

REVIEW ARTICLE

OPEN



The role of untranslated region variants in Mendelian disease: a review

Nechama Wieder^{1,2}, Elston N. D'Souza^{1,2}, Ruebena Dawes^{1,2}, Alexander Chan  ³, Alexandra Martin-Geary  ^{1,2} and Nicola Whiffin  ^{1,2,4}✉

© The Author(s) 2025

Untranslated regions (UTRs) flank the protein-coding sequence of a gene. 5'UTR and 3' UTR sequences mediate post-transcriptional regulation via linear and structural elements, controlling RNA stability, cellular localisation and the rate of protein translation. Variants within both 5' and 3' UTRs have been shown to cause disease through a variety of diverse mechanisms. However, for these variants to be routinely annotated and interpreted in clinical genetic testing, we need a better understanding of these regions and the spectrum of disease-causing variants within them. In this review, we systematically assess previously identified Mendelian disease-causing variants within UTRs and catalogue their underlying mechanisms. With genome sequencing becoming readily available and increasingly incorporated in diagnostic settings, this review will provide a valuable resource for the consideration and interpretation of UTR variants.

European Journal of Human Genetics (2025) 33:1096–1105; <https://doi.org/10.1038/s41431-025-01905-x>

INTRODUCTION

Ascertaining the genetic basis of disease is important in the clinical management of patients, potentially provides therapeutic targets, and informs patients and their families about reproductive decisions. To date, clinical genetics has mostly focused on variants in the regions of the genome that directly encode protein, with associated genetic diagnostic rates for severe rare diseases of only ~30–50% [1]. Although it is known that variants in non-coding regions can also cause disease, this is a comparatively under-studied area. Indeed, only recently have recommendations been published to support routine clinical classification of variants in non-coding regions [2].

This review focuses on variants within untranslated regions (UTRs) and their role in Mendelian disease. While UTRs have long been named 'untranslated', we now know that regulatory elements within both 5' and 3' UTRs undergo active translation. Consequently, these regions may more appropriately be referred to as 'leader' (5'UTR) and 'trailer' (3'UTR) sequences. However, we use the more familiar UTR terminology here. UTRs are the non-coding regions directly flanking the protein-coding sequence (CDS) of a gene, which are transcribed into mRNA but not translated into the canonical protein. They play an important role in gene regulation and are highly variable across genes [3]. Understanding functional regulatory elements within UTRs is pertinent in understanding their role in translation and function, and how perturbation can cause disease. By increasing knowledge of these regions, we can begin to incorporate analysis of them in clinical genetics and improve diagnostic rates [4, 5]. Hence, here we systematically review regulatory features in both 5' and 3' UTRs that, when disrupted, have been shown or hypothesised to

cause Mendelian disease. We exclude more common, complex disorders that have been reviewed previously [6, 7].

UTRS ARE IMPORTANT REGULATORY ELEMENTS

Protein production is a fine-tuned process in cells; too much or too little can disrupt cellular processes and lead to disease. 5' and 3' UTRs together play a vital role in the regulation of protein production by controlling transcript stability and the rate and location of protein synthesis.

The primary function of 5'UTRs is in translational regulation. Translation of mRNAs into protein typically begins with ribosome recruitment at the 5'cap of the mRNA, ribosome scanning along the mRNA in a 5' to 3' direction (i.e. through the transcript leader), and initiation of translation at a start codon, which is usually an AUG [8]. This process is regulated through key sequence elements such as upstream open reading frames (uORFs) and structural features which influence the amount and speed of protein production, usually by interacting with, or pausing the scanning ribosome [9]. The length of the 5'UTR, as well as the amount and type of various translational control elements within it, varies widely between genes [3], providing each with the correct combination of required elements for normal protein production. Although 5'UTRs are on average very short (mean ~200 bps), in some genes they are longer than the CDS. Typically, genes where careful control of dosage is important have longer and more complex 5'UTRs [3]. Alternative splicing and different transcription start site (TSS) usage can lead to diverse 5'UTR isoforms containing different combinations of translational control elements, enabling different levels of protein translation across tissues and developmental stages [10].

¹Big Data Institute, University of Oxford, Oxford, UK. ²Centre for Human Genetics, University of Oxford, Oxford, UK. ³Department of Biomedical Engineering, University of Waterloo, Waterloo, ON, Canada. ⁴Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA. ✉email: nwhiffin@well.ox.ac.uk

Received: 22 May 2025 Accepted: 10 June 2025
Published online: 3 July 2025

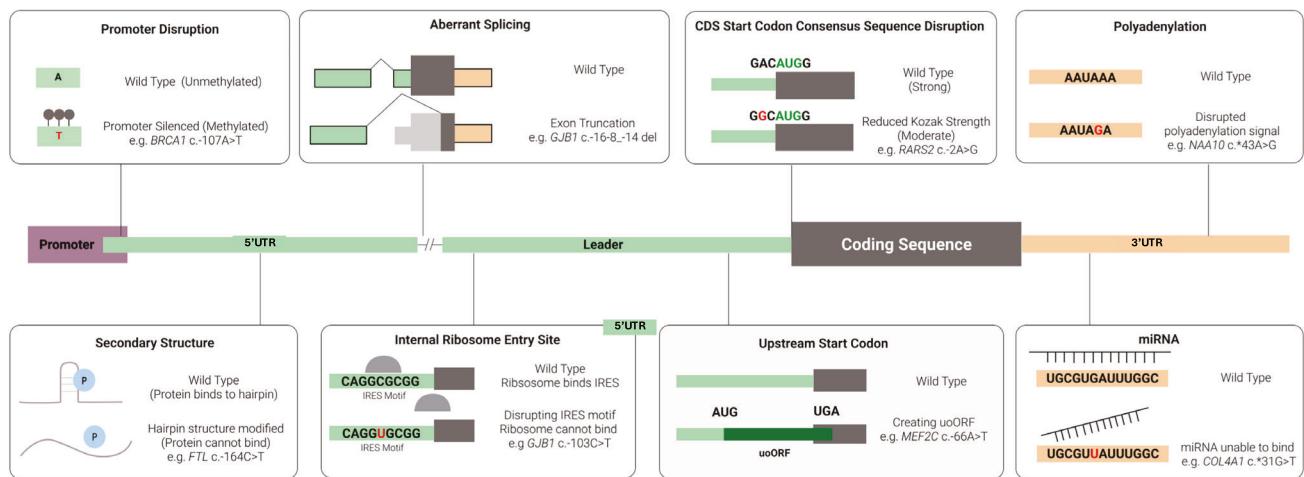


Fig. 1 Illustration of UTR regulatory elements and examples of disruptions that lead to disease. Gene and protein expression is a tightly controlled process. Elements on the DNA level (e.g. promoter) and mRNA level (e.g. uAUG codon) influence this process. When these elements are disrupted, they can lead to disease.

The primary roles of 3'UTRs are in regulating the stability of the mRNA molecule, the rate at which it is degraded, and where it is located within the cell, although, similar to 5'UTRs, they also play a role in the regulation of translation (reviewed by Mayr [11]). Much of this regulation by 3'UTRs is mediated through interactions with additional factors, namely microRNAs (miRNAs) and RNA-binding proteins (RBPs) that bind to motifs and structural elements within 3'UTRs to mediate their effect. While miRNAs are generally repressive, RBPs have a range of diverse regulatory roles (reviewed by Hentze et al. [12]). As with 5'UTRs, the precise combinations of regulatory elements vary widely between genes and between different transcript isoforms of the same gene. Alternative polyadenylation creates distinct transcript isoforms with different length 3'UTRs containing different numbers of binding sites for regulatory miRNAs and RBPs.

Given their crucial roles in gene regulation, perturbing one or more of the regulatory elements within UTRs through genetic variation can have a dramatic impact on protein production and lead to severe disease (Fig. 1). The various underlying mechanisms that have been uncovered to date are reviewed in the following sections. All the variants that are used as examples are listed in Table 1.

VARIANTS THAT CREATE UPSTREAM START CODONS DECREASE NORMAL CDS TRANSLATION

Upstream AUG (uAUG) triplets are commonly observed within 5' UTRs. uAUGs, or other near-cognate codons (most commonly CUG), may be recognised by the scanning ribosome as start codons which can initiate translation [13]. Translation from an upstream start codon may have one of multiple effects (Fig. 2A). uORFs are encoded when a start codon has an in-frame stop codon before the start of the CDS (i.e. within the 5'UTR). Translation of a uORF can be followed either by ribosome dissociation from the mRNA (therefore decreased CDS translation) or continued scanning before re-initiation of translation at the downstream CDS [14]. If there is no in-frame stop codon, translation from an upstream start would overlap the CDS. If this upstream start is in-frame with the CDS start codon, translation from it will result in an N-terminally elongated protein (N-terminal extension, NTE). Alternatively, if the upstream start is out-of-frame with the CDS, an upstream overlapping ORF (uoORF) will be translated out-of-frame with the CDS, terminating within the CDS or past the CDS stop codon [15, 16]. Translation from upstream start codons is generally assumed to repress CDS translation, as active translation of uORFs has been shown to reduce downstream CDS translation by up to 80% [15]. The prospect

of a uAUG initiating translation by influencing the recognition of the start codon by the ribosome is dependent on the local sequence context, known as the Kozak consensus sequence [16]. Around 43% of genes have one or more uAUGs in their 5'UTR [3], and these uAUGs are conserved to a significantly greater degree than any other triplet in 5'UTRs [17]. Creation of new uAUGs has been shown to be under strong negative selection. In particular, variants that create uoORFs or NTEs are, on aggregate, as deleterious as missense variants [18].

Variants that create out-of-frame uoORFs have been found across a range of diseases. This includes the recessive gene *SLC22A*, causing carnitine deficiency (*SLC22A*:NM_003060.4:c.-149G > A), where multiple patients were compound heterozygous, with one uAUG-creating variant in the 5'UTR and one CDS variant both predicted to reduce protein levels [19]. Other examples include *MEF2C* (NM_002397.5:c.-66A > T) and *NF1* (NM_001042492.3:c.-280C > T), where uoORF-creating variants were found to cause severe developmental disorders [4] and neurofibromatosis type I [18], respectively. In each case, translation of a uoORF leaves the ribosome unavailable to translate the CDS. If the uoORF is translated at high levels due to creation of an upstream start codon into a favourable sequence context, this can cause a complete loss of CDS translation. Conversely, if initiation of translation at the upstream start codon is incomplete, due to a process termed 'leaky scanning' [9], the variant will only result in a partial decrease in protein levels. Whether such hypomorphic variants have a large enough effect to cause disease is dependent on the level of dosage sensitivity of each individual gene, complicating variant interpretation.

uORF-creating variants have also been reported to cause disease, again acting to lower CDS translation. Examples include variants in *NIPBL* causing Cornelia de Lange syndrome (NM_133433.3:c.-457_456delinsAT) [20] and in *TWIST1* (NM_000474.3:c.-263C > A) causing Saethre-Chotzen syndrome [21]. uORF-creating variants are, however, even more difficult to interpret as the effect of introducing a new uORF on CDS translation is hard to predict. Not only does the uAUG context impact the strength of translation of the uORF, but also the length of the uORF created, and the distance from the end of the uORF to the CDS affect the chance of the ribosome re-initiating translation downstream at the CDS start codon. Notably, the interpretation of all uAUG-creating variants is further complicated by the complex makeup of 5'UTRs. For example, if a uAUG is created within an already highly translated uORF, it will likely not be 'seen' by scanning ribosomes as a potential start site. Alternatively, creation of a uORF which then prevents translation of an existing uORF with a more repressive effect on translation, could result in up-regulation of CDS translation.

Table 1. Examples of UTR variants in Mendelian disease.

Region	Mechanism	Example (Variant) GRCh38	Example (Disease)	ClinVar ID	Reference
5'UTR	Promoter disruption through methylation	<i>BRCA1</i> :NM_007294.4:c.-107A > T chr17:43125358-T-A	Breast and ovarian cancer	1685584	[45]
5'UTR	Altered Transcription Start Site	<i>MERTK</i> :NM_006343.3:c.-125G > A chr2:111898611-G-A	Inherited retinal disease	NA	[5]
5'UTR	Aberrant Splicing	<i>GJB1</i> :NM_000166.6:c.-16-8_-14del chrX:71223684_71223694del	Charcot-Marie-Tooth Disease	NA	[28]
5'UTR	Aberrant Splicing	<i>PAX6</i> :NM_001368894.2:c.-128-2del chr11:31806926-CT-C	Aniridia	430969	[27]
5'UTR	Create NTE	<i>IFTM5</i> :NM_001025295.3:c.-14C > T chr11:299504-G-A	Osteogenesis imperfecta Type V	37143	[22]
5'UTR	Create NTE	<i>MEF2C</i> :NM_002397.5:c.-8C > T chr5:88119613-G-A	MEF2C haploinsufficiency	1202675	[4]
5'UTR	Create uoORF	<i>MEF2C</i> :NM_002397.5:c.-66A > T chr5:88823854-T-A	MEF2C haploinsufficiency	1272072	[4]
5'UTR	Create uoORF	<i>SLC22A</i> :NM_003060.4:c.-149G > A chr5:132369824-G-A	Carnitine deficiency	25340	[19]
5'UTR	Create uoORF	<i>NF1</i> :NM_001042492.3:c.-280C > T chr17:31095030-C-T	Neurofibromatosis type I	2697306	[25]
5'UTR	Create uORF	<i>TWIST1</i> :NM_000474.3:c.-263C > A chr7:9117584-G-T	Saethre-Chotzen syndrome	NA	[21]
5'UTR	Create uORF	<i>NIPBL</i> :NM_133433.3:c.-457_-456delinsAT chr5:36876903-36876904delinsAT	Cornelia de Lange syndrome	1300231	[20]
5'UTR	uoORF removing	<i>EPHB1</i> :NM_004441.5:c.-211A > G chr3:134795421-A-G	Breast and colon cancer	NA	[23]
5'UTR	uORF removing	<i>MAP2K6</i> :NM_002758.4:c.-245T > C chr17:69414740-T-C	Colon cancer	NA	[23]
5'UTR	uoORF -> uORF	<i>THPO</i> :NM_000460.4:c.-31G > T chr3:184376290-C-A	Hereditary thrombocythemia	9510	[24]
5'UTR	uORF -> uoORF	<i>NF1</i> :NM_001042492.3:c.-272G > C chr17:31095038-G-C	Neurofibromatosis type 1	1379698	[25]
5'UTR	uORF -> uoORF	<i>NF2</i> :NM_000268.4:c.-66_-65insT chr22:29999922-A-AT	Neurofibromatosis type 2	NA	[18]
5'UTR	IRES Disrupting	<i>GJB1</i> :NM_000166.6:c.-103C > T chrX:71223249-C-T	Charcot-Marie-Tooth Disease	217166	[33]
5'UTR	IRES Disrupting	<i>c-MYC</i> :NM_002467.6:c.577C > T chr8:127738794-C-T	Multiple myeloma	NA	[35]
5'UTR	Repeat Expansion	<i>FMR1</i> :NM_002024.6:c.-128GGC[200] chrX:147912049-147912050	Fragile X	183387	[38]
5'UTR	Repeat Expansion	<i>PTCH1</i> :NM_000264.5:c.-6_-4dup chr9:95508364-T-TGCC	Basal cell carcinoma	132684	[40]

Table 1. continued

Region	Mechanism	Example (Variant) GRCh38	Example (Disease)	ClinVar ID	Reference
5'UTR	Secondary Structure Altered	<i>FTL</i> :NM_000146.4:c.-164C > T chr19:48965344-C-T	Hyperferritinemia/ cataract syndrome	96691	[49]
5'UTR	Secondary Structure Altered	<i>ADAR1</i> :NM_001111.5:c.-60A > G chr1:154608066-T-C	Dyschromatosis symmetrica hereditaria	NA	[50]
5'UTR	CDS start codon consensus sequence altered	<i>RARS2</i> :NM_020320.5:c.2T > C chr6:87589956-A-G	Pontocerebellar hypoplasia	1452675	[52]
3'UTR	Disrupted Polyadenylation and cleavage	<i>NAA10</i> :NM_003491.4:c.*43A > G chrX:153929948-T-C	Syndromic microphthalmia	617462	[54]
3'UTR	Disrupted Polyadenylation cleavage signal	<i>F2</i> :NM_000506.5:c.*97G > A chr11:46739505-G-A	Thrombophilia	13310	[55]
3'UTR	CDS stop codon readthrough affecting poly(A) tail	<i>HBA1</i> :NM_000558.5:c.427T > C chr16:177409-T-C	α-thalassemia	15624	[56]
3'UTR	miRNA binding site altered	<i>COL4A1</i> :NM_001845.6:c.*31G > T chr13:110150332-C-A	Cerebral small vessel disease	689433	[59]
3'UTR	miRNA binding site altered	<i>REEP1</i> :NM_001371279.1:c.808C > T chr2:86217086-G-A	Hereditary spastic paraparesis type 31	1166866	[60]
3'UTR	Repeat Expansion	<i>DMPK</i> :NM_004409.5:c.*224CTG[330] chr19:45770204-45770205[GCA]	Myotonic dystrophy type 1	810828	[42]
3'UTR	Repeat Expansion	<i>ATXN8OS</i> :NR_002717.2:n.1103CTG[107_127] chr13:70139384-70139386[CTG]	Spinocerebellar ataxia type 8	562101	[43]
3'UTR	Secondary Structure Altered	<i>SEPN1</i> :NM_020451.3:c.*1107T > C chr1:25816825-T-C	SEPN1-related myopathy	844320	[51]
3'UTR	Aberrant Splicing	<i>F8</i> :NM_000132.4:c.*56G > T chrX:154837541-C-A	Haemophilia A	1163207	[31]

Summary of identified UTR variants, their predicted mode of disruption, and corresponding ClinVar accession IDs.

While uORF and uoORF create variants that impact translational regulation, variants that create uAUGs in frame with the CDS that result in NTE may be more likely to disrupt protein function. For example, a variant (NM_001025295.3:c.-14C > T) that extends the N-terminus of *IFTM5* by five amino acids was found recurrently in individuals with osteogenesis imperfecta type V [22]. In this case, the addition of the extra amino acids renders the protein to be non-functional. Similarly, two distinct variants that add three (NM_002397.5:c.-8C > T) and nine (NM_002397.5:c.-26C > T) amino acids, respectively, to the start of the MEF2C protein cause MEF2C haploinsufficiency and severe developmental disorders [4]. *MEF2C* cannot tolerate the addition of amino acids at the N-terminus, as this likely disrupts the binding of this transcription factor to DNA, abolishing its function.

VARIANTS IMPACT EXISTING UORFS OR UOORFS CAN DISRUPT TRANSLATIONAL REGULATION

The above examples of uAUG-creating variants are all predicted to decrease CDS translation. Conversely, removing an existing

upstream start codon is more likely to result in an increase in protein levels and have a gain-of-function effect. Upstream start-codon removing variants in *EPHB1* (breast and colon) (NM_004441.5:c.-211A > G) and *MAP2K6* (colon) (NM_002758.4:c.-245T > C) in cancer samples (removing an uoORF and uORF, respectively) were found to be associated with enhanced translation, suggesting that loss-of-uAUG mediated translational increase of the downstream main protein-coding sequence may contribute to carcinogenesis [23].

Variants may also alter the inhibitory effect of uORFs or uoORFs. For example, a variant in the 5'UTR of *THPO* (NM_000460.4:c.-31G > T) turns a uoORF into a uORF through the creation of a stop codon within the uoORF sequence. The native uoORF is strongly translated, leading to a necessary low production of *THPO*. *THPO* encodes thrombopoietin, required for normal functioning of the pathway controlling the production of platelets. Turning the uoORF into a uORF (with the possibility of re-initiation) increases protein levels, causing hereditary thrombocytopenia [24]. Contrastingly, changing a highly translated uORF into a uoORF can

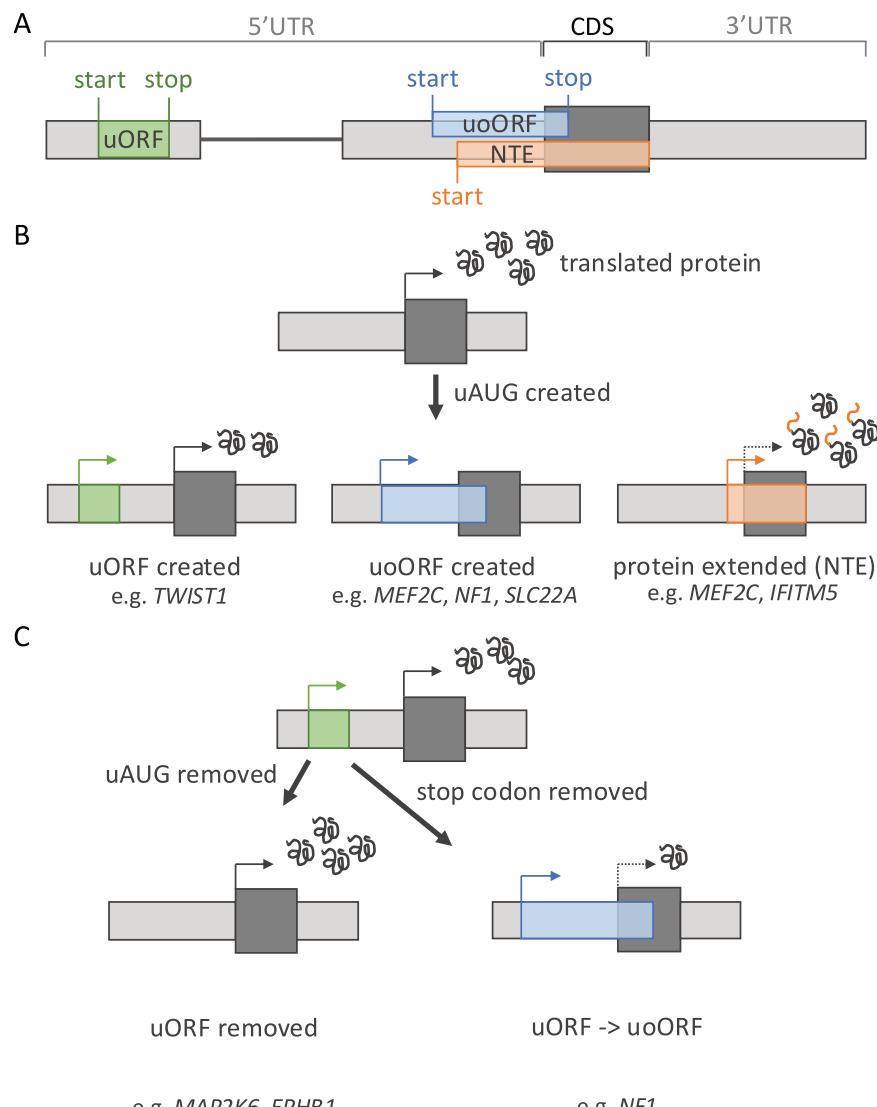


Fig. 2 Creation and disruption of upstream open reading frames (uORFs) in disease. **A** Schematic of the different elements encoded by an upstream start codon. uORFs are encoded by an upstream start with an in-frame stop codon also within the 5'UTR. When there is no in-frame stop within the 5'UTR, either an upstream overlapping ORF (uoORF) is formed when the start codon is in a different reading frame to the coding sequence (CDS), or the CDS is extended at the N-terminus (N-terminal extension; NTE). **B** Depiction of the different impacts of variants that create uORFs, uoORFs, or NTEs on CDS translation. **C** Depiction of two types of uORF-perturbing variants on CDS translation. Removing the start codon of a uORF (left) is predicted to increase CDS translation, whereas removing the stop codon of a uORF, resulting in its extension to form a uoORF (i.e. if there is no alternative in-frame stop codon), is predicted to decrease, or abolish, CDS translation.

increase its inhibitory effect. This can be achieved through two different mechanisms. An example of the first is seen in *NF1*, where variants (e.g. NM_001042492.3:c.-272G > C) that remove the stop codon of a native uORF, transforming it into an uoORF (as there are no alternative in-frame stop codons before the CDS), have been observed in patients with neurofibromatosis type I [25]. The second mechanism is seen in *NF2*, the 5'UTR of which contains a native uORF with prior evidence of translation and a strong predicted Kozak consensus. A single base insertion (NM_000268.4:c.-66_-65insT) changes the frame of the uORF, causing it to bypass the downstream stop codon and create an out-of-frame uoORF. Translation of this uoORF is predicted to lower translation of *NF2*, consistent with haploinsufficiency causing neurofibromatosis type 2 [18]. Similarly to uAUG-creating variants, interpretation of variants that alter existing uORFs can be complex. The difference in translational repression conveyed by uORFs and uoORFs, and uORFs of different lengths, is difficult to predict.

VARIANTS CAN CAUSE DISEASE BY ALTERING UTR SPlicing

Variants within UTRs that interfere with splicing can cause disease through a variety of mechanisms (Fig. 3). Approximately 38% of 5' UTRs contain introns, with the number of introns ranging to 11 [3]. Alternative splicing of 5'UTRs is known to impact mRNA stability and translation [26].

Several variants affecting splicing within the 5'UTR of *PAX6* are considered pathogenic for aniridia [27]. These variants (e.g. NM_001368894.2:c.-128-2del) induce exon skipping or splicing errors around exons 2 and 3 of the 5'UTR. The hypothesised mechanism of disease is through uORF dysregulation (Fig. 3A); there is a uORF crossing these exons, and the variants reportedly change the uORF frame, which, similar to the *NF2* example above, turns the uORF into a more inhibitory uoORF, leading to loss-of-function and disease [27].

Altering 5'UTR splicing can also impact the CDS sequence. For example, variants that disrupt the final acceptor site in the 5'UTR can lead to skipping or truncation of the exon containing the

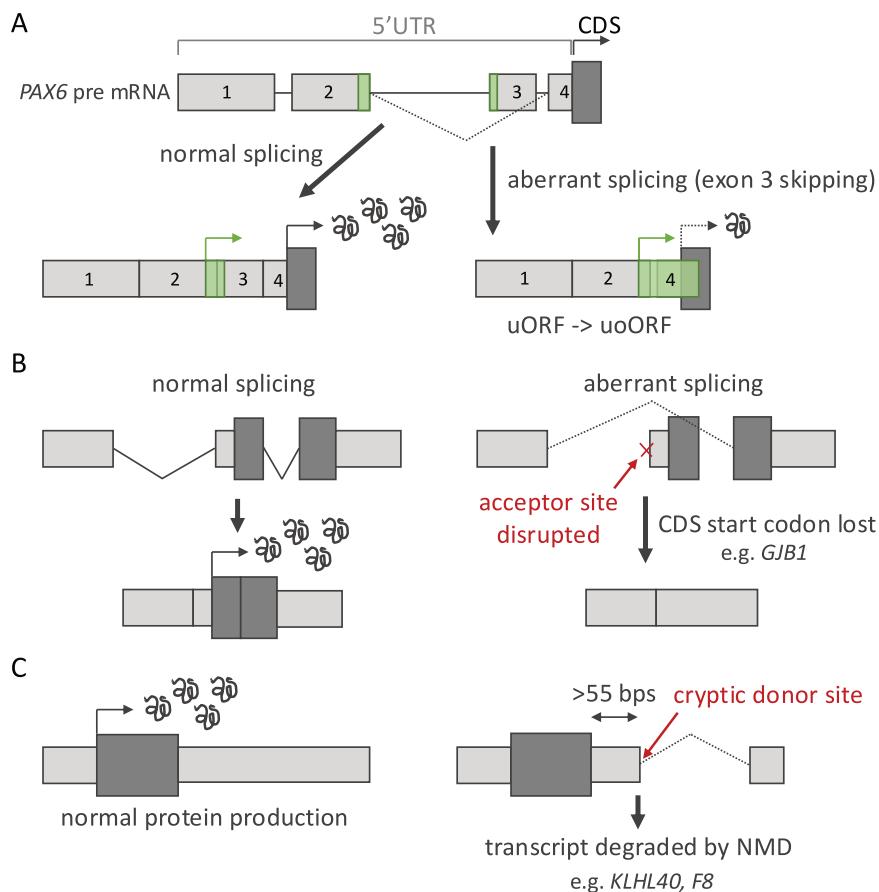


Fig. 3 Schematic showing how UTR variants can cause disease through disrupting RNA splicing. **A** 5'UTR variants in *PAX6* are a common cause of aniridia. These variants are thought to cause skipping of exon three, which contains the stop codon of an upstream open reading frame (uORF), converting it into an upstream overlapping ORF (uoORF) and resulting in lower translation of the *PAX6* coding sequence (CDS). **B** Variants that disrupt the acceptor splice site of the final exon in the 5'UTR remove the start codon of the CDS. **C** 3'UTR variants that create cryptic donor splice sites further than 55 bps downstream of the end of the CDS are predicted to lead to transcript degradation through nonsense-mediated decay (NMD).

canonical CDS start codon. One such example is in *GJB1* (NM_000166.6:c.-16-8_-14del), where this mechanism causes a significant amount (278 bps) of the following exon to be removed, including 262 bps of the CDS (31%; Fig. 3B). This variant is reported to lead to Charcot-Marie-Tooth disease [28].

While splicing is common in 5'UTRs, only around 6% of 3'UTRs contain introns [29]. The vast majority of these introns are very close to the CDS stop codon, as introns >50/55 bps downstream of the CDS would be predicted to cause transcript degradation through the nonsense-mediated decay (NMD) pathway [30]. Variants that create new introns into 3'UTRs can trigger NMD and result in loss of protein expression. An example is a variant in *F8* (NM_000132.4:c.*56G > T) that creates a new donor splice site, resulting in a 159 bp deletion in the 3'UTR, demonstrated to reduce expression and therefore thought to cause mild haemophilia A [31].

VARIANTS IN INTERNAL RIBOSOME ENTRY SITES (IRES) MAY IMPACT RIBOSOMAL RECRUITMENT

A subset of mRNAs can initiate translation in a cap-independent manner via internal ribosome entry sites (IRESs). These are specialised sequences within 5'UTRs that can directly recruit 40S ribosomal subunits to initiate scanning from within the 5'UTR independently of the 5'cap [32]. Notably, IRES motifs differ between genes so are difficult to predict, limiting our ability to annotate and interpret variants in these elements. In addition,

while there are documented examples of IRES in many organisms, the evidence supporting their widespread existence in humans is limited. Potentially as a consequence, there are very few mentions of IRES's role in disease in the published literature. Exceptions include a variant (NM_000166.6:c.-103C > T) in *GJB1* segregating with Charcot-Marie-Tooth disease. The native IRES of *GJB1* is essential for translation of connexin-32 mRNA in nerve cells, and the variant in this case was reported to abolish the IRES's function, leading to no translation, as was discovered via in-depth in vivo analysis using bicistronic reporters [33]. However, the existence of this IRES was thrown into doubt by more recent experimental work [34]. A further example is in the proto-oncogene *c-myc*, where in patients with multiple myeloma, a C > T substitution (NM_002467.6:c.577 C > T) within a reported IRES causes higher IRES activity and increased production of *c-myc* protein. The underlying mechanism is not fully understood, but experimental data suggested that *c-myc* IRES trans-acting factors (Y-box binding protein 1 (YB-1) and polypyrimidine tract-binding protein 1 (PTB-1)) bind more strongly to the mutated version of *c-myc* IRES [35]. Of note, newer transcript annotations categorise this variant as a missense variant within the CDS.

REPEAT EXPANSIONS IN UTRS CAN DISRUPT REGULATION AND CAUSE TOXIC PEPTIDE PRODUCTION

Repeat expansions (also known as microsatellites or simple sequence repeats) are a unique class of variation. The nucleotide

sequence, location within the gene, range of repeat length and clinical outcomes vary between repeats. Repeat expansions can be pathogenic when located in non-coding regions, causing so-called non-coding repeat expansion disorders [36], with many examples within UTRs. The pathogenic mechanisms vary, but the main possible mechanisms are as follows (reviewed in [37]): repeat sequences can form intramolecular structures that can influence transcription, translation and binding to various RBPs; GC-rich repeat sequences are also prone to hypermethylation that can cause gene silencing; more rarely, through repeat associated non-ATG (RAN) translation, the repeat RNA itself might be unconventionally translated into toxic peptides.

CGG repeat expansions in the 5'UTR of *FMR1* (NM_002024.6:c.-128GGC[200]) lead to hypermethylation and silencing of the gene. This results in insufficient amounts of *FMR1* protein that is required for neuronal development and causes Fragile X syndrome [38]. A CGG repeat expansion in the *GIPC1* 5'UTR is associated with oculopharyngodistal myopathy. The number of CGG repeats is <30 in controls but >60 in affected individuals [39]. However, genes can be even more sensitive to the number of repeats; insertion of a surplus eighth CGG triplet in the *PTCH1* 5'UTR (NM_000264.5:c.-6_-4dup) represses protein translation drastically compared to the wild-type seven-time repeat sequence. This non-coding variant is reported to predispose to basal cell carcinoma [40].

There are also multiple instances of pathogenic repeat expansions within 3'UTRs. For example, a large CTG repeat (NM_004409.5:c.*224CTG[330]) in the 3'UTR of *DMPK* causes myotonic dystrophy type 1 (DM1) via a toxic gain-of-function mechanism [41]. It is hypothesised that the mutant *DMPK* transcripts form aberrant structures and anomalously associate with RBPs [42]. Similarly, spinocerebellar ataxia type 8 (SCA8) can be caused by a CTG expansion (NR_002717.2:n.1103CTG[107_127]) in the 3'UTR of *ATXN8OS* [43]. Here, there is evidence for both toxic RNA and toxic protein effects, as RAN translation can generate a polyglutamate protein from the antisense strand [37].

5'UTR VARIANTS THAT OVERLAP THE PROMOTER CAN DISRUPT TRANSCRIPTION INITIATION

Promoters are sequences of DNA that initiate transcription of a gene. Relevant proteins, termed transcription factors, bind to promoter motifs to initiate transcription [44]. Promoters typically span the TSS of a gene, which also marks the beginning of the 5'UTR. Hence, the portion of the promoter that is downstream of the TSS may also overlap the 5'UTR. A variant (NM_007294.4:c.-107A > T) in the 5'UTR-overlapping promoter region of *BRCA1* observed in two unrelated families is associated with epigenetic silencing through allele-specific promoter methylation and is thought to lead to grade 3 breast cancer or high-grade serous ovarian cancer [45]. However, another study questions this: the authors looked for this variant in a larger cohort of patients with *BRCA1* allele-specific promoter methylation and did not find this variant [46]. Another example is a variant (NM_006343.3:c.-125G > A) reported to overlap the TSS of *MERTK* was hypothesised to disrupt transcription, alter secondary structure and cause inherited retinal disease [5]. This was supported by a reduction in mRNA levels in a luciferase reporter assay.

VARIANTS CAN CAUSE DISEASE BY DISRUPTING mRNA SECONDARY STRUCTURE

mRNA is a single-stranded RNA sequence capable of folding and forming secondary structures through complementary base-pairing with itself. These structures can impact translation by causing inefficient ribosomal scanning and affecting mRNA stability [47]. There are several types of secondary structures, including pseudoknots, G-quadruplexes and hairpin structures

(also referred to as stem-loops), which occur when the mRNA strand folds and base pairs with an adjacent section. It is difficult to predict the exact conformation and structure of mRNA based on sequence alone and secondary structures are often dynamic.

One of the best-studied examples of an important stem-loop is the iron-responsive element (IRE), which affects translation of mRNAs important for iron homeostasis. Cellular iron uptake must be tightly regulated, as insufficient or excess levels of iron can be damaging. A single conserved IRE stem-loop close to the 5'cap of mRNA is bound by iron-regulatory protein 1 or 2 (IRP1/IRP2). IRP binding represses translation initiation by blocking ribosome access to the 5'UTR. The IRPs register iron availability through direct interactions with iron in the cytosol and alter their binding and hence translational inhibition in response [48]. L-ferritin binds and stores iron in the cell. Hyperferritinemia/cataract syndrome (HHCS) is caused by mutations (e.g. NM_000146.4:c.-164C > T) in the IRE in the 5'UTR of the L-ferritin (*FTL*) gene, preventing interaction with the IRPs and leading to dysregulated high levels of L-ferritin production [49].

Another example of altering mRNA secondary structure involves a heterozygous variant (NM_001111.5:c.-60A > G) in the 5'UTR of *ADAR1* that reduces gene expression and is reported to cause dyschromatosis symmetrica hereditaria. The variant does not appear to disrupt any known regulatory features in that region; however, this single-nucleotide change is thought to be sufficient to alter the structure of the mRNA [50]. Exactly how this structural change results in reduced gene expression is not currently known.

SEPN1-related myopathy consists of four autosomal recessive disorders. *SEPN1* produces selenoprotein, which is required for normal muscle development. A unique feature of all selenoproteins is the presence of the amino acid selenocysteine, unusually encoded by the CDS stop codon, UGA. This re-coding of UGA is made possible by a highly-conserved stem-loop structure that starts 6 bp downstream of the CDS UGA codon, in the 3'UTR. This region is known as the selenocysteine insertion sequence (SECIS). A SECIS RBP (SBP2), which binds SECIS, is vital in the redefinition of the UGA to a selenocysteine codon, preventing termination at the UGA. Termination at the UGA triggers NMD and leads to insufficient protein. Three variants (e.g. NM_020451.3:c.*1107T > C) specifically within the stem-loop structure of the SECIS have been linked to *SEPN1*-related myopathy by interfering with SBP2 binding, significantly reducing both mRNA and protein levels [51].

5'UTR VARIANTS CAN ALTER THE EFFICIENCY OF TRANSLATION INITIATION AT THE CDS START CODON

The exact context surrounding the CDS start codon dictates the strength of translation from that AUG and hence the amount of protein produced. Variants within the 5'UTR that disrupt this Kozak consensus sequence can influence disease risk. A recent paper by Nicolle et al. describes a variant in the 5'UTR of *RARS2* (NM_020320.5:c.2T > C) which alters the Kozak sequence of the CDS start codon, reducing protein production and causing pontocerebellar hypoplasia (PCH) [52]. The authors also assessed ClinVar for other variants that are predicted to alter the Kozak sequence of the CDS and found 20, most of which are denoted as variants of unknown significance. They conclude that variants of this class are likely an underappreciated mechanism of disease.

VARIANTS IN 3'UTRS CAN AFFECT POLYADENYLATION

For most protein-coding genes, the 3' end of the pre-mRNA is formed by consecutive cleavage and polyadenylation that occurs co-transcriptionally. This is a widely studied phenomenon and is reviewed by Curinha et al. [53] and summarised here. 3'UTRs contain a polyadenylation site (polyA site (PAS)) that directs the addition of several hundred adenine residues to create the polyA tail at the end stage of transcription. The polyA tail has an

important role in nuclear export, translation and the stability of the mRNA. Usage of a specific PAS in the pre-mRNA is directed by RNA cis-elements and several trans-acting factors. The most important cis-element is the polyA signal, a hexamer (usually AAUAAA or a close variant such as UAUAAA) located ~10–35 bp upstream of the PAS.

Multiple variants affecting polyadenylation are associated with disease [53]. For example, several variants within the PAS of *NAA10* (e.g. NM_003491.4:c.*43A > G) have been linked to syndromic X-linked microphthalmia. In vitro studies demonstrated that these variants disrupt cleavage and polyadenylation and lead to reduced mRNA levels [54]. Conversely, a gain-of-function single-nucleotide variant at the terminal end of the *F2* 3'UTR (NM_000506.5:c.*97G > A) causes elevated prothrombin plasma levels. The wild-type polyadenylation cleavage signal is inefficient, but the variant increases cleavage site recognition, increases 3' end processing, mRNA accumulation and protein synthesis, leading to thrombophilia [55].

While not strictly a 3'UTR variant, a single-nucleotide variant that changes the α -globin (*HBA1*) CDS stop codon from a UAA to CAA (NM_000558.5:c.427T > C) allows the translating ribosomes to proceed into the 3'UTR. This is associated with a substantial decrease in α -globin mRNA half-life. Further experiments elucidated that there are C-rich regions that interact with a ribonucleoprotein complex, the α -complex, containing proteins that prevent degradation of the mRNA. The α -complex is thought to protect the poly(A) tail and stabilise the mRNA. If this interaction is prevented, such as in this case, the poly(A) tail undergoes accelerated shortening; the mRNA is prematurely degraded and α -thalassemia ensues [56].

VARIANTS DISRUPTING MICRORNAs AND THEIR BINDING SITES ALTER RNA SILENCING

miRNAs are non-coding RNAs ~22 bp long which bear one or more hairpin loops. They are involved in RNA silencing and post-transcriptional regulation of gene expression. miRNAs base pair to complementary sequences on the mRNA, usually within 3'UTRs, and can silence and inhibit protein production, for example, through mRNA deadenylation and decapping [57].

There are multiple examples of pathogenic variants within miRNAs themselves, specifically within the 'seed region' that is responsible for mediating mRNA binding [58]. In addition, variants in miRNA-binding sites within 3'UTRs can disrupt miRNA-mediated regulation. For example, variants within a 7 bp region of the *COL4A1* 3'UTR (e.g. NM_001845.6:c.*31G > T) abolish a binding site for the miR-29 miRNA, leading to increased levels of *COL4A1* mRNA and leading to cerebral small vessel disease [59]. Conversely, two variants in the miR-140 binding site in the 3' UTR of *REEP1* (e.g. NM_001371279.1:c.808C > T) are predicted to suppress miRNA-mediated effects on translation, leading to a decrease in *REEP1* protein. This causes hereditary spastic paraparesis type 31 [60].

DISCUSSION

Here, we systematically reviewed important regulatory elements in 5' and 3' UTRs that, when perturbed, can cause Mendelian disease. These regions have historically been overlooked, but with the advancement and inclusion of whole-genome sequencing in clinical diagnostic sequencing, these regions are becoming more accessible. Recent recommendations for clinical interpretation of non-coding variants [2] aid in classifying UTR variants in a clinical setting; however, annotating and interpreting these variants remains a considerable challenge.

Annotating variants within UTRs currently requires using a range of bioinformatics tools and datasets. Each tool can annotate variants according to a particular hypothesised effect, for example,

UTRannotator [61], but there is no single solution that combines all known regulatory elements and variant mechanisms. Each variant may also have multiple predicted effects, for example, creating a uAUG and altering a transcription regulatory factor binding site [62]. Determining exactly how a variant is mediating an effect on protein expression can be difficult without extensive functional characterisation.

Currently, UTR variants are most often annotated with respect to known regulatory elements; however, these elements may be incompletely annotated, especially if they are tissue or temporally specific. In addition, there may be further, as yet unknown categories of regulatory elements through which variants may mediate their effect. This is highlighted by reportedly disease-causing UTR variants that act via unknown mechanisms. An example is a 96 bp deletion in the 3'UTR of *VMA21* that segregates with X-linked myopathy with excessive autophagy (XMEA) and was demonstrated to reduce mRNA quantity. The underlying mechanism is unknown but may involve destabilisation of the mRNA [63].

A key challenge in interpreting UTR variants, and non-coding variants more broadly, is that they often have incomplete, or hypomorphic, effects. In addition, these effects can be in either direction, resulting in either an increase or a decrease in protein levels, as demonstrated by many of the variants detailed above. For example, uORF-creating variants in *MEF2C* decrease protein expression [4] and variants in the *THPO* 5'UTR increase protein expression [24], both leading to respective diseases. The threshold at which increased or decreased protein expression leads to disease is highly gene-specific, but these thresholds are not known for the vast majority of genes. For partial effect non-coding variants that do cause disease, they may result in milder phenotypes, for example, variants in the CDS of *KLHL40* are linked to severe forms of nemaline myopathy [64], whereas a variant in the 5'UTR of *KLHL40* is linked to a milder form of disease [65]. Additionally, partial effect non-coding variants that do cause disease may also result in later disease onset and/or have reduced penetrance [66].

Interpretation of UTR variants is also complicated by the need to consider wider sequence context. UTRs are complex regulatory elements with precise combinations of regulatory elements. Similarly to the redundancy observed in enhancers [67], other regulatory elements may be able to compensate if one element is disrupted by a genetic variant. This may be the case for miRNA-binding sites, for example, where multiple binding sites for the same miRNA can be found within the same 3'UTR [68]. Consideration of wider sequence context is also critical in deciphering the impact of 5'UTR variants that disrupt translational regulation. For example, not all upstream start codon creating variants are created equal: the chance of translation from an upstream start is dependent not only on the surrounding sequence match to the Kozak consensus, but also on the position of other translated uORFs [69]. For example, if the start is created within a ubiquitously translated uORF it will not be 'seen' by a scanning ribosome and hence will not have any impact on downstream CDS translation.

Given the challenges in the interpretation of UTR variants, functional characterisation is important. In addition to experiments to decipher the effect of individual variants observed in patients, large-scale multiplexed assays of variant effect hold great promise not only for variant interpretation but also for us to learn more about gene regulation [70]. Importantly, however, variants in UTRs can act through multiple different mechanisms: impacting transcript, RNA-processing, RNA stability and translation. A variant impact cannot be ruled out unless all these mechanisms have been captured in the experiments performed [2]. Alternatively, if an assay focuses on a downstream impact such as cell viability, the exact mechanism of effect may remain unclear. Further, as discussed above, UTRs act as entire functional elements. Including

the entire sequence in its native context is likely important for accurate functional characterisation.

Variants in 5' and 3' UTRs are a rare but underappreciated cause of Mendelian disease. Annotating and interpreting these variants in clinical settings is challenging but can result in a diagnosis for patients and, in turn, increase our knowledge of UTR-mediated gene regulation.

REFERENCES

- 100,000 Genomes Project Pilot Investigators, Smedley D, Smith KR, Martin A, Thomas EA, McDonagh EM, et al. 100,000 genomes pilot on rare-disease diagnosis in health care—preliminary report. *N Engl J Med.* 2021;385:1868–80.
2. Ellingford JM, Ahn JW, Bagnall RD, Baralle D, Barton S, Campbell C, et al. Recommendations for clinical interpretation of variants found in non-coding regions of the genome. *Genome Med.* 2022;14:73.
3. Wieder N, D’Souza EN, Martin-Geary AC, Lassen FH, Talbot-Martin J, Fernandes M, et al. Differences in 5' untranslated regions highlight the importance of translational regulation of dosage sensitive genes. *Genome Biol.* 2024;25:111.
4. Wright CF, Quaife NM, Ramos-Hernández L, Danecek P, Ferla MP, Samocha KE, et al. Non-coding region variants upstream of MEF2C cause severe developmental disorder through three distinct loss-of-function mechanisms. *Am J Hum Genet.* 2021;108:1083–94.
5. Dueñas Rey A, del Pozo Valero M, Bouckaert M, Wood KA, Van den Broeck F, Daich Varela M, et al. Combining a prioritization strategy and functional studies nominates 5'UTR variants underlying inherited retinal disease. *Genome Med.* 2024;16:7.
6. Steri M, Idda ML, Whalen MB, Orrù V. Genetic variants in mRNA untranslated regions. *WIREs RNA.* 2018;9:e1474.
7. Griesemer D, Xue JR, Reilly SK, Ulirsch JC, Kukreja K, Davis JR, et al. Genome-wide functional screen of 3'UTR variants uncovers causal variants for human disease and evolution. *Cell.* 2021;184:5247–60.e19.
8. Shirokikh NE, Preiss T. Translation initiation by cap-dependent ribosome recruitment: recent insights and open questions. *Wiley Interdiscip Rev RNA.* 2018;9:e1473.
9. Hinnebusch AG, Ivanov IP, Sonenberg N. Translational control by 5'-untranslated regions of eukaryotic mRNAs. *Science.* 2016;352:1413–6.
10. Floor SN, Doudna JA. Tunable protein synthesis by transcript isoforms in human cells. *eLife.* 2016;5:e10921.
11. Mayr C. Regulation by 3'-untranslated regions. *Annu Rev Genet.* 2017;51:171–94.
12. Hentze MW, Castello A, Schwarzl T, Preiss T. A brave new world of RNA-binding proteins. *Nat Rev Mol Cell Biol.* 2018;19:327–41.
13. Chothani SP, Adami E, Widjaja AA, Langley SR, Viswanathan S, Pua CJ, et al. A high-resolution map of human RNA translation. *Mol Cell.* 2022 [cited 2022 Jul 25]. Available from: [https://www.cell.com/molecular-cell/abstract/S1097-2765\(22\)00606-2](https://www.cell.com/molecular-cell/abstract/S1097-2765(22)00606-2).
14. Gunišová S, Hronová V, Mohammad MP, Hinnebusch AG, Valášek LS. Please do not recycle! Translation reinitiation in microbes and higher eukaryotes. *FEMS Microbiol Rev.* 2018;42:165–92.
15. Zhang H, Wang Y, Wu X, Tang X, Wu C, Lu J. Determinants of genome-wide distribution and evolution of uORFs in eukaryotes. *Nat Commun.* 2021;12:1076.
16. Kozak M. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* 1987;15:8125–48.
17. Churbanov A, Rogozin IB, Babenko VN, Ali H, Koonin EV. Evolutionary conservation suggests a regulatory function of AUG triplets in 5'-UTRs of eukaryotic genes. *Nucleic Acids Res.* 2005;33:5512–20.
18. Whiffin N, Karczewski KJ, Zhang X, Chothani S, Smith MJ, Evans DG, et al. Characterising the loss-of-function impact of 5' untranslated region variants in 15,708 individuals. *Nat Commun.* 2020;11:2523.
19. Ferdinandusse S, te Brinke H, Ruiter JPN, Haasjes J, Oostheim W, van, et al. A mutation creating an upstream translation initiation codon in SLC22A5 5' UTR is a frequent cause of primary carnitine deficiency. *Hum Mutat.* 2019;40:1899–904.
20. Coursimault J, Rovelet-Lecrux A, Cassinari K, Brischoux-Boucher E, Saugier-Veber P, Goldenberg A, et al. uORF-introducing variants in the 5'UTR of the NIPBL gene as a cause of Cornelia de Lange syndrome. *Hum Mutat.* 2022;43:1239–48.
21. Zhou Y, Koelling N, Fenwick AL, McGowan SJ, Calpene E, Wall SA, et al. Disruption of TWIST1 translation by 5' UTR variants in Saethre-Chotzen syndrome. *Hum Mutat.* 2018;39:1360–5.
22. Cho TJ, Lee KE, Lee SK, Song SJ, Kim KJ, Jeon D, et al. A Single Recurrent Mutation in the 5'-UTR of IFITM5 Causes Osteogenesis Imperfecta Type V. *Am J Hum Genet.* 2012;91:343–8.
23. Schulz J, Mah N, Neuenschwander M, Kischka T, Ratei R, Schlag PM, et al. Loss-of-function uORF mutations in human malignancies. *Sci Rep.* 2018;8:2395.
24. Ghilardi N, Wiestner A, Kikuchi M, Ohsaka A, Skoda RC. Hereditary thrombocythaemia in a Japanese family is caused by a novel point mutation in the thrombopoietin gene. *Br J Haematol.* 1999;107:310–6.
25. Evans DG, Bowers N, Burkitt-Wright E, Miles E, Garg S, Scott-Kitching V, et al. Comprehensive RNA analysis of the NF1 gene in classically affected NF1 affected individuals meeting NIH criteria has high sensitivity and mutation negative testing is reassuring in isolated cases with pigmentary features only. *EBioMedicine.* 2016;7:212–20.
26. Rosenstiel P, Huse K, Franke A, Hampe J, Reichwald K, Platzer C, et al. Functional characterization of two novel 5' untranslated exons reveals a complex regulation of NOD2 protein expression. *BMC Genom.* 2007;8:472.
27. Filatova AY, Vasilyeva TA, Marakhonov AV, Sukhanova NV, Voskresenskaya AA, Zinchenko RA, et al. Upstream ORF frameshift variants in the PAX6 5'UTR cause congenital aniridia. *Hum Mutat.* 2021;42:1053–65.
28. Li M, Yin M, Yang L, Chen Z, Du P, Sun L, et al. A novel splicing mutation in 5'UTR of GJB1 causes X-linked Charcot–Marie–tooth disease. *Mol Genet Genom Med.* 2023;11:e2108.
29. Bicknell AA, Cenik C, Chua HN, Roth FP, Moore MJ. Introns in UTRs: Why we should stop ignoring them. *BioEssays.* 2012;34:1025–34.
30. Nagy E, Maquat LE. A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends Biochem Sci.* 1998;23:198–9.
31. Pezeshkpour B, Berkemeier AC, Czogalla KJ, Oldenburg J, El-Maari O. Evidence of pathogenicity of a mutation in 3' untranslated region causing mild haemophilia A. *Haemophilia.* 2016 [cited 2023 Jul 13];22. Available from: <https://pubmed.ncbi.nlm.nih.gov/27216882/>.
32. Hellen CUT, Sarnow P. Internal ribosome entry sites in eukaryotic mRNA molecules. *Genes Dev.* 2001;15:1593–612.
33. Hudder A, Werner R. Analysis of a Charcot-Marie-Tooth disease mutation reveals an essential internal ribosome entry site element in the connexin-32 gene *. *J Biol Chem.* 2000;275:34586–91.
34. Grosz BR, Svaren J, Perez-Siles G, Nicholson GA, Kennerson ML. Revisiting the pathogenic mechanism of the GJB1 5' UTR c.-103C > T mutation causing CMTX1. *Neurogenetics.* 2021;22:149–60.
35. Cobbold LC, Wilson LA, Sawicka K, King HA, Kondrashov AV, Spriggs KA, et al. Upregulated c-myc expression in multiple myeloma by internal ribosome entry results from increased interactions with and expression of PTB-1 and YB-1. *Oncogene.* 2010;29:2884–91.
36. Swinnen B, Robberecht W, Van Den Bosch L. RNA toxicity in non-coding repeat expansion disorders. *EMBO J.* 2020;39:e101112.
37. Paulson H. Repeat expansion diseases. *Handb Clin Neurol.* 2018;147:105–23.
38. Zhou Y, Kumari D, Sciascia N, Usdin K. CGG-repeat dynamics and FMR1 gene silencing in fragile X syndrome stem cells and stem cell-derived neurons. *Mol Autism.* 2016;7:42.
39. Xi J, Wang X, Yue D, Dou T, Wu Q, Lu J, et al. 5' UTR CGG repeat expansion in GPC1 is associated with oculopharyngodistal myopathy. *Brain.* 2021;144:601–14.
40. Tietze JK, Pfob M, Eggert M, von Preußen A, Mehraein Y, Ruzicka T, et al. A non-coding mutation in the 5' untranslated region of patched homologue 1 predisposes to basal cell carcinoma. *Exp Dermatol.* 2013;22:834–5.
41. Timchenko LT, Miller JW, Timchenko NA, DeVore DR, Datar KV, Lin L, et al. Identification of a (CUG)n triplet repeat RNA-binding protein and its expression in myotonic dystrophy. *Nucleic Acids Res.* 1996;24:4407–14.
42. Cruchten RTP, van Wieringa B, Wansink DG. Expanded CUG repeats in DMPK transcripts adopt diverse hairpin conformations without influencing the structure of the flanking sequences. *RNA.* 2019;25:481–95.
43. Daughters RS, Tuttle DL, Gao W, Ikeda Y, Moseley ML, Ebner TJ, et al. RNA gain-of-function in spinocerebellar atrophy type 8. *PLoS Genet.* 2009;5:e1000600.
44. Haberle V, Stark A. Eukaryotic core promoters and the functional basis of transcription initiation. *Nat Rev Mol Cell Biol.* 2018;19:621–37.
45. Evans DGR, van Veen EM, Byers HJ, Wallace AJ, Ellingford JM, Beaman G, et al. A dominantly inherited 5' UTR variant causing methylation-associated silencing of BRCA1 as a cause of breast and ovarian cancer. *Am J Hum Genet.* 2018;103:213–20.
46. de Jong VMT, Pruntel R, Steenbruggen TG, Bleeker FE, Nederlof P, Hogervorst FBL, et al. Identifying the BRCA1 c.-107A > T variant in Dutch patients with a tumor BRCA1 promoter hypermethylation. *Fam Cancer.* 2023;22:151–4.
47. Andrzejewska A, Zawadzka M, Pacholska-Wieczorek K. On the way to understanding the interplay between the RNA structure and functions in cells: a genome-wide perspective. *Int J Mol Sci.* 2020;21:6770.
48. Mw H, Mu M, Nc A. Balancing acts: molecular control of mammalian iron metabolism. *Cell.* 2004 [cited 2023 Aug 9];117. Available from: <https://pubmed.ncbi.nlm.nih.gov/15109490/>.
49. Luscetti S, Tolle G, Aranda J, Campos CB, Risse F, Morán É, et al. Novel mutations in the ferritin-L iron-responsive element that only mildly impair IRE binding cause hereditary hyperferritinemia cataract syndrome. *Orphanet J Rare Dis.* 2013;8:30.

50. Suganuma M, Kono M, Yamanaka M, Akiyama M. Pathogenesis of a variant in the 5' untranslated region of ADAR1 in dyschromatosis symmetrica hereditaria. *Pigment Cell Melanoma Res.* 2020;33:591–600.
51. Maiti B, Arbogast S, Allamand V, Moyle MW, Anderson CB, Richard P, et al. A mutation in the SEPN1 selenocysteine redefinition element (SRE) reduces selenocysteine incorporation and leads to SEPN1-related myopathy. *Hum Mutat.* 2009;30:411–6.
52. Nicolle R, Altin N, Squier-Pernet K, Salignac S, Blanc P, Munnich A, et al. A non-coding variant in the Kozak sequence of RARS2 strongly decreases protein levels and causes pontocerebellar hypoplasia. *BMC Med Genom.* 2023;16:143.
53. Curinha A, Oliveira Braz S, Pereira-Castro I, Cruz A, Moreira A. Implications of polyadenylation in health and disease. *Nucleus.* 2014;5:508–19.
54. Johnston JJ, Williamson KA, Chou CM, Sapp JC, Ansari M, Chapman HM, et al. NAA10 polyadenylation signal variants cause syndromic microphthalmia. *J Med Genet.* 2019;56:444–52.
55. Gehring NH, Frede U, Neu-Yilik G, Hundsdoerfer P, Vetter B, Hentze MW, et al. Increased efficiency of mRNA 3' end formation: a new genetic mechanism contributing to hereditary thrombophilia. *Nat Genet.* 2001;28:389–92.
56. Morales J, Russell JE, Liebhäber SA. Destabilization of human α -globin mRNA by translation anti-termination is controlled during erythroid differentiation and is paralleled by phased shortening of the poly(A) tail *. *J Biol Chem.* 1997;272:6607–13.
57. O'Brien J, Hayder H, Zayed Y, Peng C. Overview of MicroRNA biogenesis, mechanisms of actions, and circulation. *Front Endocrinol.* 2018 [cited 2024 Feb 7];9. Available from: <https://www.frontiersin.org/journals/endocrinology/articles/10.3389/fendo.2018.00402>.
58. Jedlickova J, Vajter M, Barta T, Black GCM, Perveen R, Mares J, et al. MIR204 n.37C>T variant as a cause of chorioretinal dystrophy variably associated with iris coloboma, early-onset cataracts and congenital glaucoma. *Clin Genet.* 2023;104:418–26.
59. Verdura E, Hervé D, Bergametti F, Jacquet C, Morvan T, Prieto-Morin C, et al. Disruption of a miR-29 binding site leading to COL4A1 upregulation causes pontine autosomal dominant microangiopathy with leukoencephalopathy. *Ann Neurol.* 2016;80:741–53.
60. Beetz C, Schüle R, Deconinck T, Tran-Viet KN, Zhu H, Kremer BPH, et al. REEP1 mutation spectrum and genotype/phenotype correlation in hereditary spastic paraparesis type 31. *Brain.* 2008;131:1078–86.
61. Zhang X, Wakeling M, Ware J, Whiffin N. Annotating high-impact 5' untranslated region variants with the UTRannotator. *Bioinformatics.* 2021;37:1171–3.
62. Soukarieh O, Tillet E, Proust C, Dupont C, Jaspard-Vinassa B, Soubrier F, et al. uAUG creating variants in the 5'UTR of ENG causing Hereditary Hemorrhagic Telangiectasia. *NPJ Genom Med.* 2023;8:32.
63. Ruggieri A, Ramachandran N, Wang P, Haan E, Kneebone C, Manavis J, et al. Non-coding VMA21 deletions cause X-linked myopathy with excessive autophagy. *Neuromuscul Disord.* 2015;25:207–11.
64. Ravenscroft G, Miyatake S, Lehtokari VL, Todd EJ, Vornanen P, Yau KS, et al. Mutations in KLHL40 are a frequent cause of severe autosomal-recessive nemaline myopathy. *Am J Hum Genet.* 2013;93:6–18.
65. Dofash LNH, Monahan GV, Servián-Morilla E, Rivas E, Faiz F, Sullivan P, et al. A KLHL40 3' UTR splice-altering variant causes milder NEM8, an under-appreciated disease mechanism. *Hum Mol Genet.* 2023;32:1127–36.
66. Castel SE, Cervera A, Mohammadi P, Aguet F, Reverter F, Wolman A, et al. Modified penetrance of coding variants by cis-regulatory variation contributes to disease risk. *Nat Genet.* 2018;50:1327–34.
67. Gasperini M, Tome JM, Shendure J. Towards a comprehensive catalogue of validated and target-linked human enhancers. *Nat Rev Genet.* 2020;21:292–310.
68. Fang Z, Rajewsky N. The impact of miRNA target sites in coding sequences and in 3'UTRs. *PLoS ONE.* 2011;6:e18067.
69. Michel AM, Andreev DE, Baranov PV. Computational approach for calculating the probability of eukaryotic translation initiation from ribo-seq data that takes into account leaky scanning. *BMC Bioinform.* 2014;15:380.
70. Fowler DM, Adams DJ, Glynn AL, Hahn WC, Marks DS, Muffley LA, et al. An atlas of variant effects to understand the genome at nucleotide resolution. *Genome Biol.* 2023;24:147.

AUTHOR CONTRIBUTIONS

Work was led by NWieder with contributions from END, RD, AC and AMG. The project was conceived and supervised by NWhiffin. All authors read and approved the final manuscript.

FUNDING

NWhiffin is supported by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (220134/Z/20/Z). The research was supported by grant funding from the Rosetrees Trust (PGL19-2/10025) and the Wellcome Trust Core Award Grant Number 203141/Z/16/Z.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to Nicola Whiffin.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025