ 23 900
660 CPDGSPGCAP EGF- 710 AHGGTCYPQP 760 YYCTCPPSHT I I 810 RCEGKLRPSC EGF- 860 KPCPRNSHCL	670 PEDNCTCHHG 11ke 18 720 SGYNCTCPTG 1 1 770 GPQCQTSTDY 820 ADSPCRNRAT -11ke 22 870 QTGPSFHCLC	680 HCQRSSCVCD Y 1 1 730 YTGPTCS:EM Y EGF-1 800 CVSAPCFNGG S30 CQDSPQGPRC Y 1 1 880 LQGWTGPLCN EGF-11k	690 VGWTGFECEA FGF-like 740 TACHSGPCLN Like 20 790 TCVNRPGTFS 840 LCPTGYTGGS EGF-like 890 LPLSSCQKAA e 24	700 ELGGCISAPC 19 750 GGSCNPSPGG 800 CLCAMGFQGP 1 1 850 CQTLMDLCAQ 23 900 LSQGIDVSSL
660 CPDGSPGCAP EGF- 710 AHGGTCYPQP 760 YYCTCPPSHT I I 810 RCEGKLRPSC EGF- 860 KPCPRNSHCL 910	670 PEDNCTCHHG 11ke 18 720 SGYNCTCPTG I I 770 GPOCQTSTDY 820 ADSPCRNRAT -like 22 870 QTGPSFHCLC I I 920	680 HCQRSSCVCD 7 1 1 730 YTGPTCSEEM 6 EGF-1 830 CVSAPCFNGG EGF-1ike 21 830 CQDSPQGPRC 1 830 LQGWTGPLCN EGF-1ik. 930	690 VGWTGPECEA EGF-like 740 TACHSGPCLN Like 20 790 TCVNRPGTES 840 LCPTGYTGGS EGF-like 890 LPLSSCQKAA e 24 940	700 ELGGCISAPC 19 750 GGSCNPSPGG 800 CLCAMGFQGP I I 850 CQT LMDLCAQ 23 900 LSQGIDVSSL 950
660 CPDGSPGCAP EGF- 710 AHGGTCYPQP 760 YYCTCPPSHT I I 810 RCEGKLRPSC EGF- 860 KPCPRNSHCL 910 CHNGGLCVDS	670 PEDNCTCHHG 11ke 18 720 SGYNCTCPTG 1 1 770 GPQCQTSTDY 820 ADSPCRNRAT 11ke 22 870 QTGPSFHCLC 1 1 920 GPSYFCHCPP 1 1	680 RCQRSSCVCD 7 1 1 730 YTGPTCS EM EGF-1 780 CVSAPCFNGG 7 EGF-1ike 21 830 CQDSPQGPRC 1 830 CQDSPQGPRC 1 880 LQGWTGPLCN EGF-1ike 930 GPQGSLCQDH 7 EGF-1i	690 VGWTGFECEA EGF-like 740 TACHSGPCLN 1ike 20 790 TCVNRPGTFS 840 LCPTGVTGGS EGF-like 890 LPLSSCQKAA e 24 940 VNPCESRPCQ ke 25	700 ELGGCISAPC 19 750 GGSCNPSPGG 800 CLCAMGFQGP I I 850 CQTLMDLCAQ 23 900 LSQGIDVSSL 950 NGATCMAQPS
660 CPDGSPGCAP EGF- 710 AHGGTCYPQP 760 YYCTCPPSHT I I 810 RCEGKLRPSC EGF- 860 KPCPRNSHCL 910 CHNGGLCVDS 960	670 PEDNCTCHHG -1ike 18 720 SGYNCTCPTG I 1 770 GPQCQTSTDY 820 ADSPCRNRAT -1ike 22 870 QTGPSFHCLC I 1 920 GPSYFCHCPP I 1 970	680 HCQRSSCVCD Y 730 YTGPTCSEEM EGF-1 780 CVSAPCFNGG CVSAPCFNGG 1 1 830 CQDSPQGPRC 1 1 880 CQDSPQGPRC 1 1 880 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 CQDSPQGPRC 1 80 C CQDSPQGPRC 1 80 C C C C C C C C C C C C C C C C C C	690 VGWTGFECEA EGF-like 740 TACHSCPCLN Like 20 790 TCVNRPGTFS 840 LCPTGYTGGS EGF-like 890 LPLSSCQKAA e 24 940 VNPCESRPCQ .ke 25 990	700 ELGGCISAPC 19 750 GGSCNPSPGG 800 CLCAMGFQGP I I 850 CQTLMDLCAQ 23 900 LSQGIDVSSL 950 NGATCMAQPS 1000
660 CPDGSPGCAP EGF- 710 AHGGTCYPQP 760 YYCTCPPSHT I I 810 RCEGKLRPSC EGF- 860 KPCPRNSHCL 910 CHNGGLCVDS 960 GYLCQCAPGY I I	670 PEDNCTCHHG 720 SGYNCTCPTG 1 1 770 GPQCQTSTDY 820 ADSPCRNRAT -1ike 22 870 QTGPSFHCLC 1 2 920 GPSYFCHCPP 1 1 970 DGQNCS: ELD E	680 HCQRSSCVCD 7 730 YTGPTCSEEM 6 EGF-1 780 CVSAPCFNGG 7 EGF-1ike 21 830 CQDSPQGPRC 1 830 CQDSPQGPRC 1 840 CQDSPQGPRC 1 840 CQDSPQ 1 840 CQD 1 8	690 VGWTGPECEA EGF-like 740 730 TACHSGPCLN Like 20 790 TCVNRPGTFS 840 LCPTGYTGGS EGF-like 890 LPLSSCCKAA 940 VNPCESRPCQ .ke 25 990 GTCTPKPGGF	700 ELGGCISAPC 19 750 GGSCNPSPGG 800 CLCAMGFQGP 1 1 850 CQTLMDLCAQ 23 900 LSQGIDVSSL 950 NGATCMAQPS 1000 HCACPPGFVG 1 1
660 CPDGSPGCAP EGF- 710 AHGGTCYPQP 760 YYCCPPSHT I I 810 RCEGKLRPSC EGF- 860 KPCPRNSHCL 910 CHNGGLCVDS 960 GYLCQCAPGY I I 1010	670 PEDNCTCHHG 11ke 18 720 SGYNCTCPTG 1 1 770 GPQCQTSTDY 820 ADSPCRNRAT 11ke 22 870 QTGPSFHCLC 1 1 920 GPSYFCHCPP 1 1 970 DGQNCSKELD E 1020 LDOPCHPTGT	680 HCQRSSCVCD 7 730 YTGPTCSEM 2 EGF-11ke 21 830 CQDSPQGPRC 1 830 CQDSPQGPRC 1 830 CQDSPQGPRC 2 1 1 880 LQGWTGPLCN 2 EGF-11k 930 GFQGSLCQDH 2 EGF-11k 930 ACQSQPCHNH 3 GF-11ke 26 1030 AACHSLANAF 2 AACHSLANAF 2 AACHSLA	690 VGWTGPECEA EGF-like 740 TACHSGPCLN Like 20 790 TCVNRPGTES 840 LCPTGYTGGS EGF-like 890 LPLSSCQKAA e 24 940 VNPCESRPCQ .ke 25 990 GTCTPKPGGF 1040	700 ELGGCISAPC 19 750 GGSCNPSPGG 6GSCNPSPGG 800 CLCAMGFQGP 1 1 850 CQTLMDLCAQ 23 900 LSQGIDVSSL 950 NGATCMAQPS 1000 HCACPPGFVG 1 1 1050 000000000000000000000000000000
660 CPDGSPGCAP EGF- 710 AHGGTCYPQP 760 YYCTCPPSHT I I 810 RCEGKLRPSC EGF- 860 KPCPRNSHCL 910 CHNGGLCVDS 960 GYLCQCAPGY I I 1010 LRCEGDVDEC	670 PEDNCTCHHG 11ke 18 720 SGYNCTCPTG 1 1 770 GPQCQTSTDY 820 ADSPCRNRAT 11ke 22 870 QTGPSFHCLC 1 1 920 GPSYFCHCPP 1 1 970 DGQNCSKELD E 102 102 102 102 102 102 102 102	680 RCQRSSCVCD 7 730 YTGPTCS EM EGF-1 780 CVSAPCFNGG EGF-1ike 21 830 CQDSPQGPRC 1 830 CQDSPQGPRC 1 830 LQGWTGPLCN EGF-1ike 930 GFQGSLCQDH EGF-1i 980 ACQSQPCHNH GF-1ike 26 1030 AACHSLANAF 27	690 VGWTGFECEA EGF-like 740 TACHSGPCLN like 20 790 TCVNRPGTFS 840 LCPTGYTGGS EGF-like 890 LPLSSCQKAA e 24 940 VNPCESRPCQ ke 25 990 GTCTPRPGGF 1040 YCQCLPGHTG I I	700 ELGGCISAPC 19 750 GGSCNPSPGG 800 CLCAMGFQGP I I 850 CQTLMDLCAQ • 23 900 LSQGIDVSSL 950 NGATCMAQPS 1000 HCACPPGFVG I I 1050 QWCEVEIDPC
660 CPDGSPGCAP EGF- 710 AHGGTCYPQP 760 YYCTCPPSHT I I 810 RCEGKLRPSC EGF- 860 KPCPRNSHCL 910 CHNGGLCVDS 910 CHNGGLCVDS 910 GYLCQCAPGY I I 1010 LRCEGDVDEC 1060 HSOPCFHGGT	670 PEDNCTCHHG 720 SGYNCTCPTG 1 1 770 GPQCQTSTDY 820 ADSPCRNRAT -1ike 22 870 QTGPSFHCLC 1 920 GPSYFCHCPP 1 920 GPSYFCHCPP 1 020 LDQPCHPTGT EGF-Like 1070 CEATAGSPLG	680 HCQRSSCVCD 7 730 YTGPTCS EM EGF-1 780 CVSAPCFNGG 7 EGF-11ke 21 830 CQDSPQGPRC 1 830 CQDSPQGPRC 1 880 LQGWTGPLCN EGF-11k 930 GFQGSLCQDH 7 EGF-11 980 ACQSQPCHNH GF-11ke 26 1030 AACHSLANAF 27 1080 FICHCPRGFE	690 VGWTGFECEA EGF-like 740 TACHSGPCLM 1ike 20 790 TCVNRPGTFS 840 LCPTGYTGGS EGF-like 890 LPLSSCQKAA e 24 940 VNPCESPCQ 1040 YCQCLPGHTG I 1 1090 GFTCSHRAPS	700 ELGGCISAPC 19 750 GGSCNPSPGG 800 CLCAMGFQGP I I 850 CQTLMDLCAQ 23 900 LSQGIDVSSL 950 NGATCMAQPS 1000 HCACPPGFVG I 1050 QWCEVEIDPC 1100 CGFHHCHHGG
660 CPDGSPGCAP EGF- 710 AHGGTCYPQP 760 YYCTCPPSHT I I 810 RCEGKLRPSC EGF- 860 KPCPRNSHCL 910 CHNGGLCVDS 960 GYLCQCAPGY I I 1010 LRCEGDVDEC 1060 HSQPCFHGGT EGF	670 PEDNCTCHHG 720 SGYNCTCPTG 1 1 770 GPOCQTSTDY 820 ADSPCRNRAT -like 22 870 QTGPSFHCLC 1 1 920 GPSYFCHCPP 1 1 970 DGQNCSKELD 2000 LD0PCHPTGT EGF-like 1070 CEATAGSPLG -like 28 1020	680 HCQRSSCVCD 7 730 YTGPTCSEEM 6 EGF-1 780 CVSAPCFNGG 7 EGF-1ike 21 830 CQDSPQGPRC 1 830 CQDSPQGPRC 1 880 LQGWTGPLCN 6 EGF-1ik 930 GFQGSLCQDH 1 980 ACQSQPCHNH 6 GF-1ike 26 1030 AACHSLANAF 27 1080 FICHCPKGFE 1 110 110 110 110 110 110 110 1	690 VGWTGPECEA EGF-like 740 TACHSGPCLN Like 20 790 TCVNRPGTES 840 LCPTGYTGGS EGF-like 890 LPLSSCQKAA e 24 940 VNPCESRPCQ Lke 25 990 GTCTPKPGGF 1040 YCQCLPGHTG 1090 GPTCSHRAPS EGF-like	700 ELGGCISAPC 950 GGSCNPSPGG CLCAMGFQGP I I 850 CQTLMDLCAQ 23 900 LSQGIDVSSL 950 NGATCMAQPS 1000 HCACPPGFVG I I 1050 QWCEVEIDPC 1100 CGFHHCHHGG 29
660 CPDGSPGCAP EGF- 710 AHGGTCYPQP I I 810 RCEGK LRPSC EGF- 860 KPCPRNSHCL 910 CHNGGLCVDS 960 GYLCQCAPGY I I 1010 LRCEG DVDEC 1060 HSQPCFHGGT 1110 LCLPSPKPGF	670 PEDNCTCHHG 11ke 18 720 SGYNCTCPTG 1 1 770 GPQCQTSTDY 820 ADSPCRNAT -like 22 870 QTGPSFHCLC 1 1 920 GPSYFCHCPP 1 1 970 DGQNCSKELD E 1020 LDQPCHTGT EGF-like 1070 CEATAGSPLG -like 28 1120	680 HCQRSSCVCD 7 730 YTGPTCS EM 6 EGF-1 780 CVSAPCFNGG 6 EGF-1ike 21 830 CQDSPQGPRC 7 I 1 880 LQGWTGPLCN 6 EGF-1ike 930 GFQGSLCQDH 6 EGF-1ike 930 ACQSQPCHNH 6 GF-1ike 26 1030 ACQSQPCHNH 6 GF-1ike 26 1030 ACCSQPCHNH 6 1030 ACCSQPCHNH 7 1030 ACCSQPCHNH 7 1030 ACCSQPCH	690 VGWTGPECEA EGF-like 740 TACHSGPCLN Like 20 790 TCVNRPGTFS 840 LCPTGYTGGS EGF-like 890 LPLSSCQKAA e 24 940 VNPCESRPCQ 1040 YCQCLPGHTG 1090 GTCTPKPGGF 1040 PCGCPPSPC	700 ELGGCISAPC 19 750 GGSCNPSPGG 800 CLCAMGFQGP I I 850 CQTLMDLCAQ 900 LSQGIDVSSL 950 NGATCMAQPS 1000 HCACPPGFVG I I 1050 QWCEVEIDPC 1100 CGFHHCHHGG 29 1150
660 CPDGSPGCAP EGF- 710 AHGGTCYPQP 760 YYCTCPPSHT I I 810 RCEGKLRPSC EGF- 860 KPCPRNSHCL 910 CHNGGLCVDS 960 GYLCQCAPGY I I 1010 LRCEGDVDEC 1060 HSQPCFHGGT EGF 1110 LCLPSPKPGF	670 PEDNCTCHHG 120 SGYNCTCPTG 1 1 770 GPQCQTSTDY 820 ADSPCRNRAT 11ke 22 870 QTGPSFHCLC 1 1 920 GPSYFCHCPP 1 1 970 DGQNCS: ELD 020 LDQPCHPTGT EGF-1ike 1070 CEATAGSPLG -1ike 28 1120 PPRCACLSGY 1170	680 RCQRSSCVCD 7 730 YTGPTCS EM EGF-1 780 CVSAPCFNGG 7 EGF-1ike 21 830 CQDSPQGPRC 7 I 1 880 LQGWTGPLCN EGF-1ike 930 GFQGSLCQDH 7 GF-1ike 26 1030 AACHSLANAF 27 1080 FICHCPKGFE 1 1130 GGPDCL TPPA	690 VGWTGPECEA EGF-like 740 TACHSGPCLN like 20 790 TCVNRPGTFS 840 LCPTGYTGGS 840 LPLSSCQKAA e 24 940 VNPCESRPCQ L040 YCQCLPGHTG 1040 YCQCLPGHTG 1090 GTCTRRPGGF 1040 YCQCLPGHTG 110 1090 GPTCSHRAPS EGF-like 1140	700 ELGGCISAPC 19 750 GGSCNPSPGG 800 CLCAMGFQGP I I 850 CQTLMDLCAQ 950 NGATCMAQPS 1000 HCACPPGFVG 1050 QWCEVEIDPC 29 1150 LYNGSCSETT 1200
660 CPDGSPGCAP EGF- 710 AHGGTCYPQP 760 YYCTCPPSHT I I 810 RCEGKLRPSC EGF 860 KPCPRNSHCL 910 CHNGGLCVDS 960 GYLCQCAPGY I I 1010 LRCEGDVDEC 1060 HSQPCFHGGT EGF 1110 LCLPSPKPGF	670 PEDNCTCHHG 720 SGYNCTCPTG 1 1 770 GPQCQTSTDY 820 ADSPCRNRAT -1ike 22 870 QTGPSFHCLC 1 920 GPSYFCHCPP 1 920 GPSYFCHCPP 1 020 LDQPCHPTGT EGF-1ike 1020 CEATAGSPLG -1ike 28 1120 PPRCACLSGY 1170 CPHSSPGPRC	680 HCQRSSCVCD 7 730 YTGPTCS EM EGF-1 780 CVSAPCFNGG 7 EGF-1ike 21 830 CQDSPQGPRC 1 830 CQDSPQGPRC 1 880 LQGWTGPLCN EGF-1ike 26 1030 ACQSQPCHNH 9 80 ACQSQPCHNH 6 1030 AACHSLANAF 27 1080 FICHCPKGFE 6 1130 GGPDCL TPPA 1180 QKPGAKGCEG	690 VGWTGFPECEA EGF-like 740 TACHSGPCLN 1ike 20 790 TCVNRPGTFS 840 LCPTGYTGGS EGF-like 890 LPLSSCQKAA e 24 940 VNPCESPCQ 1040 YCQCLPGHTG I 1 1090 GFTCSHRAPS EGF-like 1140 PKGCGPPSPC LNR 1 1190	700 ELGGCISAPC 19 750 GGSCNPSPGG 800 CLCAMGFQGP I I 850 CQTLMDLCAQ 23 950 NGATCMAQPS 1000 HCACPPGFVG I 1050 QWCEVEIDPC 1100 CGFHHCHHGG 29 1150 LYNGSCSETT 1200 CSGFCGNWDG
660 CPDGSPGCAP EGF- 710 AHGGTCYPQP 760 YYCTCPPSHT I I 810 RCEGKLRPSC EGF- 860 KPCPRNSHCL 910 CHNGGLCVDS 960 GYLCQCAPGY I I 1010 LRCEGDVDEC 1060 HSQPCFHGGT EGF 1110 LCLPSPKPGF 11210	670 PEDNCTCHHG 11ke 18 720 SGYNCTCPTG 1 1 770 GPOCQTSTDY 820 ADSPCRNAT -like 22 870 QTGPSFHCLC 1 1 920 GPSYFCHCPP 1 3 970 DGQNCSKELD EGF-1ike 1020 LDQPCHPTGT EGF-1ike 28 1120 PPRCACLSGY 11 1170 CHSSPGPRC	680 HCQRSSCVCD 7 730 YTGPTCSEEM 6 EGF-1 780 CVSAPCFNGG EGF-1ike 21 830 CQDSPQGPRC 7 830 CQDSPQGPRC 7 830 LQGWTGPLCN 8 EGF-1ike 930 GFQGSLCQDH 7 930 GFQGSLCQDH 7 930 GFQGSLCQDH 7 880 LQGWTGPLCN 8 EGF-1ike 26 1030 AACHSLANAF 2 1130 GGPDCL TPPA 7 1180 CKPGAKGCEG 1 LNR 2 1230	690 VGWTGPECEA EGF-like 740 TACHSGPCLN Like 20 790 TCVNRPGTES 840 LCPTGYTGGS EGF-like 890 LPISSCQKAA e 24 940 VNPCESRPCQ Lke 25 990 GTCTPKPGGF 1040 YCQCLPGHTG 1090 GPTCSHRAPS EGF-like 1140 PKGCGPPSPC LNR 1 1190 RSGDGACDAG	700 ELGGCISAPC 19 750 GGSCNPSPGG 800 CLCAMGFQGP I I 850 CQT LMDLCAQ 23 900 LSQGIDVSSL 950 NGATCMAQPS 1000 HCACPPGFVG I 1 1050 QWCEVEIDPC 1100 CGFHHCHHGG 29 1150 LYNGSCSETT 1200 CSGPGGNWDG I 1250
660 CPDGSPGCAP EGF 710 AHGGTCYPQP I I 810 RCEGK LRPSC EGF 860 KPCPRNSHCL 910 CHNGGLCVDS 960 GYLCQCAPGY I I 1010 LRCEG DVDEC 1060 HSQPCFHGGT EGF 1110 LCLPSPKPGF 1150 GLGGPGFRCS 1210 GDCSLGVPDP	670 PEDNCTCHHG 11ke 18 720 SGYNCTCPTG 1 1 770 GPQCQTSTDY 820 ADSPCRNAT 11ke 22 870 QTGPSFHCLC 1 1 920 GPSYFCHCPP 1 1 970 DGQNCSKELD E 1020 LDQPCHTGT EGF-11ke 28 1120 PPRCACLSGY 11 1170 CPHSSPGPRC	680 HCQRSSCVCD 7 730 YTGPTCS EM 6 EGF-1 780 CVSAPCFNGG 21 830 CQDSPQGPRC 1 830 CQDSPQGPRC 1 880 LQGWTGPLCN 6 EGF-1ike 26 1030 ACQSQPCHNH 6 GF-1ike 26 1030 ACQSQPCHNH 6 CVSAPCC 7 1030 ACQSQPCHNH 6 CVSAPCC 7 1030 ACQSQPCHNH 6 CVSAPCC 7 1030 ACQSQPCHNH 7 1030 ACQSQPCHNH 7 1030 FICHCPKGFE 1 1130 GQPDCL 7 1180 CVSAPCC 7 1230 CVSAPCC 7 1230 CVSAPC 7 1330 CVSAPC 7 1330 CVSAPC 7 1330 CVSAPC	690 VGWTGPECEA EGF-like 740 TACHSGPCLN Like 20 790 TCVNRPGTFS 840 LCPTGYTGGS EGF-like 890 LPLSSCQKAA e 24 940 VNPCESRPCQ L040 YCQCLPGHTG 1 1 1090 GTCTPKPGGF 1040 YCQCLPGHTG 1 1 1090 GFTCSHRAPS EGF-like 1140 PKGCGPPSPC LNR 1 1190 RSGDGACDAG	700 ELGGCISAPC 19 750 GGSCNPSPGG 800 CLCAMGFQGP I I 850 CQTLMDLCAQ 950 NGATCMAQPS 1000 HCACPPGFVG 1000 HCACPPGFVG 1050 QWCEVEIDPC 1100 CGFHHCHHGG 29 1150 LYNGSCSETT 1200 CSGPGGNWDG I 1250 DGYDCETPPA
660 CPDGSPGCAP EGF- 710 AHGGTCYPQP I I I I RCEGKLRPSC EGF- 810 KPCPRNSHCL 910 CHNGGLCVDS 960 GYLCQCAPGY I I 010 CHNGGLCVDS 960 GYLCQCAPGY I I 0100 LRCEGDVDEC EGF 1110 LCLPSPKPGF 1110 GLGGPGFRCS I 1210 GDCSLGVPDP 1 1260	670 PEDNCTCHHG 120 SGYNCTCPTG 1 1 770 GPQCQTSTDY 820 ADSPCRNRAT 11ke 22 870 QTGPSFHCLC 1 1 920 QTGPSFHCLC 1 1 970 DGQNCSKELD E 1020 LDQPCHFGT EGF-1ike 1070 CEATAGSPLG 1120 PPRCACLSGY 11 1170 CPHSSPGPRC 1220 WKGCPSHSRC	680 HCQRSSCVCD 7 730 YTGPTCS EM EGF-1 780 CVSAPCFNGG 7 EGF-1ike 21 830 CQDSPQGPRC 1 830 CQDSPQGPRC 1 830 LQGWTGPLCN EGF-1ike 26 1030 ACCSQPCHNH 6 GF05LCQDH 7 EGF-1ike 26 1030 AACHSLANAF 7 1080 FICHCPKGFE 1 1130 GGPDCL TPPA 7 1180 QKPGAKGCEG 1 1180 QKPGAKGCEG 1 1180 QKPGAKGCEG 1 1230	690 VGWTGPECEA EGF-like 740 TACHSGPCLN like 20 790 TCVNRPGTFS 840 LCPTGYTGGS EGF-like 890 LPLSSCQKAA e 24 940 VNPCESRPCQ Lke 25 990 GTCTPKPGGF 1040 YCQCLPGHTG 1 1 1090 GFTCSHRAPS EGF-like 1140 PKCGGPSPC LNR 1 1190 RSGDGACDAG 1240 PQCDSEECLF 1290	700 ELGGCISAPC 19 750 GGSCNPSPGG 800 CLCAMGFQGP I I 850 CQTLMDLCAQ 950 NGATCMAQPS 1000 HCACPPGEVG 1000 HCACPPGEVG 1050 QWCEVEIDPC 1100 CGFHHCHHGG 29 1150 LYNGSCSETT 1200 CSGPGGNWDG I 1250 DGYDCETPPA I 1300

Figure 5. Notch4 protein domains. Colors represent protein domains.

	Notch1							
A/A	SNP ID	Nt Change NG_007458.1	Mutation	Domain	Phenotype	Accession Number		
1	_	_	Y2116>TERM	TAD	Mental retardation, autosomal dominant	CM171556		
2	rs79782048	36150G>A	E694>K	EGF-like 18	Alzheimer's disease			

 Table 4. SNPs and mutations in Notch1 associated with neurodegenerative diseases.

 Table 5. SNPs and mutations in Notch2 associated with neurodegenerative diseases.

	Notch2							
A/A	SNP ID	Nt Change NG_008163.1	Mutation	Domain	Phenotype	Accession Number		
1	rs61752484	148130A>G	D1327G	EGF-like 34	autism multiplex			

 Table 6. SNPs and mutations in Notch3 associated with neurodegenerative diseases.

	Notch3											
A/A	SNP ID	nt Change NG_009819.1	Codon Change	Mutation	Domain	Phenotype	Accession Number					
1	N/A		TGC-GGC	C43>G			CM052264					
2	N/A		TGC-TCC	C43>S	-		CM1712096					
3	N/A		TGC-TTC	C43>F	-		CM035648					
4	N/A		TGT-CGT	C49>R	-		CM106868					
5	N/A		TGT-GGT	C49>G	-		CM073243					
6	rs193921045	8431G>A	TGT-TAT	C49>Y	-		CM971054					
7	rs193921045	8431G>T	TGT-TTT	C49>F	-		CM052265					
8	N/A		GGT-TGT	G53>C	-		CM106869					
9	N/A		CGT-TGT	R54>C	-		CM003012					
10	N/A		TGC-GGC	C55>G	-	CADACI	CM1714276					
11	N/A		TCC-TGC	S60>C	ECE like 1		CM052266					
12	N/A		CGG-TGG	R61>W	- EGF-like I	CADASIL	CM0910982					
13	N/A		TGC-GGC	C65>G	-		CM108953					
14	N/A		TGC-TAC	C65>Y	-		HM070086					
15	N/A		TGC-TCC	C65>S	-		CM052267					
16	N/A		TGC-AGC	C67>S	-		CM092086					
17	N/A		TGC-TAC	C67>Y	-		CM034666					
18	rs28937321	13478G>T	TGG-TGT	W71>C	-							
19	N/A		CGG-CCG	R75>Q	-		CM156963					
20	N/A		CGG-CCG	R75>P	-		CM061880					
21	N/A		TGT-CGT	C76>R	-		CM023649					
22	N/A		TGT-TGG	C76>W	-		CM052268					
23	N/A		TGT-TAT	C76>Y	-		CM1615079					

A/A

24

25

SNP ID

N/AN/A

	Not	ch3			
nt Change NG_009819.1	Codon Change	Mutation	Domain	Phenotype	Accession Number
	GAC-GGC	D80>G			CM150655
	TGT-CGT	C82>R			CM148255
	TGT-CGT	C87>R			CM052269
	TGT-TAT	C87>Y			CM044913
	TGT-TTT	C87>F			CM1714277
	GGC-TGC	G89>C			CM128781
	CGT-TGT	R90>C			CM971055
	TGC-GGC	C93>G			CM1714278
	TGC-TAC	C93>Y			CM023650

26	N/A	TGT-CGT	C87>R	-		CM052269
27	N/A	TGT-TAT	C87>Y	-		CM044913
28	N/A	TGT-TTT	C87>F	-		CM1714277
29	N/A	GGC-TGC	G89>C	-		CM128781
30	N/A	CGT-TGT	R90>C	-		CM971055
31	N/A	TGC-GGC	C93>G			CM1714278
32	N/A	TGC-TAC	C93>Y			CM023650
33	N/A	TGC-TTC	C93>F			CM001263
34	N/A	CGA-TGA	R103>Term			CM1313313
35	N/A	TGC-CGC	C106>R			CM138997
36	N/A	TGC-GGC	C106>G	EGF-like 2	CADASIL	CM156032
37	N/A	TGC-TGG	C106>W	-		CM044912
38	N/A	CGG-TGG	R107>W	-		CM159172
39	N/A	TGC-AGC	C108>S			CM125146
40	N/A	TGC-CGC	C108>R			CM045072
41	N/A	TGC-TAC	C108>Y			CM052270
42	N/A	TGC-TGG	C108>W			HM040053
43	N/A	TGC-TTC	C108>F			CM1714279
44	N/A	CGT-TGT	R110>C			CM971056
45	N/A	TGC-CGC	C117>R			CM078103
46	N/A	TGC-TAC	C117>Y			CM0910785
47	N/A	TGC-TCC	C117>S			CM132414
48	N/A	TGC-TTC	C117>F			CM001264
49	N/A	TCC-TGC	S118>C			HM040099

Notch3									
	A/A	SNP ID	nt Change NG_009819.1	Codon Change	Mutation	Domain	Phenotype	Accession Number	
	50	N/A		TGC-TAC	C123>Y			CM003013	
	51	N/A		TGC-TTC	C123>F	_		CM001265	
	52	N/A		TGT-GGT	C128>G	_		CM064156	
	53	N/A		TGT-TAT	C128>Y	_		CM023651	
	54	N/A		TGT-TTT	C128>F	_		CM125156	
	55	N/A		GGT-TGT	G131>C	_		HM070153	
	56	rs137852642	13740C>T	CGC-TGC	R133>C	_		CM971057	
	57	rs137852642	13740C>A	CGC-AGC	R133>S	_			
	58	N/A		TGC-GGC	C134>G	_		CM156958	
	59	N/A		TGC-TAC	C134>Y	_		CM125159	
	60	N/A		TGC-TGG	C134>W	_		CM014589	
	61	N/A		GAT-GGT	D139>V	_		CM1615017	
	62	N/A		CGC-TGC	R141>C	_		CM971058	
	63	N/A		TTC-TGC	F142>C	_		CM023652	
	64	N/A		TGC-TCC	C144>S	_		CM159357	
	65	N/A		TGC-TAC	C144>Y	EGF-like 3	CADASIL	CM001267	
	66	N/A		TGC-TTC	C144>F	_		CM003947	
	67	N/A		TCC-TGC	S145>C	_		CM044908	
	68	N/A		TGC-CGC	C146>R	_		CM971059	
	69	N/A		TGC-TAC	C146>Y	-		CM045307	
	70	N/A		TGC-TGG	C146>W	-		CM1213543	
	71	N/A		GGC-GTC	G149>V	_		CM147538	
	72	N/A		GGC-TGC	G149>C	_		CM052271	
	73	N/A		TAC-TGC	Y150>C	_		CM001268	
	74	rs371491165	13794C>G	CAG-GAG	Q151>E	_		CM0910786	
	75	rs797045014	13800C>T	CGC-TGC	R153>C	_		CM971060	
	76	rs797045014	13800C>A	CGC-AGC	R153>S	_			
	77	N/A		TGC-AGC	C155>S	_		CM0910788	
	78	N/A		TGC-TAC	C155>Y	-		CM125157	
	79	N/A		TGC-TCC	C155>S	-		CM044910	
	80	N/A		TGC-TGG	C155>W	-		CM159358	

Table 6. Cont.

Tahl	e 6	Cont	

Notch3										
A/A	SNP ID	nt Change NG_009819.1	Codon Change	Mutation	Domain	Phenotype	Accession Number			
81	N/A		TGC-AGC	C162>S			CM003014			
82	N/A		TGC-CGC	C162>R	_		CM086704			
83	N/A		TGC-TAC	C162>Y	_		CM170225			
84	N/A		TGC-TGG	C162>W	_	CADASIL	CM035650			
85	N/A		GGT-TGT	G165>C	_		CM0910787			
86	rs28933696	13848C>T	CGC-TGC	R169>C	_		CM961043			
87	rs147373451	13852A>G	CAT-CGT	H170>R	_	CADASIL, Alzheimer's disease	CM107598			
88	rs147373451	13852A>T	CAT-CTT	H170>L	_	CADASIL, Alzheimer's disease				
89	N/A		GGT-TGT	G171>C	_		CM971061			
90	N/A		TGC-AGC	C174>S	_		CM125160			
91	N/A		TGC-CGC	C174>R	_		CM033795			
92	N/A		TGC-TAC	C174>Y	_		CM001269			
93	N/A		TGC-TTC	C174>F	_		CM014211			
94	N/A		TCC-TGC	S180>C	-	ECE-like 1	CM003015			
95	N/A		TTC-TGC	F181>C	- FGF-like 4		CM095734			
96	rs28933697	13887C>T	CGC-TGC	R182>C		CADASIL, Alzheimer's disease	CM961044			
97	N/A		TGC-AGC	C183>S	-		CM001271			
98	N/A		TGC-CGC	C183>R	-		CM001270			
99	N/A		TGC-TAC	C183>Y	_		CM1615080			
100	N/A		TGC-TTC	C183>F	-		CM052272			
101	N/A		TGT-AGT	C185>S	-		CM1010137			
102	N/A		TGT-CGT	C185>R	_		CM971062			
103	N/A		TGT-GGT	C185>G	_		CM014590			
104	N/A		TGT-TAT	C185>Y	_		CM147761			
105	N/A		TAC-TGC	Y189>C	_		CM042442			
106	N/A		TGT-AGT	C194>S	_		CM042443			
107	N/A		TGT-CGT	C194>R	_		CM023653			
108	N/A		TGT-GGT	C194>G	_		CM150660			
109	N/A		TGT-TCT	C194>S	_		CM1010138			
110	N/A		TGT-TTT	C194>F	-		CM001272			
111	N/A		TGT-TAT	C194>Y	-		CM003016			

			Not	ch3			
A/A	SNP ID	nt Change NG_009819.1	Codon Change	Mutation	Domain	Phenotype	Accession Number
112	N/A		GCG-ACG	A198>T		White matter lesions?	CM0911492
113	N/A		TGT-CGT	C201>R	_		CM065340
114	N/A		TGT-TAT	C201>Y	_	·	CM044914
115	N/A		GCG-GTG	A201>V	_	CADASIL	CM121679
116	N/A		TGC-CGC	C206>R			CM055455
117	N/A		TGC-TAC	C206>Y	_	CADASIL	CM003017
118	N/A		CGT-CAT	R207>H	White matter lesions?	White matter lesions?	CM0911493
119	N/A		CGT-TGT	R207>C	_	-	CM003018
120	N/A		TGC-AGC	C212>S	_		CM971063
121	N/A		TGC-CGC	C212>R	_		CM1714280
122	N/A		TGC-TAC	C212>Y	_		CM110280
123	N/A		TGC-TGG	C212>W	_		CM132412
124	N/A		TGC-TTC	C212>F	_		CM1515586
125	N/A		AGG-AAG	R213>K	_		CM033796
126	N/A		TAC-TGC	Y220>C	_		HM0657
127	N/A		TGT-AGT	C222>S	- EGF-like 5		BM1486714
128	N/A		TGT-CGT	C222>R	_		BM1496778
129	N/A		TGT-GGT	C222>G	_	CADASIL	CM971064
130	N/A		TGT-TAT	C222>Y	_		CM023654
131	N/A		TGT-TCT	C222>S	_		CM106870
132	N/A		TGT-CGT	C224>R	_		CM159178
133	N/A		TGT-TAT	C224>Y	_		CM971065
134	N/A		TTT-TGT	F228>C			CM1414770
135	N/A		TGT-AGT	C233>S			CM023655
136	N/A		TGT-CGT	C233>R			CM1010139
137	N/A		TGT-TAT	C233>Y		CM052273	
138	N/A		TGT-TGG	C233>W	_		CM035651

Table 6. Cont.

			Not	ch3			
A/A	SNP ID	nt Change NG_009819.1	Codon Change	Mutation	Domain	Phenotype	Accession Number
139	N/A		GTG-ATG	V237>M			CM025913
140	N/A		TGT-TCT	C240>S	_		CM052274
141	N/A		TGT-AGT	C245>S	-		CM056015
142	N/A		TGT-CGT	C245>R	_		CM052275
143	N/A		TGT-TAT	C245>Y	_		CM1213544
144	N/A		TGC-AGC	C251>S	- FGE-like 6	CADASIL	CM035643
145	N/A		TGC-CGC	C251>R			CM023656
146	N/A		TGC-GGC	C251>G	-		CM077212
147	N/A		TGC-TAC	C251>Y	-		CM092087
148	N/A		GTG-ATG	V252>M	-		CM150661
149	N/A		TAT-TGT	Y258>C	-		CM971066
150	N/A		TGC-CGC	C260>R	-		CM1414771
151	N/A		TGC-GGC	C260>G			CM095351
152	N/A		TGC-TAC	C260>Y	-		CM052276
153	N/A		TGC-TTC	C260>F	-	CADASIL	CM1213545
154	N/A		TGC-TTC	C271>F	-		HM060011
155	N/A		TGT-CGT	C278>R	-	CADASIL	CM155988
156	N/A		TGC-TAC	C291>Y	- EGF-like 7	CADASIL	CM156956
157	N/A		GGT-TGT	G296>C	-	CADASIL	CM108668
158	N/A		AGC-TGC	S299>C	_	CADASIL	CM046101
159	N/A		AGA-AAG	G309>K	-	White matter lesions?	CM0911494
160	N/A		TGC-AGC	C311>S	_	CADASIL	CM1414772
161	N/A		TGT-TTT	C318>F			CM1615081
162	N/A		GTG-ATG	V322>M	-		CM150662
163	N/A		TGC-AGC	C323>S	-		CM159171
164	N/A		TGC-TGG	C323>W	-		CM167812
165	N/A		TGC-TAC	C329>Y	_		CM1213546
166	rs137852641	14516C>T	CGC-TGC	R332>C	-		CM014070
167	N/A		TCT-TGT	S335>C	EGF-like 8	CADASIL	CM052277
168	N/A		TAC-TGC	Y337>C	-		CM052278
169	N/A		TGT-CGT	C338>R			CM056018
170	N/A		TGT-TCT	C338>S			CM1718752
171	N/A		TGT-TTT	C338>F			CM150663
172	N/A		TGC-TAC	C340>Y			CM1413344
173	N/A		TGC-TTC	C340>F	-		CM1615082

			Not	ch3			
A/A	SNP ID	nt Change NG_009819.1	Codon Change	Mutation	Domain	Phenotype	Accession Number
174	N/A		TGC-CGC	C360>R			CM140149
175	N/A		TGT-CGT	C366>R	_		CM139199
176	N/A		TGT-TGG	C366>W	_		CM082988
177	N/A		TGT-CGT	C379>R	EGF-like 9	CADASIL	HM090054
178	N/A		TGT-TCT	C379>S	_		CM052279
179	N/A		GGC-TGC	G382>C	_		CM035644
180	N/A		TGT-TAT	C388>Y	_		CM064155
181	N/A		TGC-CGC	C395>R			CM052280
182	rs863225297	16704C>G	TCT-TGT	S396>C	_	CADACI	CM125147
183	N/A		TGC-TAC	C408>Y	_	CADASIL	CM159179
184	N/A		TCC-TGC	S414>C	_		CM140150
185	N/A		TGC-TGG	C419>W	_		CM159180
186	N/A		GGT-TGT	G420>C	EGF-like 10		CM023652
187	N/A		CGT-TGT	R421>C	 		CM05228
188	N/A		CGC-TGC	R427>C			CM108354
190	N/A		TGT-CGT	C428>R			CM056019
191	N/A		TGT-TAT	C428>Y	_		CM052282
192	rs267606915	16897T>A	TGT-TCT	C428>S	-		CM014592
193	N/A		TGT-CGT	C435>R			CM03564
194	N/A		TGC-AGC	C440>S	=		CM056012
195	N/A		TGC-CGC	C440>R	_		CM052283
196	N/A		TGC-GGC	C440>G	_		CM023658
197	N/A		TGC-GGC	C446>G	_		CM141477
198	N/A		TGC-TCC	C446>S	_		CM04491
199	N/A		TGC-TTC	C446>F	EGF-like 11	CADASIL	CM035649
200	N/A		CGC-TGC	R449>C	_		CM02365
201	rs28933698	16978T>A	TGT-AGT	C455>S	_		
202	rs28933698	16978T>C	TGT-CGT	C455>R	_		CM021648
203	N/A		TGT-TAT	C455>Y	_		CM14015
204	N/A		TGT-TCT	C457>S	<u> </u>	CM101014	
205	N/A		TAT-TGT	Y465>C	-		CM035642
206	N/A		TGC-TAC	C466>Y	-		CM111183

19 of 36

			Not	tch3			
A/A	SNP ID	nt Change NG_009819.1	Codon Change	Mutation	Domain	Phenotype	Accession Number
207	N/A		TGC-TAT	C478>Y		CADASIL	CM150326
208	N/A		TGC-GGC	C484>G	_	CADASIL	CM125158
209	N/A		TGC-TAC	C484>Y	_	CADASIL	CM044909
210	N/A		TGC-TTC	C484>F	_	CADASIL	CM052284
211	N/A		TGC-TAC	C495>Y	_	CADASIL	CM052285
212	rs11670799	17742C>T	CCC-CTC	P496>L	EGF-like 12	Ischemic stroke, Alzheimer's disease, cerebral small-vessel disease	
213	rs114207045	17745C>G	TCG-TGG	S497>W	_	CADASIL, white matter lesions ?	
214	rs114207045	17745C>T	TCG-TTG	S497>L	-	CADASIL, white matter lesions ?	CM119547
215	N/A		TGT-CGT	C504>R	_	CADASIL	CM0911339
216	N/A		TGC-CGC	C511>R	_	CADASIL	CM052286
217	N/A		TGC-TAC	C511>Y	_	CADASIL	HM050011
218	N/A		TGC-TTC	C511>F		CADASIL	CM115174
219	N/A		GGC-TGC	G528>C		CADASIL	CM056020
220	N/A		TGC-GGC	C531>G		Leukoencephal vascular	opathy CM175386
221	N/A		TGC-TCC	C531>S	_	CADASIL	HM0684
222	N/A		CGC-TGC	R532>C			HM070085
223	N/A		TGT-CGT	C542>R			CM140152
224	N/A		TGT-TAT	C542>Y	_		CM961045
225	N/A		CGC-TGC	R544>C	- ECE l:l., 12	CADASI	CM994179
226	N/A		TGC-CGC	C549>R	- EGF-like 13	CADASIL	CM035645
227	N/A		TGC-TAC	C549>Y	_		CM052287
228	N/A		CGC-TGC	R558>C	_		CM961046
229	N/A		TGT-TAT	C568>Y	_		HM0710
230	N/A		TAC-TGC	Y574>C			HM0685
231	N/A		ACA-GCA	T577>A	ECE 1:1 14	CADACH	HM0718
232	N/A		CGC-TGC	R578>C	EGF-like 14	CADASIL	CM961047
233	N/A		TGC-CGC	C579>R			CM121680

Notch3										
A/A	SNP ID	nt Change NG_009819.1	Codon Change	Mutation	Domain	Phenotype	Accession Number			
234	N/A		CGC-TGC	R587>C			CM061879			
235	N/A		TGC-CGC	C591>R	_	CADASIL	CM125164			
236	N/A		CGC-AGC	R592>S	_	White matter lesions?	CM0911495			
237	N/A		CGC-TGC	R592>C	_		CM107182			
238	N/A		TGC-TAC	C597>Y			CM1615083			
239	N/I		TGC-TCC	C597>S	- EGF-like 15		CM119361			
240	N/A		TGC-TGG	C597>W	_		CM1615018			
241	N/A		TGC-CGC	C606>R	_	CADASIL	CM125148			
242	N/A		CGC-CAC	R607>H	_		CM1615019			
243	N/A		CGC-TGC	R607>C	_		CM003019			
244	N/A		TGC-TGG	C617>W	_		CM1610629			
245	N/A		GTC-GTT	V633>V		CADASIL	CM124698			
246	N/A		CGC-TGC	R640>C		CADASIL	CM125168			
247	N/A		GTC-GAC	V644>D	– EGF-like 16	White matter lesions?	CM0911496			
248	N/A		GGC-TGC	G667>C		CADASIL	CM125169			
249	rs10406745	20390G>C	CGC-CCC	R680>P	EGF-like 17	CADASIL				
250	rs10406745	20390G>A	CGC-CAC	R680>H	_	CADASIL				
251	N/A		CGC-TGC	R680>C	_	CADASIL with intracerebral haemor- rhage?	CM122007			
252	N/A		CCT-ACT	P685>T		CADASIL	CM111340			
253	N/A		TAT-TGT	Y710>C			CM1313312			
254	N/A		CGC-TGC	R717>C			CM1414856			
255	N/A		CGC-TGC	R728>C	EGF-like 18	CADASIL	CM971067			
256	N/A		TGC-GGC	C729>G	_		CM1714281			
257	N/A		GTC-GCC	V764>A	EGF-like 19	White matter lesions?	CM119548			
258	N/A		TGC-TCC	C775>S		CADASIL	CM052288			
259	N/A		CGC-TGC	R785>C	EGF-like 20	CADASIL	CM1413345			
260	N/A		GAG-GAA	E813>E	EGF-like 21	CADASIL	CM1615020			
261	N/A		TAC-TGC	Y916>C	EGF-like 23	CADASIL	CM1414857			
262	N/A		TGT-CGT	C939>R	EGF-like 24	CADASIL	CM125149			
263	N/A		GGC-TGC	G953>C	EGE-like 24	CADASIL	CM023660			

Notch3										
A/A	SNP ID	nt Change NG_009819.1	Codon Change	Mutation	Domain	Phenotype	Accession Number			
264	N/A		TGC-TGA	C966>TERM		Arteriopathy and cavitating leukoen- cephalopa- thy	CM152731			
265	N/A		GGC-TGC	G975>C		CADASIL with intracerebral haemorrhage	CM067439			
266	N/A		TGC-AGC	C977>S	-		HM050017			
267	N/A		TGC-GGC	C977>G	EGF-like 25		CM1615021			
268	N/A		AGC-CGC	S978>R	-		HM0711			
269	N/A		GCC-TCC	A979>S	-		CM1313735			
270	N/A		TTC-TGC	F984>C		CADACII	CM003020			
271	N/A		CGC-TGC	R985>C		CADASIL	CM971068			
272	N/A		TGC-CGC	C986>R	_		CM1414774			
273	N/A		TGC-TAC	C988>Y	_		CM062927			
274	N/A		TGC-TTC	C988>F	_		CM1714282			
275	N/A		TGC-GGC	C997>G	-		HM070152			
276	N/A		TGC-GGC	C1004>G	EGF-like 26	_	CM1714283			
277	N/A		TGC-TAC	C1004>Y			CM082987			
278	N/A		CGC-TGC	R1006>C			CM971069			
279	N/A		CCT-TCT	P1008>S			CM148526			
280	N/A		GGT-TGT	G1013>C			CM125150			
281	N/A		TGC-AGC	C1015>S	_		CM156030			
282	N/A		TGC-CGC	C1015>R	_	CADASIL	CM994180			
283	rs35769976	25217G>T	GCC-TCC	A1020>S	EGF-like 26	CADASIL	_			
284	rs35769976	25217G>A	GCC-ACC	A1020>T	_					
285	rs35769976	25217G>C	GCC-CCC	A1020>P			CM085589			
286	N/A		TAT-TGT	Y1021>C			CM023661			
287	N/A		TGC-TTC	C1022>F	_		CM118356			
288	N/A		TGG-TCG	W1028>C			CM078549			
289	N/A		CGC-TGC	R1031>C	-		CM971070			
290	N/A		GGT-TGT	G1058>C			CM014592			
291	N/A		TGT-TAT	C1061>Y	FCE-like 27	CADASII	CM127977			
292	N/A		TAC-TGC	Y1069>C	- EGI-IIKE 2/	CIDAJIL	CM092088			
293	N/A		CGT-TGT	R1076>C	_		CM056016			

Notch3							
A/A	SNP ID	nt Change NG_009819.1	Codon Change	Mutation	Domain	Phenotype	Accession Number
294	N/A		ACC-TCC	T1098>S			CM106871
295	N/A		TGC-TAC	C1099>Y	_	CADASIL	HM0709
296	N/A		TAC-TGC	Y1106>C	EGF-like 28		CM1615022
297	N/A		ATG-GTG	M1107>V		Autism spectrum disorder?	CM187124
298	N/A		TGC-TGG	C1131>W		CADASIL	CM081358
299	rs112197217	26557C>G	CAC-CAG	H1133>Q	_	CADASIL	
300	rs112197217	26557C>A	CAC-CAA	H1133>Q	EGF-like 29	Alzheimer's disease, ischemic stroke, CADASIL	
301	rs112197217	26557C>T	CAC-CAT	H1133>H	_	CADASIL	
302	N/A		GGG-CGG	G1134>R	_	AUTISM -	CM124589
303	N/A		CGC-TGC	R1143>C	_	CADASIL	CM1715079
304	N/A		CGG-TGG	R1175>W		CADASIL -	CM159359
305	rs10408676	26786G>A	GTG-ATG	V1183>M	EGF-like 30	Alzheimer disease, modifier of?	CM186058
306	N/A		GAC-GAG	D1184>E	_	CADASIL	CM159173
307	N/A		CGC-TGC	R1210>C			CM1615084
308	rs199638166	26903T>G	TGC-GGC	C1222>G		CADASIL	CM1414775
309	rs201680145	26930C>T	CGT-TGT	R1231>C	_	CADASIL	CM971071
310	rs201680145	26930C>A	CGT-AGT	R1231>S	EGF-like 31	Alzheimer's disease	
311	N/A		CAT-CTT	H1235>L		White matter lesions?	CM0911497
312	N/A		TGC-TGG	C1250>W	_	CADASIL	HM090055
313	N/A		TGC-CGC	C1261>R	_	CADASIL	CM961048
314	N/A		TGC-TAC	C1261>Y	EGF-like 32	CADASIL	CM052289
315	N/A		CGT-CTT	R1262>L		White matter lesions?	CM119549
316	N/A		TGC-TGG	C1293>W	_	CADASIL	CM186164
317	N/A		TGC-TTC	C1298>F	EGF-like 33	CADASIL with haemor- rhagic strokes	CM135094
318	N/A		TGC-TAC	C1315>Y		CADASIL	CM116899
319	N/A		GCC-ACC	A1450>T	LNR2	White matter lesions ?	CM119550
320	N/A		CTC-CCC	L1515>P	NOD	CADASIL	CM081357

Notch3							
A/A	SNP ID	nt Change NG_009819.1	Codon Change	Mutation	Domain	Phenotype	Accession Number
321	rs141320511	31730C>A	CTG-ATG	L1518>M	NOD	White matter lesions?	CM119551
322	rs141320511	31730C>T	CTG-TTG	L1518>L	NOD	White matter lesions, CADASIL, vascular dementia	
323	rs367543285	31734T>C	CTG-CCG	L1519>P	NOD	Migraine and white matter lesions	
324	N/A		CCA-CCG	P1521>P	NOD	Alzheimer disease, modifier of?	CM186056
325	rs78501403	31857G>C	CGG-CCG	R1560>P	NOD	Ischemic stroke, cerebral small-vessel disease	
326	rs78501403	31857G>A	CGG-CAG	R1560>Q	NOD	Ischemic stroke, cerebral small-vessel disease	
327	N/A		GGT-GAT	G1710>D	RAM	White matter lesions?	CM119552
328	N/A		GTG-ATG	V1762>M	RAM	CADASIL	CM1212986
329	N/A		GAC-AAC	D1823>N	NLS	White matter lesions?	CM119553
330	rs115582213	43458G>A	GTG-ATG	V1952>M	ANK	Alzheimer disease (modifier of?)	CM186057
331	rs142007575	44385G>A	GTA-ATA	V2011>1	NLS	Ischemic stroke	
332	rs114447350	44575C>T	CCG-CTG	P2074>L	unknown	Alzheimer's disease	
333	rs114447350	44575C>A	CCG-CAG	P2074>Q	unknown	Alzheimer's disease	
334	rs1044009	45022C>G	GCG-GGG	A2223>G	PEST	Alzheimer's disease, CADASIL	
335	rs1044009	45022C>T	GCG-GTG	A2223>V	PEST	Alzheimer's disease, CADASIL	

Notch4						
A/A	Mutation	Domain	Phenotype	Accession Number		
1	EGF-like 6	Schizophrenia	CM099076	EGF-like 6		
2	EGF-like 8	Multiple sclerosis	CM133099	EGF-like 8		
3	EGF-like 21	Migraine severity	CM134239	EGF-like 21		
4	NOD	Migraine duration	CM134237	NOD		

Table 7. Mutations in Notch4 associated with neurodegenerative diseases.



Figure 6. Number of associated and unassociated mutations with known SNPs.

The percentage of mutations in Notch3 associated with neurodegenerative diseases is displayed graphically in a pie chart in Figure 7. According to this study, 90% of Notch3 mutations lead to CADASIL disease, 4% of Notch3 mutations lead to Alzheimer's disease, and 4% of Notch3 mutations lead to white matter lesions. Only 2% of Notch3 mutations are associated with other neurodegenerative diseases such as the small-vessel disease of the brain, ischemic stroke, migraine, and autism. Since CADASIL is represented by 90% of mutations in Notch3, a guide map (Figure 4) was created for Notch3 mutations. Mutations in the amino acid sequence of the Notch3 protein are marked in bold red.



Figure 7. Percentage of Notch3 mutations associated with a specific neurodegenerative disease.

There have been reported to be 310 mutations in Notch3 that cause CADASIL syndrome (Table 7). The majority of the mutations (305), as shown in the amino acid sequence of Figure 4, are found in the EGF domain. The distribution of mutations in the EGF-like repeats and other protein domains is illustrated through a chart (Figure 8). The highest concentration of mutations is observed in the EGF-like 3 and EGF-like 4 repeats while the lowest numbers of mutations are found in the NOD, RAM, and PEST domains. In addition, more than 60% of Notch3 protein mutations that lead to CADASIL disease occur at the cysteine residue (Figure 9).



Figure 8. Demonstration of mutation number per Notch3 domain.



Figure 9. Mutated amino acid residues in Notch3 protein associated with CADASIL.

3.4. Mutation Analysis

Since 90% of the mutations identified in Notch3 and related to neurodegenerative diseases were located in the EGF region, the mutation analysis was mainly focused on this specific region. The multiple-sequence alignment (MSA) of amino acid sequences of Notch3 EGF-like repeats was performed to identify highly conserved amino acids within 34 EGF-like repeats. EGF-like repeats have a significant role in Notch signaling [2]. Six cysteine residues in each EGF repeat generate disulfide bonds affecting their native three-dimensional structure. Consequently, they are a crucial component of the EGF domain, and mutations in these residues lead to a pathological phenotype, specifically in CADASIL syndrome [2,30,49]. As shown by the present multiple alignment in the visualized results of the histogram "conservation" (Figure 10), cysteine residues are conserved within all EGF-like repeats. More particularly, the "consensus" histogram (Figure 8) shows the percentage of conserved amino acids at each position. Based on the data, cysteine residues are 100% conserved at positions 27, 41, 43, and 52 while positions 6 and 21 are 97% and 94% conserved, respectively. Additionally, glycine is conserved at positions 49 (97%), 46 (91%), 25 (71%), and 24 (76%), and proline is conserved at position 20 (74%). The greatest concentration of mutations appears at the cysteine residues. Almost 60% of cysteines at position 27 and 50% at positions 6 and 52 were identified as mutated. Results for the identified mutation percentage in each amino acid position of the EGF sequences

	10	20	30	40	50
1/1-38	- APPCLDGS	P C A N G	RCTQLPS	REAACLCPF	GWVGERCQ
2/1-41	- L E D P C H S G	P C A G R (VCQSSVV-A	GTARFSCRCPF	CFRGPDCS
3/1-38	- L P D P C L S S	P C A H G A	RCSVGPD	G R F L C S C P F	'GYQGR SCR
4/1-38	- DVDECRVG	EPCRHGO	TCL-NTP	GSFRCQCPA	AGYTGPLCE
5/1-38	- P A V P C A P S	P C R N G (TCRQSGD	- LT - YDCACLF	GFEGQNCE
6/1-37	- NVDDC - PG	HRCLNG	TCV-DGV	NTYNCQCPF	EWTGQFCT
7/1-39	- DVDECQLQP	NACHNGO	TCF-NTL	GGHSCVCVN	GWT G E S C S
8/1-37	- N I D D C A T A	VC F H G A	TCH-DRV	ASFYCACPN	AGKTGLLCH
9/1-39	- LDDACVSN	P C H E D A	LICDTNPV-N	IG RAICTCPF	GFTGGACD
10/1-39	- DVDECSIGA	NP C E H L C	GRCV-NTQ	GSFLCQCGF	GYTGPRCE
11/1-37	- DVNECLSGP	CRNQAT-	- CL - DRI	GQ FTCICMA	GFTGTYCE
12/1-37	- D I D E C Q S S	PCVNGO	GVCK-DRV	NGFSCTCP	GFSGSTCQ
13/1-37	- DV D E C A S T	P C R N G A	4 K C V - DQ P	DGYECRCAE	GFEGTLCD
14/1-36	- N V D D C S P D	P C H H G -	RCV-DGI	ASFSCACAF	GYTGTRCE
15/1-37	- Q V D E C R S Q	P C R H G (KCL-DLV	DKYLCRCPS	GTTGVNCE
16/1-36	- N I D D C A S N	P C T F G -	VCR - DGI	NRYDCVCQF	GFTGPLCN
17/1-37	– E I N E C A S S – – – – – –	P C G E G (5 C V - D G E	NGFRCLCP	GSLPPLCL
18/1-36	- P S H P C A H E	P C S H G -	ICYDAP	G G F R C V C E F	GWSGPRCS
19/1-37	- A R D A C E S Q	P C R A G (GTCSSDGN	1G F H C T C P F	GVQGRQCE
20/1-38	- L L S P C T P N	P C E H G (G R C E S A P	- GQLPVCSCPC	GWQGPRCQ
21/1-38	- DVDECAGP	A P C G P H (GICT - NLA	GSFSCTCH0	GYTGPSCD
22/1-37	- D I N D C D P N	P C L N G (5 S C Q - D G V	G S F S C S C L F	GFAGPRCA
23/1-36	- DVDECLS	NPCGP - 0	GT <u>C</u> T - DHV	A S F T C T C P F	GYGGFHCE
24/1-37	- D L P D C S P S	S C F N G 🤆	5TCV - DGV	NSFSCLCRF	GYTGAHCQ
25/1-37	- EADPCLSR	<u>P</u> C L H G C	GVCSAAH	P G F R C T C L E	SFTGPQCQ
26/1-35	- L V DWC S RQ	PCQNG	RCVQTG	AYCLCPF	GWSGRLCD
27/1-47	- R S L P C R E A A A Q I G V	RLEQLCQAG	5 Q C V D E D S	SHYCVCP	E G R T G S H C E
28/1-37	- EVDPCLAQ	P <u>C Q H</u> G (GTCR GY	- MGGYMCECLF	GYNGDNCE
29/1-38	DDVDECASQ	PCQHG(5 S C I - D L V	ARYLCSCPF	GTLGVLCE
30/1-44	- NEDDCGPGPPLD	- S G P <mark>R</mark> C L H N (GTCV-DLV	GGFRCTCPF	GYTGLRCE
31/1-40	- D I N E C R S G A	CHAAHT F	R DC LQ DP G	G G F R C L C H /	A G F S G P R C Q
32/1-42	- V L S P C E S Q	P <u>C</u> QHG(GQ <mark>C</mark> RPSPGP0	GGLT FTCH <u>C</u> AC) P FWGPRCE
33/1-37	- V A R S C R E L	Q C P V G \	/ P C Q Q T P	R G P R C A C P F	GLSGPSCR
34/1-39	- SNASCAAA	P C L H G (SCRPAP	LAPFFRCACAC	GWTGPRCE

are demonstrated in the histogram "mutations" (Figure 10). Most of the mutations are identified in positions 6, 27, and 52 of EGF-like repeats.



Figure 10. Conserved amino acids based on the sequence alignment of the EGF-like repeats. Each EGF-like repeat is presented with the specific number and the sequence length. The amino acids marked with the red square are the mutated ones.

The cysteine residues with the greatest level of conservation were analyzed to determine how frequently a certain amino acid change occurs at these positions (Table 8). The conserved cysteine residues are C21, C27, C41, C43, and C52. The frequency of each mutation is represented graphically in Figure 11. Based on the genetic background, more particularly based on the triplets that code for the amino acids replacing the cysteine, the appearance of the specific mutations was expected. The changing of one nucleotide in the triplets coding for cysteine (TGT, TGC) leads to the coding of another amino acid that has two common nucleotides with cysteine. Despite the fact that all cases of cysteine mutations associated with CADASIL syndrome have been reported (Table 8), it is vital to note that no nonsense mutation (TGA) has been reported as a cause of CADASIL. Although, based on the genetic code, the specific mutations are expected, their different frequencies of occurrence lead to the conclusion that these mutations are also related to genetic drift. Most frequently, cysteines were found to be mutated into arginine and tyrosine. This analysis also revealed that cysteines C27S/R/Y/W/F (EGF 5), C43S/R/Y/W/F (EGF 2), and C52S/R/G/F/Y (EGF 4) appeared to be more sensitive to pathogenic changes.

Table 8. Conserved amino acid changes in cysteine residues at positions (C6, C21, C27, C41, C43, and C52) of the EGF-like repeats of the Notch3 protein, accompanied with the frequency of their appearances and nucleotide changes (marked in red).

A/A		Amino Acid	Frequency	Codons
1		С	Given	TGT, TGC
2		R	8	C GT, C GC
3	9	Y	6	T <mark>A</mark> T, TAC
4	tion	F	3	Т <mark>Т</mark> Т, Т <mark>Т</mark> С
5	Posi	S	3	A GT TCT, A GC, TCC
6		W	3	TG <mark>G</mark>
7		G	0	G GT, GGC
1	_	С	Given	TGT, TGC
2		R	6	C GT, C GC
3	21	Y	6	T <mark>A</mark> T, TAC
4	tion	F	4	Т <mark>Т</mark> Т, Т <mark>Т</mark> С
5	Posi	S	2	A GT TCT, A GC, TCC
6		W	2	TG <mark>G</mark>
7		G	2	<mark>G</mark> GT, GGC
1		С	Given	TGT, TGC
2		R	7	C GT, C GC
3	27	Y	13	T <mark>A</mark> T, TAC
4	tion (F	5	Т <mark>Т</mark> Т, Т <mark>Т</mark> С
5	Posi	S	7	A GT TCT, A GC, TCC
6	_ · · _	W	4	TG <mark>G</mark>
7		G	8	<mark>G</mark> GT, GGC

A/A		Amino Acid	Frequency	Codons
1		С	Given	TGT, TGC
2		R	8	C GT, C GC
3	41 —	Y	7	T <mark>A</mark> T, TAC
4	lion	F	5	Т <mark>Т</mark> Т, Т <mark>Т</mark> С
5	Posit	S	6	A GT TCT, A GC, TCC
6		W	1	TG <mark>G</mark>
7		G	4	G GT, GGC
1		С	Given	TGT, TGC
2		R	5	C GT, C GC
3	43	Y	9	T <mark>A</mark> T, TAC
4	lion -	F	3	Т <mark>Т</mark> Т, Т <mark>Т</mark> С
5	 Posit	S	5	A GT TCT, A GC, TCC
6		W	3	TG <mark>G</mark>
7		G	1	G GT, GGC
1		С	Given	TGT, TGC
2		R	8	C GT, C GC
3	22 –	Y	9	T <mark>A</mark> T, TAC
4	lion	F	3	Т <mark>Т</mark> Т, Т <mark>Т</mark> С
5	Posit	S	6	A GT TCT, A GC, TCC
6		W	4	TG <mark>G</mark>
7		G	3	G GT, GGC



Figure 11. Presentation of cysteine mutations at positions (C6, C21, C27, C41, C43, and C52) based on the frequency of occurrence of a specific mutation. The chart's bold-colored columns represent the total number of conserved cysteine mutations of 34 EGF-like repeats.

3.5. Structural Analysis

The structural analysis of the mutations made it possible to understand the implications of inserting specific mutations into the amino acid sequence of the Notch3 EGF domain. Each EGF-like repeat consists of a set of two anti-parallel β -sheets (Figure 12). The structural stability of the Notch3 protein is maintained via disulfide bridges established by a set of strategically positioned cysteine residues [50]. As the key element of the domain, the six cysteine residues of the EGF-like repeat are crucial for the creation of disulfide bonds determining the native 3D structure of Notch proteins (Figure 12B).



Figure 12. Structural representation of the EGF domain and the EGF-like 2 repeat of the Notch3 protein. (**A**) Structure of EGF domain. (**B**) Structure of EGF-like 2 wild-type repeat.

Mutations in the cysteine residue at position 27 of EGF2 partly rearranged and destabilized the structure of EGF-like repeat due to the destruction of the disulfide bridge between the mutant cysteine and another cysteine residue (Figure 13). A change in the structure of the EGF 2 repeat was induced in each case of cysteine mutation.

Cysteine residue has a polar, uncharged side chain. The thiol group imparts polarity to cysteine. The induced C27R mutation leads to a broken disulfide bond between the mutant cysteine in one of the β sheets and its interacting cysteine in the coil opposite the β -sheet (Figure 13A). Arginine is a positively charged amino acid with a long side chain that carries a charged guanidine group. The full positive charge of the arginine side chain interacted with the partial negative charge of the cysteine side chain in this study. Even though there was no change in the β -sheet structure, this mutation seemed to affect the coil where the cysteine is located, causing the expansion of the coil because arginine is larger than cysteine. Thus, with the opening of the peptide chain, the space that could accommodate the arginine was created, causing the disturbance of the EGF-like 2 structures.

The C27Y and C27F mutations did not significantly change the structure of EGF despite breaking the disulfide bond. The C27Y mutation results in the differentiation of the coil opposite to tyrosine since tyrosine is larger than cysteine (Figure 13B). Tyrosine is a non-polar amino acid that contains an aromatic side chain. Due to the presence of the hydroxyl group in the side chain, tyrosine is predicted to interact via hydrogen bonding with cysteine. C27F mutation leads to a similar differentiation of the EGF 2 coil since phenylalanine is also an aromatic amino acid with a larger side chain than cysteine (Figure 13C). Phenylalanine is a non-polar amino acid and does not interact with cysteine.



Figure 13. Structural representation of the mutated EGF2 repeat of the Notch3 protein in the cases where the cysteine at position 27 mutated: (**A**) C27R, (**B**) C27Y, (**C**) C27F, (**D**) C27W, (**E**) C27S, and (**F**) C27G. In the EGF structure, the anti-parallel β -sheets are shown in yellow and the location of the mutated cysteine is shown in orange.

The C27W mutation destroys the disulfide bond and causes a partial loss of the structure of both anti-parallel β sheets (Figure 13D). The large, conjugated side chain of tryptophan got away from the EGF2 core. Due to the lack of available space, 27W was expected to be outside the β -sheet structure. On the contrary, it was found in the available space at the same level defined by the two β -sheets. As a result, the two β -sheets lost their original form as pleated surfaces and were converted into a coil structure.

The C27S mutation destroyed a disulfide bond in the EGF-like 2 structure (Figure 13E). This led to the rearrangement of the coil structure where the cysteine, which interacts with the serine, is located. Dipolar–dipolar interactions between serine and cysteine may

have caused this rearrangement. Consequently, an α helix is formed in this region of the EGF2 structure.

The C27G mutation destroys the disulfide bond and causes the partial loss of the two anti-parallel β -sheet structures (Figure 13F). Glycine is a non-polar amino acid that carries a non-polar aliphatic side chain. In the absence of a side group, there is no stere-ochemical barrier for glycine, allowing it to adopt a variety of conformations that could result in polypeptide chain curving and enhanced flexibility. This feature of glycine may also be responsible for the partial loss of β -sheets and rearrangement into coils.

4. Discussion

Decades of research have shown the significance of the Notch signaling pathway in neural development. More recent studies have proven that Notch receptors continue to be expressed and active in numerous areas of the adult central nervous system [51–53].

Adult neurogenesis, memory, synaptic plasticity, acute brain trauma, and chronic neurological diseases have all been linked to Notch signaling [22]. The analysis of mutation datasets revealed that human Notch1, Notch2, and Notch4 proteins are not significantly associated with neurodegenerative diseases [21]. On the other hand, most mutations in Notch3 lead to neurodegenerative diseases, mainly CADASIL syndrome [21,33]. Consequently, the current in silico study yielded new insights that might contribute to a better understanding of the correlation between neurodegenerative disorders and the human Notch family.

Considerable focus was given to analyzing Notch3 protein mutations associated with CADASIL disease. The study of mutations in the Notch3 protein is crucial because it could contribute to a better understanding of the molecular mechanisms that cause the disease, which is easier to study due to its monogenic nature. Even though the majority of mutations are point mutations, the effect of each on the three-dimensional structure of the Notch3 protein is significant [30,54]. Clinical genetics databases, including disease-specific mutation databases and genotype-phenotype research, provide a large amount of data on bioinformatics. Nevertheless, there is a scientific gap in linking the data provided by disease mutation databases and polymorphism databases [55]. Developing a database that provides a unified mapping of nucleotide sequences, protein sequences, and their protein domains, as well as polymorphisms and mutations related to human diseases, may pose a challenge for computational biology.

To date, a series of pathogenic mutations in Notch3 affecting the number of cysteine residues in the receptor's extracellular domain and resulting in protein misfolding and receptor aggregation have been identified [54]. Cysteine is the most active amino acid since it is involved in a wide range of biological functions [56]. Within extracellular proteins, cysteines are frequently involved in disulfide bridges in which pairs of cysteines are oxidized to create a covalent bond. Disulfide bonds' primary function is to stabilize protein structures. Cysteine generally has no preference for substituting with any other amino acid [57]. The reported cysteine substitutions in Notch3 ECD that induce CADASIL disease are arginine, tyrosine, phenylalanine, serine, glycine, and tryptophane. Generally, the extremely varied functions that cysteines play in extracellular proteins explain the below preferences for substitution: Arg (-5), Gly (-6), Tyr (-4), Phe (-5), Trp (-5), and Ser (-5) [33,57].

In addition, it has been observed that most frequently, the mutations associated with CADASIL occur in the first two nucleotides and much less frequently in the third nucleotide of the triplet that codes for the cysteine amino acid. Mutations in the first and second nucleotide of the cysteine triplet in this study led to the replacement of cysteine with arginine and glycine, as well as with serine, phenylalanine, and tyrosine. Cysteine substitution to tryptophane was noticed when a mutation occurred in the third nucleotide of the triplet. This study suggested that the first and second nucleotides are sensitive to mutations whereas the third nucleotide appears more conserved. Based on the genetic code, the occurrence of specific mutations was expected [58]. Also, the different frequency

of occurrence of each of these mutations is considered linked to the genetic draft, which slowly eliminates the variability that mutations cause, thereby achieving a steady state [59] (high frequency of cysteine mutating to tyrosine and arginine). Recording only one case of nonsense cysteine mutation in the Notch3 protein leads to two possible conclusions. There is the possibility that nonsense cysteine mutations lead to diseases, but no more cases of these mutations have been identified. It is also possible that cases of nonsense cysteine mutations have not been identified. Consequently, the only mutations identified are the ones that result in non-physiological protein function and therefore cause neurodegeneration. The frequent occurrence of mutations in cysteine residues that are highly conserved in the EGF-like repeats of Notch3 leads to protein misfolding and the manifestation of CADASIL syndrome [54].

Based on the results of the present work, which stem from the specialized study of the EGF-like domain of Notch3, several beneficial conclusions emerge. The accumulation of mutations appears to be different between EGF-like repeats 1 and 34, and these mutations were significantly increased in key amino acids in each EGF-like repeat such as in cysteine, glycine, and arginine (Figures 9 and 10). Today, with the increasing number of experimental data from patients with CADASIL syndrome, it is possible to create a mathematical model through which we will be able to relate the order and the series of mutations in different EGF-like repeats based on a specific phenotype of the disease, as well as based on sex and age [38,50,60,61]. Some studies also have made this observation [39]. In addition, based on the literature, we know the different phenotype in characteristics displayed by each patient that can perhaps be explained by the use of this mathematical model and the use of the above characteristics [61]. On the other hand, several attempts have been made to treat the disease based on the key amino acids that most mutations show in the EGF-like domain of Notch3 [33,40,60,62]. This particular work presents all the candidate positions as a holistic atlas in a detailed analysis of the changes both at the nucleotide and protein level for a contribution in this effort to fight CADASIL syndrome and neurodegenerative diseases in general [21,53].

5. Conclusions

To summarize, the present in silico study focused on analyzing mutations in Notch1– Notch4 proteins correlated with neurodegenerative diseases. The Notch pathway is crucial for the nervous system's development and pathogenesis due to its strong association with stem/progenitor cell progression and extensive pleiotropy. Neurodegenerative diseases are conditions characterized by the progressive and slow degeneration of neurons resulting in aberrant cell function and cell death. So far, no therapies that cure or prevent the progression of neurodegenerative diseases by targeting their underlying causes have been developed. Current therapies for these disorders are limited to symptom treatment. The integration of molecular methods, such as nanomedicine, genomics, proteomics, bioinformatics, and the measurement of environmental toxic body burdens, holds great promise for accelerating the process of identifying specific risk factors and mechanisms of pathogenesis in order to develop effective therapies for these diseases. Due to their role in cell fate determination and cell communication, as well as the proteolytic process they undergo in the signaling pathway, Notch proteins could be used as promising therapeutic targets for neurodegenerative diseases.

The ultimate aim of the in silico study was to uncover potential CADASIL diseasecausing conserved mutations and analyze the consequences of these mutations in the protein structure. Based on the results obtained from the present work, the correlation of *Notch3* polymorphisms—mutations with neurodegenerative diseases, especially in CADASIL syndrome—are clearly evident. In particular, the results show the accumulation of most of them in the EGF region of the protein. This specific protein region appears to be very crucial in the biomolecule's functionality, with changes in the EGF region appearing to lead to neurological pathologies. Through our analysis, we studied the contribution of specific sequence alterations, their frequency of occurrence at candidate sites in each EGF-like repeat, and their frequency of occurrence at specific key amino acids that appear to be conserved in each EGF-like repeat. In this direction, detailed molecular dynamics simulations showed that these conserved mutations trigger local rearrangements in the structure of the mutant EGF-like repeat of the Notch3 protein. The identified conserved mutations of cysteine residues could be used as supplementary pharmacological targets for the development of effective therapeutic schemes against CADASIL.

Since CADASIL syndrome is a monogenic disease, the opportunity to better interpret the mode of function of Notch proteins and their association with neurodegenerative diseases through mutations occurring in *Notch3* was utilized. Therefore, we propose the creation of a mathematical model through which we will be able to study, in detail, the importance and contribution of mutations in EGF-like repeats based on both their concentration, frequency of occurrence, and mutation pattern in each specific numbered EGF-like repeat as well as their detection in specific key positions described in this work. As it is known, EGF-like domains are prevalent in numerous protein families, suggesting that the employment of this specific mathematical model could potentially implicate both other proteins in neurodegenerative diseases as well as various other disorders. Furthermore, future objectives should encompass the comprehensive examination of the mutations delineated in this study, particularly those occurring within the intracellular domain, from both evolutionary and structural perspectives.

Author Contributions: Conceptualization, L.P. (Louis Papageorgiou), L.P. (Lefteria Papa), E.P., A.M., K.D., D.C., A.B., E.B.-R., G.P.C., S.K., E.E. and D.V.; methodology, L.P. (Louis Papageorgiou), L.P. (Lefteria Papa), K.D., A.B., C.I., E.B.-R., G.P.C., S.K., E.E. and D.V.; software, L.P. (Louis Papageorgiou), C.I. and D.V.; validation, L.P. (Louis Papageorgiou), L.P. (Lefteria Papa), E.E. and D.V.; formal analysis, L.P. (Louis Papageorgiou), L.P. (Lefteria Papa), E.E. and D.V.; formal analysis, L.P. (Louis Papageorgiou), L.P. (Lefteria Papa), E.P. and D.V.; investigation, L.P. (Louis Papageorgiou), S.K., E.E. and D.V.; resources, L.P. (Louis Papageorgiou), L.P. (Lefteria Papa), E.P., A.M., K.D., E.E. and D.V.; data curation, L.P. (Louis Papageorgiou), L.P. (Lefteria Papa), A.B., E.E. and D.V.; writing—original draft preparation, L.P. (Louis Papageorgiou), L.P. (Lefteria Papa), K.D., E.E. and D.V.; visualization, L.P. (Louis Papageorgiou) and L.P. (Louis Papageorgiou), D.C., S.K., E.E. and D.V.; visualization, L.P. (Louis Papageorgiou) and L.P. (Lefteria Papa); supervision, L.P. (Louis Papageorgiou), E.E. and D.V.; project administration, D.V. All authors agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- 1. Mohr, O.L. Character Changes Caused by Mutation of an Entire Region of a Chromosome in *Drosophila*. *Genetics* **1919**, *4*, 275–282. [CrossRef] [PubMed]
- Vlachakis, D.; Papageorgiou, L.; Papadaki, A.; Georga, M.; Kossida, S.; Eliopoulos, E. An updated evolutionary study of the Notch family reveals a new ancient origin and novel invariable motifs as potential pharmacological targets. *PeerJ* 2020, *8*, e10334. [CrossRef] [PubMed]
- Artavanis-Tsakonas, S.; Muskavitch, M.A.; Yedvobnick, B. Molecular cloning of Notch, a locus affecting neurogenesis in *Drosophila* melanogaster. Proc. Natl. Acad. Sci. USA 1983, 80, 1977–1981. [CrossRef] [PubMed]
- Yochem, J.; Weston, K.; Greenwald, I. The *Caenorhabditis elegans* lin-12 gene encodes a transmembrane protein with overall similarity to *Drosophila Notch*. *Nature* 1988, 335, 547–550. [CrossRef] [PubMed]
- Zhou, B.; Lin, W.; Long, Y.; Yang, Y.; Zhang, H.; Wu, K.; Chu, Q. Notch signaling pathway: Architecture, disease, and therapeutics. Signal Transduct. Target. Ther. 2022, 7, 95. [CrossRef] [PubMed]
- 6. Zhang, R.; Engler, A.; Taylor, V. Notch: An interactive player in neurogenesis and disease. *Cell Tissue Res.* **2018**, *371*, *73–*89. [CrossRef] [PubMed]
- Milner, L.A.; Bigas, A. Notch as a mediator of cell fate determination in hematopoiesis: Evidence and speculation. *Blood* 1999, 93, 2431–2448. [CrossRef] [PubMed]

- 8. Dumortier, A.; Wilson, A.; MacDonald, H.R.; Radtke, F. Paradigms of notch signaling in mammals. *Int. J. Hematol.* 2005, 82, 277–284. [CrossRef]
- 9. Shen, W.; Huang, J.; Wang, Y. Biological Significance of NOTCH Signaling Strength. *Front. Cell Dev. Biol.* 2021, *9*, 652273. [CrossRef]
- 10. Sanchez-Martin, M.; Ferrando, A. The NOTCH1-MYC highway toward T-cell acute lymphoblastic leukemia. *Blood* 2017, 129, 1124–1133. [CrossRef]
- 11. Garg, V.; Muth, A.N.; Ransom, J.F.; Schluterman, M.K.; Barnes, R.; King, I.N.; Grossfeld, P.D.; Srivastava, D. Mutations in *NOTCH1* cause aortic valve disease. *Nature* 2005, 437, 270–274. [CrossRef] [PubMed]
- 12. Xiu, M.X.; Liu, Y.M. The role of oncogenic Notch2 signaling in cancer: A novel therapeutic target. *Am. J. Cancer Res.* 2019, *9*, 837–854. [PubMed]
- 13. Arruga, F.; Vaisitti, T.; Deaglio, S. The NOTCH Pathway and Its Mutations in Mature B Cell Malignancies. *Front. Oncol.* **2018**, *8*, 550. [CrossRef] [PubMed]
- 14. Sugimoto, K.; Maekawa, Y.; Kitamura, A.; Nishida, J.; Koyanagi, A.; Yagita, H.; Kojima, H.; Chiba, S.; Shimada, M.; Yasutomo, K. Notch2 signaling is required for potent antitumor immunity in vivo. *J. Immunol.* **2010**, *184*, 4673–4678. [CrossRef] [PubMed]
- 15. Canalis, E.; Zanotti, S. Hajdu-Cheney Syndrome, a Disease Associated with NOTCH2 Mutations. *Curr. Osteoporos. Rep.* **2016**, 14, 126–131. [CrossRef]
- 16. Turnpenny, P.D.; Ellard, S. Alagille syndrome: Pathogenesis, diagnosis and management. *Eur. J. Hum. Genet.* **2012**, *20*, 251–257. [CrossRef] [PubMed]
- Manini, A.; Pantoni, L. CADASIL from Bench to Bedside: Disease Models and Novel Therapeutic Approaches. *Mol. Neurobiol.* 2021, 58, 2558–2573. [CrossRef]
- 18. Hosseini-Alghaderi, S.; Baron, M. Notch3 in Development, Health and Disease. Biomolecules 2020, 10, 485. [CrossRef] [PubMed]
- 19. Snijders, A.M.; Schmidt, B.L.; Fridlyand, J.; Dekker, N.; Pinkel, D.; Jordan, R.C.; Albertson, D.G. Rare amplicons implicate frequent deregulation of cell fate specification pathways in oral squamous cell carcinoma. *Oncogene* **2005**, *24*, 4232–4242. [CrossRef]
- Xiu, M.; Zeng, X.; Shan, R.; Wen, W.; Li, J.; Wan, R. Targeting Notch4 in Cancer: Molecular Mechanisms and Therapeutic Perspectives. *Cancer Manag. Res.* 2021, 13, 7033–7045. [CrossRef]
- Kapoor, A.; Nation, D.A. Role of Notch signaling in neurovascular aging and Alzheimer's disease. Semin. Cell Dev. Biol. 2021, 116, 90–97. [CrossRef] [PubMed]
- Guerra, J.; Valadao, A.L.; Vlachakis, D.; Polak, K.; Vila, I.K.; Taffoni, C.; Prabakaran, T.; Marriott, A.S.; Kaczmarek, R.; Houel, A.; et al. Lysyl-tRNA synthetase produces diadenosine tetraphosphate to curb STING-dependent inflammation. *Sci. Adv.* 2020, *6*, eaax3333. [CrossRef] [PubMed]
- 23. Basak, O.; Giachino, C.; Fiorini, E.; Macdonald, H.R.; Taylor, V. Neurogenic subventricular zone stem/progenitor cells are Notch1-dependent in their active but not quiescent state. *J. Neurosci.* **2012**, *32*, 5654–5666. [CrossRef]
- Nyfeler, Y.; Kirch, R.D.; Mantei, N.; Leone, D.P.; Radtke, F.; Suter, U.; Taylor, V. Jagged1 signals in the postnatal subventricular zone are required for neural stem cell self-renewal. *EMBO J.* 2005, 24, 3504–3515. [CrossRef]
- Lv, Y.; Pang, X.; Cao, Z.; Song, C.; Liu, B.; Wu, W.; Pang, Q. Evolution and Function of the Notch Signaling Pathway: An Invertebrate Perspective. *Int. J. Mol. Sci.* 2024, 25, 3322. [CrossRef]
- 26. Sprinzak, D.; Blacklow, S.C. Biophysics of Notch Signaling. Annu. Rev. Biophys. 2021, 50, 157–189. [CrossRef]
- Urban, N.; van den Berg, D.L.; Forget, A.; Andersen, J.; Demmers, J.A.; Hunt, C.; Ayrault, O.; Guillemot, F. Return to quiescence of mouse neural stem cells by degradation of a proactivation protein. *Science* 2016, 353, 292–295. [CrossRef]
- Dugger, B.N.; Dickson, D.W. Pathology of Neurodegenerative Diseases. Cold Spring Harb. Perspect. Biol. 2017, 9, a028035. [CrossRef] [PubMed]
- Jagaran, K.; Singh, M. Nanomedicine for Neurodegenerative Disorders: Focus on Alzheimer's and Parkinson's Diseases. Int. J. Mol. Sci. 2021, 22, 9082. [CrossRef] [PubMed]
- Vlachakis, D.; Tsaniras, S.C.; Ioannidou, K.; Papageorgiou, L.; Baumann, M.; Kossida, S. A series of Notch3 mutations in CADASIL; insights from 3D molecular modelling and evolutionary analyses. *J. Mol. Biochem.* 2014, *3*, 134.
- Ables, J.L.; Breunig, J.J.; Eisch, A.J.; Rakic, P. Not(ch) just development: Notch signalling in the adult brain. *Nat. Rev. Neurosci.* 2011, 12, 269–283. [CrossRef] [PubMed]
- 32. Wang, M.M. Cadasil. Handb. Clin. Neurol. 2018, 148, 733–743. [CrossRef] [PubMed]
- 33. Mizuta, I.; Nakao-Azuma, Y.; Yoshida, H.; Yamaguchi, M.; Mizuno, T. Progress to Clarify How NOTCH3 Mutations Lead to CADASIL, a Hereditary Cerebral Small Vessel Disease. *Biomolecules* **2024**, *14*, 127. [CrossRef] [PubMed]
- 34. Louvi, A.; Arboleda-Velasquez, J.F.; Artavanis-Tsakonas, S. CADASIL: A critical look at a Notch disease. *Dev. Neurosci.* 2006, 28, 5–12. [CrossRef] [PubMed]
- 35. Kalaria, R.N.; Kittner, S.J. Top-NOTCH3 Variants in the Population at Large. Stroke 2020, 51, 3482–3484. [CrossRef] [PubMed]
- Haltom, A.R.; Jafar-Nejad, H. The multiple roles of epidermal growth factor repeat O-glycans in animal development. *Glycobiology* 2015, 25, 1027–1042. [CrossRef]
- 37. Zeng, Q.; Pan, H.; Zhao, Y.; Wang, Y.; Xu, Q.; Tan, J.; Yan, X.; Li, J.; Tang, B.; Guo, J. Association between NOTCH3 gene and Parkinson's disease based on whole-exome sequencing. *Front. Aging Neurosci.* **2022**, *14*, 995330. [CrossRef] [PubMed]

- Szymanowicz, O.; Korczowska-Lacka, I.; Slowikowski, B.; Wiszniewska, M.; Piotrowska, A.; Goutor, U.; Jagodzinski, P.P.; Kozubski, W.; Dorszewska, J. Headache and NOTCH3 Gene Variants in Patients with CADASIL. *Neurol. Int.* 2023, 15, 1238–1252. [CrossRef] [PubMed]
- Rutten, J.W.; Van Eijsden, B.J.; Duering, M.; Jouvent, E.; Opherk, C.; Pantoni, L.; Federico, A.; Dichgans, M.; Markus, H.S.; Chabriat, H.; et al. The effect of NOTCH3 pathogenic variant position on CADASIL disease severity: NOTCH3 EGFr 1-6 pathogenic variant are associated with a more severe phenotype and lower survival compared with EGFr 7-34 pathogenic variant. *Genet. Med.* 2019, 21, 676–682. [CrossRef]
- 40. Matsushima, T.; Conedera, S.; Tanaka, R.; Li, Y.; Yoshino, H.; Funayama, M.; Ikeda, A.; Hosaka, Y.; Okuzumi, A.; Shimada, Y.; et al. Genotype-phenotype correlations of cysteine replacement in CADASIL. *Neurobiol. Aging* **2017**, *50*, 169.e7–169.e14. [CrossRef]
- Young, K.Z.; Lee, S.J.; Zhang, X.; Cartee, N.M.P.; Torres, M.; Keep, S.G.; Gabbireddy, S.R.; Fontana, J.L.; Qi, L.; Wang, M.M. NOTCH3 is non-enzymatically fragmented in inherited cerebral small-vessel disease. *J. Biol. Chem.* 2020, 295, 1960–1972. [CrossRef] [PubMed]
- Dichgans, M.; Ludwig, H.; Muller-Hocker, J.; Messerschmidt, A.; Gasser, T. Small in-frame deletions and missense mutations in CADASIL: 3D models predict misfolding of Notch3 EGF-like repeat domains. *Eur. J. Hum. Genet.* 2000, *8*, 280–285. [CrossRef] [PubMed]
- Wang, W.; Ren, Z.; Shi, Y.; Zhang, J. A Novel Mutation Outside of the EGFr Encoding Exons of NOTCH3 Gene in a Chinese with CADASIL. J. Stroke Cerebrovasc. Dis. Off. J. Natl. Stroke Assoc. 2020, 29, 105410. [CrossRef] [PubMed]
- 44. Mizuno, T.; Mizuta, I.; Watanabe-Hosomi, A.; Mukai, M.; Koizumi, T. Clinical and Genetic Aspects of CADASIL. *Front. Aging Neurosci.* **2020**, *12*, 91. [CrossRef]
- Sari, U.S.; Kisabay, A.; Batum, M.; Tarhan, S.; Dogan, N.; Coskunoglu, A.; Cam, S.; Yilmaz, H.; Selcuki, D. CADASIL with Atypical Clinical Symptoms, Magnetic Resonance Imaging, and Novel Mutations: Two Case Reports and a Review of the Literature. J. Mol. Neurosci. 2019, 68, 529–538. [CrossRef]
- 46. Servito, M.; Gill, I.; Durbin, J.; Ghasemlou, N.; Popov, A.F.; Stephen, C.D.; El-Diasty, M. Management of Coronary Artery Disease in CADASIL Patients: Review of Current Literature. *Medicina* **2023**, *59*, 586. [CrossRef]
- Mitchell, A.; Chang, H.Y.; Daugherty, L.; Fraser, M.; Hunter, S.; Lopez, R.; McAnulla, C.; McMenamin, C.; Nuka, G.; Pesseat, S.; et al. The InterPro protein families database: The classification resource after 15 years. *Nucleic Acids Res.* 2015, 43, D213–D221. [CrossRef] [PubMed]
- Chillakuri, C.R.; Sheppard, D.; Lea, S.M.; Handford, P.A. Notch receptor-ligand binding and activation: Insights from molecular studies. Semin. Cell Dev. Biol. 2012, 23, 421–428. [CrossRef] [PubMed]
- 49. Papakonstantinou, E.; Bacopoulou, F.; Brouzas, D.; Megalooikonomou, V.; D'Elia, D.; Bongcam-Rudloff, E.; Vlachakis, D. NOTCH3 and CADASIL syndrome: A genetic and structural overview. *EMBnet J.* **2019**, *24*, e921. [CrossRef]
- Muino, E.; Gallego-Fabrega, C.; Cullell, N.; Carrera, C.; Torres, N.; Krupinski, J.; Roquer, J.; Montaner, J.; Fernandez-Cadenas, I. Systematic Review of Cysteine-Sparing NOTCH3 Missense Mutations in Patients with Clinical Suspicion of CADASIL. *Int. J. Mol. Sci.* 2017, 18, 1964. [CrossRef]
- 51. Salazar, J.L.; Yang, S.A.; Yamamoto, S. Post-Developmental Roles of Notch Signaling in the Nervous System. *Biomolecules* **2020**, 10, 985. [CrossRef] [PubMed]
- 52. Ho, D.M.; Artavanis-Tsakonas, S.; Louvi, A. The Notch pathway in CNS homeostasis and neurodegeneration. *Wiley Interdiscip. Rev. Dev. Biol.* **2020**, *9*, e358. [CrossRef] [PubMed]
- Lampada, A.; Taylor, V. Notch signaling as a master regulator of adult neurogenesis. Front. Neurosci. 2023, 17, 1179011. [CrossRef] [PubMed]
- Lee, S.J.; Zhang, X.; Wu, E.; Sukpraphrute, R.; Sukpraphrute, C.; Ye, A.; Wang, M.M. Structural changes in NOTCH3 induced by CADASIL mutations: Role of cysteine and non-cysteine alterations. *J. Biol. Chem.* 2023, 299, 104838. [CrossRef] [PubMed]
- 55. Webb, E.A.; Smith, T.D.; Cotton, R.G. Difficulties in finding DNA mutations and associated phenotypic data in web resources using simple, uncomplicated search terms, and a suggested solution. *Hum. Genom.* **2011**, *5*, 141–155. [CrossRef] [PubMed]
- Nallapareddy, V.; Bogam, S.; Devarakonda, H.; Paliwal, S.; Bandyopadhyay, D. DeepCys: Structure-based multiple cysteine function prediction method trained on deep neural network: Case study on domains of unknown functions belonging to COX2 domains. *Proteins* 2021, *89*, 745–761. [CrossRef] [PubMed]
- 57. Barnes, M.R. *Bioinformatics for Geneticists a Bioinformatics Primer for the Analysis of Genetic Data*, 2nd ed.; Wiley: Chichester, UK, 2007. [CrossRef]
- 58. Saier, M.H.J. Understanding the Genetic Code. J. Bacteriol. 2019, 201. [CrossRef] [PubMed]
- 59. Brenner, S.; Maloy, S.R.; Hughes, K.T. Brenner's encyclopedia of genetics. In *Encyclopedia of Genetics*, 2nd ed.; Academic Press: New York, NY, USA, 2013.
- Locatelli, M.; Padovani, A.; Pezzini, A. Pathophysiological Mechanisms and Potential Therapeutic Targets in Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy (CADASIL). *Front. Pharmacol.* 2020, 11, 321. [CrossRef] [PubMed]

- 61. Dupe, C.; Guey, S.; Biard, L.; Dieng, S.; Lebenberg, J.; Grosset, L.; Alili, N.; Herve, D.; Tournier-Lasserve, E.; Jouvent, E.; et al. Phenotypic variability in 446 CADASIL patients: Impact of NOTCH3 gene mutation location in addition to the effects of age, sex and vascular risk factors. *J. Cereb. Blood Flow Metab. Off. J. Int. Soc. Cereb. Blood Flow. Metab.* 2023, 43, 153–166. [CrossRef]
- Oliveira, D.V.; Coupland, K.G.; Shao, W.; Jin, S.; Del Gaudio, F.; Wang, S.; Fox, R.; Rutten, J.W.; Sandin, J.; Zetterberg, H.; et al. Active immunotherapy reduces NOTCH3 deposition in brain capillaries in a CADASIL mouse model. *EMBO Mol. Med.* 2023, 15, e16556. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.