NAR Genomics and Bioinformatics, 2025, **7**, lqaf017
https://doi.org/10.1093/nargab/lqaf017
Application Notes

OXFORD

Enhancing the annotation of small ORF-altering variants using MORFEE: introducing MORFEEdb, a comprehensive catalog of SNVs affecting upstream ORFs in human 5′UTRs

Caroline Meguerditchian[1], David Baux[2,3,4], Thomas E. Ludwig[5,6], Emmanuelle Genin[5,6], David-Alexandre Trégouët[1,*], Omar Soukarieh[1,7,*]

[1]Université de Bordeaux, INSERM, Bordeaux Population Health Research Center, UMR 1219, F-3000 Bordeaux, France
[2]Molecular Genetics Laboratory, Université de Montpellier, CHU Montpellier, F-34000 Montpellier, France
[3]Institute for Neurosciences of Montpellier (INM), Université de Montpellier, Inserm, F-34000 Montpellier, France
[4]Montpellier BioInformatique pour le Diagnostic Clinique (MOBIDIC), CHU Montpellier, F-34000 Montpellier, France
[5]Université de Brest, Inserm, EFS, UMR 1078, GGB, F-29200 Brest, France
[6]CHRU Brest, F-29200 Brest, France
[7]Université de Bordeaux, INSERM, Biology of Cardiovascular Diseases, U1034, F-33600 Pessac, France

[*]To whom correspondence should be addressed. Tel: +33 5 57 89 01 06; Email: omar.soukarieh@inserm.fr
Correspondence may also be addressed to Alexandre Trégouët. Email: david-alexandre.tregouet@u-bordeaux.fr

Abstract

Non-canonical small open reading frames (sORFs) are among the main regulators of gene expression. The most studied of these are upstream ORFs (upORFs) located in the 5′-untranslated region (UTR) of coding genes. Internal ORFs (intORFs) in the coding sequence and downstream ORFs (dORFs) in the 3′UTR have received less attention. Different bioinformatics tools permit the prediction of single nucleotide variants (SNVs) altering upORFs, mainly those creating AUGs or deleting stop codons, but no tool predicts variants altering non-canonical translation initiation sites and those altering intORFs or dORFs. We propose an upgrade of our MORFEE bioinformatics tool to identify SNVs that may alter all types of sORFs in coding transcripts from a VCF file. Moreover, we generate an exhaustive catalog, named MORFEEdb, reporting all possible SNVs altering existing upORFs or creating new ones in human transcripts, and provide an R script for visualizing the results. MORFEEdb has been implemented in the public platform Mobidetails. Finally, the annotation of ClinVar variants with MORFEE reveals that > 45% of UTR-SNVs can alter upORFs or dORFs. In conclusion, MORFEE and MORFEEdb have the potential to improve the molecular diagnosis of rare human diseases and to facilitate the identification of functional variants from genome-wide association studies of complex traits.

Introduction

Non-canonical small open reading frames (sORFs) located in untranslated regions (UTRs) surrounding the coding sequence (CDS, main ORF) of mRNAs are the main regulators of translation and can themselves be translated [1, 2]. Upstream ORFs (upORFs) located in the 5′UTR have been the most studied sORFs for their role in the regulation of translation initiation. The number of identified disease-causing variants creating new upORFs or altering existing ones has increased in the last years [3, 4]. Different kinds of variants have been identified in a wide range of rare genetic diseases, most of which are associated with decreased protein levels, characterizing them as loss-of-function (LOF) mutations. These variants can (i) create a new upstream translation initiation site (uTIS) [3–5], (ii) delete upstream stop codons (uStops) [3], and/or (iii) create new upstream stop codons [6]. The presence of downstream ORFs (dORFs) in the 3′UTR of mRNA as regulators of CDS translation has also been described [7]. Even though less studied than variants affecting upORFs, variants creating or altering dORFs have also been associated with rare human diseases [8]. Different bioinformatics tools have been developed to annotate UTR variants from high-throughput sequencing (HTS) data, primarily aimed at aiding the molecular diagnosis of rare genetic diseases [3, 9–11]. These tools focus on the annotation of 5′UTR variants creating or disrupting upORFs [3, 9, 11] and/or 3′UTR variants altering different regulatory elements (i.e. polyadenylation sites) [10, 11]. Despite significant efforts to study non-coding variants in human diseases [12–14], no bioinformatics tools were specifically designed to annotate variants that can create or disrupt dORFs or those that could create near-cognate TISs. Near-cognate TISs, also referred to as non-canonical TISs, are those differing by one nucleotide from the AUG codon. However, different studies have demonstrated that translation can be initiated with near-cognate TISs, with CUG being the most efficient one (i.e. the closest translation efficiency to AUG) [2, 15, 16]. There is, moreover, increasing evidence for the implication of variants creating near-cognate TISs in human diseases [4, 17]. In addition to upORFs and dORFs, ribosome profiling analyses have identified small ORFs located within the coding sequence

Received: November 12, 2024. **Revised:** January 17, 2025. **Editorial Decision:** February 11, 2025. **Accepted:** February 17, 2025

[internal ORFs (intORFs)], but their role in the regulation of translation and their implications in diseases are still unknown [1].

Creation or disruption of non-canonical ORFs in rare diseases has been associated not only with rare constitutional variants but also with somatic and *de novo* variants [14, 18]. Moreover, several examples of common variants altering non-canonical ORFs and associated with complex traits have been reported [19–21]. Altogether, these observations illustrate the necessity of revealing DNA variants that could be able to create or to disrupt small ORFs along a given transcript and to characterize them in order to help in understanding gene regulation in rare and common diseases.

We here updated our bioinformatics tool MORFEE [22] to predict single nucleotide variants (SNVs) that could create or disrupt non-canonical ORFs along a given transcript. In its initial version, MORFEE was used to identify pathogenic uAUG-creating variants in the 5′UTR of *PROS1* and *ENG* in protein S deficiency and hereditary hemorrhagic telangiectasia (HHT), respectively [5, 23]. It is here extended to predict the creation of non-canonical TISs and of non-canonical ORFs in 5′UTRs but also in the CDS and the 3′UTR of human transcripts. Furthermore, we applied this updated version on all possible SNVs in 5′UTRs resulting from an *in silico* mutational saturation of all human transcripts to build the MORFEEdb database cataloging all SNVs that could alter upORFs. MORFEE was also deployed on a set of variants from the ClinVar database.

Materials and methods

MORFEE workflow

MORFEE is written in R and can run on all operating systems that have an R interpreter (including Linux, macOS, and Windows). First, MORFEE uses ANNOVAR [24] to annotate minimal VCF files (i.e. with at least the *chr*, *position*, *reference allele*, and *alternate allele* fields) containing SNVs in order to obtain the information needed about the variants (i.e. gene, transcript, or nomenclature). Then, the sequence of all transcripts is downloaded from GENCODE (GRCh37 or GRCh38) and variants are set on corresponding transcripts. At this step, MORFEE extracts all TISs and stop codons existing on transcripts in the absence or presence of the annotated variants and compares the information between both sequences to identify (i) newly created canonical and non-canonical TISs; (ii) newly created stop codons; and (iii) deleted stop codons, along a given transcript (Fig. 1A). It is worth noting that alterations of AUGs are also annotated by MORFEE, and are noted as "canonical_to_non_canonical_TIS". Finally, MORFEE provides specific characteristics such as sORF type, size, position on the transcript, and Kozak sequence {together with its strength and Kozak similarity score (KSS) [17]}, among others. MORFEE annotations, as detailed in the readme available on https://doi.org/10.5281/zenodo.14864790, are assembled in an excel file.

Annotation of all possible upORF-SNVs in the 5′UTR of human transcripts

In order to generate an exhaustive database containing all possible SNVs creating or altering upORFs (upORF-SNVs), we performed a mutational saturation of the 5′UTR of 55 304 transcripts from ∼17 000 coding genes (DOI 10.5281/zenodo.14604077). First, the most recent version of the Ensembl database (GRCh38.p14) was downloaded by using BioMart package version 2.50.0 to extract the identity of all transcripts (ENST). The sequence of all transcripts was then extracted from Gencode (release 43, GRCh38.p13), and 5′UTRs were defined based on the position of the main AUG. Finally, each position in the 5′UTR was *in silico* mutated with the three alternative nucleotides, and the generated VCF files for all transcripts were annotated with the MORFEE workflow described above.

Annotation of all ClinVar variants

All variants reported in the ClinVar database by September 2023 were downloaded. Variants were filtered to keep only SNVs in the 5′UTR, CDS, and the 3′UTR (i.e. in coding and non-coding exons; $n = 1\ 048\ 573$ SNVs; Supplementary Tables S1 and S2), which were annotated with MORFEE as described above.

Results and discussion

MORFEEdb, an exhaustive database for upORF-SNVs in the 5′UTR of coding genes

While most available tools mainly focus on upORFs initiated with a canonical TIS [9, 11], there is accumulating evidence that some non-canonical TISs (at least CUG, GUG, UUG, and ACG) can initiate translation [6, 25] and be associated with disease [4, 6]. We have upgraded the MORFEE tool, which was initially developed to annotate SNVs that create uAUG or delete uStop in the 5′UTR. We used MORFEE to generate the MORFEEdb database reporting all possible SNVs predicted to create upORFs (i.e. creation of canonical and non-canonical uTISs and of uStop) or to modify existing ones (i.e. uAUG and uStop deletion).

We *in silico* mutated the 5′UTR of 55 304 transcripts, among which 16 981 are considered as MANE Select in the Ensembl database GRCh38.p14 (Supplementary Table S3), from 17 281 coding genes (Supplementary Table S4). We annotated the resulting file containing 41 536 234 SNVs with MORFEE and identified 18 886 702 upORFs as the result of uTIS creations, uStop creations, and/or uStop deletions in the totality of transcripts (Supplementary Table S5). The identity of transcripts, the size of their 5′UTR, the number of variants from the mutational saturation, and MORFEE annotations are detailed in Supplementary Table S4. According to the results of MORFEE, 14 338 457 upORFs could result from the creation of a new uTIS, 2 788 881 from the creation of a new stop codon in the 5′UTR, and 4 798 264 from the deletion of a stop codon. This means that at least 2 503 387 upORFs are annotated on two or more isoforms and/or are generated from SNVs with multiple consequences (i.e. uTIS creation, uStop deletion, and/or new stop creation). While 1 472 632 of the uTIS-creating variants annotated with MORFEE could be at the origin of uAUGs in the annotated transcripts, 13 615 818 create non-canonical uTISs. The annotated variants constitute the MORFEEdb database, henceforth available on Zenodo at DOI 10.5281/zenodo.14604077 and via this link from the Mobidetails platform (https://mobidetails.iurc.montp.inserm.fr/MD/static/resources/morfeedb/morfee.20231213.txt.gz). MORFEEdb has been implemented in the public database Mobidetails and is

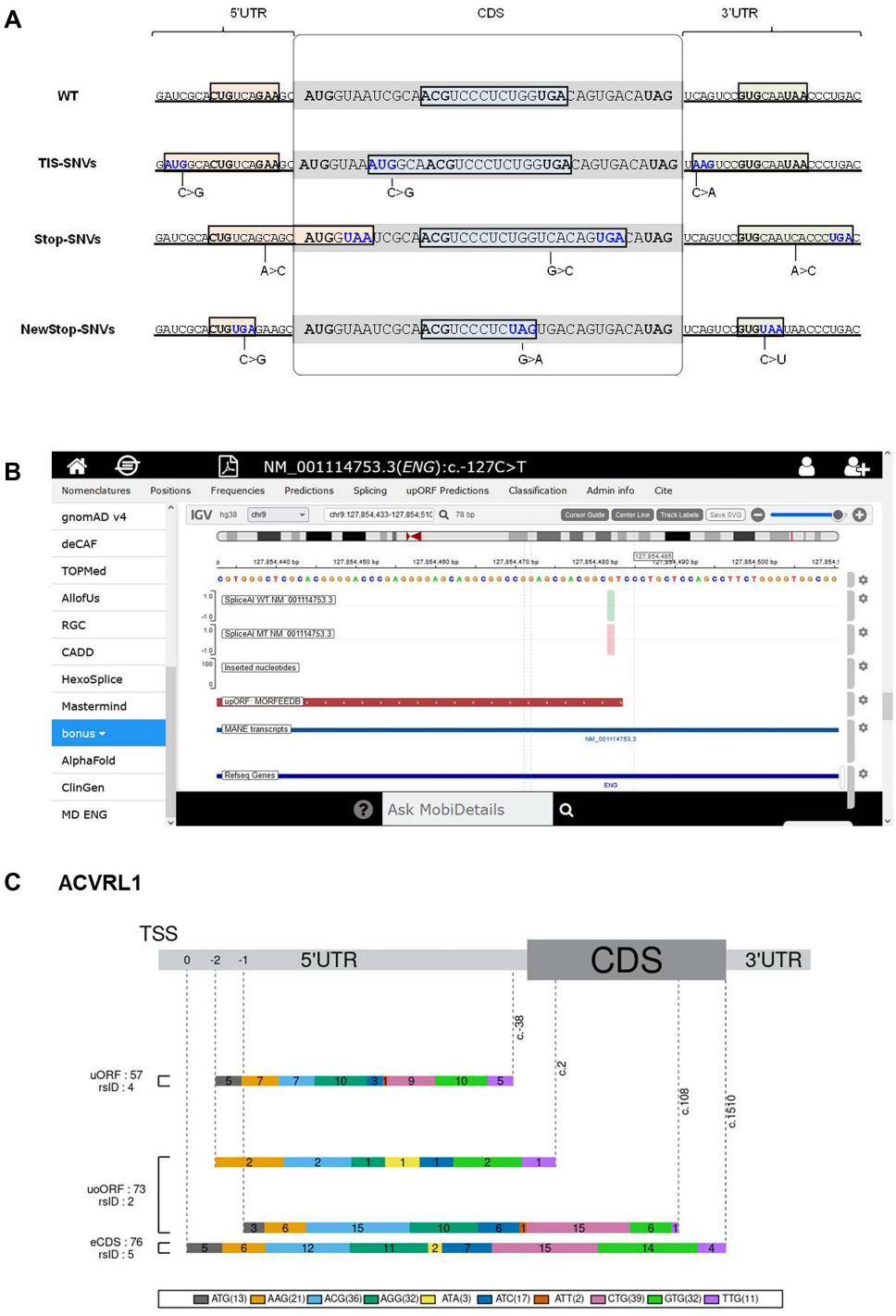

Figure 1. Description of the MORFEE tool and its associated database MORFEEdb. (**A**) Annotations of single nucleotide variants (SNVs) dedicated to small open reading frames (sORFs) along a given human transcript available in MORFEE tool. The first line corresponds to a wild-type (WT) transcript sequence containing sORFs in the 5′-untranslated region (UTR), the coding sequence (CDS), and the 3′UTR. MORFEE annotates SNVs that could create new translation initiation sites (TIS-SNVs), delete stop codons (Stop-SNVs), or create new stop codons (NewStop-SNVs). These types of variants can create new sORFs or modify existing ones. (**B**) MORFEEdb output as illustrated in the public database MobiDetails. The example of the c.-127C>T variant located in the 5′UTR of *ENG* is shown. This variant is predicted to create an upstream AUG resulting in an overlapping upstream ORF (upORF) on the MANE transcript as shown on the red bar. A table containing detailed information about the created uAUG and upORF can also be found on MobiDetails. (**C**) Illustration of upORFs resulting from all possible uTIS-SNVs on a given transcript extracted from MORFEEdb. A specific script allowing this graphic presentation to be obtained for all human transcripts is available on https://doi.org/10.5281/zenodo.14864790. The shown example corresponds to the MANE transcript (ENST00000388922.9/NM_000020.3) of the *ACVRL1* gene. The total number of annotated upORFs and those resulting from variants reported in the dbSNP database (https://www.ncbi.nlm.nih.gov/snp/) are indicated. The frame and position of stop codons associated with the different upORFs are shown. TSS, transcription start site; uORF, fully upstream ORF; uoORF, upstream overlapping ORF; eCDS, elongated CDS.

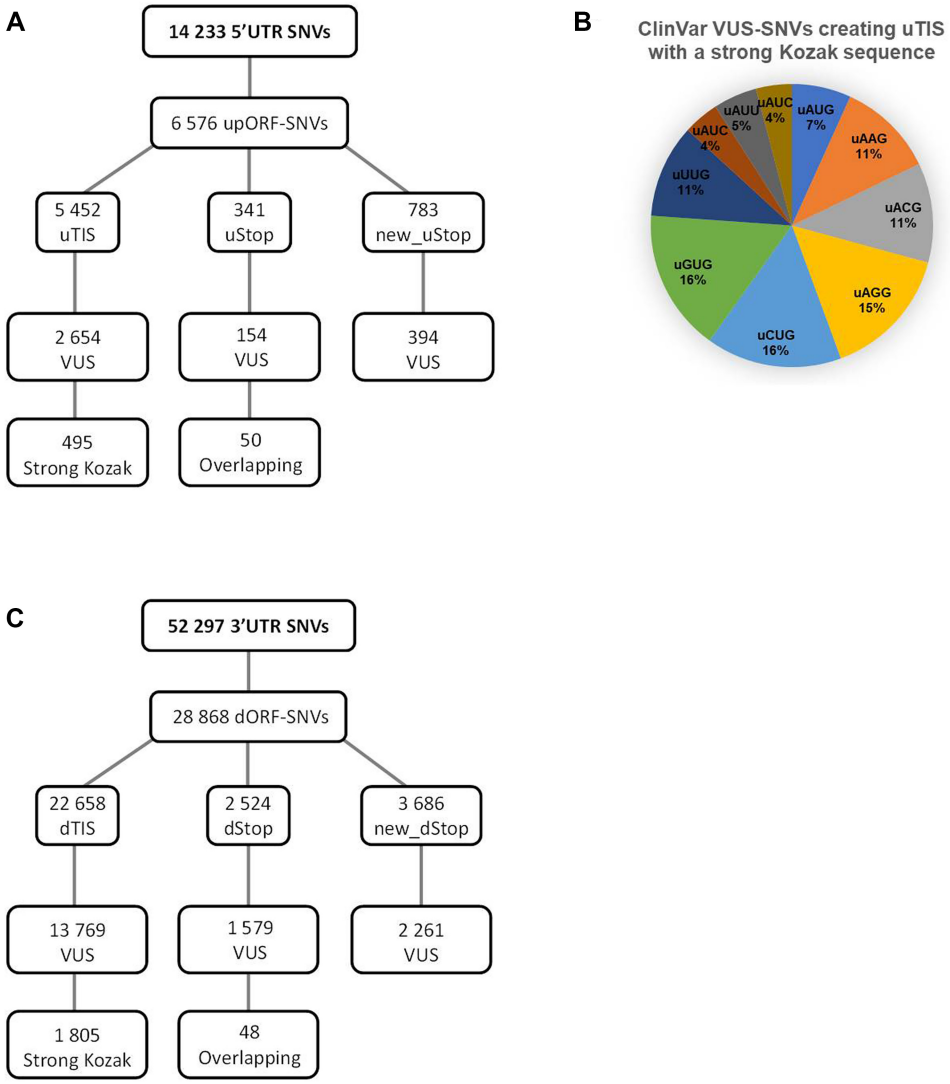

Figure 2. Summary of ClinVar UTR annotations with MORFEE. (**A**) Single nucleotide variants (SNVs) annotated with MORFEE to create or alter upstream ORFs (upORFs-SNVs) among 5′UTR variants reported in ClinVar by September 2023 ($n = 14\,233$). Variants creating new upstream translation initiation sites (uTISs), new stop codons (New_uStop), or deleting stop codons (uStop) are indicated. Variants of unknown significance (VUS) were then filtered among each category, as indicated. Strong Kozak corresponds to ClinVar SNVs classified as VUS and creating uTISs surrounded with a strong Kozak sequence and a KSS score ≥ 0.6. Overlapping corresponds to ClinVar SNVs classified as VUS and transforming fully upstream ORFs into overlapping ones by deleting stop codons. (**B**) Filtered 5′UTR SNVs classified as VUS and predicted to create canonical and non-canonical upstream translation initiation sites (uTISs) with a strong Kozak sequence and a KSS score ≥ 0.6. (**C**) Single nucleotide variants (SNVs) annotated with MORFEE to create or alter downstream ORFs (dORF-SNVs) among 3′UTR variants reported in ClinVar by September 2023 ($n = 52\,297$). A similar analysis to that of 5′UTR variants described in (A) was performed. dTIS, variants creating downstream translation initiation sites; dStop, variants deleting downstream stop codons; new_dStop, variants creating new stop codons. Overlapping corresponds to ClinVar 3′UTR SNVs deleting stop codons and associated with overlapping dORFs.

represented schematically under the IGV panel and in a dedicated table with detailed information about the associated upORF(s) (https://mobidetails.iurc.montp.inserm.fr/MD/, Fig. 1B). Some characteristics of the corresponding upORF(s), including the position and sequence of the uTIS and uStop, the nature of the upORF (i.e. overlapping or not overlapping), and the strength of the Kozak sequence surrounding the uTIS, are provided. Although these parameters are not exhaustive, they should help to predict the potential functional effect of upORF-SNVs. For example, 5′UTR SNVs creating overlapping upORFs are the most frequent compared with those creating other types of upORFs [3–5, 23]. Also, uTISs surrounded with strong Kozak sequences are more likely to be translated and then alter the translation of the main ORF [4, 5, 26]. While Whiffin *et al.* [3] have already annotated all possible uAUG-creating variants and uStop-deleting variants in human canonical gene transcripts, our work is the first to annotate all types of upORF-altering SNVs including those creating non-canonical uTISs. Moreover, a graphic presentation of all upORFs resulting from uTIS SNVs can be generated by using a specific script available on https://doi.org/10.5281/zenodo.14864790. Figure 1C shows the example of upORFs in the 5′UTR of *ACVRL1*.

By making MORFEEdb available, with a visual and facilitated presentation on the popular Mobidetails platform, we

contribute to improving the identification of upORF-SNVs in any gene of interest associated with human diseases.

A high proportion of ClinVar SNVs can alter or create non-canonical ORFs

Despite increasing success [14, 18, 27], the identification and characterization of non-coding variants causing rare genetic diseases or responsible for statistical associations observed in genome-wide genotyping/sequencing studies of complex traits remains challenging. Recent advances in ribosome profiling techniques and functional assays shed light on the role of disease-associated sORFs in translation and/or RNA stability [2, 18]. The improved version of MORFEE now annotates all types of sORF-altering SNVs along a given coding transcript and is the first tool annotating dORFs and intORF-SNVs. This aims to contribute to resolving molecular challenges in genetic diseases and to facilitating the identification of variants responsible for genetic association signals. This updated version of MORFEE has already enabled us to identify disease-causing uTIS-creating variants in the 5′UTR and 3′UTR of *TP53* (Li–Fraumeni syndrome) [28], in the 5′UTR *of CDKN1C* with isoform-dependent effects (Beckwith–Wiedemann syndrome) [29], and in the 5′UTR of *ENG* (HHT) [4]. We here used MORFEE to annotate all possible sORF-SNVs among ClinVar variants.

Overall, 1 048 573 SNVs were reported in ClinVar by September 2023 in CDSs and UTRs of human transcripts (Supplementary Tables S1 and S2). MORFEE annotated 6 576 upORF-SNVs in the 5′UTR, among which 5 452 create new uTISs, 341 delete upstream stop codons, and 783 create new stop codons (Supplementary Table S6). Around 50% (3 202/6 576) of these variants are of unknown significance (VUS), reflecting the lack of information about this kind of variant in human diseases (Fig. 2A). For instance, 15% (n = 495 variants, Supplementary Table S7) of the identified VUS among upORF-SNVs create uTISs encompassed by a strong Kozak sequence (i.e. when it contains a purine at position −3 and a guanine at position +4) and with high Kozak scores (KSS score from TIS predictor $\geq$ 0.6). Only 38 of the created uTISs correspond to AUG and 457 correspond to non-canonical uTISs, including 87 CUGs (Fig. 2B). In addition, 50 VUS are predicted to delete existing stop codons and are at the origin of overlapping upORFs. Of note, while the majority of pathogenic/likely pathogenic upORF-SNVs in ClinVar create uAUGs, the second most frequent pathogenic/likely pathogenic upORF-SNVs are those creating uCUG.

In parallel, MORFEE identified 28 868 dORF-SNVs from ClinVar 3′UTR variants (Supplementary Table S8). These variants correspond to 22 658 dTISs, 2524 dStop SNVs, and 3686 new_dStop-SNVs among which 60% (n = 17 609 variants) are VUS with 1805 variants creating dTISs (277 dAUG and 1528 non-dAUGs) surrounded with strong Kozak sequences (Fig. 2C).

Finally, 995 770 intORFs have been annotated with MORFEE from coding ClinVar variants (Supplementary Table S9).

It is worth noting that some of the annotated variants are predicted to have multiple consequences (TIS creation, stop deletion, and/or stop creation) (Supplementary Table S10) or isoform-dependent consequences resulting from alternative 5′UTRs [3] (i.e. alternative transcription start sites or alternative non-coding exons) and/or alternative main AUG (Supplementary Table S11).

As a side note, it has been shown that variants that could directly alter Kozak sequences are also able to alter translation efficiency [26]. Furthermore, the presence of single nucleotide polymorphisms could also modify the access of ribosomes to the 5′UTR [30]. In consequence, it would be very useful to add new functions to MORFEE to annotate variants that could alter Kozak sequences of a given TIS on one hand, and the potential modification of the effect of an sORF-SNV by a nearby second sORF-SNV (i.e. haplotype effect) on the other hand. Finally, while MORFEE has been developed to annotate variants on existing transcripts reported in the Ensembl database, it needs to be updated to also cover newly identified unannotated transcripts (e.g. resulting from long-read RNA sequencing).

Conclusion

We here propose an exhaustive catalog of all possible upORF-SNVs in human 5′UTRs and an easy-to-use tool to guide experimental analysis contributing to determining the pathogenicity of disease-associated variants. This would improve the molecular diagnosis of rare human diseases and would also identify potential common variants in complex diseases.

Acknowledgements

Statistical analyses benefited from the CBiB computing center of the University of Bordeaux.

Author contributions: O.S. and D.-A.T. conceived the project. C.M. upgraded MORFEE and performed the mutational saturation and variant annotation with MORFEE. O.S. analyzed the data. D.B., T.E.L., and E.G. participated in the bioinformatics development of MORFEEdb and in its implementation in Mobidetails. O.S. and D.-A.T. drafted the manuscript that was further shared to co-authors who read/corrected/and approved the final manuscript.

Supplementary data

Supplementary data are available at NAR Genomics & Bioinformatics Online.

Conflict of interest

The authors declare that they have no known competing financial or non-financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

INSERM GOLD Cross-Cutting program (E.G., T.E.L., D.-A.T., and O.S); The French National Research Agency (ANR) ANR-23-CE17-0042-01 program as part of the ENDOMORF project (D.-A.T. and O.S.); and the France 2030 investment plan [Precision and Global Vascular Brain Health Institute funded as part of the IHU3 initiative, ANR-23-IAHU-0001].

Data availability

The used version of the MORFEE tool is available at https://doi.org/10.5281/zenodo.14864790.

SNVs resulting from the *in silico* mutational saturation of human 5′UTRs from 55 304 transcripts and the MORFEEdb database are available on Zenodo, under the following DOI 10.5281/zenodo.14604077.

SNVs reported in ClinVar and annotated with MORFEE to possibly create or alter internal ORFs in the coding sequence are available at 10.5281/zenodo.13872114.

References

1. Chothani SP, Adami E, Widjaja AA *et al.* A high-resolution map of human RNA translation. *Mol Cell* 2022;**82**:2885–99. https://doi.org/10.1016/j.molcel.2022.06.023
2. Rodriguez CM, Chun SY, Mills RE *et al.* Translation of upstream open reading frames in a model of neuronal differentiation. *BMC Genomics* 2019;**20**:391. https://doi.org/10.1186/s12864-019-5775-1
3. Whiffin N, Karczewski KJ, Zhang X *et al.* Characterising the loss-of-function impact of 5' untranslated region variants in 15,708 individuals. *Nat Commun* 2020;**11**:2523. https://doi.org/10.1038/s41467-019-10717-9
4. Munsch G, Proust C, Deiber C *et al.* Unveiling Endoglin non canonical regulation: spotlight on the new role of the uPAR pathway. medRxiv, https://doi.org/10.1101/2024.01.28.24301864, 26 July 2024, preprint: not peer reviewed.
5. Soukarieh O, Tillet E, Proust C *et al.* uAUG creating variants in the 5'UTR of ENG causing hereditary hemorrhagic telangiectasia. *NPJ Genom Med* 2023;**8**:32. https://doi.org/10.1038/s41525-023-00378-5
6. Lee DSM, Park J, Kromer A *et al.* Disrupting upstream translation in mRNAs is associated with human disease. *Nat Commun* 2021;**12**:1515. https://doi.org/10.1038/s41467-021-21812-1
7. Wu Q, Wright M, Gogol MM *et al.* Translation of small downstream ORFs enhances translation of canonical main open reading frames. *EMBO J* 2020;**39**:e104763. https://doi.org/10.15252/embj.2020104763
8. Steri M, Idda ML, Whalen MB *et al.* Genetic variants in mRNA untranslated regions. *WIREs RNA* 2018;**9**:e1474. https://doi.org/10.1002/wrna.1474
9. Zhang X, Wakeling M, Ware J *et al.* Annotating high-impact 5'untranslated region variants with the UTRannotator. *Bioinformatics* 2021;**37**:1171–3. https://doi.org/10.1093/bioinformatics/btaa783
10. Romo L, Findlay SD, Burge CB. Regulatory features aid interpretation of 3'UTR variants. *Am Hum Genet* 2024;**111**:350–63. https://doi.org/10.1016/j.ajhg.2023.12.017
11. Liu Y, Dougherty JD. utr.Annotation: a tool for annotating genomic variants that could influence post-transcriptional regulation. *Bioinformatics* 2021;**37**:3926–8. https://doi.org/10.1093/bioinformatics/btab635
12. Martin-Geary AC, Blakes AJM, Dawes R *et al.* Systematic identification of disease-causing promoter and untranslated region variants in 8,040 undiagnosed individuals with rare disease. medRxiv, https://doi.org/10.1101/2023.09.12.23295416, 2 July 2017, preprint: not peer reviewed.
13. Lim Y, Arora S, Schuster SL *et al.* Multiplexed functional genomic analysis of 5' untranslated region mutations across the spectrum of prostate cancer. *Nat Commun* 2021;**12**:4217. https://doi.org/10.1038/s41467-021-24445-6
14. Plassmeyer SP, Florian CP, Kasper MJ *et al.* A massively parallel screen of 5'UTR mutations identifies variants impacting translation and protein production in neurodevelopmental disorder genes. medRxiv, https://doi.org/10.1101/2023.11.02.23297961, 3 November 2023, preprint: not peer reviewed.
15. Kearse MG, Goldman DH, Choi J *et al.* Ribosome queuing enables non-AUG translation to be resistant to multiple protein synthesis inhibitors. *Genes Dev* 2019;**33**:871–85. https://doi.org/10.1101/gad.324715.119
16. Dever TE, Ivanov IP, Hinnebusch AG. Translational regulation by uORFs and start codon selection stringency. *Genes Dev* 2023;**37**:474–89. https://doi.org/10.1101/gad.350752.123
17. Gleason AC, Ghadge G, Chen J *et al.* Machine learning predicts translation initiation sites in neurologic diseases with nucleotide repeat expansions. *PLoS One* 2022;**17**:e0256411. https://doi.org/10.1371/journal.pone.0256411
18. Schuster SL, Arora S, Wladyka CL *et al.* Multi-level functional genomics reveals molecular and cellular oncogenicity of patient-based 3' untranslated region mutations. *Cell Rep* 2023;**42**:112840. https://doi.org/10.1016/j.celrep.2023.112840
19. Soukarieh O, Meguerditchian C, Proust C *et al.* Common and rare 5'UTR variants altering upstream open reading frames in cardiovascular genomics. *Front Cardiovasc Med* 2022;**9**:841032. https://doi.org/10.3389/fcvm.2022.841032
20. Somers J, Wilson LA, Kilday J-P *et al.* A common polymorphism in the 5' UTR of ERCC5 creates an upstream ORF that confers resistance to platinum-based chemotherapy. *Genes Dev* 2015;**29**:1891–6. https://doi.org/10.1101/gad.261867.115
21. Dasgupta A, Prensner JR. Upstream open reading frames: new players in the landscape of cancer gene regulation. *NAR Cancer* 2024;**6**:zcae023. https://doi.org/10.1093/narcan/zcae023
22. Aïssi D, Soukarieh O, Proust C *et al.* MORFEE: a new tool for detecting and annotating single nucleotide variants creating premature ATG codons from VCF files. bioRxiv, https://doi.org/10.1101/2020.03.29.012054, 30 March 2020, preprint: not peer reviewed.
23. Labrouche-Colomer S, Soukarieh O, Proust C *et al.* A novel rare c.-39C>T mutation in the PROS1 5'UTR causing PS deficiency by creating a new upstream translation initiation codon. *Clin Sci (Lond)* 2020;**134**:1181–90. https://doi.org/10.1042/CS20200403
24. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010;**38**:e164. https://doi.org/10.1093/nar/gkq603
25. Loughran G, Zhdanov AV, Mikhaylova MS *et al.* Unusually efficient CUG initiation of an overlapping reading frame in POLG mRNA yields novel protein POLGARF. *Proc Natl Acad Sci USA* 2020;**117**:24936–46. https://doi.org/10.1073/pnas.2001433117
26. Pisarev AV, Kolupaeva VG, Pisareva VP *et al.* Specific functional interactions of nucleotides at key –3 and +4 positions flanking the initiation codon with components of the mammalian 48S translation initiation complex. *Genes Dev* 2006;**20**:624–36. https://doi.org/10.1101/gad.1397906
27. Trotman J, Armstrong R, Firth H *et al.* The NHS England 100,000 Genomes Project: feasibility and utility of centralised genome sequencing for children with cancer. *Br J Cancer* 2022;**127**:137–44. https://doi.org/10.1038/s41416-022-01788-5
28. Abstracts from the 57th European Society of Human Genetics (ESHG) Conference: e-posters. *Eur J Hum Genet* 2024;**32**:904–1230. EP01.095. https://doi.org/10.1038/s41431-024-01733-5
29. Abstracts from the 57th European Society of Human Genetics (ESHG) Conference: e-posters. *Eur J Hum Genet* 2024;**32**:904–1230. EP20.019. https://doi.org/10.1038/s41431-024-01733-5
30. Niederer RO, Rojas-Duran MF, Zinshteyn B *et al.* Direct analysis of ribosome targeting illuminates thousand-fold regulation of translation initiation. *Cell Syst* 2022;**13**:256–64. https://doi.org/10.1016/j.cels.2021.12.002

Received: November 12, 2024. **Revised:** January 17, 2025. **Editorial Decision:** February 11, 2025. **Accepted:** February 17, 2025