Accurately modelling RNase H-mediated antisense oligonucleotide efficacy

Barney Hill[1,3,5*], Maisie R. Jaques[2], Remya R. Nair[2], Nicola Whiffin[1,3,4], Matthew J. A. Wood[1,5], Stephan J. Sanders[1,5,6,7], Peter L. Oliver[2], Alyssa C. Hill[1,5], Carlo Rinaldi[1,5,*]

Abstract

By modulating gene expression, antisense oligonucleotides (ASOs) represent a powerful class of drugs with potential to treat a wide range of human diseases. However, prediction of ASO efficacy remains inaccurate, often necessitating large-scale and costly experimental screens to identify optimal candidates for specific targets. To address this challenge, we introduce *ASO Atlas*, a dataset comprising 188,521 RNase H-mediated ASOs with corresponding inhibition measurements, extracted from published patents targeting 334 unique genes. *ASO Atlas* provides the first systematic resource to rigorously evaluate hypotheses regarding key parameters in ASO design, including sequence composition, chemical backbone modifications, and target region selection. Using *ASO Atlas* we have trained *OligoAI*, a deep learning model capable of jointly modelling ASO sequence, RNA target context, sugar and backbone chemistries, and dosage to predict inhibition. We demonstrate the utility of *OligoAI* in accelerating ASO screening, achieving a 5.72-fold reduction in screening effort compared to random selection when selecting top-scored candidates that inhibit *KCNT2* expression. *ASO Atlas* and *OligoAI* have been made freely accessible through an online web-tool, facilitating the accelerated optimisation of ASO design.

[1]Department of Paediatrics,
University of Oxford, OX3 7TY Oxford, United Kingdom.
[2]MRC Nucleic Acid Therapy Accelerator,
Research Complex at Harwell, Harwell Research and Innovation Campus, Oxfordshire, OX11 0FA, UK
[3]Big Data Institute,
University of Oxford, Oxford, UK.
[4]Broad Center for Mendelian Genomics, Program in Medical and Population Genetics
Broad Institute of MIT and Harvard, Cambridge, MA, USA.
[5]Institute of Developmental and Regenerative Medicine (IDRM),
IMS-Tetsuya Nakamura Building, Old Road Campus, OX3 7TY Oxford, United Kingdom.
[6]New York Genome Center,
New York, NY 10013, USA.
[7]Department of Psychiatry and Behavioral Sciences,
UCSF Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, CA 94178, USA.
*Correspondence to: barney.hill@merton.ox.ac.uk, carlo.rinaldi@idrm.ox.ac.uk

Introduction

Antisense oligonucleotides (ASOs) are short, synthetic, single-stranded nucleic acids designed to selectively hybridise with target RNA sequences and modulate gene expression[1]. ASOs can operate through various mechanisms, ranging from reducing levels to altering splicing patterns of target RNA transcripts. The former is achieved by recruitment of RNase H enzymes, which cleave the RNA strand within the RNA-DNA heteroduplex formed by the ASO and its target[2,3].

RNase H-mediated ASOs typically employ a "gapmer" design. This architecture features a central gap of unmodified DNA nucleotides, essential for RNase H recognition and cleavage, flanked by 'wings' containing chemically modified nucleotides to enhance the ASO's drug-like properties[3]. Common examples include phosphorothioate (PS) linkages throughout the backbone, which replace non-bridging oxygen atoms with sulfur to confer nuclease resistance, and sugar modifications in the wings, such as 2'-O-Methoxyethyl (2'-MOE) or constrained Ethyl (cEt) BNA, which increase binding affinity and stability[4]. Gapmer structures are often denoted by the lengths of the 5' wing, gap, and 3' wing (e.g., a '5-10-5' configuration). This strategy effectively balances the requirement for RNase H activity with the need for enhanced stability and affinity.

Initially proposed as therapeutics over four decades ago[5], ASOs utilizing gapmer designs have recently gained significant clinical momentum, with landmark approvals such as volanesorsen for Familial Chylomicronemia Syndrome and inotersen for Transthyretin Amyloidosis demonstrating their viability for treating rare genetic disorders[6,7]. As their safety profile becomes increasingly established, opportunities for treating ultra-rare genetic diseases continue to expand[8–11].

Despite these clinical successes, the determinants of ASO efficacy remain poorly understood. Effective ASO design requires optimising two key components: the nucleotide sequence, which dictates target specificity, and the pattern of chemical modifications, which influence the drug-like properties. However, predicting the combined impact of sequence and chemistry on efficacy remains a significant challenge. With gene transcripts containing thousands of potential target sites, and there being many potential combinations of ASO length and chemical composition, current development strategies are heavily burdened by costly experimental screening efforts. Enhancing our understanding of the specific molecular features that govern ASO efficiency is therefore essential to reduce the burden of large-scale experimental screening and unlock their full potential for tackling genetic diseases. Computational methods such as ASOptimizer have sought to establish structure–activity relationships, but current predictive models fail to capture the complex interplay between sequence and chemistry, leaving ASO design largely dependent on extensive empirical testing[12].

Research in ASO design has been hampered by the lack of standardised, large-scale datasets linking sequence and chemistry to efficacy, with existing computational tools constrained by proprietary data that cannot be independently evaluated. In this study, we

present *ASO Atlas*, an extensive resource compiled from ASO screening data reported throughout the patent literature. Leveraging this comprehensive dataset, we have developed a machine-learning approach to jointly model how ASO sequence and chemical modifications influence efficacy, demonstrating a data-driven approach that enhances the identification of potent ASO candidates.

Results

ASO Atlas: A large-scale dataset of gapmer ASOs with corresponding in vitro efficacy measurements

The ASO Atlas comprises a diverse collection of 188,521 RNase H-mediated gapmer ASOs extracted from 417 USPTO (United States Patent and Trademark Office) patents published between 2001 and 2025. To construct this dataset, we implemented a multi-stage computational pipeline that began by manually labelling tables containing ASO inhibition data from patent applications made by IONIS Pharmaceuticals in USPTO. h. This dataset contains sequence composition, chemical modifications, target gene, and corresponding qRT-PCR mRNA abundance measurements across multiple cell lines. Manual review of 100 randomly selected entries across 100 patents revealed no numerical transcription errors, while 4% of chemical modification transcriptions contained at least one error.

The dataset features experiments conducted in a diverse range of cell-types including A-431 (*N*=44,855), HepG2 (*N*=23,428) and SH-SY5Y (*N*=18,192) and features 306 unique gene targets. The gene targets with the most ASOs were *F12* (*N* = 5,874), *DGAT2* (*N* = 5,326), and *PMP22* (*N* = 5,273) (Fig 1c). Among ASOs with chemistry annotations, 51.9% were 5-10-5 2'-MOE PS gapmers and 30.0% were 3-10-3 cEt PS gapmers. While the data spans from 2001 to 2025, we observed a notable increase in the volume of patented sequences in recent years (Fig 1b). The dataset exhibits a broad distribution of inhibition values (Fig 1d) with median inhibition of 45.0%. Importantly, 17.4% of ASOs show <10% inhibition, indicating substantial inclusion of ineffective sequences. This representation across the full efficacy spectrum suggests that patent filings document comprehensive screening results rather than selectively reporting successful candidates, providing robust training signal for machine learning models.

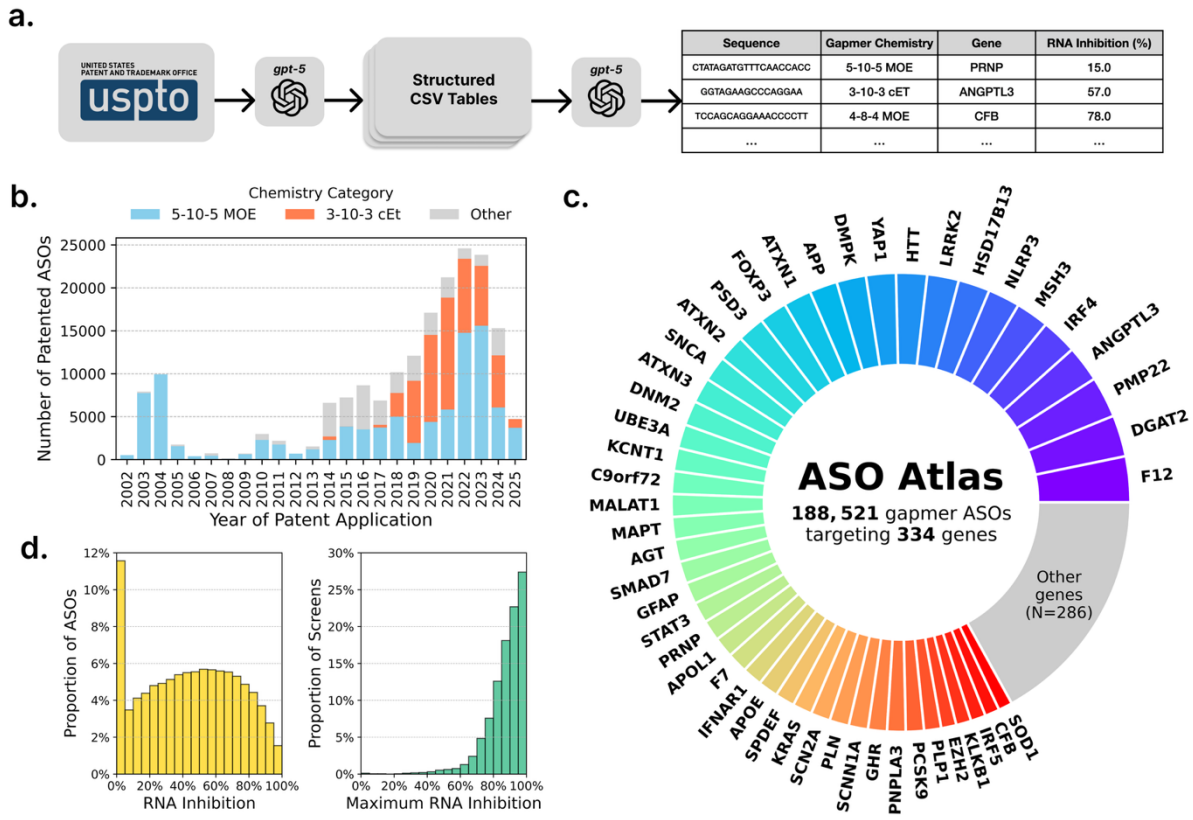


Figure 1. Construction and composition of the ASO Atlas dataset. *(a) Pipeline for creating the ASO Atlas dataset, showing extraction of ASO data from USPTO patents using gpt-5 for initial structured table extraction, followed by annotation of sequence, chemistry, target gene, and inhibition measurements. (b) Temporal distribution of patented ASO sequences from 2001-2025, showing a peak in patented sequences in 2023. Blue represents 5-10-5 MOE modified ASOs, orange represents 3-10-3 cEt ASOs and grey represents other chemistries. (c) Donut chart of ASO distribution across genes, demonstrating concentrated development efforts on select therapeutic targets. Major genes with ≥1,500 ASOs are labeled individually, while the "Other" category encompasses 286 genes with < 1,500 ASOs. (d) RNA inhibition distributions for all ASOs (left, yellow) showing a broad efficacy range including ineffective ASOs, and maximum inhibition per screen (right, green), indicating that a large proportion of screens achieve high inhibition with at least one ASO.*

Target site characteristics influence gapmer ASO efficacy

We systematically evaluated multiple factors influencing ASO efficacy, including sequence composition, genomic location, regulatory element overlap, RNA secondary structure. To identify position-specific nucleotide patterns associated with efficacy, we performed multivariate linear regression analysis for each nucleotide at each position along the ASO sequence, separately for 5-10-5 MOE (all PS) and 3-10-3 cEt (all PS) chemistries. After Bonferroni correction, we identified 83 significant position-specific nucleotide effects that

varied substantially between chemistry types (Fig 2a,b). For example, cytosines showed opposite effects between chemistries: 13/20 positions were significantly positively associated with inhibition in 5-10-5 MOE ASOs, while 12/16 positions showed negative association in 3-10-3 cEt ASOs. The only pattern shared across both chemistry types was thymine showing positive association with inhibition at positions 7-12 (5-10-5 MOE) and 1-9 (3-10-3 cEt) suggesting chemistry-dependent modulation of sequence composition effects.

Analysis of ASO efficacy across different genomic target region classifications revealed statistically significant variation in target inhibition ($p < 0.001$, Kruskal-Wallis test). We observed that ASOs targeting splice junction regions exhibited the lowest median inhibition (32.0% [IQR: 10.0%–57.0%]), followed by intron-targeting ASOs (40.0% [IQR: 17.0%–63.0%]) and 5'UTR-targeting ASOs (44.0% [IQR: 22.0%–65.0%]). In contrast, both 3'UTR and exon-targeting ASOs demonstrated the highest median inhibitions (53.0% [IQR: 28.0%–73.0%] and 53.0% [IQR: 29.0%–72.0%] respectively). We note that ASOs targeting coding exons exhibited similar inhibition to those targeting 3'UTR regions, while intron-targeting ASOs showed lower efficacy, likely due to exons being present in both pre-mRNA and mature mRNA transcripts (Fig 2a).

To further investigate the mechanisms underlying these regional differences, we first examined overlaps with miRNA and protein binding sites, given their enrichment in untranslated regions where we observed the varying efficacy[13]. After accounting for regional variation in ASO efficacy with a multivariate linear regression analysis controlling for genomic region type, only miRNA binding site overlap ($\beta = 4.43\%, p < 0.001$) was identified as a significant predictor of inhibition.

Next, we investigated the role of RNA secondary structure on efficacy based on overlap with ViennaRNA-predicted pre-mRNA secondary structure elements[14]. After controlling for screen-specific effects, ASOs targeting regions with fully accessible RNA showed 8.5 percentage points higher inhibition efficacy compared to those targeting fully bound regions, indicating that secondary structure accessibility significantly influences ASO performance ($p < 0.001$, linear regression). Predicted structure types showed distinct effects on efficacy. Hairpin structures were associated with decreased median inhibition (43.0% vs 45.0%, $N = 18{,}459$ vs $141{,}051$, $p < 0.001$, Mann-Whitney U test) whereas multiloop structures enhanced performance (49.0% vs 45.0%, $N = 8{,}197$ vs $153{,}313$, $p < 0.001$, Mann-Whitney U test). Exterior loop structures also improved efficacy (49.0% vs 45.0%, $N = 712$ vs $158{,}798$, $p = 0.0036$, Mann-Whitney U test) (Fig 2d).

Finally, to identify sequence motifs that influence ASO inhibition efficacy, we analysed all possible 4-mer nucleotide patterns while controlling for variability between experimental screens. The 'GGGG' motif showed the most pronounced significant association; ASOs containing this sequence exhibited a markedly lower median inhibition (24.0%), a finding consistent with previous studies[15] (Fig 2c). Several other motifs, including 'AAAA' (29.0%), 'TAAA' (32.0%), 'CTAA' (34.0%), and 'CCTA' (35.0%), were also significantly associated with lower inhibition (all $p < 0.001$). In contrast, motifs such as 'TTGT'

(53.0%), 'GTAT' (54.0%), 'CGTA' (54.0%), 'GTCG' (54.7%), and 'GCGT' (57.0%) were significantly associated with higher inhibition.

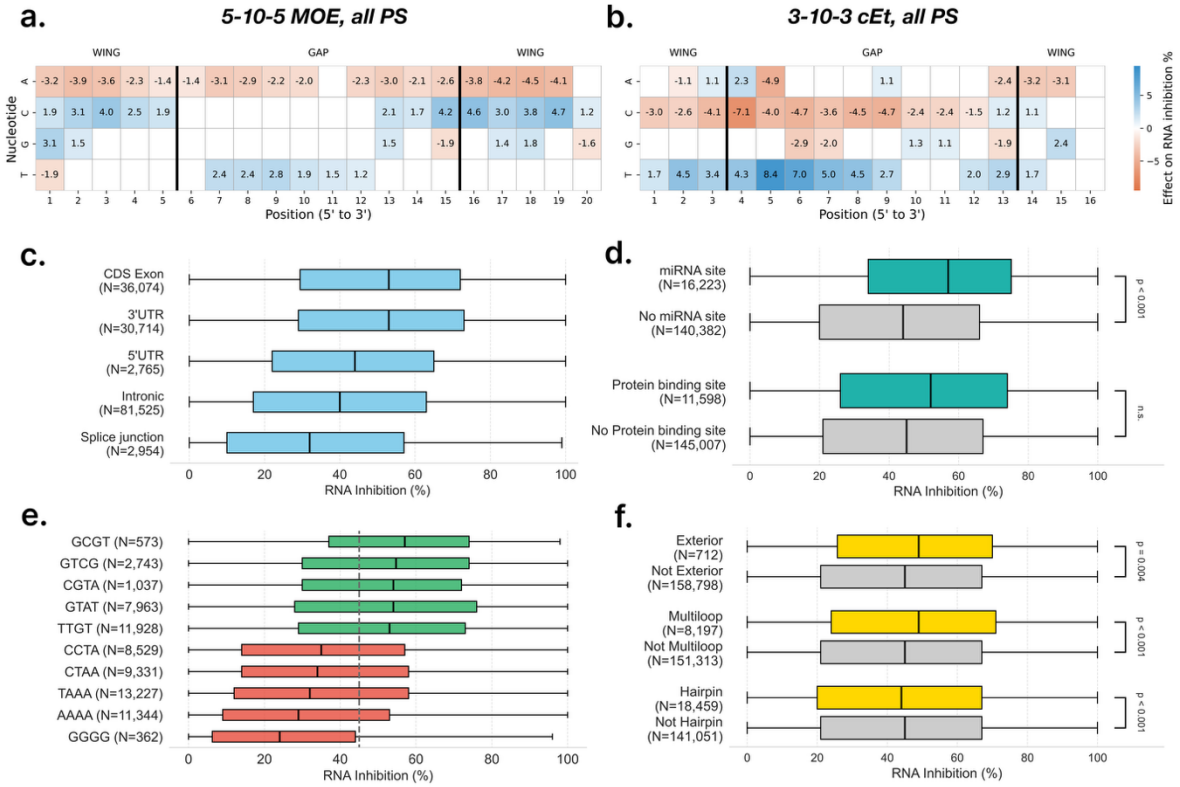


Figure 2. Analysis of ASO inhibition efficacy across genomic regions and sequence contexts. *(a,b) Heatmaps of position-specific nucleotide effects on ASO inhibition for 5-10-5 MOE, all PS and 3-10-3 cEt, all PS chemistries, respectively. Effects from linear regression are shown only where significant after Bonferroni correction. Colors indicate effect magnitude on RNA inhibition percentage, with blue representing positive effects and orange representing negative effects on inhibition. (c) Comparison of inhibition distributions between ASOs overlapping different genomic regions. Regions are presented in order of median inhibition and all group differences are significant (Mann-Whitney U, see Methods). (d) Comparison of inhibition distributions between ASOs overlapping miRNA binding sites as reported by TarBase v9 (Skoufos et al. 2024) and Protein binding sites as reported by RBP-Tar (Gresova et al. 2024). (e) Bar chart displaying top/bottom-5 median inhibition differences of 4-mer motifs (All significant after accounting for screen confounding with linear model, see Methods), with red bars showing inhibitory motifs and blue bars showing enhancing motifs. (f) Comparison of inhibition distributions between ASOs overlapping RNA secondary structure elements including multiloop, external, hairpin as annotated by ViennaRNA 2.0 (Lorenz et al. 2011) using ± 150 bp context around the ASO target.*

OligoAI: A deep learning model for ASO efficacy prediction

To leverage the sequence, structural, and positional determinants identified in ASO Atlas, we developed OligoAI (https://sitlabs.org/OligoAI), a transformer-based model that jointly encodes ASO nucleotide sequence, target RNA context, chemical modifications, and experimental conditions to predict in vitro target RNA knockdown (Fig. 3a)[18]. The model takes as input the ASO sequence and its target RNA context ($\pm$50 nucleotides flanking the hybridization site), position-specific chemical modifications including 2'-O-Methoxyethyl (2'-MOE) and constrained Ethyl (cEt) sugar modifications as well as phosphorothioate (PS) and phosphodiester (PO) backbone linkages, and experimental parameters such as dosage and transfection method.

The architecture integrates these feature modalities through a multi-stage encoding process. ASO sequences and their target RNA contexts are independently encoded using RiNALMo-giga, a 650-million parameter bidirectional transformer pre-trained on 36 million non-coding RNA sequences, to capture local secondary structure accessibility and sequence context effects. Chemical modifications are represented through learned embeddings and concatenated with the RiNALMo-encoded ASO sequence representations. These combined features are processed through a bottleneck network to learn cross-modal sequence-chemistry interactions. The resulting ASO and context representations are then globally pooled and integrated with the dosage-scaled transfection method embeddings through a multi-layer perceptron to generate final efficacy predictions

When evaluated across 299 held-out test screens, OligoAI achieved a median Spearman correlation of 0.419 [IQR: 0.290–0.533] between predicted and measured inhibition values, with a 3.14× enrichment factor for identifying high-performing ASOs (meaning the top 10% of OligoAI predictions contained 3.14 times more ASOs that were truly in the top 10% by measured inhibition compared to random selection. Table 1). This substantially outperforms previous state-of-the-art thermodynamics-based models ASOptimizer ($\rho = 0.076$, 1.53× enrichment) and OligoWalk ($\rho = 0.147$, 1.55× enrichment), which rely primarily on hybridisation energetics and target site accessibility and do not jointly model chemical modifications[12,19].

To assess OligoAI's practical utility for identifying hits for in vitro gene screens with a large number of possible sequence designs, we evaluated its performance on a large held-out screen targeting PSD3 ($N = 2{,}234$ ASOs, unseen gene). The model demonstrated robust discriminative ability: while the overall screen achieved a median inhibition of 65.0%, OligoAI's top 1% scored ASOs ($N = 22$) achieved a median inhibition of 92.5%, ($p < 0.001$, one-sample Wilcoxon test) (Fig. 3b) highlighting OligoAI's potential to substantially reduce experimental screening burden in therapeutic ASO development.

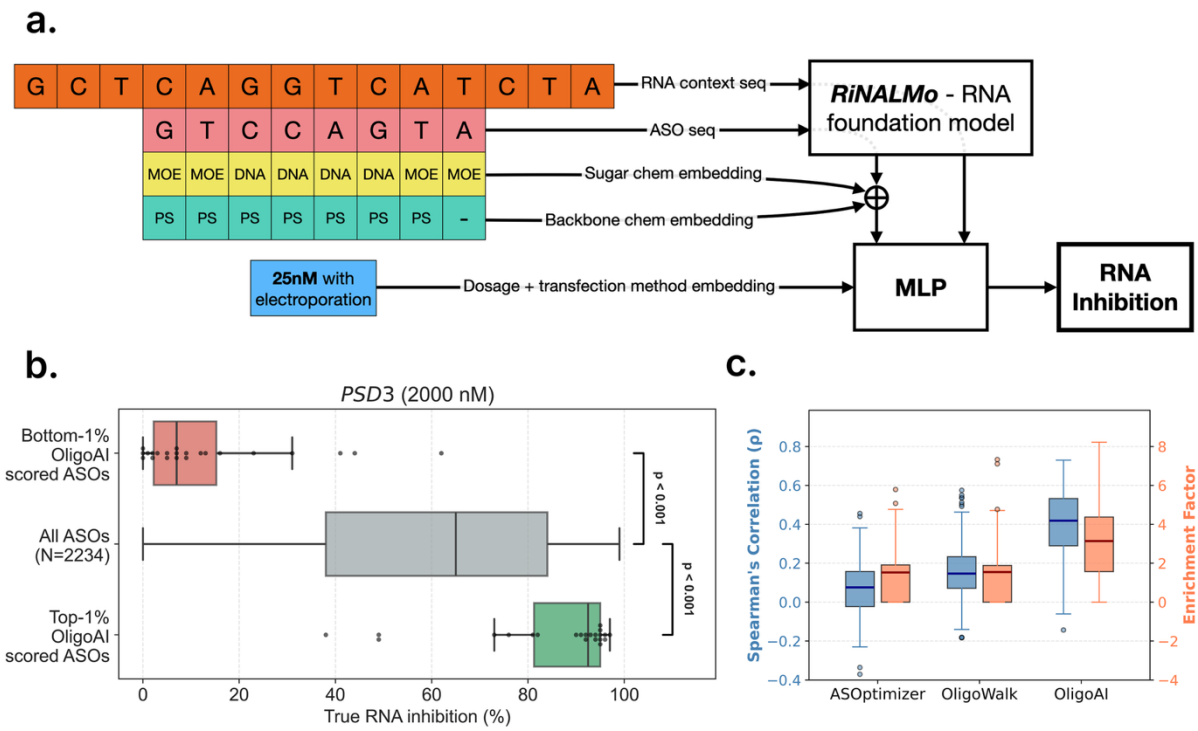


Figure 3. OligoAI model architecture and performance evaluation. *(a) Architecture of OligoAI showing the integration of multiple feature modalities. ASO sequences and their target RNA context (±50 nucleotides) are encoded using the pre-trained RiNALMo-giga transformer model. Position-specific sugar modifications and backbone chemistry are encoded through learned embeddings and concatenated with RiNALMo sequence representations. The final prediction is generated by a MLP that integrates the pooled ASO representation, pooled context representation, and dosage-scaled transfection method information to predict percent inhibition. (b) Validation on a large held-out screen (PSD3: $N = 2{,}234$) showing that OligoAI-predicted top 1% ASOs achieve significantly higher inhibition than screen average, while predicted bottom 1% ASOs show significantly lower inhibition (One-Sample Wilcoxon Test). (c) Distribution of Spearman correlation (ρ) and enrichment factor across $N = 299$ held-out test screens, comparing OligoAI performance to baseline methods (ASOptimizer, OligoWalk). Enrichment factor is the ratio of hit rate in the predicted top 10% to the baseline hit rate of 10%, where hits are defined as ASOs in the true top 10% by measured inhibition.*

Table 1: *Model performance comparison (median [IQR] values reported across validation test set, N=299 qRT-PCR screens)*

Model	Spearman ρ	Enrichment factor
ASOptimizer	0.076 (-0.023–0.158)	1.53× (0.00–1.92)
OligoWalk	0.147 (0.071–0.234)	1.55× (0.00–1.89)
OligoAI	**0.419 (0.290–0.533)**	**3.14× (1.57–4.38)**

Experimental validation of OligoAI prioritisation in *KCNT2* screening

To test the utility of OligoAI in a real-world experimental screening setting, we generated a virtual library of 200,374 20-mer unique gapmer ASOs targeting the Potassium Sodium-Activated Channel Subfamily T Member 2 (*KCNT2*), a gene associated with a rare developmental epileptic encephalopathy[20, 21]. Two groups of ASOs were defined for in vitro testing: 32 randomly sampled from the entire library and 18 randomly sampled from the 1% of ASOs with the highest OligoAI scores. To determine a suitable experimental dose for the ASO testing, a pilot experiment using two randomly chosen 'active' ASOs was carried out by transfection in HeLa cells. After testing a range of doses and incubation times, 30 nM dosing for 48-hours provided the most suitable dynamic range of *KCNT2* knockdown by the 'active' ASOs with no likely floor effects (Supplementary Fig. 1). In the subsequent complete library screening experiment, the relative *KCNT2* knockdown level was determined for all 50 ASOs in parallel, providing a ranking of *in vitro* efficacy (Supplementary Fig. 2).

ASOs ranked in the top 1% by OligoAI achieved superior target inhibition (median relative *KCNT2* expression = 0.19 [IQR: 0.12–0.36]) compared to randomly selected ASOs (median = 0.64 [IQR: 0.38–0.89]) (Mann-Whitney U test, $p = 1.1 \times 10^{-4}$). These results demonstrate OligoAI's ability to effectively identify high-performing ASO candidates (Fig. 4). Follow-up dose-response studies confirmed these results were reproducible, with IC50s in the low nanomolar range (Supplementary Fig. 3).

To quantify the practical screening efficiency gains, we performed bootstrap analysis (N = 10000 iterations) to estimate the screening effort required for random selection to achieve equivalent performance to OligoAI's top candidates. Our analysis revealed that to match the median inhibition efficacy of our 18 top scored ASOs, a standard random screening approach would require testing 103 ASOs (95% CI: 53–179). This represents a 5.72-fold reduction in required screening effort (95% CI: 2.94–9.94-fold), demonstrating substantial cost savings and accelerated candidate identification for ASO development workflows.

To confirm that OligoAI prioritises the previously identified biologically relevant features, we examined whether the top 1% ranked KCNT2 ASOs were enriched for genomic and sequence characteristics associated with efficacy. Top-ranked candidates showed significant enrichment for exonic targets ($n = 118/2006$ in top 1% vs $n = 5510/198368$ in remaining sequences; fold enrichment=2.12, $p = 1.43 \times 10^{-13}$, Fisher's exact test), presence of top-5 sequence motifs (TTGT, GTAT, CGTA, GTCG, GCGT; $n = 570/2006$ vs $n = 32707/198368$; fold enrichment=1.72, $p = 2.61 \times 10^{-40}$), and miRNA binding site overlap ($n = 22/2006$ vs $n = 658/198368$; fold enrichment=3.31, $p = 2.35 \times 10^{-6}$). Finally, proportion of unbound target RNA structure showed little difference between groups (mean=0.423 vs 0.421; fold change=1.00, $p = 0.048$, Mann–Whitney U test).

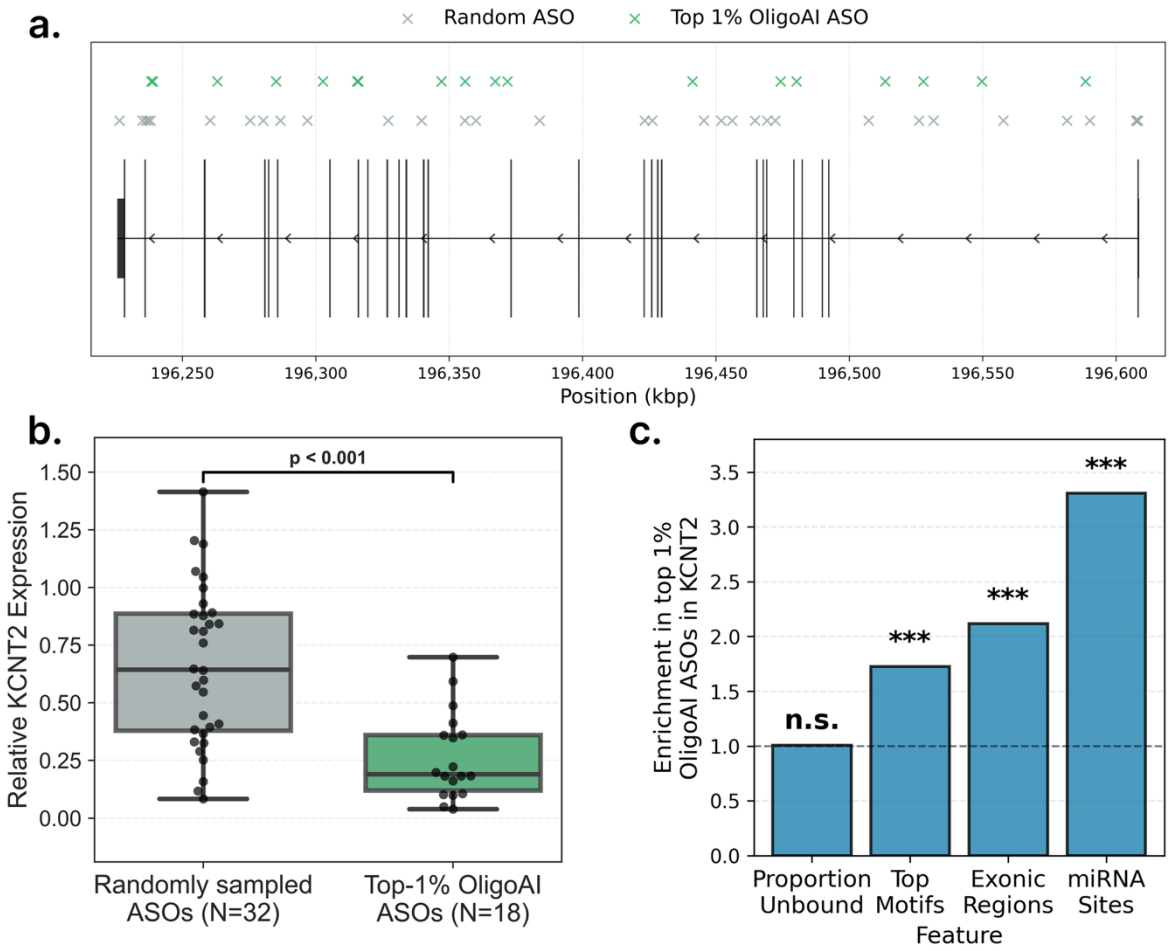


Figure 4. Experimental validation of OligoAI prioritisation in KCNT2 screening. *(a) Schematic representation of the KCNT2 transcript structure showing the genomic locations of ASOs sampled from two groups: randomly selected ASOs ($N = 32$), and top 1% OligoAI-ranked ASOs ($N = 18$). (b) Comparison of KCNT2 knockdown efficacy across the three ASO groups, expressed as relative target gene expression following transfection in HeLa cells. Top 1% OligoAI-ranked ASOs demonstrated significantly superior performance compared to randomly selected ASOs (Mann-Whitney U test, $p = 1.1 \times 10^{-4}$), with median relative expression levels of 0.19 versus 0.64, respectively. (c) Barplot showing the enrichment of biologically relevant features (exonic targeting, top-5 motifs, miRNA binding sites, and unbound RNA structure) within the top 1% ranked KCNT2 ASOs compared to remaining sequences. Asterisks indicate statistical significance after Bonferroni correction (***: p<0.001; n.s.: not significant).*

Discussion

Our systematic analysis of 188,521 gapmer ASOs reveals key principles governing RNase H-mediated ASO efficacy and establishes a predictive framework for therapeutic design.

The observed variation in efficacy across genomic regions, combined with secondary structure analysis, points to RNA accessibility as a central determinant of ASO performance. The reduced efficacy observed for hairpin-targeting ASOs compared to those targeting multiloop and exterior loop regions suggests that stable secondary structures present thermodynamic barriers to ASO hybridization. This interpretation is further supported by the positive association between miRNA binding site overlap and ASO efficacy, since miRNA target sites must be structurally accessible to enable miRNA binding, suggesting that ASOs similarly benefit from targeting accessible regions[22]. Together, these findings indicate that successful ASO design must consider not only sequence complementarity but also the structural and genomic context of the target site.

Following the understanding that ASO efficacy is influenced by not just the ASO sequence but also the RNA target context, we incorporated these factors alongside chemistry, dosage, and transfection method into OligoAI, a machine learning model that achieves strong predictive performance in predicting ASO in vitro efficacy. In a rigorous validation screen targeting *KCNT2*, OligoAI's top 1% prioritised candidates achieved significantly superior knockdown efficacy compared to random selection, with bootstrap analysis revealing that random screening would require testing 103 ASOs to match the performance of just 18 OligoAI-prioritised candidates, representing a 5.72-fold reduction in screening effort. These advances potentially address a significant bottleneck in antisense therapeutics, represented by the current reliance on large-scale screens for effective ASO sequences.

ASO design has historically relied on heuristic trial-and-error approaches with limited datasets, with computational tools primarily focusing on thermodynamic parameters while largely ignoring the chemical modifications that define modern therapeutics[23,24]. Existing algorithms such as OligoWalk predict hybridisation stability based on nearest-neighbour thermodynamics but fail to account for the phosphorothioate backbones and sugar modifications (2'-MOE, cEt) that comprise the majority of clinical ASO candidates[19]. More recently, ASOptimizer has emerged as a two-component framework that addresses both sequence selection and chemical modification optimization separately[12]. Its sequence engineering module employs a linear regression model incorporating three thermodynamic features: target site binding free energy, off-target binding potential, and secondary structure accessibility. ASOptimizer also includes a separate chemistry optimization module based on graph neural networks that optimizes modification patterns given a fixed sequence. However, this bifurcated approach has notable limitations. Our systematic analysis revealed strong interactive effects between sequence composition and chemistry type: position-specific nucleotide associations with efficacy were almost entirely different between 5-10-5 MOE and 3-10-3 cEt chemistries (Fig 2a,b), with only thymine showing consistent patterns across both. These chemistry-specific sequence preferences demonstrate that nucleotide composition and chemical modifications do not act independently but rather likely interact to determine ASO efficacy. By treating sequence selection and chemistry optimization as independent problems, ASOptimizer cannot capture these complex, non-linear interactions between nucleotide sequence, chemical modifications, and target RNA structure that govern ASO efficacy. In our

comparative evaluation across 299 held-out screens, ASOptimizer's sequence module achieved only modest predictive performance (Spearman $\rho = 0.076$), substantially lower than OligoAI ($\rho = 0.419$). OligoAI's improved performance likely stems from several key architectural differences including jointly modelling sequence and chemistry rather than treating them as independent optimization problems and leveraging a pre-trained RNA language model (RiNALMo) which has been shown to perform across various RNA modelling tasks.

While OligoAI represents a significant advance in ASO design prediction, several important limitations constrain its current utility and highlight areas for future development. First, the current model is limited to predicting RNase-H mediated ASOs. Since this modality requires more extensive screening efforts than others (for example, splice-switching ASOs that target a single exon), it provides a greater opportunity for model training and a stronger need for optimisation. Additionally, OligoAI's predictive accuracy, though substantially improved over baseline approaches, achieves only modest correlation with experimental outcomes, indicating that substantial variance in ASO efficacy remains unexplained. Our reliance on automatic extraction of patent-derived data represents a constraint, with LLM-based annotation introducing chemistry annotation errors in 4% of cases. Finally, our dataset lacks information on off-target effects, cytotoxicity, and *in vivo* efficacy: critical factors that ultimately determine therapeutic success but are not captured by our current model.

In summary, we have created the *ASO Atlas*, the largest publicly available database of ASOs with experimentally validated efficacy data, comprising 188,521 gapmers targeting 334 genes. Through systematic analysis of this resource, we uncovered the sequence, positional, and chemical modification features that govern ASO performance. We translated these insights into *OligoAI*, a deep learning prediction tool that achieves a 5.72-fold reduction in experimental screening burden, as rigorously validated through experimental testing in inhibiting the *KCNT2 transcript*. By making both the *ASO Atlas* database and *OligoAI* prediction tool freely available to the research community, we provide resources that will substantially reduce the cost and time associated with ASO development, accelerating the translation of antisense therapeutics from target identification to clinical application. When combined with emerging experimental platforms such as patient-derived organoid screening systems[25], this work enables more rapid progression toward personalised ASO therapeutics, ultimately enhancing the accessibility of precision medicine approaches for patients with rare genetic disorders.

Methods

Dataset collection and processing

Patent selection and initial data extraction

We constructed ASO Atlas by mining the USPTO Bulk Data Storage System (BDSS) for ASO efficacy data from 2001-2025. After filtering for patents published by either ISIS or IONIS Pharmaceuticals, we manually annotated tables containing inhibition data. From these tables we also included the preceding five paragraphs as contextual information.

Table structure normalisation and validation

We developed a multi-stage natural language processing pipeline to convert unstructured XML patent tables into standardised formats. The full prompts provided to each model and the final dataset conforms to a structured schema which are both fully defined in the supplementary. For table structure recognition, we employed OpenAI's gpt-526 with verbosity='low' and reasoning='low' to create a Python script to convert the raw XML table into normalised CSV format [26]. Subsequently, we performed schema validation using gpt-5 to generate SQL commands that concatenate the ASO sequence information and quantitative efficacy measurements (either inhibition percentage or UTC percentage) across tables.

Chemistry annotation and sequence characterization

For each validated entry, we implemented comprehensive annotation across three main stages. First, we used gpt-5 to systematically extract and standardise chemical modification patterns from patent text through chemistry annotation. Our annotation framework captured five primary modification types: phosphorothioate (PS) backbones, and 2'-MOE, cEt sugar modifications. For each ASO, we recorded both modification types and their precise positions along the sequence, with the model being provided with the five previous text paragraphs and row-specific data for this task. Second, we performed target information extraction using gpt-5 to annotate entries with metadata including target mRNA name, HGNC gene symbol, experimental cell line, ASO dosage (nM), cell density, and transfection method extracted from the five previous text paragraphs. Finally, we conducted genomic target mapping by mapping each ASO sequence to its target genomic location using the Ensembl 110 human genome assembly [27]. For each target gene, we selected the Ensembl canonical transcript, and ASO target sites were identified by searching for reverse complementary sequence matches between the ASO sequence and the pre-mRNA transcript.

Dataset quality assurance

We implemented rigorous quality control measures throughout the curation process. For sequence filtering, we restricted our dataset to ASOs with sequence lengths between 12-30 nucleotides and removed tables with duplicate sequence-inhibition pairs to prevent data redundancy. To ensure extraction accuracy, we performed manual verification by

cross-referencing 100 randomly selected entries across 100 different patents. No errors were detected in ASO sequences and inhibition values in our validation sample, with a 94% accuracy rate in chemistry annotation. Regarding model selection rationale, we employed gpt-5 for annotation tasks based on benchmarking that demonstrated superior task accuracy and cost-effectiveness compared to alternative models such as gpt-5-mini. The final curated dataset comprises 188,521 ASOs from 417 patents, with comprehensive annotation of sequences, chemical modifications, target genes, and experimental conditions (Table 2).

Table 2: *Antisense oligonucleotide modifications present in ASO Atlas*

Modification Type	ASO Count	Unique Patents
PS Backbone	185,585	409
2'-MOE	126,620	391
cEt	77,516	53
Any	**188, 521**	**417**

Position-wise nucleotide association testing

For each position along the ASO sequence and each nucleotide (A, C, G, T), we constructed a linear regression model with inhibition percentage as the dependent variable. The independent variables included: (1) a binary indicator for the presence of the specific nucleotide at the given position, (2) target gene identity, (3) transfection method, and (4) dosage (nM). This approach controlled for confounding effects of experimental conditions while isolating the contribution of each position-nucleotide combination to ASO efficacy.

Position-nucleotide combinations with insufficient variation (fewer than 5 ASOs containing or lacking the nucleotide at that position) were excluded from analysis. For each regression model, we extracted the coefficient and p-value corresponding to the nucleotide indicator variable. To account for multiple hypothesis testing across all position-nucleotide combinations, we applied Bonferroni correction, with statistical significance defined as corrected $p < 0.05$.

Comparison of inhibition efficacy by genomic target region

ASO target regions were classified hierarchically into mutually exclusive categories with the following priority: 5'UTR, 3'UTR, splice junction (overlapping an exon-intron boundary), CDS exonic and intronic as defined by Ensembl 110 canonical transcripts. ASOs overlapping miRNA or protein binding sites were defined by at least 10 nucleotides shared between the ASO target region and experimentally validated binding sites to reduce false positives. miRNA binding site regions were sourced from TarBase v9 using PAR-CLIP or HITS-CLIP data[16], while protein binding sites were obtained from RBP-Tar using eCLIP data ($p < 0.05$, Bonferroni-corrected for 515,644 sites)[17].

Comparisons of efficacy across different genomic target regions (5' UTR, Splice Junction, Intron, Exon, 3' UTR) utilised the Kruskal-Wallis H-test for an overall assessment of differences among the five region types. Following a statistically significant result, post-hoc pairwise comparisons between specific region types were performed using Dunn's test, with p-values adjusted for multiple comparisons using the Bonferroni correction method. All statistical analyses were performed using Python (v3.10) with the SciPy library (v1.15.1)[28] for the Kruskal-Wallis and Mann-Whitney U tests.

Motif Analysis

To identify ASO motifs associated with inhibition efficacy while controlling for experimental batch effects, we performed a systematic analysis of all 256 possible 4-mer motifs. For each motif, a multiple linear regression model was fitted to predict inhibition percentage based on the motif's presence, with the experimental screen (`custom_id`) included as a categorical covariate to account for screen-to-screen variability. To ensure adequate statistical power, only motifs present in at least 15 sequences were analysed. We applied a strict Bonferroni correction for multiple comparisons, establishing a significance threshold of $p < 0.05/256$ ($\approx 1.95 \times 10^{-4}$). The direction of a motif's effect (positive or negative) was determined from its regression coefficient, and all motifs presented in this study were significant under these criteria.

Computational models for ASO efficacy prediction

Deep learning model architecture

We implemented *OligoAI*, a transformer-based architecture that leverages the pre-trained RNA language model RiNALMo to predict ASO efficacy. The model processes multiple input modalities to capture both sequence and experimental context:

- **Sequence representations**: ASO sequences and their target RNA context (ASO binding site $\pm$50 flanking nucleotides) are encoded using RiNALMo-giga, a pre-trained bidirectional transformer model specialised for RNA sequences. Representations are extracted from the final transformer layer, yielding 1,280-dimensional embeddings per nucleotide position.
- **Chemistry track**: Position-specific sugar modifications (MOE, cEt, DNA) are encoded through learned embeddings (16 dimensions)
- **Backbone track**: Phosphorothioate (PS) or phosphodiester (PO) linkages are represented through separate embeddings (8 dimensions)
- **Experimental conditions**: Transfection method (Electroporation, Gymnosis, Lipofection, Other) is encoded through learned 4-dimensional embeddings and element-wise multiplied by log-transformed dosage (log1p) to capture method-specific dose-response relationships

The architecture integrates these features through a multi-stage process. First, ASO sequence representations from RiNALMo (1,280-dim) are concatenated with chemistry

(16-dim) and backbone embeddings (8-dim)[18]. These combined features (1,304-dim total) pass through a bottleneck network (linear 1,304-dim to 128, ReLU, dropout p=0.2, linear 128-dim to 1,280, ReLU) to learn cross-modal representations[29]. Both ASO and context representations are then processed through global pooling layers that project to 64 dimensions while handling variable-length sequences via masked mean pooling. The final prediction is made by a 3-layer MLP with 128 hidden dimensions that combines the pooled ASO representation (64-dim), pooled context representation (64-dim), and method-scaled dosage embedding (4-dim) through fully-connected layers with ReLU activations and dropout (p=0.3).

Training and evaluation methodology

The model was trained to directly predict percent inhibition values using a mean squared error (MSE) regression objective. Target inhibition values were standardized (zero mean, unit variance) using a StandardScaler fitted on the training set, with predictions inverse-transformed to the original scale for evaluation. We initialised the model with pre-trained RiNALMo-giga weights and employed a gradual unfreezing schedule during fine-tuning to preserve learned RNA representations while adapting to the ASO efficacy prediction task. At epoch 0, only randomly initialized components were unfrozen (prediction head, chemistry embedder, backbone embedder, transfection method embedder). At epoch 3, we unfroze the pre-trained RiNALMo final layer normalization and transformer blocks 6-39.

OligoAI was optimised using the Adam optimizer with a learning rate of 5×10^{-5}, weight decay of 0, and a linear learning rate schedule decaying to 5×10^{-6} (end factor = 0.1) over the total training steps[30]. Training proceeded for 10 epochs with a batch size of 64 samples, employing gradient clipping (max norm = 0.5) to ensure stable convergence. The model was trained with mixed precision (16-bit) on NVIDIA L40S GPUs (48GB memory) to improve computational efficiency. The final model was selected based on the epoch achieving the lowest validation set MSE.

We evaluated the model using an 80/10/10 train/validation/test split, with stratification performed at the patent level to prevent data leakage between related ASO sequences. This patent-level splitting ensures that all ASOs from a given patent appear in only one split, addressing the tendency for patents to contain structurally and functionally related oligonucleotides. Model performance was assessed using multiple metrics: coefficient of determination (R^2), mean absolute error (MAE), root mean squared error (RMSE), and Spearman rank correlation (ρ) calculated per experimental screen and averaged across screens. Enrichment factor was defined as the ratio of the hit rate in the predicted top 10% to the baseline hit rate of 10%, where hits were defined as ASOs in the true top 10% by measured inhibition.

Model comparisons

To compare the relative performance of OligoAI we used ASOptimizer and OligoWalk[12 19]. Specifically, from ASOptimizer we implemented the linear regression model from the sequence engineering module with the published regression parameters: $\hat{y} = a_0 + a_1 x_1 +$

$a_2x_2 + a_3x_3$, where $a_0 = -1.077$, $a_1 = -0.037$, $a_2 = 0.019$, $a_3 = 1.422$, x_1 represents the target Gibbs free energy change calculated by miRanda[31], x_2 represents the mean of the top-10 off-target Gibbs free energy calculated against canonical protein coding pre-mRNA transcripts by miRanda[31], and x_3 represents the secondary structure accessibility (ratio of unhybridised nucleotides) calculated by mFold[32]. From OligoWalk we used the `oligowalk_overall` metric, which incorporates multiple thermodynamic terms including duplex stability (ΔG_{duplex}), target mRNA secondary structure stability ($\Delta G_{target\_structure}$), oligo self-structure ($\Delta G_{intra\_oligomer}$), and the oligo-oligo dimer ($\Delta G_{inter\_oligomer}$) to predict overall binding affinity.

Cell culture, ASO treatment and gene expression analysis

HeLa cells were grown on 96-well plates (Thermo Scientific, cat no: 167008) at a seeding density of 10,000 cells per well in 100 μl media (DMEM with GlutaMAX (Gibco, cat no: 61965026), 10% FBS (Gibco, cat no: A3840402) and 1% Penicillin Streptomycin (Gibco, cat no: 15070063). The following day, cells were treated with individual ASOs at the doses indicated. The media was replaced with 100 μl fresh media before dosing and 20 μl of transfection mix (0.15 μl Lipofectamine RNAiMAX (Thermo Scientific, cat no: 13778150) with the ASO dose in OptiMEM (Gibco, cat no: 31985062) media). For the initial quantitative (q)RT-PCR assay optimisation, ASOs were dosed for 24, 48, or 72 hours at 3, 10, 30 or 100 nM. For the ASO library screen, ASOs were dosed for 48 hours at 30 nM. Three biological replicate wells were used for each treatment. Controls included untransfected cells (OptiMEM), RNAiMAX transfection with no ASO and a non-targeting control with a matching 5-10-5 2'-MOE gapmer design containing PS bonds throughout and 5'methyl-C / dC bases (5'-AGTCGCACACGTCTATACGC-3'). For dose-response analysis, a log10 dosing scale was used from 0.1 nM to 1 μM for 48 hours as above. ASOs were synthesised by ATDBio with QC by HPLC and LC / MS.

After ASO dosing, total RNA was extracted using the MagMAX mirVana Total RNA Isolation Kit (Applied Biosystems, cat no: A27828). RNA concentration was measured using the Qubit RNA High Sensitivity Assay Kit (Invitrogen, cat no: Q32855) and cDNA synthesis was carried out using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems, cat no: 4368813). Quantitative RT-PCR was performed using Fast SYBR Green Master Mix (Applied Biosystems, cat no: 4385618) on a CFX Opus 384 Real-Time PCR System (Bio-Rad). The following cycling conditions were used: enzyme activation at 95°C for 20 seconds, followed by 40 cycles of denaturation at 95°C for 3 seconds, annealing and extension at 60°C for 30 seconds. The qRT-PCR primer sequences were as follows: 5'-GTGCAGACACTCTTCAGGTTG-3' and 5'-AGCCTCTCTCCCGTTCTTTC-3' for *KCNT2*; 5'-AGTTCTGTGGCCATATGCTTAGTAG-3' and 5'-AAACAACAATCCGCCCAAAGG-3' for the reference gene *HPRT*. qRT-PCR data were analysed using the comparative Ct ($\Delta\Delta Ct$) method: the Ct values of the target gene were normalised to the Ct values of the reference gene. Relative gene expression levels were calculated using the formula $2^{-\Delta\Delta Ct}$. All reactions were performed in technical triplicates per biological replicate and the average Ct values used for analysis.

Data and Code Availability

The ASO Atlas dataset scripts and OligoAI training code are available at https://github.com/barneyhill/aso_atlas and https://github.com/barneyhill/OligoAI. The trained model can be downloaded at https://huggingface.co/barneyhill/OligoAI. An interactive web interface for ASO design is available at https://sitlabs.org/OligoAI.

Supplementary Material

ASO Atlas schema

ASO Atlas Dataset Schema. *Description of the fields, data types, and definitions for each entry in the ASO Atlas dataset.*

Field Name	Data Type	Description
`aso_sequence_5_to_3`	String	The nucleotide sequence of the ASO.
`inhibition_percent`	Float	The measured percentage of target RNA inhibition. Typically 0-100 although there are some values < 0 where upregulation was observed.
`chemistry`	Object	A object containing a list of all sugar and backbone modifications
`custom_id`	String	A string refering to the location of the referenced patent table file.
`target_mrna`	String	The name used to refer to the target mRNA.
`target_gene`	String	The HUGO gene name corresponding the the target_mrna
`cell_line`	String	The cell-line used for the screen.
`dosage`	Float	The dosage of the administered ASO in nM.
`cells_per_well`	Integer	The number of cells per well used in the screen.
`transfection_method`	String	The method of introducing the ASO into the cells.

LLM Prompts for Data Extraction

To ensure the reproducibility of our data extraction pipeline, this section details the verbatim prompts used for table normalisation and merging with `gpt-5`.

Prompt 1

```
## Task
Write a Python 3.11 script to convert OCR-extracted table XML data into a s
tructured CSV format.
```

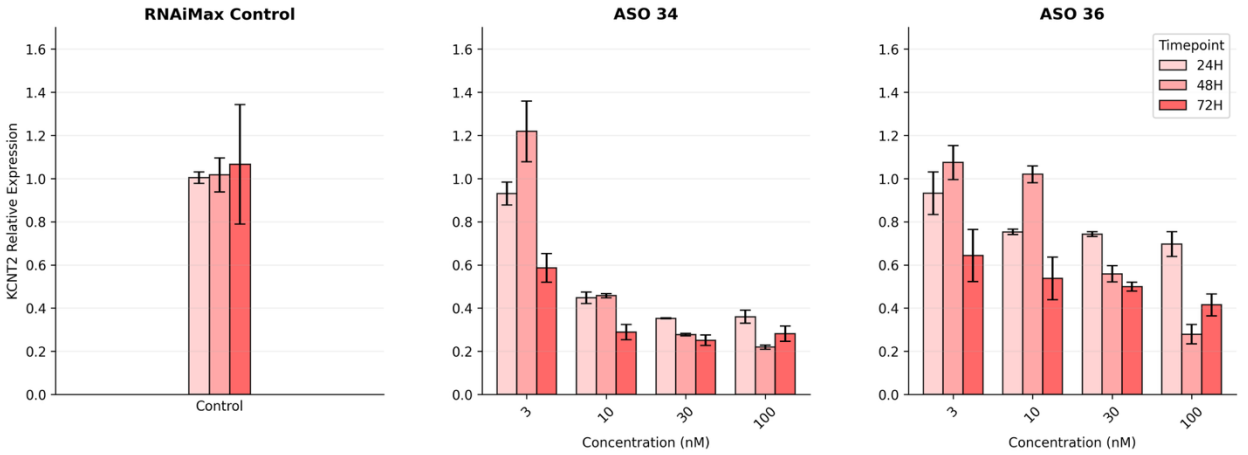
```
## Output Format
Return a Script object with:
- `pyscript`: Complete Python conversion script as a string

## Function Requirements
- The Python script should contain a function `xml_to_csv(xml_str: str) ->
str`
- You do need to return if __name__ ... w/ example usage - just our functio
n.

## Technical Requirements
### Dependencies
- Uses only Python standard library + `re` module

### Column Name Rules
- Preserve meaning of original column headers in new csv. i.e "UTC Untreate
d control group (%)" to "utc_untreated_control_group_pct" / "Inhibition (%)
" to "inhibition_pct"
- Make SQL-compatible: underscores for spaces, no dots, lowercase
- Hardcode column names (no need to dynamically generate)

### Data Handling
- Use "NA" for missing/empty cells
- Properly escape CSV values (quotes, commas, newlines)
- Use comma delimiter
- Sometimes   is used in the XML, this should be replaced with a spa
ce in the CSV.

## Domain Context
- UTC = "Untreated Control" percentage - this is not inhibition
- Preserve scientific notation and decimal precision
- Don't interpret abbreviations unless obvious

## Quality control
- To ensure correct rows let's strip newspace and capitalise the sequence c
olumn. It should have >=8 ATGC characters. If not skip the row.

## Preview of the input XML Structure (xml_str) - we will use the full vers
ion as input for your script (do not return this in your output!):
Your goal is to produce a dataset of antisense-oligonucleotide sequences an
d their inhibition percentages.
To do so you must stack the secondary_table with the primary_table using a
SQL command.

Required Columns in secondary_table:
1. ASO sequence (case insensitive)
2. One of:
   - inhibition/knockdown/reduction percentage
   -  UTC (Untreated Control) / RNA percentage
```

```
Transformation Rules:
- Numeric columns -> DOUBLE
- inhibition_percent =
    - Direct copy from inhibition/knockdown columns
    - 100 - UTC(%) for untreated control
- CONCAT two columns if ASO sequence is split (e.g.,sequence_part_one, sequ
ence_part_two)
- When using CAST be careful, some rows may not be castable to double, henc
e use TRY_CAST.

Task:
1. Generate SQL:
   - Stack (INSERT INTO) secondary_table onto primary_table
   - Apply transformations as needed
   - **CRITICAL: Always use "secondary_table" as the table name in your FRO
M clause**
   - **When referencing columns from secondary_table, use the exact column
names shown in the schema, including any special characters or numbers.**

Output Format:
- sql_command: string containing complete SQL command to stack the secondar
y_table onto primary_table

Data:
primary_table:
Schema:
- aso_sequence_5_to_3 (VARCHAR): 5'-3' ASO nucleotide sequence
- inhibition_percent (DOUBLE): target inhibition percentage, range 0-100
```

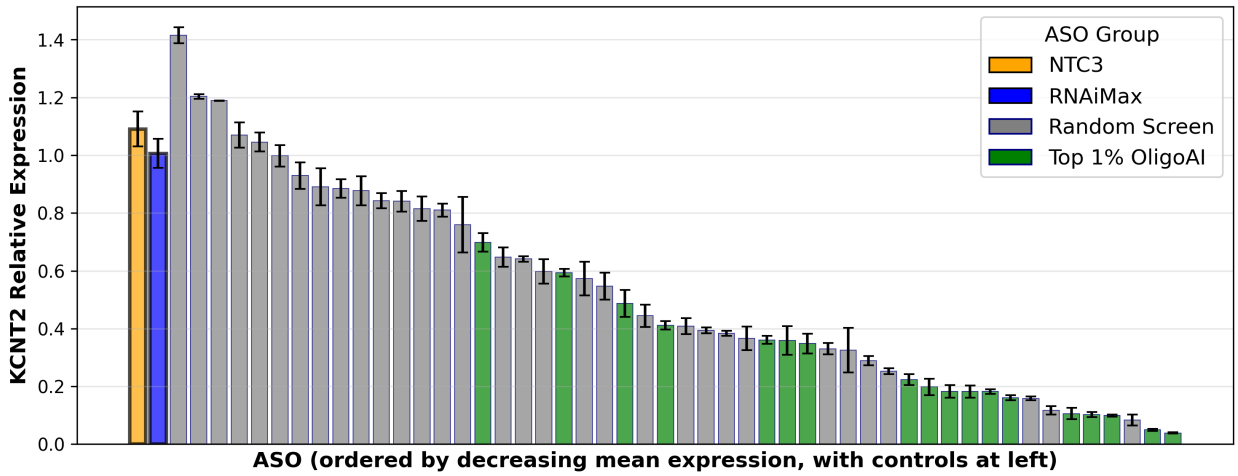
Prompt 2

```
Your goal is to produce a dataset of antisense-oligonucleotide sequences an
d their inhibition percentages.
To do so you must stack the secondary_table with the primary_table using a
SQL command.

Required Columns in secondary_table:
1. ASO sequence (case insensitive)
2. One of:
   - inhibition/knockdown/reduction percentage
   -  UTC (Untreated Control) / RNA percentage

Transformation Rules:
- Numeric columns -> DOUBLE
- inhibition_percent =
    - Direct copy from inhibition/knockdown columns
    - 100 - UTC(%) for untreated control
- CONCAT two columns if ASO sequence is split (e.g.,sequence_part_one, sequ
```

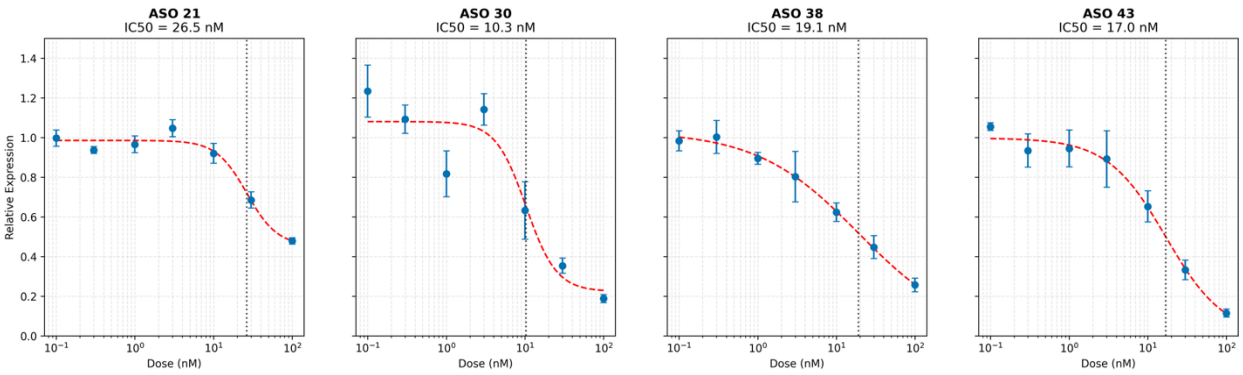
```
ence_part_two)
- When using CAST be careful, some rows may not be castable to double, henc
e use TRY_CAST.

Task:
1. Generate SQL:
   - Stack (INSERT INTO) secondary_table onto primary_table
   - Apply transformations as needed
   - **CRITICAL: Always use "secondary_table" as the table name in your FRO
M clause**
   - **When referencing columns from secondary_table, use the exact column
names shown in the schema, including any special characters or numbers.**

Output Format:
- sql_command: string containing complete SQL command to stack the secondar
y_table onto primary_table

Data:
primary_table:
Schema:
- aso_sequence_5_to_3 (VARCHAR): 5'-3' ASO nucleotide sequence
- inhibition_percent (DOUBLE): target inhibition percentage, range 0-100
```



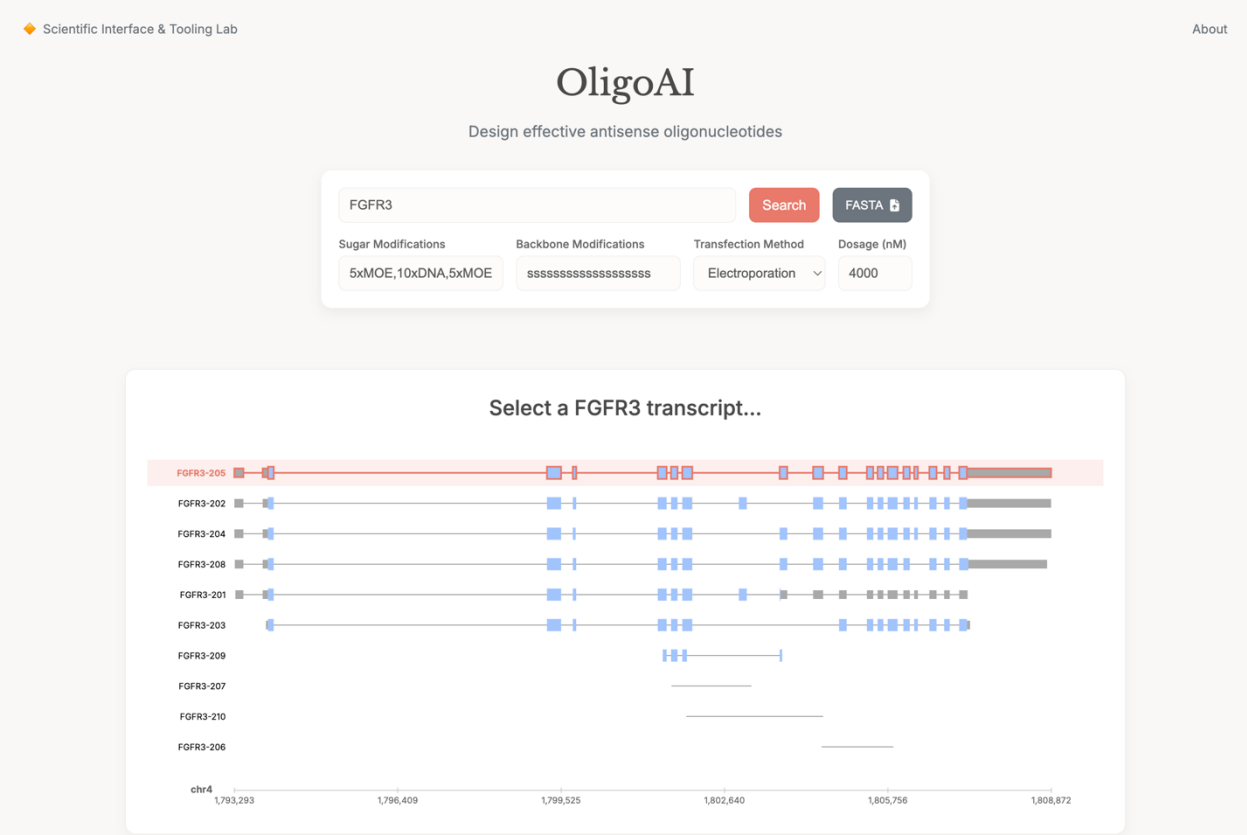
Supplementary Figure 1. Dose- and time-dependent knockdown of KCNT2 expression by selected ASOs in HeLa cells. *Individual panels show the relative activity for two ASOs. Cells were treated at concentrations of 3, 10, 30 and 100 nM for 24, 48 or 72 hours. KCNT2 mRNA levels were measured by quantitative RT-PCR and normalised to HPRT expression using the* $2^{-\Delta\Delta C_t}$ *method. Data represent mean* $\pm$ *SEM of three biological replicates per condition. Based on these results, 30 nM treatment for 48 hours was selected as the optimal condition. Data are shown relative to transfection reagent alone (RNAiMAX).*



Supplementary Figure 2. *Bar plot showing relative KCNT2 expression levels for all 80 screened ASOs ordered by decreasing mean expression (most effective knockdown on the left). HeLa cells were transfected with individual ASOs at 30 nM for 48 hours. Relative gene expression was determined by qRT-PCR using the ΔΔCt method with HPRT as the reference gene. Error bars represent SEM from three biological replicates. Bars are colored by ASO selection group: randomly selected ASOs (grey bars, standard screen, N = 32) and ASOs from the top 1% of OligoAI scores (green bars, predicted active, N = 18). The distribution demonstrates that OligoAI predictions correlate with experimental efficacy, with most top 1% ASOs showing greater knockdown activity (lower relative expression) compared to the random ASOs. Data are shown relative to transfection reagent alone (RNAiMAX) and a non-targeting control (NTC) ASO of matching chemical composition.*



Supplementary Figure 3. Individual dose-response curves showing relative KCNT2 expression levels following ASO treatment in HeLa cells. *Cells were transfected with ASOs at concentrations ranging from 0.1 nM to 1 μM and incubated for 48 hours. Each data point represents the mean ± SEM of three biological replicates. Relative gene expression was determined by qRT-PCR using the ΔΔCt method with HPRT as the reference gene. Red dashed lines show 4-parameter logistic curve fits (Hill equation). Black dotted vertical lines indicate calculated IC_{50} values. IC_{50} values demonstrate knockdown activity in the low nanomolar range.*



Supplementary Figure 4. OligoAI online portal. *Users can either supply a target RNA via FASTA file upload or select a Ensembl human transcript. When a target is specified, the OligoAI model processes the target RNA using serverless GPU inference. These results are returned to the user on the same page or via a email reminder and are available to download.*

References

1. Crooke, S. T., Baker, B. F., Crooke, R. M. & Liang, X.-H. Antisense technology: An overview and prospectus. *Nature Reviews. Drug Discovery* **20**, 427–453 (2021).

2. Liang, X.-H., Sun, H., Nichols, J. G. & Crooke, S. T. RNase H1-Dependent Antisense Oligonucleotides Are Robustly Active in Directing RNA Cleavage in Both the Cytoplasm and the Nucleus. *Molecular Therapy* **25**, 2075–2092 (2017).

3. Crooke, S. T. Molecular Mechanisms of Antisense Oligonucleotides. *Nucleic Acid Therapeutics* **27**, 70–77 (2017).

4. Khvorova, A. & Watts, J. K. The chemical evolution of oligonucleotide therapies of clinical utility. *Nature Biotechnology* **35**, 238–248 (2017).

5. Zamecnik, P. C. & Stephenson, M. L. Inhibition of Rous sarcoma virus replication and cell transformation by a specific oligodeoxynucleotide. *Proceedings of the National Academy of Sciences* **75**, 280–284 (1978).

6. Paik, J. & Duggan, S. Volanesorsen: First Global Approval. *Drugs* **79**, 1349–1354 (2019).

7. Keam, S. J. Inotersen: First Global Approval. *Drugs* **78**, 1371–1376 (2018).

8. Crooke, S. T. *et al.* The Effects of 2'-O-Methoxyethyl Containing Antisense Oligonucleotides on Platelets in Human Clinical Trials. *Nucleic Acid Therapeutics* **27**, 121–129 (2017).

9. Baker, B. F. *et al.* Safety and Tolerability of GalNAc3-Conjugated Antisense Drugs Compared to the Same-Sequence 2'-O-Methoxyethyl-Modified Antisense Drugs: Results from an Integrated Assessment of Phase 1 Clinical Trial Data. *Nucleic Acid Therapeutics* **34**, 18–25 (2024).

10. Kim, J. *et al.* Patient-Customized Oligonucleotide Therapy for a Rare Genetic Disease. *New England Journal of Medicine* **381**, 1644–1652 (2019).

11. O'Connor, D. J. *et al.* The Rare Therapies Launchpad: A pilot program for individualized medicines in the UK. *Nature Medicine* 1–2 (2025) doi:10.1038/s41591-025-03547-4.

12. Hwang, G. *et al.* ASOptimizer: Optimizing antisense oligonucleotides through deep learning for IDO1 gene regulation. *Molecular Therapy Nucleic Acids* **35**, (2024).

13. Mayya, V. K. & Duchaine, T. F. Ciphers and Executioners: How 3'-Untranslated Regions Determine the Fate of Messenger RNAs. *Frontiers in Genetics* **10**, (2019).

14. Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms for Molecular Biology* **6**, 26 (2011).

15. Matveeva, O. V. *et al.* Identification of sequence motifs in oligonucleotides whose presence is correlated with antisense activity. *Nucleic Acids Research* **28**, 2862–2865 (2000).

16. Skoufos, G. *et al.* TarBase-v9.0 extends experimentally supported miRNA–gene interactions to cell-types and virally encoded miRNAs. *Nucleic Acids Research* **52**, D304–D310 (2024).

17. Gresova, K. *et al.* RBP-Tar – a searchable database for experimental RBP binding sites. *F1000Research* **12**, 755 (2024).

18. Penić, R. J., Vlašić, T., Huber, R. G., Wan, Y. & Šikić, M. RiNALMo: General-Purpose RNA Language Models Can Generalize Well on Structure Prediction Tasks. *Nature Communications* **16**, 5671 (2025).

19. Lu, Z. J. & Mathews, D. H. Efficient siRNA selection using hybridization thermodynamics. *Nucleic Acids Research* **36**, 640–647 (2008).

20. Barcia, G. *et al.* De novo gain-of-function KCNT1 channel mutations cause malignant migrating partial seizures of infancy. *Nature Genetics* **44**, 1255–1259 (2012).

21. Heron, S. E. *et al.* Missense mutations in the sodium-gated potassium channel gene KCNT1 cause severe autosomal dominant nocturnal frontal lobe epilepsy. *Nature Genetics* **44**, 1188–1190 (2012).

22. Long, D. *et al.* Potent effect of target structure on microRNA function. *Nature Structural & Molecular Biology* **14**, 287–294 (2007).

23. Ho, S. P. *et al.* Mapping of RNA accessible sites for antisense experiments with oligonucleotide libraries. *Nature Biotechnology* **16**, 59–63 (1998).

24. Shao, Y., Wu, Y., Chan, C. Y., McDonough, K. & Ding, Y. Rational design and rapid screening of antisense oligonucleotides for prokaryotic gene modulation. *Nucleic Acids Research* **34**, 5660–5669 (2006).

25. Means, J. C. *et al.* Rapid and scalable personalized ASO screening in patient-derived organoids. *Nature* **638**, 237–243 (2025).

26. Introducing GPT-5. (2025).

27. Dyer, S. C. *et al.* Ensembl 2025. *Nucleic Acids Research* **53**, D948–D957 (2025).

28. Virtanen, P. *et al.* SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods* **17**, 261–272 (2020).

29. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research* **15**, 1929–1958 (2014).

30. Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. (2017) doi:10.48550/arXiv.1412.6980.

31. John, B., Enright, A. J., Aravin, A., Tuschl, T., Sander, C. & Marks, D. S. Human MicroRNA Targets. PLoS Biology 2, e363 (2004). doi:10.1371/journal.pbio.0020363

32. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Research 31, 3406–3415 (2003). doi:10.1093/nar/gkg595

33. Goldfarb, T. et al. NCBI RefSeq: reference sequence standards through 25 years of curation and annotation. Nucleic Acids Research 53, D243–D257 (2025).