🔓 OPEN ACCESS

# Massive mutagenesis reveals an incomplete amyloid motif in Bri2 that turns amyloidogenic upon C-terminal extension

Mariano Martín[a] (ID) and Benedetta Bolognesi[a,1] (ID)

Stop-loss mutations cause over twenty different diseases. The effects of stop-loss mutations can have multiple consequences that are, however, hard to predict. Stop-loss in *ITM2B/BRI2* results in C-terminal extension of the encoded protein and, upon furin cleavage, in the production of two 34 amino acid long peptides, ADan and ABri, that accumulate as amyloids in the brains of patients affected by familial Danish and British Dementia. To systematically explore the consequences of Bri2 C-terminal extension, here, we use a yeast-based massively parallel assay to measure amyloid formation for 676 ADan substitutions and identify the region that forms the putative amyloid core of ADan fibrils, located between positions 20 and 26, where stop-loss occurs. Moreover, we measure amyloid formation for ~18,000 random C-terminal extensions of Bri2 and find that ~32% of these sequences can nucleate amyloids. We find that the amino acid composition of these nucleating sequences varies with peptide length and that short extensions of two specific amino acids (Aliphatics, Aromatics, and Cysteines) are sufficient to generate de novo amyloid cores. Overall, our results show that the C-terminus of Bri2 contains an incomplete amyloid motif that can turn amyloidogenic upon extension. C-terminal extension with de novo formation of amyloid motifs may thus be a widespread pathogenic mechanism resulting from stop-loss, highlighting the importance of determining the impact of these mutations for other sequences across the genome.

amyloid | stop-loss | dementia | deep mutagenesis | random sequences

Frameshift mutations and stop-loss mutations make up ~0.2% of codon-changing mutations (1) and often result in mRNA degradation by the non-stop-mediated-decay (NMD) machinery (2). However, in the presence of an alternative in-frame stop codon within 50 nt from the original one, NMD is not triggered and aberrant proteins with an extended C-terminus can be expressed (1). The C-terminal extension resulting from stop-loss can lead to both loss or gain of protein function (3–10). with over 20 of these mutations proposed as causes of genetic disease (6).

The molecular mechanisms underlying pathogenicity of stop-loss mutations are diverse (3–10). The extended protein products resulting from stop-loss can be depleted co- or posttranslationally (3, 4). Stop-loss mutations in genes *PNPO, HSD3B2*, and *SMAD4* for example produce a novel degron in the extended C-terminal and are targeted to the ubiquitin–proteasome system (5, 6). On the other hand, amorphous aggregates formed by C-terminal extensions of CEP97 are instead cleared by lysosomal degradation (3). Other mechanisms of stop-loss pathogenicity include mislocalization, with C-terminal extensions causing the transcription factor MITF to mislocalize from the nucleus to the cytoplasm and rhodopsin to occupy different layers of the plasma membrane (7, 8).

Finally, the physico-chemical properties encoded by extended sequences can occasionally lead to aggregation of the mutated protein into amyloid fibrils and thus cause or contribute to the pathogenicity induced by stop-loss mutations. Stop-loss mutations in *NEFH* and *REEP1* allow translation of 3′ UTRs that contain amyloidogenic motifs, leading to protein aggregation and neurodegeneration in Charcot–Marie–Tooth disease and peripheral neuropathy, respectively (9, 11). *ITM2B*, also known as *BRI2*, is instead mutated in familial British (FBD) (10) and Danish Dementia (FDD) (12) and also encodes amyloids upon stop-loss. In the last 3 y, two additional stop-loss mutations in *ITM2B/BRI2* have been discovered, leading to Korean (FKD) (13) and Chinese Dementia (FCD) (14). These examples suggest that cryptic amyloid formation could be a more general mechanism of pathogenicity induced by stop-loss and consequent C-terminal extension.

The C-terminus of the transmembrane protein encoded by *ITM2B/BRI2* is cleaved by furin, releasing the 23 amino acid long peptide known as Bri2 (Fig. 1*A*). Upon stop-loss mutations, furin cleavage produces aberrant 34 amino acid long peptides with distinct amino acid sequences in different types of dementia (Fig. 1*B*). In FBD, stop-loss results from a single nucleotide variant (SNV), c.799T>A, and produces a peptide known as Abri (10).

## Significance

Stop-loss mutations that do not trigger the non-stop-mediated-decay machinery can result in proteins with aberrant extended sequences at their C-terminus. At least 20 of these mutations are known to cause human disease. Stop-loss in *ITM2B/BRI2* leads to the production of aberrant peptides that deposit as amyloids and cause different familial forms of dementia. Thanks to deep mutagenesis and random sequence extension, we identify the amyloid core that is generated upon stop-loss in *ITM2B/BRI2* and find that 30% of 18,000 randomly extended sequences form de novo amyloids. We also find that extensions of just two residues are enough to generate novel amyloids, highlighting the importance of measuring, understanding, and predicting the impact of C-terminal extension of protein sequences.
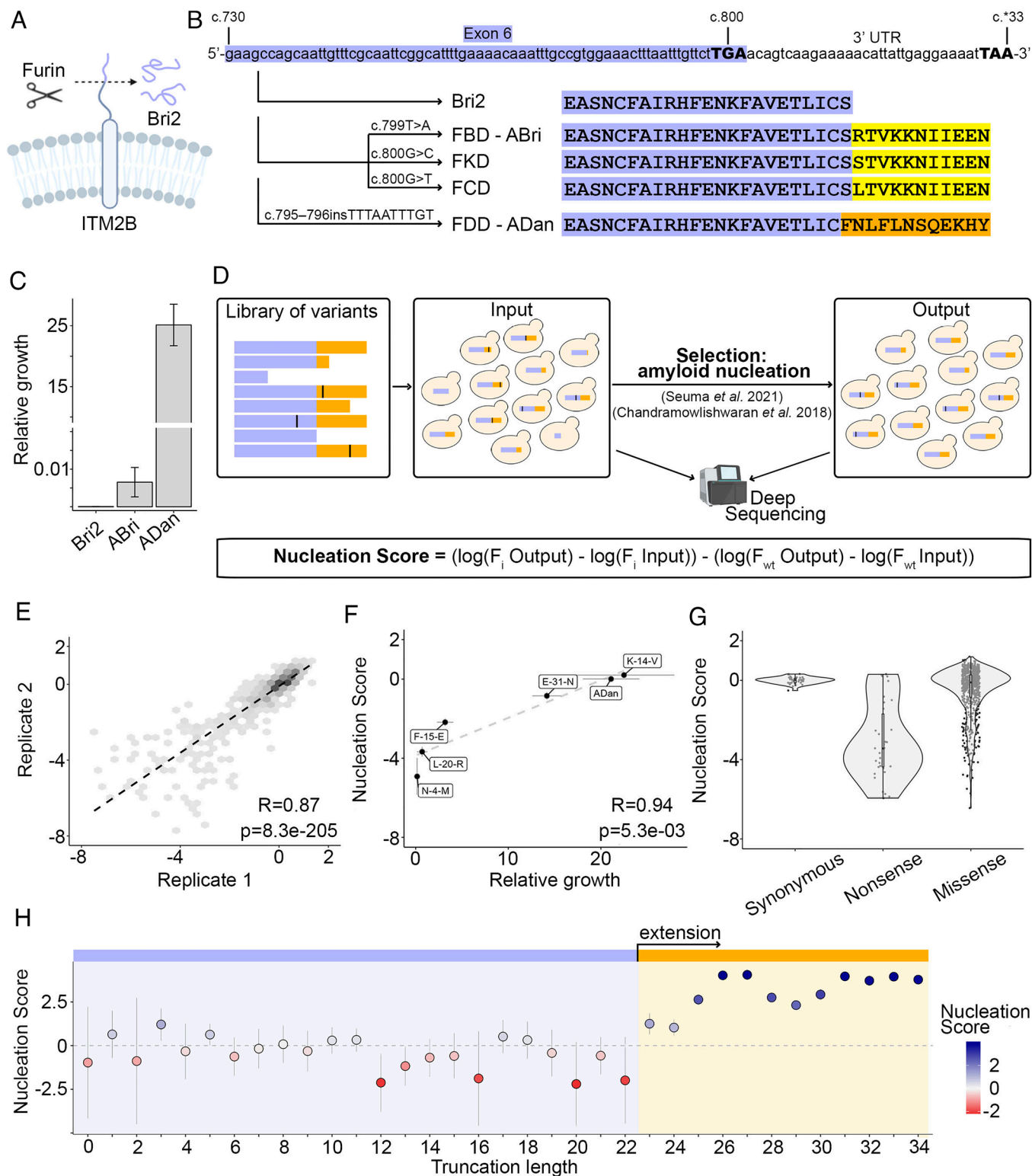
**Fig. 1.** ADan and its truncations are strong nucleators. (*A*) Schematics of ITMB2/BRI2 cleavage by the furin protease. (*B*) Schematics of *ITMB2/BRI2* and dementia-related mutations at the nucleotide and protein level. cDNA coordinates are detailed at the top and stop codons (conventional or alternative) 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. (*C*) 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) (15). Error bars indicate the SD of the experiments (n = 3). (*D*) Deep mutagenesis scheme. A library of sequences of interest is expressed in yeast and selected for amyloid nucleation as reported by Seuma et al. (16). Amyloid nucleation scores are obtained through deep sequencing of input and output samples. (*E*) Correlation of nucleation scores between two biological replicates for variants in the ADan library. Pearson correlation coefficient and *P*-value are indicated. (*F*) Correlation between nucleation scores obtained from selection and deep sequencing with relative growth of individual variants in selective over nonselective conditions (n = 5). Vertical and horizontal error bars indicate estimated sigma errors and SD of the experiments (n = 3), respectively. Pearson correlation coefficient and *P*-values are indicated. (*G*) Distribution of nucleation scores for each class of mutations in ADan library. (*H*) 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% CI of the mean.

FCD and FKD mutations (c.800G>C and c.800G>T, respectively) generate peptides that differ from ABri only in one amino acid at position 24 of the extended peptide (Fig. 1*B*) (13, 14). In FDD, a 10 nt duplication (c.795-796insTTTAATTTGT) that causes frameshifting and stop loss, leads to an aberrant peptide known as ADan, whose last 12 residues are entirely distinct from ABri (Fig. 1*B*) (12). Thus, multiple C-terminal extensions of Bri2 peptide can lead to amyloid formation in the context of autosomal dementia, suggesting that the Bri2 sequence holds an amyloid potential which can be unlocked by specific peptide extensions.

Here, we employ deep mutational scanning (DMS) to scan the impact of all ADan mutations on amyloid nucleation so as to identify the putative amyloid core formed by this peptide. To then explore how likely C-terminal extension in *ITM2B/BRI2* is to generate amyloid cores we use random sequence extension (RSE) and systematically measure the consequences of ~18,000 random C-terminal extensions of Bri2. Our massive dataset reveals that a large fraction (~32%) of these extended peptides nucleate amyloids. Among the strongest nucleators we find C-terminal extensions of just two amino acids. We further find that the novel amyloids display specific amino acid enrichments that depend on sequence length.

Altogether, our results suggest that the C-terminus of Bri2 contains an incomplete amyloid motif that can turn amyloidogenic upon extension. They also demonstrate that massively parallel quantitative assays, such as the one presented here, provide a robust strategy to map amyloid formation from novel areas of the sequence space and that they can generate valuable insights into the general consequences of C-terminal extension arising from mutations, drug-induced readthrough, or aberrant 3′ UTR translation.

## Results

**Quantifying Amyloid Nucleation of Bri2 Extensions.** To measure the ability of Bri2 extensions to nucleate amyloid aggregates we employed a cell-based selection assay where sequences are fused to the nucleation domain (SupN; residues 1 to 123) of the yeast prion Sup35 and amyloid nucleation of the fused peptides seeds the aggregation of endogenous Sup35 (*SI Appendix*, Fig. S1*A*) (15). This process is required for yeast growth in selective conditions and we previously showed that growth rates in this assay correlate with rates of amyloid nucleation in a test tube (16).

Cells expressing the fusion protein SupN-Bri2 do not grow in these conditions, revealing that the unextended peptide does not nucleate amyloid aggregation (Fig. 1*C*). In sharp contrast, the expression of ADan fusion leads to 25.14 ± 3.41% of the cells growing in selection media over nonselective conditions, a twofold increase compared to the growth rate induced in the same conditions by the amyloid-beta peptide (Aβ42), the peptide that forms amyloids in Alzheimer's disease and is mutated in familial forms of the disease (16) (Fig. 1*C* and *SI Appendix*, Fig. S1*B*). Cells expressing ABri fusions 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. 1*C* and *SI Appendix*, Fig. S1*B*). Growth assays of cells expressing Bri2, ABri, or ADan fusions in nonselective conditions revealed that their growth rates were not confounded by differential toxicity (*SI Appendix*, Fig. S1*C*).

**Quantifying the Effect of ADan Substitutions and Truncations on Amyloid Nucleation at Scale.** To further investigate the sequence-nucleation landscape of the ADan peptide we designed and synthesized a library containing all ADan single amino acid substitutions and truncations (n = 680), and expressed it fused to SupN in yeast cells that went under selection as above (*SI Appendix*, Fig. S1*A*). We quantified growth rates and therefore amyloid nucleation for each variant in the library by deep sequencing before and after selection (Fig. 1*D*) (16). Growth rates, henceforth "nucleation scores," from this high-throughput assay were reproducible between replicates (Fig. 1*E* and *SI Appendix*, Fig. S1*D*) and correlate well with the effects of variants quantified individually (Pearson Correlation, R = 0.94, *P*-value: 5.3e-3; Fig. 1*F*). Altogether, we measured amyloid nucleation for 642 single amino acid substitutions (99% of possible amino acid substitutions) and all 34 truncations. In this dataset, synonymous mutations have a nucleation score close to 0, i.e similar to WT ADan, nonsense variants (truncations) have a unimodal distribution with a peak at a nucleation score of ~–4 and the distribution of missense variants (amino acid substitutions) has a strong bias toward reduced amyloid nucleation (Fig. 1*G*). In addition, monitoring the aggregation of synthetic unfused peptides in vitro confirmed faster aggregation kinetics for ADan compared to ABri and Bri2 (*SI Appendix*, Fig. S1*D* and Dataset S6). In vitro, a substantial increase in peptide concentration is necessary to obtain formation of Thioflavin-T binding aggregates for ABri and Bri2 (*SI Appendix*, Fig. S1*E* and Dataset S6).

**A four Amino Acid Extension of Bri2 has the Same Propensity to Nucleate Amyloids than the Full-Length ADan Peptide.** Pathogenic Bri2 extensions share the first 22 residues and differ only in the last 12 amino acids (Fig. 1*B*), suggesting it is the specific extension of ADan that encodes amyloid potential. Measuring the nucleation score of all possible ADan C-terminal truncations in our library, we find that a truncated version of ADan ending at position 25 (equivalent to an extension of Bri2 by F23, N24, and L25) nucleates amyloids significantly faster than Bri2 [Fig. 1*H*; Z-test, false discovery rate (FDR) = 0.01, *P*-adjusted: 1.038e-297, Benjamini & Hochberg correction]. This effect becomes stronger with a peptide of 26 amino acids (Bri2 extension by F23, N24, L25, and F26), leading to a similar nucleation score as that of the full-length ADan peptide (Z-test, FDR = 0.01, *P*-adjusted: 1.544e-02, Benjamini & Hochberg correction). Together with the previous observation that a 28 residue long ADan truncation forms beta-sheet rich aggregates in vitro (17), these findings suggest that the amyloid potential of ADan is encoded by the first residues that get translated upon stop-loss. For two of the variants increasing nucleation, one truncation (L27*) and one substitution (N13G), we obtained synthetic, unfused peptides and confirmed that they form amyloids faster than WT ADan in vitro by monitoring Thioflavin- T fluorescence over time (*SI Appendix*, Fig. S3*C* and Dataset S6).

**Deep Mutagenesis of ADan Uncovers Its Amyloid Core and Residues that Act as Gatekeepers of Amyloid Nucleation.** The distribution of mutational effects for ADan nucleation reveals that 44% of single amino acid substitutions reduce nucleation and only 25% increase it (Z-test, FDR = 0.1, Fig. 2*A*) while 30% of these variants have a nucleation score similar to WT ADan.

Inspecting the heatmap of mutational effects for amino acid changes at all positions in ADan reveals that 26% of the mutations that decrease amyloid nucleation are localized in a short region of 7 amino acids between residues L20-F26 (Fig. 2*B* and *SI Appendix*, Fig. S3*A*). In this continuous stretch of residues most mutations decrease amyloid nucleation (an average of 55% of mutations per position decrease nucleation) and all substitutions to prolines disrupt it (except for L20P that increases nucleation and N24P that is WT-like), suggesting tight structural
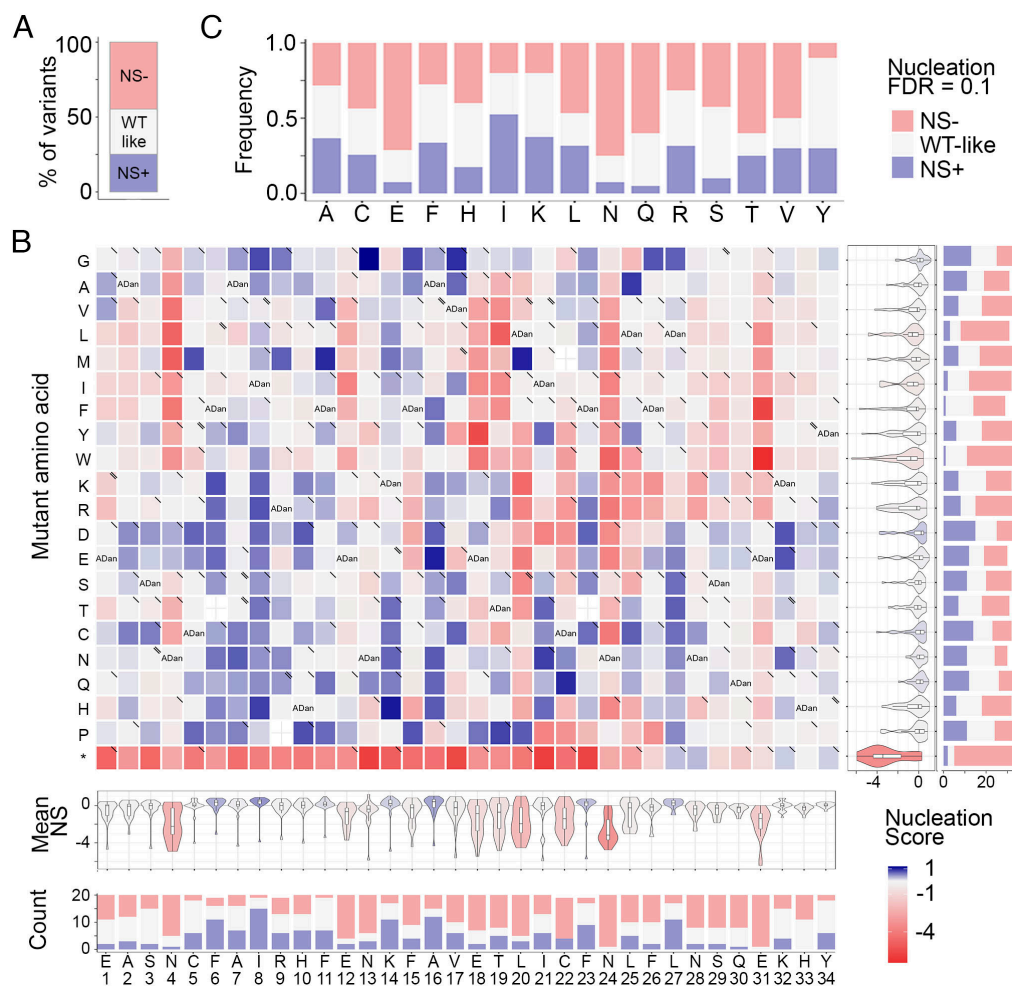
**Fig. 2.** Deep mutagenesis of ADan. (*A*) Frequency of variants increasing or decreasing nucleation at FDR = 0.1 and variants that are WT-like. (*B*) Heatmap of nucleation scores for ADan single amino acid substitutions. The WT amino acid and position are indicated in the *x* axis and the mutant amino 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 stacked 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 stacked bars at the *Right*. Mutation to stop codons are indicated with an "*". (*C*) Frequency of single amino acid substitutions that are increasing or decreasing nucleation (FDR = 0.1) upon substituting specific WT amino acid.

constraints (Fig. 2*B* and *SI Appendix*, Fig. S3*A*). Thus, we propose that this region forms the inner core of the nucleating ADan fibrils and assign a key role in this process to N24 as all mutations (except N24P) at this position lead to a decrease in nucleation. Charges in this stretch are not tolerated, except at position 23 and change F26E. Outside of this region, at positions E18, T19, N4, and E31, mutations to hydrophobic amino acids 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 (75% and 71% of substitutions decrease nucleation, respectively; Fig. 2*C*).

We identify five gatekeepers, positions where mutations are more likely to increase nucleation than decrease it (18–20): 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. 2*B*). Surprisingly, and in contrast to what was previously found for other amyloids (18–20), including Aβ42 (16, 21), four out of five gatekeepers are aliphatic residues except for K14.

We also find that 7 of the reported SNPs in *ITM2B/BRI2* [GnomAD database (22)] decrease nucleation, 10 do not affect nucleation and only 1 of them (rs1272422512), which results in mutation R9Q, increases the nucleation of the ADan peptide (Fig. 2*B* and *SI Appendix*, Table S1). However, we identify another 53 SNVs in ADan that significantly increase nucleation (Fig. 2*B* and *SI Appendix*, Fig. S3*A*).

**Multiple *ITM2B/BRI2* Stop-Loss Mutations Increase Nucleation.** We then employed the same strategy to test whether specific genetic variants in ABri can boost its weak amyloid nucleation propensity. After data processing and filtering (*Methods* and *SI Appendix*, Fig. S2*A*), we found 121 variants that were able to increase nucleation (*SI Appendix*, Figs. S2 and S3*B*). The heatmap of mutational effects for amino acid changes at all positions in ABri reveals that substitutions R24L and R24S increase nucleation (*SI Appendix*, Figs. S2*D* and S3*B*), meaning that the peptides that are a product of FCD and FKD mutations are more prone to aggregate than both Bri2 and ABri. Moreover, the R24C mutation, equivalent to a single stop-loss mutation leading to a Cys codon (c.801A>C: TGA>TGC or c.801A>T: TGA>TGT), significantly increases nucleation (*SI Appendix*, Figs. S2 *B* and *D* and S3*B*).

Also, substitution by Cys at position 20 (L20C) significantly increases nucleation, inducing a relative growth in selection media over nonselective conditions of 63.1 ± 7.1%, that is 2.7 times more than ADan peptide nucleation (23.1 ± 3.6%).

To explore whether the low nucleation scores could be due to differential protein expression, we assessed that variants of both ABri and ADan, with different nucleation scores as quantified from the selection assay, were able to sustain growth without adenine when expressed in a [PIN+] strain, where the rate of nucleation of SupN is not limited by the fused sequence (15) (*SI Appendix*, Fig. S4A and *Methods*). In essence, assessing growth in the lack of adenine for this strain allows us to control that these sequences are expressed and do not lead to degradation of SupN. All of the variants tested are able to sustain growth of [PIN+] strain, regardless of whether they can nucleate. Finally, to evaluate whether nucleation scores could be due to specific interactions, we measured the relative growth of variants of both ABri and ADan when knocking out *Cur1, Btn2, Hsp104, Ssz1,* and *Upf1* (23). *Hsp104* KO inhibited growth for all of the variants tested, as expected as Hsp104 is instrumental for prion fragmentation and propagation (24). While deletion of other anti-prions resulted in changes in growth in selective conditions, these changes were not variant-specific (*SI Appendix*, Fig. S4, Dataset S7, and *Methods*).

### Different Types of Mutations Increase the Nucleation of ADan, ABri, and Aβ42 Amyloids.

As Alzheimer's disease and *ITM2B/BRI2*-related familial dementias share common clinical and neuropathological features, including amyloid depositions (12, 25–27), we evaluated whether the impact of mutations on amyloid nucleation of ADan depends on similar properties as those driving Aβ42 nucleation. There is no significant correlation between the nucleation scores of variants in these peptides when comparing mutations of the same amino acids (Fig. 3A) or mutations to the same amino acid (Fig. 3B). In addition, to compare mutational effects by position, we aligned ADan and Aβ42 peptides in three different ways, at the N-terminal, C-terminal, and with ClustalO (28). Regardless of the strategy, no significant correlation at aligned positions was detected (*SI Appendix*, Fig. S5 A–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. 3 C and D and *SI Appendix*, Fig. S5D). We also tested the correlation in those positions with at least 5 confident measurements in the ABri library, uncovering significant positive correlations only for positions 13 (N), 26 (F and V), and 32 (K and E, ADan and ABri respectively; Fig. 3E).

Surprisingly, changes in hydrophobicity can explain very little of the variance in the ADan dataset (*SI Appendix*, Fig. S6A). Moreover, state-of-the-art amyloid predictor scores poorly correlate with ADan nucleation scores (*SI Appendix*, Fig. S6A) and analysis of the net and total charge of ADan variants gives little insight in the process of ADan amyloid nucleation (*SI Appendix*, Fig. S6B).

### Bri2 Random Extensions Reveal the Properties of De Novo Amyloid Nucleating Peptides.

To better understand the sequence determinants of amyloid core generation upon Bri2 C-terminal extension, we generated three libraries where the Bri2 peptide (22 amino acids) is extended by 12 random degenerate (NNK, where N = A/C/G/T and K = G/T) codons, in a way that is similar to what happens in ADan biogenesis (ADan-like extensions, Fig. 4A). Each library was selected independently, and sequencing was used to quantify the relative amyloid nucleation scores for a total of
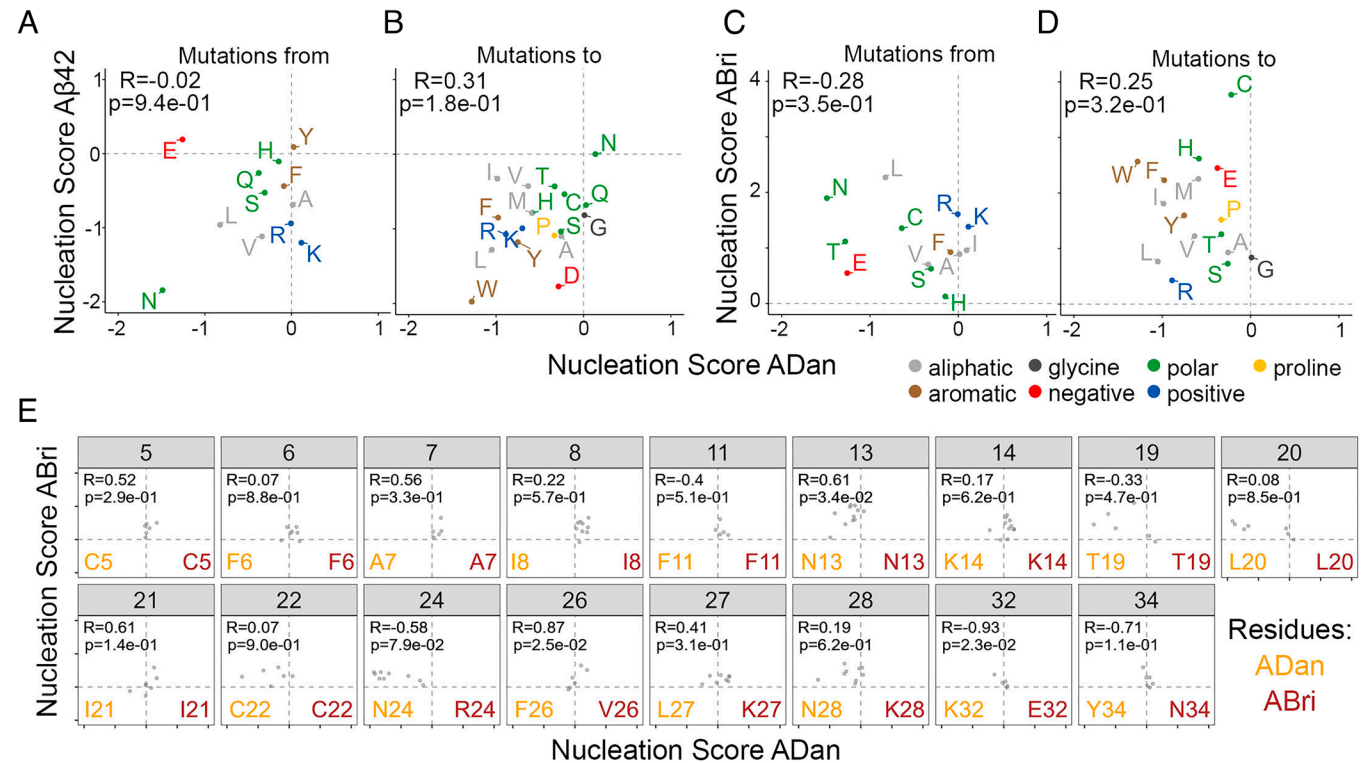


**Fig. 3.** Comparing the mutational effects of single AA variants in ADan, Aβ42, and ABri libraries. (*A* and *B*) Correlation of average nucleation scores for each type of substitution of the same amino acid (*A*) or to the same amino acid (*B*) in ADan and Aβ42. (*C* and *D*) Correlation of average nucleation scores for each substitution to the same amino acid (*C*) or of the same amino acid (*D*) in ADan and ABri. (*E*) Correlation of ADan and ABri nucleation scores for positions where more than four ABri variants were measured. ADan residues are detailed in orange and ABri residues in red. Pearson correlation coefficient and *P*-value are indicated in each plot.

17,952 unique amino acid sequences corresponding to 4.68e-18 of the possible sequence space $(20^{12})$ (Fig. 4B and *SI Appendix,* Fig. S8A). The β-sheet propensity and hydrophobicity, relevant features in amyloid nucleation, of this experimental sequence space overlap with those of the human proteome (Fig. 4D). After data processing and quality control, the vast majority of sequences had a nucleation score that is not significantly different from 0. Consequently, we classified sequences with a nucleation score significantly greater than 0 as nucleators (n = 5,678; 31.6% of the sequences; Z-test, FDR = 0.05), and all other sequences as nonnucleators (n = 12,274; 68.4%; Fig. 4B). The resulting nucleation scores correlate well with the effect of sequences quantified individually (Fig. 4C). In parallel, we also confirmed that two of the strongest novel nucleators identified here, resulting from a 12 (EASNCFAIRHFENKFAVETLICSQLIMIYEDRKG) and a 3 (EASNCFAIRHFENKFAVETLICSIV) amino acid extension respectively, rapidly form amyloids in vitro at 6 and 3 μM concentration (Fig. 4F and Dataset S6).

We examined the differences in amino acid frequency between nucleating and nonnucleating sequences with an extension of 12 amino acids (Fig. 4E). Although the differences were modest, several of these were statistically significant (*t*-test), owing to the large sample size of our data. The most relevant differences are in Asn (*P*-value = 4.28e-56, Cohen's d effect size = 0.30, difference in frequency = 1.50), Cys (*P*-value = 6.69e-110, effect size = 0.43, difference = 2.71), and Arg (*P*-value = 5.23e-30, effect size = 0.22, difference = −1.87; Fig. 4E, *SI Appendix,* Fig. S7C, and Dataset S3). We also analyzed the amino-acid composition position-wise, finding that toward the N-terminus of the extension (close to the end of the Bri2 sequence) nucleators are significantly enriched (chi-squared test) in polar residues (Fig. 4 F and G; min. *P*-value position 24 = 3.53e-11, Cohen's d effect size = 0.25, difference in frequency = 0.056), Cys, Phe, and/ or Val (Fig. 4F, *SI Appendix,* Fig. S7C, and Dataset S4). They are instead depleted in charged residues and aliphatics (Fig. 4 F and G and Dataset S5). The enrichment in polars and Cys is maintained all along the whole extension while the depletion in aliphatics becomes even stronger toward the C-terminus (Fig. 4 F and G and Datasets S4 and S5). In addition, at the C-terminus, nucleators are enriched in Gly (Fig. 4 F and G; min. *P*-value position 33 = 1.54e-5, effect size = 0), aromatics as well as positive and negative charges (Fig. 4 F and G and Datasets S4 and S5).

Among the random C-terminal extensions, 1,824 sequences (~10% of the total) have a Ser residue at position 23, mimicking the effect of stop-loss mutations that lead to production of ABri and effectively generating a subset of ABri-like extensions. We performed the same amino acid enrichment analysis on this subset of extensions, finding similar results as the ones found for all the other random extensions (*SI Appendix,* Fig. S8).

We next evaluated whether common amyloid aggregation predictors are able to predict the nucleation of these sequences and whether they can discriminate between nonnucleators and nucleators (*SI Appendix,* Fig. S9A), or between nonnucleators and the 10% strongest nucleators in each random library (*SI Appendix,* Fig. S9B). All of the predictors show a poor or limited performance on the prediction task (*SI Appendix,* Fig. S9). In both cases CamSol outperforms the rest of the predictors with an AUC of 0.5660±0.0086 for discriminating nonnucleators versus nucleators and an AUC of 0.6482±0.0154 for nonnucleators versus the 10% strongest nucleators.

**Bri2 Extensions of Two Specific Amino Acids Are Sufficient to Nucleate Amyloids.** Our random libraries also included shorter C-terminal extensions due to premature stop codons in the 12 degenerate-codon sequence (Fig. 5A; number of sequences with an extension <12 amino acids: 5,110). We analyzed these sequences and found that the resulting peptide length is not correlated with the strength of amyloid nucleation (Fig. 5B). Moreover, we found that extending Bri2 by only two specific amino acids can create strong nucleating peptides among the 10% strongest nucleators in the dataset (Fig. 5 C and E). We analyzed the position-specific differences between nucleators and nonnucleators composition and found that the amino acid composition of C-terminal extensions with length 8 to 11 amino acids (peptides length of 30 to 33 residues) is similar to that of the 12 amino acid C-terminal extensions (*SI Appendix,* Fig. S10). However, for C-terminal extensions of length <8 amino acids (peptides length of 24 to 29 residues), nucleators exhibit a significant enrichment in aromatics at the N-terminus (Fig. 5D and *SI Appendix,* Fig. S10), revealing that amyloid nucleation can be promoted by distinct types of residues depending on peptide length. In ABri-like extensions (i.e. sequences with a Ser at position 23), we find that a C-terminal extension of two specific amino acids, where at least one of them is aliphatic, creates strong nucleators (Fig. 5E).

## Discussion

*ITM2B/BRI2* mutations affecting its stop codon are associated with familial forms of dementia and deposits containing amyloid forms of the peptides which are generated upon stop-loss have been observed in vivo (10, 12, 31). Here, we systematically addressed how sequences resulting from stop-loss in *ITM2B/BRI2* can nucleate amyloids using deep and random mutagenesis combined to a selection assay that reports on the amyloid nucleation of thousands of sequences of different length in parallel.

The ADan peptide very rapidly forms amyloids in this cell-based assay, two-fold faster than Aβ42 for reference, but ABri is at the lower end of the detection capability of our assay. While this sharp difference could be due to limitations in this specific assay, it is in line with results in transgenic mice and drosophila models expressing ABri, which do not exhibit amyloid deposits (32) and feature low neurotoxicity, in contrast with those expressing Adan (31, 33).

We identify two important caveats in our approach: 1) the cytosolic expression of sequences that are found in the secretory pathway and 2) the potential impact of the SupN fusion on nucleation and the resulting phenotype. We cannot completely exclude that specific variants in our libraries may feature reduced amyloid nucleation in the intracellular context or as a result of cellular toxicity or susceptibility to anti-prions. Small-scale in vitro validation as well as testing individual variants for their expression, toxicity, and interactions suggests this is not the case. Previous studies using the same reporter (15, 21) also showed that the ability to promote yeast growth in selective conditions was not dependent on toxicity of the fused protein. The in vitro fibril formation experiments also contribute to address the second caveat as peptides in these experiments are not fused to SupN. However, we cannot exclude that the interaction of the fused sequences with SupN may impact nucleation and the final phenotypic outcome in the yeast assay, at least for some of the variants.

We identify a continuous stretch of residues (L20-F26) in ADan which are crucial for amyloid nucleation and likely constitute the core of ADan amyloids. This identification of a putative amyloid core involving residues just after the Bri2 stop codon further suggests that amyloid nucleation in the assay is driven by specific sequence extensions as SupN-Bri2 fusions do not result in the ability to grow in selective conditions. Surprisingly, changes in hydrophobicity can explain very little of the variance in the ADan dataset (*SI Appendix,* Fig. S6). It would have been impossible to predict the mutational impact in ADan on the basis of simple
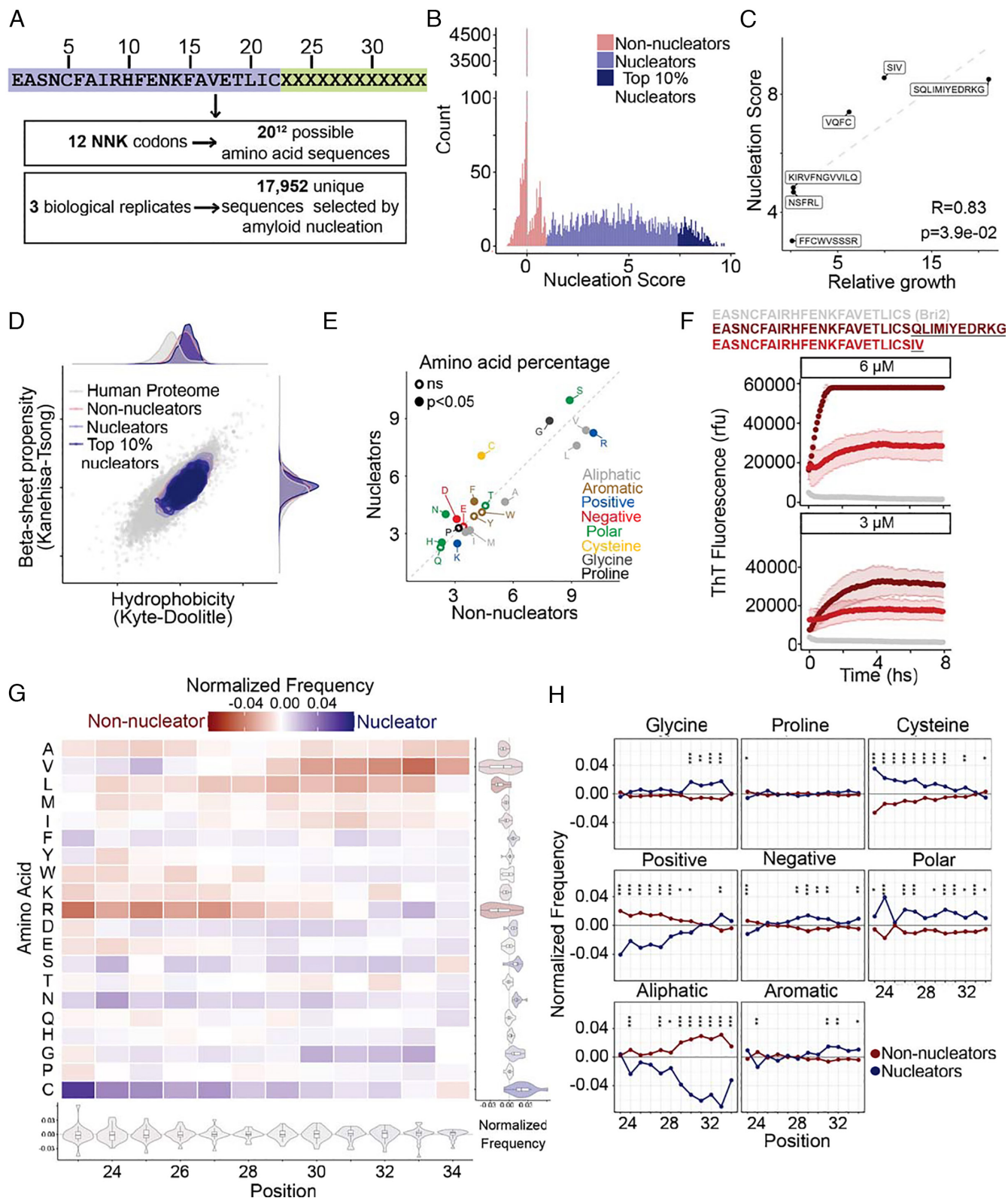
**Fig. 4.** Amyloid nucleation by Bri2 random extensions of 12 amino acids. (*A*) Random extensions design and sequence space. A library of random sequences encoded by 12 degenerated NNK codons was expressed in yeast and selected for amyloid nucleation. (*B*) Nucleation scores distribution of random sequences present in replicate 1. Variants are classified as nonnucleators (red), nucleators (blue), or top 10% nucleators (dark blue). (*C*) Correlation between nucleation scores obtained from selection and deep sequencing with relative growth of individual variants in selective over nonselective conditions (n = 6). Pearson correlation coefficient and *P*-values are indicated. (*D*) Hydrophobicity [Kyte-Doolittle (29)] and beta-sheet propensity [Kanehisa-Tsong (30)] of assayed sequences relative to human proteome. (*E*) The percent composition of residues grouped by their physicochemical properties in nucleators and nonnucleators sequences. (*F*) The aggregation of Bri2 and Bri2 extensions EASNCFAIRHFENKFAVETLICSQLIMIYEDRKG and EASNCFAIRHFENKFAVETLICSIV was followed using a continuous ThT binding assay. The concentration of the peptides is detailed in the upper part of each panel. Error bars indicate the SD of the replicates (n = 3). (*G*) Heatmap of normalized frequencies for Bri2 random extensions. The distribution of normalized frequencies for each position is summarized in the violin plots below the heatmap and the distribution of normalized frequencies for each mutation is summarized in the violin plots at the right-hand side of the heatmap. (*H*) The position-specific differences in amino acid type frequencies across nucleating and nonnucleating sequences. Asterisks indicate marginal *P*-value (chi-square test). *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.
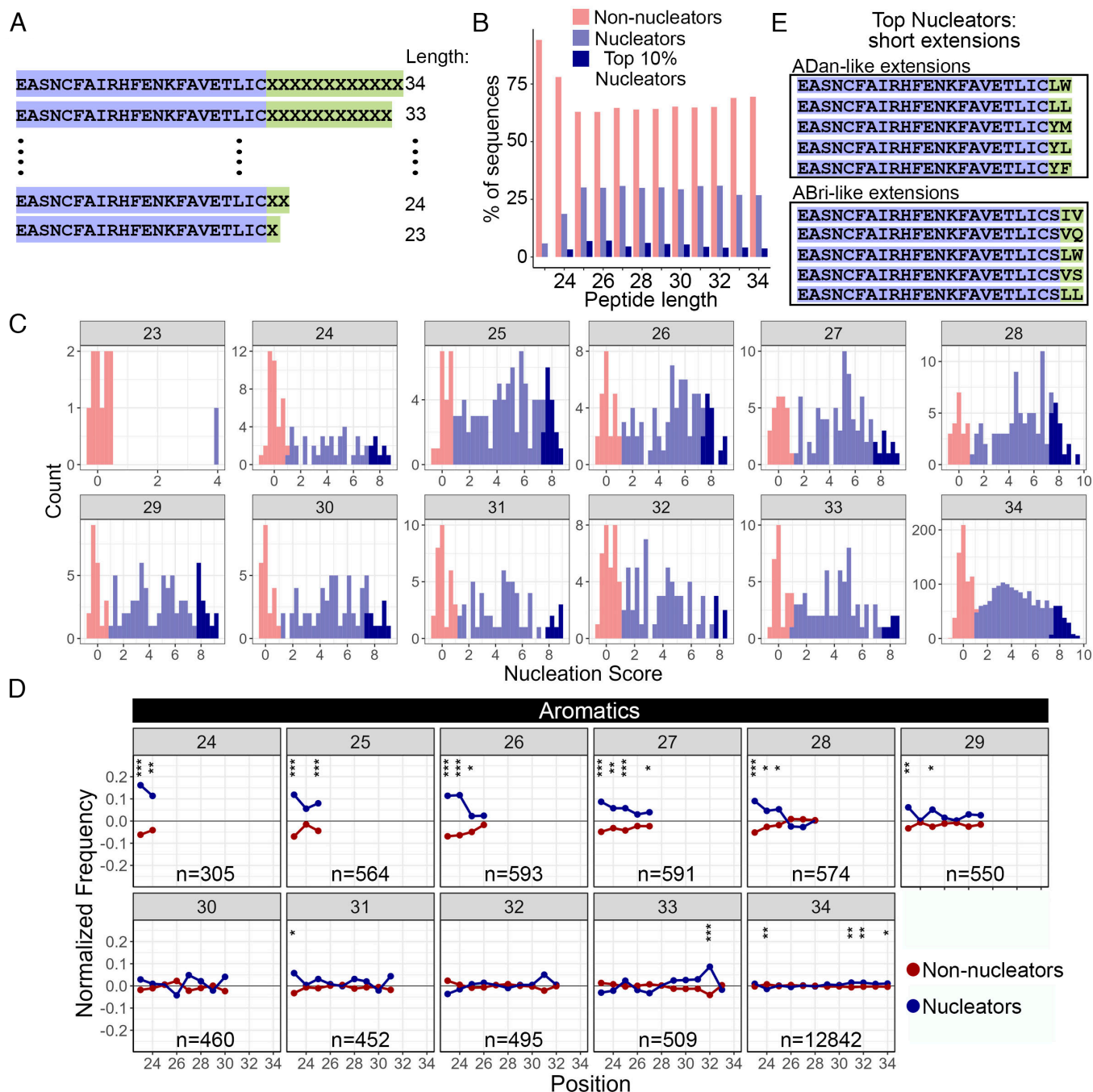
**Fig. 5.** Nucleators resulting from Bri2 random extensions of different lengths have different specific amino acid enrichment. (*A*) Range of lengths for the Bri2 random extensions measured in the assay. (*B*) Percentage of nonnucleators, nucleators, and top 10% nucleators sequences for each peptide length. (*C*) Nucleation scores distribution of random extended sequences present in replicate 1. The length of the peptide is indicated inside the gray box of each plot. Variants are classified as nonnucleators (red), nucleators (blue), or top 10% nucleators (dark blue). (*D*) The position-specific differences in aromatic amino acids frequencies across nucleating and nonnucleating sequences in truncations of different length. The length of the peptide is indicated inside the gray box of each plot. Asterisks indicate marginal *P*-value (chi-square test). \*$P < 0.05$; \*\*$P < 0.01$; \*\*\*$P < 0.001$. (*E*) Top nucleating sequences of ADan-like extensions (*Top*) and ABri-like extensions (*Bottom*) of two amino acids. Bri2 sequence is highlighted in purple and the extended sequence in green.

physicochemical properties, aggregation predictors, or even the latest generation of variant effect predictors (34). What is more, it would have also been impossible to predict mutational impact in ADan on the basis of the complete mutational landscape of another extracellular amyloid peptide of similar length, Aβ42 (21), suggesting that our understanding of the process of amyloid formation by novel peptides is still data limited and highlighting the need of systematically measuring amyloid nucleation for these and other peptides that form amyloids and lead to neurodegeneration (35, 36).

By employing unbiased random sequence extension, we expanded the probed sequence space both in terms of length and identity, quantifying amyloid nucleation for ~18,000 sequences. To our knowledge, this is the largest dataset of protein sequence extensions currently available. ~32% of them are able to nucleate amyloids. This dataset also reveals that a C-terminal amino acid extension of just two residues is enough to significantly increase the propensity of Bri2 to nucleate amyloids. Other short C-terminal extensions (peptide length < 34 amino acids) also generate strong nucleators. These results are in agreement with a

previous study showing that an ADan truncation at position 28 can form beta-sheet rich aggregates in vitro (17). They also suggest that an "incomplete amyloid motif" exists at the end of the Bri2 sequence, and that specific C-terminal extensions upon stop loss or due to insertions before the stop codon can turn it into an effective driver of amyloid nucleation.

Cryptic amyloid motifs have been discovered in the 3' UTRs of at least three genes that are mutated in neurodegenerative disease (9–11). Our datasets can now contribute to build better models that can identify and predict more of these motifs. In addition, while stop-codon readthrough of TGA codons happens at a low frequency of 0.01 to 0.1% (37, 38) this phenomenon can be promoted by small molecules (39) such as Gentamicin or ELX-02, which are under clinical trial for the treatment of Herlitz junctional epidermolysis bullosa and Cystic Fibrosis, respectively (clinical trials: NCT04140786 and NCT04135495). Along this line, the possibility of quantifying the impact of C-terminal sequence extensions provides a systematic way to uncover incomplete or hidden amyloids across the genome, gaining the ability to preventively interpret the impact of stop-codon readthrough on de novo amyloid formation.

## Methods

**Library Design.** The designed libraries contain a total of 1,088 unique ADan nucleotide variants (680 single protein variants encoded by 1, 2, and 3 nucleotide changes) and 1,088 unique ABri nucleotides variants (680 single protein variants encoded by 1, 2, and 3 nucleotide changes). The random libraries were designed with the sequence encoding for the Bri2 peptide (22 amino acids) at the N-terminus followed by random extensions of 12 amino acids encoded by degenerate codons (NNK, where N = A/C/G/T and K = G/T).

**Plasmid Library Construction.** Libraries were synthesized by Integrated DNA Technologies (IDT) as oligo pools covering the ADan and/or ABri 102 nt sequence or as an ultramer of 12 NNK codons for the random libraries (36 nucleotides), flanked by 25 nt upstream and 21 nt downstream constant regions for cloning. 2 μL of 10 μM library pool were extended into a double-stranded DNA by a single cycle PCR (Q5 high-fidelity DNA polymerase, NEB) with primers annealing to the constant regions (primers MM_01, Dataset S2). The product was purified from a 2% agarose gel (QIAquick Gel Extraction Kit, Qiagen). In parallel, the pCUP1-SupN plasmid was linearized by PCR (Q5 high-fidelity DNA polymerase, NEB; primers MM_02-04, Dataset S2). 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 at 50 °C 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 *Escherichia coli* (NEB), by electroporation at 2.0 kV, 200 Ω, 25 μF (BioRad GenePulser machine). For the random libraries, the electroporation was performed three times independently to obtain three random libraries. 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 >1 M transformants were estimated, meaning that each variant in the ADan and ABri libraries are represented ~1,000 times. For the three random libraries, ~5 M, ~8 M and ~2 M transformants were obtained, respectively. 5 mL of overnight cultures were harvested to purify the ADan and ABri libraries with a mini prep (QIAprep Miniprep Kit, Qiagen). 50 mL of overnight culture were harvested to purify each random library with a midi prep (Plasmid MIDI Kit, Qiagen).

**Yeast Transformation.** *Saccharomyces cerevisiae* [psi-pin-] (MATα ade1–14 his3 leu2-3,112 lys2 trp1 ura3–52) provided by the Chernoff lab (15) was used in all experiments in this study. Yeast cells were transformed with the ADan and ABri 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 to 5 h. When cells reached the exponential phase (OD ~ 0.8 to 0.9) cells were harvested at 3,000 × g for 5 min, washed with 50 mL milliQ, centrifuged at 3,000 × 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 μL 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 μL 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 (SC-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 to 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.

**Large-Scale Yeast Transformation of Random Libraries.** Yeast cells were transformed with the plasmid random libraries midi prep in three independent experiments. For each of the experiments, an overnight pregrowth culture in 25 mL of YPDA medium at 30 °C was diluted to OD600 = 0.3 in 175 mL YPDA and incubated at 30 °C 200 rpm for ~4 h. When cells reached the exponential phase, they were harvested, washed with milliQ, and resuspended in sorbitol mixture (100 mM LiOAc, 10 mM Tris pH 8, 1 mM EDTA, 1M sorbitol). After a 30 min incubation at room temperature (RT), 4 μg of plasmid library and 175 μL of ssDNA (UltraPure, Thermo Scientific) were added to the cells. YTB-PEG mixture (100 mM LiOAc, 10 mM Tris pH 8, 1 mM EDTA pH 8, 40% PEG3350) was also added and cells were incubated for 30 min at RT and heat-shocked for 15 min at 42 °C in a water bath. Cells were harvested, washed, resuspended in 250 mL recovery medium (YPD, sorbitol 0.5M, 70 mg/L adenine) and incubated for 1 h at 30 °C 200 rpm. After recovery, cells were resuspended in 350 mL SC-URA plasmid selection medium and allowed to grow for 50 h. Transformation efficiency was calculated for each tube of transformation by plating an aliquot of cells in −URA plates. Each of the three random libraries was bottlenecked to ~500,000 yeast transformants. Two days after transformation, the culture was diluted to OD600 = 0.08 in 500 mL of SC-URA medium and grown until exponential phase. At this stage, cells were harvested and stored at −80 °C in 25% glycerol.

**Selection Experiments.** In vivo selection assays were performed in three independent biological replicates for ADan and ABri libraries. For each 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 (SC-URA, 2% glucose, 100 μM Cu$_2$SO$_4$) at OD = 0.1. After 24 h the 40 mL input pellets were collected, and cells were plated on SC-ADE-URA selection medium in 145-cm$^2$ plates (Nunc, Thermo Scientific). Plates were incubated at 30 °C for 7 d. Finally, colonies were scraped off the plates with PBS 1× and harvested by centrifugation to collect the output pellets. Both input and output pellets were stored at −20 °C before DNA extraction. Three input and three output samples were processed for sequencing. For each of the three random libraries, one selection experiment was performed, and one input sample and three technical replicates of the output pellet were processed for sequencing as described above.

**DNA Extraction and Sequencing Library preparation.** Input and output pellets were resuspended in 0.5 mL extraction buffer (2% Triton-X, 1% SDS, 100 mM NaCl, 10 mM Tris-HCl pH 8, 1 mM EDTA pH 8). They were then frozen for 10 min in an ethanol-dry ice bath and heated for 10 min at 62 °C. This cycle was repeated twice. 0.5 mL 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 2,000 × 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 2,000 × 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 1× 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-SupN plasmid at 58 °C for 40 cycles (primers MM_05-06, Dataset S2). 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 (ADan and ABri libraries) or 300 million molecules (random libraries) were amplified for 15 cycles with frame-shifted primers with homology to Illumina sequencing primers (primers MM_07-26, Dataset S2). The products were treated with ExoSAP treatment (Affymetrix) and purified by column purification (MinElute PCR Purification Kit, Qiagen). They were then amplified for 12 cycles in PCR2 with Illumina-indexed primers (primers MM_27-67, Dataset S2). For ADan and ABri libraries, the six samples (one input–output pair per biological 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 three random libraries were purified individually. The libraries were sent for 125 bp paired-end sequencing in an Illumina HiSeq2500 sequencer at the CRG Genomics core facility. In total, >10 million paired-end reads were obtained for each of ADan and ABri libraries, representing >1,000× read coverage. Each of the random libraries obtained >40 million paired-end reads.

**Individual Variant Testing.** Selected variants were obtained via PCR mutagenesis or ultramer amplification followed by Gibson assembly. Plasmids were transformed into *E. coli* and then yeast, with mutations verified by Sanger sequencing. Yeast expressing each variant were grown in induction media (SC-URA 2% glucose 100 µM Cu$_2$SO$_4$), then plated on control (SC-URA) and selective (SC-URA-ADE) plates for colony counting. For toxicity measurements, cells were grown in non-inducing and inducing protein expression media. Growth rates were measured in a microplate reader and data was analyzed using GrowthCurver in R.

**Thioflavin T Binding Assay.** Synthetic purified peptides were purchased from Bachem and Genescript as TFA salts. The peptides were resuspended in TFA, sonicated, frozen, and lyophilized. After HFIP treatment, peptides were stored at −80 °C. For the Thioflavin-T (ThT) assay, peptides were resuspended in NaP 50 mM pH 7.4 buffer with 20 µM ThT. Fluorescence was measured at 480 nm every 5 min at 29 °C. Peptide concentrations were determined by acid hydrolysis and amino acid analysis by the Separative Techniques Unit of the Cientific and Technological Centers of the University of Barcelona CCiTUB.

**Spotting Nucleation Assay.** Yeast strains [psi-pin−] and [psi-PIN+] were transformed with plasmids expressing SupN fused to various Bri2, ABri, and ADan variants. Yeast was grown in selective and induction media before serial dilutions were plated on control and selection plates. Growth was assessed after 7 d at 30 °C, with relative growth calculated as the percentage of colonies on selection plates compared to control plates.

**Generation of Strains Defective in Anti-Prion Systems.** Anti-prions knock-out strains were built using the tool-box system (40). KanamycinMax resistance was amplified by PCR (primers MM_92-101, Dataset S2) with flanking regions homologous to the gene of interest. The purified PCR product was transformed into yeast GT409 strain, plated in YPDA agar and incubated at 30 °C ON. The YPDA agar plate was replicated in a YPDA Kanamycin plate and incubated at 30 °C for 48 h. Eight isolated colonies (for each of the KO genes) were restreaked in YPDA Kanamycin and incubated at 30 °C ON. The correct genome editing was confirmed by colony-PCR (primers MM_102-107, Dataset S2) for each of the clones.

**Data Processing.** FastQ files from paired end sequencing of the libraries were processed using DiMSum (https://github.com/lehner-lab/DiMSum) (41), 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. Nondesigned variants were also discarded for further analysis, as well as variants with fewer than 100 (random libraries) or 200 (ADan and ABri libraries) input reads in all of the replicates.

**Nucleation Scores and Error Estimates.** The DiMSum package (https://github.com/lehner-lab/DiMSum) (41) 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 Dataset S1.

The 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. As this library has a poor signal/noise ratio (*SI Appendix*, Fig. S2A), few variants passed the filtering process. NS was obtained for 676 unique ADan variants and 679 unique ABri variants. 664 confident estimates (with low normalized sigma value) were obtained for the ADan library and 177 in the ABri library.

For the random extensions libraries, 1 input and 3 output sequencing results were used. DimSum was run using an arbitrary WT sequence chosen among those with high input counts. The sequences used were TTGTAGGTTGCTCAGGAGGTTAATACGTATACGTAG, TTTTGTTAGTTTGATGTGCATGAGTGTCTGTAGTGT, and TGGAAGGGGACGATGGTTGTGGGTAGTAATTGGCCG for experiment 1, 2, and 3, respectively.

NSs for each sequence were merged within each replicate using error-weighted mean and centered to the mode value of the distribution (as most of the sequences do not nucleate). Those sequences that are present in the input and not in the output were imputed with the mode value of the distribution. Most of the sequences have a nucleation score of 0. Consequently, we classified sequences with a nucleation score significantly (FDR = 0.05) greater than 0 as nucleators and all other sequences as nonnucleators. Duplicated sequences were merged and mean nucleation score and the mode of the nucleator status was calculated. We discarded those nonnucleators sequences that had big errors and a mean nucleation score above the minimum nucleator score. For each of the random libraries we classified the top 10% nucleator sequences as another group. Finally, variants from the three experiments were merged and duplicated variants were treated as previously described. 17,952 unique sequences with NS were obtained (7,783 experiment 1; 5,562 experiment 2 and 4,607 experiment 3). All NS estimates are available in Dataset S1.

**Aggregation Predictors.** For the aggregation predictors (Tango, Amypred, Camsol, Aggrescan) (42–45) 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 (29).

**Statistics and Reproducibility.** Based on the transformation efficiency, each variant in the designed libraries (n = 1,088 nucleotide variants each) is expected to be represented at least 10× 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) (41). The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

**Data, Materials, and Software Availability.** Raw sequencing reads and processed nucleation scores are deposited in NCBI's Gene Expression Omnibus (GEO) as GSE244612 (46) and GSE270792 (47). The processed reads and nucleation scores are also available as Dataset S1. All scripts used for downstream analysis and to reproduce all figures are in https://github.com/BEBlab/ADan-BriNNK.

1. S. E. Hamby, N. S. Thomas, D. N. Cooper, N. Chuzhanova, A meta-analysis of single base-pair substitutions in translational termination codons ('nonstop' mutations) that cause human inherited disease. *Hum. Genomics* **5**, 241 (2011).
2. L. Monaghan, D. Longman, J. F. Cáceres, Translation-coupled mRNA quality control mechanisms. *EMBO J.* **42**, e114378 (2023), 10.15252/embj.2023114378.
3. L. Kramarski, E. Arbely, Translational read-through promotes aggregation and shapes stop codon identity. *Nucleic Acids Res.* **48**, 3747–3760 (2020).
4. J. A. Arribere et al., Translation readthrough mitigation. *Nature* **534**, 719–723 (2016).
5. S. Dhamija et al., A pan-cancer analysis reveals nonstop extension mutations causing SMAD4 tumour suppressor degradation. *Nat. Cell Biol.* **22**, 999–1010 (2020).
6. N. Shibata et al., Degradation of stop codon read-through mutant proteins via the ubiquitin-proteasome system causes hereditary disorders. *J. Biol. Chem.* **290**, 28428–28437 (2015).
7. J. Sun et al., Functional analysis of a nonstop mutation in MITF gene identified in a patient with Waardenburg syndrome type 2. *J. Hum. Genet.* **62**, 703–709 (2017).
8. T. J. Hollingsworth, A. K. Gross, The severe autosomal dominant retinitis pigmentosa rhodopsin mutant Ter349Glu mislocalizes and induces rapid rod cell death. *J. Biol. Chem.* **288**, 29047–29055 (2013).
9. A. P. Rebelo et al., Cryptic amyloidogenic elements in the 3′ UTRs of neurofilament genes trigger axonal neuropathy. *Am. J. Hum. Genet.* **98**, 597–614 (2016).
10. R. Vidal et al., A stop-codon mutation in the BRI gene associated with familial British dementia. *Nature* **399**, 776–781 (1999).
11. A. S. Bock et al., A nonstop variant in *REEP1* causes peripheral neuropathy by unmasking a 3′UTR-encoded, aggregation-inducing motif. *Hum. Mutat.* **39**, 193–196 (2018).
12. R. Vidal et al., A decamer duplication in the 3′ region of the *BRI* gene originates an amyloid peptide that is associated with dementia in a Danish kindred. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 4920–4925 (2000).
13. J.-M. Rhyu et al., A novel c.800G>C variant of the ITM2B gene in familial Korean dementia. *J. Alzheimers Dis.* **93**, 403–409 (2023).
14. X. Liu et al., A novel ITM2B mutation associated with familial Chinese dementia. *J. Alzheimers Dis.* **81**, 499–505 (2021).
15. P. Chandramowlishwaran et al., Mammalian amyloidogenic proteins promote prion nucleation in yeast. *J. Biol. Chem.* **293**, 3436–3450 (2018).
16. M. Seuma, A. J. Faure, M. Badia, B. Lehner, B. Bolognesi, The genetic landscape for amyloid beta fibril nucleation accurately discriminates familial Alzheimer's disease mutations. *Elife* **10**, e63364 (2021).
17. K. Todd, J. Ghiso, A. Rostagno, Oxidative stress and mitochondria-mediated cell death mechanisms triggered by the familial Danish dementia ADan amyloid. *Neurobiol. Dis.* **85**, 130–143 (2016).
18. F. Rousseau, J. Schymkowitz, L. Serrano, Protein aggregation and amyloidosis: Confusion of the kinds? *Curr. Opin. Struct. Biol.* **16**, 118–126 (2006).
19. S. S. Bhoite, D. Kolli, M. A. Gomulinski, M. R. Chapman, Electrostatic interactions mediate the nucleation and growth of a bacterial functional amyloid. *Front. Mol. Biosci.* **10**, 1070521 (2023).
20. R. Sant'Anna et al., The importance of a gatekeeper residue on the aggregation of transthyretin. *J. Biol. Chem.* **289**, 28324–28337 (2014).
21. M. Seuma, B. Lehner, B. Bolognesi, An atlas of amyloid aggregation: The impact of substitutions, insertions, deletions and truncations on amyloid beta fibril nucleation. *Nat. Commun.* **13**, 7084 (2022).
22. S. Chen et al., A genomic mutational constraint map using variation in 76,156 human genomes. *Nature* **625**, 92–100 (2024).
23. M. Son, R. B. Wickner, Antiprion systems in yeast cooperate to cure or prevent the generation of nearly all [PSI+] and [URE3] prions. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2205500119 (2022).
24. Y. O. Chernoff, S. L. Lindquist, B. Ono, S. G. Inge-Vechtomov, S. W. Liebman, Role of the chaperone protein Hsp104 in propagation of the yeast prion-like factor [psi+]. *Science* **268**, 880–884 (1995).
25. J. L. Holton et al., Regional distribution of amyloid-Bri deposition and its association with neurofibrillary degeneration in familial British dementia. *Am. J. Pathol.* **158**, 515–526 (2001).
26. J. A. Ghiso et al., Systemic amyloid deposits in familial British dementia. *J. Biol. Chem.* **276**, 43909–43914 (2001).
27. T. Lashley et al., Molecular chaperons, amyloid and preamyloid lesions in the BRI2 gene-related dementias: A morphological study. *Neuropathol. Appl. Neurobiol.* **32**, 492–504 (2006).
28. F. Madeira et al., Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res.* **50**, W276–W279 (2022).
29. J. Kyte, R. F. Doolittle, A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982).
30. M. I. Kanehisa, T. Y. Tsong, Local hydrophobicity stabilizes secondary structures in proteins. *Biopolymers* **19**, 1617–1628 (1980).
31. R. Vidal, A. G. Barbeito, L. Miravalle, B. Ghetti, Cerebral amyloid angiopathy and parenchymal amyloid deposition in transgenic mice expressing the danish mutant form of human BRI $_2$. *Brain Pathol.* **19**, 58–68 (2009).
32. F. Pickford, J. Coomaraswamy, M. Jucker, E. McGowan, Modeling familial British dementia in transgenic mice. *Brain Pathol.* **16**, 80–85 (2006).
33. M. S. Marcora et al., Amyloid peptides ABri and ADan show differential neurotoxicity in transgenic Drosophila models of familial British and Danish dementia. *Mol. Neurodegener.* **9**, 5 (2014).
34. B. J. Livesey, J. A. Marsh, Interpreting protein variant effects with computational predictors and deep mutational scanning. *Dis. Model Mech.* **15**, dmm049510 (2022).
35. M. Seuma, B. Bolognesi, Understanding and evolving prions by yeast multiplexed assays. *Curr. Opin. Genet. Dev.* **75**, 101941 (2022).
36. M. Thompson et al., Massive experimental quantification of amyloid nucleation allows interpretable deep learning of protein aggregation. bioRxiv [Preprint] (2024). https://doi.org/10.1101/2024.07.13.603366 (Accessed 1 October 2024).
37. C. Floquet, I. Hatin, J.-P. Rousset, L. Bidou, Statistical analysis of readthrough levels for nonsense mutations in mammalian cells reveals a major determinant of response to gentamicin. *PLoS Genet.* **8**, e1002608 (2012).
38. M. Dabrowski, Z. Bukowy-Bieryllo, E. Zietkiewicz, Translational readthrough potential of natural termination codons in eucaryotes–The impact of RNA sequence. *RNA Biol.* **12**, 950–958 (2015).
39. I. Toledano, F. Supek, B. Lehner, Genome-scale quantification and prediction of pathogenic stop codon readthrough by small molecules. *Nat. Genet.* **56**, 1914–1924 (2024).
40. C. Janke et al., A versatile toolbox for PCR-based tagging of yeast genes: New fluorescent proteins, more markers and promoter substitution cassettes. *Yeast* **21**, 947–962 (2004).
41. A. J. Faure, J. M. Schmiedel, P. Baeza-Centurion, B. Lehner, DiMSum: An error model and pipeline for analyzing deep mutational scanning data and diagnosing common experimental pathologies. *Genome Biol.* **21**, 207 (2020).
42. A.-M. Fernandez-Escamilla, F. Rousseau, J. Schymkowitz, L. Serrano, Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat. Biotechnol.* **22**, 1302–1306 (2004).
43. P. Charoenkwan et al., AMYPred-FRL is a novel approach for accurate prediction of amyloid proteins by using feature representation learning. *Sci. Rep.* **12**, 7697 (2022).
44. P. Sormanni, F. A. Aprile, M. Vendruscolo, The CamSol method of rational design of protein mutants with enhanced solubility. *J. Mol. Biol.* **427**, 478–490 (2015).
45. O. Conchillo-Solé et al., AGGRESCAN: A server for the prediction and evaluation of 'hot spots' of aggregation in polypeptides. *BMC Bioinformatics* **8**, 65 (2007).
46. M. Martín, B. Bolognesi, Deep mutagenesis reveals the distinct mutational landscape of ADan and ABri amyloid nucleation. Gene Expression Omnibus. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE244612. Deposited 9 October 2023.
47. M. Martín, B. Bolognesi, Amyloids "at the border": Deep mutagenesis and random sequence extension reveal an incomplete amyloid-forming motif in Bri2 that turns amyloidogenic upon C-terminal extension. Gene Expression Omnibus. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE270792. Deposited 2 August 2024.