Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Clinical significance of mRNA nonstop decay in rare disease diagnosis and recommendations for its application in variant classification

Yue Zhou
Baylor Genetics
Dandan He
Baylor Genetics
Nikita Mehta
Memorial Sloan Kettering Cancer Center
Christian C. Taborda
Baylor Genetics
Robert Rigobello
Baylor Genetics
Morgan Driver
Baylor Genetics
John Lattier
Baylor Genetics
Ning Liu
Baylor Genetics
Yue Wang
Baylor Genetics
David Wu
University of Washington
Lucy A. Godley
Northwestern University
Liesbeth Vossaert
Baylor Genetics
Xiaonan Zhao
Baylor Genetics
Linyan Meng
Baylor Genetics
Christine M. Eng
Baylor Genetics
Fan Xia
Baylor Genetics
Xi Luo
xi.luo@bcm.edu

Baylor Genetics

Research Article

Keywords: nonstop decay, mRNA surveillance, variant classification, PVS1, loss-of-function, stop-loss, rare disease, genome sequencing

Posted Date: March 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9126009/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: Competing interest reported. Y.Z., C.C.T., R.R., M.D., J.L., N.L., Y.W., L.V., X.Z., L.M., C.M.E., F.X., and X.L. are employed by Baylor Genetics. Baylor College of Medicine and Miraca Holdings Inc. have formed a joint venture with shared ownership and governance of Baylor Genetics, which performs genetic testing and derives revenue. N.L., Y.W., L.V., X.Z., L.M., C.M.E., F.X., and X.L. are also employees of Baylor College of Medicine and derive support through a professional services agreement with Baylor Genetics.

Abstract

Background: Nonsense-mediated decay (NMD) is routinely considered during clinical interpretation of predicted loss-of-function (LoF) variants, whereas mRNA nonstop decay (NSD)—a translation-coupled surveillance pathway that degrades transcripts lacking an in-frame termination codon—is rarely considered. Variants that abolish the canonical stop codon or remove the final stop-containing exon(s) can create nonstop transcripts, but current American College of Medical Genetics and Genomics/Association for Molecular Pathology (ACMG/AMP) variant classification guidelines and the Clinical Genome Resource (ClinGen) PVS1 recommendations do not explicitly address NSD, creating a potential mechanism-specific gap in variant classification.

Methods: We developed a computational workflow to identify disease-associated genes and transcript regions in which variants are predicted to generate nonstop transcripts in any of the three reading frames. Disease genes were obtained from OMIM and transcript annotations were retrieved from Ensembl. NSD-potential genes were defined as those lacking an in-frame stop codon in the 3' UTR prior to the annotated polyadenylation site for at least one reading frame. We then screened 2,439 clinical genome sequencing (GS) variant call format (VCF) files for variants overlapping NSD-susceptible regions and performed case-level review.

Results: Among 4,954 OMIM disease-associated protein-coding genes, 333 (6.72%) had NSD potential in ≥1 reading frame, yielding 546 genomic NSD-susceptible loci spanning 34,744 bp (~0.0011% of the human genome). Screening of 2,439 GS cases identified 359 candidate NSD variants; six variants in six probands were consistent with the individuals' phenotypes and the genes' disease mechanism. In these cases, incorporation of NSD as a transcript-null mechanism would increase the weight of LoF evidence and support reclassification from variant of uncertain significance to likely pathogenic or pathogenic when combined with additional clinical and genetic evidence.

Conclusions: NSD-predicted variants are not ultra-rare among disease genes and can be clinically relevant in rare disease diagnosis. We propose practical recommendations for recognizing NSD-susceptible variants and incorporating NSD into gene-/disease-specific LoF frameworks (e.g., PVS1) to improve consistency and diagnostic yield.

Background

Messenger RNA (mRNA) quality control pathways maintain proteostasis and transcriptome integrity by recognizing and degrading aberrant RNAs that would otherwise produce potentially toxic proteins. Among the best-characterized translation-coupled pathways are nonsense-mediated decay (NMD), no-go decay (NGD), and nonstop decay (NSD). NMD targets transcripts harboring premature termination codons (PTCs), typically when translation terminates upstream of exon–exon junction complexes or other features that signal premature termination [1–3]. NSD, in contrast, is triggered when translation reaches the 3' end of an mRNA in the absence of an in-frame stop codon, leading to ribosome stalling at the poly(A) tail and recruitment of factors that couple ribosome rescue with 3'→5' mRNA degradation by the SKI complex and the cytoplasmic RNA exosome [2–4] (Figure 1).

In clinical genomic variant classification, prediction of transcript-level consequences is central to interpreting variants that reduce or abolish protein production. The American College of Medical Genetics and Genomics/Association for Molecular Pathology (ACMG/AMP) framework includes the PVS1 criterion for variants predicted to cause loss of function (LoF) in genes where LoF is a known disease mechanism [5]. Subsequently, the Clinical Genome Resource (ClinGen) Sequence Variant Interpretation (SVI) working group recommendations provide a decision tree for assigning PVS1 strength (Very Strong/Strong/Moderate/Supporting) based on variant type, location, and predicted impact (e.g., NMD-competence, critical domains, and percentage of protein truncation) [6]. However, the current ACMG/AMP and SVI PVS1 guidance does not explicitly address NSD as a mechanism by which 3'-terminal frameshift, stop-loss, or terminal exon splice or deletion variants can generate a functional null allele. Consequently, variants expected to yield nonstop transcripts are frequently treated as C-terminal protein alterations with limited predicted impact, even when the true molecular outcome is transcript degradation through NSD.

We hypothesized that NSD-susceptible variation is more prevalent among disease genes than generally appreciated and that explicit recognition of NSD would improve the consistency and accuracy of LoF evidence assignment in variant classification. Here, we (i) systematically identified disease-associated genes with transcript architectures susceptible to NSD across reading frames, (ii) defined the genomic regions in which variants could generate nonstop transcripts, (iii) assessed the prevalence of candidate NSD variants in a clinical genome sequencing (GS) cohort, and (iv) propose practical recommendations for incorporating NSD into LoF variant classification, with an exemplar implementation in *DDX41*-related hematologic malignancy predisposition.

Methods

Identification of disease-associated gene set.

Disease-associated genes were obtained from the Online Mendelian Inheritance in Man (OMIM) database (genemap2.txt; accessed January 3, 2025). Genes were filtered to retain entries with phenotype annotations, approved HGNC gene symbols, and Ensembl gene identifiers. In total, 4,954 disease-associated protein-coding genes were included. Disease-associated but non–protein-coding genes (n = 24) were excluded from downstream analyses.

Transcript assembly and annotation.

For each gene, transcript annotations were retrieved from Ensembl using the Ensembl REST API, including exon coordinates and transcript biotypes. For each gene, exon sequences for the canonical transcript were retrieved and validated for completeness; full-length transcript sequences were assembled by concatenating exon sequences in genomic order and transcriptional orientation. Coding sequence (CDS) boundaries and the terminal exon were annotated by alignment to Ensembl GRCh38 CDS FASTA references.

Polyadenylation site selection.

NSD depends on whether a compensatory stop codon is present before the poly(A) tail. Because tissue-specific alternative polyadenylation may alter the effective 3' end of transcripts, we used polyadenylation sites corresponding to RefSeq transcripts where feasible, prioritizing those with the longest annotated 3' UTR among mapped candidates. Ensembl–RefSeq mappings were obtained via BioMart; manual review was performed when automated mappings were unavailable.

Identification of NSD-potential genes and NSD-susceptible regions.

For each assembled transcript, we scanned the 3' UTR (from the canonical stop codon to the poly(A) site) in all three reading frames for in-frame stop codons. Transcripts lacking an in-frame stop codon within the 3' UTR prior to the poly(A) site in at least one reading frame were classified as NSD-potential. For NSD-potential genes, we defined the transcript intervals in which a variant could generate a nonstop transcript for each reading frame (Figure 2). In the +0 frame, only disruption of the canonical stop codon (3 nucleotides) yields an in-frame nonstop transcript. For +1 and +2 frames, we simulated a frameshifted transcript and defined the NSD-susceptible region as the interval between the last available stop codon in that shifted frame and the canonical stop codon. Transcript-relative coordinates were converted to genomic coordinates to create a unified NSD coordinate reference set.

Clinical GS patient cohort

A retrospective analysis of GS results from 2,439 patients at Baylor Genetics was performed. These individuals underwent GS either as proband-only, duo, or trio-based to investigate suspected genetic etiologies. This diagnostic laboratory is certified by the College of American Pathologists (CAP) and complies with the Clinical Laboratory Improvement Amendments. The analysis of deidentified data was approved by the Institutional Review Board at Baylor College of Medicine (IRB: H-41191)

Genome sequencing

Peripheral blood samples and buccal samples were provided in most cases, although other sources of DNA were also accepted. Library preparation was performed using a PCR-free approach following the manufacturer's protocols (Illumina, San Diego, US). The GS protocol also incorporated either a coding single-nucleotide polymorphism array procedure or a spike-in plasmid tracer for quality control and sample tracking. Each library underwent sequencing with paired-end 150bp reads to achieve an average nuclear genome coverage ≥40x.

Variant detection and classification

Variant calling was conducted using in-house and proprietary bioinformatics pipelines. The scope of variant detection included single nucleotide variants (SNV), small insertions and deletions (<50bp, indels), copy-number variants (CNV), absence of heterozygosity, uniparental disomy, tandem repeat expansions, and mitochondrial variants. For clinical reporting, the genomic findings were classified into five tiers (Pathogenic, Likely Pathogenic, Variant of Uncertain Significance, Likely Benign, and Benign) according to Baylor Genetics' classification scheme, which incorporates the guidelines recommended by the ACMG/AMP [5].

Variant screening

We screened the 2,439 GS VCFs for variants overlapping NSD-susceptible regions (Figure 3). Variants were annotated for overlapping position, predicted reading frame, and variant consequence. Candidate NSD variants underwent post-screening review to determine (i) whether LoF is an established disease mechanism for the gene, (ii) whether zygosity and inheritance were consistent with the gene–disease association, and (iii) whether the individual's phenotype was compatible with the associated disorder. This study is descriptive; no hypothesis-testing statistics were applied.

Results

Genome-wide landscape of NSD susceptibility among disease genes

Among 4,954 OMIM disease-associated protein-coding genes, 333 (6.72%) were classified as NSD-potential in at least one reading frame. Across these 333 genes, we identified 546 distinct genomic regions capable of harboring variants that generate nonstop transcripts (Supplemental Table 1). In aggregate, NSD-susceptible loci spanned 34,744 nucleotides—approximately 0.0011% of the total chromosomal sequence—indicating that while the genomic footprint is small, the number of disease genes with NSD potential is substantial (Figure 4).

Prevalence of candidate NSD variants in a clinical GS cohort

Screening of 2,439 GS cases identified 359 candidate NSD variants across NSD-susceptible loci. Most variants were observed in blood group–related genes (n = 332) and there were three genes which LoF is not the established disease mechanism (n = 3). Eighteen candidate NSD variants were found in genes associated with autosomal recessive disease; in the absence of a second reportable variant in trans and compatible phenotypes, these findings were not clinically reportable but may be relevant in the context of reproductive testing (e.g., carrier screening).

Six probands harbored NSD-predicted variants in disease-causing genes that were consistent with their clinical phenotypes and inheritance expectations (Table 1). These variants would typically receive only PVS1_Moderate or PVS1_Strong under existing PVS1 decision trees because they occur near the 3' end and are predicted to alter only a small C-terminal portion of the protein. Incorporating NSD as a transcript-null mechanism supports treating these variants as

functional null alleles (subject to gene- and transcript-specific considerations) and can increase the overall weight of pathogenic evidence when combined with additional data (e.g., de novo occurrence, phenotype specificity, segregation and allelic data, in silico predictions).

Case vignettes illustrating NSD-relevant reclassification

NDUFS7 (autosomal recessive mitochondrial complex I deficiency, nuclear type 3, OMIM: 618224). Compound heterozygous variants were identified in the *NDUFS7* gene: a paternal inherited frameshift variant NM_024407.5:c.610del (p.Glu204Serfs*?) predicted to abolish the canonical stop codon and generate a nonstop transcript, and a known pathogenic missense variant NM_024407.5:c.364G>A (p.Val122Met). Without NSD consideration, the frameshift might be interpreted as a near-terminal truncation removing nine amino acids and receive only PVS1_Moderate; with NSD incorporation, it is consistent with a null allele acting in trans with a pathogenic variant and a highly concordant phenotype with mitochondrial complex I deficiency, supporting pathogenic classification.

ISG15 (autosomal recessive immunodeficiency 38, OMIM: 616126). A homozygous NM_005101.4:c.463dup (p.Arg155Profs*?) frameshift variant predicted to create a nonstop transcript was identified in the *ISG15* gene. Because the frameshift occurs near the 3' end with only ten animo acids downstream, this variant would typically receive PVS1_Moderate; NSD provides a mechanistic rationale for transcript degradation and supports pathogenic classification in the context of a compatible phenotype with immunodeficiency and recessive inheritance in this patient.

HCN2 (autosomal dominant *HCN2*-related epilepsy, OMIM: 602477). A heterozygous de novo frameshift NM_001194.4:c.2328_2334dup (p.Ser779Profs*?) variant predicted to generate a nonstop transcript was detected in the *HCN2* gene. This variant disrupts normal transcript termination and generates a nonstop mRNA predicted to be degraded through the NSD pathway. This variant had a PVS1_Strong criteria code since the altered 3’ coding region of protein is more than 10% and was also reclassified from variant of uncertain significance (VUS) to likely pathogenic after the incorporation of NSD rules in variant classification which is consistent with the clinical diagnosis of epilepsy.

HNRNPD (autosomal dominant *HNRNPD*-related neurodevelopmental disorder, CCID: 009030). A de novo canonical splice acceptor variant in the last coding exon of the *HNRNPD* gene (NM_031369.3:c.944-1G>A) was identified. For this gene, LoF/haploinsufficiency is an established disease mechanism. Current splicing recommendations often assign PVS1_Moderate for the last-exon splice site variants in the absence of evidence for NMD [7]; NSD provides an additional pathway by which last-exon splice disruption can lead to transcript loss when the termination codon is removed and no compensatory stop codon exists prior to polyadenylation, supporting a higher PVS1 strength when transcript context is consistent.

SHANK2 (autosomal dominant *SHANK2*-related neurodevelopmental disorder, CCID: 006134). A heterozygous frameshift variant was detected in the *SHANK2* gene (NM_012309.5:c.5302_5305del; p.Leu1768Glnfs*?). In genes where haploinsufficiency is the disease mechanism, NSD prediction is consistent with a null allele and supports upgrading LoF evidence from PVS1_Moderate to PVS1. This variant was also reclassified from VUS to likely pathogenic which is consistent with the neurodevelopmental phenotype in this patient.

TYMP (autosomal recessive mitochondrial DNA depletion syndrome type 1 (MNGIE), OMIM: 603041). Compound heterozygous variants were detected in the *TYMP* gene including a frameshift duplication variant NM_001953.5:c.865_889dup (p.Ala297Glyfs*?) predicted to abolish the native stop codon and induce NSD, in trans with a likely pathogenic in-frame deletion NM_001953.5:c.1198_1203del (p.Val400_Leu401del). Considering LoF as the disease mechanism, NSD supports treating the frameshift as a null allele consistent with molecular and clinical diagnosis of mitochondrial DNA depletion syndrome type 1 (MNGIE) in this patient.

Exemplar implementation in gene-/disease-specific variant classification guidance: *DDX41*

ClinGen Variant Curation Expert Panels (VCEPs) develop gene-/disease-specific ACMG/AMP specifications to improve consistency and accuracy in variant classification. The *DDX41* gene is associated with *DDX41*-related hematologic malignancy predisposition syndrome (CCID: 004638). The condition is primarily characterized by adult-onset acute myeloid leukemia (AML) or myelodysplastic syndrome (MDS). Malignant transformation is driven by biallelic inactivation of *DDX41* with the second hit frequently being one of several highly specific somatic hotspot variants [8,9]. Given *DDX41* functions as a tumor suppressor gene, null variants are typically identified in patients at risk for this condition. While many pathogenic *DDX41* variants are truncating, a subset of terminal frameshift variants is predicted to generate nonstop transcripts depending on the location of compensatory stop codons relative to the poly(A) tail. We summarize published and database-reported individuals who carry *DDX41* frameshift variants consistent with NSD prediction (Table 2). In several reported cases, coexistence of a hotspot somatic variant provides phenotype/genotype coherence that can strengthen pathogenicity assessment when NSD is recognized as a functional null mechanism.

For *DDX41* (NM_016222.4), the annotated polyadenylation signal and insertion site imply that +1 and −1 frameshift variants in the terminal coding region can place the next in-frame stop codon downstream of the poly(A) tail, producing nonstop transcripts. In this setting, treating such variants as LoF is mechanistically and clinically reasonable, contingent on transcript selection and gene mechanism considerations. This illustrates how NSD criteria can be incorporated into ClinGen VCEP specifications alongside established PVS1 decision trees.

Recommendations for incorporating NSD into variant classification

Recommendation 1: Treat NSD as a transcript-null mechanism analogous to NMD when LoF is an established disease mechanism. NSD is expected to reduce transcript abundance through translation-coupled degradation when an mRNA lacks an in-frame termination codon and translation reaches the poly(A) tail [2–4]. For genes in which haploinsufficiency or biallelic LoF causes disease, variants predicted to trigger NSD should generally be evaluated under LoF frameworks (e.g., PVS1), rather than being discounted as minor C-terminal alterations.

Recommendation 2: Define NSD susceptibility on a transcript-specific basis. NSD prediction depends on (i) the clinically relevant transcript (preferably MANE Select and/or MANE Plus Clinical where available), (ii) the location of the poly(A) site (including alternative polyadenylation), and (iii) the presence of compensatory stop codons in the relevant reading frame between the disrupted canonical stop codon and the poly(A) tail. Laboratories should document the transcript used and the stop-codon landscape in the relevant frame as part of the interpretation rationale.

Recommendation 3: Recognize the variant classes most likely to trigger NSD. These include: (a) stop-loss variants that abolish the canonical stop codon without the alteration of upstream coding sequence; (b) frameshift variants in the terminal coding region that eliminate the canonical stop codon; (c) splice acceptor/donor or other variants that cause skipping of the last (or penultimate) coding exon containing the termination codon or cause a frameshift into a NSD-susceptible reading frame; and (d) copy-number or structural variants that delete the final stop-containing exon(s).

Recommendation 4: Calibrate strength of evidence using existing PVS1 logic, with explicit NSD checkpoints. We recommend adding an NSD checkpoint to the PVS1 decision tree [6] for variants near the 3' end that would otherwise be downgraded due to limited predicted truncation. If the variant is predicted to generate a nonstop transcript on the clinically relevant transcript and LoF is a known disease mechanism, applying PVS1 at the same strength as other null variants of comparable confidence is appropriate (Table 3). In semi-quantitative adaptations of ACMG/AMP, Very Strong pathogenic evidence corresponds to 8 points [10,11], and an NSD-triggering variant in a LoF gene can reasonably contribute at this level when prediction confidence is high. (Notably, an ACMG/AMP/CAP/ClinGen SVC v4.0 standard is in development; laboratories should align NSD implementation with the finalized framework when released.)

Recommendation 5: Distinguish canonical NSD from long 3'-tail/termination inefficiency mechanisms. Some transcripts may terminate far downstream of the canonical stop due to readthrough or creation of a distant stop codon; evidence suggests that unusually long 3' UTR translation can trigger decay mechanisms distinct from classic NSD [12]. When the next stop codon is present but far downstream (e.g., >150 nt beyond the canonical termination site), we suggest considering a reduced-strength LoF contribution (PVS1_Moderate) unless additional experimental evidence demonstrates transcript loss (Table 3). This approach parallels existing guidance to modulate PVS1 strength based on predicted molecular outcome confidence [6].

Operational checklist. When evaluating a candidate NSD variant, we recommend documenting: (1) gene–disease mechanism (LoF established?); (2) transcript selection (MANE Select/ MANE Plus Clinical/ clinically curated transcript); (3) whether the variant removes the canonical stop codon or stop-containing exon(s); (4) whether an in-frame compensatory stop codon exists prior to polyadenylation in the relevant reading frame; (5) whether alternative transcripts or polyadenylation could rescue termination; and (6) any corroborating evidence (RNA studies, segregation, de novo status, phenotype specificity).

Discussion

This study highlights NSD as an under-recognized but potentially important mechanism in clinical variant classification. Although the genomic regions capable of generating nonstop transcripts are small in aggregate, we found that ~7% of OMIM disease-associated genes have transcript architectures that permit NSD in at least one reading frame. This observation challenges the perception that NSD-relevant variants are “ultra-rare” and suggests that NSD should be considered routinely when interpreting stop-loss, terminal frameshift, or last-exon splice or deletion variants in LoF-mediated conditions.

Our case series illustrates a recurring interpretive pitfall: near-terminal variants are often downgraded because they appear to remove only a short C-terminal segment. This logic is appropriate when translation is expected to terminate soon after the variant and produce a largely intact protein, but it can be misleading when the true outcome is loss of the termination codon and degradation of the transcript. In most examples here, NSD provides a mechanistically coherent explanation for functional null alleles that would otherwise be categorized as VUS due to reduced PVS1 strength. Importantly, NSD should not be applied indiscriminately: prediction requires transcript-specific assessment of compensatory stop codons and polyadenylation sites, and clinical classification must still integrate additional evidence (inheritance, segregation, de novo status, phenotype specificity, population frequency, in silico prediction, experimental evidence) per ACMG/AMP/ClinGen guidelines [5,6].

The *DDX41* exemplar underscores how NSD can be operationalized within ClinGen VCEP specifications. In tumor suppressor genes with characteristic second-hit somatic hotspots, recognizing NSD as a LoF mechanism can align molecular predictions with observed genotype patterns. More broadly, NSD-aware LoF assessment may improve cross-laboratory consistency by reducing reliance on variable, ad hoc interpretations of terminal frameshift or stop-loss variants.

Limitations. First, NSD susceptibility was inferred computationally from reference transcript annotations and selected polyadenylation sites; alternative splicing and alternative polyadenylation could mitigate or exacerbate NSD in a tissue-specific manner. Second, we did not systematically quantify transcript abundance changes (e.g., RNA sequencing) for candidate variants; future work integrating RNA data would help calibrate prediction confidence and PVS1 strength. Third, our GS cohort screening reflects a clinical testing population and is enriched for disease-causing gene and variants which may limit generalizability of raw prevalence estimates. Finally, interpretation in oncology predisposition genes may require additional somatic context and careful distinction between germline and somatic events.

Conclusion

Variants predicted to trigger mRNA nonstop decay are present across a non-trivial fraction of disease-associated genes and can contribute directly to rare disease molecular diagnoses. Incorporating NSD into LoF variant classification—using transcript-specific checks and gene mechanism awareness—can improve the consistency and accuracy of PVS1 strength assignment and reduce the number of clinically relevant variants remaining as VUS. We recommend that NSD checkpoints be incorporated into gene-/disease-specific ACMG/AMP/ClinGen specifications and, as consensus frameworks evolve, into general variant classification guidance.

Abbreviations

ACMG: American College of Medical Genetics and Genomics; AML: Acute Myeloid Leukemia; AMP: Association for Molecular Pathology; CAP: College of American Pathologists; CCID: ClinGen Curation Identification; CDS: coding sequence; ClinGen: Clinical Genome Resource; GS: genome sequencing; LoF: loss of function; MANE: Matched Annotation from NCBI and EMBL-EBI; MDS: Myelodysplastic Syndromes; NGD: no-go decay; NMD: nonsense-mediated decay; NSD: nonstop decay; PTC: premature termination codon; SVI: Sequence Variant Interpretation; VCEP: Variant Curation Expert Panel; VCF: variant call format; VUS: variant of uncertain significance.

Declarations

Ethics approval and consent to participate

This study was conducted in accordance with protocol (H-41191) approved by the Institutional Review Board at Baylor College of Medicine. The study was also conducted in accordance with the Declaration of Helsinki.

Consent for publication

Consent was waived for the individuals included in this study per Institutional Review Board protocol.

Availability of data and materials

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

Y.Z., C.C.T., R.R., M.D., J.L., N.L., Y.W., L.V., X.Z., L.M., C.M.E., F.X., and X.L. are employed by Baylor Genetics. Baylor College of Medicine and Miraca Holdings Inc. have formed a joint venture with shared ownership and governance of Baylor Genetics, which performs genetic testing and derives revenue. N.L., Y.W., L.V., X.Z., L.M., C.M.E., F.X., and X.L. are also employees of Baylor College of Medicine and derive support through a professional services agreement with Baylor Genetics.

Fundings

No funding was received for the current study.

Authors' contributions

Study conception: Y.Z., X.L. Study design: Y.Z., X.L. Data curation: Y.Z., D.H., N.M., C.C.T., J.L., X.L. Data analysis: Y.Z., D.H., N.M., J.L., X.L. Data interpretation: Y.Z., D.H., N.M., J.L., N.L., Y.W., X.L. Writing-original draft: Y.Z., D.H., N.M., R.R., M.D., N.L., X.L. Writing-review and editing: Y.Z., D.H., N.M., C.C.T., R.R., M.D., J.L., N.L., Y.W., D.W., L.A.G., L.V., X.Z., L.M., C.M.E., F.X., and X.L.

Acknowledgements

Not applicable.

References

1. Wolin SL, Maquat LE. Cellular RNA surveillance in health and disease. Science (1979). American Association for the Advancement of Science; 2019. p. 822–7. https://doi.org/10.1126/science.aax2957
2. Monaghan L, Longman D, Cáceres JF. Translation-coupled mRNA quality control mechanisms. EMBO J. Springer Science and Business Media LLC; 2023;42. https://doi.org/10.15252/embj.2023114378

3. Powers KT, Szeto JYA, Schaffitzel C. New insights into no-go, non-stop and nonsense-mediated mRNA decay complexes. Curr. Opin. Struct. Biol. Elsevier Ltd; 2020. p. 110–8. https://doi.org/10.1016/j.sbi.2020.06.011
4. Klauer AA, van Hoof A. Degradation of mRNAs that lack a stop codon: A decade of nonstop progress. Wiley Interdiscip. Rev. RNA. 2012. p. 649–60. https://doi.org/10.1002/wrna.1124
5. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genetics in Medicine. Nature Publishing Group; 2015;17:405–24. https://doi.org/10.1038/gim.2015.30
6. Abou Tayoun AN, Pesaran T, DiStefano MT, Oza A, Rehm HL, Biesecker LG, et al. Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion. Hum Mutat. John Wiley and Sons Inc.; 2018;39:1517–24. https://doi.org/10.1002/humu.23626
7. Walker LC, Hoya M de la, Wiggins GAR, Lindy A, Vincent LM, Parsons MT, et al. Using the ACMG/AMP framework to capture evidence related to predicted and observed impact on splicing: Recommendations from the ClinGen SVI Splicing Subgroup. Am J Hum Genet. Cell Press; 2023;110:1046–67. https://doi.org/10.1016/j.ajhg.2023.06.002
8. Polprasert C, Schulze I, Sekeres MA, Makishima H, Przychodzen B, Hosono N, et al. Inherited and Somatic Defects in DDX41 in Myeloid Neoplasms. Cancer Cell. Cell Press; 2015;27:658–70. https://doi.org/10.1016/j.ccell.2015.03.017
9. Kadono M, Kanai A, Nagamachi A, Shinriki S, Kawata J, Iwato K, et al. Biological implications of somatic DDX41 p.R525H mutation in acute myeloid leukemia. Exp Hematol. Elsevier Inc.; 2016;44:745-754.e4. https://doi.org/10.1016/j.exphem.2016.04.017
10. Tavtigian S V., Greenblatt MS, Harrison SM, Nussbaum RL, Prabhu SA, Boucher KM, et al. Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework. Genetics in Medicine. Nature Publishing Group; 2018;20:1054–60. https://doi.org/10.1038/gim.2017.210
11. Tavtigian S V., Harrison SM, Boucher KM, Biesecker LG. Fitting a naturally scaled point system to the ACMG/AMP variant classification guidelines. Hum Mutat. John Wiley and Sons Inc; 2020;41:1734–7. https://doi.org/10.1002/humu.24088
12. Takata A, Hamanaka K, Matsumoto N. Refinement of the clinical variant interpretation framework by statistical evidence and machine learning. Med. Cell Press; 2021;2:611-632.e9. https://doi.org/10.1016/j.medj.2021.02.003

Tables

Table 1. Positive GS cases with NSD variants identified in disease-causing genes consistent with clinical phenotypes.

Gene	Disease	Inheritance/ Zygosity	Variant	Protein Full Length	PVS1 criteria without or with NSD incorporation	ACMG/AMP/ClinGen Classification without or with NSD incorporation
NDUFS7	Mitochondrial Complex I Deficiency, Nuclear Type 3 (OMIM: 618224).	AR/ Compound Heterozygous	NM_024407.5: c..610del (p.E204Sfs*?); c.364G>A (p.V122M)	213	PVS1_Moderate → PVS1	VUS → Pathogenic
ISG15	Immunodeficiency 38 (OMIM: 616126)	AR/ Homozygous	NM_005101.4: c.463dup (p.R155Pfs*?)	165	PVS1_Moderate → PVS1	VUS → Pathogenic
HCN2	*HCN2*-related epilepsy (OMIM: 602477)	AD/ De Novo	NM_001194.4: c.2328_2334dup, (p.S779Pfs*?)	889	PVS1_Strong → PVS1	VUS → Likely Pathogenic
HNRNPD	*HNRNPD*-related neurodevelopmental disorder (CCID: 009030)	AD/ De Novo	NM_031369.3:c.944-1G>A	355	PVS1_Moderate → PVS1	VUS → Pathogenic
SHANK2	*SHANK2*-related neurodevelopmental disorder (CCID: 006134)	AD/ Heterozygous	NM_012309.5: c.5302_5305del (p.Leu1768Glnfs*?)	1800	PVS1_Moderate → PVS1	VUS → Likely Pathogenic
TYMP	Mitochondrial DNA depletion syndrome type 1 (MNGIE) (OMIM: 603041)	AR/ Compound Heterozygous	NM_001953.5:c.865_889dup (p.Ala297Glyfs*?); c.1198_1203del (p.Val400_Leu401del)	482	PVS1_Strong → PVS1	Likely Pathogenic → Pathogenic

Table 2. *DDX41* frameshift variants reported in literature or public databases that are predicted to be subject to non-stop decay.

Germline cDNA	Germline protein	Germline VAF	Somatic cDNA	Somatic protein	Somatic VAF	Origin confirmed?	Age	Diagnosis	R
c.1791_1792del	p.(Lys597Asnfs*?)	>40%	c.1574G>A	p.Arg525His	10%	No	66	AML	[1
c.1811_1812insCATATGTGCTAT	p.(Lys604Asnfs*?)	52%	c.1574G>A	p.Arg525His	23%	No	N/A	Myeloid neoplasm	[3 33
c.1836_1837del	p.(Asp613LeufsTer?)	35-40%	N/A	N/A	N/A	No	60-65	None	[4 1 T
None	None	N/A	c.1857dup c.1589G>A	p.Met620Hisfs*? p.Gly530Asp	34.8% 30.5%	Yes	72	Myeloid neoplasm	[5
c.1814del	p.(Gln605ArgfsTer?)	41.9%	c.1574G>A	p.Arg525His	5.3%	Yes	N/A	MDS	[6
c.1721del	p.(Leu574Argfs*?)	N/A	N/A	N/A	N/A	N/A	N/A	MDS, AML	C
c.1773del	p.(Ile592Serfs*?)	45-50%	N/A	N/A	N/A	No	N/A	N/A	g 1 T
c.1843del	p.(Leu615TrpfsTer?)	45-50%	N/A	N/A	N/A	No	40-45	N/A	g 1 A

Table 3. Conceptual mapping of termination-related mRNA surveillance mechanisms and proposed evidence strength in variant classification.

Scenario	Description	Pathway Triggered	LoF evidence criteria
Premature stop codon (early termination)	Nonsense-Mediated Decay (NMD)	Upstream exon–exon junction → decay triggered	PVS1; +8 points
No stop codon → ribosome reaches poly(A)	Non-stop Decay (NSD)	Ribosome stalls at poly(A), mRNA degraded → decay triggered	PVS1; +8 points
Stop codon far downstream (>150 nt past normal site)	Extended nonstop-like surveillance	Ribosome reads through long 3′ region → decay may be triggered	PVS1_Moderate; +2 points

References

1. Ebert MS´, Passet M, Raimbault A, Rahmé R, Rahmé R, Raffoux E, et al. Germline DDX41 mutations define a significant entity within adult MDS/AML patients. Blood. 2019.

2. Duployez N, Largeaud L. Prognostic impact of DDX41 germline mutations in intensively treated acute myeloid leukemia patients: an ALFA-FILO study. Blood. American Society of Hematology; 2020;136:2125–32. https://doi.org/10.1182/BLOOD.2019000962

3. Aguilera-Diaz A, Larrayoz MJ, Palomino-Echeverría S, Vazquez I, Ariceta B, Mañú A, et al. Strategy for identification of a potential inherited leukemia predisposition in a 299 patient's cohort with tumor-only sequencing data. Leuk Res. Elsevier Ltd; 2020;95. https://doi.org/10.1016/j.leukres.2020.106386

4. Kovilakam SC, Gu M, Dunn WG, Marando L, Barcena C, England G, et al. Prevalence and significance of DDX41 gene variants in the general population.

5. Qu S, Li B, Qin T, Xu Z, Pan L, Hu N, et al. Molecular and clinical features of myeloid neoplasms with somatic DDX41 mutations. Br J Haematol. Blackwell Publishing Ltd; 2021;192:1006–10. https://doi.org/10.1111/bjh.16668

6. Maierhofer A, Mehta N, Chisholm RA, Hutter S, Baer C, Nadarajah N, et al. The clinical and genomic landscape of patients with DDX41 variants identified during diagnostic sequencing. Blood Adv. American Society of Hematology; 2023;7:7346–57. https://doi.org/10.1182/bloodadvances.2023011389

Figures

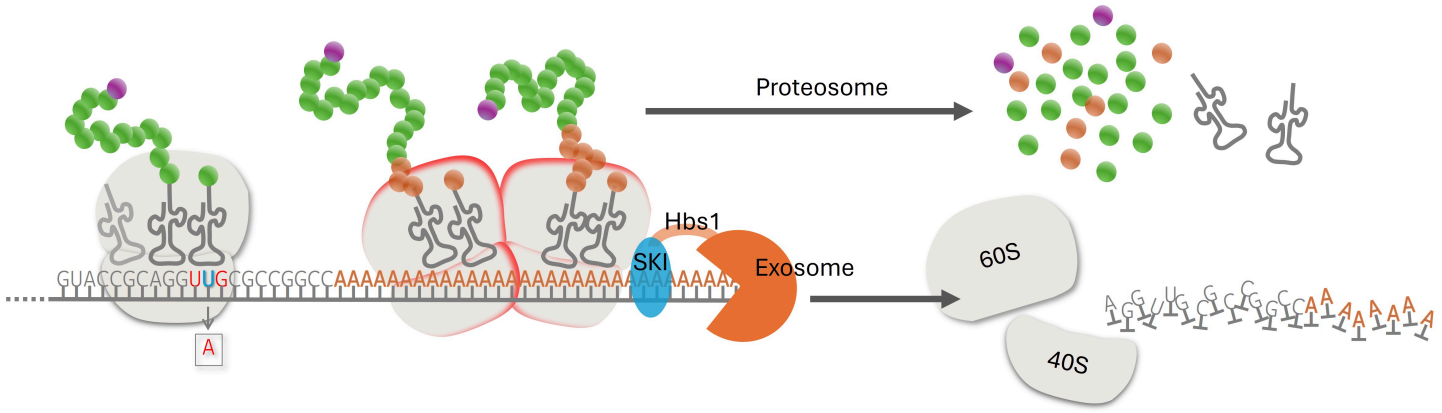


Figure 1

Nonstop decay mechanism. Translation of an mRNA lacking a stop codon leads to ribosome stalling at the poly(A) tail and recruitment of factors that couple ribosome rescue with 3'→5' mRNA degradation (SKI complex and cytoplasmic exosome) and nascent peptide handling via the proteasome.

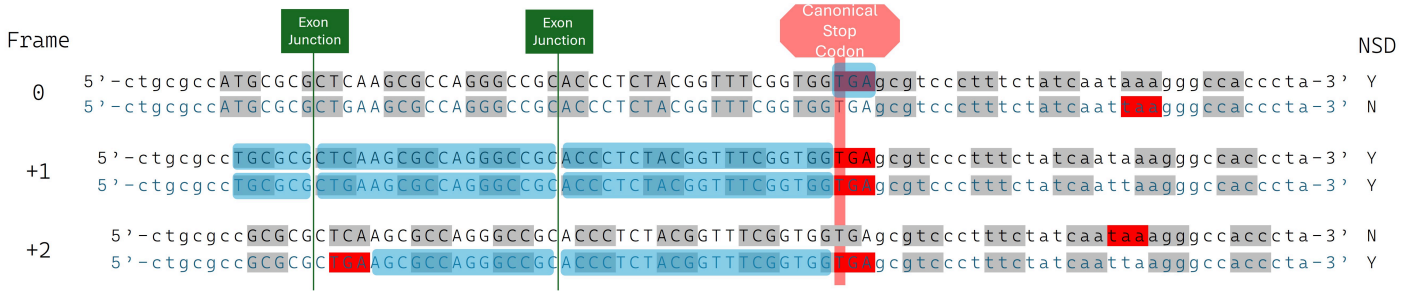


Figure 2

Sequence analysis workflow to identify NSD-susceptible genomic regions. For each gene, the mRNA is analyzed in three reading frames. For each frame, the last in-frame stop codon and the canonical stop codon are highlighted in red and transcript intervals in which variants could produce nonstop transcripts are defined in blue.

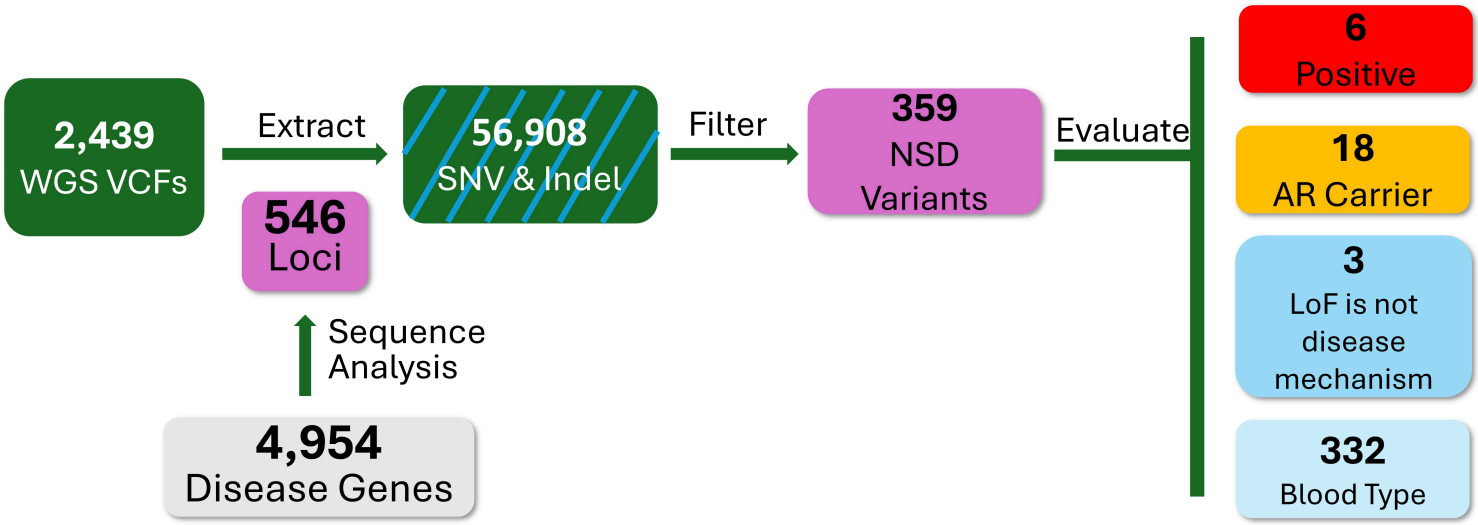


Figure 3

Flowchart of clinical GS screening for NSD candidate variants.

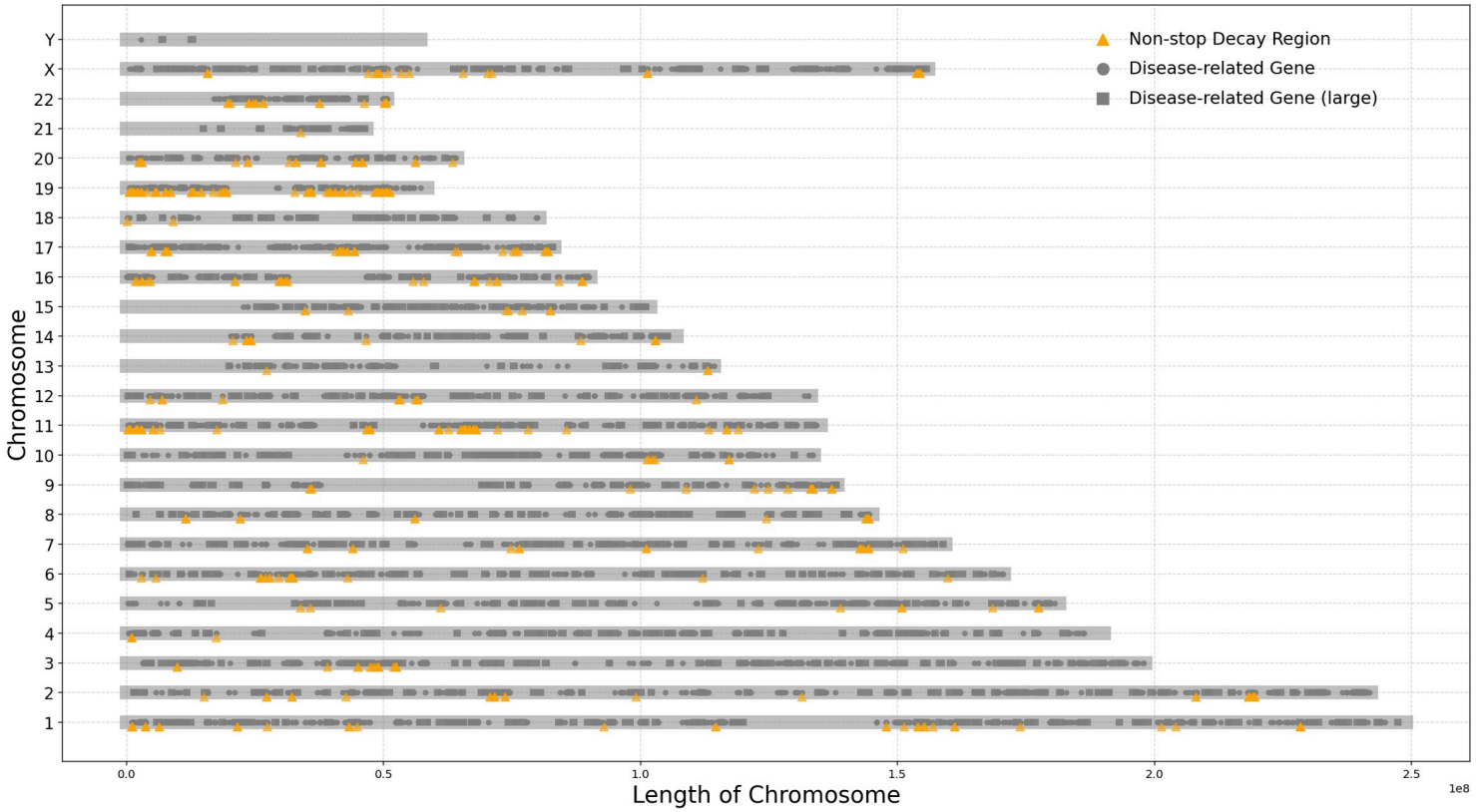


Figure 4

Distribution of NSD-capable loci throughout the genome.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- NSDSupplementalTable1.csv