

RESEARCH ARTICLE

Utargetome: A targetome prediction tool for modified U1-snRNAs to identify distal-target positions with improved selectivity

Paolo Pignini¹, Federico Manuel Giorgi², Keng Boon Wee^{1*}

¹ Institute of Molecular and Cell Biology (IMCB), Agency for Science, Technology and Research (A* STAR), Republic of Singapore, ² Department of Pharmacy and Biotechnology, University of Bologna, Bologna, Italy

* weekb@a-star.edu.sg



Abstract

The endogenous U1 small nuclear RNA (U1-snRNA) plays a crucial role in splicing initiation through base-pairing to donor splice sites (5'-SSs). Likewise, modified U1s that carry a mutation-adapted 5'-terminal sequence have been demonstrated to rescue exon splicing when this is disrupted by genetic mutations within the 5'-SS. Given the base-pairing flexibility of the endogenous U1, the selectivity of modified U1s requires investigation. We developed a computational pipeline (Utargetome) that considers combinations of mismatches and alternative annealing registers to predict the transcriptome-wide binding sites (or targetome) of a U1. The pipeline accuracy was tested by recapitulating well-established alternative annealing registers and specificity for 5'-SSs in the predicted targetome of the human endogenous U1. It was then applied to analyse the targetome of 54 modified U1s that have been demonstrated to restore exon inclusion when affected by 5'-SS pathogenic mutations. While the targetome size was found to be wide-ranging, the off-target load appeared to be reduced for U1s targeting distal sites from the canonical U1-binding position. This feature was predicted also for a large set of 30,204 newly designed U1s targeting 839 5'-SS pathogenic mutations that were expected to affect exon inclusion. Targetome analysis indeed revealed an optimal distal-targeting position at 3 nucleotides downstream from the canonical 5'-SS, for which a modified U1 is likely to have minimal off-targets at 5'-SSs and acceptor splice sites (3'-SSs). Based on these insights, we propose to implement targetome prediction in the design and optimization of therapeutic U1s with improved selectivity.

OPEN ACCESS

Citation: Pignini P, Giorgi FM, Wee KB (2025) Utargetome: A targetome prediction tool for modified U1-snRNAs to identify distal-target positions with improved selectivity. *PLoS Comput Biol* 21(9): e1013534. <https://doi.org/10.1371/journal.pcbi.1013534>

Editor: Bishoy Kamel, University of New Mexico, UNITED STATES OF AMERICA

Received: November 19, 2024

Accepted: September 17, 2025

Published: September 23, 2025

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pcbi.1013534>

Copyright: © 2025 Pignini et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution,

Author summary

In the context of evolving gene therapy technologies that demand higher precision and safety, we present Utargetome, a computational tool designed to

and reproduction in any medium, provided the original author and source are credited.

Data availability statement: All relevant data are within the manuscript and its [Supporting information](#) files. Source codes for Utargetome are available at available at <https://github.com/ppigini/utargetome>.

Funding: The study was funded by the Agency for Science, Technology & Research, A*STAR (Industry Alignment Fund - Pre-positioning Program H20H6a0027 to KBW) and the Italian Ministry of University and Research (programs PON "Ricerca e Innovazione" 2014–2020, PRIN project 2022CEHEX8, and PNRR program for HPC, Big Data, and Quantum Computing to FMG). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

predict the binding sites of modified U1-snRNAs across the transcriptome. U1 plays a key role in splicing by binding to specific sites on RNA, and modified U1s have been used to restore normal splicing in cases where splice sites are lost due to mutations. However, ensuring the selectivity of these modified U1s, particularly avoiding unintended off-target effects, is critical for their therapeutic application. Utargetome predicts both the intended (on-target) and unintended (off-target) binding sites of U1s, accounting for mismatches and alternative binding registers. Our findings from analysing more than 30,000 modified U1s show that U1s targeting positions slightly downstream of their typical binding site have fewer off-target events. The insight enables the design of precise U1-based therapies for genetic disorders caused by splicing defects, and towards the advancement of safer gene therapies.

Introduction

U1-snRNA is an evolutionarily conserved non-coding RNA. Encoded by the *RNU1-1* gene in humans, it is 164 nucleotide (nt)-long, and contains an Sm motif and four stem-loop structures (SLI, SLII, SLIII and SLIV), which interact with multiple proteins to form small nuclear ribonucleoproteins (snRNPs). U1 snRNP initiates exon splicing through base-pairing between 11 nucleotides at its 5' end (5'-AUACUUACCUG-3'), here called binding sequence, and 5'-SSs [1,2], followed by the recruitment of other snRNPs for the assembly of the spliceosome. Besides splicing, U1 prevents premature transcription cleavage and polyadenylation through binding along the nascent transcript and inhibiting nearby polyadenylation signals, a process known as "tele-scripting" [1,3]. In addition, a potential role in transcription initiation and directionality was proposed, although the exact mechanism is unclear [1]. The various functions of U1 may account for it being one of the most expressed snRNAs in the cell, with a multitude of known paralogs (over 140 in humans) [1].

An estimated 15% of pathogenic mutations result in mRNA splicing defects [4]. Mutations that occur at 5'-SSs can disrupt endogenous U1 binding with the consequence of aberrant splicing, manifested as exon skipping or intron retention [4–6]. Adaptation of the 5'-terminal sequence of the endogenous U1 to a mutant 5'-SS as a therapeutic strategy for restoring splicing and rescuing wildtype expression has been shown with the use of exogenous U1-snRNAs [4]. The minuscule sequence length of engineered U1s as compared to gene replacement and gene editing approaches confers advantages in manufacturing and for vector delivery. Despite the numerous proof-of-principles in which efficacy of engineered U1s was observed in disease models, including spinal muscular atrophy, cystic fibrosis, hemophilia and neurofibromatosis [4], no U1 to our knowledge has progressed to human study.

Non-selectivity of engineered U1s may indeed be a critical limiting factor, especially given the tolerance for base-pairing mismatches and alternative annealing registers of the endogenous U1 [7–12], whose mRNA-binding sequence is fully complementary to only 0.85% of all 5'-SS sequences [1]. Potential off-target effects

include: 1) promoting the inclusion of alternative or cryptic exons [13–16], which may function as “poison exons” [17] that affect transcript stability; 2) interfering with the activity of other splicing elements, especially in proximity of 3′-SSs [18–21]; 3) inhibiting normal transcript cleavage and polyadenylation at 3′ UTRs [22–24]. Based on a small number of transcriptome-wide studies [25–27], the preliminary conclusion is that engineered U1 off-targets are present, albeit limited. In spinal cord tissues from mice expressing a modified U1, 12 among the 12,414 investigated genes were observed to be up- or down-regulated [25]. The magnitude of differentially expressed genes was similar in liver tissues from mice treated with a modified U1, with expression changes in 13 out of ~13,000 genes and splicing changes in less than 0.1% of transcripts [26]. In human HEK293 cells expressing a modified U1, only one differentially expressed gene and two alternative splicing events were observed [27]. By contrast, an *in silico* study predicted 1,827 perfect matches to the human transcriptome for a modified U1 [28].

As off-targets are dependent on a U1 binding sequence, this warrants an in-depth investigation of their relationship. Utargetome, available at <https://github.com/ppigini/utargetome>, is a new analysis tool to predict and characterize the transcriptome-wide targets (here referred to as targetome) of a given U1 by considering, besides perfect matches, targets that originate from Watson-Crick base-pairing mismatches and alternative annealing registers, which the endogenous U1 is known to tolerate [7–9]. The targetome was analysed for the total number of targets and their relative position to splice sites. To facilitate assessment between target counts and U1 binding capability, targets were progressively filtered by decreasing the number of minimum annealed bases (MABs), defined as the minimum number of canonical Watson-Crick base-pairings between U1 and target mRNA. Accordingly, 11 MABs is most selective as every base of the 11-nt U1 antisense sequence is paired to the target strand, which can include bulges but no mismatch. Whereas at 9 MABs for instance, there are 11, 10 or 9 canonical base-pairings between U1 and target strand. As a validation, the pipeline was first applied to reproduce the targetome of the human endogenous U1. Thereafter, the targetome of 54 published U1s that have been experimentally validated for their efficacy as therapeutic candidates were investigated and it was found that the sizes of their targetome span several orders of magnitude, indicating a wide range of selectivity. Lastly, 30,204 U1s were newly designed to target 839 5′-SS pathogenic mutations that were predicted to impair exon inclusion. Analyses of their targetome revealed a specific U1 targeting window in proximity of the canonical 5′-SS position that has minimal off-targets at 5′-SS and 3′-SS. This study underscores the need for and the advantage of integrating selectivity as a parameter in U1 engineering. The pipeline developed for U1 targetome prediction could therefore facilitate the discovery of U1 therapeutic candidates with improved selectivity.

Results

Survey of potential pathogenic targets for modified U1s

Genetic variants that can potentially be rescued by engineered U1s were surveyed. Unique pathogenic variants localized within the canonical U1 binding site (base positions from -3 to +8 from the exon-intron boundary, Fig 1A) were extracted from the ClinVar database (see “Methods”). Such variants are likely to result in proximal exon skipping, alternative or cryptic donor splice site usage, or adjacent intron retention. The list excluded variants located at base positions +1 or +2, since they would abrogate the entire splicing process [2] and are thus not likely rescuable by engineered U1s. Variants located at predicted mismatched positions between the transcript and the endogenous U1 (Fig 1A) were also excluded, for they are unlikely to diminish U1 binding. A total of 839 potential target mutations were identified in 763 distinct exons encoded by 500 different genes (Fig 1B and S1 Table). This highlights the broad applicability of engineered U1s as a therapeutic strategy, with only a handful of the identified pathogenic variants that have been previously addressed in U1 therapeutic studies (see below).

Pipeline for transcriptome-wide prediction of U1 targetome

The U1 targetome prediction was based on the Watson-Crick base-pairing rule between the U1 binding sequence and the transcriptome-wide RNAs. Besides targets with perfect complementarity (labelled as “COM”), the predicted targetome includes targets that form alternative annealing registers and base-pairing mismatches (suffixed with “+mm”, Fig 2,

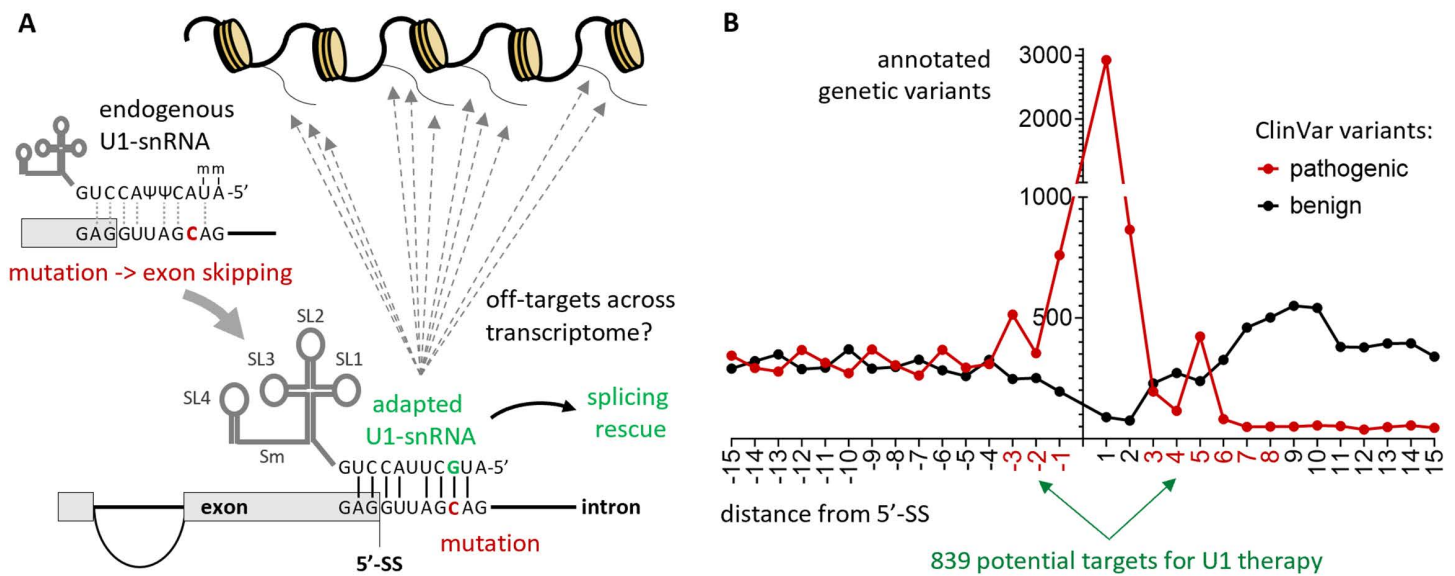


Fig 1. Design of modified U1s and landscape of 5'-SS mutations. (A) General design of a modified U1 tailored to a specific mutation. The binding sequence is engineered from the endogenous U1 in order to rescue a 5'-SS mutation. The endogenous U1 binding sequence is represented as 5'-mAmUACΨΨACCUG-3', where "m" represents 2'-O-methylation and "Ψ" is a pseudouridine. Methylation and pseudouridines are not represented in the modified U1, due to lack of prior knowledge. The canonical binding site is represented as 5'-GAGGUUAGCAG-3'. **(B)** Number of annotated "pathogenic" and "benign" mutations extracted from ClinVar and occurring in proximity of 5'-SSs.

<https://doi.org/10.1371/journal.pcbi.1013534.g001>

module A). Specifically, the annealing registers consist of single- or double-nucleotide bulges on the target RNA, on the U1 or on both strands (labelled as "BS1", "BS2", "BA1" and "BA2"), as well as asymmetric loops on both strands (labelled as "ALS" and "ALA"), which have all been shown to be utilized by the endogenous U1 [7–9]. Base-pairing mismatches are considered in the alternative annealing registers when they occur at least one nucleotide apart from a bulge or loop (Fig 2, module A, and S1 Fig).

The pipeline workflow involves the following sequential steps (Fig 2, see also "Methods"). Given a U1 sequence and the desired MAB, the targetome is built by 1) removing or inserting nucleotides in the target sequence to reproduce loops and bulges and 2) gradually inserting mismatched positions until the number of base-pairs reaches the input MAB (Fig 2, module A, and S1 Fig). Specifically, BS1 and BS2 (single- and double-nucleotide bulges on target strand) are simulated by inserting 1 and 2 nt respectively, BA1 and BA2 (single- and double-nucleotide bulges on the U1 strand) are simulated by deleting 1 and 2 nt respectively, ALS (asymmetric loops with the larger loop on the target strand) are simulated by deleting 1 nt and inserting a new combination of 2 nt, ALA (asymmetric loops with the larger loop on the U1 strand) are simulated by deleting 2 nt and inserting a new nt, and base-pairing mismatches are subsequently implemented as new combinations of nucleotides, with at least one nucleotide apart from the bulge(s)/loop (if present) to preserve their hypothetical structure. As the tolerance for G:U wobble pairing under a wide repertoire of modified U1 sequences is not known, the current implementation treats G:U as mismatches. Target sites with incidence of wobble pairs will thus have lower MABs than when wobble pairs are considered explicitly.

Next, each target sequence in the list is BLASTed against all genomic sequences that produce annotated transcripts, including introns (Fig 2, module B); mitochondrial or chloroplast (for *A. thaliana*, see below) genes are excluded, since U1 activity resides in the nucleus [2]. The hits obtained are subsequently analysed by the following criteria (Fig 2, module C): 1) hits on the same genomic position but mapped to different annotated transcripts are counted as distinct, since they would affect different RNA molecules; 2) hits produced by different annealing registers but mapped to the same position in

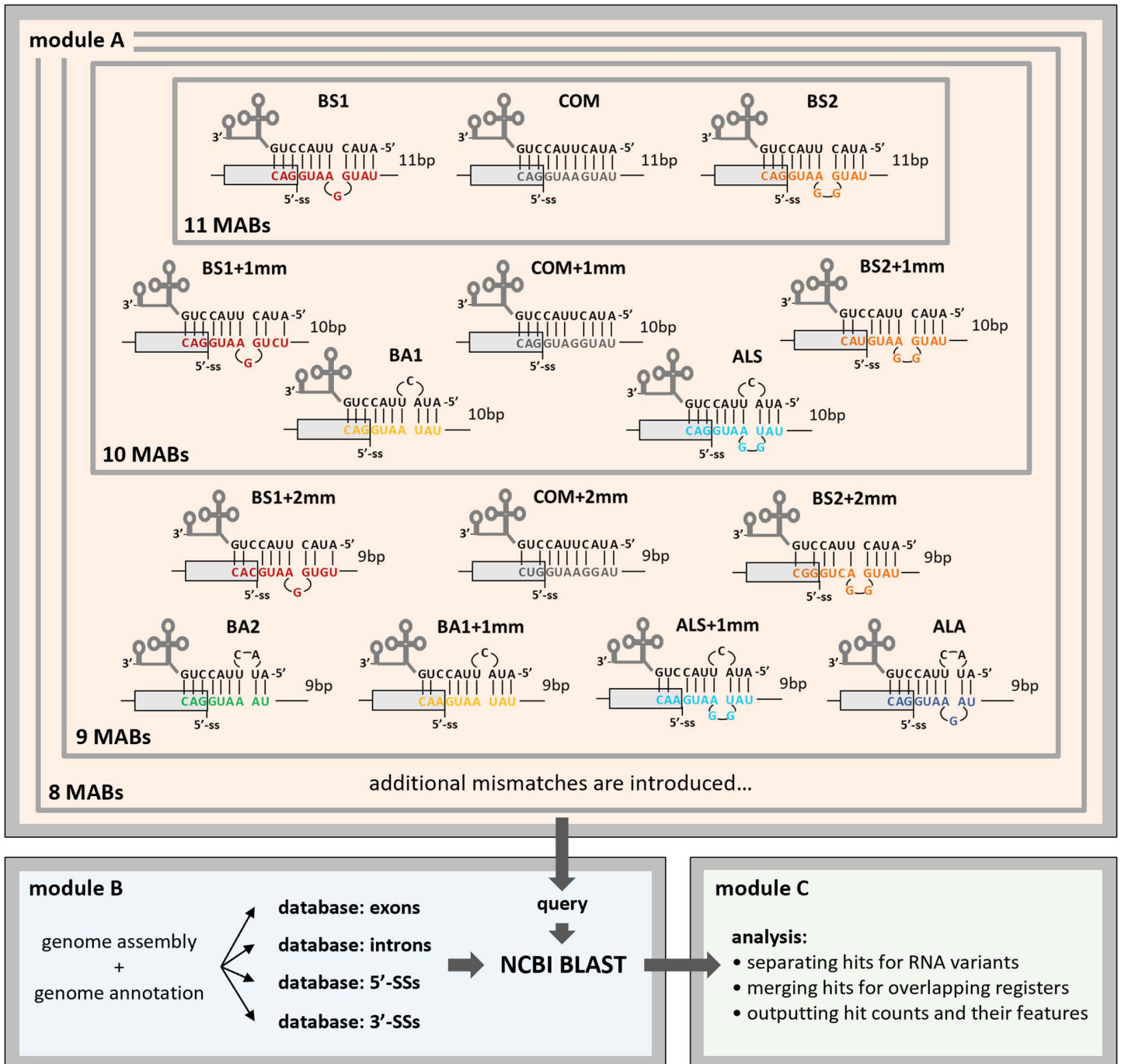


Fig 2. Pipeline for transcriptome-wide prediction of U1 targetome. (Module A) The potential targetome of the binding sequence (input sequence) of a given U1 was predicted through the consideration of six alternative annealing registers and the presence of mismatches in the input sequence or in the sequences with alternative registers. Annealing registers include: regular annealing with no alternative registers (“COM”), single- and double-nucleotide bulges on the target strand (“BS1” and “BS2”, respectively), single- and double-nucleotide bulges on the U1 strand (“BA1” and “BA2”, respectively), asymmetric loops with the larger loop on the target strand (“ALS”), asymmetric loops with the larger loop on the U1 strand (“ALA”), and mismatched positions (“mm”). The number of annealed base-pairs (indicated as bp) for each combination of annealing register and mismatches corresponds to the number of vertical lines connecting the annealed strands in module A; representative MABs of 11, 10 and 9 are depicted. **(Module B)** Every possible target sequence generated from the input sequence was BLASTed against a database of exons, introns, 5'-SSs or 3'-SSs. **(Module C)** BLASTed hits obtained in module B were processed as follows: 1) hits on the same genomic location but on different transcript variants were considered

as distinct; 2) hits on the same position of the same transcript variant produced from different registers were merged. The processed hits were classified and broken down by their annealing registers and by their relative position within exons, introns, 5'-SSs or 3'-SSs.

<https://doi.org/10.1371/journal.pcbi.1013534.g002>

a transcript (i.e., sharing the same 5'-most base position) are counted as a unique hit, since their effect on the transcript is expectedly identical. Finally, the hits are classified by the genomic annotation of their loci into four categories: hits mapped entirely in exons or introns are classified as “exonic” or “intronic targets” respectively; hits overlapping with 5'-SSs or 3'-SSs are classified as “5'-SS” or “3'-SS targets” respectively. The position of a predicted target in reference to a nearby 5'- or 3'-SS always refers to the position of the 5'-most nucleotide on the target sequence, i.e., the 3'-most nucleotide on the U1 antisense sequence, regardless of whether the base-pairing consists of a canonical Watson-Crick or a mismatched pairing (S1 Fig). The sum of all targets from these four categories constitutes the full size of the U1 targetome.

Targetome prediction of the endogenous U1

The pipeline was applied to obtain the targetome of the human endogenous U1 (*RNU1-1*) and for mapping and characterizing its transcriptome-wide binding sites as a means for method validation. The 5'-first 11 nucleotides of the *RNU1-1* transcript (5'-AUACUUACCUG-3') were used as input binding sequence. To recapitulate the flexibility of *RNU1-1* in base-pairing mismatches and alternative annealing registers, the targetome size was analysed at decreasing numbers of annealed bases, down to 6 MABs, for which every target in the targetome has at least six bases annealed to the U1 binding sequence. This ~55% minimum complementarity corresponds to the minimal 14–15 hydrogen bonds required for functional U1 binding [10–12,15]. For comparative controls, targetomes were obtained for the endogenous U1 of two evolutionarily distant eukaryotic species, the plant *Arabidopsis thaliana* and the amoeba *Dictyostelium discoideum*, both of which have identical binding sequence as *RNU1-1* [29]. As a negative control in each species, the complementary sequence of the input sequence, c(*RNU1-1*), was used.

Both the number of *RNU1-1* and c(*RNU1-1*) targets increases logarithmically by five orders of magnitude in each species when MABs decrease from 11 to 6 (Figs 3A, S2A and S3A). Fig 3B depicts the human *RNU1-1* targetome composition by both target locations and annealing registers as a function of MABs. While most targets are localized in introns, which are generally one order of magnitude longer than exons [30], 5'-SS targets disproportionately constitute the next largest fraction from 11 to 9 MABs, suggesting *RNU1-1* selectivity for these sites. The same trend was also observed in the respective targetome of *A. thaliana* and *D. discoideum* (S2B and S3B Figs); of note, the fraction of exonic targets is significantly larger than the fraction of intronic targets in both species, which is probably due to the higher representation of exon-coding regions in their genomes. Below 9 MABs in the three species, more targets localize to exons than at both splice sites combined, which is in concordance with the relative proportions of exonic and splice site positions in the genome. With regard to the annealing registers, BS1 and BS2 (single- and double-nucleotide bulges on the target strand, respectively) are the most common (Figs 3B, S2B and S3B, rectangular bars). As the variety of registers increases, they appear to be evenly distributed at the four target locations (S4 Fig). Importantly, the pipeline was able to identify six known 5'-SS targets that each interacts with the endogenous U1 via alternative registers [7,9] (S5 Fig and S2 Table).

Considering that canonical 5'-SS binding sites (located between position -3 and +8 from the exon-intron boundary) constitute the key functional targetome of the endogenous U1, targets located at this position were further analysed. Target counts at canonical 5'-SSs in *H. sapiens*, *A. thaliana* and *D. discoideum* follow a sigmoidal trend with decreasing MABs, and respectively cumulate to 96.7%, 97.6% and 96.5% of the corresponding total annotated 5'-SSs at 6 MABs (Figs 3C, S2C and S3C). By contrast, canonical 5'-SS target counts for c(*RNU1-1*) are not significant in each species. This further corroborates the minimum of 14–15 hydrogen bonds (approximately 6 MABs) required for the functional binding of the endogenous U1 [10–12,15]. In the remaining small percentage of annotated 5'-SSs that were not matched to the canonical 5'-SS targets in the targetome, many of the splice sites carry an alternative dinucleotide instead of the typical

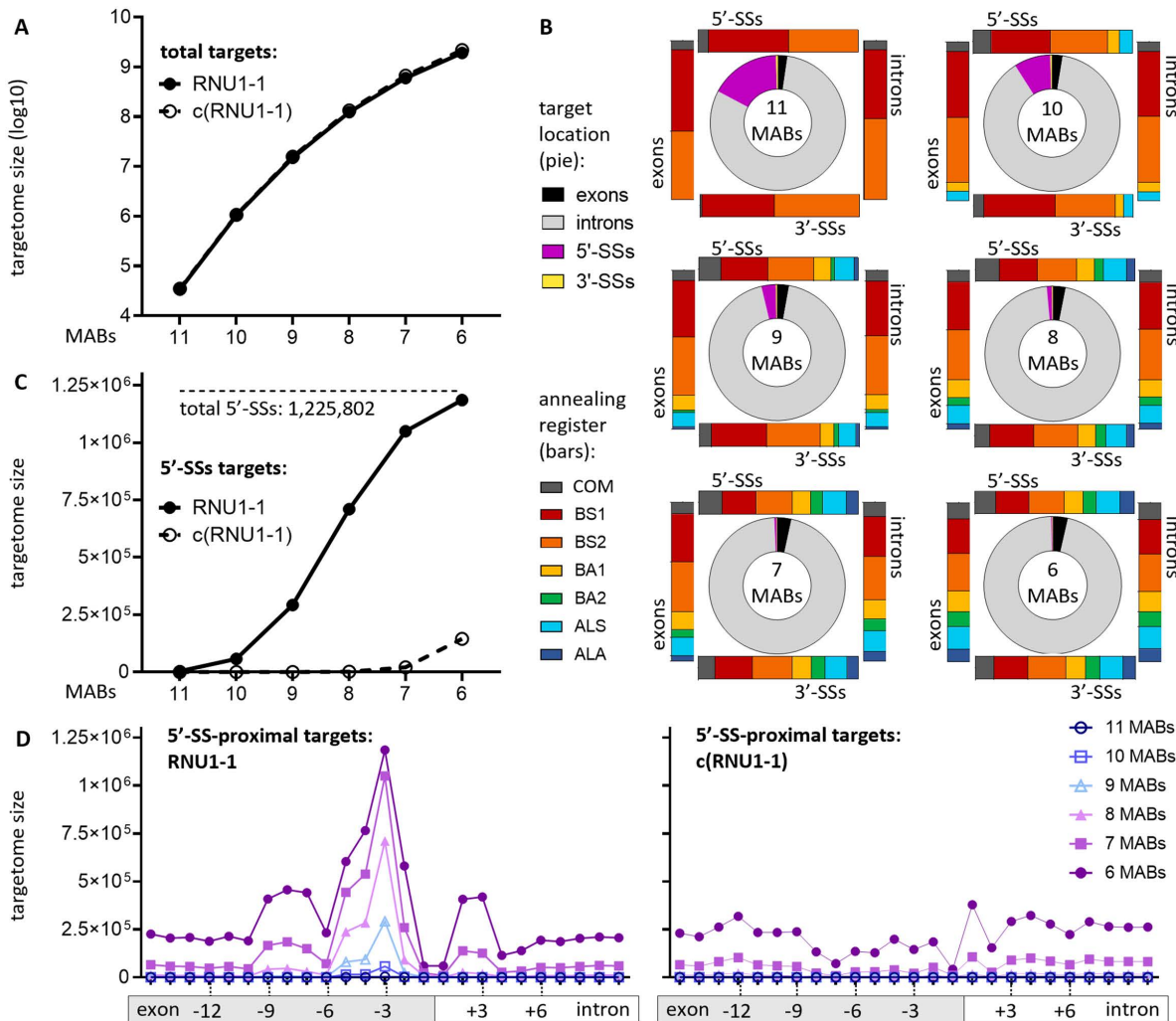


Fig 3. Targetome of the human endogenous U1. The 5'-terminal nucleotides of *RNU1-1* transcript (5'-AUACUUACCUG-3') were used as the input binding sequence for the pipeline. Targetome of its complementary sequence was analysed as control. MABs from 11 to 6 (corresponding to 100% and 55% complementarity respectively) were considered. **(A)** Targetome size for the endogenous U1, labelled as "RNU1-1", and the control, labelled as "c(RNU1-1)", with decreasing MABs. **(B)** Breakdown of the targetome composition of the endogenous U1 by target locations (pies) and annealing registers (rectangular bars). **(C)** Target counts overlapping canonical 5'-SSs (from position -3 to +8 from the exon-intron boundary) for the endogenous U1 and c(RNU1-1) with decreasing MABs. The dashed horizontal line indicates the total number of annotated 5'-SSs in all transcripts and variants as a reference. **(D)** Target distribution in the proximities of canonical 5'-SSs in the targetome of the endogenous U1 (left) and c(RNU1-1) (right) as a function of MABs. Sites are 1 nt apart, ranging from 15 nt up- to 10 nt down-stream of the exon-intron boundary. All positions are referenced to the 5'-most position of the target sequence.

<https://doi.org/10.1371/journal.pcbi.1013534.g003>

GT at positions +1 and +2 (S6 Fig), and thus possibly suggests an alternative splicing mechanism [31]. As this can also be attributed to the endogenous U1 mediating splicing at non-canonical positions [31,32], targets in the proximity of 5'-SSs, which lie within positions -15 to +10 from the splice site (referred to the 5'-most position of the target sequence from the exon-intron boundary), were interrogated. In addition to the highest enrichment at the canonical position as expected, 5'-SS targets of the endogenous U1, but not c(RNU1-1), were enriched consistently at specific distal positions across the three species (Figs 3D, S2D and S3D). Further analysis of the nucleotide compositions at these distal positions revealed a CAG motif among the three species (S7 Fig), which may be an evolutionary conserved binding motif for mediating splicing

at non-canonical sites. Similarly, targets in proximity of 3'-SSs were also found in all three species, with a significant enrichment between positions -3 and +8 from the intron-exon boundary (S8 Fig), which is consistent with previous findings indicating possible direct binding of the endogenous U1 to 3'-SSs [33]. In conclusion, the pipeline is able to capture the essential characteristics of the targetome of the endogenous U1, which shall be the benchmark for the analysis of the targetome of modified U1s.

Lastly, the Gibbs free energy (ΔG) of U1:target duplexes were evaluated for targets of the endogenous human U1 predicted at positions overlapping with 5'-SSs, 3'-SSs or exonic regions at MABs from 11 to 6 [7]. As expected, ΔG is inversely correlated with MABs (S9A Fig). ΔG of duplexes at 5'-SS are generally lower than at the other two positions (S9B Fig), which may suggest U1 preferential binding to 5'-SSs, a well-established fact.

Targetome analysis of modified U1s

The human targetome of 54 modified U1s was predicted with the pipeline described in the previous paragraph. Their efficacy in restoring exon inclusion, affected by 23 different 5'-SS pathogenic mutations, has been validated in 16 different studies as a potential therapeutic strategy for a variety of diseases (S3 Table). Each U1 carried a binding sequence of 11 bases with no dinucleotide TT, GA, or GG at the 5'-end, which negatively affect its stability [34] and could therefore influence its on- or off-target activity. In all the studies, the modified U1s were expressed from a plasmid vector containing the promoter, scaffold and terminator sequences of human *RNU1-1*. Selectivity of each U1 was inferred from both its full targetome size and number of targets at splice sites, which mediate the default mechanism of action of engineered U1s [13–16, 18–21]. The full targetome size across the U1s was found to span four orders of magnitude at perfect complementarity, from 15 to 15,817 targets, and from 2,314,404–54,757,018 targets at 9 MABs, which are dominated by intronic targets (Fig 4A and S3 Table). For targets overlapping 5'-SSs, 0–737 perfect hits or 5,202–507,381 hits at 9 MABs were predicted across the modified U1s (Fig 4B and S3 Table). By comparison, the *RNU1-1* targetome size lies approximately at the median of the modified U1s, and it has more 5'-SS targets than 51 of the modified U1s. The latter trend is consistent when targets at distal positions from the canonical 5'-SSs (from 10 nt up- to 25 nt downstream of the exon-intron boundary) [35–37] were included, since their contribution is not significant (S10 Fig). There are significantly fewer targets overlapping 3'-SSs than 5'-SSs, with 0–109 at perfect complementarity, and 2,365–71,121 at 9 MABs (Fig 4C and S3 Table).

The pattern of target counts among the 54 U1s is generally similar between perfect complementarity and 9 MABs (Fig 4A, 4B and 4C). The trend is still broadly conserved from 11 to 7 MABs among four representative U1s, each with a unique targetome characteristic (S11 Fig and S4 Table), namely U1-1 (smallest targetome), U1-36 (highest 5'-SS target count), U1-49 (highest 3'-SS target count), and U1-54 (largest targetome). Another observation is that while modified U1s with small targetome size usually have low target counts at both splice sites, this is not always the case. For example, U1-52 possesses the second largest targetome size but has one of the lowest number of 5'-SS targets, whereas U1-36 has the most 5'-SSs targets although its targetome size is the 19th largest (Fig 4A and 4B). Given the essential role of 5'-SSs in mediating U1 function, U1-52 is expected to be more selective than U1-36. Hence, targetome size is not a definite indicator of selectivity.

In order to show that predicted targets are biologically relevant, the 5'-SS targets of U1-36 were further interrogated (Fig 4D and S5 Table). As U1-36 was validated to rescue the skipping of *RHO* exon 4 when affected by c.936G>A mutation in retinopathies [38], RNAseq datasets originating from human retinas were analysed to identify the most probable off-target exons. These exons, besides containing a binding site for U1-36 at their 5'-SS, were selected for having an average percent-splice-in (PSI) of less than 0.5, i.e., they are spliced-out in more than 50% of the coding transcripts (see Methods). Given the relatively low inclusion levels of these exons, a significant increase in PSI induced by U1-36 is likely to have biological implications. A total of 39 candidate exons were identified from 737 perfectly complementary 5'-SS targets (Fig 4E and S6 Table) on which U1-36 can potentially increase their PSI and affect the transcript stability

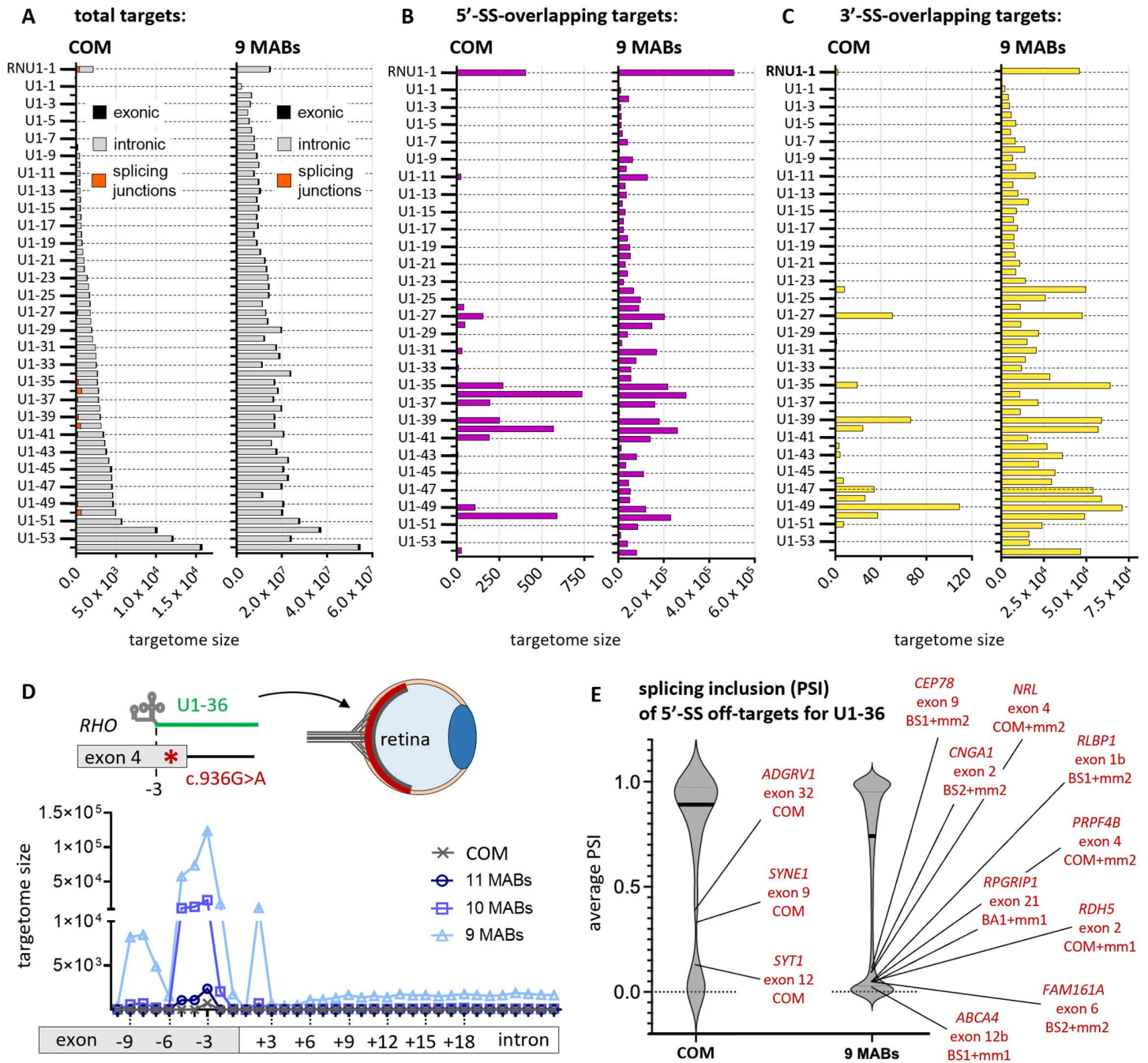


Fig 4. Targetome analysis of the 54 modified U1s targeting human transcripts. The targetome of every modified U1 was predicted by the pipeline for the register with perfect complementarity (COM) or for all registers with 9 MABs. (A) Targetome size of every modified U1 (refer to S3 Table for the description). Total target counts are broken down by their location within exons or introns or overlapping with splicing junctions. Target counts at (B) 5'-SSs and (C) 3'-SSs, which include all positions overlapping with the respective splice junction. (D) Positional distribution of 5'-SS target counts of U1-36 (S3 Table) from the canonical 5'-SS. Position intervals are 1 nt apart, ranging from 10 nt up- to 25 nt down-stream of the exon-intron junction (with reference to the 5'-most position of the target sequence). The analysis was performed for the register with perfect complementarity (COM) and for all annealing registers from 11 to 9 MABs. (E) Average percent-splice-in (PSI) for exons expressed in human retina whose 5'-SS was found in the targetome of U1-36. All targets, with perfect complementarity or 9 MABs, were located from 10 nt up- to 25 nt down-stream of the exon-intron junction. Average PSI values were calculated from eight different RNA-seq datasets derived from human eye retinas. The annealing register with mismatch (if any) for each targets are provided.

<https://doi.org/10.1371/journal.pcbi.1013534.g004>

or function. Three exons in particular, *ADGRV1* exon 32, *SYNE1* exon 9 and *SYT1* exon 12, are associated with important retina functions [39,40]; in comparison, RNU1-1 is predicted to bind to these exons at non-perfect complementarity with no more than 10 annealed bases (~90% complementarity). As anticipated, the number of candidate exons grows exponentially with decreasing complementarity, with 46,217 candidate exons found from 297,962 5'-SS targets at 9 MABs (Fig 4E and S7 Table). The sheer number of biologically relevant exons predicted from the targetome inevitably increases the probability of actual off-target events. These include exons from several genes with critical roles in retinal biology, such as *ABCA4* [41], *CEP78* [42], *CNGA1* [43], *FAM161A* [44], *NRL* [45], *PRPF4B* [46], *RDH5* [47], *RLBP1* [48], and *RPGRIP1* [49]. Among them, BS1 and BS2 are most common annealing registers which is similar to the endogenous U1 at 9 MABs (Fig 3B). Utargetome can thus be useful to direct experimental efforts in the evaluation of off-targets for a modified U1.

Lastly, it is important to note that modified U1s addressing the same genetic mutation but carrying different binding sequences have distinct targetomes (S12 Fig). An exemplary case is U1-30, which addresses the same mutation as U1-36, but which has one of the least 5'-SSs targets among the 54 U1s (Fig 4A and 4B). Moreover, total targetome size is not an accurate proxy for selectivity – for instance, U1-52, despite having one of the largest targetomes has a relatively small set of 5'-SS targets. In summary, the results suggested that modified U1s can be classified based on their selectivity, and is highly (binding) sequence-dependent, indicating that the design of the U1 binding sequence can be leveraged to minimise off-targets, as shown below.

Distal targeting strategy to mitigate U1 off-targeting

Considering that sequences are generally less conserved at distal positions than at canonical 5'-SSs, the well-established mechanism of U1 distal targeting [37] was investigated as a strategy to design U1s with reduced off-targets. This idea is supported by the targetome analysis of distal-targeting U1s amongst the 54 modified U1s (S8 Table). The first case study involves three U1s that rescue the effect of c.9726 +5G>A mutation in *F7* exon 9 [35,50]. Their respective target positions, with reference to the 5'-most position of the target sequence, are +17 (U1-8), -10 (U1-14) and -3 (U1-50). With both perfect complementarity and 9 MABs, distal-targeting U1-8 and U1-14 have substantially smaller targetome size and target counts at both 5'-SSs and 3'-SSs than the canonical-targeting U1-50 (Fig 5A and S8 Table). The second case study considered three distal-targeting U1s rescuing the effect of c.669A>T mutation in *F8* exon 6, with respective target positions at +1 (U1-29), +7 (U1-38) and +16 (U1-52) [51]. Although all three U1s showed no 5'-SS targets at perfect complementarity, U1-38 is most selective when considering both the targetome size and 5'-SS targets with 9 MABs (Fig 5A and S8 Table).

With the aim of testing the distal targeting strategy on additional mutations to identify possible optimal distal positions, *de-novo* U1s were designed for the 839 unique 5'-SS mutations that were identified previously in this study (Fig 1B and S1 Table). Specifically, 35 U1s were designed for every mutation by walking their 11 nt-long binding sequence, at one nucleotide resolution, from 10 nt upstream to 25 nt downstream of the exon-intron boundary (with reference to the 5'-most position of the target sequence, Fig 5B and S9 Table), which defines an optimal range for mediating exon rescue [35,36]. A U1 carrying the binding sequence of RNU1-1 but with a single-nucleotide adaptation to each mutation was also included, as it can be effective in some cases (e.g., U1-27, U1-40 and U1-48 in S3 Table). The targetome of each of the 30,204 newly-designed U1s was subsequently predicted by the pipeline at perfect complementarity. Fig 5C depicts the target counts at exons, introns, 5'-SSs and 3'-SSs for each U1 targeting each walking position (also S9 and S10 Tables). An optimal distal position was observed at 3 nt downstream from the canonical 5'-SS or equivalently, at position +1, given that both 5'-SS and 3'-SS targets are the lowest for U1s targeting at this position, showing both the lowest median and the lowest count for the U1 with the highest number of targets (S10 Table); p-values < 0.01 when counts at this position are compared to counts for U1s targeting most of the other considered positions in all target categories (S10 Table). This was further confirmed when all targets with 10 MABs were considered for position ± 1 and the two nearest positions (S13 Fig).

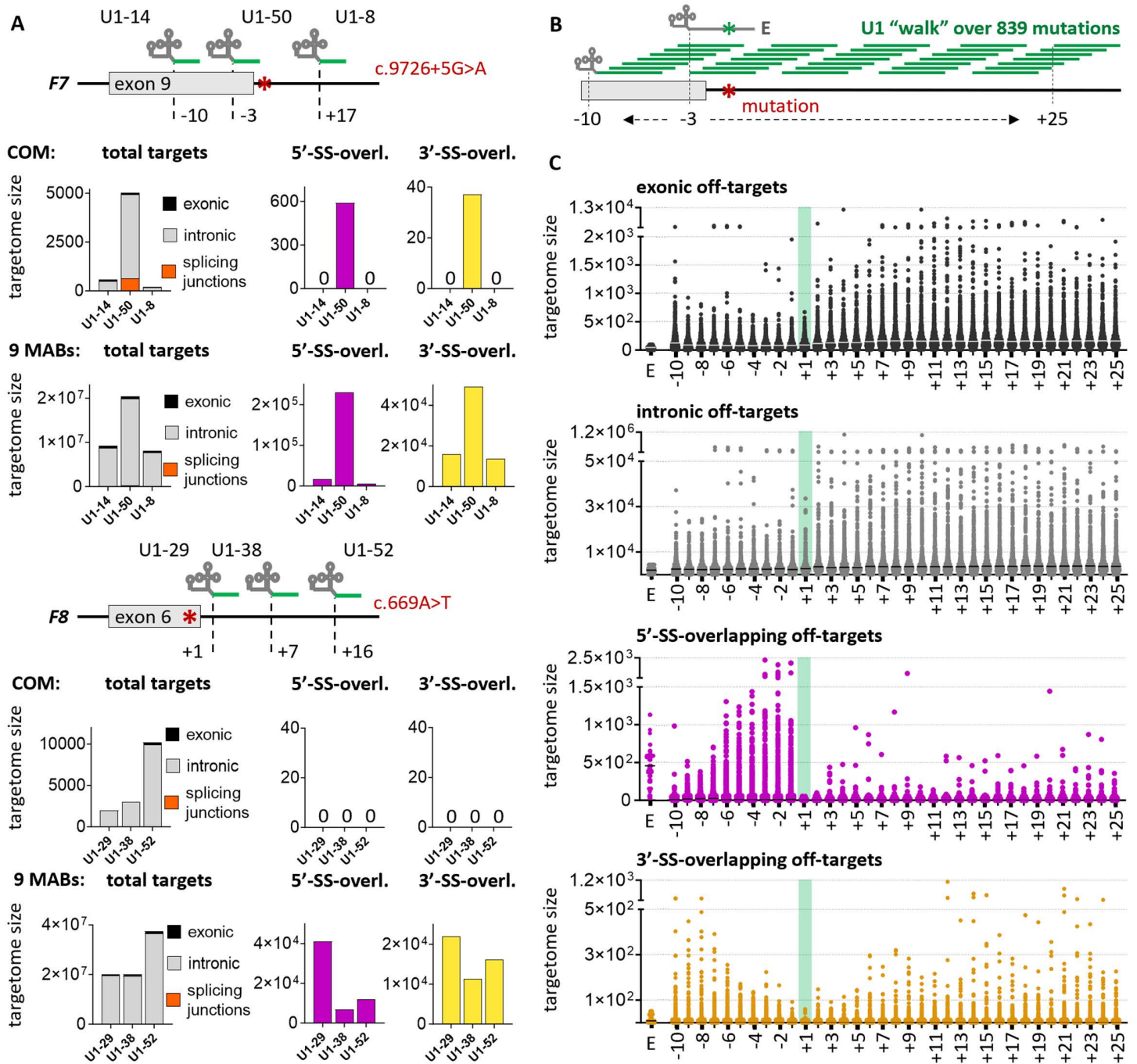


Fig 5. Targetome analysis of distal-targeting modified U1s. (A) Target counts of selected U1s that have previously been demonstrated to efficiently rescue exon skipping resulting from mutations *F7* c.9726+5G>A (top) or *F8* c.669A>T (bottom). Except for U1-50, they have been each designed to target distal positions from the canonical 5'-SS. Positions indicate the 5'-most nucleotide of the U1 target sequence, where the canonical site corresponds to "-3". The total targetome size (broken down by their location within exons or introns or overlapping with splicing junctions) and targets overlapping 5'-SSs, or 3'-SSs were predicted for perfectly complementary targets (COM) or with 9 MABs. (B) De-novo design of 30,204 U1s targeting 839 unique 5'-SS mutations from ClinVar database that were predicted to affect splicing. The U1s were "walked" along their target transcripts, from 10 nt up- to 25 nt downstream of the exon-intron junction (with reference to the 5'-most position of the target sequence). The endogenous U1 carrying a single-nucleotide adaptation to the mutation (labelled as "E") was also designed. (C) Targetome analysis of the 30,204 newly designed U1s at each target position. Counts of perfectly complementary targets are depicted for exonic, intronic, as well as 5'-SSs and 3'-SSs overlapping targets. An optimal target position is located 1 nt downstream of the exon-intron junction (highlighted in green).

<https://doi.org/10.1371/journal.pcbi.1013534.g005>

On the other hand, the lowest exonic and intronic targets are observed for U1s targeting positions -9, -8, -4, and +1 ([S10 Table](#)), which further affirm position +1 as an optimal target position.

Finally, the 23 5'-SS mutations targeted by the 54 published U1s ([S3 Table](#)) were revisited, for which optimized U1s targeting position +1 were designed and their targetome predicted at both perfect complementarity and 9 MABs. Consistently, the optimized U1s have significantly lower target counts at both 5'-SSs and 3'-SSs than the 54 published U1s, while no considerable differences in their exonic and intronic target counts were discerned ([S14 Fig](#)). In conclusion, the distal targeting strategy is demonstrated to improve the selectivity of modified U1s through reducing off-targeting events, especially at the splice sites, even when allowing mismatches and without the cost of increased off-targets at exons or introns. This forms the basis for the rational design of modified U1s that can be facilitated by Utargetome.

Discussion

We developed Utargetome, a computational pipeline for predicting the transcriptome-wide binding sites, or targetome, of a U1. In addition to base-pairing mismatches between the U1 binding sequence and target RNAs, the pipeline considers six known alternative annealing registers, which include single- and double-nucleotide bulges and asymmetric loops between the base-pairing strands. Also, the pipeline annotates the relative location of the predicted targets, whether they are exonic, intronic, or overlapping with 5'-SSs or 3'-SSs. The pipeline was first tested to predict the targetome of the human U1, *RNU1-1*, and of both *Arabidopsis thaliana* and *Dictyostelium discoideum*, which all share the same U1 binding sequence. Once validated, the pipeline was applied to obtain and analyse the human targetome of 54 modified U1s, which had been experimentally validated to be effective in restoring exon inclusion as a potential therapeutic strategy for 5'-SS pathogenic mutations. The selectivity of the U1s, which was inferred from both their full targetome size and targets at splice sites, was found to be wide-ranging. However, six modified U1s that were designed to bind distal positions were observed to have significantly reduced off-target counts at both 5'-SSs and 3'-SSs, suggesting that distal targeting can improve U1 selectivity. This evidence was then leveraged for the design of 30,204 U1s targeting 839 unique 5'-SS pathogenic mutations, which collectively implicate the splicing of 763 exons encoded in 500 genes. Analysis of the 30,204 targetomes predicted by the pipeline converge to an optimal U1 distal-targeting position at 3 nt downstream from the canonical 5'-SS, which leads to minimised off-target events, especially at 5'-SSs and 3'-SSs. The results justify the rationale of considering this particular distal position in the design of a U1 binding sequence.

Comprehensive combinations of mismatches and annealing registers were considered by the pipeline as the base-pairing mechanisms between a U1 binding sequence and its RNA binding site. Detailed analysis of the targetome predicted for the endogenous human U1 recapitulated not only the well-established specificity for 5'-SSs and alternative annealing registers, but also less studied aspects such as the enrichment of target sites at 3'-SSs and non-canonical 5'-SSs positions. Future transcriptome-wide RNA–RNA interaction data using psoralen-based crosslinking techniques [[33](#)] will be required to ascertain whether these target sites are indeed sites of base-pairing interactions with U1. Nonetheless, a very small fraction of annotated 5'-SSs in the human transcriptome was not found in the predicted targetome. This is inferred to be due to mechanisms not considered by the current pipeline, such as: 1) unelucidated alternative annealing registers; 2) alternative splicing mechanisms for 5'-SSs that harbour atypical dinucleotides at positions +1 and +2 [[31](#)]; 3) intron processing through the minor spliceosome, which uses different snRNAs and a distinct 5'-SS motif [[31](#),[33](#)]; and 4) additional protein factors possibly mediating U1 binding to RNA targets [[2](#)]. Although the predicted targetome for the human endogenous U1 is likely underestimated, it may not necessarily be the case for modified U1s, as they may not utilize such alternative base-pairing mechanisms.

Analysis of the human targetome of 54 modified U1s showed a wide range of off-target counts, spanning a few orders of magnitude in some cases. This facilitates, for the first time, the classification of U1s based on their selectivity and therefore provides a metric for selectivity in the design of the U1 binding sequence. Of relevance, the full targetome size of a modified U1 may have an indirect effect on its efficiency due to a “sponge” effect, in which the on-target site needs to

compete with the transcriptome-wide off-target sites for binding. At the same time, the full targetome size cannot be a definite indicator of selectivity, as it does not always correlate with the number of target sites at both the 5'-SSs and 3'-SSs. Based on the essential role of 5'-SSs in mediating both the on- and off-target functions of a modified U1, the target count therein was consequently chosen as a selectivity indicator, also considering that the major off-target effect would be the increased inclusion levels of alternatively or poorly spliced exons. In support of this, such biologically relevant off-target exons were identified in the targetome of the least selective U1, U1-36. Of note, the pattern of selectivity across different U1s seemed to be generally preserved even when considering alternative registers and mismatches.

The distal-targeting strategy was inspired from the superior selectivity of specific U1s, amongst the published 54 modified U1s, that were designed to bind distal positions. This was corroborated from the targetome of 30,204 newly designed U1s targeting 839 unique 5'-SS pathogenic mutations at distal positions. For these, an optimal target site at position +1 was identified to be associated with the lowest off-target events at both 5'-SSs and 3'-SSs. Of note, novel U1 design at such position also improved the selectivity of most of the 54 modified U1s that were initially evaluated from literature. However, as the binding sequence of a U1 influences its efficiency and selectivity simultaneously, and depending on the specific mutation to rescue, positions outside of +1 may need to be screened to identify optimal U1 candidates. Furthermore, in the presence of proximal splice sites, it is possible that distal targeting may result in the use one of the inactive putative sites [52], again suggesting that each scenario should be evaluated individually.

Utargetome is showed to recapitulate with fidelity major known features of the human endogenous U1. Because of the infidelity of U1 binding, it is tricky to rank or discriminate target sites for their likelihood of binding by either MABs or Gibbs free energy of U1:target duplexes. Nonetheless future experimental validation shall be useful to quantitate the actual off-target rates as a direct comparison with RNAseq-based off-target analyses remains limited by the current literature. The few available transcriptome-wide studies of engineered U1s have reported relatively modest gene expression or splicing changes [25–27]. Furthermore, most of these studies were conducted in mouse tissues. Although one study performed RNAseq in human cells, it did not provide a comprehensive list of off-target events for direct comparison. At the same time, non-sequence based factors such as chemical modifications of the binding sequence (including pseudouridines and methylation) [1], and splicing factors or spliceosome components (such as U1-C, U5 and U6) [2] could be considered for implementing the methodology. In endogenous U1s, two conserved pseudouridines at positions 5 and 6 in the 5' end are known to enhance the thermodynamic stability of base-pairing with 5'-SSs and contribute to accurate splice site recognition and spliceosome assembly [53–55]. It is plausible that co-transcriptional pseudouridylation may occur on modified U1s as they are being expressed from plasmids or viral vectors [56]. {R1Q1} Analogously, the role of both U5 and U6 snRNAs is likely required for the efficacy of modified U1s including distal-targeting ones since during canonical splicing, they displace U1 and U2 snRNPs from the splice site, stabilize the spliceosomal complex and trigger the ligation of the two exons and intron removal [1,2]. U5 interacts with the last nucleotides of the upstream exon while U6 base-pairs with the intronic side of the 5' splice site. Future studies are thus required to clarify their mechanistic action and also investigate pseudouridylation on modified U1s and the effect on binding kinetics.

Overall, Utargetome can be readily applied to prioritise or drop out U1s prior to a screen by comparing the selectivity of their predicted targetomes. Moreover, predicted off-targets with biological relevance can inform and guide the analysis of RNAseq data from U1-treated disease models. In conclusion, we propose the application of targetome prediction in the design of U1 binding sequences and expedite the discovery and validation of optimal U1s.

Methods

ClinVar analysis

The entire ClinVar database was downloaded (release version in S11 Table), and annotated variants were analysed with Python3.9. Only “pathogenic” variants associated with degenerative disorders were considered, whereas variants labelled as “likely pathogenic”, “likely benign”, or lacking information about the related pathology and/or their genetic coordinates,

as well as variants related to “cancer”, were filtered out. The relative positions of the variants from their respective nearest annotated 5′-SS were determined. Only variants located within the canonical binding site of the endogenous U1, or equivalently from positions -3 to +8 from the exon-intron boundary, were selected. Variants located at positions +1 or +2 were further excluded as they are generally not rescuable by U1 approach [2]. The final selection of variants was based on whether they potentially affect the binding of the endogenous U1. The script and its manual are available for download in GitHub repository, under the name “uvariants” (link in [S11 Table](#)).

U1 targetome prediction pipeline

The targetome prediction pipeline was coded in Python3.9 under Linux and MacOSX environments and integrated with the BLASTn algorithm [24,57,58] (version in [S11 Table](#)). The implementation steps are summarized as follows. First, the input list of U1 binding sequences is converted to target sequences by complementarity. Second, annealing registers ([Figs 2](#) module A and [S1](#)) are implemented by deleting or inserting new nucleotide combinations at every possible position, except for the first and last positions: BS1 and BS2 (single- and double-nucleotide bulges on the RNA target strand) are implemented by inserting 1 and 2 nt respectively; BA1 and BA2 (single- and double-nucleotide bulges on the U1 strand) are implemented by deleting 1 and 2 nt respectively; ALS (asymmetric loops with the larger loop on the target RNA strand) are implemented by deleting 1 nt and inserting a new combination of 2 nt; ALA (asymmetric loops with the larger loop on the U1 strand) are implemented by deleting 2 nt and inserting 1 new nt. Base-pairing mismatches between the two strands are subsequently implemented as new combinations of nucleotides, with at least one nucleotide apart from the bulge(s)/loop (if present) to preserve their hypothetical structure. Collectively, the full set of target sequences generated constitutes all possible targets of the given U1. Third, BLASTn is used to search every query sequence from the full set of target sequences generated above, in one or more databases. Each database is built with BLASTn (command `makeblastdb`) from a list of exons, introns, 5′-SSs and 3′-SSs. These sequences are extracted by the pipeline from the annotation and assembly files of a given species (*H. sapiens*, *D. discoideum* or *A. thaliana* in this study, release versions in [S11 Table](#)). Mitochondrial genes (and chloroplast genes in *A. thaliana*) are excluded from the databases. Targets sequences are then searched using BLASTn (command `blastn`) with the following settings: `-task blastn-short`; `-word_size` as the exact length of the input sequence (depending on the annealing register, see [Figs 2](#) and [S1](#) for examples); `-strand plus`; `-evaluate 1000000000`; `-max_target_seqs 1000000000`. Fourth, BLASTed hits are filtered based on the following criteria: 1) hits on different annotated transcripts but corresponding to the same genomic sequence are counted as distinct, as they might induce different effects; 2) hits produced by different annealing registers but located in the same position of the same transcript (i.e., sharing the same 5′-most position) are counted as single hits. Finally, filtered targets are counted and classified as either exonic or intronic (i.e., fully residing within the exon or intron), or overlapping with 5′-SSs or 3′-SSs. 5′-SS and 3′-SS targets are additionally examined within custom ranges or specific positions from the splice site. The position of a predicted target in reference to a nearby 5′- or 3′-SS always refers to the position of the 5′-most nucleotide on the target sequence, i.e., the 3′-most nucleotide on the U1 antisense sequence, regardless of whether the base-pairing consists of a canonical Watson-Crick or a mismatched pairing ([S1 Fig](#)). The command-line version of the pipeline and its manual are available for download in GitHub repository, under the name “utargetome” (link in [S11 Table](#)).

Gibbs free energy (ΔG) of U1:target duplexes

The ΔG of RNA:RNA duplex formation between a given U1 sequence and its predicted target site were estimated using RNAcofold (version 2.7.0) from the ViennaRNA Package [59]. Sequences were aligned in 5′–3′ orientation and input in the standard format. Only base-pairing interactions between U1 and the target were considered. All predictions were performed using default thermodynamic parameters and temperature settings (37°C). The output ΔG values were directly used to assess the relationship between the number of predicted base pairs and the strength of RNA-RNA binding.

U1 sequences

The binding sequences of the 54 literature-validated U1s were gathered from each reference ([S3 Table](#)). The promoter, scaffold and terminator sequences of the U1 cassette were used to deduce the start and end sites of the binding sequences [60]. Novel U1s were designed for the ClinVar dataset of mutations either by extracting the complementary sequence of the target site from the mutated exon sequences, or by adapting the endogenous U1 sequence to the mutation ([S9 Table](#)).

RNAseq analysis

The following RNAseq datasets of human retinas were downloaded from the SRA repository ([S11 Table](#)): SRR15431770, SRR15431758, SRR15351389, SRR15351390, ERR5236661, SRR17467505, SRR15539412, SRR16846779 [61–64]. They were converted to FASTQ format using the SRA toolkit and aligned with the HISAT2 program [65] using the in-built human genome index (build GRCh38, index version 2.0.2-beta). The aligned SAM files were converted to BAM format and then sorted using the SAMtools package [66]. TPM values were calculated with StringTie [67] using the human genome annotation (release version in [S11 Table](#)). Total gene expression was calculated as the sum of the TPM values of all transcript variants. Transcripts with TPM < 0.5 and novel transcripts were excluded from the analysis. PSI values for a given exon were calculated as the ratio between the sum of the TPM values of all transcript variants carrying the given exon and the sum of the TPM values of all transcript variants of the same gene. Pathway and ontology enrichment analyses were performed with DAVID [68] ([S11 Table](#)).

Bioinformatics and statistical analysis

The total number of 5'-SSs in the genomes of *H. sapiens*, *D. discoideum* and *A. thaliana* were calculated from the respective annotation files (release versions in [S11 Table](#)) using Python3.9. All annotated transcript variants were counted, with same 5'-SSs from different variants constituting separate counts, consistently with the targetome prediction pipeline. Positional Weight Matrices (PWM) for [S7 Fig](#) were generated from the U1 target sequences of interest by WebLogo web tool ([S11 Table](#)). Statistical analysis was performed with GraphPad Prism 9. For the free energy calculation of RNA-RNA interactions ([S9B Fig](#)), simple linear regression was performed with 95% confidence intervals, and the Pearson correlation coefficient was also calculated using 95% confidence intervals. For the targetome size of the 30,204 de-novo designed U1s ([Figs 5C](#) and [S13 Fig](#)), statistical P-values were calculated with one-way ANOVA for repeated values using Tukey's correction, assuming Gaussian distribution of residuals, and performing multiple comparisons between the mean off-target count of each target position and the mean of every other position. For the targetome size of U1s with optimized design ([S14 Fig](#)), statistical P-values were calculated with paired t-test and corrected for False Discovery Rate using the Benjamini-Hochberg method.

Supporting information

S1 Fig. Representation of targetome creation for a given U1 in the Utargetome pipeline.

(DOCX)

S2 Fig. Targetome of the endogenous U1 of *A. thaliana*.

(DOCX)

S3 Fig. Targetome of the endogenous U1 of *D. discoideum*.

(DOCX)

S4 Fig. Proportion of alternative annealing registers in the targetomes of the endogenous U1 in *H. sapiens*, *A. thaliana* and *D. discoideum*.

(DOCX)

S5 Fig. Reported donor splice sites bound by the endogenous U1 through alternative annealing registers.
(DOCX)

S6 Fig. Dinucleotide combinations at positions +1 and +2 of 5'-SS targets of the endogenous U1 in *H. sapiens*, *A. thaliana* and *D. discoideum*.
(DOCX)

S7 Fig. Potential binding motifs of the endogenous U1 at distal positions in *H. sapiens*, *A. thaliana* and *D. discoideum*.
(DOCX)

S8 Fig. Target distribution in proximity of 3'-SSs for the endogenous U1 and c(RNU1-1) in *H. sapiens*, *A. thaliana* and *D. discoideum*.
(DOCX)

S9 Fig. Analysis of free energy of binding for the predicted targets of the human endogenous U1 at positions overlapping with 5'-SSs, 3'-SSs and exonic regions.
(DOCX)

S10 Fig. 5'-SS targets for the 54 modified U1s.
(DOCX)

S11 Fig. Correlation between targetome size and decreasing complementarity for U1-1, U1-36, U1-49 and U1-54.
(DOCX)

S12 Fig. Targetome of the 54 modified U1s grouped by target mutation. Refer to [Fig 4A–4C](#) for the legend.
(DOCX)

S13 Fig. Targetome analysis at 10 MABs for de-novo designed U1s targeting 839 unique 5'-SS mutations at selected distal positions (-1, +1 and +2).
(DOCX)

S14 Fig. Comparison of targetome size between modified U1s validated in literature and the newly designed U1s targeting the distal position +1.
(DOCX)

S1 Table. Potential ClinVar pathogenic variants amenable for U1 therapy.
(XLSX)

S2 Table. Predicted 5'-SS targets (canonical position) with 9 MABs for the human endogenous U1 snRNA.
(XLSX)

S3 Table. Predicted targetome for the 54 modified U1s validated in literature.
(XLSX)

S4 Table. Predicted targetome for 4 selected modified U1s validated in the literature (U1-1, U1-36, U1-49 and U1-54).
(XLSX)

S5 Table. Targetome of U1-36 ("RHO_4_-1G>A_1").
(XLSX)

S6 Table. Average PSI of exons expressed in human retina whose 5'-SS was found in the targetome of U1-36 with perfect complementarity.

(XLSX)

S7 Table. Average PSI of exons expressed in human retina whose 5'-SS was found in the targetome of U1-36 with 9 MABs.

(XLSX)

S8 Table. Targetome of literature-validated modified U1s targeting distal positions.

(XLSX)

S9 Table. Targetome of 30,204 newly designed U1s targeting distal positions.

(XLSX)

S10 Table. Median and maximum target counts and statistical significance of 30,204 newly designed U1s.

(XLSX)

S11 Table. Web resources.

(XLSX)

Acknowledgments

We acknowledge Prof. Francesc Xavier Roca Castella for inspiring this work, and Wei Yuan Cher for advising the development of the targetome prediction pipeline.

Author contributions

Conceptualization: Paolo Pigni, Federico Manuel Giorgi, Keng Boon Wee.

Data curation: Paolo Pigni.

Formal analysis: Paolo Pigni.

Funding acquisition: Keng Boon Wee.

Investigation: Paolo Pigni, Keng Boon Wee.

Methodology: Paolo Pigni, Federico Manuel Giorgi, Keng Boon Wee.

Project administration: Paolo Pigni, Keng Boon Wee.

Resources: Paolo Pigni.

Software: Paolo Pigni, Federico Manuel Giorgi.

Supervision: Keng Boon Wee.

Validation: Paolo Pigni.

Visualization: Paolo Pigni.

Writing – original draft: Paolo Pigni.

Writing – review & editing: Paolo Pigni, Federico Manuel Giorgi, Keng Boon Wee.

References

1. Guiro J, O'Reilly D. Insights into the U1 small nuclear ribonucleoprotein complex superfamily. *Wiley Interdiscip Rev RNA*. 2015;6(1):79–92. <https://doi.org/10.1002/wrna.1257> PMID: 25263988

2. Wilkinson ME, Charenton C, Nagai K. RNA Splicing by the Spliceosome. *Annu Rev Biochem.* 2020;89:359–88. <https://doi.org/10.1146/annurev-biochem-091719-064225> PMID: 31794245
3. Venters CC, Oh J-M, Di C, So BR, Dreyfuss G. U1 snRNP Telescripting: Suppression of Premature Transcription Termination in Introns as a New Layer of Gene Regulation. *Cold Spring Harb Perspect Biol.* 2019;11(2):a032235. <https://doi.org/10.1101/cshperspect.a032235> PMID: 30709878
4. Hwu W-L, Lee Y-M, Lee N-C. Gene therapy with modified U1 small nuclear RNA. *Expert Rev Endocrinol Metab.* 2017;12(3):171–5. <https://doi.org/10.1080/17446651.2017.1316191> PMID: 30063459
5. Scotti MM, Swanson MS. RNA mis-splicing in disease. *Nat Rev Genet.* 2016;17(1):19–32. <https://doi.org/10.1038/nrg.2015.3> PMID: 26593421
6. Anna A, Monika G. Splicing mutations in human genetic disorders: examples, detection, and confirmation. *J Appl Genet.* 2018;59(3):253–68. <https://doi.org/10.1007/s13353-018-0444-7> PMID: 29680930
7. Tan J, Ho JXJ, Zhong Z, Luo S, Chen G, Roca X. Noncanonical registers and base pairs in human 5' splice-site selection. *Nucleic Acids Res.* 2016;44(8):3908–21. <https://doi.org/10.1093/nar/gkw163> PMID: 26969736
8. Roca X, Krainer AR, Eperon IC. Pick one, but be quick: 5' splice sites and the problems of too many choices. *Genes Dev.* 2013;27(2):129–44. <https://doi.org/10.1101/gad.209759.112> PMID: 23348838
9. Roca X, Akerman M, Gaus H, Berdeja A, Bennett CF, Krainer AR. Widespread recognition of 5' splice sites by noncanonical base-pairing to U1 snRNA involving bulged nucleotides. *Genes Dev.* 2012;26(10):1098–109. <https://doi.org/10.1101/gad.190173.112> PMID: 22588721
10. Freund M, Asang C, Kammler S, Konermann C, Krummheuer J, Hipp M, et al. A novel approach to describe a U1 snRNA binding site. *Nucleic Acids Res.* 2003;31(23):6963–75. <https://doi.org/10.1093/nar/gkg901> PMID: 14627829
11. Freund M, Hicks MJ, Konermann C, Otte M, Hertel KJ, Schaal H. Extended base pair complementarity between U1 snRNA and the 5' splice site does not inhibit splicing in higher eukaryotes, but rather increases 5' splice site recognition. *Nucleic Acids Res.* 2005;33(16):5112–9. <https://doi.org/10.1093/nar/gki824> PMID: 16155183
12. Roca X, Sachidanandam R, Krainer AR. Determinants of the inherent strength of human 5' splice sites. *RNA.* 2005;11(5):683–98. <https://doi.org/10.1261/rna.2040605> PMID: 15840817
13. Singh NN, Singh RN, Androphy EJ. Modulating role of RNA structure in alternative splicing of a critical exon in the spinal muscular atrophy genes. *Nucleic Acids Res.* 2007;35(2):371–89. <https://doi.org/10.1093/nar/gkl1050> PMID: 17170000
14. Fernandez Alanis E, Pinotti M, Dal Mas A, Balestra D, Cavallari N, Rogalska ME, et al. An exon-specific U1 small nuclear RNA (snRNA) strategy to correct splicing defects. *Hum Mol Genet.* 2012;21(11):2389–98. <https://doi.org/10.1093/hmg/dds045> PMID: 22362925
15. Singh NN, Del Rio-Malewski JB, Luo D, Ottesen EW, Howell MD, Singh RN. Activation of a cryptic 5' splice site reverses the impact of pathogenic splice site mutations in the spinal muscular atrophy gene. *Nucleic Acids Res.* 2017;45(21):12214–40. <https://doi.org/10.1093/nar/gkx824> PMID: 28981879
16. Yanaizu M, Sakai K, Tosaki Y, Kino Y, Satoh J-I. Small nuclear RNA-mediated modulation of splicing reveals a therapeutic strategy for a TREM2 mutation and its post-transcriptional regulation. *Sci Rep.* 2018;8(1):6937. <https://doi.org/10.1038/s41598-018-25204-2> PMID: 29720600
17. García-Moreno JF, Romão L. Perspective in Alternative Splicing Coupled to Nonsense-Mediated mRNA Decay. *Int J Mol Sci.* 2020;21(24):9424. <https://doi.org/10.3390/ijms21249424> PMID: 33321981
18. Pagani F, Buratti E, Stuani C, Bendix R, Dörk T, Baralle FE. A new type of mutation causes a splicing defect in ATM. *Nat Genet.* 2002;30(4):426–9. <https://doi.org/10.1038/ng858> PMID: 11889466
19. Denti MA, Rosa A, D'Antona G, Sthandier O, De Angelis FG, Nicoletti C, et al. Body-wide gene therapy of Duchenne muscular dystrophy in the mdx mouse model. *Proc Natl Acad Sci U S A.* 2006;103(10):3758–63. <https://doi.org/10.1073/pnas.0508917103> PMID: 16501048
20. Ferri L, Covello G, Caciotti A, Guerrini R, Denti MA, Morrone A. Double-target Antisense U1snRNAs Correct Mis-splicing Due to c.639 + 861C>T and c.639 + 919G>A GLA Deep Intronic Mutations. *Mol Ther Nucleic Acids.* 2016;5(10):e380. <https://doi.org/10.1038/mtna.2016.88> PMID: 27779620
21. Covello G, Ibrahim GH, Bacchi N, Casarosa S, Denti MA. Exon Skipping Through Chimeric Antisense U1 snRNAs to Correct Retinitis Pigmentosa GTPase-Regulator (RPGR) Splice Defect. *Nucleic Acid Ther.* 2022;32(4):333–49. <https://doi.org/10.1089/nat.2021.0053> PMID: 35166581
22. Fortes P, Cuevas Y, Guan F, Liu P, Pentlicky S, Jung SP, et al. Inhibiting expression of specific genes in mammalian cells with 5' end-mutated U1 small nuclear RNAs targeted to terminal exons of pre-mRNA. *Proc Natl Acad Sci U S A.* 2003;100(14):8264–9. <https://doi.org/10.1073/pnas.1332669100> PMID: 12826613
23. Abad X, Vera M, Jung SP, Oswald E, Romero I, Amin V, et al. Requirements for gene silencing mediated by U1 snRNA binding to a target sequence. *Nucleic Acids Res.* 2008;36(7):2338–52. <https://doi.org/10.1093/nar/gkn068> PMID: 18299285
24. Goracznik R, Behlke MA, Gunderson SI. Gene silencing by synthetic U1 adaptors. *Nat Biotechnol.* 2009;27(3):257–63. <https://doi.org/10.1038/nbt.1525> PMID: 19219028
25. Rogalska ME, Tajnik M, Licastro D, Bussani E, Camparini L, Mattioli C, et al. Therapeutic activity of modified U1 core spliceosomal particles. *Nat Commun.* 2016;7:11168. <https://doi.org/10.1038/ncomms11168> PMID: 27041075
26. Balestra D, Scalet D, Ferrarese M, Lombardi S, Ziliotto N, C Croes C, et al. A Compensatory U1snRNA Partially Rescues FAH Splicing and Protein Expression in a Splicing-Defective Mouse Model of Tyrosinemia Type I. *Int J Mol Sci.* 2020;21(6):2136. <https://doi.org/10.3390/ijms21062136> PMID: 32244944

27. Donadon I, Bussani E, Riccardi F, Licastro D, Romano G, Pianigiani G, et al. Rescue of spinal muscular atrophy mouse models with AAV9-Exon-specific U1 snRNA. *Nucleic Acids Res.* 2019;47(14):7618–32. <https://doi.org/10.1093/nar/gkz469> PMID: [31127278](https://pubmed.ncbi.nlm.nih.gov/31127278/)
28. Del Corpo O, Goguen RP, Malard CMG, Daher A, Colby-Germinario S, Scarborough RJ, et al. A U1i RNA that Enhances HIV-1 RNA Splicing with an Elongated Recognition Domain Is an Optimal Candidate for Combination HIV-1 Gene Therapy. *Mol Ther Nucleic Acids.* 2019;18:815–30. <https://doi.org/10.1016/j.omtn.2019.10.011> PMID: [31734561](https://pubmed.ncbi.nlm.nih.gov/31734561/)
29. Schwartz SH, Silva J, Burstein D, Pupko T, Eyraas E, Ast G. Large-scale comparative analysis of splicing signals and their corresponding splicing factors in eukaryotes. *Genome Res.* 2008;18(1):88–103. <https://doi.org/10.1101/gr.6818908> PMID: [18032728](https://pubmed.ncbi.nlm.nih.gov/18032728/)
30. Schwartz S, Oren R, Ast G. Detection and removal of biases in the analysis of next-generation sequencing reads. *PLoS One.* 2011;6(1):e16685. <https://doi.org/10.1371/journal.pone.0016685> PMID: [21304912](https://pubmed.ncbi.nlm.nih.gov/21304912/)
31. Erkelenz S, Poschmann G, Ptok J, Müller L, Schaal H. Profiling of cis- and trans-acting factors supporting noncanonical splice site activation. *RNA Biol.* 2021;18(1):118–30. <https://doi.org/10.1080/15476286.2020.1798111> PMID: [32693676](https://pubmed.ncbi.nlm.nih.gov/32693676/)
32. Roca X, Krainer AR. Recognition of atypical 5' splice sites by shifted base-pairing to U1 snRNA. *Nat Struct Mol Biol.* 2009;16(2):176–82. <https://doi.org/10.1038/nsmb.1546> PMID: [19169258](https://pubmed.ncbi.nlm.nih.gov/19169258/)
33. Engreitz JM, Sirokman K, McDonel P, Shishkin AA, Surka C, Russell P, et al. RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent Pre-mRNAs and chromatin sites. *Cell.* 2014;159(1):188–99. <https://doi.org/10.1016/j.cell.2014.08.018> PMID: [25259926](https://pubmed.ncbi.nlm.nih.gov/25259926/)
34. Yeh C-S, Chang S-L, Chen J-H, Wang H-K, Chou Y-C, Wang C-H, et al. The conserved AU dinucleotide at the 5' end of nascent U1 snRNA is optimized for the interaction with nuclear cap-binding-complex. *Nucleic Acids Res.* 2017;45(16):9679–93. <https://doi.org/10.1093/nar/gkx608> PMID: [28934473](https://pubmed.ncbi.nlm.nih.gov/28934473/)
35. Pinotti M, Rizzotto L, Balestra D, Lewandowska MA, Cavallari N, Marchetti G, et al. U1-snRNA-mediated rescue of mRNA processing in severe factor VII deficiency. *Blood.* 2008;111(5):2681–4. <https://doi.org/10.1182/blood-2007-10-117440> PMID: [18156490](https://pubmed.ncbi.nlm.nih.gov/18156490/)
36. Balestra D, Maestri I, Branchini A, Ferrarese M, Bernardi F, Pinotti M. An Altered Splicing Registry Explains the Differential ExSpeU1-Mediated Rescue of Splicing Mutations Causing Haemophilia A. *Front Genet.* 2019;10:974. <https://doi.org/10.3389/fgene.2019.00974> PMID: [31649737](https://pubmed.ncbi.nlm.nih.gov/31649737/)
37. Singh RN, Singh NN. A novel role of U1 snRNP: Splice site selection from a distance. *Biochim Biophys Acta Gene Regul Mech.* 2019;1862(6):634–42. <https://doi.org/10.1016/j.bbagr.2019.04.004> PMID: [31042550](https://pubmed.ncbi.nlm.nih.gov/31042550/)
38. Tanner G, Glaus E, Barthelmes D, Ader M, Fleischhauer J, Pagani F, et al. Therapeutic strategy to rescue mutation-induced exon skipping in rhodopsin by adaptation of U1 snRNA. *Hum Mutat.* 2009;30(2):255–63. <https://doi.org/10.1002/humu.20861> PMID: [18837008](https://pubmed.ncbi.nlm.nih.gov/18837008/)
39. Potter C, Zhu W, Razafsky D, Ruzycycki P, Kolesnikov AV, Doggett T, et al. Multiple Isoforms of Nesprin1 Are Integral Components of Ciliary Rootlets. *Curr Biol.* 2017;27(13):2014–2022.e6. <https://doi.org/10.1016/j.cub.2017.05.066> PMID: [28625779](https://pubmed.ncbi.nlm.nih.gov/28625779/)
40. Grassmeyer JJ, Cahill AL, Hays CL, Barta C, Quadros RM, Gurumurthy CB, et al. Ca²⁺ sensor synaptotagmin-1 mediates exocytosis in mammalian photoreceptors. *Elife.* 2019;8:e45946. <https://doi.org/10.7554/eLife.45946> PMID: [31172949](https://pubmed.ncbi.nlm.nih.gov/31172949/)
41. Al-Khuzaei S, Broadgate S, Foster CR, Shah M, Yu J, Downes SM, et al. An Overview of the Genetics of ABCA4 Retinopathies, an Evolving Story. *Genes (Basel).* 2021;12(8):1241. <https://doi.org/10.3390/genes12081241> PMID: [34440414](https://pubmed.ncbi.nlm.nih.gov/34440414/)
42. Nikopoulos K, Farinelli P, Giangreco B, Tsika C, Royer-Bertrand B, Mbefo MK, et al. Mutations in CEP78 Cause Cone-Rod Dystrophy and Hearing Loss Associated with Primary-Cilia Defects. *Am J Hum Genet.* 2016;99(3):770–6. <https://doi.org/10.1016/j.ajhg.2016.07.009> PMID: [27588451](https://pubmed.ncbi.nlm.nih.gov/27588451/)
43. Kandaswamy S, Zobel L, John B, Santhiya ST, Bogedein J, Przemeczek GKH, et al. Mutations within the cGMP-binding domain of CNGA1 causing autosomal recessive retinitis pigmentosa in human and animal model. *Cell Death Discov.* 2022;8(1):387. <https://doi.org/10.1038/s41420-022-01185-0> PMID: [36115851](https://pubmed.ncbi.nlm.nih.gov/36115851/)
44. Matsevich C, Gopalakrishnan P, Obolensky A, Banin E, Sharon D, Beryozkin A. Retinal Structure and Function in a Knock-in Mouse Model for the FAM161A-p.Arg523 * Human Nonsense Pathogenic Variant. *Ophthalmol Sci.* 2023;3(1):100229. <https://doi.org/10.1016/j.xops.2022.100229> PMID: [36420180](https://pubmed.ncbi.nlm.nih.gov/36420180/)
45. Maggi J, Hanson JVM, Kurmann L, Koller S, Feil S, Gerth-Kahlert C, et al. Retinal Dystrophy Associated with Homozygous Variants in NRL. *Genes (Basel).* 2024;15(12):1594. <https://doi.org/10.3390/genes15121594> PMID: [39766861](https://pubmed.ncbi.nlm.nih.gov/39766861/)
46. Chen X, Liu Y, Sheng X, Tam POS, Zhao K, Chen X, et al. PRPF4 mutations cause autosomal dominant retinitis pigmentosa. *Hum Mol Genet.* 2014;23(11):2926–39. <https://doi.org/10.1093/hmg/ddu005> PMID: [24419317](https://pubmed.ncbi.nlm.nih.gov/24419317/)
47. Bianco L, Antropoli A, Benadji A, Condroyer C, Antonio A, Navarro J, et al. RDH5 and RLBP1-Associated Inherited Retinal Diseases: Refining the Spectrum of Stationary and Progressive Phenotypes. *Am J Ophthalmol.* 2024;267:160–71. <https://doi.org/10.1016/j.ajo.2024.06.016> PMID: [38945349](https://pubmed.ncbi.nlm.nih.gov/38945349/)
48. Burstedt M, Whelan JH, Green JS, Holopigian K, Spera C, Greco E, et al. Retinal Dystrophy Associated With RLBP1 Retinitis Pigmentosa: A Five-Year Prospective Natural History Study. *Invest Ophthalmol Vis Sci.* 2023;64(13):42. <https://doi.org/10.1167/iovs.64.13.42> PMID: [37883093](https://pubmed.ncbi.nlm.nih.gov/37883093/)
49. Beryozkin A, Aweidah H, Carrero Valenzuela RD, Berman M, Iguzquiza O, Cremers FPM, et al. Retinal Degeneration Associated With RPGRIP1: A Review of Natural History, Mutation Spectrum, and Genotype-Phenotype Correlation in 228 Patients. *Front Cell Dev Biol.* 2021;9:746781. <https://doi.org/10.3389/fcell.2021.746781> PMID: [34722527](https://pubmed.ncbi.nlm.nih.gov/34722527/)
50. Pinotti M, Balestra D, Rizzotto L, Maestri I, Pagani F, Bernardi F. Rescue of coagulation factor VII function by the U1+5A snRNA. *Blood.* 2009;113(25):6461–4. <https://doi.org/10.1182/blood-2009-03-207613> PMID: [19387004](https://pubmed.ncbi.nlm.nih.gov/19387004/)

51. Donadon I, Pinotti M, Rajkowska K, Pianigiani G, Barbon E, Morini E, et al. Exon-specific U1 snRNAs improve ELP1 exon 20 definition and rescue ELP1 protein expression in a familial dysautonomia mouse model. *Hum Mol Genet.* 2018;27(14):2466–76. <https://doi.org/10.1093/hmg/ddy151> PMID: [29701768](https://pubmed.ncbi.nlm.nih.gov/29701768/)
52. Brackenridge S, Wilkie AOM, Sreaton GR. Efficient use of a “dead-end” GA 5′ splice site in the human fibroblast growth factor receptor genes. *EMBO J.* 2003;22(7):1620–31. <https://doi.org/10.1093/emboj/cdg163> PMID: [12660168](https://pubmed.ncbi.nlm.nih.gov/12660168/)
53. Xu H, Kong L, Cheng J, Al Moussawi K, Chen X, Iqbal A, et al. Absolute quantitative and base-resolution sequencing reveals comprehensive landscape of pseudouridine across the human transcriptome. *Nat Methods.* 2024;21(11):2024–33. <https://doi.org/10.1038/s41592-024-02439-8> PMID: [39349603](https://pubmed.ncbi.nlm.nih.gov/39349603/)
54. Kierzek E, Malgowska M, Lisowiec J, Turner DH, Gdaniec Z, Kierzek R. The contribution of pseudouridine to stabilities and structure of RNAs. *Nucleic Acids Res.* 2014;42(5):3492–501. <https://doi.org/10.1093/nar/gkt1330> PMID: [24369424](https://pubmed.ncbi.nlm.nih.gov/24369424/)
55. Lin T-Y, Mehta R, Glatt S. Pseudouridines in RNAs: switching atoms means shifting paradigms. *FEBS Lett.* 2021;595(18):2310–22. <https://doi.org/10.1002/1873-3468.14188> PMID: [34468991](https://pubmed.ncbi.nlm.nih.gov/34468991/)
56. Martinez NM, Su A, Burns MC, Nussbacher JK, Schaening C, Sathe S, et al. Pseudouridine synthases modify human pre-mRNA co-transcriptionally and affect pre-mRNA processing. *Mol Cell.* 2022;82(3):645–659.e9. <https://doi.org/10.1016/j.molcel.2021.12.023> PMID: [35051350](https://pubmed.ncbi.nlm.nih.gov/35051350/)
57. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215(3):403–10. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2) PMID: [2231712](https://pubmed.ncbi.nlm.nih.gov/2231712/)
58. Worley-Morse TO, Gunsch CK. A computational analysis of antisense off-targets in prokaryotic organisms. *Genomics.* 2015;105(2):123–30. <https://doi.org/10.1016/j.ygeno.2014.11.010> PMID: [25486012](https://pubmed.ncbi.nlm.nih.gov/25486012/)
59. Lorenz R, Bernhart SH, Höner Zu Siederdisen C, Tafer H, Flamm C, Stadler PF, et al. ViennaRNA Package 2.0. *Algorithms Mol Biol.* 2011;6:26. <https://doi.org/10.1186/1748-7188-6-26> PMID: [22115189](https://pubmed.ncbi.nlm.nih.gov/22115189/)
60. Lund E, Dahlberg JE. True genes for human U1 small nuclear RNA. Copy number, polymorphism, and methylation. *J Biol Chem.* 1984;259(3):2013–21. PMID: [6198328](https://pubmed.ncbi.nlm.nih.gov/6198328/)
61. Sterling JK, Baumann B, Foshe S, Voigt A, Guttha S, Alnemri A, et al. Inflammatory adipose activates a nutritional immunity pathway leading to retinal dysfunction. *Cell Rep.* 2022;39(11):110942. <https://doi.org/10.1016/j.celrep.2022.110942> PMID: [35705048](https://pubmed.ncbi.nlm.nih.gov/35705048/)
62. Parekh M, Ferrari S, Di Iorio E, Barbaro V, Camposampiero D, Karali M, et al. A simplified technique for in situ excision of cornea and evisceration of retinal tissue from human ocular globe. *J Vis Exp.* 2012;(64):e3765. <https://doi.org/10.3791/3765> PMID: [22733120](https://pubmed.ncbi.nlm.nih.gov/22733120/)
63. Wolf J, Boneva S, Rosmus D-D, Agostini H, Schlunck G, Wieghofer P, et al. In-Depth Molecular Profiling Specifies Human Retinal Microglia Identity. *Front Immunol.* 2022;13:863158. <https://doi.org/10.3389/fimmu.2022.863158> PMID: [35371110](https://pubmed.ncbi.nlm.nih.gov/35371110/)
64. Zhou J, Flores-Bellver M, Pan J, Benito-Martin A, Shi C, Onumere O, et al. Human retinal organoids release extracellular vesicles that regulate gene expression in target human retinal progenitor cells. *Sci Rep.* 2021;11(1):21128. <https://doi.org/10.1038/s41598-021-00542-w> PMID: [34702879](https://pubmed.ncbi.nlm.nih.gov/34702879/)
65. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. *Gigascience.* 2021;10(2):giab008. <https://doi.org/10.1093/gigascience/giab008> PMID: [33590861](https://pubmed.ncbi.nlm.nih.gov/33590861/)
66. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol.* 2019;37(8):907–15. <https://doi.org/10.1038/s41587-019-0201-4> PMID: [31375807](https://pubmed.ncbi.nlm.nih.gov/31375807/)
67. Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol.* 2015;33(3):290–5. <https://doi.org/10.1038/nbt.3122> PMID: [25690850](https://pubmed.ncbi.nlm.nih.gov/25690850/)
68. Sherman BT, Hao M, Qiu J, Jiao X, Baseler MW, Lane HC, et al. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res.* 2022;50(W1):W216–21. <https://doi.org/10.1093/nar/gkac194> PMID: [35325185](https://pubmed.ncbi.nlm.nih.gov/35325185/)