Nonsense-mediated decay is highly stable across individuals and tissues

Nicole A. Teran,1,2,6 Daniel C. Nachun,1,6 Tiffany Eulalio,1,3 Nicole M. Ferraro,1,3 Craig Smail,1,3,4 Manuel A. Rivas,5 and Stephen B. Montgomery1,2,*

Summary

Precise interpretation of the effects of rare protein-truncating variants (PTVs) is important for accurate determination of variant impact. Current methods for assessing the ability of PTVs to induce nonsense-mediated decay (NMD) focus primarily on the position of the variant in the transcript. We used RNA sequencing of the Genotype Tissue Expression v.8 cohort to compute the efficiency of NMD using allelic imbalance for 2,320 rare (genome aggregation database minor allele frequency ≤ 1%) PTVs across 809 individuals in 49 tissues. We created an interpretable predictive model using penalized logistic regression in order to evaluate the comprehensive influence of variant annotation, tissue, and inter-individual variation on NMD. We found that variant position, allele frequency, the inclusion of ultra-rare and singleton variants, and conservation were predictive of allelic imbalance. Furthermore, we found that NMD effects were highly concordant across tissues and individuals. Due to this high consistency, we demonstrate in silico that utilizing peripheral tissues or cell lines provides accurate prediction of NMD for PTVs.

Introduction

RNA expression is not only regulated by transcription but also by degradation;1 RNA transcripts with protein-truncating variants (PTVs) are often targeted for degradation by the nonsense-mediated decay (NMD) pathway.2 The accurate identification of PTV-harboring transcripts that are successfully cleared by NMD can have a large effect on disease outcome. Some nonsense mutations lead to dominant-negative effects where the truncated allele can impede the function of the full-length allele.3 Mendelian disease diagnostics can benefit from the identification of the PTVs that escape NMD and may therefore lead to truncated peptides and corresponding gain-of-function effects.4 To be able to identify the PTVs that undergo or escape NMD, existing tools have integrated variant-level annotations which provide a prediction of the NMD efficiency or ability for a PTV containing transcript to be targeted and degraded by the NMD machinery as measured by the relative amount of a PTV containing transcript as compared to the wild-type.5–8

Position explains most variation in NMD efficiency, summarized by the 50 nucleotide (50nt) rule: if the variant occurs farther upstream than 55 to 50 nucleotides before the last exon junction, it will be targeted for degradation. Additional analysis in cancer has indicated that falling within 150 nt of the start of a gene or in a long exon (>407 base pairs) impedes degradation and a simple decision tree, called NMDetective-B, which utilizes these rules can explain 68% of the variation in NMD efficiency.8 These existing approaches have benefited from measuring NMD effects through allele-specific measurement of RNA-sequencing (RNA-seq) read counts overlying PTV variants. However, there is evidence that the ratio of the RNA read counts from the aberrant allele to that of the wild-type allele can vary between tissues, which would not be expected if variant position was the only determining factor.7,9 Furthermore, it remains unclear as to the degree NMD efficiency varies across individuals.

We utilized the Genotype Tissue Expression (GTex) dataset to assess the impact of tissue type on NMD efficiency.10 We measured the functional impact of 2,320 rare (genome aggregation database [gnomAD] minor allele frequency [MAF] ≤ 1%) PTVs from 809 individuals across 49 different tissues. We observed that, in addition to position, allele frequency, including rare, ultra-rare (MAF < 0.001%), and singleton alleles predict NMD efficiency. However, tissue is not predictive of NMD efficiency and PTVs showed more consistent allelic imbalance across tissues than any other type of coding transcript variant. Using this information, we demonstrate that accurate identification of PTVs that either undergo or escape NMD can be further achieved in peripheral tissues or cell lines.

Material and methods

Calling nonsense-mediated decay from allele-specific expression

The set of calls generated from binary alignment map files (BAMs) aligned with STAR using the WASP method for allelic mapping bias1 was used. Allele-specific expression (ASE) was called from
GTEx version 8 data using ASEAlleleCounter from GATK, analyzing only bi-allelic heterozygous SNPs. Nonsense-mediated decay (NMD) from ASE calls was defined as occurring at a protein-truncating variant (PTV) if the ratio of reference reads to the total number of reads was greater than 0.65.

**Variant annotation**

Variants were annotated using Variant Effect Predictor with Ensembl v.88, the same annotation used for other analyses in GTEx v.8, except to obtain gnomAD allele frequencies, for which version 97 of the Ensembl annotation was used. The LOFTEE plugin for VEP was used to determine if a variant was predicted to be loss of function and if the variant fell after 50 nt before the last exon junction. Conservation scores and GC content for each variant was looked up in the existing resource Combined Annotation Dependent Depletion (CADD). The canonical isoform was used to select a single annotation for each variant, with variants which were annotated to multiple genes, had multiple predicted consequences, or were annotated as intergenic being removed. NMD was only considered for variants which were exclusively annotated as “stop gained.” The additional categories “missense_variant,” “synonymous_variant,” “intron_variant,” “3_prime_UTR_variant,” “5_prime_UTR_variant,” and “non_coding_transcript_exon_variant” were used for the comparison of different classes of variants described below.

**Multi-tissue allele-specific expression**

Using the proportion of reference reads as a measurement for NMD efficiency does not account for the uncertainty of low read coverage, nor does it exploit the availability of gene expression from multiple tissues in a subject to increase certainty. We used a procedure described in Rivas et al. to integrate this information and compute a probability of ASE. This normally produces probabilities for no ASE, moderate ASE, or strong ASE. We disabled the estimation of strong ASE in each tissue as this usually exhibited a low probability, that is, a variant was unlikely to be predicted to undergo strong ASE. Using only the moderate ASE measurement provided one probability for the presence of ASE per sample. For every variant, an individual probability was given for each tissue from an individual. NMD was defined as occurring in a given variant if the ASE probability was greater than 0.8 and the proportion of reference reads was greater than 0.5 (to remove variants that exhibited a bias toward the alternate allele).

**Predictive models**

NMD efficiency was predicted as either a categorical outcome, using the proportion of reference reads or the multi-tissue ASE probabilities, or a continuous outcome, using L1-penalized regression with the glmnet R package on scaled and centered predictors. For the categorical outcomes, the logistic family was used, with area under the curve (AUC) used as the performance metric, while for the continuous outcomes, the Gaussian family was used on the logit-transformed probabilities, with a correction applied to adjust probabilities of 0.0 or 1.0, and root mean square error (RMSE) as the performance metric. The penalization parameter lambda was optimized across a range of values chosen by glmnet using cross-validation using chromosomes as the folds to guarantee that all variants in the test set were never used in the training set. The best lambda was chosen as the value for which the mean performance metric was best across all 22 folds. p values and confidence intervals were estimated using the selectiveInference package using the optimal lambda identified in the cross-validation.

**Correlation of ASE across tissues**

We estimated the Pearson’s correlation of ASE across pairs of tissues by matching all instances of a variant being shared between two tissues in the same subject and correlating the corresponding proportions of reference reads. Separate correlations were performed for each of the following classes of variants: stop-gain, missense variant, synonymous variant, intronic variant, non-coding transcript exon variant, 5’ UTR variant, and 3’ UTR variant.

**Assessing NMD variants using ClinVar pathogenicity information**

PTV pathogenicity was assessed using the September 2020 release of the ClinVar Variant Summary table. Data were accessed on September 6, 2020 and filtered for premature termination variants. We then intersected the variants with ClinVar pathogenicity information with the NMD variants recovering 309/2,320 (13%). This accounted for 157 out of 1,189 observations (13.2%) in fibroblasts and 164 out of 1,013 observations (16.2%) in whole blood.

**Results**

**Identifying NMD-targeted variants in GTEx**

In order to evaluate NMD rules in and between normal human tissues, we annotated the proportion of expressed reference reads for rare PTV sites across the GTEx dataset. We used the genomes of 809 individuals of European descent to identify 2,320 different PTVs with an allele frequency from the gnomAD database less than or equal to one percent. Rare variants were selected to enrich for variants likely to undergo NMD and in order to prevent inclusion of common stop-gain variants that may have arisen due to selection and adaptation favoring the truncated transcript, as has been observed for “poison exons.” The proportion of expressed reference reads was calculated by dividing the number of RNA-seq reads that map to a variant site containing the reference allele by the total number of RNA-seq reads overlapping the variant site in a single sample (i.e., one tissue in one person).

We analyzed RNA sequencing data from 49 distinct tissues where each individual had a median number of 17 tissues and a median of five expressed PTVs that were testable in at least one GTEx tissue. Each variant was testable in 11 samples (a unique individual-tissue combination), with a median of eight observations per variant within an individual. We calculated the proportion of reference reads for each variant in each tissue for a total of 40,402 variant-tissue-subject observations from the 13,849 tissue-subject samples. From these 40,402 observations, 55% (22,301) were predicted to be targeted by the NMD machinery according to rules in Lindeboom et al. of not being near the start of a gene (within 150 nt), in a long exon (more than 407 nt), or after 55–50 nucleotides before the last exon junction. The remaining 45% (18,629) of PTVs were predicted to escape NMD. These rules, on a
whole, provided good separation of variants that showed allelic imbalance: 52% of observations of variants that were predicted to be targeted by NMD showed allelic imbalance (reference read proportion > 65%) compared to only 20% of those predicted to escape (Figure 1A).

**Ultra-rare protein-truncating variants have increased allelic imbalance**

Previous studies have reported that rarer PTVs are more likely to trigger NMD,7,8,20 including in an earlier version of GTEx.7 This initial exploration of NMD in GTEx analyzed 4,584 PTVs across the allele frequency spectrum acquired from 173 individuals. Given our increased sample size, more extensive whole-genome data in GTEx, and the availability of precise allele frequency information from gnomAD,13 we set out to evaluate this effect with more granularity in the rare allele frequency spectrum. Here, we evaluated the allelic imbalance for PTVs predicted to be NMD targets versus those predicted to be NMD escapees stratified by allele frequency. For rare variants, we saw significant separation between the predicted NMD escapees and the predicted NMD targets; strikingly, this separation was significantly more pronounced (by a difference in differences test) at ultra-rare (MAF < 0.001%) allele frequencies (Figure 1B).

By combining gnomAD allele frequency information with the whole-genome sequencing samples from GTEx, we were able to further investigate the allelic imbalance of ultra-rare variants seen in gnomAD against the 504 novel PTVs that were unobserved in gnomAD but present in GTEx (Figure 1B). These novel PTVs showed increased allelic imbalance, indicating that there is not a plateauing of the NMD effect for ultra-rare PTVs.

**NMD efficiency is primarily determined by mutation location, allele frequency, and conservation**

The primary means of detecting NMD has been the 50 nt rule in which a transcript will be degraded if the PTV occurs upstream of the point 55 to 50 nucleotides prior to the last exon-exon junction.5,21 However, as the 50 nt rule alone is not a perfect predictor of NMD efficiency (Figure S1), we wanted to investigate whether there were more subtle regulatory, tissue-specific, or inter-individual effects that could be detected using the multi-tissue, population design of GTEx. We chose to use the set of predictors previously described in Rivas et al.7 as they had been shown to have predictive power for NMD. Motivated by our previous findings (Figure 1B) and further leveraging the unique capabilities of GTEx, we added gnomAD MAF, tissue, and subject as predictors in the models to test their effects on NMD efficiency.

Initially, we constructed our model to predict allelic imbalance as defined by the binary classification of proportion of reference reads greater than or equal to 0.65 or less than 0.65, a threshold that has previously been shown to provide the best replication of effects.20 Notably, we found that including tissue as a predictive variable did not significantly improve the model (Figure 2A), and including individuals as a predictive variable actually decreased performance (Figure S2). Although we did see suggestive evidence for differences in median NMD efficiency between tissues (Figure S3), modeling tissue did not increase predictive performance when combined with other information about the PTV. This is similar to what was observed in Rivas et al.,7 where some samples showed differences in NMD efficiency. Despite our increase in sample size, we were not able to identify a systematic pattern. We also
includes 50nt rule, long exon, and near start (yellow). “Cons(ervation) Incl(uded)” model also includes the distance to the end (canonical stop), GC content, position in the coding sequence (from start), gnomAD allele frequency, vertebrate phyloP score, RNA integrity number, and total read depth at the site of interest (pink). “Tissue Incl(uded)” adds tissue to the conservation included model (green). (B) Forest plot of effect sizes and 95% confidence intervals for features that were chosen by the model with the optimal lambda penalty value as measured by AUC using the multi-tissue moderate ASE outcome. An odds ratio under one indicates the variable is associated with a decrease in the odds of the variant being imbalanced, while an odds ratio over one indicates the variable is associated with an increase in the odds of the variant being imbalanced. An odds ratio of one (gray line) indicates no association, farther from one indicates stronger association.

tested whether tissue was predictive of NMD in the subset of variants predicted to undergo NMD by the 50 nt rule, analogous to the tissue heterogeneity analysis in Rivas et al.,7 and observed uniformly poor predictive performance (best AUC = 0.56, Figure S4).

We were further able to leverage the multi-tissue design of the GTEx project to improve performance by predicting the incidence of allelic-specific expression (MODASE, equal to 1 − [probability of no ASE]) using a Bayesian stratification approach that reduces noise by including information from multiple observations of a PTV in one individual across tissues.7 Given the integration of multiple tissue information, this approach may reduce our ability to detect tissue-specific differences. We proceeded to use MODASE because tissue was not a predictive variable of the proportion of reference reads and the noise reduction led to an improvement in our predictive power (Figure S2).

In order to disentangle the often-correlated biological predictors, we chose to use the LASSO penalized logistic regression model implemented by the R package glmnet to produce a sparse and interpretable model.16 In addition to the canonical 50 nt rule, long exon, and start proximal predictors identified by Lindeboom et al.,8 we found that the distance to the canonical stop, GC content, position in the coding sequence (the distance from the start), gnomAD allele frequency, vertebrate phyloP score, RNA integrity number, and total read depth at the site of interest were significant predictors of MODASE status (Figure 2B, definitions in Table S1). Unsurprisingly, NMD was easier to detect in samples with higher RNA quality, as denoted by RNA integrity number, and for variants with higher read count.

In order to test the impact of additional factors, we included additional variant and subject level information in our model. Sample-level variables that were dropped from the model include age, sex, cause of death (Hardy scale), and postmortem interval.

Allelic imbalance of PTVs is consistent across tissues
Based on our observations that tissue was not predictive of allelic imbalance for PTVs, we wanted to evaluate the consistency of allelic imbalance for PTVs across tissues and within an individual. For each individual subject that had the same variant expressed in multiple tissues, we performed a pairwise correlation of the allelic ratio of that variant in those tissues. We were able to investigate most tissue combinations, but we did not have individuals that were sampled for both male-specific tissues (prostate and testis) and female-specific tissues (ovary and vagina) or in two of the lower sampled tissues analyzed (small intestine - terminal ileum and brain - amygdala). We also computed intra-individual, cross-tissue pairwise correlations of a variant’s allelic ratios for missense and synonymous coding variants and non-coding variants in introns, untranslated regions (UTRs), and non-coding exons (Figure S5 and Table S2, summarized in Figure 3).

We found a significantly stronger correlation between the proportion of reference reads for all PTVs, with a median Pearson correlation of 0.508, than for any other coding transcript variant, with a median Pearson correlation of 0.131 for synonymous variants and 0.204 for missense variants, or non-coding variants in introns or UTRs (median correlation of 0.176 for 3’ UTRs, 0.243 for 5’ UTRs, and 0.257 for intronic variants). Additionally, PTVs that were predicted to escape NMD showed lower correlation (median 0.449) than those that were predicted to undergo NMD (median 0.552), suggesting the consistency of NMD across tissues.

Intriguingly, noncoding transcripts were the only transcripts that showed higher between-tissue allelic
correlations. This higher correlation was not attributable to a systematic difference in read depth (Figure S6) or the distribution of the proportion of reference reads between noncoding transcripts and other gene biotypes (Figure S7). It was also not attributable to differences in conservation scores or GC content between noncoding transcript exon variants and other noncoding variants (Figure S8). Only 130 of 58,133 noncoding transcript exon variants were found to fall within any annotated CDS regions, confirming that these variants are not being expressed in any alternate coding transcripts.

*Unpredicted allelic balance is consistent across tissues*

Using these models to predict the efficiency of NMD and the RNA-sequencing data to verify the effects, we were able to discern which PTVs showed unpredicted allelic balance—that is, PTVs which are predicted to undergo NMD and are therefore expected to show allelic imbalance but instead show allelic balance. Of the 2,320 rare variants, we found 23.4% (543) variants that showed unpredicted allelic balance at least once, with 7.5% (173) variants showing unpredicted allelic balance in more than 90% of observations.

One of the most consistent unpredicted allelic balance variants is rs141826798 in *EGFL8* for which we observe the variant in 7 individuals and 44 different tissues for 131 total observations and 98% (128 of the 131) of the observations have a proportion of reference reads below 0.65 (Figure 4A). Inspection of the PTV location shows that it is more than 150 nt from start, not in a long exon, or not within 50 nt before the last exon junction (Figure 4B) and is not associated with any aberrant splicing events. The variant, rs141826798, was further previously identified as a risk variant for psoriasis in a UK BioBank genome-wide association study.22 This further suggests that identification of unpredicted allelic balance can be identified empirically by performing RNA sequencing from a patient of a readily accessible tissue or cell line.

In order to determine whether utilizing cell lines or readily accessible tissues could provide improved accuracy for determining NMD in all tissues, we looked at the similarity between the allelic imbalance of variants expressed in an easily accessible tissue and cell line (whole blood and fibroblasts) and all other tissues. The category determined by the majority of observations of a variant in fibroblasts and whole blood was a marginally better predictor than our best predictive model using only genomic data. Using the variants observed in fibroblasts to predict the categorical outcome of the variants in other tissues, we found a balanced accuracy of 0.767 as compared to 0.728 for the prediction for the same variants from our genomic model. The balanced accuracy of the predictions derived from blood was 0.745 as compared to 0.734 for the same variants using the predictions from our genomic model (Figure 4C). We found a similar marginal improvement when using Cohen’s Kappa to measure reliability (0.539 versus 0.458 for fibroblast and 0.496 versus 0.465 for whole blood, Figure S9).

Because tissue and cell line observations provided additional information for cross-tissue NMD predictions, we wanted to analyze the pathogenicity of the variants for which predicted and observed ASE classification differed. We observed that imbalanced variants were more likely to be annotated as pathogenic in ClinVar (Figure 4D, p < 0.0001 Fisher’s exact test), possibly because GTEx individuals were not selected based on a specific disease phenotype and many imbalanced pathogenic variants are likely recessive.

**Discussion**

We analyzed rare protein-truncating variants (PTVs) across individuals and tissues using GTEx v.8 project data. Using RNA-seq-based measurements of allelic imbalance of PTVs as a measure for NMD efficiency, we observed that, in addition to the position of the variant in the transcript, both allele frequency and conservation were predictive of NMD efficiency. Previous studies have demonstrated increased allelic imbalance of rare versus common PTVs.7,20,23 We observed that these effects did not plateau in the rare portion of the allele frequency spectrum; ultra-rare (MAF < 0.001%) and novel PTVs showed evidence of increased NMD efficiency. Additionally, we observed that GC content impacted NMD efficiency, suggesting an additional role of RNA structure. By combining these factors, we were able to improve
our ability to predict whether a PTV would show allelic imbalance beyond the 50 nt rule.

Strikingly, NMD efficiency is highly consistent across tissues and individuals, indicating the fundamental importance of this cellular machinery. This is consistent with previous observations of when the NMD machinery fails: individuals with mutations in one of the key NMD factors, UPF3B, show severe intellectual disability and the variant persists in these families only because it is X-linked.\textsuperscript{24,25} As the GTEx individuals were not selected for any phenotypic abnormality, we did not observe any missense or nonsense mutations in the core NMD proteins which may have otherwise provided variation in NMD efficiency between individuals. The relative lack of variation between tissues may be attributed to a finely tuned autoregulatory feedback loop as several of the core NMD proteins are known to be upregulated when NMD is inhibited.\textsuperscript{26,27} In Rivas et al.,\textsuperscript{7} tissue identity was not directly tested as a predictor of NMD, but it was observed that among variants predicted to undergo NMD based on the 50 nt rule, there was evidence of heterogeneity of tissue effects in 17\% of variants tested. While we also observed suggestive evidence of tissue heterogeneity, we did not find that this heterogeneity could be used to improve predictions of NMD.

For classifying novel variants, especially for rare disease diagnostic purposes, it is very promising that NMD is consistent across tissues, age, and sex. The high tissue-sharing of NMD efficiency further indicates that potential gain-of-function effects of NMD escapees, such as those reported by Coban-Akdemir et al.,\textsuperscript{4} are unlikely to manifest in a single tissue when the target gene is expressed across multiple tissues.\textsuperscript{4} This provides confidence that tissue-agnostic predictive tools such as NMDetective\textsuperscript{B8} provide equal predictive power regardless of the tissue in which a gene is expressed. Further, future studies may benefit from testing in easily biopsied tissues or synthetically testing PTVs in cell lines. 70\% of disease-associated genes from the Online Mendelian Inheritance in Man database are expressed in blood, allowing for their measurement, although this does not account for potential tissue-specific splicing.\textsuperscript{28} This is especially valuable given the importance of collecting high-quality RNA with high coverage at the site of interest. Given datasets like GTEx, it is possible to assess the degradation of many rare variants for appropriate classification without further experiments. To this end, we provide the classification for rare variants identified in this study (Table S3) for future research.

**Data and code availability**

GTEx data are available at dbGaP under the accession phs000424.v8.p2 and the GTEx Portal https://gtexportal.org/home/.

**Supplemental information**

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2021.06.008.

**Acknowledgments**

N.A.T. is supported by NIH grants DK107437, HL142015, and DK112348 and the Stanford School of Medicine Department of Pathology. D.N. is supported by 1T32AG047126-01. T.E. is supported

Figure 4. Additional disease-relevant information may be gathered by analyzing readily available tissues
(A) Proportion of reference reads against the total number of reads covering the variant for predicted NMD targets (light purple fill) and variants predicted to escape NMD (light blue fill). The variant rs141826798 in Epidermal Growth Factor Like Domain Multiple 8 (EGFL8), which has been implicated in psoriasis, is highlighted with a navy outline.
(B) The premature termination variant rs141826798 in EGFL8 occurs 3 nt from the end of exon 4. It is not in a long (>407 nt) exon, proximal to the start of the gene (within 150 nt of start), in the last exon, or 50 nt before the last exon junction.
(C) Balanced accuracy of the predictive ability in all other tissues of variants observed in whole blood or fibroblasts using our best predictive model utilizing genomic annotations (pink) or the classification called by the majority of observations across individuals in the indicated tissue (green).
(D) Proportion of variants observed in fibroblasts and blood that are pathogenic (dark purple) or benign (dark blue) as determined by ClinVar. Brackets show results of Fisher’s exact test. ****p < 0.00001.
by NLM training grant LM007033. N.M.E. is supported by a National Science Foundation Graduate Research Fellowship, grant no. DGE–1656518, and a graduate fellowship from the Stanford Center for Computational, Evolutionary and Human Genomics. C.S. is supported by NIH grant T32LM121409. M.A.R. is in part supported by the NHGRI of the NIH under award R01HG010140 (M.A.R.) and an NIH Center for Multi- and Trans-ethnic Mapping of Mendelian and Complex Diseases grant (SU01HG009080). S.B.M. is supported by NIH grants R01AG066490, R01MH125244, U01HG009431, R01HL142015, R01HG008150, and U01HG009080. This work in part used supercomputing resources provided by the Stanford Genomics Bioinformatics Service Center, supported by National Institutes of Health S10 Instrumentation Grant S10OD023452. Thanks to the members of the Montgomery Lab, the Rivas Lab, and Sarah Teran for their support and critical feedback in the preparation of this manuscript. The authors thank the GTEx and UKBB participants and their families.

Declaration of interests

M.A.R. is on the SAB of 54 gene and Related Sciences and has advised BioMarin, Third Rock Ventures, and MazeTx. S.B.M. is on the SAB of Myome Inc.

Received: February 4, 2021
Accepted: June 9, 2021
Published: July 2, 2012

Web resources

ClinVar data (September 6, 2020), https://ftp.ncbi.nlm.nih.gov/pub/clinvar/tab_delimited/variant_summary.txt.gz
Code used to perform these analyses, https://github.com/smontgomlab/nmd_2021/

References

1. Pai, A.A., Cain, C.E., Mizrahi-Man, O., De Leon, S., Lewellen, N., Veyrieras, J.-B., Degner, J.F., Gaffney, D.J., Pickrell, J.K., Stephens, M., et al. (2012). The contribution of RNA variation to trait loci to inter-individual variation in steady-state gene expression levels. PLoS Genet. 8, e1003000.
2. Kurosaaki, T., Popp, M.W., and Maquat, L.E. (2019). Quality and quantity control of gene expression by nonsense-mediated mRNA decay. Nat. Rev. Mol. Cell Biol. 20, 406–420.
3. Khajavi, M., Inoue, K., and Lupski, J.R. (2006). Nonsense-mediated mRNA decay modulates clinical outcome of genetic disease. Eur. J. Hum. Genet. 14, 1074–1081.
4. Coban-Akdemir, Z., White, J.J., Song, X., Jhangiani, S.N., Faith, J.M., Gambin, T., Bayram, Y., Chinn, I.K., Karaca, E., Punetha, J., et al.; Baylor-Hopkins Center for Mendelian Genomics (2018). Identifying Genes Whose Mutant Transcripts Cause Dominant Disease Traits by Potential Gain-of-Function Alleles. Am. J. Hum. Genet. 103, 171–187.
5. Nagy, E., and Maquat, L.E. (1998). A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. Trends Biochem. Sci. 23, 198–199.
6. Lindeboom, R.G.H., Supek, F., and Lehner, B. (2016). The rules and impact of nonsense-mediated mRNA decay in human cancers. Nat. Genet. 48, 1112–1118.
7. Rivas, M.A., Pirinen, M., Conrad, D.F., Lek, M., Tsang, E.K., Karczewski, K.J., Mallor, J.B., Kukurba, K.R., DeLuca, D.S., Fromer, M., et al.; GTEx Consortium; and Geuvadis Consortium (2015). Human genomics. Effect of predicted protein-truncating genetic variants on the human transcriptome. Science 348, 666–669.
8. Lindeboom, R.G.H., Vermeulen, M., Lehner, B., and Supek, F. (2019). The impact of nonsense-mediated mRNA decay on genetic disease, gene editing and cancer immunotherapy. Nat. Genet. 51, 1645–1651.
9. Zetoune, A.B., Fontanière, S., Magnin, D., Anczuków, O., Busson, M., Zhang, C.X., and Mazoyer, S. (2008). Comparison of nonsense-mediated mRNA decay efficiency in various murine tissues. BMC Genet. 9, 83.
10. GTEx Consortium (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science 369, 1318–1330.
11. van de Geijn, B., McVicker, G., Gilad, Y., and Pritchard, J.K. (2015). WASP: allele-specific software for robust molecular quantitative trait locus discovery. Nat. Methods 12, 1061–1063.
12. Castel, S.E., Levy-Moonshine, A., Mohammadi, P., Banks, E., and Lappalainen, T. (2015). Tools and best practices for data processing in allelic expression analysis. Genome Biol. 16, 195.
13. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Larichchia, K.M., Ganna, A., Birnbaum, D.P.; et al.; Genome Aggregation Database Consortium (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 581, 434–443.
14. Kircher, M., Witten, D.M., Jain, P., O’Roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. Nat. Genet. 46, 310–315.
15. Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J., and Kircher, M. (2019). CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res. 47 (D1), D886–D894.
16. Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. J. Stat. Softw. 33, 1–22.
17. Smithson, M., and Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. Psychol. Methods 11, 54–71.
18. Tibshirani, R.J., Taylor, J., Lockhart, R., and Tibshirani, R. (2016). Exact Post-Selection Inference for Sequential Regression Procedures. J. Am. Stat. Assoc. 111, 600–620.
19. Carvill, G.L., and Mefford, H.C. (2020). Poisson exons in neurodevelopment and disease. Curr. Opin. Genet. Dev. 65, 98–102.
20. Kukurba, K.R., Zhang, R., Li, X., Smith, K.S., Knowles, D.A., How Tan, M., Piskol, R., Lek, M., Snyder, M., Macarthur, D.G., et al. (2014). Allelic expression of deleterious protein-coding variants across human tissues. PLoS Genet. 10, e1004304.
21. Popp, M.W., and Maquat, L.E. (2016). Leveraging Rules of Nonsense-Mediated mRNA Decay for Genome Engineering and Personalized Medicine. Cell 165, 1319–1322.
22. Emdin, C.A., Khera, A.V., Chaffin, M., Klarin, D., Natarajan, P., Aragam, K., Haas, M., Bick, A., Zekavat, S.M., Nomura, A., et al. (2018). Analysis of predicted loss-of-function variants in UK Biobank identifies variants protective for disease. Nat. Commun. 9, 1613.
23. Lappalainen, T., Sammeth, M., Friedländer, M.R., ’t Hoen, P.A.C., Monlong, J., Rivas, M.A., González-Porta, M., Kurbatova, N., Griebel, T., Ferreira, P.G., et al.; Geuvadis Consortium (2013). Transcriptome and genome sequencing uncovers functional variation in humans. Nature 501, 506–511.

24. Laumonnier, F., Shoubridge, C., Antar, C., Nguyen, L.S., Van Esch, H., Kleeftstra, T., Briault, S., Fryns, J.P., Hamel, B., Chelly, J., et al. (2010). Mutations of the UPF3B gene, which encodes a protein widely expressed in neurons, are associated with nonspecific mental retardation with or without autism. Mol. Psychiatry 15, 767–776.

25. Tejada, M.I., Villate, O., Ibarluzea, N., de la Hoz, A.B., Martínez-Bouzas, C., Beristain, E., Martínez, F., Friez, M.J., Sobrino, B., and Barros, F. (2019). Molecular and Clinical Characterization of a Novel Nonsense Variant in Exon 1 of the UPF3B Gene Found in a Large Spanish Basque Family (MRX82). Front. Genet. 10, 1074.

26. Yepiskoposyan, H., Aeschimann, F., Nilsson, D., Okoniewski, M., and Mühlemann, O. (2011). Autoregulation of the nonsense-mediated mRNA decay pathway in human cells. RNA 17, 2108–2118.

27. Huang, L., Lou, C.-H., Chan, W., Shum, E.Y., Shao, A., Stone, E., Karam, R., Song, H.-W., and Wilkinson, M.F. (2011). RNA homeostasis governed by cell type-specific and branched feedback loops acting on NMD. Mol. Cell 43, 950–961.

28. Frévard, L., Smail, C., Ferraro, N.M., Teran, N.A., Li, X., Smith, K.S., Bonner, D., Kernohan, K.D., Marwaha, S., Zappala, Z., et al.; Undiagnosed Diseases Network; and Care4Rare Canada Consortium (2019). Identification of rare-disease genes using blood transcriptome sequencing and large control cohorts. Nat. Med. 25, 911–919.