Deep mutagenesis reveals the distinct mutational landscape of ADan and ABri amyloid nucleation

Mariano Martín[1] and Benedetta Bolognesi[1]

1Institute for Bioengineering of Catalonia (IBEC), The Barcelona Institute of Science and Technology, Baldiri Reixac 10-12, 08028 Barcelona, Spain.

Abstract

Different forms of dementia are caused by stop-loss mutations in the ITM2B gene, also known as Bri2, which result in the expression of 34 amino acid long peptides that accumulate as amyloids in human brains. In order to gather mechanistic insights into the formation of amyloids by two of these peptides, ADan and ABri - hallmarks of Danish and British dementia respectively - we employed saturation mutagenesis combined to a massively parallel selection assay that reports on amyloid nucleation. Our results reveal that ADan aggregates into amyloids remarkably faster than both the unextended peptide Bri2 and the extended ABri sequence. The complete mutational landscape of ADan reveals asparagines and charged residues as key players in the nucleation process in addition to aliphatic residues within positions 20-25. What is more, we show that extending Bri2 with just two specific residues is enough to generate a novel amyloid core which we suggest builds the structured core of ADan fibrils. On the other hand, only a handful of mutations can boost the ability of ABri to nucleate amyloids, including a SNV replacing the Bri2 stop codon by a Cys codon. Overall, the remarkably different aggregation profiles and mutational landscapes for the two peptides suggest that different disease mechanisms underlie disease in Danish and British dementia and highlight the importance of accurately measuring the impact of stop extension mutations for these and other sequences across the genome.

Intro

Amyloid deposits are the pathological signature of a wide range of human diseases mainly affecting the nervous system. The proteins that deposit in pathology are encoded by genes that are the ones mutated in familial forms of neurodegenerative diseases. Such is the case of mutations in the gene APP, the precursor of amyloid beta peptide (Aβ42), that cause familial forms of Alzheimer's disease.
Frameshift mutations and stop-loss mutations, also known as nonstop mutations, have been pin-pointed as causative variants in several pathologies (Mignogna *et al*, 2022; Bauer *et al*, 2023; Vaché *et al*, 2023; Rebelo *et al*, 2016). Nonstop mutations constitute ~0.2% of codon-changing mutations (Hamby *et al*, 2011) and often lead to mRNA degradation by the non-stop mediated-decay machinery (Monaghan *et al*, 2023). However, if an alternative in frame stop codon appears shortly afterward (within 50 nt), the 3'UTR can be translated resulting in an aberrant protein with an extended C-term

(Hamby *et al*, 2011). This is the case of a set of NEFH gene mutations that cause Charcott-Marie-Tooth disease, the most common inherited motor neuron disorder. Stop-loss mutations in the NEFH gene allow translation of cryptic amyloidogenic elements (CAEs) contained in its 3' UTR, leading to protein aggregation and neurodegeneration (Rebelo *et al*, 2016). In addition, nonstop mutations that lead to the translation of 3'UTR CAEs have been identified in the REEP1 gene in patients with peripheral neuropathy (Bock *et al*, 2018).

Preventively knowing the phenotypic impact of any genetic mutation has become a major goal of precision medicine in the era of genomics and machine learning. Many Variant Effect Predictors (VEPs) have been developed in the last years to predict the impact of missense variants (Frazer *et al*, 2021; Adzhubei *et al*, 2010; Ng & Henikoff, 2001; Brandes *et al*, 2023). However, most VEPs rely on multiple sequence alignments and cannot capture the impact of stop-loss mutations. On the other hand, amyloid aggregation predictors have also limited predictive power on these sequences. These algorithms are largely based on hydrophobicity or beta sheet propensity - features that correlate only moderately with amyloid nucleation (Seuma *et al*, 2021) - and they have been exclusively trained on single amino acid substitutions. As a result, with the current tools, it is challenging to predict the appearance of CAEs due to stop-loss and therefore to identify potential molecular drivers of disease.

Stop-loss mutations in the gene ITM2B, also known as *bri2*, have been reported in patients with rare autosomal dominant forms of dementia. Familial British Dementia (FBD), Familial Danish Dementia (FDD), as well as the recently reported Familial Chinese Dementia (FCD) and Familial Korean Dementia (FKD), are caused by mutations in this gene (Vidal *et al*, 1999, 2000; Liu *et al*, 2021; Rhyu *et al*, 2023). In FDD, a 10 nt duplication produces a C-term frameshift and concomitant stop codon loss, leading to the formation of an aberrant 34 amino acid long peptide known as ADan. In FBD, a single nucleotide variant (SNV) affects the stop codon, resulting in a stop-loss with the consequent formation of an aberrant 34 amino acid long peptide known as ABri. FCD and FKD are also caused by stop-loss mutations that produce peptides that differ from ABri only in one amino acid. Amyloid deposits formed by these aberrant peptides have been detected in patients (Holton *et al*, 2001; Ghiso *et al*, 2001; Lashley *et al*, 2006; Vidal *et al*, 2000). Both loss of function and amyloid toxicity have so far been proposed as molecular mechanisms underlying the disease, contributing to disease onset and progression (Coomaraswamy *et al*, 2010; Marcora *et al*, 2014; Tamayev *et al*, 2010a; Cantlon *et al*, 2015).

Several *in vivo* and *in vitro* approaches have been developed with the aim of understanding the molecular mechanism of aggregation that underlie these diseases. Transgenic mice expressing ADan peptide present amyloid depositions (Vidal *et al*, 2009) in contrast to mice expressing ABri peptide where no amyloids are detected (Pickford *et al*, 2006). None of the two mouse models developed for FDD recapitulates fully patient's phenotype (Vidal *et al*, 2009; Tamayev *et al*, 2010b).

In vitro studies have revealed that ABri and ADan, but not Bri2, can form toxic oligomeric species and mature fibrillar aggregates (El-Agnaf *et al*, 2001a; Gibson *et al*, 2003; El-Agnaf *et al*, 2001b; Todd *et al*, 2016; Schlenzig *et al*, 2009; Cantlon *et al*, 2015). Both

peptides exhibit an oxidized and a reduced form *in vitro*. While ADan can form toxic oligomers regardless of oxidation state, only oxidized ABri aggregates in toxic soluble oligomers (El-Agnaf *et al*, 2001a; Gibson *et al*, 2003). In solution both peptides lack a defined structure but after a few hours they evolve into beta-sheet rich structures (Todd *et al*, 2016, 2014).

While Alzheimer's disease and Bri2-related familial dementias share common clinical and neuropathological features, including amyloid depositions, the primary sequences of the amyloid-forming peptides (Aβ42 or ADan/ABri) are substantially different, making it challenging to transfer mechanistic insights acquired for Aβ42 to ADan and ABri. What is more, in contrast with Aβ42, no high-resolution structure of ADan and ABri fibrils, *in vitro* or *ex-vivo*, has been solved.

Here, we employ deep mutational scanning (DMS) to systematically study the formation of amyloids by ADan and ABri and measure the effect of all possible mutations on the respective peptide amyloid nucleation. We also quantify the effect of new stop-loss mutations, generating datasets that can be used to guide clinical interpretation of novel mutations identified in the human population. Moreover, we report that a two amino acid extension of Bri2 C-term is enough for producing an amyloid core that makes the peptide prone to nucleate aggregates. Finally, the resulting data provide numerous mechanistic insights into the distinct process underlying ADan and ABri fibril nucleation.

Results

Bri2 extensions differ in their nucleation propensity

We measured the effect of Bri2 extensions (Fig. 1a) on the ability of these sequences to form amyloid aggregates using a cell-based selection assay where sequences are fused to the nucleation domain of the yeast prion Sup35 (SupN) and the aggregation of the fused peptides nucleates the aggregation of endogenous Sup35, a process required for yeast growth in selective conditions (Chandramowlishwaran *et al*, 2018). Cells expressing Bri2 do not grow in these conditions, revealing that the unextended peptide does not nucleate amyloid aggregation. In sharp contrast, the expression of ADan leads to 25.14 ± 3.41% of the cells growing in selection media over non-selective conditions, a 2-fold increase compared to Aβ42 (Seuma *et al*, 2021) (Fig. 1, Supp. Fig. 1a). Cells carrying ABri show very limited growth in the assay (0.0065 ± 0.0039%) with no significant difference compared to expression of the reporter without any fusion (SupN) (Fig. 1b, Supp. Fig. 1a). Growth of cells expressing Bri2, ABri or ADan fusions in non-selective conditions revealed that their growth rates were not confounded by differential toxicity (Supp. Fig. 1b).

ADan and ABri share the first 22 residues and differ only in the last 12 amino acids (Fig. 1a) suggesting it is the specific extension of ADan that encodes amyloid potential. Hence, we evaluated all possible ADan C-term truncations using the selection assay described above combined to deep sequencing. We designed and synthesized a

library of ADan single amino acid substitutions and truncations and expressed it fused to SupN in yeast cells that went under selection. We quantified an amyloid nucleation score for each variant by deep sequencing before and after selection (Fig. 1c) (Seuma *et al*, 2021). We found that a Bri2 extension of just 2 amino acids (N24 and L25) is enough to produce a significant increase in amyloid nucleation in comparison to Bri2 (Fig. 1d). This effect becomes stronger with a 3 amino acid extension (N24, L25 and F26), leading to a similar nucleation score as that of the full-length ADan peptide. Together with the previous observation that a 28 residue long ADan truncation forms beta-sheet rich structures *in vitro* (Todd *et al*, 2016), these findings suggest that the amyloid potential of ADan is encoded by the first residues that get translated upon stop-loss.

Deep mutagenesis of ADan

To investigate and compare the effects of diverse genetic variants on the aggregation of ADan we then looked at the impact of all possible single amino acid changes in the ADan sequence by the same approach. The resulting nucleation scores are reproducible between replicates (Fig. 2a, Supp. Fig. 1d) and correlate well with the effects of variants quantified individually ($R = 0.94$, p-value<0.01, Fig.2b).

Altogether, we measured amyloid nucleation for 677 variants. The distribution of mutational effects for ADan nucleation has a strong bias towards reduced amyloid nucleation, with 44% of single amino acids substitutions reducing nucleation but only 25% increasing it (Z-test, false discovery rate [FDR] = 0.1, Fig. 2c,d). Among the variants that decrease nucleation, 66 have nucleation scores similar to mean Bri2-truncations nucleation score. Finally, 200 variants have a nucleation score similar to WT ADan.

ADan nucleation core is a product of Bri2 extension

Inspecting the heatmap of mutational effects for amino acid changes at all positions in ADan reveals that most of the mutations that decrease amyloid nucleation are localized between residues L20-L25 (Fig. 2e, Supp. Fig. 2). In this continuous stretch of residues most mutations decrease amyloid nucleation and all substitutions to prolines disrupt it, suggesting tight structural constrains. Thus, we propose that this region forms the inner core of Adan fibrils. In particular, the most important residue of the core seems to be N24 as all mutations at this position lead to a decrease in nucleation. Charges in this region are not tolerated, except at position 23. The two residues that precedes the core (E18 and T19) are also important as many of the mutations (mainly to hydrophobic amino acids) at these positions decrease nucleation. Albeit, distant from this region, residues N4 and E31 are also crucial for the nucleation of Adan as most substitution of these residues decrease nucleation suggesting longer distance interactions with residues of the amyloid core or a role in modulating the conformations of the monomeric Adan ensemble. Overall, asparagines and glutamic acids are key in Adan nucleation as most substitutions of these amino acids decrease nucleation (Fig. 2f).

ADan gatekeeper residues

We consider as gatekeepers those positions where mutations are more likely to increase nucleation rather than decrease it. We identified 5 residues as gatekeepers: F6 (12 mutations increasing nucleation and 2 decreasing), I8 (15 increasing), K14 (11 increasing and 2 decreasing), A16 (12 increasing and 5 decreasing) and L27 (11 increasing and 3 decreasing) (Fig.2e). Surprisingly, and in contrast to what was previously found for other amyloids (Rousseau *et al*, 2006; Bhoite *et al*, 2023; Sant'Anna *et al*, 2014), including Aβ42 (Seuma *et al*, 2021), the gatekeepers are all aliphatic residues with the exception of K14.

Combination of SNPs and FDD mutation affects ADan nucleation propensity

14 SNPs are reported for *bri2* in the GnomAD database (Chen *et al*, 2022) that affect the C-term of this protein and 8 SNPs in the 3' UTR (up to ABri stop codon) (Supp. Table 1). We asked what the impact of these SNPs would be in combination with the 10 nt duplication that leads to the ADan peptide formation. Most of the SNPs in the context of ADan peptide decrease nucleation or do not affect the nucleation of the peptide. Only two of the reported SNPs (rs1272422512 and rs1455609180), that result in R9Q and K32T mutations, increase nucleation. However, we identify another 53 SNVs that are able to significantly increase nucleation in ADan, providing a comprehensive atlas for the evaluation of genetic variation in this sequence.

Bri2 stop-loss mutations, different than those causing FBD, increase nucleation

We then employed the same strategy for studying the effect of diverse ABri genetic variants on amyloid nucleation. As the expression of ABri - and its saturation mutagenesis library - result in limited growth in the selection assay, we applied a stringent filter in order to proceed with analysis only with those variants that have a low normalized sigma value (see Methods, Supp. Fig. 1c) and thus, an accurate estimate. Most of the confident variants are those that increase nucleation. The resulting nucleation scores are reproducible between replicates (Fig. 3a, Supp. Fig. 1e) and correlate well with the effects of variants quantified individually (R = 0.83, Fig.3b). The heatmap of mutational effects for amino acid changes at all positions in ABri reveals that substitutions R24L and R24S increase nucleation (Fig. 3c, Supp. Fig. 2b), meaning that the peptides that are a product of FCD and FKD mutations are more prone to aggregate than ABri. ABri stop-loss due to SNVs can lead to stop codon substitution by R (FBD), L (FCD), S (FKD), G, W or C. SNV stop-loss leading to G at position 24 of the extended peptide does not differ from ABri nucleation propensity. However, a C at position 24 significantly increases nucleation. When tested individually, the strength of R24C nucleation is half of Aβ42 nucleation and 25% of that of ADan (Fig. 3b,c). 10 out of 11 substitutions, beyond SNVs, at position 24 increase nucleation, revealing that R24 may act as a gatekeeper of ABri nucleation. In addition, residue N13 is also a candidate gatekeeper as any change at this position increases ABri nucleation.

Mutation L20C severely increases ABri nucleation

A particular mutation in ABri is the strongest nucleator of this library: L20C (Fig. 3c). Cells expressing this ABri mutant have a relative growth rate (colonies in selective condition / colonies in non-selective conditions) of 63.1 ± 7.1% (-URA-Ade/-URA), 2.7 times more than ADan peptide (23.1 ± 3.6%), confirming that this sequence is a very fast nucleator.

ADan nucleation is different from Aβ42 and ABri nucleation

In order to evaluate if the impact of mutations on amyloid nucleation of ADan depends on similar properties as those driving Aβ42 nucleation, we compared the nucleation scores obtained after mutagenizing both peptides. There is no significant correlation between the nucleation of these peptides when comparing the effect of specific amino acid changes (Fig. 4a), as well as mutations from the same amino acid (Fig. 4b) and mutations to the same amino acid (Fig. 4c). In addition, to compare mutational effect per position, we aligned ADan and Aβ42 peptides in three different ways, at the N-term, C-term and with ClustalO (Madeira *et al*, 2022). Regardless of the strategy, no significant correlation at the aligned positions was detected (Supp. Fig. 3a,b,c).
Similarly, we compared ADan and ABri nucleation scores. None of the comparisons between ADan and ABri revealed significant correlations between the nucleation scores of these peptides (Fig. 4d,e; Supp. Fig. 3d). We also tested the correlation in those positions where we have at least 5 confident measurements in the ABri library (Fig. 4f) uncovering significant positive correlations only for positions 13 (N), 26 (F and V) and 32 (K and E, ADan and ABri respectively).

Structural insights may explain the difference in aggregation propensity between Bri2 extended peptides

In ADan, substitutions to prolines increase nucleation up to residue 21 but between residues 21-26 (where we propose that the amyloid core is located) decrease nucleation. This increase in nucleation due to proline substitution in the first part of the peptide may suggest the presence of secondary structure elements in monomeric ADan that would be able to prevent nucleation. Thereafter, we used state-of-the-art structure predictors for predicting Bri2, ABri and ADan structures. Alphafold and ESM (Jumper *et al*, 2021; Lin *et al*, 2023) predict Bri2 structure as a beta-hairpin and ABri structure as a beta-sheet or a beta-hairpin with a disordered tail (Supp. Fig. 4a,b). While there is consensus between predictors for Bri2 and ABri, the outcome is different when it comes to ADan, that is predicted as a beta-hairpin with a disordered tail with a small helix (low confidence in this region) by Alphafold and as two helices (low confidence) by ESM. This may suggest that Bri2 and ABri have more stable native structures, while ADan may sample a higher energy ensemble of conformations. In order to explore these ensembles of configurations, we used the CALVADOS coarse-grained force field (Tesei & Lindorff-Larsen, 2023) to run dynamic simulations of Bri2, ABri and ADan peptides.The nonionic energy contact map for ADan reveals a cluster of interactions between residues 5-11 with residues 19-28 (Supp. Fig. 4C). In contrast, the energy contact map for ABri shows a smaller cluster of interactions that comprehends residues 6-11 and 19-24. Bri2, has an even smaller cluster of

interactions that is similar to the one of ABri. While residues F6, A7 and I8 are close to residues L20 and I21 in the predicted ABri three dimensional structures, when it comes to ADan a larger set of favorable pair-wise interactions are predicted by CALVADOS force field and may contribute to overcome the energy barrier to ADan nucleation.

Discussion

Bri2 gene mutations affecting its stop codon are associated with familial forms of dementia: FBD, FDD, FCD and FKD. Upon furin processing and cleavage, these mutations lead to the production of aberrant C-term extended peptides known as ABri and ADan (FBD and FDD, respectively) which form amyloid aggregates in the affected individuals (Vidal *et al*, 1999, 2000).

The resulting toxicity is proposed to depend both on the reduced expression of the precursor protein (Tamayev *et al*, 2010b, 2010a) and by the fate of the cleaved sequences. Both ADan and ABri, as well as the peptides generated in FCD and FKD, are 34 amino acids long. While the sequence of the first 22 residues is common, these peptides differ in the last 12 C-terminal residues.

The clinical features of these forms of dementias are strikingly similar to those of AD, suggesting a similar gain of toxic function mechanism upon mutation, but the primary sequences of ADan and ABri differ entirely from that of Aβ42, posing the question of how these sequences aggregate. In order to systematically measure the formation of amyloids by these two distinct peptides and to gather mechanistic insights on how they turn into amyloids upon extension, we combined DMS to a selection assay that reports on amyloid nucleation and obtained a comprehensive mutational landscape of the two peptides. The reporter assay revealed that Bri2, ABri and ADan peptides have different nucleation propensity. Bri2 has no signal in the cell-based assay, meaning that the 23 amino acid long peptide is not nucleating amyloid aggregates. While ADan peptide very rapidly forms amyloids in the cell-based assay, 2-fold faster than Aβ42 for reference, ABri is at the lower end of the detection capability of the assay.

In both FBD and FDD patients amyloid depositions formed by the aberrant peptides have been detected (Vidal *et al*, 1999, 2000), revealing that both peptides can nucleate amyloid aggregation. However, the difference we measure is in line with mice models for FBD and FDD. Transgenic mice expressing ABri do not exhibit amyloid depositions (Pickford *et al*, 2006) while those expressing ADan feature parenchymal amyloid deposition and cerebral amyloid angiopathy (Vidal *et al*, 2009). Furthermore, there's evidence for ADan leading to higher toxicity than ABri in a *Drosophila* model, possibly due to a faster rate of aggregation of the ADan peptide (Marcora *et al*, 2014).

The low nucleation rate of the ABri peptide in our assay suggests a more drastic difference in aggregation rates of the two peptides than what has been proposed from *in vitro* experiments so far. The ability of both peptides to aggregate into beta-sheet rich aggregates has been observed *in vitro* (Todd *et al*, 2016, 2014), but there is no quantitative comparison reporting on the aggregation kinetics of these two peptides,

i.e. on how fast they nucleate amyloids. The aggregation propensity of these peptides is also thought to be affected by post-translational modifications such as the N-term glutamate to pyroglutamate conversion, disulfide bond formation and truncations. However, pyroglutamate formation is not compatible with the nucleation assay employed here as the peptide N-term is engaged in the SupN fusion. Moreover, the reporter assay takes place in the yeast cytosol making the formation of disulfide bonds unlikely. *In vitro*, it has been shown that only the oxidized form of ABri aggregates in toxic soluble oligomers (El-Agnaf *et al*, 2001a). In contrast, both ADan oxidized and reduced forms can form oligomers, with the reduced form resulting in higher neurotoxicity (Gibson *et al*, 2003).

In ADan, we identify a continuous stretch of residues (L20-L25) that are crucial for amyloid nucleation as most amino acid changes at these positions decrease nucleation. This region contains four aliphatic residues that together with polar and charged residues (N4, E12, N13 and E31) are likely to be involved in the amyloid interface at the core of the aggregates formed by this peptide. Surprisingly, and in contrast to Aβ42, gatekeeper residues in ADan are mainly hydrophobic (F6, I8, A16 and L27), with the exception of K14.
Overall, correlating nucleation scores for the same types of mutation reveals that the nucleation of amyloids by ADan differs from that of ABri, as well as from Aβ42. Changes in hydrophobicity can explain very little of the variance in the Adan dataset (Supp. Fig. 5), where the largest changes in nucleation involve polar and negatively charged residues. It would have been impossible to predict the types of changes that increase or decrease nucleation in ADan even on the basis of the complete mutational dataset of another amyloid peptide, Aβ42 (Seuma *et al*, 2022), highlighting the need of systematically measuring amyloid nucleation for these and other peptides that are known to result in similar phenotypes (Seuma & Bolognesi, 2022).

The low signal/noise ratio for ABri also makes it difficult to identify positions where mutations disrupt amyloid formation and therefore where a likely fibril core could be established. However, we can still confidently identify those positions where single amino acid substitutions are more likely to increase ABri nucleation. Of special interest and concern are mutations affecting the position R24 as this is the position coded by the mutation in the original stop codon. FCD and FKD mutations (R24L and R24S) increase nucleation of ABri peptide. The R24C substitution here, which results from a SNV affecting the stop codon, severely increases nucleation generating one of the strongest nucleators in the ABri library together with L20C. Both substitutions occur at a two residue distance from Cys22, forming a Cys-X-Cys motif. This type of motif can bind metal ions as zinc or copper (Scheller *et al*, 2018) and modulate amyloid nucleation. Alternatively, Cys-Cys ladders in the fibril axis could stabilize the amyloid conformation, as reported in other amyloid-forming sequences (Mompeán *et al*, 2018). Overall, ~90% of the mutations of R24 quantified here increase nucleation, with 8 mutations, besides those leading to FCD and FKD, that are likely to be pathogenic.

This suggests that the majority of extensions of Bri2 resulting from stop-loss are of concern.

Circular Dichroism experiments revealed that, before evolving into beta-sheet rich structures, ABri and ADan are largely disordered (Todd *et al*, 2016, 2014). On the other hand, structure predictions using Alphafold and ESM show a consensus for Bri2 and ABri peptides, with predicted beta strands, but not for ADan. It is possible that Bri2 and ABri may populate a more stable ensemble of conformations in their monomeric state, while ADan could more easily overcome the energy barrier to amyloid nucleation. Dynamic simulations using the CALVADOS force field (Tesei & Lindorff-Larsen, 2023) reveal energetically favorable interactions between residues 5-11 and 19-28 in ADan and between shorter stretches in ABri (residues 6-11 and 19-24), some of which could be prone to nucleate amyloid fibril formation. In the lack of a clear consensus, even on the basis of the most advanced structure prediction tools (Jumper *et al*, 2021; Lin *et al*, 2023), datasets such as the ones generated here can be of great help to guide or re-train structural and molecular dynamics models for these peptides.

Several SNPs in Bri2 are reported in GnomAD (Chen *et al*, 2022). Among them, 14 occur before position 23 of the cleaved peptide and 8 in the 3'UTR, the region that gets translated as the novel C-term of the extended peptides. While these SNPs are benign, they could turn out to be pathogenic in individuals carrying such SNPs in combination with a stop-loss mutation. SNPs mutation in the ADan background do not increase nucleation, except for R9Q (rs1272422512) and K32T (rs1455609180). In contrast, in the ABri background, 4 SNPs increase nucleation: A7S (rs578006622), I8V (rs139394220), L20S (rs1428007068) and K28R (rs760120359). The remaining SNPs do not alter the nucleation of these peptides or can even decrease it, therefore acting as protective variants.
TGA is the most susceptible codon to stop-codon readthrough, a phenomenon which happens at low frequency of 0.01-0.1% (Floquet *et al*, 2012; Dabrowski *et al*, 2015), and that can cause an extension of Bri2 at the C-terminus. In this scenario the resulting Bri2 extended peptide would be similar to the ABri peptide. This mechanism can be promoted by small molecules such as Gentamycin or 5-Fluoruridine, drugs that are already used as treatment for different conditions (Toledano *et al*, 2023). Patients under treatment with these drugs may be suffering stop-codon readthrough and, the translation of CAEs that may drive amyloid fibrils aggregation.
Mapping the effects of all ADan truncation shows that a Bri2 two amino acid extension is enough for increasing nucleation in comparison to Bri2 peptide. A three amino acid extension severely increases nucleation. A similar increase in nucleation is seen for the remaining truncations. These results are in agreement with a previous study showing that an ADan truncation at position 28 can form beta-sheet rich structures *in vitro* (Todd *et al*, 2016). Our results reveal that the nucleation core is likely formed by this short extension itself and highlights the importance of studying other possible Bri2 extensions that may play important roles in the development of dementia.

Furthermore, our research reinforces the importance of studying CAEs coded in 3' UTRs as firstly proposed by Rebelo *et al* (Rebelo *et al*, 2016). Stop-loss can lead to the translation of CAEs that may be the driving force of neurodegenerative disease or systemic amyloidosis. The authors identified CAEs in NEFH, NEFL and FUS 3' UTRs by using a range of amyloid aggregation predictors. However, the output scores of these tools poorly correlate with nucleation scores. This is likely due to the fact that these predictors were trained on a small number of sequences and are based exclusively on hydrophobicity scores or beta-sheet propensity features that alone cannot explain amyloid nucleation. Indeed, here we show the relevance of polar and charged residues in ADan amyloid nucleation. Thus, a systematic and high-throughput identification of CAEs is needed as it will open new avenues for the understanding of familial amyloidosis.

Methods

Library design

The designed libraries contain a total of 1088 unique ADan nucleotide variants (680 single protein variants encoded by 1, 2, and 3 nucleotide changes) and 1088 unique ABri nucleotides variants (680 single protein variants encoded by 1, 2, and 3 nucleotide changes).

Plasmid Library construction

Libraries were synthesized by Integrated DNA Techonologies (IDT) as oligo pools covering the ADan and/or ABri 102 nt sequence, flanked by 25 nt upstream and 21 nt downstream constant regions for cloning. 10 ng of the library were amplified by PCR (Q5 high-fidelity DNA polymerase, NEB) for 12 cycles with primers annealing to the constant regions (primers MM_01-02, Supplementary Data 2). The product was purified from a 2% agarose gel (QIAquick Gel Extraction Kit, Qiagen). In parallel, the PCUP1-Sup35N plasmid was linearized by PCR (Q5 high-fidelity DNA polymerase, NEB; primers MM_03-04, Supplementary Data 2). The product was purified from a 1% agarose gel (QIAquick Gel Extraction Kit, Qiagen). The oligo pool was then ligated into 200 ng of the linearized plasmid in a 1:10 (vector:insert) ratio by a Gibson approach with 3 h of incubation followed by dialysis for 3 h on a membrane filter (MF-Millipore 0.025 µm membrane, Merck) and vacuum concentration. The product was transformed into 10-beta Electrocompetent *E. coli* (NEB), by electroporation with 2.0 kV, 200 Ω, 25 µF (BioRad GenePulser machine). Cells were recovered in SOC medium for 30 min and grown overnight in 30 ml of LB ampicillin medium. A small amount of cells were also plated in LB ampicillin plates to assess transformation efficiency. A total of >1M transformants were estimated, meaning that each variant in the library is represented ~1000 times. 5 ml of overnight culture were harvested to purify the library with a mini prep (QIAprep Miniprep Kit, Qiagen).

Yeast transformation

Saccharomyces cerevisiae [psi-pin-] (MATα ade1–14 his3 leu2-3,112 lys2 trp1 ura3–52) provided by the Chernoff lab (Chandramowlishwaran *et al*, 2018) was used in all experiments in this study. Yeast cells were transformed with the libraries in three biological replicates. Per replicate, an individual colony was grown overnight in 3 ml YPDA medium at 30 °C and 4 g. Cells were diluted in 40 ml to OD600 = 0.3 and grown for 4–5 h. When cells reached the exponential phase (OD~0.8–0.9) cells were harvested at 3000 × g for 5 min, washed with 50 ml milliQ, centrifuged at 3000 × g for 5 min and washed with 25 ml SORB buffer (100 mM LiOAc, 10 mM Tris pH 8.0, 1 mM EDTA, 1 M sorbitol). Cells were resuspended in 1.4 ml of SORB and incubated 30 min on an orbital shaker. After incubation, 800 ng of library and 30 ul of ssDNA (UltraPure, Thermo Scientific) were added and incubated for 5 min at room temperature and then 10 min on an orbital shaker. 6 ml of YTB-PEG (100 mM LiOAc, 10 mM Tris pH 8.0, 1 mM EDTA, 40% PEG 3350) and 580 ul of DMSO were added to the cells. Heat-shock was performed at 42 °C for 20 min in a liquid bath with intermittent shaking. Cells were harvested and incubated in 50 ml of recovery medium (YPDA medium + Sorbitol 0.5 M) for 1 h at 30 °C. Cells were harvested and grown in 50 ml plasmid selection medium (-URA, 2% glucose) for 50 h at 30 °C. A small amount of cells were also plated in plasmid selection solid medium to assess transformation efficiency. 60,000-100,000 transformants were estimated for each biological replicate, meaning that each variant in the library is represented at least 50 times. After 50 h, cells were diluted in 50 ml plasmid selection medium to OD= 0.05 and grown exponentially for 15 h. Finally, the culture was harvested and stored at −80 °C in 25 % glycerol.

Selection experiments

In vivo selection assays were performed in three biological replicates. For each technical replicate, cells were thawed from −80 °C in 50 ml plasmid selection medium at OD = 0.05 and grown until exponential for 15 h. At this stage, cells were harvested and resuspended in 50 ml protein induction medium (-URA, 2% glucose, 100 μM Cu2SO4) at OD = 0.1. After 24 h the 40 ml input pellets were collected and cells were plated on -ADE-URA selection medium in 145-cm2 plates (Nunc, Thermo Scientific). Plates were incubated at 30 °C for 7 days. Finally, colonies were scraped off the plates with PBS 1x and harvested by centrifugation to collect the output pellets. Both input and output pellets were stored at −20 °C before DNA extraction.

DNA extraction and sequencing library preparation

One input and one output pellets for each biological replicate were resuspended in 0.5ml extraction buffer (2% Triton-X, 1% SDS, 100 mM NaCl, 10 mM Tris-HCl pH 8, 1 mM EDTA pH 8). They were then freezed for 10 min in an ethanol-dry ice bath and heated for 10 min at 62 °C. This cycle was repeated twice. 0.5ml of phenol:chloroform:isoamyl (25:24:1 mixture, Thermo Scientific) was added together with glass beads (Sigma). Samples were vortexed for 10 min and centrifuged for 30 min at 2000 × g. The aqueous phase was then transferred to a new tube, and mixed again with phenol:chloroform:isoamyl, vortexed and centrifuged for 45 min at 2000 ×

g. Next, the aqueous phase was transferred to another tube with 1:10 V 3 M NaOAc and 2.2 V cold ethanol 96% for DNA precipitation. After 30 min at −20 °C, samples were centrifuged and pellets were dried overnight. The following day, pellets were resuspended in 0.3 ml TE 1X buffer and treated with 10 µl RNAse A (Thermo Scientific) for 30 min at 37 °C. DNA was finally purified using 10 µl of silica beads (QIAEX II Gel Extraction Kit, Qiagen) and eluted in 30 µl elution buffer. Plasmid concentrations were measured by quantitative PCR with SYBR green (Merck) and primers annealing to the origin of replication site of the PCUP1-Sup35N plasmid at 58 °C for 40 cycles (primers MM_05-06, Supplementary Data 2). The library for high-throughput sequencing was prepared in a two-step PCR (Q5 high-fidelity DNA polymerase, NEB). In PCR1, 50 million of molecules were amplified for 15 cycles with frame-shifted primers with homology to Illumina sequencing primers (primers MM_07–20, Supplementary Data 2). The products were purified with ExoSAP treatment (Affymetrix) and by column purification (MinElute PCR Purification Kit, Qiagen). They were then amplified for 12 cycles in PCR2 with Illumina-indexed primers (primers MM_21–37, Supplementary Data 2). The six samples of each replicate were pooled together equimolarly and the final product was purified from a 2% agarose gel with 20 µl silica beads (QIAEX II Gel Extraction Kit, Qiagen). The library was sent for 125 bp paired-end sequencing in an Illumina HiSeq2500 sequencer at the CRG Genomics core facility. In total, >50 million paired-end reads were obtained representing >1000x read coverage for each designed variant in the library.

Individual variant testing

Selected variants for individual testing were obtained by PCR linearisation (Q5 high-fidelity DNA polymerase, NEB) with mutagenic primers (primers MM_38-59, Supplementary Data 2). PCR products were treated with Dpn1 overnight and transformed in DH5α competent *E. coli*. Plasmids were purified by mini prep (QIAprep Miniprep Kit, Qiagen) and transformed into yeast cells. All mutated plasmids were verified by Sanger sequencing. For measuring growth in selective conditions, yeast cells expressing individual variants were grown overnight in a plasmid selection medium (-URA 2% glucose). They were then diluted to OD 0.1 in protein induction medium (-URA 2% glucose 100 µM Cu_2SO_4) and grown for 24 h. Cells were plated on -URA (control) and -ADE-URA (selection) plates in three independent replicates and allowed to grow for 7 days at 30 °C. Adenine growth was calculated as the percentage of colonies in -ADE-URA relative to colonies in -URA.

For individual growth rate measurements, yeast cells expressing individual variants were grown overnight in plasmid selection medium (-URA 2% glucose) and diluted to OD 0.2 until exponential. They were then diluted again to OD 0.1 in non-inducing (-URA 2% glucose) and inducing (-URA 2% glucose 100 µM Cu_2SO_4) protein expression mediums. Cell growth was measured at 30 °C for >48 h at 10 min intervals in a microplate reader (Spark, Tecan) in three biological replicates. Growth rates were calculated using the GrowthCurver package in R.

Data processing

FastQ files from paired end sequencing of the libraries were processed using DiMSum (https://github.com/lehner-lab/DiMSum) (Faure *et al*, 2020), an R pipeline for analyzing deep mutational scanning data. 5′ and 3′ constant regions were trimmed, allowing a maximum of 20% of mismatches relative to the reference sequence. Sequences with a Phred base quality score below 30 were discarded. Non-designed variants were also discarded for further analysis, as well as variants with fewer than 200 input reads in all of the replicates. Estimates from DiMSum were used to choose the filtering thresholds.

Nucleation scores and error estimates

The DiMSum package (https://github.com/lehner-lab/DiMSum) (Faure *et al*, 2020) was also used to calculate nucleation scores (NS) and their error estimates for each variant in each biological replicate as:

$$\text{Nucleation score} = ESi - ESwt$$

Where ESi = log(Fi OUTPUT) - log(Fi INPUT) for a specific variant and ESwt = log(Fwt OUTPUT) - log(Fwt INPUT) for ADan or ABri WT.

NSs for each variant were merged across biological replicates using error-weighted mean and centered to the WT NS. All NS and associated error estimates are available in Supplementary Data 1.

ADan library has a good signal/noise ratio due to its significant growth in the selection assay. After NSs centering, each variant was tested against the WT NS at FDR=0.1 and classified in three possible groups: WT-like, NS_inc and NS_dec. Sigma values were normalized to the interquartile range and WT-like variants with a normalized sigma value above a cut-off of 0.2 were excluded. The same procedure was part of ABri library processing, but as it has a poor signal/noise ratio a more stringent filtering process was applied. Only those variants with a normalized sigma value <=0.2 were kept and analyzed.

Data analysis

Variants in the library. NS was obtained for 677 unique ADan variants and 679 unique ABri variants. 664 confident estimates were obtained for the ADan library and 177 in the ABri library.

Aggregation and variant effect predictors. For the aggregation predictors (Tango, Amypred, Camsol) (Fernandez-Escamilla *et al*, 2004; Charoenkwan *et al*, 2022; Sormanni *et al*, 2015) scores were calculated per sequence and, if necessary, individual residue level scores were summed to obtain a score per single AA mutation sequence. We also used the Kyte-Doolittle hydrophobicity scale (Kyte & Doolittle, 1982).

Structural analysis

Structure prediction was performed using AlphaFold2 colab (https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb#scrollTo=33g5IIegij5R) (Mirdita *et al*, 2022) and ESM (https://esmatlas.com/) (Lin *et al*, 2023). 3D structures were visualized using Visual Molecular Dynamics software (VMD) (Humphrey *et al*, 1996).

Molecular dynamic simulations were run in the IDR.lab colab (https://colab.research.google.com/github/KULL-Centre/EnsembleLab/blob/main/IDRLab.ipynb#scrollTo=-SCC7r8ltUTy) (Tesei *et al*, 2023).

Statistics and reproducibility

Based on the transformation efficiency, each variant in the designed libraries (n = 1088 nucleotide variants each) is expected to be represented at least 10x at each step-in selection experiments and library preparation. In terms of sequencing, reads that did not pass the QC filters using the DiMSum package were excluded (https://github.com/lehner-lab/DiMSum) (Faure *et al*, 2020). The experiments were not randomized. The Investigators were not blinded to allocation during experiments and outcome assessment.

Data availability

Raw sequencing reads and processed nucleation scores are deposited in NCBI's Gene Expression Omnibus (GEO) as GSE244612. The processed reads and nucleation scores are also available as Supplementary Table 1.

Code availability

All scripts used for downstream analysis and to reproduce all figures are in https://github.com/BEBlab/ADanABri.

Acknowledgements

Work in the lab of BB is supported by the la Caixa Research Foundation project 'DeepAmyloids' (LCF/PR/HR21/52410004), by the Spanish Ministry of Science, Innovation and Universities (PID2021-127761OB-I00 and RYC2020-028861-I, funded by MCIN/AEI/ 10.13039/501100011033, "ERDF A way of making Europe" and "ESF Investing in your future"). IBEC is a member of the CERCA Program / Generalitat de Catalunya. We thank the Chernoff lab for providing strains and plasmids and the CRG Genomics core technology for sequencing.

References

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS & Sunyaev SR (2010) A method and server for predicting damaging missense mutations. *Nat Methods* 7: 248–249

Bauer AH, Basta DW, Hornick JL & Dong F (2023) Loss of function SMAD4 nonstop mutations in human cancer. *Histopathology* 82: 1098–1104

Bhoite SS, Kolli D, Gomulinski MA & Chapman MR (2023) Electrostatic interactions mediate the nucleation and growth of a bacterial functional amyloid. *Front Mol Biosci* 10

Bock AS, Günther S, Mohr J, Goldberg L V., Jahic A, Klisch C, Hübner CA, Biskup S & Beetz C (2018) A nonstop variant in *REEP1* causes peripheral neuropathy by unmasking a 3′UTR-encoded, aggregation-inducing motif. *Hum Mutat* 39: 193–196

Brandes N, Goldman G, Wang CH, Ye CJ & Ntranos V (2023) Genome-wide prediction of disease variant effects with a deep protein language model. *Nat Genet* 55: 1512–1522

Cantlon A, Frigerio CS, Freir DB, Boland B, Jin M & Walsh DM (2015) The Familial British Dementia Mutation Promotes Formation of Neurotoxic Cystine Cross-linked Amyloid Bri (ABri) Oligomers. *Journal of Biological Chemistry* 290: 16502–16516

Chandramowlishwaran P, Sun M, Casey KL, Romanyuk A V., Grizel A V., Sopova J V., Rubel AA, Nussbaum-Krammer C, Vorberg IM & Chernoff YO (2018) Mammalian amyloidogenic proteins promote prion nucleation in yeast. *Journal of Biological Chemistry* 293: 3436–3450

Charoenkwan P, Ahmed S, Nantasenamat C, Quinn JMW, Moni MA, Lio' P & Shoombuatong W (2022) AMYPred-FRL is a novel approach for accurate prediction of amyloid proteins by using feature representation learning. *Sci Rep* 12: 7697

Chen S, Francioli LC, Goodrich JK, Collins RL, Kanai M, Wang Q, Alföldi J, Watts NA, Vittal C, Gauthier LD, *et al* (2022) A genome-wide mutational constraint map quantified from variation in 76,156 human genomes. *bioRxiv*: 2022.03.20.485034

Coomaraswamy J, Kilger E, Wölfing H, Schäfer C, Kaeser SA, Wegenast-Braun BM, Hefendehl JK, Wolburg H, Mazzella M, Ghiso J, *et al* (2010) Modeling familial Danish dementia in mice supports the concept of the amyloid hypothesis of Alzheimer's disease. *Proceedings of the National Academy of Sciences* 107: 7969–7974

Dabrowski M, Bukowy-Bieryllo Z & Zietkiewicz E (2015) Translational readthrough potential of natural termination codons in eucaryotes – The impact of RNA sequence. *RNA Biol* 12: 950–958

El-Agnaf OMA, Nagala S, Patel BP & Austen BM (2001a) Non-fibrillar oligomeric species of the amyloid ABri peptide, implicated in familial British dementia, are

more potent at inducing apoptotic cell death than protofibrils or mature fibrils11Edited by J. Karn. *J Mol Biol* 310: 157–168

El-Agnaf OMA, Sheridan JM, Sidera C, Siligardi G, Hussain R, Haris PI & Austen BM (2001b) Effect of the Disulfide Bridge and the C-Terminal Extension on the Oligomerization of the Amyloid Peptide ABri Implicated in Familial British Dementia. *Biochemistry* 40: 3449–3457

Faure AJ, Schmiedel JM, Baeza-Centurion P & Lehner B (2020) DiMSum: an error model and pipeline for analyzing deep mutational scanning data and diagnosing common experimental pathologies. *Genome Biol* 21: 207

Fernandez-Escamilla A-M, Rousseau F, Schymkowitz J & Serrano L (2004) Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat Biotechnol* 22: 1302–1306

Floquet C, Hatin I, Rousset J-P & Bidou L (2012) Statistical Analysis of Readthrough Levels for Nonsense Mutations in Mammalian Cells Reveals a Major Determinant of Response to Gentamicin. *PLoS Genet* 8: e1002608

Frazer J, Notin P, Dias M, Gomez A, Min JK, Brock K, Gal Y & Marks DS (2021) Disease variant prediction with deep generative models of evolutionary data. *Nature* 599: 91–95

Ghiso JA, Holton J, Miravalle L, Calero M, Lashley T, Vidal R, Houlden H, Wood N, Neubert TA, Rostagno A, *et al* (2001) Systemic Amyloid Deposits in Familial British Dementia. *Journal of Biological Chemistry* 276: 43909–43914

Gibson G, Gunasekera N, Lee M, Lelyveld V, El-Agnaf OMA, Wright A & Austen B (2003) Oligomerization and neurotoxicity of the amyloid ADan peptide implicated in familial Danish dementia. *J Neurochem* 88: 281–290

Hamby SE, Thomas NS, Cooper DN & Chuzhanova N (2011) A meta-analysis of single base-pair substitutions in translational termination codons (‘nonstop’ mutations) that cause human inherited disease. *Hum Genomics* 5: 241

Holton JL, Ghiso J, Lashley T, Rostagno A, Guerin CJ, Gibb G, Houlden H, Ayling H, Martinian L, Anderton BH, *et al* (2001) Regional Distribution of Amyloid-Bri Deposition and Its Association with Neurofibrillary Degeneration in Familial British Dementia. *Am J Pathol* 158: 515–526

Humphrey W, Dalke A & Schulten K (1996) VMD: Visual molecular dynamics. *J Mol Graph* 14: 33–38

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, *et al* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* 596: 583–589

Kyte J & Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. *J Mol Biol* 157: 105–132

Lashley T, Holton JL, Verbeek MM, Rostagno A, Bojsen-Møller M, David G, van Horssen J, Braendgaard H, Plant G, Frangione B, *et al* (2006) Molecular chaperons, amyloid and preamyloid lesions in the BRI2 gene-related dementias: a morphological study. *Neuropathol Appl Neurobiol* 32: 492–504

Lin Z, Akin H, Rao R, Hie B, Zhu Z, Lu W, Smetanin N, Verkuil R, Kabeli O, Shmueli Y, *et al* (2023) Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science (1979)* 379: 1123–1130

Liu X, Chen K-L, Wang Y, Huang Y-Y, Chen S-D, Dong Q, Cui M & Yu J-T (2021) A Novel ITM2B Mutation Associated with Familial Chinese Dementia. *Journal of Alzheimer’s Disease* 81: 499–505

Madeira F, Pearce M, Tivey ARN, Basutkar P, Lee J, Edbali O, Madhusoodanan N, Kolesnikov A & Lopez R (2022) Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res* 50: W276–W279

Marcora MS, Fernández-Gamba AC, Avendaño LA, Rotondaro C, Podhajcer OL, Vidal R, Morelli L, Ceriani MF & Castaño EM (2014) Amyloid peptides ABri and ADan show differential neurotoxicity in transgenic Drosophila models of familial British and Danish dementia. *Mol Neurodegener* 9: 5

Mignogna ML, Ficarella R, Gelmini S, Marzulli L, Ponzi E, Gabellone A, Peschechera A, Alessio M, Margari L, Gentile M, *et al* (2022) Clinical characterization of a novel *RAB39B* nonstop mutation in a family with ASD and severe ID causing RAB39B downregulation and study of a *Rab39b* knock down mouse model. *Hum Mol Genet* 31: 1389–1406

Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S & Steinegger M (2022) ColabFold: making protein folding accessible to all. *Nat Methods* 19: 679–682

Mompeán M, Li W, Li J, Laage S, Siemer AB, Bozkurt G, Wu H & McDermott AE (2018) The Structure of the Necrosome RIPK1-RIPK3 Core, a Human Hetero-Amyloid Signaling Complex. *Cell* 173: 1244-1253.e10

Monaghan L, Longman D & Cáceres JF (2023) Translation-coupled mRNA quality control mechanisms. *EMBO J*

Ng PC & Henikoff S (2001) Predicting Deleterious Amino Acid Substitutions. *Genome Res* 11: 863–874

Pickford F, Coomaraswamy J, Jucker M & McGowan E (2006) Modeling Familial British Dementia in Transgenic Mice. *Brain Pathology* 16: 80–85

Rebelo AP, Abrams AJ, Cottenie E, Horga A, Gonzalez M, Bis DM, Sanchez-Mejias A, Pinto M, Buglo E, Markel K, *et al* (2016) Cryptic Amyloidogenic Elements in the 3′ UTRs of Neurofilament Genes Trigger Axonal Neuropathy. *The American Journal of Human Genetics* 98: 597–614

Rhyu J-M, Park J, Shin B-S, Kim Y-E, Kim E-J, Kim KW & Cho YG (2023) A Novel c.800G>C Variant of the ITM2B Gene in Familial Korean Dementia. *Journal of Alzheimer's Disease* 93: 403–409

Rousseau F, Schymkowitz J & Serrano L (2006) Protein aggregation and amyloidosis: confusion of the kinds? *Curr Opin Struct Biol* 16: 118–126

Sant'Anna R, Braga C, Varejão N, Pimenta KM, Graña-Montes R, Alves A, Cortines J, Cordeiro Y, Ventura S & Foguel D (2014) The Importance of a Gatekeeper Residue on the Aggregation of Transthyretin. *Journal of Biological Chemistry* 289: 28324–28337

Scheller JS, Irvine GW & Stillman MJ (2018) Unravelling the mechanistic details of metal binding to mammalian metallothioneins from stoichiometric, kinetic, and binding affinity data. *Dalton Transactions* 47: 3613–3637

Schlenzig D, Manhart S, Cinar Y, Kleinschmidt M, Hause G, Willbold D, Funke SA, Schilling S & Demuth H-U (2009) Pyroglutamate Formation Influences Solubility and Amyloidogenicity of Amyloid Peptides. *Biochemistry* 48: 7072–7078

Seuma M & Bolognesi B (2022) Understanding and evolving prions by yeast multiplexed assays. *Curr Opin Genet Dev* 75: 101941

Seuma M, Faure AJ, Badia M, Lehner B & Bolognesi B (2021) The genetic landscape for amyloid beta fibril nucleation accurately discriminates familial Alzheimer's disease mutations. *Elife* 10

Seuma M, Lehner B & Bolognesi B (2022) An atlas of amyloid aggregation: the impact of substitutions, insertions, deletions and truncations on amyloid beta fibril nucleation. *Nat Commun* 13: 7084

Sormanni P, Aprile FA & Vendruscolo M (2015) The CamSol Method of Rational Design of Protein Mutants with Enhanced Solubility. *J Mol Biol* 427: 478–490

Tamayev R, Giliberto L, Li W, d'Abramo C, Arancio O, Vidal R & D'Adamio L (2010a) Memory Deficits Due to Familial British Dementia *BRI2* Mutation Are Caused by Loss of *BRI2* Function Rather than Amyloidosis. *The Journal of Neuroscience* 30: 14915–14924

Tamayev R, Matsuda S, Fà M, Arancio O & D'Adamio L (2010b) Danish dementia mice suggest that loss of function and not the amyloid cascade causes synaptic plasticity and memory deficits. *Proceedings of the National Academy of Sciences* 107: 20822–20827

Tesei G & Lindorff-Larsen K (2023) Improved predictions of phase behaviour of intrinsically disordered proteins by tuning the interaction range. *Open Research Europe* 2: 94

Tesei G, Trolle AI, Jonsson N, Betz J, Pesce F, Johansson KE & Lindorff-Larsen K (2023) Conformational ensembles of the human intrinsically disordered proteome: Bridging chain compaction with function and sequence conservation. *bioRxiv*: 2023.05.08.539815

Todd K, Fossati S, Ghiso J & Rostagno A (2014) Mitochondrial dysfunction induced by a post-translationally modified amyloid linked to a familial mutation in an alternative model of neurodegeneration. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* 1842: 2457–2467

Todd K, Ghiso J & Rostagno A (2016) Oxidative stress and mitochondria-mediated cell death mechanisms triggered by the familial Danish dementia ADan amyloid. *Neurobiol Dis* 85: 130–143

Toledano I, Supek F & Lehner B (2023) Genome-scale quantification and prediction of pathogenic stop codon readthrough by small molecules. *bioRxiv*: 2023.08.07.552350

Vaché C, Cubedo N, Mansard L, Sarniguet J, Baux D, Faugère V, Baudoin C, Moclyn M, Touraine R, Lina-Granade G, *et al* (2023) Identification and in vivo functional investigation of a HOMER2 nonstop variant causing hearing loss. *European Journal of Human Genetics* 31: 834–840

Vidal R, Barbeito AG, Miravalle L & Ghetti B (2009) Cerebral Amyloid Angiopathy and Parenchymal Amyloid Deposition in Transgenic Mice Expressing the Danish Mutant Form of Human BRI $_2$. *Brain Pathology* 19: 58–68

Vidal R, Frangione B, Rostagno A, Mead S, Révész T, Plant G & Ghiso J (1999) A stop-codon mutation in the BRI gene associated with familial British dementia. *Nature* 399: 776–781

Vidal R, Révész T, Rostagno A, Kim E, Holton JL, Bek T, Bojsen-Møller M, Braendgaard H, Plant G, Ghiso J, *et al* (2000) A decamer duplication in the 3′ region of the *BRI* gene originates an amyloid peptide that is associated with dementia in a Danish kindred. *Proceedings of the National Academy of Sciences* 97: 4920–4925

Figures Legends

Figure 1. Amyloid nucleation of Bri2, ABri and ADan. A. Schematics of *Bri2 gene* and dementia-related mutations at the nucleotide and protein level. cDNA coordinates are detailed at the top and stop codons (conventional or alternative) are represented with bold fonts and upper case. The exon and its translated protein product is highlighted in purple. Extended products resulting from mutations are highlighted in yellow (ABri and ABri-like) and orange (ADan). Familial dementia mutations are detailed over the arrows. **B.** Relative growth of cells expressing Bri2, ABri or ADan fusions calculated as number of colonies growing in the selective conditions of the amyloid nucleation assay (-URA–Ade) over colonies growing in the absence of selection (-URA) (Chandramowlishwaran *et al*, 2018). Error bars indicate the standard deviation of the experiments (n=3).
C. Deep mutagenesis scheme. A library of sequences of interest is expressed in yeast and selected for amyloid nucleation as reported in (Seuma *et al*, 2021). Amyloid nucleation scores are obtained through deep sequencing of input and output samples.
D. ADan truncations nucleation propensity. The horizontal line indicates the weighted mean nucleation score of Bri2 truncations (up to position 22). Vertical error bars indicate 95% confidence interval of the mean.

Figure 2. Deep mutagenesis of ADan. A. Correlation of nucleation scores between two biological replicates for variants in the ADan library. Pearson correlation coefficient and p-value are indicated. **B.** Correlation between nucleation scores obtained from selection and deep sequencing and relative growth of individual variants in selective conditions (n=5). Vertical and horizontal error bars indicate sigma errors. Pearson correlation coefficient and p-values are indicated. **C.** Distribution of nucleation scores for each class of mutations in ADan library. **D.** Frequency of variants increasing or decreasing nucleation at FDR=0.1 and variants that are WT-like. **E.** Heatmap of nucleation scores for ADan single amino acid substitutions. The WT amino acid and position are indicated in the x-axis and the mutant amino acid is indicated in the y-axis. Synonymous variants are indicated as “ADan”, missense variants due to SNVs are indicated with “\” and SNVs present in GnomAD with a “\\” in the upper right corner of the cell. The distribution of nucleation scores for each position is summarized in the violin plots below the heatmap. The number of variants increasing or decreasing nucleation at FDR=0.1 per position are indicated as stack bars at the bottom. The distribution of nucleation scores for each mutation is summarized in the violin plots at the right-hand side of the heatmap. The number of variants increasing or decreasing nucleation at FDR=0.1 per mutation are indicated as stack bars at the right. **F.** Frequency of single amino acid substitutions that are increasing or decreasing nucleation (FDR=0.1) upon substituting specific WT amino acid.

Figure 3. Deep mutagenesis of ABri. A. Correlation of nucleation scores between two biological replicates for variants in the ABri library. Pearson correlation coefficient and p-value are indicated. **B.** Correlation of nucleation scores obtained from selection

and deep sequencing with relative growth of individual variants in selective conditions (n=5). Vertical and horizontal error bars indicate sigma errors. Pearson correlation coefficient and p-values are indicated. **C.** Heatmap of nucleation scores for ABri single amino acid substitutions. The WT amino acid and position are indicated in the x-axis and the mutant amino acid is indicated in the y-axis. Synonymous variants are indicated as “ABri”, missense variants due to SNVs are indicated with “\” and SNVs present in GnomAD with a “\\” in the upper right corner of the cell. Missing variants have been excluded on the basis of poor sequencing quality and low normalized sigma values (see Methods).

Figure 4. Comparing the mutational effects of single AA variants in ADan, Aβ42 and ABri libraries. A-B. Correlation of average nucleation scores for each type of substitution of the same amino acid (**A**) or to the same amino acid (**B**) in ADan and Aβ42. **C.** Correlation of average nucleation scores per amino acid change between ADan and Aβ42. Points are colored on the basis of the type of change. **D-E.** Correlation of average nucleation scores for each substitution to the same AA (**D**) or of the same AA (**E**) in Adan and ABri. **F.** Correlation of ADan and ABri nucleation scores for positions with more than 4 values in ABri library. ADan residues are detailed in orange and ABri residues in red. Pearson correlation coefficient and p-value are indicated in each plot.

Supplementary Figure 1. ADan and ABri nucleation. A. Inidvidual testing of SupN (negative control), Aβ42 (positive control), Bri2, ABri and ADan in the reporter assay. T-test significance is indicated. **B.** Individually measured growth rates for Bri2, ABri and ADan, in non-inducing (no Cu^{2+}) and inducing (Cu^{2+}) protein expression conditions (n=3 biological replicates/variant). One-way ANOVA with Dunnett’s multiple comparisons test against Bri2 Cu^{2+}. **C.** Normalized sigma distribution for variants increasing, decreasing or not affecting nucleation at FDR=0.1. Vertical line indicates threshold value for considering low sigma. **D.** Correlation of nucleation scores between biological replicates for variats in the ADan library. Pearson correlation coefficient and p-value are indicated. **E.** Correlation of nucleation scores between biological replicates for variants in the ABri library. Pearson correlation coefficient and p-value are indicated.

Supplementary Figure 2. Mutational effects of single AA substitutions. Heatmap of FDR categories for the effects of single amino acid substitutions. The WT amino acid and position are indicated in the x-axis and the mutant amino acid is indicated in the y-axis. Synonymous mutants are indicated with “ADan” (**A**) or “ABri” (**B**). Missense variants due to SNVs present in GnomAD are indicated with “\” in the upper right corner of the cell. The number of variants increasing or decreasing nucleation at different FDRs per position are indicated as stack bars at the bottom.

Supplementary Figure 3. Comparing the mutational effects of single amino acid variants in ADan, Aβ42 and ABri libraries. Correlation of nucleation scores between

ADan and Aβ42 aligned at N-term (**A**), at C-term (**B**) and with ClustalO (**C**), and between ADan and ABri (**D**).

Supplementary Figure 4. Structural insights into Bri2, ABri and ADan. A. Bri2, ABri and ADan structure prediction by Alphafold and MetaESM. Model confidence is colored with a divergent palette: Blue = High, Green = Mid and Red = Low. **B.** Non-ionic energy contact maps of Bri2, ABri and ADan calculated with CALVADOS force-field. Color scale indicates the interaction energy between residues.

Supplementary Figure 5. Physicochemical features partially explain nucleation. A. ADan nucleation scores correlation with amyloid propensity predictors (Tango and Amypred), solubility predictors (CamSol) and Kyte-Doolittle hydrophobicty score. Red dots indicate the values for the WT sequence. Pearson correlation coefficient and p-values are indicated. **B.** Nucleation score distributions arranged by the number of charged residues (y-axis) and the total net charge (x-axis) in ADan. Numbers inside each cell indicate the number of positive and negative residues. The horizontal line indicates WT nucleation score (0).

Supplementary Table 1. SNVs in Bri2 gene reported in GnomAD database.

ID	HGSVc	HGVS Bri2	HGVS ABri	HGVS ADan	AF	NS ABri	NS ADan
rs1425712363	c.730G>A	p.Glu244Lys	p.Glu1Lys	p.Glu1Lys	3.99e-6	-	-1.20
rs1302882644	c.737G>A	p.Ser246Asn	p.Ser3Asn	p.Ser3Asn	3.99e-6	-	0.0448
rs1365902742	c.743G>A	p.Cys248Tyr	p.Cys5Tyr	p.Cys5Tyr	3.99e-6	-	-0.00901
rs765077508	c.747C>G	p.Phe249Leu	p.Phe6Leu	p.Phe6Leu	3.99e-6	-0.109	-0.500
rs578006622	c.748G>T	p.Ala250Ser	p.Ala7Ser	p.Ala7Ser	4.38e-5	0.962	0.0708
rs578006622	c.748G>A	p.Ala250Thr	p.Ala7Thr	p.Ala7Thr	7.97e-6	-	-0.0425
rs139394220	c.751A>G	p.Ile251Val	p.Ile8Val	p.Ile8Val	1.24e-4	1.31	0.0841
rs1272422512	c.755G>A	p.Arg252Gln	p.Arg9Gln	p.Arg9Gln	1.20e-5	-	0.474
rs1380885043	c.769A>G	p.Lys257Glu	p.Lys14Glu	p.Lys14Glu	6.37e-5	-	-0.00624
rs1290822136	c.776C>T	p.Ala259Val	p.Ala16Val	p.Ala16Val	7.97e-6	-	-0.233
rs373968526	c.778G>A	p.Val260Met	p.Val17Met	p.Val17Met	1.59e-5	-	-1.07
rs1428007068	c.787T>G	p.Leu263Val	p.Leu20Val	p.Leu20Val	3.99e-6	-	-1.46
rs747798420	c.788T>C	p.Leu263Ser	p.Leu20Ser	p.Leu20Ser	7.97e-6	1.84	-3.53
rs771676204	c.790A>G	p.Ile264Val	p.Ile21Val	p.Ile21Val	3.99e-6	0.211	-0.0338
rs772731795	c.*3A>G	-	p.Thr25Thr	p.Ser29Gly	3.99e-6	-	-0.160
rs1163491396	c.*8A>G	-	p.Lys27Arg	p.Gln30Gln	3.99e-6	0.343	-
rs760120359	c.*11A>G	-	p.Lys28Arg	p.Glu31Glu	1.20e-5	0.873	-
rs1455609180	c.*13A>C	-	p.Asn29His	p.Lys32Thr	4.00e-6	-	0.234
rs770862229	c.*15C>T	-	p.Asn29Asn	p.His33Tyr	4.00e-6	-	-0.228
rs776617945	c.*18T>C	-	p.Ile30Ile	p. Tyr34His	4.00e-6	-	0.0246
rs759493813	c.*20T>C	-	p.Ile31Thr	p. Tyr34Tyr	7.10e-6	-	-

Figure 1

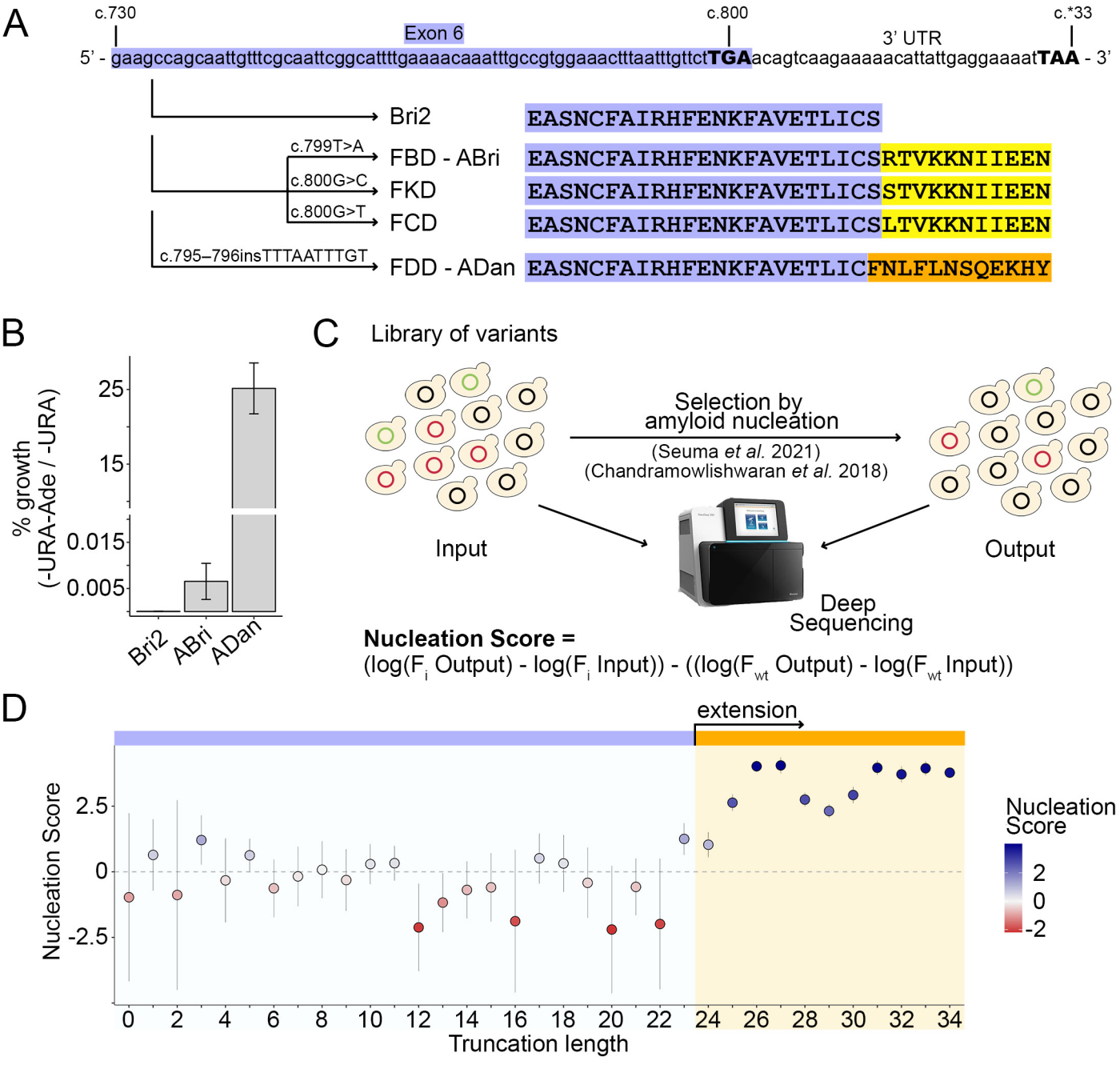

A
c.730
Exon 6
c.800
3' UTR
c.*33
5' - gaagccagcaattgtttcgcaattcggcattttgaaaacaaatttgccgtggaaactttaatttgttctTGAacagtcaagaaaaacattattgaggaaaatTAA - 3'
Bri2
EASNCFAIRHFENKFAVETLICS
c.799T>A
FBD - ABri
EASNCFAIRHFENKFAVETLICSRTVKKNIIEEN
c.800G>C
FKD
EASNCFAIRHFENKFAVETLICSSTVKKNIIEEN
c.800G>T
FCD
EASNCFAIRHFENKFAVETLICSLTVKKNIIEEN
c.795–796insTTTAATTTGT
FDD - ADan
EASNCFAIRHFENKFAVETLICFNLFLNSQEKHY
B
% growth
(-URA-Ade / -URA)
25
15
0.015
0.005
Bri2
ABri
ADan
C
Library of variants
Selection by
amyloid nucleation
(Seuma et al. 2021)
(Chandramowlishwaran et al. 2018)
Input
Output
Deep
Sequencing
Nucleation Score =
(log(F_i Output) - log(F_i Input)) - ((log(F_wt Output) - log(F_wt Input))
D
extension
Nucleation Score
2.5
0
-2.5
0
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32
34
Truncation length
Nucleation
Score
2
0
-2

Figure 2

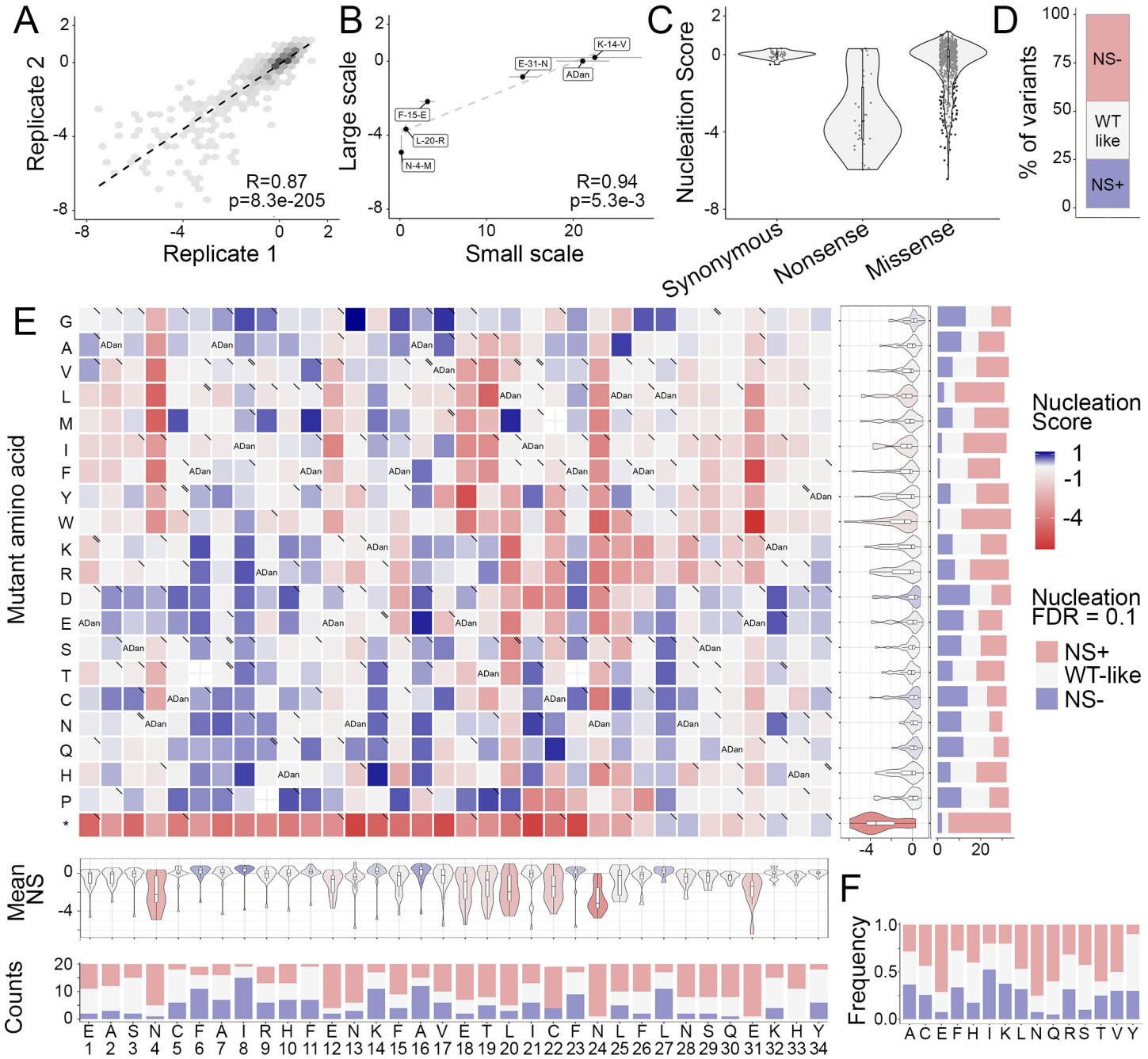

A
Replicate 2
Replicate 1
R=0.87
p=8.3e-205
B
Large scale
Small scale
R=0.94
p=5.3e-3
K-14-V
E-31-N
ADan
F-15-E
L-20-R
N-4-M
C
Nucleaition Score
Synonymous
Nonsense
Missense
D
% of variants
NS-
WT like
NS+
E
Mutant amino acid
Nucleation Score
Nucleation FDR = 0.1
NS+
WT-like
NS-
Mean NS
Counts
F
Frequency

Figure 3

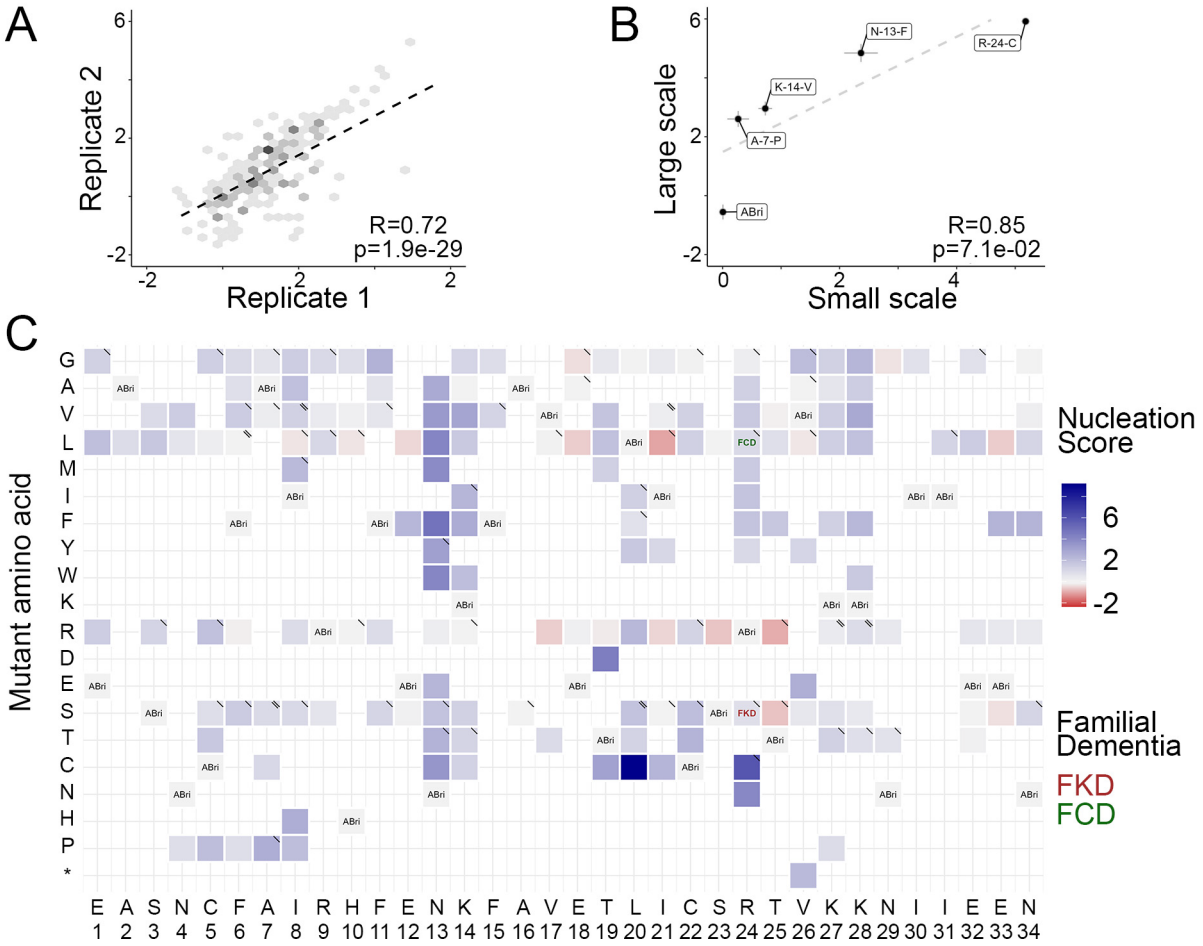
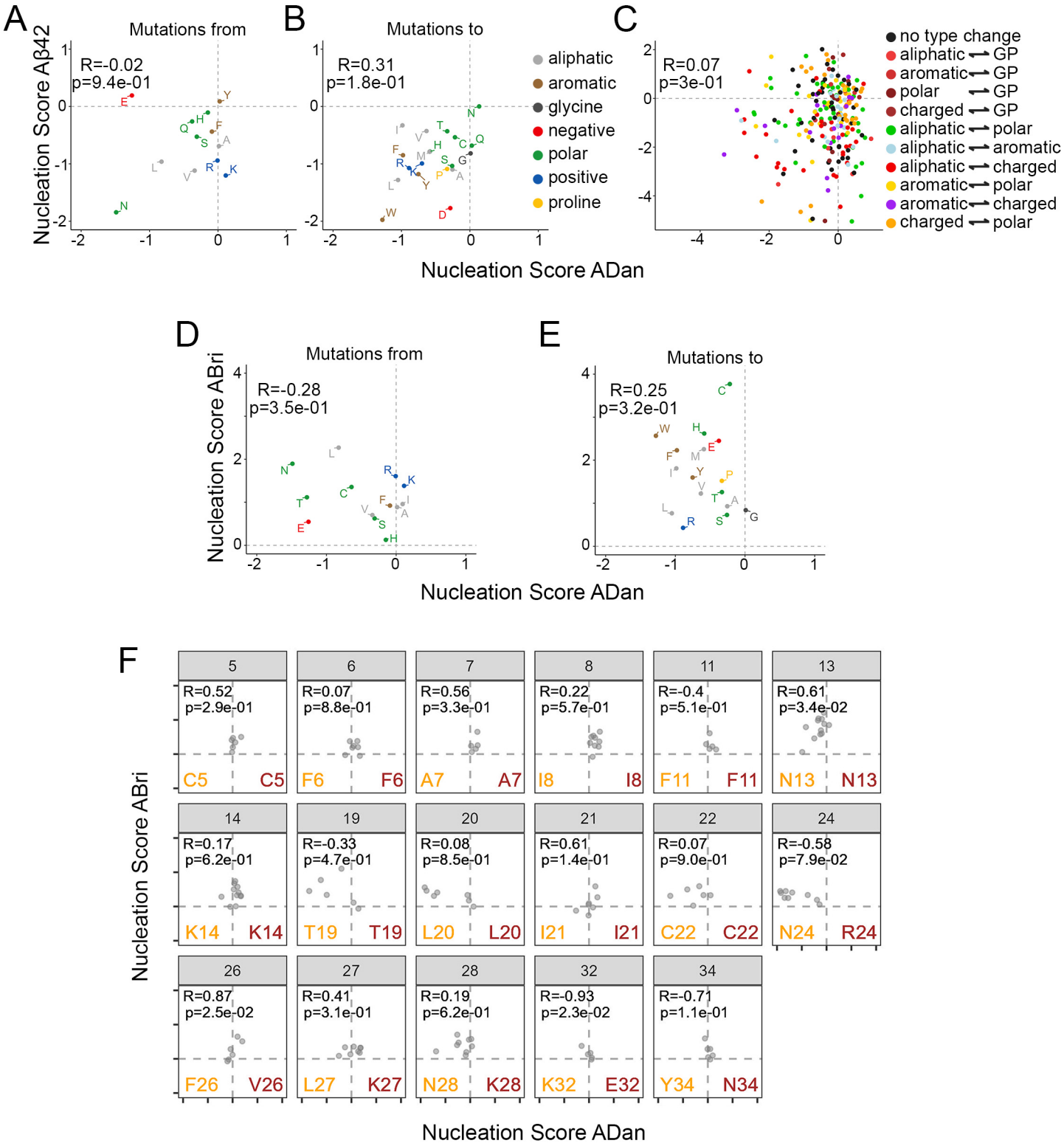

A
Replicate 2
Replicate 1
R=0.72
p=1.9e-29
B
Large scale
Small scale
N-13-F
R-24-C
K-14-V
A-7-P
ABri
R=0.85
p=7.1e-02
C
Mutant amino acid
Nucleation Score
Familial Dementia
FKD
FCD

Figure 4


A
Mutations from
R=-0.02
p=9.4e-01
Nucleation Score Aβ42
B
Mutations to
R=0.31
p=1.8e-01
aliphatic
aromatic
glycine
negative
polar
positive
proline
C
R=0.07
p=3e-01
no type change
aliphatic ⇄ GP
aromatic ⇄ GP
polar ⇄ GP
charged ⇄ GP
aliphatic ⇄ polar
aliphatic ⇄ aromatic
aliphatic ⇄ charged
aromatic ⇄ polar
aromatic ⇄ charged
charged ⇄ polar
Nucleation Score ADan
D
Mutations from
R=-0.28
p=3.5e-01
Nucleation Score ABri
E
Mutations to
R=0.25
p=3.2e-01
Nucleation Score ADan
F
5 R=0.52 p=2.9e-01 C5 C5
6 R=0.07 p=8.8e-01 F6 F6
7 R=0.56 p=3.3e-01 A7 A7
8 R=0.22 p=5.7e-01 I8 I8
11 R=-0.4 p=5.1e-01 F11 F11
13 R=0.61 p=3.4e-02 N13 N13
14 R=0.17 p=6.2e-01 K14 K14
19 R=-0.33 p=4.7e-01 T19 T19
20 R=0.08 p=8.5e-01 L20 L20
21 R=0.61 p=1.4e-01 I21 I21
22 R=0.07 p=9.0e-01 C22 C22
24 R=-0.58 p=7.9e-02 N24 R24
26 R=0.87 p=2.5e-02 F26 V26
27 R=0.41 p=3.1e-01 L27 K27
28 R=0.19 p=6.2e-01 N28 K28
32 R=-0.93 p=2.3e-02 K32 E32
34 R=-0.71 p=1.1e-01 Y34 N34
Nucleation Score ABri
Nucleation Score ADan