Mapping genetic variants for nonsense-mediated mRNA decay regulation across human tissues

Bo Sun[1] and Liang Chen[1]*

*Correspondence:
liang.chen@usc.edu

[1] Department of Quantitative and Computational Biology, University of Southern California, 1050 Childs Way, Los Angeles, CA 90089, USA

Abstract

Background: Nonsense-mediated mRNA decay (NMD) was originally conceived as an mRNA surveillance mechanism to prevent the production of potentially deleterious truncated proteins. Research also shows NMD is an important post-transcriptional gene regulation mechanism selectively targeting many non-aberrant mRNAs. However, how natural genetic variants affect NMD and modulate gene expression remains elusive.

Results: Here we elucidate NMD regulation of individual genes across human tissues through genetical genomics. Genetic variants corresponding to NMD regulation are identified based on GTEx data through unique and robust transcript expression modeling. We identify genetic variants that influence the percentage of NMD-targeted transcripts (pNMD-QTLs), as well as genetic variants regulating the decay efficiency of NMD-targeted transcripts (dNMD-QTLs). Many such variants are missed in traditional expression quantitative trait locus (eQTL) mapping. NMD-QTLs show strong tissue specificity especially in the brain. They are more likely to overlap with disease single-nucleotide polymorphisms (SNPs). Compared to eQTLs, NMD-QTLs are more likely to be located within gene bodies and exons, especially the penultimate exons from the 3' end. Furthermore, NMD-QTLs are more likely to be found in the binding sites of miRNAs and RNA binding proteins.

Conclusions: We reveal the genome-wide landscape of genetic variants associated with NMD regulation across human tissues. Our analysis results indicate important roles of NMD in the brain. The preferential genomic positions of NMD-QTLs suggest key attributes for NMD regulation. Furthermore, the overlap with disease-associated SNPs and post-transcriptional regulatory elements implicates regulatory roles of NMD-QTLs in disease manifestation and their interactions with other post-transcriptional regulators.

Keywords: Nonsense-mediated mRNA decay, Genetic variants, Genetical genomics, GTEx, NMD-QTL, Post-transcriptional regulation, RNA binding proteins, miRNA, Disease SNPs, Decay efficiency

Background

Nonsense-mediated mRNA decay (NMD) is an important post-transcriptional regulatory mechanism defining gene expression. It was originally conceived as an mRNA surveillance mechanism that recognizes and degrades transcripts harboring a premature termination codon (PTC) to prevent the production of potentially deleterious truncated proteins [1, 2]. These mRNAs with PTCs originate from various sources (e.g., nonsense mutations, aberrant alternative splicing, and DNA rearrangements) [3]. The broader impact of NMD has been revealed in genetic disease, gene editing, and cancer immunotherapy [4]. Nonsense mutations account for 20.3% of all disease-associated single base-pair mutations and many of them introduce PTCs [5]. While NMD is a protective mechanism against the production of C-terminal truncated proteins, NMD can either aggravate or alleviate the effects of those PTCs that cause genetic diseases [6].

Research also shows NMD is an important post-transcriptional gene regulation mechanism targeting many non-aberrant mRNAs [7–11]. Transcriptome-wide studies suggest NMD accounts for autoregulation of 3–10% of normal transcripts in different cell types [12–14]. Our own analysis of the GENCODE annotation showed that about 12% of natural transcripts are presumably targeted by NMD regulation, since they follow the 50-nt rule (i.e., the presence of a terminal codon > 50 nt upstream of the last exon-exon junction). Accordingly, NMD factors are implicated in cellular homeostasis, stress response, cell cycle, differentiation, development, neural activity, immunity, and spermatogenesis [15–24]. NMD typically downregulates its substrate by 2–20 folds [25, 26], generating significant functional and clinical implications. Although the overall degradation process is believed to be similar for both erroneous and normal transcripts, factors involved in NMD recognition and regulation of degradation efficiency are poorly understood.

Variations in NMD efficiency are observed across tissues and individuals [27, 28]. Such variations can influence the outcome of PTC read-through therapeutic strategies aimed at suppressing nonsense mutations to restore full-length proteins, as the efficacy of such nonsense suppression largely depends on NMD efficiency (reviewed in [29]). Thus, a thorough examination of natural genetic variants affecting NMD efficiency is essential in understanding not only NMD regulation but also phenotypic variability and different degrees of expressivity manifested in various nonsense-mutation-triggered diseases.

Genetic variants associated with changes in gene expression (expression quantitative trait loci, eQTLs) are being cataloged for human tissues on an unprecedented scale (e.g., the Genotype-Tissue Expression (GTEx) project) [30, 31]. Genetical genomics has been focused on identifying genetic loci linked to variation(s) in mRNA expression levels. Recent studies have extended the analysis to identify genetic variants associated with other molecular phenotypes (e.g., ribosome occupancy, protein abundance, allele-specific expression, and alternative splicing) [32–34]. eQTLs mapped to NMD factors (e.g., *SMG7*, *UPF3B*, *MAGOH*, and so on) provide valuable information about how genetic variants may affect the NMD pathway through their cis-acting regulation on NMD factors. Several studies examined how loss-of-function variants and protein-truncating variants affect their own gene expression, which implicated the involvement of surveillance NMD [35–37]. However, no systematic survey of *cis* genetic variants on individual NMD substrates and their impact on NMD regulation has been done.

Here we utilize the GTEx resources and build a genome-wide landscape of tissue-specific genetic variants involved in NMD regulations. To distinguish the NMD effect on steady-state mRNA levels from that of transcriptional regulation, we built a novel and robust statistical model. We focused on natural mRNAs targeted by NMD according to the 50-nt rule instead of aberrant mRNAs caused by DNA mutations or RNA processing errors. We identified natural genetic variants controlling the percentage of NMD-targeted transcripts, and variants regulating the decay efficiency of NMD-targeted transcripts.

Results

Modeling of NMD-QTLs

We propose a statistical model to identify and characterize genetic variants associated with NMD and apply it to human tissue data (Fig. 1a). The mRNA transcripts of a gene were divided into two categories: NMD-targeted and non-NMD transcripts, based on whether they were annotated by the "nonsense_mediated_decay" tag in the GENCODE annotation. Genes with both NMD-targeted and non-NMD transcripts were referred as NMD genes. We used α and θ to parameterize NMD regulation: α represents the percent of transcribed transcripts targeted by NMD, and θ represents the percent of
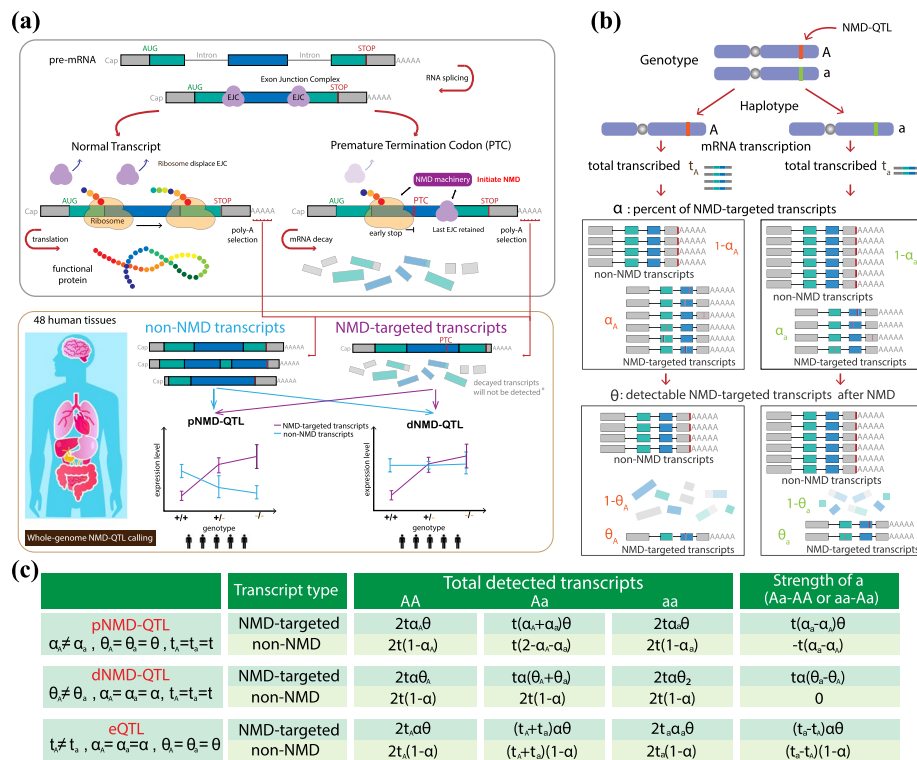
Fig. 1 Modeling of NMD-QTLs. **a** Identification of NMD-QTLs across human tissues based on the GTEx data. **b** Parameterization of the expression of NMD genes. *t* represents the total amount of transcribed transcripts, *α* represents the percentage of transcribed transcripts targeted by NMD, and *θ* represents the percentage of NMD-targeted transcripts that remained after degradation. The parameters can be different for different alleles of a genetic variant. **c** Allelic effect sizes of a pNMD-QTL, dNMD-QTL, or eQTL for NMD-targeted transcripts and non-NMD transcripts

NMD-targeted transcripts remaining after degradation. As illustrated in Fig. 1b, gene haplotypes bearing different alleles ("A" or "a") may generate different amounts of transcribed mRNAs (t_A or t_a), different percentages of transcribed mRNAs (α_A or α_a) may be targeted by NMD regulation, and these NMD-targeted transcripts may be degraded at different efficiencies ($1 - \theta_A$ or $1 - \theta_a$). Note that degradation efficiency is relative to the regular degradation of non-NMD transcripts. Generally, genetic variants identified from a regular eQTL analysis are those with allele-specific transcription levels (i.e., $t_A \neq t_a$) without taking into consideration NMD-transcripts (α) and their degradation ($1 - \theta$). Here, we aim to capture genetic variants resulting in different α or θ values. The former, called pNMD-QTLs, controls the different percentages of NMD-targeted transcript isoforms ($\alpha_A \neq \alpha_a$). The latter, called dNMD-QTLs, regulates the *d*egradation efficiency of NMD-targeted isoforms ($\theta_A \neq \theta_a$). Herein they are jointly referred to as NMD-QTLs.

As shown in Fig. 1c, pNMD-QTLs and dNMD-QTLs can be identified and distinguished from regular eQTLs by separately examining their allelic effect sizes (i.e., the magnitude of the effect of an allele has on the expression level, or slope estimations of the linear regression models) for NMD-targeted transcripts and non-NMD transcripts. For a pNMD-QTL, the allelic effect sizes are in the opposite direction of NMD-targeted and non-NMD transcripts, and the observed regulatory strength is smaller for NMD-targeted transcripts as fewer are detected after degradation. For a dNMD-QTL, the significant association is only observable in NMD-targeted transcripts. Examples of a pNMD-QTL and a dNMD-QTL are shown in Fig. 2a. A positive effect size is observed for the alternative allele of the pNMD-QTL when considering NMD-targeted transcripts (i.e., higher transcript amount when an individual contains more copies of the alternative allele). However, the allelic effect size is negative for non-NMD transcripts. For the dNMD-QTL, the copy number of the alternative allele only affects the amounts of NMD-targeted transcripts.

Evaluation of NMD-QTL Modeling

To evaluate the performance of our models, we estimated precision and recall for the identification of pNMD-QTLs and dNMD-QTLs from simulations. A total of 10,000 genetic variants were simulated; 5% were pNMD-QTLs ($\alpha_A \neq \alpha_a$, $\theta_A = \theta_a$, $t_A = t_a$), 5% were dNMD-QTLs ($\alpha_A = \alpha_a$, $\theta_A \neq \theta_a$, $t_A = t_a$), and the remainder (90%) were null cases for NMD-QTLs ($\alpha_A = \alpha_a$, $\theta_A = \theta_a$). The null cases were either regular eQTLs or variants not associated with expression depending on the randomly assigned differences between t_A and t_a. When the allelic effect size of NMD-QTLs was between 10 and 15% (i.e., $|\alpha_A - \alpha_a| = 0.1$–$0.15$ or $|\theta_A - \theta_a| = 0.1$–$0.15$), at a recall rate of 0.8, the precision was 0.93 and 0.84 for the identification of pNMD-QTLs and dNMD-QTLs, respectively (Fig. 2b). More detailed results and results for other simulations with smaller (5–10%) or larger (15–20%) allelic effect sizes can be found in Additional file 1: Fig. S1. The simulations verified we can distinguish NMD-QTLs from regular eQTLs by jointly considering the significance and directions of the allelic effect sizes for NMD-targeted and non-NMD transcripts. To ensure the accuracy of the effect size estimation, we applied an additional quantile regression model (see Methods) as quantile regression has the desired robustness to outliers and dropouts in transcriptomics data, and it significantly improves eQTL
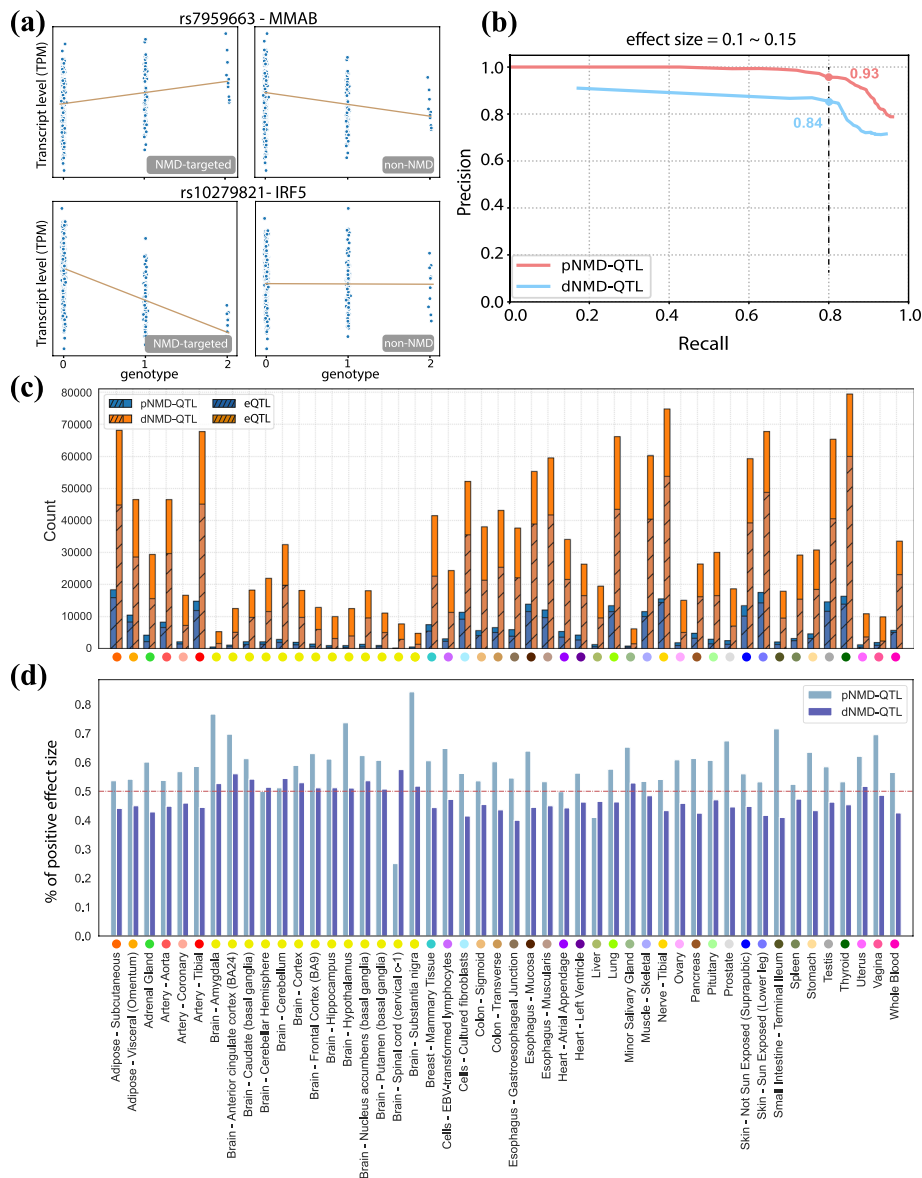
Fig. 2 Identification of pNMD-QTLs and dNMD-QTLs. **a** Examples of a pNMD-QTL and a dNMD-QTL. A variant is identified as a pNMD-QTL if the allelic effect size is in opposite directions for NMD-targeted transcripts and non-NMD transcripts. A variant is identified as a dNMD-QTL if the association is only observed in NMD-targeted transcripts. **b** Precision-recall curves for identifying pNMD-QTLs and dNMD-QTLs. The effect size was simulated between 0.1 and 0.15. **c** The number of pNMD-QTLs and dNMD-QTLs detected for each of the 48 tissues. The shaded areas represent the number of variants also identified as eQTLs in GTEx. The color scheme used for tissues is the same as the one used in GTEx. **d** The fractions of positive effect sizes of pNMD-QTLs and dNMD-QTLs for NMD-targeted transcripts in each tissue

mapping with an accurate effect size estimation [38], which enables us to better identify NMD-QTLs.

## Identification of NMD-QTLs across different human tissues

We identified pNMD-QTLs and dNMD-QTLs in *cis* for each of the 48 tissues in GTEx v7 with the sample size ≥ 70. On average, 6033 pNMD-QTLs and 34,049 dNMD-QTLs

were detected for a tissue (Fig. 2c). Usually more dNMD-QTLs were discovered than pNMD-QTLs (4–11 folds in different tissues). The thyroid and tibial nerves had the highest numbers of NMD-QTLs (99,053 and 95,803 respectively), while the minor salivary gland and vagina tissues had the lowest (7044 and 12,148 respectively). Worth noting is that GTEx, which models the overall expression as phenotypes without dissecting the NMD regulation, did not identify 7.3–93.0% of pNMD-QTLs and 24.4–76.0% of dNMD-QTLs as eQTLs (Fig. 2c, non-shaded areas). NMD-QTLs were identified for many well-known NMD targets. For example, we discovered 11 dNMD-QTLs for *BAG1*, ten dNMD-QTLs and one pNMD-QTL for *HNRNPL*, seven dNMD-QTLs for *GADD45B*, and three dNMD-QTLs for *SRSF10*.

To distinguish NMD-QTLs promoting or inhibiting NMD, we examined the effect size for NMD-targeted transcripts (Fig. 2d). Since the genotype was coded as the number of alternate non-reference alleles, a positive effect size suggests that $\alpha_{alt} > \alpha_{ref}$ (pNMD-QTLs) or $1 - \theta_{alt} < 1 - \theta_{ref}$ (dNMD-QTLs). An average of 59% pNMD-QTLs across tissues had positive signs. Thus, the alternate non-reference alleles promote the generation of NMD-targeted transcripts. Usually < 50% of dNMD-QTLs had positive signs (average: 47%). Thus, the non-reference alleles of these dNMD-QTLs inhibit the decay of NMD-targeted transcripts. Equivalently, an average of 53% dNMD-QTLs had negative signs and their non-reference alleles promote the decay of NMD-targeted transcripts.

Tissue specificity of NMD-QTLs

To examine the tissue similarity of NMD regulation, we considered a similarity score reflecting the enrichment of shared NMD-QTLs between two tissues (details in Methods). As shown in Fig. 3a, brain tissues had unique lists of NMD-QTLs: the NMD-QTL
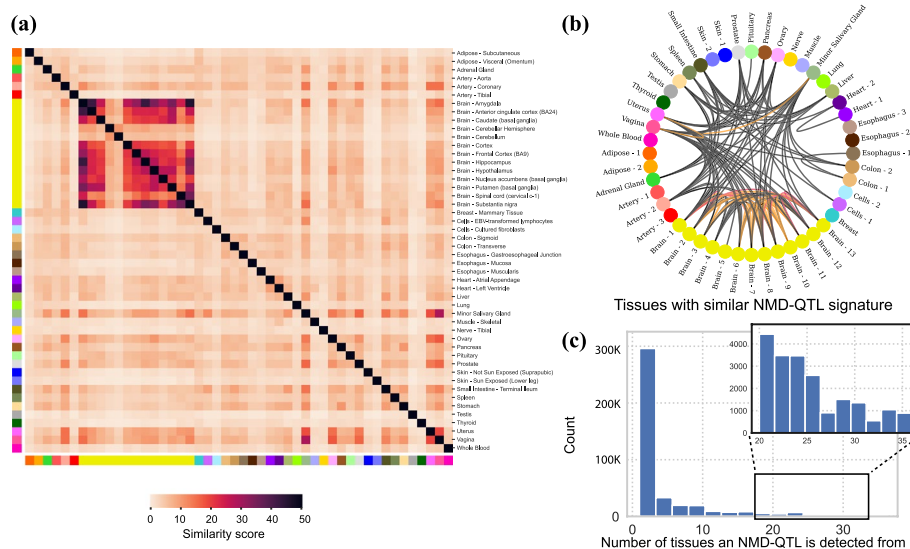
Fig. 3 Tissue specificity of NMD-QTLs. **a** The similarity of NMD-QTL signatures between tissue pairs. The similarity score s* reflects the enrichment of shared NMD-QTLs among different tissues. The detected NMD-QTLs from brain tissues are very distinctive from other tissue. **b** Tissue interaction map. No edge is drawn if the similarity score s* between the two tissues < 10. Black edges: 10 ≤ s* < 20; orange edges: 20 ≤ s* < 30; and red edges: 30 ≤ s*. **c** The number of tissues in which an NMD-QTL variant is detected. Brain tissues are combined as one tissue

similarities were generally high among different brain tissues (average of 24.7), but the average similarity between brain and any other tissues was as low as 6.6. This suggests that NMD-QTLs were shared within different brain sub-regions but were substantially different from those in other tissues.

The tissue interaction map based on the similarities of NMD-QTLs (Fig. 3b) clearly shows the distinction of brain tissues as well as the connections among major viscera (small intestine, spleen, stomach) and reproductive organs (uterus, vagina, ovary). The strong tissue specificity of NMD-QTLs is also supported by the observation that 54.3% of NMD-QTLs were discovered only in a single tissue (Fig. 3c, n.b. brain sub-regions were combined as one tissue). Taken together, NMD-QTLs exist as ubiquitously across the human body as NMD regulation itself, and many demonstrate strong tissue specificity.

A handful of NMD-QTLs were discovered in all tissues (Fig. 3c). These are located in important genes such as *TP53I13* (a tumor suppressor gene), *ST7L* (homologous to tumor suppressor gene *ST7*), *NDUFAF7* (an enzyme involved in the assembly and stabilization of Complex I, an important complex for mitochondrial respiratory chain), *IRF5* (belongs to a transcription factor family with diverse roles including virus-mediated activation of interferon, cell growth, differentiation, apoptosis, and immune system activity), *MMAB* (an enzyme that catalyzes the final step in the conversion of vitamin B_{12} into adenosylcobalamin), and three ribosomal proteins (S19, L13, L27a). It is known that NMD has high inter-individual variability and closely affects the clinical outcome for genetic disease [39, 40]. Therefore, these natural variations of NMD-QTLs impacting tumor suppressors and important metabolic genes across all tissues may be exploited to enhance personalized cancer immunotherapy and to treat a wide range of genetic diseases.

Remarkably, we found the direction of the allelic effect to be highly consistent across tissues for NMD-QTLs identified in many tissues. Among 62.9% (1840 out of 2923) of the pNMD-QTLs identified in more than 20 tissues, the sign of allelic effect was the same across tissues. Thus, the allelic effect sizes of a particular pNMD-QTL were all negative or all positive in the tissues that it was detected. For dNMD-QTLs identified in more than 20 tissues, the allelic effect was even more consistent across tissues. About 82.1% (14,547 out of 17,709) of them exhibited the same sign across all the tissues that they were detected.

Overlap of NMD-QTLs and disease SNPs

Functional variants are well-known to play an important role in disease etiology. Utilizing the collection of disease-related SNPs in DisGeNET [41], we explored the overlap of NMD-QTLs and disease-related SNPs. Specifically, out of the 54,643 non-redundant pNMD-QTLs and the 319,608 non-redundant dNMD-QTLs, 1151 (2.10%) and 3952 (1.24%) of them were found to be disease SNPs reported in DisGeNET, respectively (Fig. 4a). Compared to regular eQTLs discovered in GTEx for NMD genes (0.82% were found to be disease SNPs), NMD-QTLs were much more likely to overlap with disease SNPs: the odds ratio was 2.54 for pNMD-QLTs vs. eQTLs (*p*-value of the Fisher's exact test: 8.5×10^{-154}); and 1.49 for dNMD-QTLs vs. eQTLs (*p*-value of the Fisher's exact test: 4.2×10^{-94}).
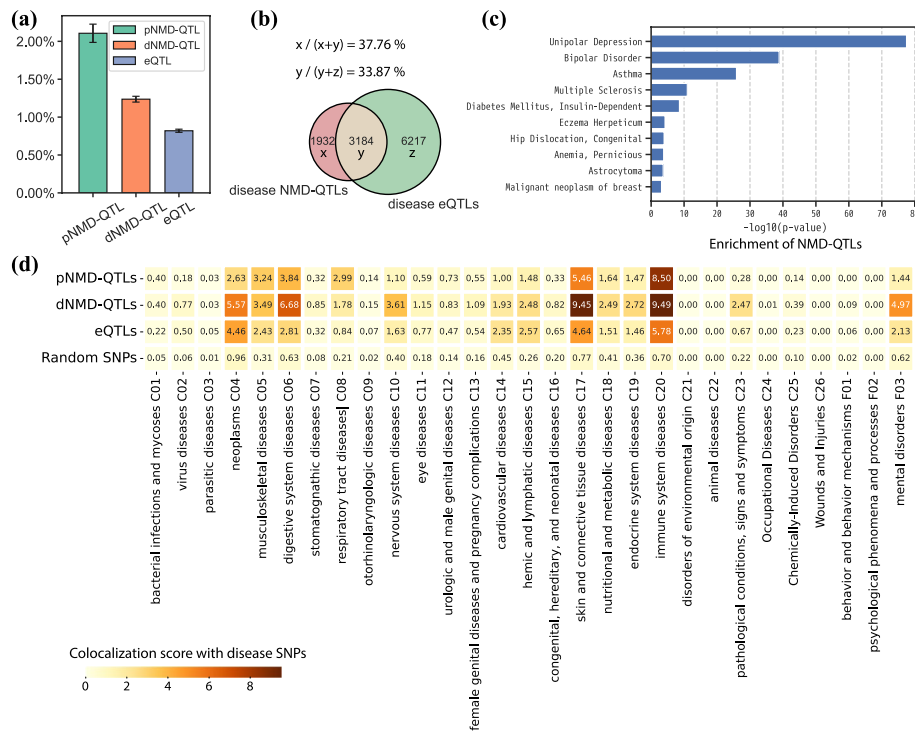
Fig. 4 Variant-disease association analysis for NMD-QTLs. **a** Percentages of pNMD-QTLs, dNMD-QTLs, and eQTLs overlapping with disease SNPs. **b** Venn diagram among disease-associated NMD-QTLs and disease-associated eQTLs. **c** Top 10 diseases associated with enriched NMD-QTLs. They are ranked by the p values of the proportion tests. **d** Overlap with disease SNPs under each MeSH term. The overlap scores are compared among pNMD-QTLs, dNMD-QTLs, eQTLs, and randomly sampled SNPs (the number is the same as that of NMD-QTLs)

We observed a tendency for NMD-QTLs to be located in closer proximity to genes' TSS compared to eQTLs, with a notable enrichment within a 500-kbp region of the TSS (Additional file 1: Figs. S2a−b). Considering that GWAS SNPs also tend to be located near the gene's TSS [42], we investigated whether the enrichment of disease-associated NMD-QTLs could be attributed to their location preferences. To examine this, we categorized NMD-QTLs and eQTLs into three groups based on their distance from the TSS: < 100 k, 100−500 k, and 500−1 M. It was consistently observed that NMD-QTLs showed a higher likelihood of overlapping with disease SNPs (Additional file 1: Figs. S2c−e), although the enrichment signal for dNMD-QTLs was slightly weaker in the > 500 k group (Additional file 1: Fig. S2e).

Although the total number of non-redundant NMD-QTLs was only 11.53% of the total number of non-redundant eQTLs (331,233 vs. 2,872,430 for NMD genes), we also discovered 33.87% disease eQTLs acting as NMD-QTLs (Fig. 4b). More importantly, 37.76% disease NMD-QTLs were not detected as eQTLs (Fig. 4b). A total of 737 diseases were associated with NMD-QTLs, suggesting that NMD regulation may be an important component in maintaining the normal function of cells.

Among the 8073 total diseases considered in DisGeNET, we identified 43 with prominent NMD-QTLs enrichment (*p*-values < 0.05 for both the proportion test and the hypergeometric test, details in Additional file 1: Supplementary Texts).

Remarkably, 9 out of the 43 (20.9%) diseases were categorized as "mental disorders (F03)" under the "psychiatry and psychology (F)" category of the Medical Subject Headings (MeSH) system. The most significant was "unipolar depression", followed by "bipolar disorder". The top 10 diseases are shown in Fig. 4c and the complete list of 43 diseases is in Additional file 2: Table S1. Mutations in core NMD pathway genes are implicated in multiple mental disorders [43–46]. Together with our aforementioned findings that NMD-QTLs show special distinctiveness and tissue-specificity in brain tissues, the brain disease association further suggests the critical role of NMD regulation for brain function and malfunction.

The diseases in DisGeNet were categorized into 29 MeSH terms. We then studied the detailed overlap status of disease SNPs and NMD-QTLs for each specific MeSH term. The calculation of the overlap enrichment score is described in Methods. As shown in Fig. 4d, dNMD-QTLs were observed to be enriched in nervous system diseases (C10) and mental disorders (F03). The fold-change of the enrichment for dNMD-QTLs vs. eQTLs was 2.2 for C10 and 2.3 for F03. The fold change was as high as 9.0 for C10 and 8.0 for F03 when comparing dNMD-QTLs with randomly sampled SNPs. Contrariwise, pNMD-QTLs were less enriched than eQTLs for these two terms. But the fold change was still as high as 2.8 and 2.3, respectively, when compared to randomly selected SNPs. This suggests regulating the decay efficiency of NMD-targeted mRNA substrates could be an important part of NMD regulation.in the brain and neural systems. Additionally, dNMD-QTLs but not pNMD-QTLs were more enriched than eQTLs in terms such as neoplasm (C04) and pathological conditions (C23). Both pNMD-QTLs and dNMD-QTLs were more enriched than eQTLs for terms such as musculoskeletal diseases (C05), digestive system diseases (C06), respiratory tract diseases (C08), skin and connective tissue diseases (C17) and immune system diseases (C20).

Some NMD-QTLs were related to many diseases. For example, the pNMD-QTL SNP rs5443 was associated with 48 out of the total 8073 diseases. It acted as a pNMD-QTL to regulate the fraction of NMD-targeted transcripts of the DNA repair gene *XRCC3* in 18 tissues including 6 brain tissues. This implies that NMD regulation may interplay closely with DNA repair to ensure genome integrity and preventing genetic disease. The dNMD-QTL SNP rs1800629, regulating the degradation efficiency of NMD-transcripts of genes *HLA-C* and *ATP6V1G2-DDX39B* (read-through transcription between neighboring genes *ATP6V1G2* and *DDX39B*), was related to 58 diseases, 18 of which were related to neoplastic processes.

To further underscore the significance of NMD-QTLs in complex traits, we performed colocalization analysis between NMD-QTLs and GWAS signals using the ezQTL tool [47] (https://analysistools.cancer.gov/ezqtl/#/home). For example, we performed colocalization analysis between dNMD-QTLs derived from adipose (subcutaneous) tissue, targeting gene ENSG00000243696, and GWAS signals associated with body mass index (BMI) [48]. As shown in Additional file 1: Fig. S3, our analysis highlighted SNP rs6769789 as a potential causal NMD-QTL for BMI, with a posterior probability of 0.846 using the HyPrColoc [49] colocalization analysis (cutoff: 0.5), and a posterior probability of 0.023 using the eCAVIAR [50] colocalization analysis (cutoff: 0.01).
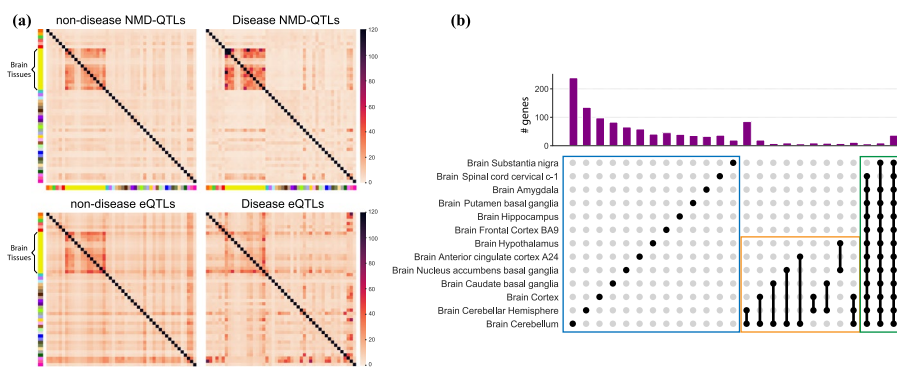
Disease-associated NMD-QTLs in brain tissues

Compared to non-disease NMD-QTLs, disease-associated NMD-QTLs showed an even stronger tissue specificity in brain sub-regions (Fig. 5a). By contrast, brain compartmentalization was weaker in disease eQTLs than in non-disease eQTLs. This reaffirms that NMD-QTLs play a critical role in maintaining neural system functionalities and can be valuable for examining brain and psychiatric pathogenesis.

We examined the sub-region membership of genes with at least one discovered NMD-QTL (i.e., the specific set of brain sub-regions from which NMD-QTLs were discovered for a given gene). The 25 most frequent brain sub-region memberships were displayed via an UpSet plot (Fig. 5b). The most frequent membership pattern appeared to be genes with NMD-QTLs identified in a single or a few sub-regions (blue and orange boxes in Fig. 5b), as well as genes with NMD-QTLs discovered in most brain sub-regions (green box in Fig. 5b).

NMD-QTLs are more likely to be located within gene bodies and exons compared to eQTLs

The 50-nt rule was applied to tag NMD-targeted transcripts in GENCODE because the rule has the strongest predictive value for NMD susceptibility [51]. In this study, NMD-targeted transcripts all follow this rule but the percentage of transcripts undergoing NMD and their decay efficiency can vary. The genomic positions of NMD-QTLs can inform additional determinants of NMD regulation.

First, we examined whether NMD-QTLs had any preference for residing inside a gene body or in the intergenic regions. For each gene, we counted the numbers of NMD-QTLs located within ($N_{i,g}$) or outside ($N_{o,g}$) of the gene body. Then a linear model was fit between $N_{i,g}$ and $N_{o,g}$ taking the gene length (l_g) into account: $N_{i,g} = w_0 + w_1 N_{o,g} + w_2 log_2(l_g)$. We modeled each tissue separately and compared the results for NMD-QTLs and eQTLs. As shown in Fig. 6a, the estimations of $\widehat{w}_1$ for NMD-QTLs were consistently larger than those for eQTLs, especially for brain tissues (Fig. 6c). This suggests that, compared to eQTLs, NMD-QTLs are more likely to function within a gene body than in intergenic regions. However, the estimations of $\widehat{w}_2$ for NMD-QTLs and eQTLs were similar (*p*-value = 0.37, Wilcoxon test, Fig. 6b). This suggests that the gene length factor was controlled at a similar level for both.

Fig. 5 Disease NMD-QTLs in brain tissues. **a** Tissue pair-wise similarity of non-disease and disease QTL signatures. Disease NMD-QTLs but not disease eQTLs are more similar in brain tissues. **b** UpSet plot showing the top 25 most frequent brain sub-region memberships displayed by genes with NMD-QTLs
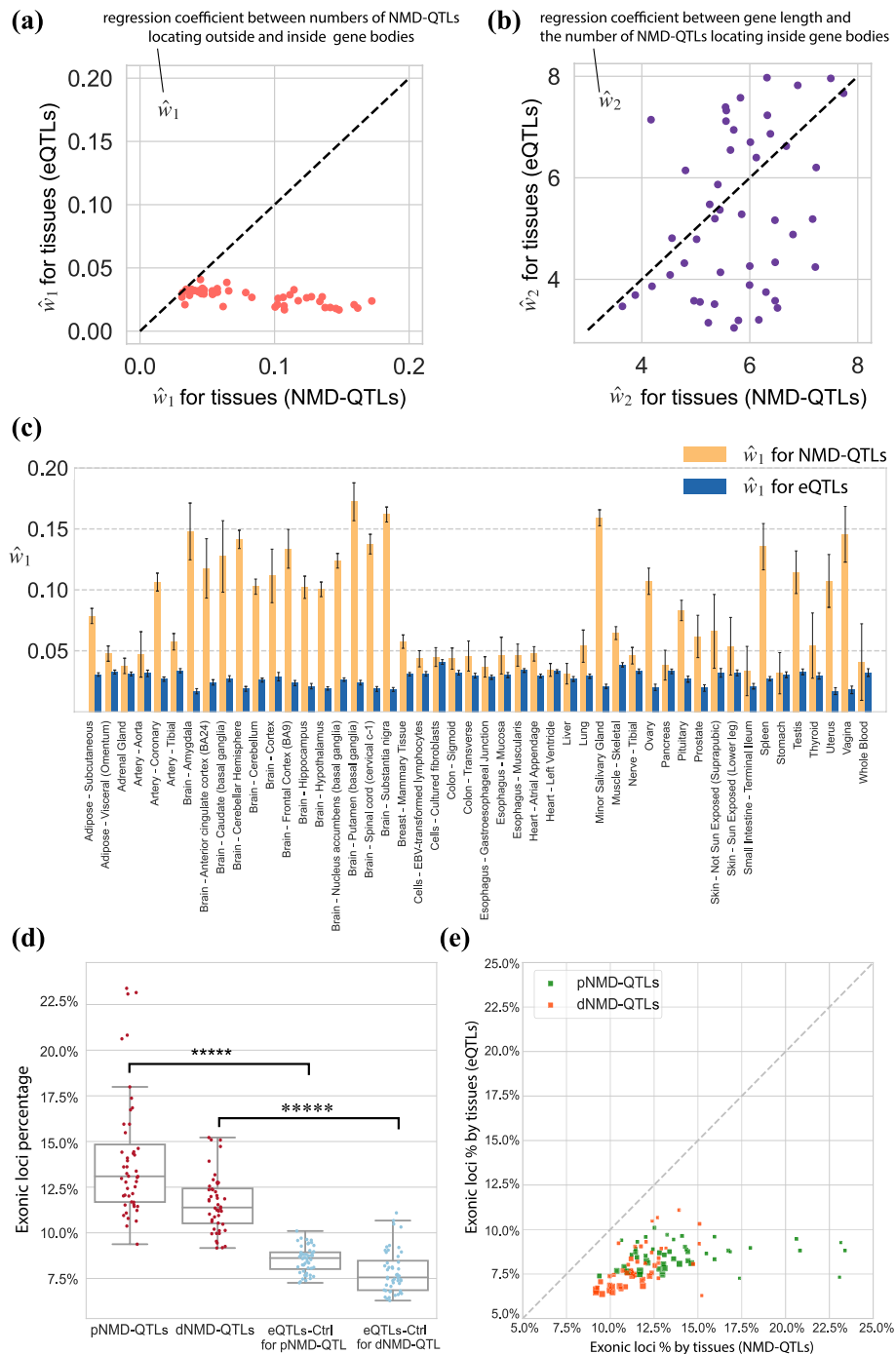
Fig. 6 Genomic locations of NMD-QTLs. **a** The coefficient estimations for $\widehat{w}_1$. **b** The coefficient estimations for $\widehat{w}_2$. $\widehat{w}_1$ and $\widehat{w}_2$ are for the model $N_{i,g} = w_0 + w_1 N_{o,g} + w_2 log_2(l_g)$, where $N_{i,g}$ is the number of NMD-QTLs (or eQTLs) inside a gene, $N_{o,g}$ is the number of NMD-QTLs (or eQTLs) outside of any gene, and l_g is the gene length. **c** The $\widehat{w}_1$ estimation for each tissue. Error bars: 95% confidence intervals. The $\widehat{w}_1$ estimations for NMD-QTLs are consistently larger than those for eQTLs, meaning that NMD-QTLs prefer locating inside a gene body. **d** The proportion of genic variants that are located on exons. The exonic proportions are significantly higher for NMD-QTLs, compared to eQTLs discovered for the same genes. **e** The scatter plots of the exonic proportions. The size of the symbol is proportional to the sample size for a specific tissue

For those NMD-QTLs residing within a gene body, we further examined whether they prefer exon or intron regions. Compared to eQTLs, NMD-QTLs were more likely to be in exonic positions (Fig. 6d, p-values from Wilcox tests $< 1 \times 10^{-6}$). The tissue-wise exonic proportions of NMD-QTLs and eQTLs are compared in Fig. 6e, further supporting the exonic location preference of NMD-QTLs. Regulatory elements for mRNA degeneration are usually exonic after mRNA splicing. In addition, exon-junction complex (EJC) has been reported to be involved in the initiation of NMD [52, 53]. It is possible that NMD-QTLs in exonic regions interact more effectively with the deposition of EJC to tag NMD regulation for natural transcripts.

Ordinal positions of NMD-QTLs along exons and introns

Since the location of EJC is essential for NMD regulation, we next studied the relative positions of NMD-QTLs within a gene body by defining an ordinal rank for exons and introns. For each gene, the combined exon intervals from the collapsed gene model (i.e., combining all isoforms of a single gene into a single transcript) used in GTEx were considered. The intron intervals are deduced from those exon intervals. We ranked these exon and intron intervals separately from both 5′-end and 3′-end, starting from 0. For all genes with NMD-QTLs, the frequency count and the median length of the exon (or intron) intervals on each ordinal position are shown in Additional file 1: Figs. S4a–b. We recorded exon (or intron) ordinal positions for all genic NMD-QTLs; the raw count distributions are shown in Additional file 1: Figs. S4c–d.

To remove counting bias, we normalized the raw count by the number of exon or intron intervals at each ordinal position (Fig. 7a, b), and further by the median length of exon or intron intervals at each ordinal position (Fig. 7c, d). Both pNMD-QTLs and dNMD-QTLs were more likely to be located in the last exon interval (rank 0 at 3′, Fig. 7a). This was partly due to the fact that the last exon interval is usually much longer than other exons (median length of 831 vs. 187). Taking the exon interval length into
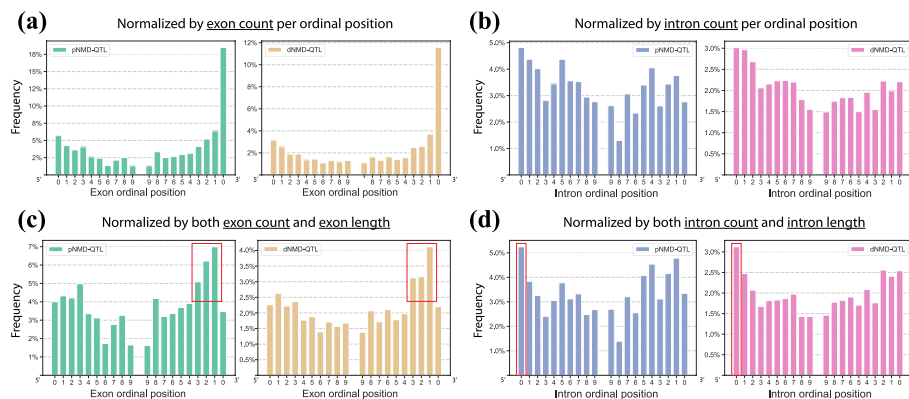
Fig. 7 Ordinal positioning of NMD-QTLs along exon and intron intervals. The exon and intron intervals are deduced from the collapsed gene model and ranked from 0 for both 5′ and 3′ ends. **a** Distribution of the count of NMD-QTLs on each exon ordinal position normalized by the exon count per ordinal position. **b** Distribution of the count of NMD-QTLs on each intron ordinal position normalized by the intron count per ordinal position. **c** Distribution of NMD-QTL counts on each exon ordinal position normalized by both the exon count and exon length per ordinal position. **d** Distribution of NMD-QTL counts on each intron ordinal position normalized by both the intron count and intron length per ordinal position

account, the chance of an NMD-QTL locating in an exonic position of the last exon interval (rank 0 at 3′) was actually much lower (Fig. 7c). The highest frequency was within the penultimate exon from the 3′ end (i.e., rank 1), followed by the two upstream exons (i.e., ranks 1–2). These 3′ exon ordinal positions may influence the formation of the last EJC or affect the interaction between PTC and downstream EJC. As regards the intron ordinal positions for NMD-QTLs, both pNMD-QTLs and dNMD-QTLs preferred the first intron interval (rank 0 at 5′, Fig. 7b, d).

Location preference of NMD-QTLs at the base-pair (bp) resolution

In addition to the ordinal positions along exons or introns, we examined the base-pair (bp) distance of NMD-QTLs to the exon or intron boundary. Both exonic NMD-QTLs and eQTLs favored the 0–50-bp regions close to exon boundaries (Fig. 8a for pNMD-QTLs and Additional file 1: Fig. S5a for dNMD-QTLs). On the other hand, intronic NMD-QTLs were more likely to be near intron boundaries than intronic eQTLs (Fig. 8b and Additional file 1: Fig. S5b).

The location of PTC is tightly related to NMD regulation. We therefore considered the distances between NMD-QTLs and PTCs of NMD-targeted transcripts (vs. regular stop codons of non-NMD transcripts). For exonic NMD-QTLs, we calculated the distances on the transcripts without considering intronic positions (Fig. 8c). Compared to eQTLs, pNMD-QTLs were more concentrated in the $(-75\text{ bp}, +150\text{ bp})$ regions around PTCs (orange line vs. gray line in the upper panel of Fig. 8c). Interestingly, pNMD-QTLs were also more concentrated around the $(-25\text{ bp}, +50\text{ bp})$ regions around regular stop codons than eQTLs (red line vs. light blue line). We also observed another pNMD-QTLs
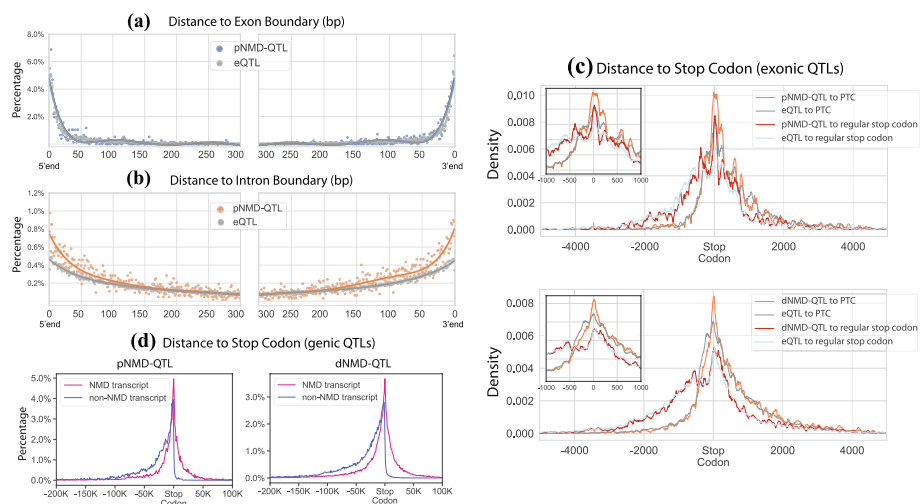
Fig. 8 Distances of NMD-QTLs to boundaries of exons, introns, and stop codons. **a** Distances to exon boundaries for variants located on an exon interval. The comparison is made between pNMD-QTLs and eQTLs. **b** Distances to intron boundaries for variants located on an intron interval. The comparison is made between pNMD-QTLs and eQTLs. Similar results for dNMD-QTLs can be found in Additional file 1: Fig. S5. **c** Distances of exonic NMD-QTLs to PTCs of NMD-targeted transcripts, and to regular stop codons of non-NMD transcripts. Only exonic positions are counted in the distance calculation. **d** Distance between genic NMD-QTLs and the stop codons. Both exonic and intronic positions are counted in the distance calculation. Note that the distance between an NMD-QTL (or an eQTL) and each transcript isoform's PTC (or regular stop codon) was calculated separately. All these distances are shown here

peak (red line) around 400 bp upstream of the regular stop codons. Since the median distance between the upstream PTC and the downstream regular stop codon of the same gene is 822 bp (Additional file 1: Fig. S6), the peak on -400 bp of regular stop codons usually corresponds to positions downstream of PTCs.

The exonic dNMD-QTLs were only slightly more concentrated around PTCs (orange line vs. gray line in the lower panel of Fig. 8c). Their distribution is almost indistinguishable around regular stop codons (red line vs. light blue line in the lower panel of Fig. 8c). We also considered the genomic distances (both exonic positions and intronic positions) between genic NMD-QTLs and PTCs (or regular stop codons). As shown in Fig. 8d, NMD-QTLs can be located upstream or downstream of PTCs but are rarely downstream of regular stop codons.

The relationship between NMD-QTLs and miRNAs or RNA-binding proteins (RBPs)

We wonder how these *cis*-acting NMD-QTLs interact with *trans*-acting factors to impact NMD regulation. We focused on miRNAs and RBPs. miRNAs recognize complementary mRNA molecules and function in RNA silence and gene regulation [54, 55]. RBPs are critical for post-transcriptional RNA processing and modification (e.g., splicing, mRNA stabilization, and mRNA localization) [56, 57]. We hypothesize some NMD-QTLs could function by affecting the miRNA or RBP binding sites. Here we examined how many NMD-QTLs were located at miRNA target regions predicted by TargetScan [58] as well as how many NMD-QTLs were located at RBP binding regions surveyed in ENCODE [59].

When we considered the targets of conservative miRNAs, 7.6‰ nonredundant pNMD-QTLs and 6.3‰ nonredundant dNMD-QTLs were found to be located in the miRNA target regions. Both were higher than the 4.2‰ level observed for regular eQTLs (for pNMD-QTLs vs. eQTLs, odds ratio $= 1.79$, Fisher's exact test p-value $= 6.3 \times 10^{-29}$, for dNMD-QTLs vs. eQTLs, odds ratio $= 1.49$, Fisher's exact test p-value $= 1.5 \times 10^{-62}$). The top 20 miRNAs whose targets were enriched with pNMD-QTLs or dNMD-QTLs are shown in Fig. 9a. Seven of them were shared between pNMD-QTLs and dNMD-QTLs (orange colored). The full list of miRNAs can be found in Additional file 3: Table S2. When the overlap percentages were calculated for individual tissues, NMD-QTLs still had higher percentages than eQTLs (Fig. 9b). The same conclusion can be made when only NMD-QTLs (or eQTLs) located in the 3' untranslated regions (UTRs) were used to calculate the overlap percentages (Additional file 1: Fig. S7).

A total of 12.7% nonredundant dNMD-QTLs and 14.9% nonredundant pNMD-QTLs were found in the binding sites of RBPs in the K562 cell line. Similarly, 14.9% nonredundant dNMD-QTLs and 16.4% nonredundant pNMD-QTLs were found in the binding sites of RBPs in the HepG2 cell line. These percentages were higher than those of eQTLs (10.3% and 12.3% in K562 and HepG2, respectively). For pNMD-QTLs vs. eQTLs, the odds ratio was 2.84 for K562 and 3.13 for HepG2 (Fisher's exact test p-values $< 10^{-16}$). For dNMD-QTLs vs. eQTLs, the odds ratio was 2.42 for K562 and 2.84 for HepG2 (Fisher's exact test p-values $< 10^{-16}$). Nine of the top 20 RBPs whose targets in the K562 cell line were enriched with pNMD-QTLs or dNMD-QTLs were shared (Fig. 9c, orange colored). The top 20 RBPs in HepG2 were similar between pNMD-QTLs and dNMD-QTLs (17 out of 20 were shared and orange
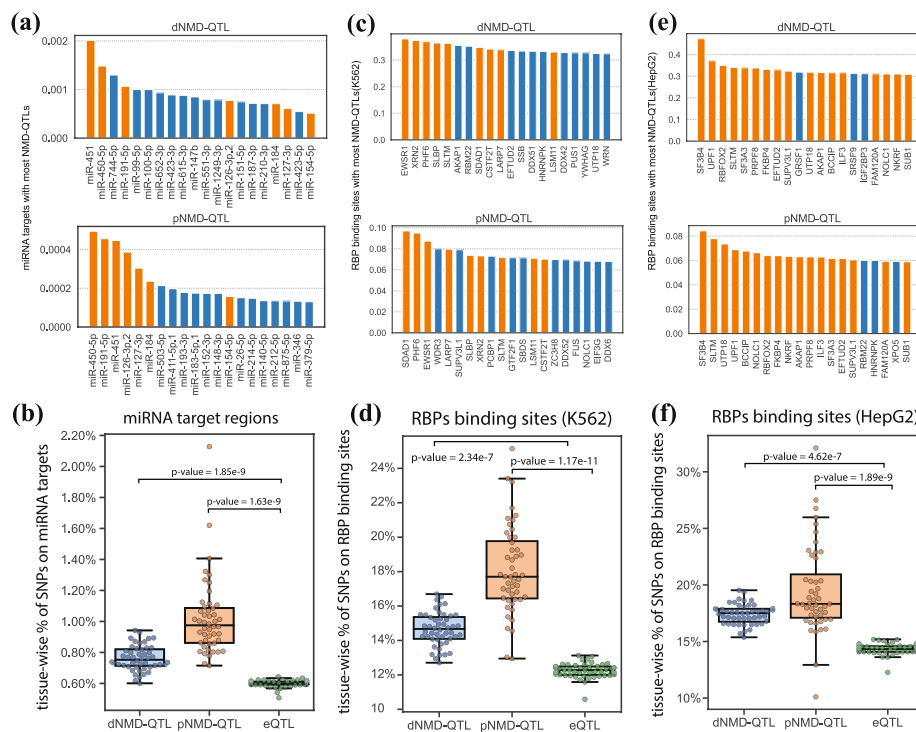
Fig. 9 Relationship between NMD-QTLs and miRNAs or RBPs. **a** Top 20 miRNAs whose target regions are enriched with dNMD-QTLs or pNMD-QTLs. Shared miRNAs between dNMD-QTLs and pNMD-QTLs are orange colored. The *Y* axis shows the percentage of binding site positions that are NMD-QTLs. NMD-QTLs from all tissues are combined. **b** Tissue-wise percentages of NMD-QTLs in miRNA targets compared to those of eQTLs. The *Y* axis shows the percentage of NMD-QTLs locating in miRNA targets among all NMD-QTLs identified from a specific tissue. **c** Top 20 RBPs whose binding regions profiled in K562 are enriched with dNMD-QTLs or pNMD-QTLs. Shared RBPs between dNMD-QTLs and pNMD-QTLs are orange colored. **d** Tissue-wise percentages of NMD-QTLs in binding regions of RBPs in K562 compared to those of eQTLs. **e** Top 20 RBPs whose binding regions profiled in HepG2 are enriched with dNMD-QTLs or pNMD-QTLs. **d** Tissue-wise percentages of NMD-QTLs in binding regions of RBPs in HepG2 compared to those of eQTLs

colored, Fig. 9e), including UPF1, the key effector of the NMD pathway. Other RBPs could function as co-factors or modulators of the NMD regulation pathways. Full lists of RBPs are provided in Additional file 4: Table S3. When we considered NMD-QTLs discovered from each individual tissue instead of pooling them together, we still found that NMD-QTLs, especially pNMD-QTLs, were more likely to be in the binding sites of RBPs compared to eQTLs (Fig. 9d, f).

RBPs such as HNRNPL [60] and PTBP1[61] protect mRNAs from degradation, and their binding sites are expected to be de-enriched with NMD-QTLs. Upon checking, the binding sites of the two RBPs are indeed de-enriched with NMD-QTLs. For example, in the K562 cell line, the total length of PTBP1 binding sites accounted for 0.6% of all RBP binding positions, but only 0.16% of pNMD-QTLs and 0.21% of dNMD-QTLs residing in RBP binding sites were on PTBP1 binding sites (*p*-values for the de-enrichment $< 2 \times 10^{-10}$). Similarly, the total length of HNRNPL binding sites accounted for 2.5% of all RBP binding positions, but only 1% pNMD-QTLs and 1% dNMD-QTLs residing in RBP binding sites were on HNRNPL binding sites (*p*-values $< 2 \times 10^{-16}$). Similar de-enrichment signals were observed in the HepG2 cell line (*p*-values $< 2 \times 10^{-16}$).
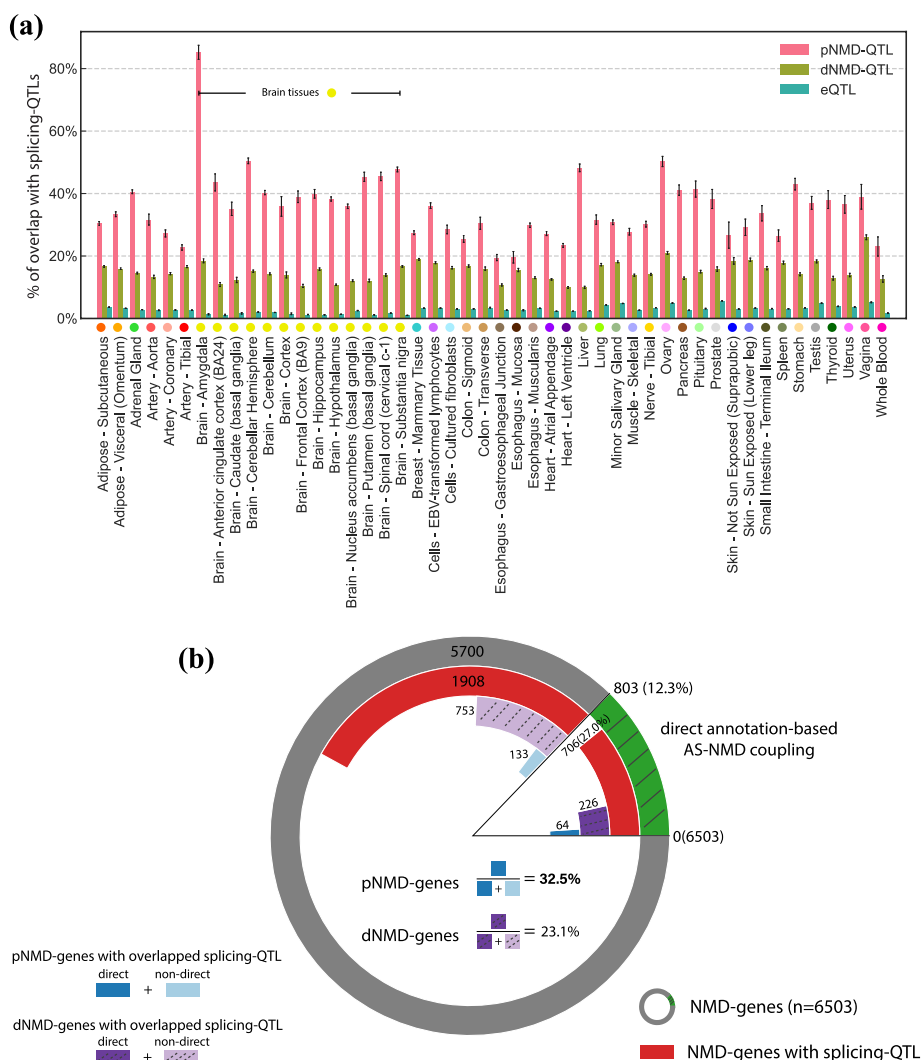
Overlap of splicing-QTLs and NMD-QTLs

In mammalian cells, alternative splicing coupled to nonsense-mediated decay (i.e., AS-NMD) is a conserved post-transcriptional regulation mechanism in which alternative splicing can alter the reading frame to produce a PTC-containing transcript subject to NMD [11, 24, 62–65]. We therefore analyzed the overlap of NMD-QTLs and splicing-QTLs. Here, we identified splicing-QTLs by mapping variants associated with the inclusion ratios of annotated cassette exons (details in Methods). As expected, NMD-QTLs were much more likely to overlap with splicing-QTLs than with eQTLs (average odds ratios: 15.9 for pNMD-QTLs vs. eQTLs; 6.1 for dNMD-QTLs vs. eQTLs). Compared to dNMD-QTLs, pNMD-QTLs were more likely to act as splicing-QTLs simultaneously since alternative splicing is coupled with NMD through "producing" transcripts subject to NMD. Such overlap of pNMD-QTLs and splicing-QTLs was more prominent in brain tissues (average odds ratio in brain tissues for pNMD-QTL vs. eQTLs: 30.8, Fig. 10a). The odds ratio was as high as 60.8 in the amygdala which defines and regulates human emotions. The results suggest that AS-NMD can be a prevalent post-transcriptional regulation mechanism for brain tissues, especially in the amygdala.

To validate this finding, we focused on NMD genes with direct annotation-based evidence of AS-NMD coupling: given a cassette exon, all NMD-targeted transcripts include the cassette exon and all non-NMD transcripts skip that cassette exon, or conversely, all NMD transcripts skip the cassette exon and all non-NMD transcripts include it. As shown in Fig. 10b, out of all 6503 NMD genes, 803 (12.3%) of them had direct annotation evidence for AS-NMD coupling (the green arc). Out of the 2614 NMD genes with splicing-QTLs (the red arcs), 706 genes (27.0%) showed direct evidence of AS-NMD coupling. For genes regulated by pNMD-QTLs which overlap splicing-QTLs (the blue arcs), the percentage of genes with direct AS-NMD coupling was even higher (64 out of 197, 32.5%). However, this percentage was only 23.1% (226 out of 979) for genes with dNMD-QTLs which overlap splicing-QTLs (the purple arcs). Therefore, NMD genes with direct AS-NMD coupling were more likely to have pNMD-QTLs overlapping splicing-QTLs (Fisher's exact tests p-values $< 10^{-16}$). This corroborates our hypothesis that alternative splicing couples with NMD through "producing" NMD transcripts and suggests that alternative splicing itself may not influence decay efficiency of NMD transcripts.

Discussion

To uncover the etiological path from SNPs to resultant phenotypic traits, it is essential to understand the mechanisms underlying gene expression variation. The current gene expression regulatory landscape derived from eQTL studies is far from complete. Traditional eQTL studies have focused on the identification of genetic loci linked to variations in the overall mRNA expression [31, 66, 67], and more recently to variations in mRNA splicing [68–70]. A special interest of the current human eQTL studies is tissue-specific genetic impacts such as the efforts in GTEx [30, 31, 71, 72], because they can provide valuable insights into disease phenotypes manifested on particular tissues. However, transcriptional and post-transcriptional regulations jointly determine tissue-specific gene expression, which needs a more careful integrated examination. In this study, we analyzed the post-transcriptional processing of RNA and proposed a more fine-grained

Fig. 10 Overlap of NMD-QTLs and splicing-QTLs. **a** Percentages of NMD-QTLs and eQTLs overlapping with splicing-QTLs across tissues. Error bars: 95% confidence intervals. **(b)** Nested pie chart showing the relative proportion of each category. The most outer circle shows NMD genes with and without annotation-based AS-NMD coupling (i.e., direct or non-direct). The inner circles from outside to inside show NMD genes with splicing-QTLs (red arcs), dNMD-genes (i.e., genes with dNMD-QTLs) with overlapped splicing-QTLs (purple arcs), and pNMD-genes (i.e., genes with pNMD-QTLs) with overlapped splicing-QTLs, respectively (blue arcs)

regulatory QTL model, the NMD-QTL model, to pinpoint genetic variants that function via regulating the production percentage of NMD-transcripts (pNMD-QTLs) and the decay efficiency of targeted RNAs (dNMD-QTLs).

Through joint consideration of allelic effect sizes for both NMD-transcripts and normal transcripts, we distinguish pNMD-QTLs and dNMD-QTLs from regular eQTLs. Our NMD-QTL identification is on the gene level, but we divided transcripts into two groups: NMD-targeted transcripts and non-NMD transcripts. With the advances of accurate expression quantification for transcript isoforms, we may consider individual NMD transcript isoforms in the future to obtain an even more fine-grained identification of isoform-specific NMD-QTLs. The modeling is validated by extensive simulations mimicking real data sets. Taking advantage of the unprecedented resource available in

GTEx, we reveal the genome-wide landscape of genetic variants affecting NMD regulation which can be missed in regular GTEx eQTLs mapping, and uncover its tissue specificity and its variant-disease associations.

Our results show NMD-QTLs are valuable in dissecting the variant-disease associations and to characterize disease susceptibility, especially for brain and mental disorders. The higher overlapping rates of disease SNPs with NMD-QTLs than with eQTLs point to the need to include NMD-QTLs in prioritizing and interpreting variants discovered from the genome-wide association studies. The enrichment of NMD-QTLs was observed for a variety of diseases, including musculoskeletal diseases, digestive system diseases, respiratory tract diseases, skin and connective tissue diseases, and immune system diseases. The link between NMD regulation and various disease susceptibility stresses the fact that NMD functions as an intricate gene regulation mechanism across human tissues. In particular, brain tissues exhibit unique NMD-QTL signatures, and these NMD-QTLs can uncover prominent variant-disease associations. NMD is critical for proper neuronic development and function, and that disruption of this process can lead to neurological disorders. Further research is needed to fully understand the specific mechanisms by which NMD factors and mutations impact brain function, and to develop targeted therapies for these disorders.

Approximately 3 million individuals in the USA alone are afflicted with genetic diseases caused by nonsense mutations (PTC diseases) [73]. Additionally, deficiencies in NMD factors are linked to neurological disorders, immune diseases, and cancers [20, 74, 75]. An NMD-based therapeutic strategy will be impossible without a comprehensive understanding of NMD regulation. First, NMD efficiency influences patients' response to nonsense-suppression drugs [76], but NMD efficiency modulation is unclear. Second, NMD inhibitors have been proposed to treat PTC diseases, but potential side effects are awaiting to be addressed [29, 73, 77]. Our discovered NMD-QTLs may assist in the design of gene-specific NMD inhibition to treat PTC diseases, mitigating the global side effects caused by manipulations of core NMD factors [78]. For example, by considering the allele configuration of NMD-QTLs for a specific disease gene, we can help to judge whether a patient will be more or less responsive to the PTC read-through treatment.

Little is known about the determinants of NMD efficiency. The enrichment of NMD-QTLs within gene bodies and exons, especially the penultimate exons from the 3′ end, suggests embedded sequence features for NMD regulation. Regions around PTC are important regulatory regions since NMD-QTLs tend to be located around PTCs. The enrichment of pNMD-QTLs in near proximity to regular stop codons and their upstream 400-bp positions implies additional determinants producing NMD-targeted transcripts.

The location preferences of NMD-QTLs in the targets of miRNAs and RBPs show that NMD regulation is not fixated but can be modulated by miRNAs and RBPs. The mammalian core NMD machinery includes a PIKK complex (SMG1, SMG8, SMG9) and other SMG proteins (SMG5, SMG6, SMG7), UPF proteins (UPF1, UPF2, UPF3A, UPF3B), eukaryotic release factors (eRF1, eRF3), and exon junction complex (EJC) members (eIF4A3, RBM8A, MAGOH, and MLN51) [14, 66, 67]. Enrichments of NMD-QTLs overlapping with RBP binding sites and miRNA targets suggest that these *cis*-acting regulatory elements interact with RBPs and miRNAs to modulate the generation and

decay efficiency of NMD transcripts. Additionally, the overlap of NMD-QTLs, especially pNMD-QTLs, with splicing-QTLs reinforces the notion that AS-NMD coupling is pervasive in NMD regulation. At the same time, it demonstrates that our identified NMD-QTLs are most likely to be true positives since they exhibit a strong association with alternative splicing, as expected.

Conclusions

We explored the post-transcriptional regulatory genetic variants in human tissues and built a genome-wide landscape of genetic variants affecting the percentage of NMD-targeted transcripts (pNMD-QTLs), as well as genetic variants regulating the decay efficiency of NMD-targeted transcripts (dNMD-QTLs). These NMD-QTLs demonstrate a strong association with disease SNPs and exhibit position characteristics. These findings are valuable in understanding post-transcriptional RNA regulation, establishing the etiological roadmap from SNPs to resultant phenotypic traits, and facilitating the research of NMD-based therapeutic strategies for genetic disease.

Methods

Data acquisition

For NMD-QTL mapping, we obtained expression data at the transcript level and genotype data from the Genotype-Tissue Expression (GTEx) project (dbGaP accession: phs000424.v7.p2). We grouped transcripts tagged by "nonsense_mediated_decay" in the GENCODE annotation (release 19, GRCh37.p13) as NMD-targeted transcripts for a gene. For disease SNP analysis, we downloaded the curated variant-disease associations from the DisGeNET database (GRCh38) [79]. Genome coordinates from different assembly versions were converted to hg19 through segment_liftover [80]. For miRNA targets, we considered the genome coordinates of targets (conserved or not) for conservative and broad conservative miRNAs obtained in TargetScan [58] which included a total of 257 miRNAs. For RNA binding protein (RBP) binding sites, eCLIP experiments for 120 RBPs from the K562 cell line and 103 RBPs from the HepG2 cell line from ENCODE were used.

Modeling of NMD-QTLs

Assuming the amount of mRNAs being transcribed is t, the percentage of transcribed mRNAs that are subject to NMD is α, and the NMD degradation efficiency relative to regular mRNA degradation is $(1-\theta)$, the amount of NMD-targeted transcripts observed from the RNA-seq experiment is the result of $t \cdot \alpha \cdot \theta$. To apply an additive genetic model, the total level for NMD-targeted transcripts can be written as $T = t_A \cdot \alpha_A \cdot \theta_A + t_a \cdot \alpha_a \cdot \theta_a$, where the subscript A and a denote the two alleles for a biallelic SNP. Two scenarios were our major interests: (i) pNMD-QTL where the genetic variant affects the percentage of transcripts subject to NMD ($\alpha_A \neq \alpha_a$), (ii) dNMD-QTL where the genetic variant regulates the NMD efficacy ($\theta_A \neq \theta_a$). These two cases can be distinguished from regular eQTLs by examining their effect sizes obtained in our NMD-aware QTL mapping procedure as shown in Fig. 1.

Identification of NMD-QTLs

Different from a typical gene-level analysis where all isoforms of a gene are considered together, the mRNA transcripts of a gene were divided into two categories: NMD-targeted and non-NMD transcripts, based on whether they were annotated by a "nonsense_mediated_decay" tag in the GENCODE annotation. A transcript is tagged as NMD if (i) there is a PTC located > 50 bp from a downstream splice site; or (ii) if the transcript does not cover the full reference coding sequence and NMD is unavoidable (i.e., no matter what the exon structure of the missing portion is, the transcript will be subject to NMD).

We summarized transcripts level by adding up the transcripts per million (TPM) of all isoforms from the NMD-targeted group and the non-NMD group respectively. We excluded genes with all or none NMD-tagged transcripts. The summarized expressions were normalized within samples by a normalizing factor calculated based on the median of the geometric mean of all genes considered and were then normalized across samples by a rank-based inverse normal transformation. Detailed formulas for normalization can be found in Additional file 1: Supplementary Texts.

QTL mappings were performed with FastQTL [81] on both the NMD-targeted and non-NMD groups. According to our modeling, pNMD-QTLs and dNMD-QTLs can be identified by comparing their regulatory effect sizes (Fig. 1). For pNMD-QTLs, the significant allelic regulatory effect (q-values ≤ 0.05) for NMD-targeted transcripts and non-NMD transcripts are in the opposite direction. For dNMD-QTLs, the regulatory effect is only observable for NMD-targeted transcripts (q-value ≤ 0.05) but not for non-NMD transcripts (q-value > 0.8). In our QTL mapping (NMD-QTLs and splicing-QTLs), similar to the approach in GTEx, variants in the up- or down-stream 1 Mb *cis* window around the TSS were considered. The same covariates used in GTEx eQTL analysis were used. They included genotyping principal components, sequencing platform, sequencing protocol, sex, and a set of covariates identified using the Probabilistic Estimation of Expression Residuals (PEER) method [82]. As in the GTEx eQTL analysis, the permutation mode in FastQTL was used with the setting "−permute 1000 10,000" and the false discovery rate (FDR) threshold of 0.05 was applied in our studies.

Effect size confirmation with quantile regression

RNA-seq quantifications tend to be heavy-tailed and with inflated number of zeros. The effect size estimation in the ordinary linear square fit in FastQTL is sensitive to such non-normality and outliers, even with the applied inverse normal transformation. To reduce false positives and increase the model robustness, for pNMD-QTLs and dNMD-QTLs discovered from FastQTL, we further fit quantile regression models on TPM values without the inverse normal transformation as we did previously [38]. Only pNMD-QTLs and dNMD-QTLs still exhibiting desired effect size properties in quantile regression models were in our final discovery list. Thus, for pNMD-QTLs, both quantile regression nominal p-values of the effect sizes should be smaller than 0.05, and opposite signs of effect sizes for NMD-targeted and non-NMD transcripts should be observed and the former absolute value was smaller. For dNMD-QTLs, significant effect size (quantile regression nominal p-value ≤ 0.05) was only observed for NMD-targeted transcripts, but not for non-NMD transcripts (quantile regression nominal p-value > 0.95).

Simulation studies to test our model

To validate our NMD-QTL models, we estimated the precision and recall for identifying pNMD-QTLs and dNMD-QTLs via massive amount of simulations. We simulated the expression levels (total detected transcripts) for NMD-targeted transcripts and non-NMD transcripts by formulas in Fig. 1c for different scenarios. Then we performed QTL calling for both NMD-targeted and non-NMD transcripts through regression models. We declared pNMD-QTL or dNMD-QTL findings according to the desired effect size and directions.

Specifically, we performed three rounds of simulations, each round with different regulatory strength and 10,000 simulated genetic variants. For each genetic variant, n genotypes, $X = [x_1, x_2, \ldots, x_n]$, were simulated with a minor allele frequency of m, where $n = 500$ and $m = 0.05$. The total detected transcripts, $\mathbf{Y}^c = \left[y_1^c, y_2^c, \ldots, y_n^c \right]$, were simulated as a function of genotype, transcript type c (NMD-targeted transcripts or not), and plus a Gaussian random noise: $y_i^c = f_c(x_i; \Theta_i(t, \alpha, \theta)) + \sigma_i$.

For the simulated expression and genotype data, we performed a linear regression $\mathbf{Y}_c = \beta_c \mathbf{X} + \epsilon$ for each SNP and inspected the coefficient β_c. If β_c s were significant (*p*-value less than a particular threshold) for both NMD-targeted and non-NMD transcripts and were in opposite directions, the SNP was identified as a pNMD-QTL. If β_c was significant for NMD-targeted transcripts but not for non-NMD transcripts, the SNP was identified as a dNMD-QTL. The detailed parameter settings for simulations with different regulatory strengths can be found in Additional file 1: Supplementary Texts.

Control experiment to assess false discovery rate

We randomly selected 6000 non-NMD genes. The transcript isoforms from these genes were randomly grouped into two groups to mimic the NMD-targeted transcript and non-NMD-transcript groups. Then we performed the same NMD-QTLs calling procedures for this control setting. We tested the adipose subcutaneous tissue. Using the same criteria mentioned for our NMD-QTL identification, 2228 pNMD-QTLs were identified in the random control. This is about 12% of the total number of pNMD-QTLs discovered for adipose subcutaneous in our results (18,571). Thus, the FDR is controlled by 0.12. For dNMD-QTLs, the estimated FDR based on the control set is about 0.17 for adipose subcutaneous.

Genome-wide NMD-QTL mapping for GTEx tissues

The analysis power for NMD-QTL detection was linearly related to the sample size (Additional file 1: Fig. S8a, the regression coefficient of the sample size was 28.8 for pNMD-QTLs and 117.0 for dNMD-QTLs), but less sensitive than that for eQTL detection (Additional file 1: Fig. S8b, regression coefficient: 2938). The testis and whole blood tissues were drastically deviated from this linear trend (Additional file 1: Fig. S8a), suggesting that the testis had many more NMD-QTLs while the whole blood had many less NMD-QTLs than expected. Such deviations were also observed in the eQTLs mapping (Additional file 1: Fig. S8b).

NMD-QTL signature similarity between a tissue pair

To measure the similarity of identified NMD-QTLs between tissues, we calculated $s_{ij} = n_{ij}/(n_i - n_{ij})/(n_j/(n - n_j))$ where n_{ij} is the number of NMD-QTLs shared between tissue i and j; n_i and n_j are the numbers of NMD-QTLs discovered in respective tissues; and n is the total number of nonredundant NMD-QTLs discovered across all tissues. Thus, s_{ij} compares the odds of observing tissue j's NMD-QTLs in tissue i with the odds of observing tissue j's NMD-QTLs in all NMD-QTLs. Note that s_{ij} is not necessarily equal to s_{ji}, we calculated $s_{ij}^* = \frac{s_{ij} + s_{ji}}{2}$ as the final similarity score for NMD-QTL signatures between two tissues. If two tissues share many NMD-QTLs, s_{ij}^* is large.

Overlap of NMD-QTLs with disease-associated SNPs

Considering the disease-associated variants reported in DisGeNET, we compared how likely the regular eQTLs reported by GTEx and our NMD-QTLs to overlap with disease SNPs. We devised a strategy with the combination of a proportion test and a hyper-geometric test to examine the significance of the overlap between disease SNPs and NMD-QTLs. Here, we pooled NMD-QTLs identified across different tissues. Thus, one NMD-QTL could be counted multiple times if it was identified in multiple tissues. Detail of such two tests can be found in Additional file 1: Supplementary Texts.

Overlap enrichment score for MeSH

For a disease category represented by a MeSH term (C01-C26, F01-F03), the overlap enrichment score was calculated as the number of pNMD-QTLs or dNMD-QTLs which were also associated with the disease SNP in the DisGeNET database for the MeSH term, further normalized by the total number of pNMD-QTLs or dNMD-QTLs and the total number disease SNPs annotated by the MeSH term in DisGeNET. The same calculation was performed for eQTL and randomly sampled SNPs.

Enrichment of NMD-QTLs in the targets of miRNAs and RBPs

miRNA and RBPs were ranked by the number of nonredundant pNMD-QTLs or dNMD-QTLs across all tissues which were located in their target regions and normalized by the total base-pair length of the target regions. To compare NMD-QTLs to eQTLs concerning their overlap with miRNA and RBP binding sites, the percentage of NMD-QTLs (or eQTLs) residing in the target regions among all NMD-QTLs (or eQLTs) identified from a specific tissue was calculated.

Splicing-QTL mapping

We identified all cassette exons from the GENCODE annotation. A cassette exon is defined as a non-boundary exon between two other exons and can be either included or skipped to generate two distinct transcripts in alternative splicing. To simplify, we considered transcripts either using or skipping the cassette exon entirely. In other words, transcripts partially overlap with the cassette exons were excluded. To perform

splicing-QTL mapping, for a given cassette exon, we used the inclusion ratio of that cassette exon as the quantitative trait. The inclusion ratio of a given cassette exon is calculated as follows:

$$\frac{\sum_{j=1}^{M} \mathrm{TPM}\left(t_{inc}^{(j)}\right)}{\sum_{j=1}^{M} \mathrm{TPM}\left(t_{inc}^{(j)}\right) + \sum_{k=1}^{N} TPM\left(t_{exc}^{(k)}\right)}$$

where $t_{inc}^{(j)}$ are transcripts including the cassette exon, $t_{exc}^{(k)}$ are transcripts excluding the cassette exon, and TPM() is the detected transcripts per million for a transcript. We provide the list of cassette exons and their corresponding transcripts in Additional file 5: Table S4.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13059-023-03004-w.

Additional file 1. Supplementary Texts and Supplementary Figures S1-S8.

Additional file 2: Table S1. The 43 diseases with significant NMD-QTLs enrichment.

Additional file 3: Table S2. The list of miRNAs and the number of dNMD-QTLs or pNMD-QTLs in their target regions.

Additional file 4: Table S3. The list of RBPs and the number of dNMD-QTLs or pNMD-QTLs in their binding sites in the K562 or the HepG2 cell lines.

Additional file 5: Table S4. The list of cassette exons considered in splicing-QTL mapping.

Additional file 6. Review history.

Authors' contributions
BS and LC conceived and designed the work. BS conducted formal analyses. LC supervised the study. BS and LC wrote the manuscript. All authors read and approved the final manuscript.

Availability of data and materials
Original data were obtained from the following online sources: the GTEx project (analysis V7) [83], the GENCODE annotation file (release 19) [84], the Variant-disease association [85], the Genome assembly conversion tool [86], miRNA targets from TargetScan (release 7.2) [87], RBP binding sites from ENCODE (eCLIP data in K562 and HepG2) [88], NMD-QTLs identified from this project and our source code (under the MIT license) have been deposited on Open Science Framework https://osf.io/hts7w/ with https://doi.org/10.17605/OSF.IO/HTS7W [89].

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

References

1. Behm-Ansmant I, Kashima I, Rehwinkel J, Sauliere J, Wittkopp N, Izaurralde E. mRNA quality control: an ancient machinery recognizes and degrades mRNAs with nonsense codons. FEBS Lett. 2007;581:2845–53.
2. Bhuvanagiri M, Schlitter AM, Hentze MW, Kulozik AE. NMD: RNA biology meets human genetic medicine. Biochem J. 2010;430:365–77.
3. Huang L, Wilkinson MF. Regulation of nonsense-mediated mRNA decay. Wiley Interdiscip Rev RNA. 2012;3:807–28.
4. Lindeboom RGH, Vermeulen M, Lehner B, Supek F. The impact of nonsense-mediated mRNA decay on genetic disease, gene editing and cancer immunotherapy. Nat Genet. 2019;51:1645–51.
5. Mort M, Ivanov D, Cooper DN, Chuzhanova NA. A meta-analysis of nonsense mutations causing human genetic disease. Hum Mutat. 2008;29:1037–47.
6. Supek F, Lehner B, Lindeboom RGH. To NMD or Not To NMD: Nonsense-Mediated mRNA Decay in Cancer and Other Genetic Diseases. Trends Genet. 2021;37:657–68.
7. Mendell JT, Sharifi NA, Meyers JL, Martinez-Murillo F, Dietz HC. Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. Nat Genet. 2004;36:1073–8.
8. Rehwinkel J, Letunic I, Raes J, Bork P, Izaurralde E. Nonsense-mediated mRNA decay factors act in concert to regulate common mRNA targets. RNA. 2005;11:1530–44.
9. Tani H, Imamachi N, Salam KA, Mizutani R, Ijiri K, Irie T, Yada T, Suzuki Y, Akimitsu N. Identification of hundreds of novel UPF1 target transcripts by direct determination of whole transcriptome stability. RNA Biol. 2012;9:1370–9.
10. Weischenfeldt J, Damgaard I, Bryder D, Theilgaard-Monch K, Thoren LA, Nielsen FC, Jacobsen SE, Nerlov C, Porse BT. NMD is essential for hematopoietic stem and progenitor cells and for eliminating by-products of programmed DNA rearrangements. Genes Dev. 2008;22:1381–96.
11. Zheng S. Alternative splicing and nonsense-mediated mRNA decay enforce neural specific gene expression. Int J Dev Neurosci. 2016;55:102–8.
12. Kurosaki T, Maquat LE. Nonsense-mediated mRNA decay in humans at a glance. J Cell Sci. 2016;129:461–7.
13. Lelivelt MJ, Culbertson MR. Yeast Upf proteins required for RNA surveillance affect global expression of the yeast transcriptome. Mol Cell Biol. 1999;19:6710–9.
14. Mitrovich QM, Anderson P. mRNA surveillance of expressed pseudogenes in C. elegans. Curr Biol. 2005;15:963–7.
15. Bao J, Tang C, Yuan S, Porse BT, Yan W. UPF2, a nonsense-mediated mRNA decay factor, is required for pre-pubertal Sertoli cell development and male fertility by ensuring fidelity of the transcriptome. Development. 2015;142:352–62.
16. Bao J, Vitting-Seerup K, Waage J, Tang C, Ge Y, Porse BT, Yan W. UPF2-Dependent Nonsense-Mediated mRNA Decay Pathway Is Essential for Spermatogenesis by Selectively Eliminating Longer 3'UTR Transcripts. PLoS Genet. 2016;12: e1005863.
17. Han X, Wei Y, Wang H, Wang F, Ju Z, Li T. Nonsense-mediated mRNA decay: a "nonsense" pathway makes sense in stem cell biology. Nucleic Acids Res. 2018;46:1038–51.
18. Kurosaki T, Sakano H, Proschel C, Wheeler J, Hewko A, Maquat LE. NMD abnormalities during brain development in the Fmr1-knockout mouse model of fragile X syndrome. Genome Biol. 2021;22:317.
19. Li Z, Vuong JK, Zhang M, Stork C, Zheng S. Inhibition of nonsense-mediated RNA decay by ER stress. RNA. 2017;23:378–94.
20. Lykke-Andersen S, Jensen TH. Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes. Nat Rev Mol Cell Biol. 2015;16:665–77.
21. Ottens F, Gehring NH. Physiological and pathophysiological role of nonsense-mediated mRNA decay. Pflugers Arch. 2016;468:1013–28.
22. Raimondeau E, Bufton JC, Schaffitzel C. New insights into the interplay between the translation machinery and nonsense-mediated mRNA decay factors. Biochem Soc Trans. 2018;46:503–12.
23. Shum EY, Jones SH, Shao A, Chousal JN, Krause MD, Chan WK, Lou CH, Espinoza JL, Song HW, Phan MH, et al. The Antagonistic Gene Paralogs Upf3a and Upf3b Govern Nonsense-Mediated RNA Decay. Cell. 2016;165:382–95.
24. Zhao J, Li Z, Puri R, Liu K, Nunez I, Chen L, Zheng S. Molecular profiling of individual FDA-approved clinical drugs identifies modulators of nonsense-mediated mRNA decay. Mol Ther Nucleic Acids. 2022;27:304–18.
25. Gudikote JP, Imam JS, Garcia RF, Wilkinson MF. RNA splicing promotes translation and RNA surveillance. Nat Struct Mol Biol. 2005;12:801–9.
26. Usuki F, Yamashita A, Higuchi I, Ohnishi T, Shiraishi T, Osame M, Ohno S. Inhibition of nonsense-mediated mRNA decay rescues the phenotype in Ullrich's disease. Ann Neurol. 2004;55:740–4.
27. Huang L, Lou CH, Chan W, Shum EY, Shao A, Stone E, Karam R, Song HW, Wilkinson MF. RNA homeostasis governed by cell type-specific and branched feedback loops acting on NMD. Mol Cell. 2011;43:950–61.
28. Zetoune AB, Fontaniere S, Magnin D, Anczukow O, Buisson M, Zhang CX, Mazoyer S. Comparison of nonsense-mediated mRNA decay efficiency in various murine tissues. BMC Genet. 2008;9:83.
29. Miller JN, Pearce DA. Nonsense-mediated decay in genetic disease: friend or foe? Mutat Res Rev Mutat Res. 2014;762:52–64.
30. Consortium GT. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science. 2015;348:648–60.
31. Consortium GT, Laboratory DA, Coordinating Center -Analysis Working G, Statistical Methods groups-Analysis Working G, Enhancing Gg, Fund NIHC, Nih/Nci, Nih/Nhgri, Nih/Nimh, Nih/Nida, et al. Genetic effects on gene expression across human tissues. Nature. 2017;550:204–13.
32. Hsiao YH, Bahn JH, Lin X, Chan TM, Wang R, Xiao X. Alternative splicing modulated by genetic variants demonstrates accelerated evolution regulated by highly conserved proteins. Genome Res. 2016;26:440–50.
33. Wu L, Candille SI, Choi Y, Xie D, Jiang L, Li-Pook-Than J, Tang H, Snyder M. Variation and genetic control of protein abundance in humans. Nature. 2013;499:79–82.
34. Zhang H, Dou S, He F, Luo J, Wei L, Lu J. Genome-wide maps of ribosomal occupancy provide insights into adaptive evolution and regulatory roles of uORFs during Drosophila development. PLoS Biol. 2018;16: e2003903.

35. MacArthur DG, Tyler-Smith C. Loss-of-function variants in the genomes of healthy humans. Hum Mol Genet. 2010;19:R125-130.
36. Rivas MA, Pirinen M, Conrad DF, Lek M, Tsang EK, Karczewski KJ, Maller JB, Kukurba KR, DeLuca DS, Fromer M, et al. Human genomics. Effect of predicted protein-truncating genetic variants on the human transcriptome. Science. 2015;348:666–9.
37. Teran NA, Nachun DC, Eulalio T, Ferraro NM, Smail C, Rivas MA, Montgomery SB. Nonsense-mediated decay is highly stable across individuals and tissues. Am J Hum Genet. 2021;108:1401–8.
38. Sun B, Chen L. Quantile regression for challenging cases of eQTL mapping. Brief Bioinform. 2020;21:1756–65.
39. Khajavi M, Inoue K, Lupski JR. Nonsense-mediated mRNA decay modulates clinical outcome of genetic disease. Eur J Hum Genet. 2006;14:1074–81.
40. Nguyen LS, Wilkinson MF, Gecz J. Nonsense-mediated mRNA decay: inter-individual variability and human disease. Neurosci Biobehav Rev. 2014;46(Pt 2):175–86.
41. Pinero J, Bravo A, Queralt-Rosinach N, Gutierrez-Sacristan A, Deu-Pons J, Centeno E, Garcia-Garcia J, Sanz F, Furlong LI. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. Nucleic Acids Res. 2017;45:D833–9.
42. Brodie A, Azaria JR, Ofran Y. How far from the SNP may the causative genes be? Nucleic Acids Res. 2016;44:6046–54.
43. Addington AM, Gauthier J, Piton A, Hamdan FF, Raymond A, Gogtay N, Miller R, Tossell J, Bakalar J, Inoff-Germain G, et al. A novel frameshift mutation in UPF3B identified in brothers affected with childhood onset schizophrenia and autism spectrum disorders. Mol Psychiatry. 2011;16:238–9.
44. Laumonnier F, Shoubridge C, Antar C, Nguyen LS, Van Esch H, Kleefstra T, Briault S, Fryns JP, Hamel B, Chelly J, et al. 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. 2010;15:767–76.
45. Nguyen LS, Kim HG, Rosenfeld JA, Shen Y, Gusella JF, Lacassie Y, Layman LC, Shaffer LG, Gecz J. Contribution of copy number variants involving nonsense-mediated mRNA decay pathway genes to neuro-developmental disorders. Hum Mol Genet. 2013;22:1816–25.
46. Tarpey PS, Raymond FL, Nguyen LS, Rodriguez J, Hackett A, Vandeleur L, Smith R, Shoubridge C, Edkins S, Stevens C, et al. Mutations in UPF3B, a member of the nonsense-mediated mRNA decay complex, cause syndromic and nonsyndromic mental retardation. Nat Genet. 2007;39:1127–33.
47. Zhang T, Klein A, Sang J, Choi J, Brown KM. ezQTL: A Web Platform for Interactive Visualization and Colocalization of QTLs and GWAS Loci. Genomics Proteomics Bioinformatics. 2022;20:541–8.
48. Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, Frayling TM, Hirschhorn J, Yang J, Visscher PM, Consortium G. Meta-analysis of genome-wide association studies for height and body mass index in approximately 700000 individuals of European ancestry. Hum Mol Genet. 2018;27:3641–9.
49. Foley CN, Staley JR, Breen PG, Sun BB, Kirk PDW, Burgess S, Howson JMM. A fast and efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits. Nat Commun. 2021;12:764.
50. Hormozdiari F, van de Bunt M, Segre AV, Li X, Joo JWJ, Bilow M, Sul JH, Sankararaman S, Pasaniuc B, Eskin E. Colocalization of GWAS and eQTL Signals Detects Target Genes. Am J Hum Genet. 2016;99:1245–60.
51. Colombo M, Karousis ED, Bourquin J, Bruggmann R, Muhlemann O. Transcriptome-wide identification of NMD-targeted human mRNAs reveals extensive redundancy between SMG6- and SMG7-mediated degradation pathways. RNA. 2017;23:189–201.
52. Brogna S, Wen J. Nonsense-mediated mRNA decay (NMD) mechanisms. Nat Struct Mol Biol. 2009;16:107–13.
53. Le Hir H, Gatfield D, Izaurralde E, Moore MJ. The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. EMBO J. 2001;20:4987–97.
54. Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell. 2009;136:215–33.
55. Bartel DP. Metazoan MicroRNAs. Cell. 2018;173:20–51.
56. Hogan DJ, Riordan DP, Gerber AP, Herschlag D, Brown PO. Diverse RNA-binding proteins interact with functionally related sets of RNAs, suggesting an extensive regulatory system. PLoS Biol. 2008;6: e255.
57. Lunde BM, Moore C, Varani G. RNA-binding proteins: modular design for efficient function. Nat Rev Mol Cell Biol. 2007;8:479–90.
58. Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. Elife. 2015;4:e05005. https://doi.org/10.7554/eLife.05005, https://www.ncbi.nlm.nih.gov/pubmed/26267216.
59. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489:57–74.
60. Kishor A, Ge Z, Hogg JR. hnRNP L-dependent protection of normal mRNAs from NMD subverts quality control in B cell lymphoma. EMBO J. 2019;38(3):e99128. https://doi.org/10.15252/embj.201899128, https://www.ncbi.nlm.nih.gov/pubmed/30530525.
61. Ge Z, Quek BL, Beemon KL, Hogg JR. Polypyrimidine tract binding protein 1 protects mRNAs from recognition by the nonsense-mediated mRNA decay pathway. Elife. 2016;5:e11155. https://doi.org/10.7554/eLife.11155, https://www.ncbi.nlm.nih.gov/pubmed/26744779.
62. Yap K, Makeyev EV. Regulation of gene expression in mammalian nervous system through alternative pre-mRNA splicing coupled with RNA quality control mechanisms. Mol Cell Neurosci. 2013;56:420–8.
63. Zheng S, Black DL. Alternative pre-mRNA splicing in neurons: growing up and extending its reach. Trends Genet. 2013;29:442–8.
64. Zheng S, Gray EE, Chawla G, Porse BT, O'Dell TJ, Black DL. PSD-95 is post-transcriptionally repressed during early neural development by PTBP1 and PTBP2. Nat Neurosci. 2012;15(381–388):S381.
65. Giorgi C, Yeo GW, Stone ME, Katz DB, Burge C, Turrigiano G, Moore MJ. The EJC factor eIF4AIII modulates synaptic strength and neuronal protein expression. Cell. 2007;130:179–91.
66. Morley M, Molony CM, Weber TM, Devlin JL, Ewens KG, Spielman RS, Cheung VG. Genetic analysis of genome-wide variation in human gene expression. Nature. 2004;430:743–7.
67. Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, Redon R, Bird CP, de Grassi A, Lee C, et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science. 2007;315:848–53.

68. Garrido-Martin D, Borsari B, Calvo M, Reverter F, Guigo R. Identification and analysis of splicing quantitative trait loci across multiple tissues in the human genome. Nat Commun. 2021;12:727.

69. Montgomery SB, Sammeth M, Gutierrez-Arcelus M, Lach RP, Ingle C, Nisbett J, Guigo R, Dermitzakis ET. Transcriptome genetics using second generation sequencing in a Caucasian population. Nature. 2010;464:773–7.

70. Pickrell JK, Marioni JC, Pai AA, Degner JF, Engelhardt BE, Nkadori E, Veyrieras JB, Stephens M, Gilad Y, Pritchard JK. Understanding mechanisms underlying human gene expression variation with RNA sequencing. Nature. 2010;464:768–72.

71. Consortium GT. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013;45:580–5.

72. Mele M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M, Young TR, Goldmann JM, Pervouchine DD, Sullivan TJ, et al. Human genomics. The human transcriptome across tissues and individuals. Science. 2015;348:660–5.

73. Keeling KM, Bedwell DM. Suppression of nonsense mutations as a therapeutic approach to treat genetic diseases. Wiley Interdiscip Rev RNA. 2011;2:837–52.

74. Balistreri G, Bognanni C, Muhlemann O. Virus Escape and Manipulation of Cellular Nonsense-Mediated mRNA Decay. Viruses. 2017;9(1):24. https://doi.org/10.3390/v9010024, https://www.ncbi.nlm.nih.gov/pubmed/28124995.

75. Sartor F, Anderson J, McCaig C, Miedzybrodzka Z, Muller B. Mutation of genes controlling mRNA metabolism and protein synthesis predisposes to neurodevelopmental disorders. Biochem Soc Trans. 2015;43:1259–65.

76. Linde L, Boelz S, Nissim-Rafinia M, Oren YS, Wilschanski M, Yaacov Y, Virgilis D, Neu-Yilik G, Kulozik AE, Kerem E, Kerem B. Nonsense-mediated mRNA decay affects nonsense transcript levels and governs response of cystic fibrosis patients to gentamicin. J Clin Invest. 2007;117:683–92.

77. Keeling KM. Nonsense Suppression as an Approach to Treat Lysosomal Storage Diseases. Diseases. 2016;4(4):32. https://doi.org/10.3390/diseases4040032, https://www.ncbi.nlm.nih.gov/pubmed/28367323.

78. Nomakuchi TT, Rigo F, Aznarez I, Krainer AR. Antisense oligonucleotide-directed inhibition of nonsense-mediated mRNA decay. Nat Biotechnol. 2016;34:164–6.

79. Pinero J, Ramirez-Anguita JM, Sauch-Pitarch J, Ronzano F, Centeno E, Sanz F, Furlong LI. The DisGeNET knowledge platform for disease genomics: 2019 update. Nucleic Acids Res. 2020;48:D845–55.

80. Gao B, Huang Q, Baudis M. segment_liftover : a Python tool to convert segments between genome assemblies. F1000Res. 2018;7:319.

81. Ongen H, Buil A, Brown AA, Dermitzakis ET, Delaneau O. Fast and efficient QTL mapper for thousands of molecular phenotypes. Bioinformatics. 2016;32:1479–85.

82. Stegle O, Parts L, Durbin R, Winn J. A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. PLoS Comput Biol. 2010;6: e1000770.

83. Consortium GT: The Genotype-Tissue Expression (GTEx) project. 2013. https://gtexportal.org/home/datasets.

84. Frankish A, Diekhans M, Ferreira AM, Johnson R, Jungreis I, Loveland J, Mudge JM, Sisu C, Wright J, Armstrong J, et al: GENCODE reference annotation for the human and mouse genomes. 2019, https://www.gencodegenes.org/human/release_19.html.

85. Pinero J, Ramirez-Anguita JM, Sauch-Pitarch J, Ronzano F, Centeno E, Sanz F, Furlong LI: The DisGeNET knowledge platform for disease genomics: 2019 update. 2020, https://www.disgenet.org/downloads.

86. Gao B, Huang Q, Baudis M: segment_liftover : a Python tool to convert segments between genome assemblies. 2018, https://github.com/baudisgroup/segment-liftover.

87. Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. 2015. http://www.targetscan.org/cgi-bin/targetscan/data_download.vert72.cgi.

88. Luo Y, Hitz BC, Gabdank I, Hilton JA, Kagda MS, Lam B, Myers Z, Sud P, Jou J, Lin K, et al: New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. 2020, https://www.encodeproject.org/search/?type=Experiment&control_type!=*&status=released&perturbed=false&assay_title=eCLIP.

89. Sun B, Chen L: Mapping genetic variants for nonsense-mediated mRNA decay regulation across human tissues. 2023, https://osf.io/hts7w/, 10.17605/OSF.IO/HTS7W.

Publisher's Note