Unveiling the evolutionary code of NOTCH3: mammalian bioinformatics sheds light on human pathogenicity

Khawar Sohail Siddiqui[1], Haluk Ertan[2,3], Yi Ren[4], Anne Poljak[1,5], Tharusha Jayasena[1], Perminder Sachdev[1]

[1]Center for Healthy Brain Aging (CHeBA), School of Medical Sciences, The University of New South Wales (UNSW) Sydney, NSW 2052, Australia.

[2]School of Biotechnology and Biomolecular Sciences (BABS), The UNSW Sydney, NSW 2052, Australia.

[3]Department of Molecular Biology and Genetics, Istanbul University, Istanbul 34134, Turkiye

[4]Water Research Center (WRC), School of Civil and Environmental Engineering, The University of New South Wales (UNSW) Sydney, NSW 2052, Australia.

[5]Bioanalytical Mass Spectrometric Facility (BMSF), School of Medical Sciences, The UNSW Sydney, NSW 2052, Australia.

Abstract

NOTCH3 is a highly conserved transmembrane receptor implicated in CADASIL, a hereditary small vessel disease driven by mutations in its extracellular EGF-like repeats. The mechanism by which these mutations cause pathology remains unclear. We present the first large-scale comparative bioinformatic analysis of NOTCH3 across 113 mammalian species, uncovering three novel insights: i) a remarkable evolutionary conservation of all 204 cysteines, with the only exception being eight naturally occurring cysteine mutations in jaguar (EGFr13-15); ii) a unique deletion in Brandt's bat regulatory region, which may expose it to proteases, potentially altering signaling; iii) a rare human NOTCH3-X1 isoform, absent in most mammals but shared with select primates, a bat, and elephants, involving a cysteine-depleting deletion spanning EGFr20-22. These features provide novel evolutionary insights into human pathogenicity and suggest testable targets for *in vivo* experiments. Our study highlights the potential of comparative bioinformatics to identify previously hidden functional elements in disease-associated mammalian proteins.

Introduction

"*It is notorious that man is constructed on the same general type or model as other mammals. All the bones in his skeleton can be compared with corresponding bones in a monkey, bat, or seal. So it is with his muscles, nerves, blood-vessels and internal viscera. The brain, the most important of all the organs, follows the same law …"*[1]

NOTCH3 is a large transmembrane receptor essential for cell signaling and development in mammals. It comprises five domains: signal peptide, extracellular domain (ECD), negative regulatory region (NRR), transmembrane (TM), and intracellular domain (ICD). The ECD's 34 EGFr repeats enable ligand binding and stability, with cysteines forming disulfide bonds critical for integrity[2].

Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL) is a hereditary small-vessel disease caused by mutations in NOTCH3 EGFr[3]. Cysteine mutations disrupt disulfide bonds, leading to misfolding and GOM aggregation in brain microvasculature. Aggregation is a key factor in CADASIL pathogenesis, though its precise mechanisms remain unclear[4,5,6]. Notably, Cys>aa mutations are common and linked to more severe symptoms[7].

Recent large-scale efforts such as the Zoonomia Project have highlighted the value of mammalian comparative genomics for understanding gene function and disease evolution[8,9]. Genes that are highly conserved across diverse species are likely essential for normal biological function and, when altered, may contribute to disease. In contrast, genes that exhibit lineage-specific divergence may reflect adaptive evolution to ecological or physiological pressures[10].

Building on these large-scale efforts, we present the first comprehensive comparative bioinformatics analysis of NOTCH3 across 113 mammalian species, focusing on cysteines' roles in function and disease. Key findings include unique deletion isoforms disrupting disulfide bonds, a bat-specific deletion likely affecting the NRR activation, and jaguar-specific mutations eliminating disulfide bonds in an EGFr cluster. These results refine our understanding of the sequence–structure–function–pathogenicity relationship and challenge the notion that all cysteine disruptions are pathogenic. By highlighting cysteine's nuanced

role in NOTCH3 stability, the study informs CADASIL research and may guide future experimental and therapeutic strategies.

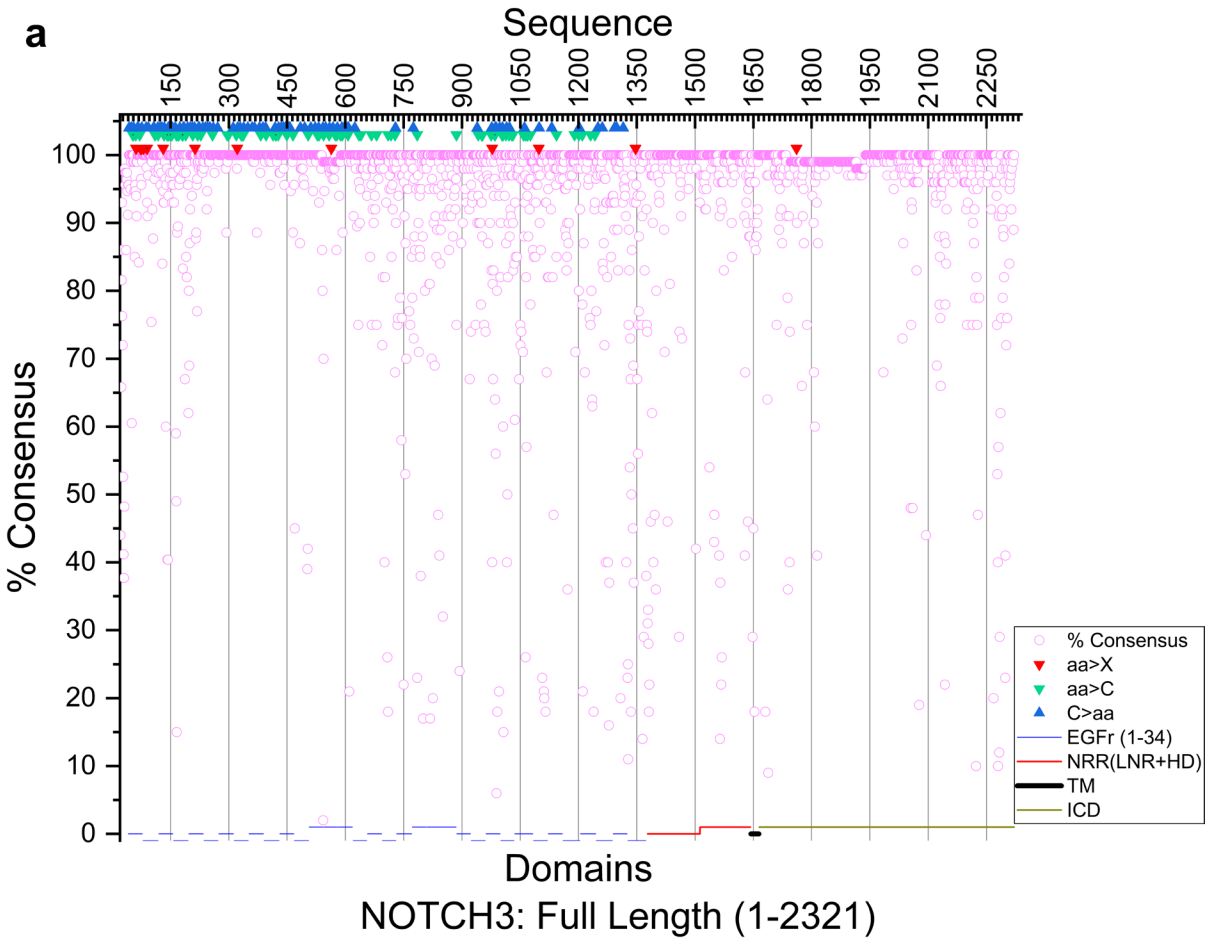
Results

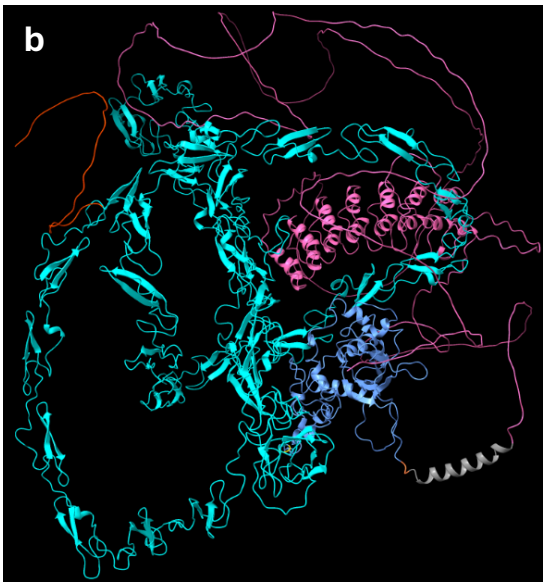
Using a comprehensive multiple-sequence alignment (MSA) of 113 mammalian NOTCH3 sequences, we uncovered novel patterns of conservation and divergence. By analyzing sequence and structural features, we gained insights into how deletions and missense mutations may impact function. Our findings reveal links between structural conservation, integrity, and disease, challenging the assumption that all EGFr cysteine mutations cause cerebrovascular pathology[11].

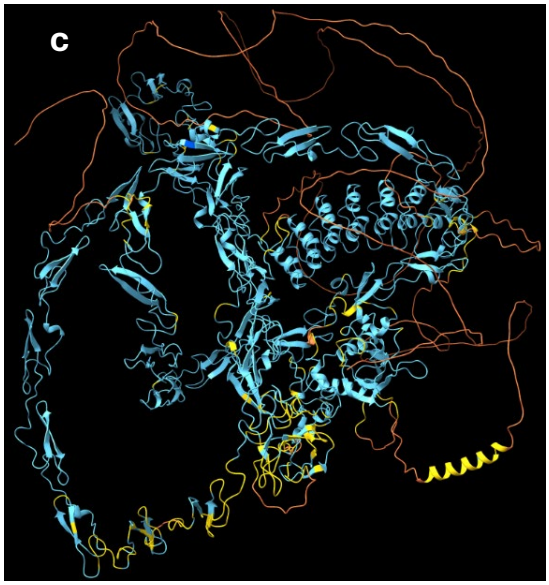
Mammalian NOTCH3 shows high sequence and structure conservation

We assessed conservation using MSA of NOTCH3 from 113 mammalian species (Supplementary Fig. 1) via Jalview (Supplementary Software 1). Combined with published pathogenicity data and domain annotations, results were visualized in Origin plots (Supplementary Software 2). Among primates, identity with human NOTCH3 ranged from 94.7% to 99.8%, while most other mammals showed ≥90% identity, except armadillo (89.7%). EGFr1–12 showed near-perfect identity, and the ICD was also highly conserved (95–100%) (Fig. 1a). The AlphaFold3-predicted full-length human NOTCH3 structure illustrates domain organization and model confidence (Fig. 1b, c), though 3D interpretation requires caution due to low-confidence inter-domain positioning. These conserved regions likely support NOTCH3's structural and functional stability across a broad range of mammals, implies a diversity of potential targets for exploring pathogenic mechanisms, and is consistent with a very large number of pathogenic mutations associated with CADASIL[11].

Fig. 1: Sequence conservation plot and structure of human NOTCH3.





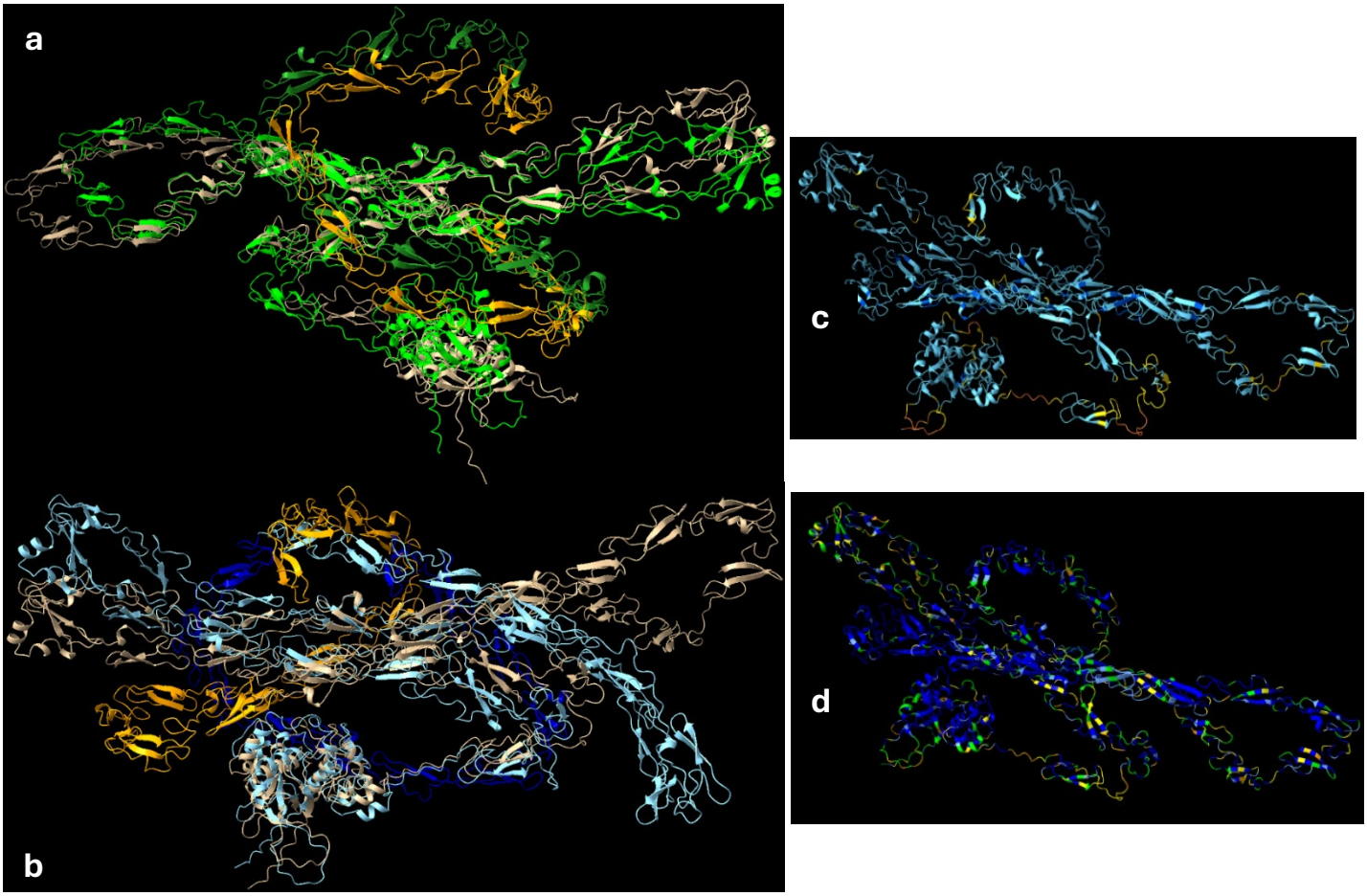


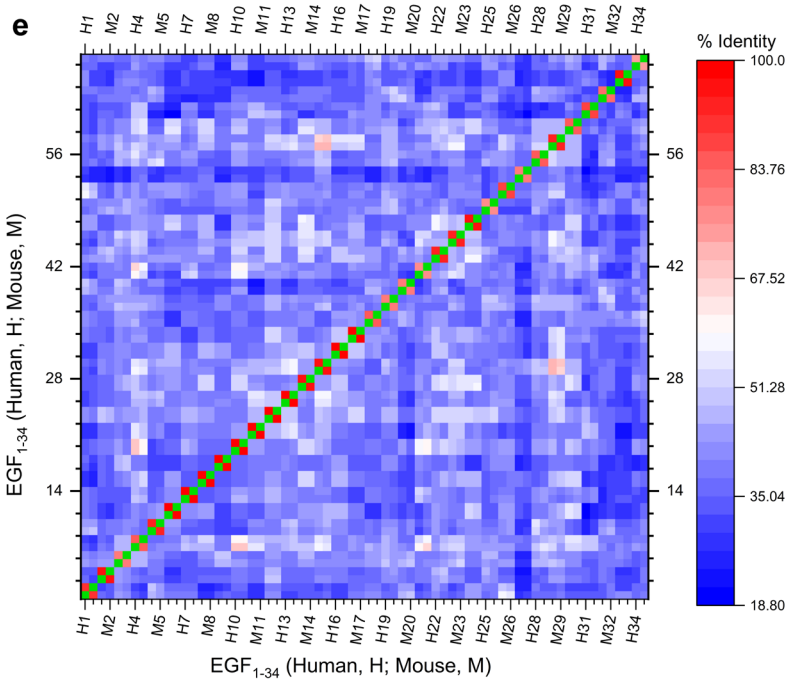
a, the Origin plot shows the NOTCH3 sequence with various domains, pathogenic mutations and % consensus with 113 mammalian NOTCH3 sequences. Domains are shown on the X-axis as colored segments. 34 EGFrs in ECD, blue; NRR (LNR + HD), red; TM, black; ICD, green; % consensus, pink circle. EGFr10,11 (391-467 aa) are involved in ligand-binding. Lifted bars indicate the EGFr region corresponding to jaguar EGFr13–15 (507-618 aa), and the X1 deletion spanning EGFr20–22 (771-885 aa). Various types of pathogenic mutations are shown as triangles: aa>X, red; aa>C, green; C>aa, blue. The interactive Origin plot features tooltips, where hovering the cursor over data points reveals the exact percent consensus, amino acid sequence, residue name, and mutation information (Supplementary Software 2). The exploded views of EGFr13-15 and EGFr20-22 are shown in Supplementary Figs 2b and 2c. **b and c,** Human full-length NOTCH3 structure predicted by AlphaFold3. **b**, color-coded domains—signal peptide (red), EGFr1–34 (cyan), NRR (blue), TM helix (grey), and ICD (pink). **c**, pLDDT confidence scores—blue: >90 (very high); light blue: 70–90 (confident); yellow: 50–70 (low); orange: <50 (very low). Although local domain structures are well defined, their spatial arrangement in the full-length model is uncertain due to low inter-domain confidence.

To explore evolutionary insights relevant to CADASIL, we selected four species—human, rhesus macaque, mouse, and naked mole-rat—for focused comparison based on biomedical relevance and unique traits (Supplementary Fig. 3). The mouse is widely used in CADASIL models; the rhesus macaque is phylogenetically close to humans; and the naked mole-rat is a long-lived rodent with notable anti-aging traits[12]. Interestingly, naked mole-rat NOTCH3 shares higher sequence identity with human (92.8%) than with mouse (91.0%), a notable exception within the typical 90–99% mammalian identity range. This trend persists in the ECD (EGFr1–34): 92.7% identity with human vs. 89.2% with mouse (Supplementary Software 2), suggesting the naked mole-rat may be a more suitable rodent model. Given the importance of structural fidelity in NOTCH3 signaling, this similarity likely preserves 3D architecture (Supplementary Fig. 3). Structural superposition confirms this: the human ECD aligns better with rhesus than with mouse, consistent with sequence identity (Figs. 2a, b). The first 11 EGF repeats in human and rhesus superimpose closely, reflecting conserved structure in CADASIL-prone and ligand-binding regions, while divergence beyond EGFr11 suggests differences in inter-repeat orientation, loop flexibility, or species-specific indels in more tolerant mid- and C-terminal regions.

The AlphaFold3-predicted structure of the human NOTCH3 ECD (Fig. 2c) shows high confidence for individual EGFr domains but low confidence in their relative orientation. A domain-wise conservation map across >113 mammals (Fig. 2d) highlights strong conservation in many EGFrs, aligning with sequence similarity and supporting their structural and functional significance. Sequence identity between corresponding EGFr domains in the human and mouse ECD was remarkably high (Fig. 2e, diagonal), suggesting evolutionary conservation driven by structural and functional constraints (Fig. 2a, b). However, comparing all 34 human and mouse EGFr domains revealed lower pairwise identity (21.6%–68.4%), with high variability across domains (Fig. 2e diagonal; Supplementary Fig. 4). The mean identity across EGFr1–34 was 41.1%; EGFr6 showed the lowest identity (21.6%) to EGFr31, while EGFr4 had the highest (68.4%) to EGFr10.

Fig. 2: Sequence and structural similarity in ECD.

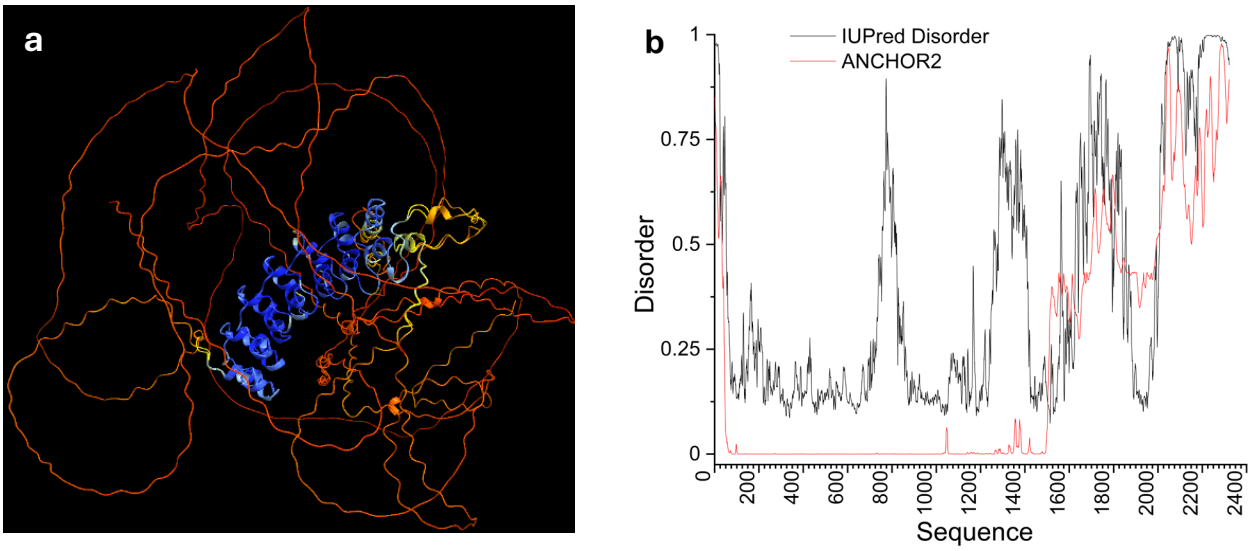




a, human and rhesus (RMSD across all 1603 residue pairs, 17.34 Å), **b**, human and mouse (RMSD across all 1603 residue pairs, 23.9 Å). Light orange/dark orange, human; light green/dark green, rhesus, light blue/dark blue, mouse. Darker shades represent EGFr1-11. **c**, human NOTCH3 extracellular domain structure predicted by AlphaFold3. **c**, colored according to plDDT confidence scores as in Fig. 1. The AlphaFold3-predicted structure of the human NOTCH3 ECD shows high confidence for individual EGFr domains but low confidence in their relative orientation. **d**, human ectodomain showing conservation among >113 mammalian species. Blue, 100%; corn-blue, 98-99%, lime, 90-97%; yellow, 50-89%; <50%, orange. Blue-light blue, high conservation; lime-yellow, medium conservation; orange, low conservation. The relative positioning of EGFr and NRR shows low confidence. **e**, sequence similarity among EGFr1–34 domains in human and mouse NOTCH3. Heatmap showing pairwise sequence identity within human EGFr domains (H vs H), within mouse EGFr domains (M vs M), and between human and mouse EGFr domains (H vs M). The green diagonal represents 100% identity for self-comparisons of the same EGFr domain; the orange-red diagonal indicates sequence identity between corresponding EGFr domains in human and mouse; off-diagonal values above and below the main diagonals show the percentage identity between non-corresponding EGFr domains within and between both species. Structurally, EGFr6 and EGFr31 had an RMSD of 3.815 Å (37 aligned residues), while EGFr4 and EGFr10 had an RMSD of 1.102 Å (38 residues). Other comparisons: EGFr4 vs. EGFr6, 41.7% identity, RMSD 3.81 Å; EGFr6 vs. EGFr10, 35.0%, RMSD 4.2 Å; EGFr4 vs. EGFr31, 50.0%, RMSD 1.0 Å.

To complement sequence analysis, we examined the 3D structures of the extracellular domain (ECD; Fig. 2a, b) and intracellular domain (ICD; Fig. 3a), along with IUPRED disorder predictions. AF3 predicted high confidence (pLDDT 70–90) for most of the first 30 EGFr domains, while EGFr20–21 and EGFr30–34 showed lower confidence (Fig. 2c), aligning with IUPRED disorder predictions (Fig. 3b). Despite high sequence conservation, the human NOTCH3 ICD appears largely disordered (Fig. 3a, orange wire), confirmed by IUPRED (Fig. 3b). Similar disorder was predicted in the ICDs of mouse (Fig. 3a), rhesus macaque, and naked mole-rat (Extended Data, Fig. 3). ANCHOR2 suggests the ICD may undergo disorder-to-order transitions upon binding partner proteins (Fig. 3b, red plot)—a dynamic feature often overlooked in classical structural studies[13]. In contrast, the EGFr1–34 and NRR domains are predicted to be largely ordered (Fig. 2c).

Fig. 3: Structure and analysis of intrinsically disordered regions.



a, superimposed human and mouse NOTCH3 ICD predicted by AF3. Refer to Fig. 1 for pIDDT confidence score. **b**, plot showing intrinsically disordered regions in the NOTCH3 sequence using two different tools: IUPred and ANCHOR2.

Exon 16 Skipping Produces ECD Deletion Isoform

The MSA revealed key differences in NOTCH3 ECD sequences across mammals, particularly in missense mutations and deletions. These deletions offer insights into how structural disruptions—especially of cysteine residues and disulfide bonds—may contribute to CADASIL pathology. A notable deletion corresponding to exon 16 was identified in several mammals, including higher primates, marmoset, elephants, and a vampire bat, but was absent in monkeys, capuchins, bats, and rodents. This suggests exon 16 skipping may exist in other, yet-unidentified, mammalian isoforms. Two whale isoforms showed a distinct upstream deletion (Fig. 4a). A conserved Gly>Asp missense mutation precedes this deletion in all affected species (Fig. 4a, arrow).

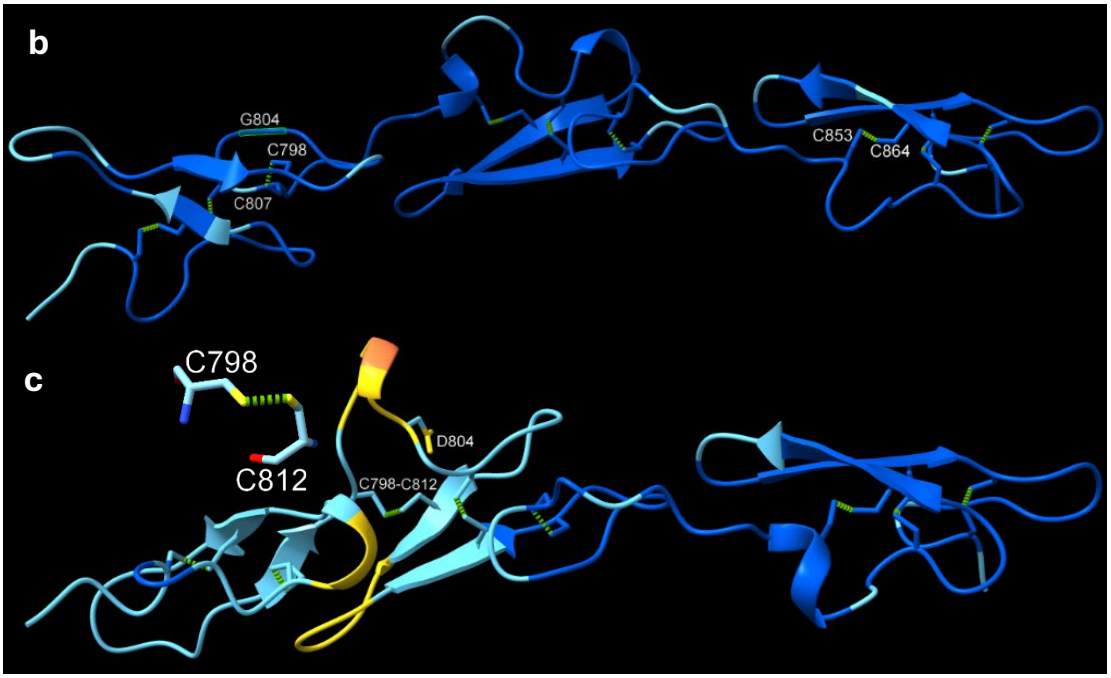
The deletion spans the final four residues of EGFr20, all of EGFr21, and the first eight of EGFr22—eliminating eight cysteines: one each in EGFr20 and EGFr22, and all six in EGFr21 (Fig. 4 abc). This disrupts all three disulfide bonds in EGFr21 and one each in EGFr20 and EGFr22. Although C1798 and C10812 appear unpaired in sequence, the deletion alters spacing and may permit novel disulfide bonding (Fig. 4c, inset). ChimeraX and Disulfide by Design 2[14] analysis showed the C1798–C10812 bond is geometrically feasible: bond distance 1.895 Å, χ3 angle –80.04°, and bond energy 2.70 kcal/mol—within tolerable limits, despite mild strain. The sigma B-factor (174.36) indicates reduced flexibility, possibly affecting redox sensitivity. Environmental factors like redox potential and pH may further modulate its stability.

Despite no free cysteines, the AF3 model of the deletion variant showed local destabilization in fused EGFr20–22 (Fig. 4c, cyan/orange/yellow) and higher aggregation propensity (A4D score –0.34 vs. WT –0.59). Such destabilization may underlie CADASIL pathology, first reported by the Hirose group in a patient pedigree[15]. Interestingly, full-length ECD modeling with AF3 did not predict the C798–C812 disulfide bond, likely due to low domain-to-domain spatial confidence in large models.

Fig. 4: Sequence and structural analysis of deletion isoforms.

a

```
                    EGFr20                          EGFr21                                  EGFr22
                          Cys                 1   ↓  2       3     4     5         6 7        8      9          10
Human-1             CEHGGRCESAPGQLPVCSCPQGWQGPRCQQDVDECAGPAPCGPHGICTNLAGSFSCTCHGGYTGPSCDQDINDCDPNPCLNGGSC 864
Human-X1            CEHGGRCESAPGQLPVCSCPQGWQD----------------------------------------------PCLNGGSC 812
Chimpanzee-1        CEHGGRCESAPGQLPVCSCPQGWQGPRCQQDVDECAGPAPCGPHGICTNLAGSFSCTCHGGYTGPSCDQDINDCDPNPCLNGGSC 864
Chimpanzee-X1       CEHGGRCESAPGQLPVCSCPQGWQGPRCQQDVDECAGPAPCGPHGICTNLAGSFSCTCHGGYTGPSCDQDINDCDPNPCLNGGSC 864
Chimpanzee-X2       CEHGGRCESAPGQLPVCSCPQGWQD----------------------------------------------PCLNGGSC 812
Pygmy-X1            CEHGGRCESAPGQLPVCSCPQGWQGPRCQQDVDECAGPAPCGPHGICTNLAGSFSCTCHGGYTGPSCDQDINDCDPNPCLNGGSC 864
Pygmy-X2            CEHGGRCESAPGQLPVCSCPQGWQD----------------------------------------------PCLNGGSC 812
Gorilla-X1          CEHGGRCESAPGQLPVCSCPQGWQGPRCQQDVDECAGPAPCGPHGICTNLAGSFSCTCHGGYTGPSCDQDINDCDPNPCLNGGSC 864
Gorilla-X2          CEHGGRCESAPGQLPVCSCPQGWQD----------------------------------------------PCLNGGSC 812
Monkey-X1           CEHGGHCESAPGQLPVCSCPQGWQGPRCQQDVDECAGPAPCGPHGICTNLAGSFSCTCHGGYTGPSCDQDINDCDPNPCLNGGSC 864
Monkey-X2           CEHGGHCESAPGQLPVCSCPQGWQGPRCQQDVDECAGPAPCGPHGICTNLAGSFSCTCHGGYTGPSCDQDINDCDPNPCLNGGSC 864
Marmoset-X1,        CEHGGRCESAPGQLAVCSCPQGWQGSRCQQDVDECASPSPCGPHGICTNLAGGFSCTCHGGYTGPSCDQNINDCDPNPCLNGGLC 862
Marmoset-X3,        CEHGGRCESAPGQLAVCSCPQGWQGSRCQQDVDECASPSPCGPHGICTNLAGGFSCTCHGGYTGPSCDQNINDCDPNPCLNGGLC 825
Mormoset-X4,        CEHGGRCESAPGQLAVCSCPQGWQD----------------------------------------------PCLNGGLC 810
CapuChin-X1         CEHGGRCESAPGQLAVCSCPQGWQGPRCQQDVDECASPSPCGPHGICTNMAGGFSCTCHGGYTGHSCDQNINDCDPNPCLNGGLC 864
CapuChin-X2         CEHGGRCESAPGQLAVCSCPQGWQGPRCQQDVDECASPSPCGPHGICTNMAGGFSCTCHGGYTGHSCDQNINDCDPNPCLNGGLC 864
Bat-1               CEHGGHCEPGPGHLAVCSCPPGWQGPRCQQDVDECAGPSPCGPHGTCTNLAGSFSCTCHGGYSGSSCDQDIDDCDPNPCLNGGSC 858
Bat-X1              CEHGGHCEPGPGHLAVCSCPPGWQGPRCQQDVDECAGPSPCGPHGTCTNLAGSFSCTCHGGYSGSSCDQDIDDCDPNPCLNGGSC 841
Bat-X3              CEHGGHCEPGPGHLAVCSCPPGWQGPRCQQDVDECAGPSPCGPHGTCTNLAGSFSCTCHGGYSGSSCDQDIDDCDPNPCLNGGSC 841
Bat-X4              CEHGGHCEPGPGHLAVCSCPPGWQGPRCQQDVDECAGPSPCGPHGTCTNLAGSFSCTCHGGYSGSSCDQDIDDCDPNPCLNGGSC 841
Vampire-X1          CEHGGHCESAPGHLAVCSCPPGWQGPRCQQDVDECAGPSPCGPHGTCTNLAGSFSCTCHGGYSGSSCDHDIDDCDPNPCLNGGSC 862
Vampire-X2          CEHGGHCESAPGHLAVCSCPPGWQD----------------------------------------------PCLNGGSC 810
Mouse-1             CEHGGHCESDPDRLTVCSCPPGWQGPRCQQDVDECAGASPCGPHGTCTNLPGNFRCICHRGYTGPFCDQDIDDCDPNPCLHGGSC 865
Mouse-X1            CEHGGHCESDPDRLTVCSCPPGWQGPRCQQDVDECAGASPCGPHGTCTNLPGNFRCICHRGYTGPFCDQDIDDCDPNPCLHGGSC 865
Mouse-X2            CEHGGHCESDPDRLTVCSCPPGWQGPRCQQDVDECAGASPCGPHGTCTNLPGNFRCICHRGYTGPFCDQDIDDCDPNPCLHGGSC 523
Hamster-X1          CEHGGHCESDPGQLTVCSCPPGWQGSRCQQDVDECAGALPCGPHGTCTNLPGSFRCLCHEGYTGPFCDQDIDDCDPNPCLNGGSC 864
Hamster-X2          CEHGGHCESDPGQLTVCSCPPGWQGSRCQQDVDECAGALPCGPHGTCTNLPGSFRCLCHEGYTGPFCDQDIDDCDPNPCLNGGSC 836
Hamster-X3          CEHGGHCESDPGQLTVCSCPPGWQGSRCQQDVDECAGALPCGPHGTCTNLPGSFRCLCHEGYTGPFCDQDIDDCDPNPCLNGGSC 864
Elephant.Ind-X1     CENGGHCESAPGQLAVCSCPPGWQGPRCQQDVDECAGPSPCGPHGTCTNLAGTFSCTCHGGYGGPSCDQDIDDCDPNPCLNGGLC 860
Elephant.Ind-X2     CENGGHCESAPGQLAVCSCPPGWQD----------------------------------------------PCLNGGLC 808
Elephant.Afr-X1     CENGGHCESAPGQLAVCSCPPGWQGPRCQQDVDECAGSSPCGPHGTCTNLAGTFSCTCHGGYGGPSCDQDIDDCDPNPCLNGGLC 860
Elephant.Afr-X2     CENGGHCESAPGQLAVCSCPPGWQD----------------------------------------------PCLNGGLC 796
Whale-X1,           CEHGGYCESAPGQLAVCSCPPGWQGSRCQQDVDECAGPSPCGPHGTCTNLAGSFSCTCHEGYSGPSCDQDIDDCDPNPCLNGGSC 863
Whale-X2,           CEHGGYCESAPGQLAVCSCPPGWQGSRCQQDVDECAGPSPCGPHGTCTNLAGSFSCTCHEGYSGPSCDQDIDDCDPNPCLNGGSC 863
Whale-X3,           ------------------------GSRCQQDVDECAGPSPCGPHGTCTNLAGSFSCTCHEGYSGPSCDQDIDDCDPNPCLNGGSC 825
Whale-X5,           --------------------------SRCQQDVDECAGPSPCGPHGTCTNLAGSFSCTCHEGYSGPSCDQDIDDCDPNPCLNGGSC 710
Whale-X6,           CEHGGYCESAPGQLAVCSCPPGWQGSRCQQDVDECAGPSPCGPHGTCTNLAGSFSCTCHEGYSGPSCDQDIDDCDPNPCLNGGSC 863
                    ** ** **          x x     ↑ **********   ****** ***  * * * ** ** *  **  * ******** ** *
```



a, MSA of isoform (X1) of human NOTCH3 compared to various mammals. Chimpanzee, *Pan troglodytes*; Pygmy chimpanzee, *Pan paniscus*; Gorilla, *Gorilla gorilla gorilla*; Monkey, *Rhinopithecus roxellana*; Marmoset, *Callithrix jacchus*; Capuchin, *Sapajus apella*; Bat, *Myotis myotis*; vampire bat, *Desmodus rotundus*; Mouse, *Mus musculus*; hamster, *Cricetulus griseus*; Indian elephant (Ind.), *Elephas maximus indicus*; African elephant (Afr.), *Loxodonta Africana*; Whale (blue), *Balaenoptera musculus*. Arrows/bold residue (D) shows the start of the deletion. *Conserved residues and x, where Cys residue is missing in some sequence. 8 Cys residues within the deleted sequence (Human) are shown in bold. There are 8 cysteines in the missing peptide (P805-N856). $C_3$814-$C_5$826, $C_4$820-$C_6$835, and $C_7$837-$C_8$846 S-S bonds are within EGFr21; In full-length NOTCH3, $C_1$798-$C_2$807 and $C_9$853-C_{10}864 S-S bonds are in EGFr20 and EGFr22 respectively. The $C_1$798 and C_{10}812 in X1 deletion mutant form a new disulfide bond and are shown in red bold. **b**, structures of human NOTCH3, EGFr20-22 (above); **c**, human NOTCH3 X1 deletion isoform (with EGFr23) is shown in plDTT colors. Dark blue, >90 (very high); light blue, 90>plDDT>70 (confident); yellow, 70>plDDT>50 (low). pLDDT scores served as a qualitative indicator of structural destabilization. All disulfide bonds are shown (green dashes). C798-C807 and C853-C864 are highlighted because C798 and C853 are within the deleted isoform. C798 forms a new disulfide with the C812 (C864 in NOTCH3 without deletion) due to the deletion (inset). The G804D mutation is also shown.

Jaguar NOTCH3 Cysteine Variants Disrupt EGFr Integrity

In humans, the three disulfide bonds in each EGFr domain are essential for stabilizing the NOTCH3 protein, and mutations in cysteine residues are linked to CADASIL disease[11]. The prevailing hypothesis posits that loss of a disulfide bond or an odd number of cysteines in ECD promotes high-molecular-weight S–S-linked aggregation[16]. However, this study suggests that such a view may be oversimplified and potentially misleading.

To assess natural variation in the 224 highly conserved NOTCH3 cysteines, we analyzed ECD sequences from 113 mammalian species using Clustal and JALVIEW (Supplementary Software 1). Unexpectedly, the jaguar sequence exhibited seven cysteine missense mutations and one deletion across EGFr13–15, eliminating all six cysteines in EGFr14. Additionally, the last cysteine of EGFr13 (C542) and the first of EGFr15 (C589) were mutated (Fig. 5a). This marks the first known case of eight consecutive missing cysteines in any mammal, in sharp contrast to their complete conservation in all other species (Supplementary Software 1).

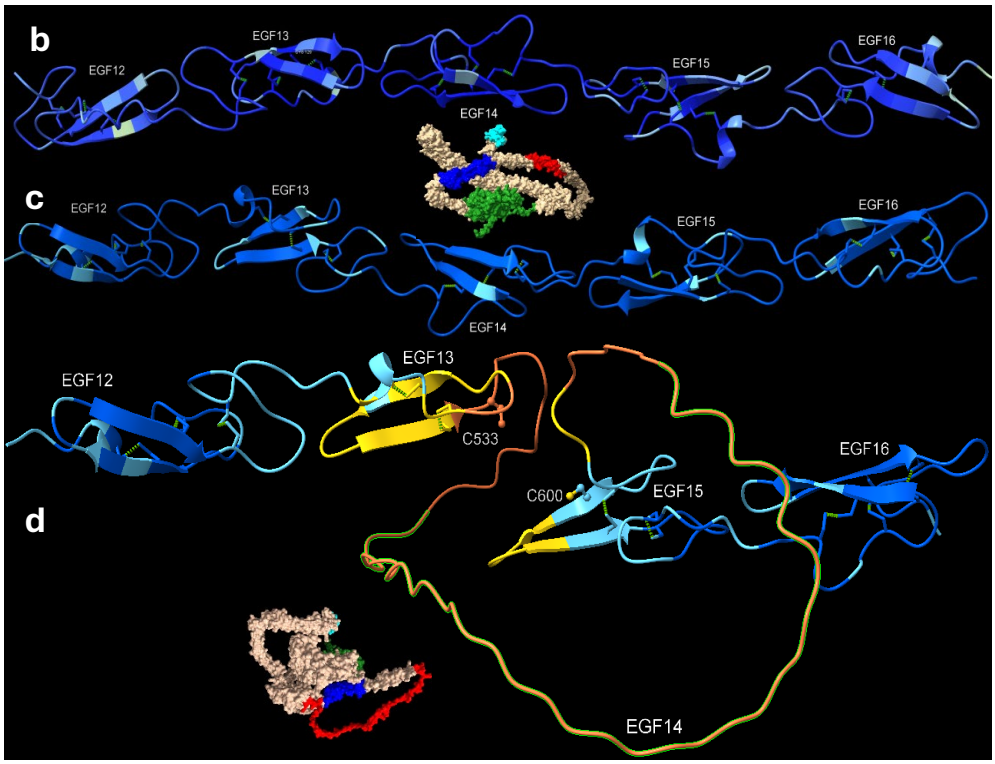
Using AF3, we modelled the structural impact of these mutations in the jaguar and compared them with structures from humans and lions (Figs. 5b–d). In humans and lions, EGFr12–16 maintained compact, folded conformations consistent with intact disulfide bonds. In contrast, the jaguar's EGFr14 appeared completely disordered, with low pLDDT scores reflecting the loss of all three disulfide bonds. EGFr13 and EGFr15 showed partial disorder due to unpaired C533 and C600, while EGFr12 and EGFr16 retained compact structures, consistent with preserved bonds (Fig. 5d).

When the WT lion EGFr12–16 region was mutated to carry all eight jaguar Cys>aa substitutions, it showed marked destabilization ($\Delta\Delta G$ = 34.85 kcal/mol), aligning with pLDDT-based disorder. However, A4D scores indicated either no change (WT vs. mutated lion: –0.84) or a slight decrease (WT jaguar: –0.68) in aggregation tendency. Restoring seven Cys residues and the original mutations in jaguar reduced $\Delta\Delta G$ to 16.34 kcal/mol and aggregation score to –0.90. Reverting all seven Cys>aa mutations and restoring the deleted cysteine further reduced destabilization to 4 kcal/mol. Yet, disulfide bridges were not reformed, as confirmed by DynaMut, which consistently indicated destabilization and increased flexibility across all aa>Cys reversions (Supplementary Table 1). These findings imply that compensatory mutations, insertions, and deletions in jaguar EGFr13–15 may block disulfide bond reformation—even with full cysteine restoration.

Fig. 5: Sequence and structural analysis of cysteine mutations in jaguar.

a

```
                Ca-EGFr13                                  Ca-EGFr14
Human    507 DVDECASTPCRNGAKCVDQPDGYECRCAEGFEGTLCDRNVDDCSPDPCHHGR--------- 558
Jaguar   507 DVDECASTPCRNGAKCVDQPDGYECRCAEGERGPLIGQQESGGGARERQKANNSLGENLGA 567
Lion     507 DVDECASTPCRNGAKCVDQPDGYECRCAEGFEGTLCERNVDDCSPDPCHHGR--------- 558
             ****************************** .* *  :: .. .    ::..
                                      Ca-EGFr15
Human    559 -CVDGIASFSCACAPGYTGTRCESQVDECRSQPCRHGGKCLDLVDKYLCRCPSGTTGVNCE 618
Jaguar   568 GLQQGRTSVEPM-------TEPGSGANDGPYQPCRHGGKCLDLVDKYLCRCPPGTTGVNCE 621
Lion     559 -CVDGIASFSCACAPGYTGTRCESQVDECRSQPCRHGGKCLDLVDKYLCRCPPGTTGVNCE 618
                :* :*..         *.  * .::   ******************** ********
```



a, MSA of jaguar EGFr13-15 with human and lion. **b**, Human; **c**, lion; **d**, jaguar AF3 structures. Seven Cys>X mutations and a deletion are shown in bold red in jaguar. In EGFr14 all 6 Cysteines are mutated, therefore, this EGFr is without any disulfide bond. One Cys (blue) residue each in EGFr13 (C533) and EGFr15 (C600) is also unpaired (their corresponding Cys partners mutated to I542 and G589 respectively, leaving both cysteines without any disulfide bridge. Bottom: AF3 structures of NOTCH3 EGFr12-16 colored according to plDDT as in the legend of Fig. 1. pLDDT scores served as a qualitative indicator of structural destabilization. Jaguar protein shows two unpaired Cys residues (C533 and C600) in EGFr13 and EGFr15. All 6 Cys residues are mutated in EGFr14 (green highlighted orange loop). Disulfide bridges are shown as green dots. EGFr12 and EGFr16

with all disulfides intact, are shown as fully-folded references. Disulfide by Design 2 tool (http://cptweb.cpt.wayne.edu/DbD2/) also shows that C533 and C600 are free. Inset figures: space-filled AF3 models of human (**b**) and Jaguar (**d**) ECD. Red, EGFr13-15; blue, ligand-binding EGFr10-11; green; first EGFr, cyan and tan, the rest of EGFr. Note the compact EGFr13-15 in humans with all disulfide bonds intact. The jaguar EGFr13-15 shows a disordered loop.

NRR Deletions in Brandt's Bat Enhance Protease Access

The NRR domain of human NOTCH3 prevents premature receptor activation. It includes three Lin12-Notch repeats (LNRs) and a heterodimerization (HD) domain, which maintain NOTCH3 in an autoinhibited state in the absence of ligand binding (Fig. 6a). The NRR shields the S2 cleavage site from ADAM proteases, ensuring NOTCH3 activation only occurs upon ligand (e.g., Jagged or Delta-like) engagement. Ligand binding induces mechanical or allosteric disruption of the NRR, exposing the S2 site for ADAM cleavage, followed by γ-secretase cleavage and release of the NOTCH3 ICD[17,18] (Fig. 6b).

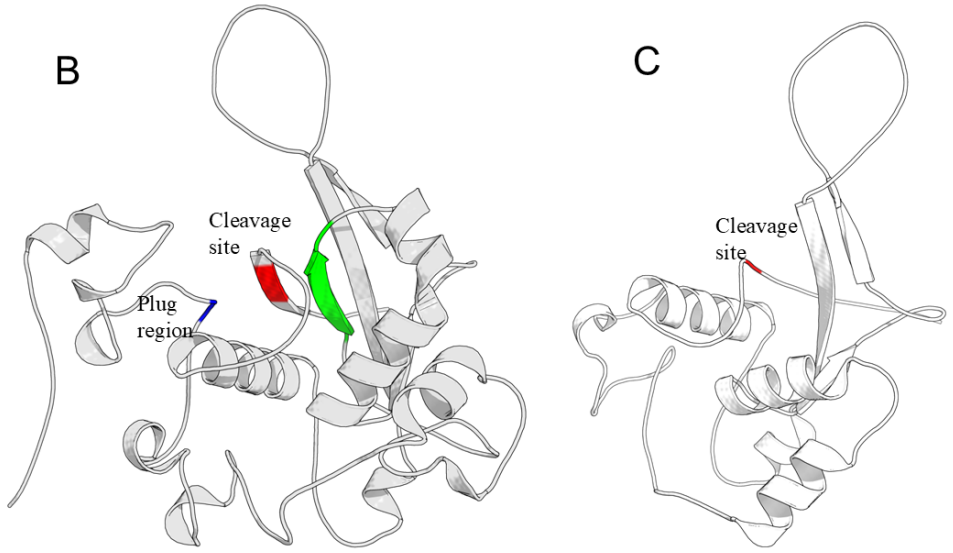
Multiple sequence alignment (MSA) of mammalian NOTCH3 (Supplementary Software 1) shows that the long-lived bat *Myotis brandtii* lacks two peptides in its NRR domain—within LNR-A and LNR-C (Fig. 6a). One missing segment in LNR-A (LNF/LSV), normally forming a loop plug that shields the S2 site, is absent in *M. brandtii*, potentially enabling easier access for ADAM10 metalloprotease (Fig. 6b, c). GETAREA analysis further supports this: human S2 residues D152 and V153 show SASA values of 21.9 Å² and 1.8 Å², while the corresponding bat residues A163 and A164 exhibit higher values of 90.3 Å² and 44.1 Å². All other examined mammals, including vampire bats and *Myotis daubentonii*, *M. myotis*, and *M. kuhlii*, retain these sequences.

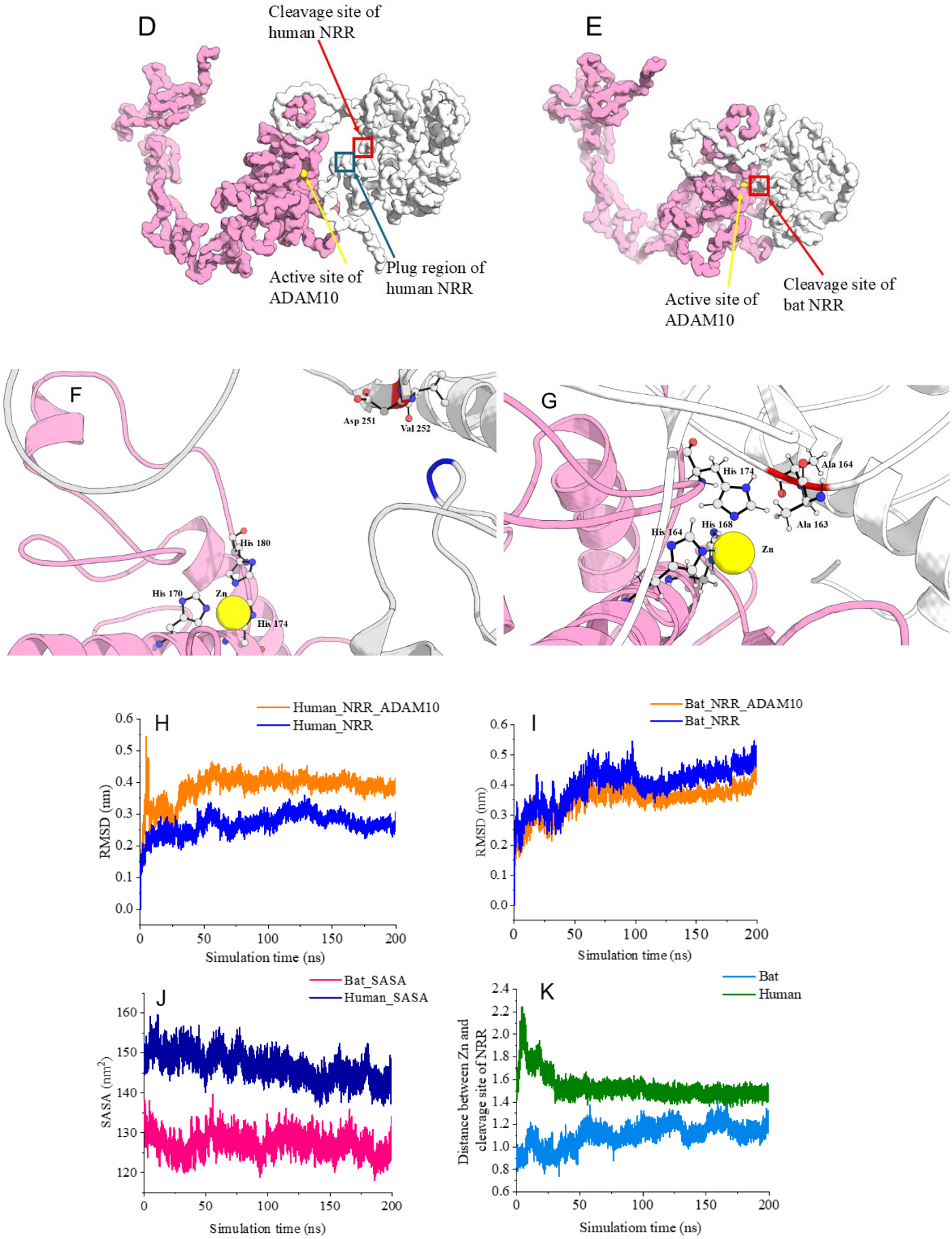
Simulations of the NRR–ADAM10 complex revealed that in the bat (Figs. 6e, g), the S2 site lies closer to the catalytic Zn ion than in the human complex (Figs. 6d, f), likely due to the LNR-A deletion. Both human and bat complexes remained stable over 200 ns (Figs. 6h, i). The human complex showed higher solvent-accessible surface area (SASA), suggesting a less compact structure and weaker interaction (Fig. 6j). The Zn–S2 cleavage site distance is consistently shorter in the bat complex (initial 8.4 Å; final 11.8 Å; average 11 Å; Figs. 6e, g, k) than in the human (initial 15.24 Å; final 15 Å; average 15.33 Å; Figs. 6d, f, k), supporting a conformation more favourable for proteolysis in *M. brandtii*.

Fig. 6. Sequence, structural, and MDS analysis of the NRR-ADAM 10 metalloprotease complexes from human and Brandt's bat.

A

```
Brandt bat      ----------------------------------------------PWRQCA-ALQCWRL    13
Human_NOTCH3    APEVSEEPRCPRAACQAKRGDQRCDRECNSPGCGWDGGDCSLSVGDPWRQCE-ALQCWRL    59
Human_NOTCH1    -----GEEACELPECQEDAGNKVCSLQCNNHACGWDGGDCSLNFNDPWKNCTQSLQCWKY    55
                                                          **::*  :****:
Brandt bat      FNNSRCDPACSSPACLYDNFDCRAGGRERTCNPVYEKYCADHFADGRCDQG---------    64
Human_NOTCH3    FNNSRCDPACSSPACLYDNFDCHAGGRERTCNPVYEKYCADHFADGRCDQGCNTEECGWD    119
Human_NOTCH1    FSDGHCDSQCNSAGCLFDGFDCQRA--EGQCNPLYDQYCKDHFSDGHCDQGCNSAECEWD    113
                *.:.:**  *.* .**:*.***: .  *  ***:*::** ***:**:****
Brandt bat      ---------------------------HFLQRLSAILRTSLRFRLDERGQAMVF    91
Human_NOTCH3    GLDCASEVPALLARGVLVLTVLLPPEELLRSSADFLQRLSAILRTSLRFRLDAHGQAMVF    179
Human_NOTCH1    GLDCAEHVPERLAAGTLVVVVLMPPEQLRNSSFHFLRELSRVLHTNVVFKRDAHGQQMIF    173
                                              .**:.** :*:*.: *: * :** *:*
Brandt bat      PYHRQDPGSEAPHLRELAPEVIGSVVMLEIDNRLCLQLPENDHCFPDAQSAADYLGALSA    151
Human_NOTCH3    PYHRPSPGSEPRARRELAPEVIGSVVMLEIDNRLCLQSPENDHCFPDAQSAADYLGALSA    239
Human_NOTCH1    PYYGM--------------DVRGSIVYLEIDNRQCVQA--SSQCFQSATDVAAFLGALAS    217
                **:                :* **:* ****** *:*   ..:** .* ..* :****::
Brandt bat      VERLDFPYPLWA↓ARGEPLEPLEPS-     175
Human_NOTCH3    VERLDFPYPLRD↓VRGEPLEPPEPS-     263
Human_NOTCH1    LGSLNIPYKIEA↓VQSETVEPPPPAQ     242
                :  *::** :  .:.* :**  *:
```




D
Cleavage site of human NRR
Active site of ADAM10
Plug region of human NRR
E
Active site of ADAM10
Cleavage site of bat NRR
F
Asp 251
Val 252
His 180
His 170
Zn
His 174
G
His 174
Ala 164
His 164
His 168
Ala 163
Zn
H
Human_NRR_ADAM10
Human_NRR
RMSD (nm)
Simulation time (ns)
I
Bat_NRR_ADAM10
Bat_NRR
RMSD (nm)
Simulation time (ns)
J
Bat_SASA
Human_SASA
SASA (nm^2)
Simulation time (ns)
K
Bat
Human
Distance between Zn and cleavage site of NRR
Simulatiom time (ns)

a, shows aligned sequences of human NOTCH1 and NOTCH3 alongside the NRR sequence of Brandt's bat. Blue, LNR-A showing boxed plug; green, LNR-B; purple, LNR-C; broken box, NRR-HD; solid box, α_3-helix; and red arrows, S2 cleavage site. **b**, human NRR; showing the S2 cleavage site (red) between a β-strand (green) and plug (blue); **c**, bat NRR structure. The α3-helix is shown below the S2 site. **d-k**, 200 ns simulations for human and bat NRR-ADAM10complexes. Grey, NRR; pink, ADAM10. **d**, overall view of Human NRR-ADAM10 complex. **e**, overall view of the bat NRR-ADAM10 complex. **f**, close-up of ADAM10-NRR complex in human. **g**, close-up of ADAM10-NRR complex in bat. Red box, S2 cleavage site in NRR. Blue box, plug region of human NRR. The yellow sphere, Zn ion indicates the location of the active site of ADAM10. All ADAM10 were open-forms. **h**, the RMSD of backbone atoms of the human NRR-ADAM10 complex and the human NRR. **i**, the RMSD of backbone atoms of the bat NRR-ADAM10 complex and bat NRR. **j**, the comparison of the solvent accessible surface area (SASA) of bat and human structures. **k**, comparison of distances between Zn ion at the active site centre of ADAM10 metalloprotease and the center of mass of the S2 cleavage site in bat (AA) and human (DV).

Discussion

Cross-species protein analyses are essential for uncovering the evolutionary conservation and pathogenic potential of vascular proteins such as NOTCH3. While aged domestic and wild animals, such as dogs and cats, can exhibit cognitive decline and microvascular changes resembling human small vessel disease, no naturally occurring animal model has been definitively identified for Vascular Cognitive Impairment and Dementia (VCID) or CADASIL. In contrast, several rodent models have been developed to replicate key features of these conditions, including inflammation, white matter damage, and cognitive impairment. Transgenic mice carrying NOTCH3 mutations further recapitulate hallmark CADASIL features such as vascular smooth muscle cell (VSMC) degeneration, white matter lesions, and GOM deposition[19]. Advances in MRI across larger mammals have also enabled cerebrovascular comparisons, laying the groundwork for translational research into how vascular defects contribute to neuronal injury and cognitive decline. These efforts may ultimately guide novel preventative or therapeutic strategies for VCI[20].

In this study, we analyzed NOTCH3 sequences from 113 mammalian species to assess conservation, structural variability, and pathological relevance. Recent large-scale efforts such as the Zoonomia Project have highlighted the power of mammalian comparative genomics for uncovering essential genes and disease mechanisms[8,10]. Our results fit within this broader framework: while NOTCH3 shows strong conservation across mammals (>90% identity), we also uncovered naturally tolerated variation, including eight cysteine mutations in jaguar (*Panthera onca*) and an EGFr XI deletion present in multiple species, including humans, without apparent ill effects. These findings emphasize that conservation alone does not dictate pathogenicity; context and compensatory mechanisms are also important.

We focused on three naturally occurring scenarios with direct implications for human CADASIL: two involving cysteine loss in EGFrs (via missense mutations or deletions), and one involving a novel deletion in the NRR. These findings form a clear basis for testable hypotheses. While sequence identity below 25% typically signals functional divergence, our MSA revealed strong conservation of NOTCH3 across mammals (>90% identity), underscoring its critical role—particularly the integrity of disulfide bonds that stabilize the 34 EGFr domains (Fig. 1). Disruption of these bonds through mutation or deletion, and the resulting unpaired cysteines in the ECD, is increasingly recognized as a key driver of CADASIL[4,5,7].

All NOTCH3 domains were well-structured except for a large portion of the intracellular domain (ICD), which appeared intrinsically disordered (Fig. 3). Intrinsically disordered proteins (IDPs) lack a stable tertiary structure under native conditions. IUPRED predicts disorder based on weak inter-residue interactions insufficient for stable folding, while ANCHOR2 identifies regions that become structured upon binding specific partners[21]. The disordered nature of the ICD allows it to adopt multiple conformations depending on binding partners, broadening its functional scope—as seen in other IDPs[22].

The human isoform X1, which likely restores the C1798–C10812 disulfide bond and eliminates free cysteines, presents a paradox: patients still show hallmark CADASIL symptoms such as GOM deposits, vascular smooth muscle degeneration, and cerebral white matter lesions[15]. This suggests that unpaired cysteines alone may not fully explain disease pathology. The recurrence of this deletion across multiple mammals implies a conserved evolutionary mechanism that may contribute to adaptive or pathological traits and warrants experimental investigation.

The discovery of cysteine mutations in the jaguar's NOTCH3 is intriguing, suggesting that not all cysteines are essential for mammalian health. Despite complete disulfide bond loss in EGFr14 and two unpaired cysteines in EGFr13 and EGFr15, the sequenced jaguar appeared healthy[23], though any late-onset neurodegeneration remains unknown. Structurally, this mutation cluster may affect ligand binding: EGFr13–15 forms an extended loop over the compact EGFr10–11 ligand-binding domains in jaguar, unlike the more sequential, compact configuration in humans and lions (Fig. 5), potentially altering ligand interactions.

Each disulfide bond contributes ~5–6 kcal/mol to protein stability[24]. The loss of three such bonds via Cys>aa substitutions in jaguar NOTCH3 led to a predicted destabilization of ~15–18 kcal/mol. Reverting all seven missense mutations in EGFr12–16 back to cysteines and restoring the deleted residue did not re-establish disulfide bonding, as destabilization persisted in A4D and DynaMut predictions (Supplementary Table 1). This implies that EGFr13–15, which harbors unique insertions, deletions, and mutations, is intrinsically suboptimal for disulfide formation. We speculate this potentially disordered region near the ligand-binding domains (EGFr10–11) may fine-tune NOTCH3 signaling in ways adapted to the jaguar’s muscular, semi-aquatic, and predatory lifestyle. Further investigation is needed

to determine if these mutations promote muscle development, ecological adaptation, or offer insights into CADASIL-relevant structural variation.

Positive selection in jaguar genes for craniofacial and limb development has been reported[23]. Since NOTCH3 loss enhances skeletal muscle in mice[25], similar effects might underlie the jaguar's exceptional musculature. Jaguars possess the highest bite force quotient among felids—capable of cracking turtle shells and caiman skulls[26]—potentially driven by adaptive NOTCH3 alterations, including loss of five disulfide bonds in EGFr13–15 and the labile DP bond in EGFr14, which exists in humans and lions. These DP bonds may promote ECD aggregation via non-enzymatic autolysis at Asp-Pro sites, a CADASIL hallmark[27]. The combined loss of eight cysteines and a DP bond is unlikely due to drift alone, though lineage-specific drift, compensatory mutations, or other evolutionary forces cannot be excluded.

Most CADASIL-associated mutations occur within the 34 EGFrs or the NRR, leading to either ECD aggregation or ligand-independent receptor activation. Although ECD aggregation is a hallmark of CADASIL, NOTCH3 signaling often remains intact, and the NRR still undergoes normal cleavage[28]. Mutations—particularly in the HD or LNR domains can destabilize the NRR, exposing the S2 cleavage site even without ligand binding.

Interestingly, a unique deletion in Brandt's bat removes a protective NRR peptide, including the LSV plug from LNR-A (Fig. 6a–c), suggesting species-specific modulation of ligand-free activation. While loss of LNR-A might appear to fully expose the S2 site, MD simulations suggest this alone is insufficient for constitutive activation. Rather, the deletion likely increases S2 accessibility and cleavability by ADAM10 compared to the human NRR.

Normally, ligand binding to the ECD initiates endocytosis, pulling LNR-A away[17]. In the bat, the absence of LNR-A implies ligand binding may no longer be necessary. With no LNR-A to displace, activation could proceed more readily[17]. Still, this deletion may not suffice, as Helix3 continues to shield the S2 site, potentially requiring subtle displacement[18]. In bat NRR, Ala143 of α3 forms hydrophobic contacts with Ala164, Arg165, and Gly166 including S2 residue Ala164 likely obstructing ADAM10 access. The shielding function of α3 in preventing cleavage has been previously reported[17,18]. Catalytically active Zn^{2+}scissile bond distances typically span ~2.1–2.4 Å (QM/MM studies of MMP-3)[29]. In our simulations, Zn–S2 distances remained substantially larger (8.5–11.8 Å), suggesting further conformational changes are needed to bring S2 into alignment with the ADAM10 active site.

A related mechanism involves a conserved calcium-binding His60 in NOTCH1, replaced by arginine in NOTCH3 (Fig. 6a). In NOTCH1, mutating this residue to proline disrupts calcium binding, causing a 20-fold increase in ligand-independent activity mimicking LNR deletion[18]. This supports the idea that structural changes in *M. brandtii* NRR may alter its activation profile, leading to excess NICD production and dysregulated transcription.

Such unregulated signaling disrupts VSMC homeostasis, leading to dysfunction, degeneration, blood–brain barrier leakage, and microangiopathy contributing to cerebral small vessel disease, stroke, vascular cognitive impairment, and dementia[7,11]. Alternatively,

the NRR deletion in the bat may support regulated signaling via distinct ligand interactions and subtle conformational changes that displace Helix3. While ligand-independent Notch activation has been associated with cancers, specific oncogenic roles for NOTCH3 remain less studied than for NOTCH1 or NOTCH2[2].

Bats are renowned for exceptional longevity relative to size[30], attributed to factors like hibernation, efficient DNA repair, mitochondrial function, low IGF-1 signaling, and enhanced autophagy[31,32]. However, the role of altered intercellular signaling—such as via NOTCH3—remains largely unexplored.

Their cerebrovascular anatomy is also unique: a well-developed vertebrobasilar network and robust pial and parenchymal arteries support high cerebral energy demands, while the internal carotid system is comparatively underdeveloped[33,34,35]. This resilience raises the question of why only Brandt's bat harbors a unique NOTCH3 NRR deletion—and whether it enables ligand-independent activation or reflects alternative regulation.

Recent studies highlight that age-related increases in NOTCH3 expression contribute to vascular aging[32], while elevated signaling promotes narrowing of cerebral arteries and reduced perfusion[11], potentially contributing to cognitive decline. These findings underscore the importance of tightly regulated NOTCH3 activity. The bat-specific NRR deletion may represent an evolutionary adaptation that enables stable, ligand-independent baseline signaling—potentially preserving vascular integrity and supporting bat longevity.

The relationship between NOTCH3 cysteine mutations and pathogenicity remains unclear, highlighting the need for species-specific responses to be validated in model organisms. In jaguars, the absence of disulfide bonds and the presence of unpaired cysteines do not cause disease, suggesting compensatory mechanisms. Disruption of EGFr14 eliminates all disulfide bonds, likely rendering it unstructured (Fig. 5). Introducing these mutations into models could clarify how disulfide loss affects NOTCH3 stability, aggregation, pathogenicity, and whether unstructured loops function as ligand-binding sites.

In the human X1 isoform, exon 16 skipping removes EGFr21 and parts of EGFr20 and EGFr22, forming a new disulfide bond between unpaired cysteines (Fig. 4). Despite this rearrangement, pathogenicity persists, implying that specific amino acid substitutions—not cysteine loss alone may drive CADASIL. Although this deletion occurs in other mammals, its role outside humans remains unknown.

Mammalian NOTCH3 sequence–structure–function relationships can reveal hidden clues to the pathogenesis of human vascular disorders, including CADASIL, highlighting how evolutionary variation can inform disease mechanisms. The three mammalian mutations identified here are particularly salient against a background of extreme conservation across 113 mammals. They may not only account for some unique physiological attributes of the three species in which they occur, but also inform the relationship between structure and pathology across species. To deepen understanding of NOTCH3 biology, various cysteine mutations, disulfide rearrangements, and NRR deletions identified in this study should be systematically modeled in cell and animal systems.

Methods

More than 100 mammalian NOTCH3 protein sequences were retrieved from the NCBI database. MSA were carried out using Clustal-Omega available at www.ebi.ac.uk/jdispatcher/msa/clustalo[36]. The mammalian NOTCH3 MSA from Clustal-Omega was exported to Jalview 2.11.3.3 available at www.jalview.org for editing and analysis[37]. The Jalview interactive alignment file is available online (see Supplementary Software1). Percentage conservation data for all 2,321 amino acids between human and mammalian NOTCH3 was imported into Origin v2020b (OriginLab Corporation, Northampton, MA, USA). All plots feature interactive tooltips for direct data visualization. The Origin interactive file is available online (see Supplementary Software2). Prediction of intrinsically disordered and binding regions in NOTCH3 was carried out using ANCHOR2 and AIUPRED tools available at https://aiupred.elte.hu[21,38]. The structures of various extracellular and cytoplasmic domains of NOTCH3 were generated using AlphaFold3, available at https://alphafoldserver.com/about[39,40]. Protein structures were analyzed and visualized using ChimeraX available at https://www.rbvi.ucsf.edu/chimerax[41]. The pathogenic mutations of NOTCH3 were retrieved from the respective publications. NOTCH3 structures, including the native form and those harboring mutations and deletions, were generated for EGFr domains using AF3 in segments of shorter repeats, as well as for the NRR region. These models were evaluated for stability and flexibility resulting from vibrational entropy changes upon missense mutations using DynaMut suite (https://biosig.lab.uq.edu.au/dynamut/) described[42]. The propensity to aggregate and stability for the WT vs deletion and missense mutant EGFr domains was analyzed using Aggrescan4D (https://biocomp.chem.uw.edu.pl/a4d/#) that includes dynamic mode into the analysis[43]. The formation of new disulfide bonds was analysed using ChimeraX and Disulfide by Design2[14] tools. The solvent accessible surface area (SASA) of residues in protein structures was calculated using GETAREA available at https://curie.utmb.edu/getarea.html[44].

Molecular Dynamics Simulations (MDS)

The crystal structure of the human NRR domain was obtained from PDB entry 4ZLP, with missing residues supplemented using AlphaFold3 (AF3) predictions. The open conformation of the human ADAM10 metalloprotease was derived from PDB entry 8ESV,

from which both the heavy and light Fab chains, as well as tetraspanin-15, were removed[45]. The NRR sequence of Brandt's bat was obtained from the genome project by[31], and its structure was predicted using AF3. The ADAM10 metalloprotease sequence of Brandt's bat (XP_005883346.1) was retrieved from the NCBI Database, and its structure was modelled based on the human open-form ADAM10 structure (8ESV).

The docking conformations between the NRR and ADAM10 metalloprotease of both human and Brandt's bat were predicted using AlphaFold3 multimer modeling[40]. Input files for molecular dynamics simulations were generated using the CHARMM-GUI platform[46]. Protein structures were prepared with the Amber ff14SB force field[47], and the protonation state was set to pH 7.0 using CHARMM-GUI. Water molecules were modeled using the TIP3P model[48]. All systems were solvated in a cubic box of TIP3P water with dimensions of 300 Å per side.

All simulations and trajectory analyses were performed using the GROMACS 2024.3-gpuvolta software package[49]. Each system was first subjected to 1000 steps of steepest-descent energy minimization. Equilibration was conducted in two phases. Initially, constant-volume (NVT) equilibration was carried out at 303.15 K using the V-rescale thermostat[50] for 4 nanoseconds, with a time step of 1 femtosecond (fs).

This was followed by constant-pressure (NPT) equilibration, with pressure maintained at 1 bar using the Parrinello–Rahman barostat[51] and semi-isotropic pressure coupling. NPT equilibration was conducted in four stages, during which position restraints were applied to both backbone and side chains, beginning with force constants of 400 and 40 kJ/mol/nm², respectively (as used in the initial NVT phase), and gradually reduced by 100 and 10 kJ/mol/nm² at each stage.

The Verlet cutoff scheme[52] was employed, with a cutoff radius of 1.2 nm for short-range van der Waals interactions. Long-range electrostatics were treated using the Particle Mesh Ewald (PME) method[53]. All bonds involving hydrogen atoms were constrained using the LINCS algorithm[54].

After equilibration, three independent 200 ns production runs were performed using a 2 fs time step. The SHAKE algorithm was employed to constrain all bonds involving hydrogen atoms[55].

Following the simulations, trajectories were corrected for periodic boundary conditions using the gmx trjconv tool with the flags -pbc mol -center. Trajectory analyses included the

calculation of backbone root mean square deviation (RMSD) using the gmx rms tool. The solvent-accessible surface area (SASA) was computed using gmx sasa, and the distance between the Zn atom of ADAM10 and the cleavage site of the NRR was measured using gmx dist.

Acknowledgements

This study was supported by grants from the NHMRC of Australia (1196150, 2006765) and the Sachdev Foundation. We acknowledge CHATGPT for editing English, word management, and formatting references.

Author contributions

K.S.S., H.E., P.S. and A.P. conceived the original idea. K.S.S., H.E., and A.P. designed the study. K.S.S. and H.E. performed the study. YR performed MDS. K.S.S., H.E., A.P., T.J. YR and P.S. discussed the data and wrote or edited the paper. All authors approved the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Data availability

Supplementary software data files are available for this paper at Figshare.com

Supplementary Software 1. Multiple Sequence Alignment (MSA) using Custal-Omega and presented using Jalview Software. http://doi.org/10.6084/m9.figshare.30223741

Supplementary Software 2. Interactive Origin plots show the NOTCH3 sequence with various domains, pathogenic mutations and % consensus with 113 mammalian NOTCH3 sequences. http://doi.org/10.6084/m9.figshare.30223747

Supplementary Material. Multiple sequence alignment, AF3 structures, and Origin plots of human NOTCH3 with 113 mammals using CLUSTAL-Omega. http://doi.org/10.6084/m9.figshare.30223750

Correspondence and requests for materials should be addressed to PS and KSS.

References

1. Darwin, C. *The Descent of Man, and selection in relation to sex* (London: John Murray, 2nd ed.) p. 6 (1874).
2. Hosseini-Alghaderi, S. & Baron, M. Notch3 in development, health and disease. *Biomolecules* **10**, 485 (2020). https://doi.org/10.3390/biom10030485
3. Boston, G., Jobson, D., Mizuno, T., Ihara, M., & Kalaria, R. N. "Most common NOTCH3 mutations causing CADASIL or CADASIL-like cerebral small vessel disease: A systematic review." *Cereb. Circ. - Cogn. Behav*. **6**, 100227 (2024). doi:10.1016/j.cccb.2024.100227
4. Young, K. Z. et al. Oligomerization, trans-reduction, and instability of mutant NOTCH3 in inherited vascular dementia. *Commun. Biol*. **5**,1 331 (2022). doi:10.1038/s42003-022-03259-2
5. Lee, S. J. et al. Structural changes in NOTCH3 induced by CADASIL mutations: role of cysteine and non-cysteine alterations. *J. Biol. Chem.* **299**, 104838 (2023). https://doi.org/10.1016/j.jbc.2023.104838
6. Hess, K. L. *et al.* Protein aggregates containing wild-type and mutant NOTCH3 are major drivers of arterial pathology in CADASIL. *J. Clin. Invest.* **134**, e175789 (2024). https://doi.org/10.1172/JCI175789
7. Mizuta, I. *et al.* Progress to clarify how NOTCH3 mutations lead to CADASIL, a hereditary cerebral small vessel disease. *Biomolecules* **14**, 127 (2024). https://doi.org/10.3390/biom14010127
8. Christmas, M.J. *et al*. Zoonomia: Evolutionary constraint and innovation across hundreds of placental mammals. *Science.* 380:366 (2023). doi:10.1126/science.abn3943
9. Romero, I.J. Seeing humans through an evolutionary lens: a collection of mammalian genomes provides insights into human biology and evolution. *Science.* 380:360–361 (2023). doi:10.1126/science.adh0745
10. Sullivan, P.F. *et al*. Zoonomia: Leveraging base-pair mammalian constraint to understand genetic variation and human disease. *Science.* 380:367 (2023). doi:10.1126/science.abn2937
11. Baron-Menguy, C., Domenga-Denier, V., Ghezali, L., Faraci, F. M. & Joutel, A. Increased Notch3 activity mediates pathological changes in structure of cerebral arteries. *Hypertension* **69**, 60–70 (2017). https://doi.org/10.1161/HYPERTENSIONAHA.116.08015
12. Oka, Kaori et al. The naked mole-rat as a model for healthy aging. *Annu. Rev. Anim. Biosci*. ***11***, 207-226 (2023).
13. Morris, O.M., Torpey, J.H. & Isaacson, R.L. Intrinsically disordered proteins: modes of binding with emphasis on disordered domains. *Open Biol*. **11**, 210222 (2021). https://doi: 10.1098/rsob.210222
14. Craig, D. B. & Dombkowski, A. A. Disulfide by Design 2.0: a web-based tool for disulfide engineering in proteins. *BMC Bioinformatics* **14**, 346 (2013). https://doi.org/10.1186/1471-2105-14-346

15. Saiki, S. et al. Varicose veins associated with CADASIL result from a novel mutation in the Notch3 gene. *Neurology* **67**, 337–339 (2006). https://doi.org/10.1212/01.wnl.0000224758.52970.19
16. Chabriat, H. *et al.* CADASIL. *Lancet Neurol.* **8**, 643–653 (2009).
17. Gordon, W. R., Arnett, K. L. & Blacklow, S. C. The molecular logic of Notch signaling—a structural and biochemical perspective. *J. Cell Sci.* **121**, 3109–3119 (2008). https://doi.org/10.1242/jcs.035683
18. Gordon, W. R. et al. Structure of the Notch1-negative regulatory region: implications for normal activation and pathogenic signaling in T-ALL. *Blood* **113**, 4381–4390 (2009). https://doi.org/10.1182/blood-2008-08-174748
19. Gong, Z. et al. Analysis of the pathogenicity and pathological characteristics of NOTCH3 gene-sparing cysteine mutations in in vitro and in vivo models. *Front. Mol. Neurosci.* **17**, 1391040 (2024). https://doi.org/10.3389/fnmol.2024.1391040
20. Hainsworth, A. H. et al. Translational models for vascular cognitive impairment: a review including larger species. *BMC Med.* **15**, 16 (2017). https://doi.org/10.1186/s12916-017-0793-9
21. Erdős, G. & Dosztányi, Z. AIUPred: combining energy estimation with deep learning for the enhanced prediction of protein disorder. *Nucleic Acids Res.* **52**, W176–W181 (2024). https://doi.org/10.1093/nar/gkae385
22. Wright, P. & Dyson, H. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* **16**:18–29 (2015). doi: 10.1038/nrm3920.
23. Figueiró, H. V. et al. Genome-wide signatures of complex introgression and adaptive evolution in the big cats. *Sci. Adv.* **3**, e1700299 (2017). https://doi.org/10.1126/sciadv.1700299
24. Zavodszky, M. et al. Disulfide bond effects on protein stability: designed variants of *Cucurbita maxima* trypsin inhibitor-V. *Protein Sci.* **10**, 149-60 (2001). doi: 10.1110/ps.26801.
25. Kitamoto, T. & Hanaoka, K. Notch3 null mutation in mice causes muscle hyperplasia by repetitive muscle regeneration. *Stem Cells* **28**, 2205–2216 (2010).
26. Hoogesteijn, R. Wild cats 101: Why jaguars hunt caimans? *Panthera Blog* (2023). https://panthera.org/blog-post/wild-cats-101-why-jaguars-hunt-caimans
27. Lee, S. J. et al. A midposition NOTCH3 truncation in inherited cerebral small vessel disease may affect the protein interactome. *J. Biol. Chem.* **299**, 102772 (2023). https://doi.org/10.1016/j.jbc.2022.102772
28. Wang, T., Baron, M. & Trump, D. An overview of Notch3 function in vascular smooth muscle cells. *Prog. Biophys. Mol. Biol.* **96**, 499–509 (2008). doi: 10.1016/j.pbiomolbio.2007.07.006
29. Feliciano, G. T. & da Silva, A. J. R. Unravelling the reaction mechanism of matrix metalloproteinase 3 using QM/MM calculations. *J. Mol. Struct.* **1091**, 125–132 (2015). https://doi.org/10.1016/j.molstruc.2015.02.079
30. Brunet-Rossinni, A. K. & Austad, S. N. Ageing studies on bats: a review. *Biogerontology* **5**, 211–222 (2004). https://doi.org/10.1023/B:BGEN.0000038022.65024.d8
31. Seim, I. et al. Genome analysis reveals insights into physiology and longevity of the Brandt's bat *Myotis brandtii*. *Nat. Commun.* **4**, 2212 (2013). doi: 10.1038/ncomms3212.

32. Ding, Yingjie et al. “Comprehensive human proteome profiles across a 50-year lifespan reveal aging trajectories and signatures.” *Cell*, S0092-8674(25)00749-4. 22 Jul. 2025, doi:10.1016/j.cell.2025.06.047
33. Gorbunova, V., Seluanov, A. & Kennedy, B. K. The world goes bats: living longer and tolerating viruses. *Cell Metab.* **32**, 31-43 (2020). https://doi.org/10.1016/j.cmet.2020.06.013
34. Hasegawa, K., Kawano, T., Oishi, M., Sunagawa, G. & Nishimura, T. The angio-architecture in the brain of bat. *Fukuoka Acta Medica* **51**, 1251–1266 (1960).
35. Andō, K. A histochemical study on the innervation of the cerebral blood vessels in bats. *Cell Tissue Res.* **217**, 55–64 (1981). https://doi.org/10.1007/BF00233826
36. Madeira, F. et al. The EMBL-EBI Job Dispatcher sequence analysis tools framework in 2024. *Nucleic Acids Res.* **52**, W521–W525 (2024). https://doi.org/10.1093/nar/gkae241
37. Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M. & Barton, G.J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics.* **25**, 1189-1191 (2009). doi: 10.1093/bioinformatics/btp033.
38. Mészáros, B., Erdős, G. & Dosztányi, Z. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.* **46**, W329-W337 (2018). https://doi.org/10.1093/nar/gky384
39. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021). https://doi.org/10.1038/s41586-020-03075-y
40. Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024). https://doi.org/10.1038/s41586-024-07487-w
41. Meng, E. C. *et al.* UCSF ChimeraX: tools for structure building and analysis. *Protein Sci.* **32**, e4792 (2023). https://doi.org/10.1002/pro.4792
42. Rodrigues, C. H. M., Pires, D. E. V. & Ascher, D. B. DynaMut2: Assessing changes in stability and flexibility upon single and multiple point missense mutations. *Protein Science* **30**, 60–69 (2021). https://doi.org/10.1002/pro.3942
43. Zalewski, M., Iglesias, V., Bárcenas, O., Ventura, S. & Kmiecik, S. Aggrescan4D: A comprehensive tool for pH-dependent analysis and engineering of protein aggregation propensity. *Protein Sci.* **33**, e5180 (2024). doi: 10.1002/pro.5180.
44. Fraczkiewicz, R. & Braun, W. Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J. Comp. Chem.* **19**, 319-333 (1998).
45. Lipper, C. H., Egan, E. D., Gabriel, K. H., & Blacklow, S. C. Structural basis for membrane-proximal proteolysis of substrates by ADAM10. *Cell* **186**, 3632–3641.e10 (2023). doi:10.1016/j.cell.2023.06.026
46. Jo, S., Kim, T., Iyer, V. G., & Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *J. Comput. Chem.* **29**, 1859–1865 (2008).
47. Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., & Simmerling, C. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput*. **11**, 3696–3713 (2015).
48. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R., & Klein, M. Refined TIP3P model for water. *J. Chem. Phys.* **79**, 926–935 (1983).

49. Abraham, M. J., et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX **1**, 19–25 (2015).
50. Bussi, G., Donadio, D., & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).
51. Ke, Q., Gong, X., Liao, S., Duan, C., & Li, L. Effects of thermostats/barostats on physical properties of liquids by molecular dynamics simulations. *J. Mol. Liq.* **365**, 120070 (2022).
52. Grubmüller, H., Heller, H., Windemuth, A., & Schulten, K. Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions. *Mol. Simul.* **6**, 121–142 (1991).
53. Petersen, H. G. Accuracy and efficiency of the particle mesh Ewald method. *J. Chem. Phys.* **103**, 3668–3679 (1995).
54. Hess, B., Bekker, H., Berendsen, H. J., & Fraaije, J. G. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472 (1997).
55. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys*. **23**, 327–341 (1977).