SAPIENZA
UNIVERSITÀ DI ROMA

Sapienza University of Rome

Department of Biochemistry Science "A. Rossi Fanelli"
PhD in Biochemistry

Thesis For The Degree Of Doctor Of Philosophy

In silico design and evaluation of exon skipping-inducing antisense oligonucleotides for a potential therapeutic intervention in cancer

Advisor
Dr. Loredana Le Pera
Dr. Teresa Colombo

Candidate
Chiara Pacelli

Academic Year 2020-2023 (XXXVI cycle)

To Emanuele

...and mine Stella

Abstract

Precision medicine in oncology has made significant progress in recent years by approving drugs that target specific genetic mutations. However, many cancer driver genes remain challenging to pharmacologically target ("undruggable"). To tackle this issue, RNA-based methods like antisense oligonucleotides (ASOs) that induce targeted exon skipping (ES) could provide a promising alternative. In this work, a comprehensive computational procedure is presented, focused on supporting the development of ES-based cancer treatments. The procedure aims to produce specific protein variants, including inactive oncogenes and partially restored tumor suppressors. This novel computational procedure encompasses target exon selection, *in-silico* prediction of ES products, and identification of best candidate ASOs for further experimental validation. The method was effectively employed on extensively mutated cancer genes, prioritised according to both their suitability for ES-based interventions and clinical relevance. Relevant cancer-related genes, such as the NRAS, BRAF and CXCL8 (or IL-8) oncogenes, and the VHL and TP53 tumor suppressors, exhibited potential for this therapeutic approach, as specific target exons were identified and optimal ASO sequences were devised to induce their skipping towards desired protein variants. To the best of our knowledge, this is the first computational procedure that encompasses all necessary steps for designing ASO sequences tailored for targeted ES, contributing with a versatile and innovative approach to address the challenges posed by undruggable cancer driver genes, and beyond.

Keywords: exon skipping; antisense oligonucleotide; ASO; splicing; RNA therapy; cancer; bioinformatics.

Abbreviations

ASO	Antisense oligonucleotide
BMD	Becker muscular dystrophy
CDS	Coding DNA sequence
DMD	Duchenne muscular dystrophy
ES	Exon skipping
ESE	Exonic splicing enhancer
ESS	Exonic splicing silencer
FDA	Food and Drug Administration
hnRNP	heterogeneous nuclear ribonucleoprotein
ISE	Intronic splicing enhancer
ISS	Intronic splicing silencer
mRNA	messanger RNA
NMD	Nonsense Mediated Decay
ONG	Oncogene
PMO	Phosphorodiamidate morpholino oligomer
PTC	Premature termination codon
siRNA	small interfering RNA
snRNA	small nuclear RNA
TPM	Transcripts per million
TSG	Tumor suppressor gene

Contents

Chapter 1

Introduction

1.1 Cancer as a global health challenge

Cancer is the second leading cause of mortality worldwide after cardiovascular diseases [1]. According to projections based on population aging and growth, the global burden of cancer is expected to increase by more than 60% in the next twenty years, raising from about 18.1 million new cases in 2018 to approximately 29.4 million cases in the year 2040. With this growing global burden, treatment of cancer is one of the most significant public health challenges of the 21st century [2]. Over the past few decades, significant advancements in DNA sequencing and a more comprehensive understanding of the cancer genome have revolutionised the cancer treatment landscape. The conventional paradigm, according to which chemotherapy drugs (primarily targeting rapidly proliferating cells) are selected based on the organ of origin, histology, and staging, has evolved towards the molecular profiling of the tumor, predominantly based on genomics. This shift guides the choice of therapeutic strategies, including chemotherapy, immunotherapy, and molecularly targeted agents [3]. Addi-

tionally, this paradigm shift has paved the way for the emergence of personalised or precision medicine approaches, which hold great promise as effective strategies for treating cancer. To develop innovative and precision therapeutic approaches, it is crucial to identify and target cancer driver genes [4], which contain mutations conferring a selective growth advantage to cells. Specifically, these genes can be classified as oncogenes or tumor suppressors based on the type and location of the genetic alterations they undergo.

1.2 Oncogenes and tumor suppressors

Oncogenes are genes that encode proteins driving the cell cycle forward, promoting uncontrolled cellular growth and division beyond the normal context of organism development. They derive from mutations occurring in their normally functioning counterparts known as proto-oncogenes. Proto-oncogenes are essential genes involved in cellular processes such as cell growth, differentiation, and signal transduction. Alterations in these genes that influence either the control of their behavior or the way that their encoded proteins are structured can lead to activation of oncogenes. When such oncogenes are formed, they go on to drive cell multiplication and assume a pivotal role in the pathogenesis of cancer [5]. Tumor suppressor genes encode for proteins that are typically involved in restraining uncontrolled cellular growth, facilitating DNA repair, activating cell cycle checkpoints and, when necessary, inducing cell death (apoptosis) [6]. Consequently, inactivation of tumor suppressor gene function due to mutations increases the selective growth advantage of the cell in which it resides, thus contributing to cancer.

In a broader sense, genes in which acquired mutations promote a selective growth advantage and thus are causally linked to cancer progression are termed cancer driver genes [4, 7]. Among well known cancer driver genes, the mutation patterns have been observed to be highly characteristic and nonrandom, thus providing a reference on which to functionally classify driver genes as oncogenes or tumor suppressors [4]. In particular, oncogenes exhibit recurrent mutations at specific amino acid positions while the occurrence of mutations in tumor suppressor genes is typically spread across their entire length [4]. Driver mutations occurring in either oncogenes or tumor suppressor genes lead to the stimulation of cell growth and division. However, in the case of oncogenes, this growth advantage to the tumor cell results from mutations that confer increased or new activity to the gene, leading to abnormal cellular processes or functions. These mutations are termed 'gain-of-function' mutations. On the other hand, mutations occurring in tumor suppressor genes disrupt their normal function, removing constraints on cell growth and division such as key checkpoints regulating cell proliferation, DNA repair, and cell cycle. These mutations are named "loss-of-function" mutations. Taking this into account, classification of cancer genes in terms of oncogenes or tumor suppressor genes can incorporate various criteria such as functional studies, somatic mutations, and copy number alterations [8–10]. In particular, the widely used "20/20 rule" combines gain-of-function and loss-of-function mutation occurrence to classify genes as oncogenes or tumor suppressors, as described in the article by Pavel *et al.,* 2016 [11] (section 4.2 in the Materials and Methods).

Overall, according to the comprehensive catalog of Somatic Mutations in Cancer (COSMIC) [12], which provides extensive information on genes with a causal impact on human cancer, there are currently 719 cancer genes

documented (COSMIC v.86, August 2018) [13]. This catalog includes details about gene contribution to disease causation, the specific types of mutations that lead to gene dysfunction in cancer, and types of cancer in which mutations have been observed at an increased frequency [14].

1.3 Challenges and perspectives about precision medicine in oncology

Despite the large number of identified cancer driver genes (i.e. over 700), approved treatments are only available for approximately 40 of them [15, 16]. Certain genes, such as the RAS family of proteins (KRAS, NRAS and HRAS, the most frequently mutated oncogenes in cancer [17]), the MYC proto-oncogene (MYC, a commonly amplified gene [18, 19]), and tumor protein 53 (TP53, the most frequently altered tumor suppressor gene in human cancer [20]), present significant challenges for pharmacological targeting and have been categorised as "undruggable" [16]. To successfully address these limitations, innovation and technological advancement are necessary [21]. The challenge of undruggability often arises from impractical localization (intracellular and often nuclear) and/or unfavorable structural features (such as lack of distinctive protein pockets into which small molecules can bind with high specificity and affinity) [22]. RNA-targeted approaches have the potential to broaden therapeutic possibilities of intervention, because cancer targets that are currently chemically intractable at the protein level can prove druggable at the messenger RNA (mRNA) level [23]. This is supported by the increasing approval of drugs utilising these technologies in various medical contexts [24].

Despite the availability of diverse, promising nucleic acid-based modal-

ities for therapeutic intervention, this work is focused on exploring the potential of antisense oligonucleotide (ASO)-mediated approaches to cancer treatment. This focus is motivated by the greater versatility [25] of ASOs compared, for instance, with small interfering RNAs (siRNAs), which have been also extensively studied in the literature [26, 27]. In fact, ASOs not only enable the inactivation of target genes but also offer the ability to modulate protein activity, representing a promising class of drugs for personalised medicine approaches [25].

1.4 Antisense oligonucleotides

ASOs are single-stranded analogues of nucleic acids able to modulate gene expression by selectively binding to target regions through Watson-Crick base pairing. The core of an ASO is formed by the sequence of nucleotide bases, with a range length of 12–25 nucleotides [28], each representing one of the four genetic building blocks: adenine (A), cytosine (C), guanine (G), and thymine (T). These molecules typically consist of a backbone composed of repeating sugar-phosphate units, but instead of the ribose sugar found in RNA, they use modified sugars for enhanced stability and resistance to cellular degradation. All oligonucleotides are negatively charged. Several ASO chemistries have been developed that possess distinct properties aimed at improving the stability, solubility, and cellular uptake of ASOs [29]. Phosphorothioate (PS) backbones, as well as 2′-O-methoxyethyl (2′-MOE) and 2′-O-methyl (2′-OMe) substituents, increase resistance to degradation and promote protein binding to target RNA. Locked nucleic acid (LNA) modification (Figure 1.1, top panel) markedly increases the binding of the oligonucleotide to the targeted mRNA. In phosphorodiamidate

morpholino oligomers (PMOs), ribose (in RNA) or deoxyribose (in DNA) is replaced with morpholine rings, and the phosphorothioate or phosphodiester (in RNA) groups are replaced with uncharged phosphorodiamidate groups, resulting in a compound that is neutral and very resistant to degradation (Figure 1.1, bottom panel). Positively charged piperazine residues in positively charged PMOs, or positively charged arginine-rich peptides in peptide-conjugated PMOs, dramatically improve the intracellular uptake of the oligomers [30]. Regarding modifications that are made to improve the delivery, stability and function of ASOs, these include chemical modifications that promote stability as well as bioconjugation to different moieties (peptides, antibodies, aptamers, lipids and sugars) and loading into delivery vehicles (DNA nanostructures, exosomes, spherical nucleic acids, lipoplexes and liposomes, stimuli-responsive nanotechnology) to promote both cellular uptake and targeting to specific cells or tissues [26].

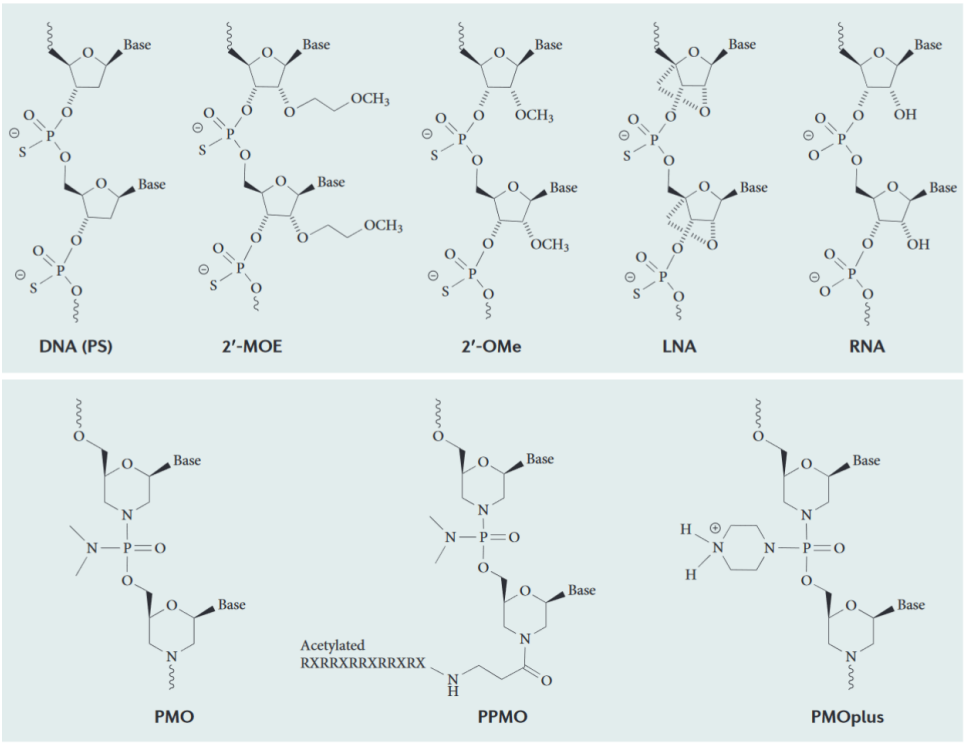

Figure 1.1: Oligonucleotide chemistries. Figure taken from Kole *et al.*, 2012. [30]

Depending on the target RNA molecule, ASOs can be designed to complementarily bind to specific regions, such as exons or splice sites, within the target RNA sequence.

1.5 Antisense oligonucleotides and modulation of RNA splicing

ASOs have the ability to interfere with RNA splicing, a crucial step in gene-expression regulation that removes intronic sequences and joins exonic regions to produce mature RNA molecules. This makes it possible to generate multiple transcript isoforms from a single gene and thus generate a wide range of protein isoforms with different functions and properties, depending on the cellular context. Specifically, RNA splicing is catalysed by

the spliceosome, a multimegadalton ribonucleoprotein complex composed of multiple small nuclear RNAs (snRNAs) and many associated protein factors [31]. The spliceosome is recruited through consensus sequence elements at the 5' and 3' splice sites (*donor* and *acceptor* sites, respectively) and branch-point sequences, and its action is further modulated by an array of cis-acting exonic and intronic splicing enhancers (ESEs and ISEs), and exonic and intronic splicing silencers (ESSs and ISSs), which are recognised by auxiliary splicing factors, including the Ser/Arg-rich (SR) proteins and heterogeneous nuclear ribonucleoproteins (hnRNPs) [32]. It is precisely by targeting these specific splice sites or splicing regulatory elements that ASOs can induce exon skipping, thereby modulating the splicing outcome [33]. The impact of ASOs on RNA splicing has specific implications in the case of mature transcripts that undergo translation into proteins.

1.6 Exon skipping

Exon skipping (ES) is a naturally occurring cellular process involved in precursor messenger RNA (pre-mRNA) splicing, contributing to the diversification of the proteome in eukaryotic organisms. During this process, specific exons within the pre-mRNA are excluded or "skipped" from the final mRNA transcript, resulting in the production of alternative splice variants with distinct protein-coding sequences (Figure 1.2). This phenomenon plays a vital role in regulating gene expression and functionality, allowing cells to generate different protein isoforms from a single gene, thus expanding their functional repertoire.

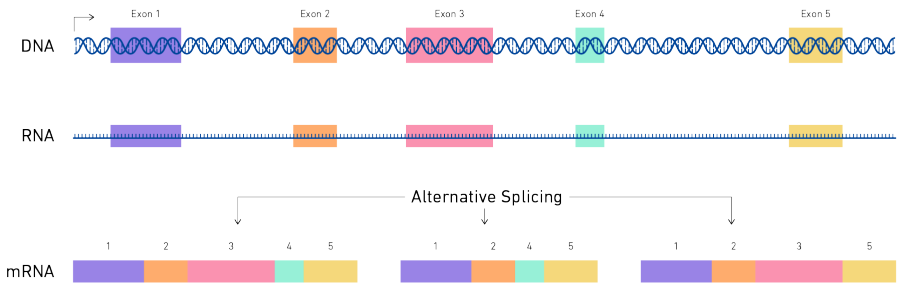

Figure 1.2: An overview of exon splicing. This process can produce a diversity of alternative transcript forms from a single gene by selective inclusion/exclusion of exons. Figure taken from Technology Networks, Jonathan Dornell, PhD (2021).

While exon skipping is a natural mechanism, researchers have harnessed this process for therapeutic purposes. By employing *ad hoc* ASOs, scientists can manipulate the splicing process to selectively skip or include certain exons, thereby modifying the transcript product and, consequently, the final protein product. Indeed, by eliminating an exon and therefore a certain number of nucleotides, the reading of the corresponding codons may or may not remain unaltered. The coding region of the mature transcript is read by the ribosome in consecutive triplets of nucleotides, which is known as the "reading frame". Each triplet, or codon, specifies an amino acid to be incorporated into the nascent polypeptide chain, thereby determining the primary sequence of the protein. The precise arrangement of amino acids subsequently dictates the three-dimensional folding of the protein, leading to its functional conformation [34]. Specifically, an ES event can lead to

either an in-frame or an out-of-frame transcript depending on the length of the exon that is skipped, with a potentially significant impact on the protein product and its biological activity [35]. In particular, if the skipped exon has a number of nucleotides that is a multiple of 3, the reading frame is maintained, and ES results in the production of a shortened and in-frame transcript. If the skipped exon has a number of nucleotides that is not a multiple of 3, the reading frame is lost and we refer to the resulting transcript as out-of-frame. In this last case, the loss of the original reading frame may lead to the generation of a premature termination codon (PTC). If the PTC is located at a position sensitive to Nonsense Mediated Decay (NMD), the transcript may be degraded through this mechanism, resulting in a loss of protein expression. On the other hand, an in-frame transcript does not affect the reading frame but generates a shorter protein (compared with the original one), which might still retain part of its function. Therefore, the effect of altered skipping on protein synthesis and function depends on the location and size of the skipped exon, as well as on the specific protein and its biological role.

1.7 Therapeutic exon skipping: the case of the Duchenne Muscular Dystrophy

In the last decade, ASOs have shown promising results when used to induce targeted ES for the treatment of certain genetic diseases, such as Duchenne muscular dystrophy (DMD; MIM #310200) [36, 37]. In fact, the FDA has authorised the use of four drugs based on this antisense approach to induce ES for therapeutic purposes in DMD [38–41]. DMD is a debilitating and progressive neuromuscular disorder that results from mutations in a sin-

gle gene, the dystrophin gene [42], encoded by a vast locus spanning over 2 million bases on chromosome X and encompassing 79 exons [43, 44]. Among the extensive number of annotated mutations (>7000) observed in DMD patients, the majority (∼80%) are large mutations, including deletions of one or more exons (68%) and large duplications (12%), while the remaining ones (∼20%) concerns small mutations, such as small in-dels and point mutations [43]. Over 90% of these mutations, which tend to cluster in the region spanning exons 45-53 [45], cause a disruption of the translational reading frame [46], ultimately leading to the complete absence of the dystrophin protein, which plays a crucial role in proper muscle function [43, 45]. A milder form of muscular dystrophy that is also linked to mutations in the dystrophin gene is known as Becker muscular dystrophy (BMD; MIM #300376). Of note, mutations found in BMD patients maintain the translational reading frame and result in a shorter yet partially functional protein [47]. In DMD, the therapeutic intervention aims to restore the dystrophin reading frame, thereby reinstating at least partial expression of dystrophin in DMD-affected muscles and consequently reducing disease severity, similar to what happens in BMD. This is achieved with the design of tailored ASOs, which selectively induce exon exclusion, resulting in the restoration of a correct reading frame [48, 49]. Consequently, a shorter yet partially functional protein is produced [49], leading to improved muscle strength and function in affected individuals. The top 5 mutated exons found in DMD patients were exon 51 (14% of total mutations/21% of deletions), exon 53 (10%/15%), exon 45 (9%/13%), exon 44 (7%/11%), and exon 43 (7%/11%) [43]. Accordingly, the FDA has authorised the use of the following four drugs which are all based on antisense approach to induce ES of the indicated exons for therapeutic purposes in DMD: Eteplirsen (exon 51),

Golodirsen and Viltolarsen (exon 53), Casimersen (exon 45) [38–41].

Figure 1.3 depicts the therapeutic intervention acted by Eteplirsen, one of the approved DMD ASO drugs. In particular, genomic deletion of exon 50 leads to an out-of-frame mRNA generating a premature termination codon. This results in the synthesis of a truncated non-functional dystrophin (left panel). Eteplirsen specifically recognises sequences of exon 51 of the DMD gene, allowing its exclusion from the mature mRNA. This restores the open reading frame, promoting the synthesis of an internally deleted but partially functional dystrophin (Figure 1.3, right panel).

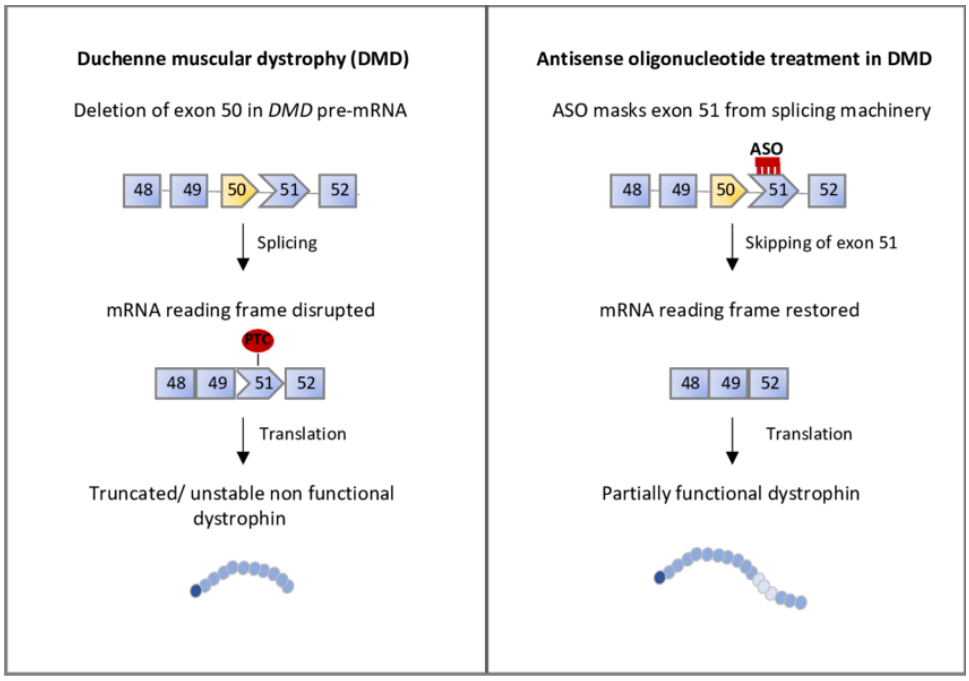

Figure 1.3: Antisense-oligonucleotide treatment for Duchenne muscular dystrophy (DMD). Patients with DMD display mutations which disrupt the open reading frame of the dystrophin pre-mRNA. Schematic representation of DMD prem-RNA from exon 48 to 52 is shown. Genomic deletion of exon 50 leads to an out-of-frame mRNA generating a premature termination codon (PTC). This results in the synthesis of a truncated non-functional dystrophin (left panel). Eteplirsen, which is an FDA-approved ASO drug, specifically recognises sequences of exon 51 of the DMD gene, and promotes its exclusion from the mature mRNA. This restores the open reading frame, resulting in the synthesis of an internally deleted, but partially functional, dystrophin (right panel). Figure taken from Verdile *et al.*, 2021 [50]. **Abbreviations**: **PTC** = Premature Termination Codon

1.8 Exon skipping in cancer

In recent years, some attempts to use ES as a therapeutic approach have also been made in cancer. For instance, ASOs have been designed to induce skipping of a specific exon (i.e., exon 4) of the ETS-Related Gene (ERG), which is an oncogene, in prostate cancer cells [51]. This generates an out-of-frame transcript with significant reduction of ERG protein levels, leading in turn

to reduced cell proliferation, increased cell death, and reduced cell migration [52]. In a more recent study involving the two isoforms of the Pyruvate Kinase M1/2 (PKM) gene, ASOs were used in one case (PKM1 isoform, a tumor suppressor) to restore its expression, and in the other (PKM2 isoform, an oncogene) to decrease it, resulting in both cases in reduced tumor growth [53]. In another study, a gene that is often improperly over-expressed in leukemia and solid tumors, the Wilms tumor 1 (WT1) gene, was considered for ES approach [54]. Specifically, ASO-mediated skipping of exon 5 resulted in decreased cell viability and survival in leukemia cell cultures [55].

To advance and expand these efforts, especially when targeting undruggable cancer-related genes, it is crucial to address the challenges associated with ASO design comprehensively.

1.9 Overview of relevant computational resources

Designing effective therapeutic ASO sequences demands careful consideration of numerous criteria. To assist in this process, some computational tools have been developed to evaluate essential nucleic-acid properties and facilitate ASO design. These tools include various functionalities, including estimation of self-complementarity and tendency to form intra-molecular hairpins, as well as the calculation or estimation of molecular weight, solution concentration, melting temperature, and absorbance coefficients [56]. Concerning the study of RNA molecules in particular, software packages such as ViennaRNA Package 2.0 [57] and RNAstructure [58] offer tools for predicting secondary structures and RNA–RNA interactions. Additionally, databases such as SpliceAid 2 [59] provide information on RNA target mo-

tifs that are bound by splicing proteins in humans. Among others, a web application and a database, namely the PFizer RNAi Enumeration and Design (PFRED) tool [60] and the eSkip-Finder database [61], have been recently developed for the design, analysis, and visualisation of antisense oligonucleotides and offer comprehensive features to aid researchers in their ASO design efforts.

Chapter 2

Aim of the work

The present thesis work focuses on the development of a computational procedure to streamline the entire process of designing *ad hoc* antisense oligonucleotides (ASOs) to induce targeted exon-skipping (ES) events in cancer genes. This procedure encompasses all the process steps, from *in silico* identifying potential target candidates to generating a list of ASO sequences tailored to produce the intended protein variants. Of note, it takes into account cancer-relevant features, such as mutation frequency in patients, which holds special value for the design of ES-based therapeutic strategies in oncology. Our approach implements state-of-the-art rules to identify the most promising candidate exons of a gene of interest, which is first classified *in silico* as either an oncogene or a tumor suppressor. Subsequently, the computational procedure designs specific ASOs, in accordance with guidelines for morpholinos, tailored for selected ES events. These events can lead to desired outcomes, such as generating oncogene variants that lack activity or tumor suppressor variants that could exhibit partially restored functionality. As a proof-of-concept, the procedure is applied to

the top 10% most mutated genes in cancer, ranking them based on their suitability for ES-targeted interventions, and further in-depth investigations are carried out on clinically relevant oncogenes and tumor suppressors as case studies. In conclusion, our computational procedure aims to provide valuable assistance to researchers in their efforts to develop innovative therapeutic interventions based on targeted ES.

Chapter 3

Results

3.1 An integrated computational procedure to support the design of exon-skipping-based therapeutic strategies in cancer

We developed a computational procedure that, given a gene of interest, initially classifies it as either an oncogene or a tumor suppressor based on annotated cancer mutations and, once all its known transcripts are collected, the pipeline proceeds to select potential target exons for exon skipping (ES), and return the expected protein variants as outcome. The pipeline predicts whether the skipping of these exons results in shortened transcripts that remain in frame or shift out of frame. To further investigate the shortened products, those labeled as out-of-frame are subjected to the prediction of their potential degradation through Nonsense Mediated Decay (NMD), while all transcript sequences are translated *in silico* into amino acid sequences. Finally, the procedure allows for the design and evaluation of candidate ASO sequences to induce the skipping of specific exons that are favorable for achieving the intended goals, such as producing inactive oncogene variants or restoring at least partially functional tumor suppressor vari-

ants. The overall procedure is illustrated in Figure 3.1 and an overview of each individual sub-task is briefly presented in the next subsections.

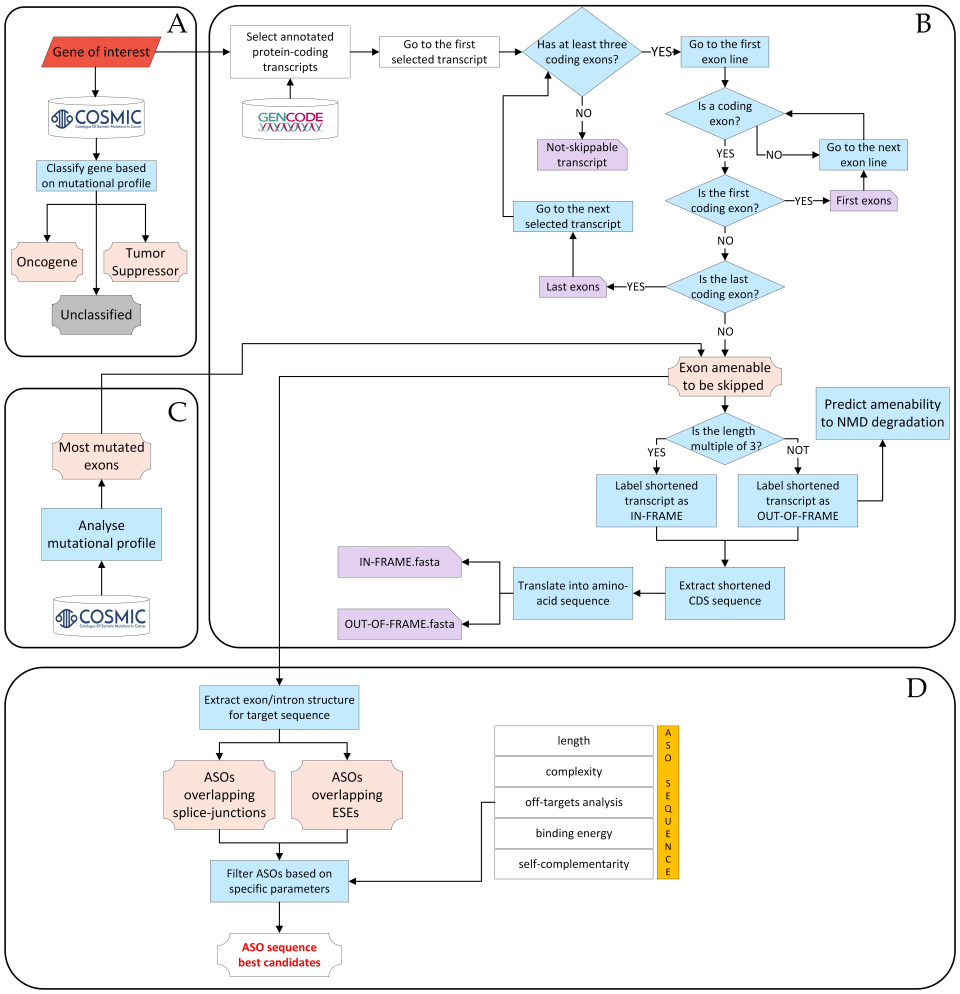

Figure 3.1: Computational procedure flowchart. The procedure begins with a gene of interest and proceeds as follows: (A) The gene is classified as either an oncogene or a tumor suppressor by analysing its mutational profile obtained from the COSMIC database. (B) All annotated transcripts of the gene are collected from the GENCODE database. Candidate exons that could undergo skipping are selected, and the corresponding ES events are classified as either in-frame or out-of-frame. For out-of-frame transcripts, the potential degradation by NMD is predicted. (C) Among exons that could undergo skipping, the procedure identifies exons that are more frequently mutated in cancer patients based on data from the COSMIC database. (D) Candidate ASOs are designed and evaluated. These sequences are differentiated based on whether they overlap splice junctions or bind to ESEs within the exons. **Abbreviations**: **ASO** = Antisense Oligonucleotide; **CDS** = Coding DNA Sequence; **ESE** = Exonic Splicing Enhancer; **NMD** = Nonsense Mediated Decay

3.1.1 Classification of a gene of interest as an oncogene or tumor suppressor

Based on COSMIC data [14] regarding annotated mutations, the computational procedure enables the *in-silico* classification of a gene of interest as either an oncogene or a tumor suppressor. This classification is based on the principles of the "20/20 rule" proposed by Pavel *et al.,* 2016 [11]. According to this rule, oncogenes tend to have more than 20% of annotated mutations occurring as gain-of-function mutations at recurrent positions, while tumor suppressors tend to have over 20% of loss-of-function mutations spread throughout the sequence (section 4.2 in the Materials and Methods). Considering this evidence, the procedure automatically assesses the likelihood of a gene belonging to either category. In particular, the classifier attempts to assign the appropriate label, either "oncogene" or "tumor suppressor" based on the evaluation of the mutational profile (Figure 3.1A) and the reference thresholds for gain-of-funtion and loss-of-function mutations (Table 4.1) [11], while leaving genes with inconclusive mutational data as “unclassified” (Figure 3.1A).

3.1.2 Identification of candidate exons to be targeted for skipping

Given a gene of interest, the procedure performs a comprehensive screening of all isoforms annotated by GENCODE [62] to identify exons amenable to be skipped (Figure 3.1B). Only protein-coding transcripts are considered for further investigations. Of all exons in the selected transcripts, those containing the start and stop codons are excluded from the procedure, as well as

any non-coding exons that may occur before and/or after them. This exclusion is made to prevent interference with the essential regions responsible for initiating and terminating protein translation [63, 64]. To identify candidate exons to induce therapeutic targeted ES, the procedure proceeds to determine the specific outcome by considering the following two scenarios, which may potentially result from ES: (1) in-frame transcript; (2) out-of-frame transcript. Each shortened transcript is then translated into the corresponding amino acid sequence. Since the skipping of an exon can result in the formation of a premature termination codon (PTC), potentially sometimes leading to the degradation of the resulting mRNA through NMD [65], the procedure assesses the likelihood of NMD occurrence for all the out-of-frame transcripts. This prediction is based on the most relevant rules associated with NMD evasion, namely, the so-called "50-55nt rule", the 'last-exon rule', and the 'start-proximal rule' (section 4.4 in the Materials and Methods). These rules evaluate the position of the PTC formed after ES and the expected outcome in terms of NMD occurrence. This step in the procedure (Figure 3.1B) yields at list of all exons deemed potential targets for ES, along with the corresponding shortened transcript and amino acid sequences obtained as a result of their exclusion. Additionally, it provides the prediction of NMD events for the subset of out-of-frame transcripts.

3.1.3 Prioritising candidate exonic targets for exon skipping via cancer mutation profiling

Aimed at selecting the best candidate exon targets for the design of ES-based therapeutic strategies, mutation profile in cancer is then evaluated based on COSMIC data [14]. The implemented procedure identifies the

exons marked by the highest mutational burden, specifically those exhibiting mutations in the largest number of cancer patients. In particular, taking into account only point mutations, the procedure automatically and carefully maps the annotated genomic mutations to the corresponding exon regions [66] and calculates the absolute frequency of mutations within each exon of the gene of interest. This is done by counting the number of patients who have at least one mutation in a specific exon. Finally, exons are ranked by the decreasing number of absolute mutation frequency, and the top 10 ranking exons are reported to proceed to ASO design (Figure 3.1C). This step of the computational procedure aims to prioritise ASO design on exons enriched in clinically relevant mutations, while limiting the list of candidate exons to a manageable number.

3.1.4 Design and evaluation of *ad hoc* antisense oligonucleotides to induce desired exon-skipping products

This step of the computational process integrates the sequences of exons susceptible to ES and those that are highly mutated, generating 25-nucleotide-long ASO sequences, which is the recommended length for morpholino oligos [63, 64], designed to induce ES. These ASOs are designed to interact, based on the principle of base complementarity, with two types of splicing regulatory sites: splice junction sites (*donor* and/or *acceptor* splice sites) and intra-exon splice sites (Figure 3.2). For each splice junction, we design a set of 7 ASOs (14 for each exon) to target the selected transcript at the overlap of the intron-exon or exon-intron border, a strategy shown to maximise the efficiency of mRNA splicing alteration using morpholinos [63]. Regarding ASO sequences designed to bind within exon regions,

only those fully overlapping at least one exonic splicing enhancer (ESE) domain are selected for further consideration. The number of ASO sequences drawn within each exon varies depending on the exon length and the number of annotated ESE regions therein. Once designed, ASO sequences finally undergo evaluation based on specificity and reference physicochemical parameters, including CG percentage, G percentage, presence of tetra G, and self-complementarity. Only ASO sequences that meet the optimal threshold values for each parameter (summarised in Table C1), as documented in the relevant literature [64], and have a unique match in the genome, are selected and included in the output (Figure 3.1D) (section 4.6 in Materials and Methods). As a point of reference for evaluating the results of our procedure in generating candidate ASOs, we compared them with four available FDA-approved ASO drugs for Duchenne muscular dystrophy (DMD) treatment [38–41]. Indeed, our procedure successfully identified ASO sequences either identical or closely matched to all drugs, as shown in Table B1.

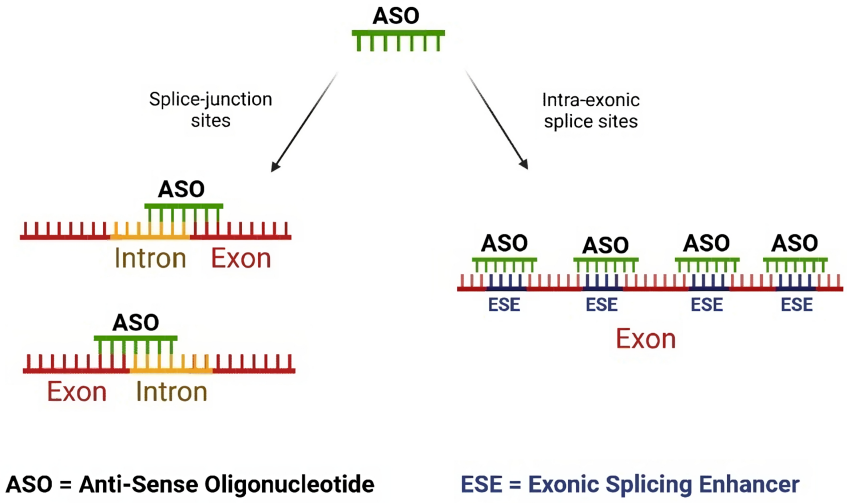

Figure 3.2: Schematic representation of the ASO design approaches. The procedure involves designing ASO sequences for a target exon using two distinct approaches. The first approach (indicated by an arrow pointing to the left in the figure) targets the splice-junction sites, including *donor* and/or *acceptor* splice sites. The second approach (indicated by an arrow pointing to the right in the figure) focuses on the ESE regions within the exon. Figure created with BioRender.com.

3.2 Application of the computational procedure to highly mutated cancer genes

To demonstrate the relevance of the entire computational procedure described previously, we applied it to a clinically relevant set of genes. Specifically, we selected the top 10% most frequently mutated genes in cancer patients based on data stored in the COSMIC database (v.96) [14]. A total of 72 genes were identified, as shown in Figure 3.3. The following subsections provide insights into the output data and biological knowledge gained by applying the procedural steps outlined in Figure 3.1 to this cancer-relevant

dataset.

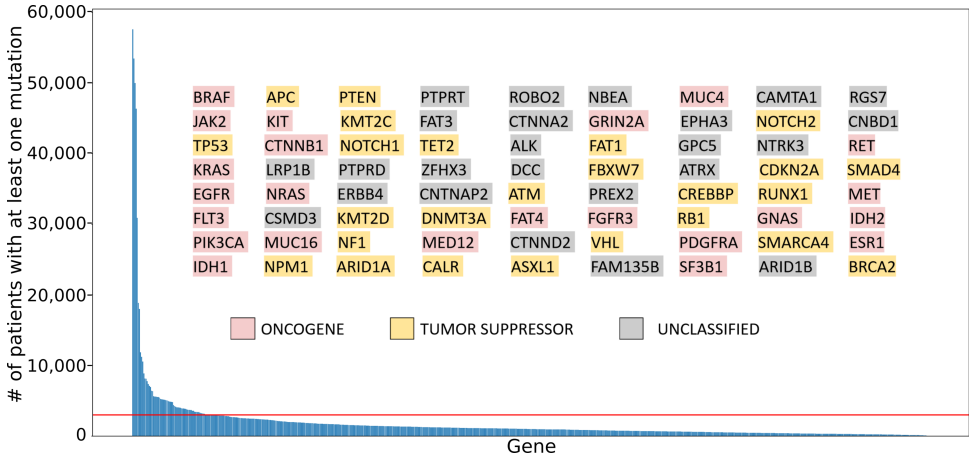

Figure 3.3: Most frequently mutated genes in cancer patients. Genes from the COSMIC cancer gene census are ranked based on the number of cancer patients with at least one mutation (Y-axis) in the indicated gene (X-axis). The horizontal red line indicates the threshold applied for selecting the top 10% most frequently mutated genes. The inset shows the selected genes (N=72), highlighted and listed in columns, arranged in descending order of mutation frequency from top to bottom and from left to right. Gene names are highlighted in the figure to depict computationally-predicted cancer gene roles, as indicated in the color legend. Prediction flow is described in section 4.2.

3.2.1 Classification of the role of selected genes in cancer

Applying the previously mentioned 20/20 rule [4, 11] (section 4.2 in the Materials and Methods), the 72 selected genes were classified according to their predicted role in cancer, as follows: 25 tumor suppressors, 23 oncogenes, and 24 unclassified genes due to inconclusive scores (Figure 3.3, inset). To validate our classification, available for 48 genes, the results were compared with two publicly available compendia of oncogenes and tumor suppressors: 1) Futreal *et al.,* 2004 [67], hereafter *MSigDB*; 2) Tokheim *et al.,* 2016 [68], hereafter *Tokheim*. Overall, 35 out of 48 were also evaluated by *MSigDB*, while *Tokheim* assessed 37 of 48, with classification concor-

dance rates of 86% and 97%, respectively. For detailed information, please refer to Appendix A and Figure A1[A-B].

3.2.2 Identification of candidate exons for targeted skipping in selected cancer genes

Analysis of the selected cancer genes (N=72) revealed an average of 10 protein-coding transcripts per gene according to annotations available in the GENCODE database (Release 43). Among these transcripts, the majority (9 out of 10) harboured at least one exon predicted to be susceptible to skipping, according to the computational procedure (Figure 3.1B). Overall, these cancer genes had an average number of 26 exons per gene as possible targets. The *in-silico* simulation of ES events demonstrated that the vast majority of these genes (68 out of 72; 94%) could yield both in-frame and out-of-frame shortened products, supporting the potential for designing therapeutic ES strategies for both inactivation and functional rescue purposes. Furthermore, for the 23 genes previously classified as oncogenes in our gene compendium (Figure 3.3), where functional inactivation is the therapeutic goal, the prediction of NMD occurrence was performed for all the out-of-frame shortened transcripts generated through *in-silico* ES. The results indicated that the majority of these genes (21/23) produced at least one out-of-frame transcript that was expected to undergo NMD.

In summary, the set of cancer genes analysed exhibited a diverse range of transcript isoforms, many of which contained exons susceptible to skipping. Importantly, a significant portion of these genes had the potential to generate both in-frame and out-of-frame transcripts when subjected to *in-silico* ES simulations. Additionally, a considerable number of oncogenes in

this dataset have the capacity to generate out-of-frame transcripts that are predicted to undergo degradation through NMD.

3.2.3 Prioritising candidate exon targets of selected cancer genes using mutation data

In the field of precision oncology, it can be valuable to devise therapeutic strategies targeting exons with the highest incidence of recurrent mutations. Therefore, our methodology involves prioritising exons within the selected cancer genes (Figure 3.3), according to this criterion. Specifically, we focused on the exons deemed potential targets for ES (from the previous section, 26 exons on average) in each cancer gene and with at least one mutation observed across the largest number of cancer patients. By analysing the mutation profiles using data from the COSMIC database (Figure 3.1C), we found that these exons showed an average of about 300 point mutations each. Subsequently, the analysed exons were ranked by frequency of mutations, and the top 10 scoring exons per gene in this list were selected as the best candidates for ASO design.

3.2.4 Design and evaluation of the best candidate antisense oligonucleotide sequences

The final step in our computational procedure involves the design of ASO sequences that hold promise for inducing desired protein variants (Figure 3.1D). In particular, to manage computational burden and prioritise experimental feasibility, we focused on designing and providing as output ASO sequences for a selected list of exons for each gene of interest. This list included the

highest-ranking exons that were evaluated as potential targets for exon skipping and were recurrently mutated in cancer patients, limited to a maximum of 10 exons per gene. Across 72 chosen cancer genes (Figure 3.3, inset), a total of 674 protein-coding exons were considered and customised ASOs were generated based on two design approaches (Figure 3.2). The first set target splicing regulatory sites located at the splice junctions, with 14 junction-exon ASOs designed per exon. The second set of ASOs focus on regulatory sites located within the exon itself. Here, an average of approximately 306 intra-exon ASOs were designed per exon. Next, both sets of ASO sequences underwent an evaluation process based on state-of-the-art knowledge regarding optimal values for a series of physicochemical parameters (Table C1). The ASO design and filtering step yielded an average of 6 ASOs at the splice junctions and 150 ASOs at the ESE sites, representing the best candidate sequences.

3.3 Proof-of-concept case studies

3.3.1 Detecting high-potential cancer genes for effective targeted exon-skipping intervention

We prioritised genes based on the availability of exons that were the most suitable for achieving effective targeted interventions. To accomplish this, we assessed the following conditions for each exon in our dataset: a) Inclusion of the exon among the ten most frequently mutated exons in cancer patients for the corresponding gene; b) Desired reading frame for the shortened transcript resulting from ES, taking into account the predicted role of the gene in cancer (oncogene or tumor suppressor) and the intended

therapeutic objective (inactivation or functional rescuing, respectively); c) Presence of at least one ASO for each design strategy (targeting splice sites within the exon or targeting splice junctions) that adheres to the recommended physicochemical parameters for effective targeted splicing. Subsequently, we ranked the cancer genes in our compendium based on the decreasing percentage of their exons that met all three criteria mentioned above (Figure 3.4). In what follows, we provide more detailed information on two exemplary cases, namely, the first oncogene (NRAS) and the first tumor suppressor (VHL), featured in this ranking.

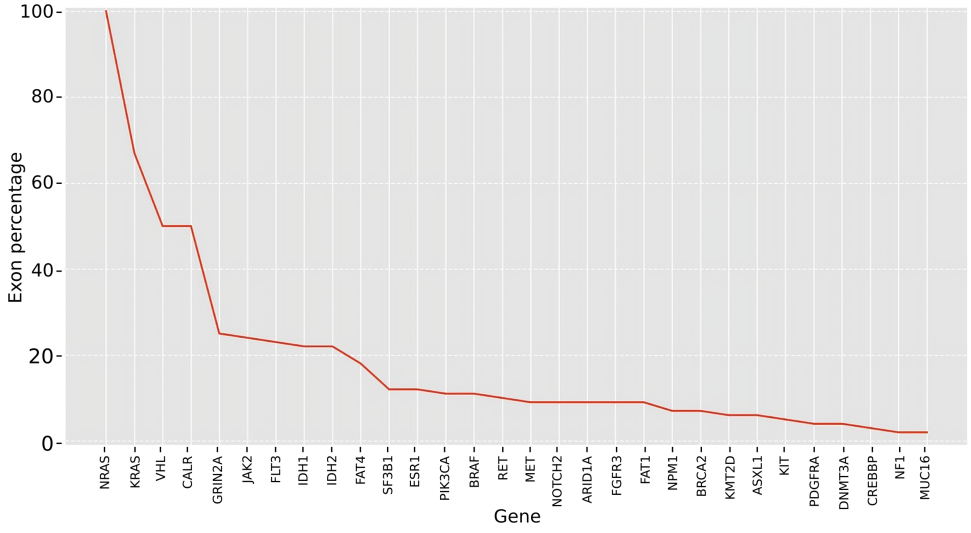

Figure 3.4: Ranking of genes based on the percentage of exons that meet all the criteria considered favorable for effective ES-based intervention. These criteria include the following: (1) Inclusion of the exon among the ten most frequently mutated exons for the corresponding gene. (2) Ensuring the correct frame is obtained following ES, depending on the predicted cancer role (specifically, an out-of-frame outcome for oncogenes and an in-frame outcome for tumor suppressors). (3) Existence of at least one antisense oligonucleotide available for both design strategies, which involve targeting splice sites within the exon or at junctions, while adhering to recommended physicochemical values.

The NRAS proto-oncogene GTPase (NRAS) has one annotated protein-

coding transcript (ENST00000369535.5) which consist of seven exons. Given its oncogenic nature, the therapeutic objective of an ES-based approach would involve promoting protein degradation in order to reduce its gain-of-function effect. To evaluate NRAS suitability as a target for this type of intervention, particularly within the context of cancer, we conducted a preliminary investigation of its expression in cancer tissues, utilising publicly available RNA-sequencing experiments [24, 69]. These data were retrieved from public genomics resources, finally collecting 1,359 tumor samples from the Pan-Cancer Analysis of Whole Genomes (PCAWG) atlas [70], along with 2,231 corresponding normal-tissue samples from the Genotype-Tissue Expression (GTEx) [71] project. The analysis was specifically focused on transcript-level expression values, which is crucial due to the well-documented evidence of different splicing isoforms exhibiting different behavior in the cancer context [72–74]. Results concerning the protein-coding transcript (ENST00000369535) annotated for the NRAS gene revealed a statistically significant upregulation (p-value = 1,2 e-243, Wilcoxon rank-sum test) in its expression level, measured in transcripts per million (TPM), in cancer tissues (TPM mean value = 29,2) compared to normal ones (TPM mean value = 12,7) (Figure 3.5). Overall, the behavior of the identified NRAS transcript appears to align with its potential suitability as a tumor target for suppression. For a more comprehensive, tissue-specific examination, please refer to Table D1 and Figure D2 in Appendix D.

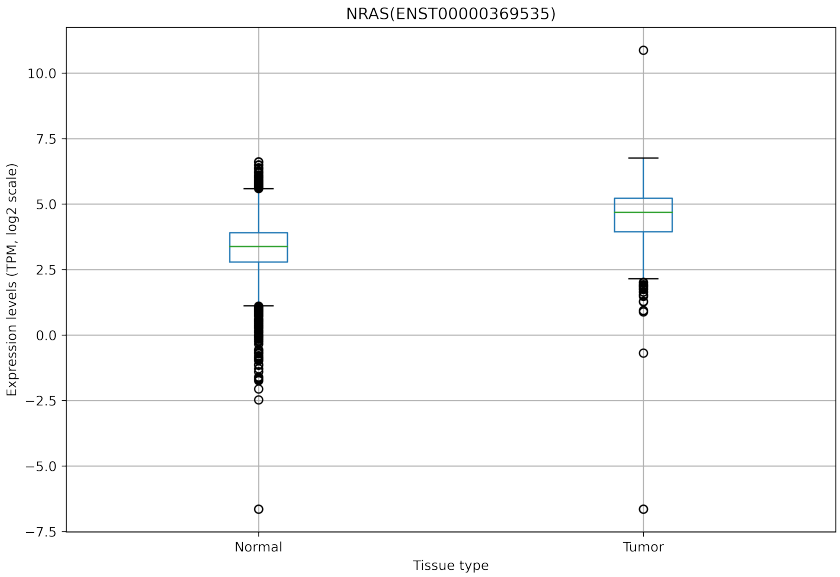

Figure 3.5: NRAS expression in cancer and normal tissues. Boxplots show expression values of the NRAS gene across 1,359 tumor samples from the PCAWG resource ("Tumor") and 2,231 normal samples from the GTEx archive ("Normal") for the protein-coding transcript annotated for this gene: ENST00000369535. Expression values are given in log2-transformed TPM counts. **Abbreviations**: **GTEx** = Genotype-Tissue Expression; **PCAWG** = Pan-Cancer Analysis of Whole Genomes; **TPM** = Transcripts per Million

From our computational analysis, two exons of the NRAS transcript ENST00000369535.5, namely, exon 3 and exon 4, emerged as potential targets for ES. Skipping either exon would result in out-of-frame transcripts. Although our computational procedure predicts that likely neither of them would undergo degradation via NMD, both transcripts would lead to significantly altered and shorter protein sequences, thus likely compromising their functionality. In terms of mutation profile analysis, both exon 3 and exon 4 rank among the top ten most mutated exons in cancer patients. Particularly, one of these exons (Ensembl ID: ENSE00001751295.1, exon 3)

(Figure 3.6) accumulates mutations in 4,052 patients, which accounts for over 60% of the total number of cancer patients with at least one annotated mutation in NRAS. Consistent with its role as an oncogene, the distribution of mutations along the gene sequence is concentrated at specific positions (Figure 3.6B) [4, 11]. Our computational procedure designed a total of 103 ASOs for NRAS exon 3. Among these ASOs, 89 were designed to target regulatory splice sites within the exon, while 14 ASOs were designed to target splice sites at the junctions with flanking introns. After filtering based on both specificity and physicochemical requirements, the procedure selected a subset of 58 ASOs, which included 52 ASOs targeting internal ESEs and 6 ASOs targeting splice sites at the junctions, with the latter being available only to target the upstream intron-exon junction (Table 3.1, Figure 3.6).

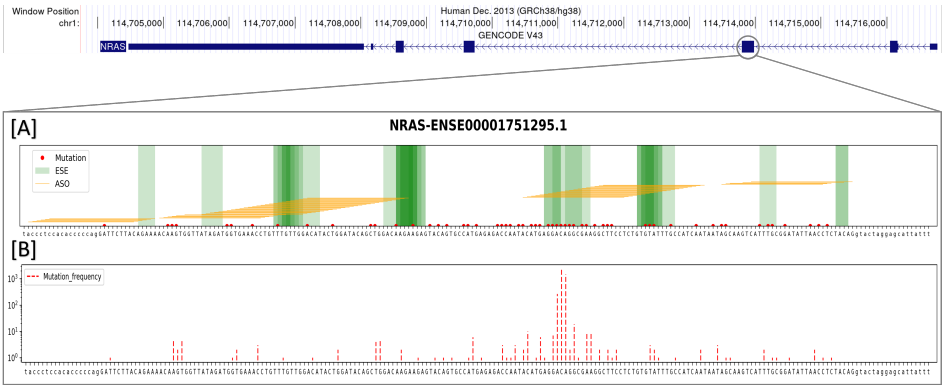

Figure 3.6: Example of a suitable candidate for designing an ES-based therapeutic approach targeting the NRAS oncogene. The figure summarises pertinent genomic (top panel), sequence (panel A), and mutation (panel B) information concerning design of an ES-based therapeutic approach targeting a specific exon (Ensembl ID: ENSE00001751295.1) of the NRAS oncogene that our computational procedure identified as a highly suitable candidate. Details are as follows: (A) Visualisation of 25-nt ASO sequences (represented by orange horizontal lines) designed to target ESE sequences (depicted by green vertical bands) or splice junctions. The objective is to induce the skipping of this particular exon in the mature transcript. The presence of mutations, as annotated in the COSMIC database for cancer patients, is indicated by red dots along the DNA sequence. The uppercase letters represent the exon, while the lowercase letters denote the flanking introns. (B) Mutation occurrences within the NRAS exon sequence. This figure panel displays red-dotted vertical bars, indicating the number of patients (on the y-axis, using a logarithmic scale) with mutations at the indicated nucleotides. **Abbreviations**: **ESE** = Exonic Splicing Enhancer; **ASO** = Antisense Oligonucleotide

Table 3.1: The best candidate ASO sequences designed at the exon 3 splice junctions of the NRAS oncogene. The table provides the ASOs that were specifically designed at the junctions of the selected exon of NRAS (Ensembl stable ID: ENSE00001751295.1, third exon in the NRAS transcript). The listed ASOs, along with their corresponding sequences indicated in the table, have successfully passed both specificity and physicochemical filters. The nomenclature used for the ASOs follows the conventions described by Mann *et al.,* 2002 [75].

Gene	Exon ID	ASO ID	ASO Sequence
NRAS	ENSE00001751295.1	H3D(13, -12)	UGCUCCUAGUACCUGUAGAGGUUAA
		H3D(12, -13)	AUGCUCCUAGUACCUGUAGAGGUUA
		H3D(11, -14)	AAUGCUCCUAGUACCUGUAGAGGUU
		H3D(10, -15)	UAAUGCUCCUAGUACCUGUAGAGGU
		H3D(9, -16)	AUAAUGCUCCUAGUACCUGUAGAGG
		H3D(8, -17)	AAUAAUGCUCCUAGUACCUGUAGAG

The tumor suppressor gene Von Hippel-Lindau (VHL) has a total of six annotated transcripts, with four of them encoding proteins. Using our computational procedure, we identified two specific exons as prime candidates for developing ES-based therapeutic approaches aimed at partially restoring their function by selecting in-frame transcripts. Skipping either of these two exons would result in shortened transcripts that maintain the correct reading frame. Among these exons, one of them (Ensembl stable ID: ENSE00003504189.1, exon 2) (Figure 3.7) exhibited a higher frequency of mutations in cancer patients, with 447 patients having at least one mutation in this exon, accounting for approximately 24% of the total mutations observed. To target the regulatory splice sites of this exon (Ensembl stable ID: ENSE00003504189.1, exon 2), our computational procedure designed a total of 113 ASOs. Among these, 99 ASOs were designed within the exon itself, while 14 ASOs targeted the junctions with adjacent introns. After filtering based on both specificity and physicochemical criteria, a subset of

61 ASOs met the requirements, including 47 ASOs targeting internal ESEs and all 14 ASOs designed at the junctions (Table 3.2, Figure 3.7).

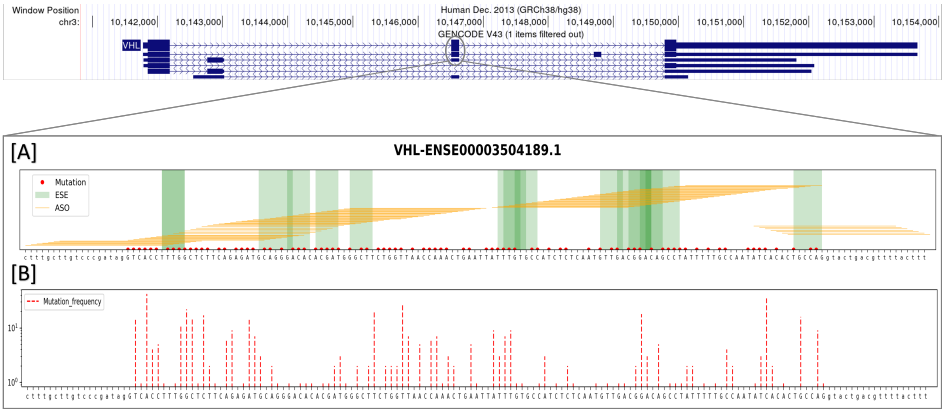

Figure 3.7: Example of a suitable candidate for designing an ES-based therapeutic approach targeting the VHL tumor suppressor. The figure summarises pertinent genomic (top panel), sequence (panel A), and mutational (panel B) information concerning design of an ES-based therapeutic approach targeting a specific exon (Ensembl ID: ENSE00003504189.1) of the VHL tumor suppressor that our computational procedure identified as a highly suitable candidate. Details are as follows: (A) Visualisation of 25-nt ASO sequences (represented by orange horizontal lines) designed to target ESE sequences (depicted by green vertical bands) or splice junctions. The objective is to induce the skipping of this particular exon in the mature transcript. The presence of mutations, as annotated in the COSMIC database for cancer patients, is indicated by red dots along the DNA sequence. The uppercase letters represent the exon, while the lowercase letters denote the flanking introns. (B) Mutation frequencies within the VHL exon sequence. The figure panel displays red-dotted vertical bars, indicating the number of patients (on the y-axis, using a logarithmic scale) with mutations at the indicated nucleotides. **Abbreviations**: **ESE** = Exonic Splicing Enhancer; **ASO** = Antisense Oligonucleotide

Table 3.2: The best candidate ASO sequences designed to target the selected exon of the VHL tumor suppresor at the splice junction. The table lists the ASOs that were specifically designed at the junctions of the selected exon of VHL (Ensembl stable ID: ENSE00003504189.1, second exon in both VHL annotated transcripts). The listed ASOs, along with their corresponding sequences indicated in the table, have successfully passed both specificity and physicochemical filters. The nomenclature used for the ASOs follows the conventions described by Mann *et al.,* 2002 [75].

Gene	Exon ID	ASO ID	ASO Sequence
VHL	ENSE00003504189.1	H2A(-18, 7)	AGGUGACCUAUCGGGACAAGCAAAG
		H2A(-17, 8)	AAGGUGACCUAUCGGGACAAGCAAA
		H2A(-16, 9)	AAAGGUGACCUAUCGGGACAAGCAA
		H2A(-15, 10)	CAAAGGUGACCUAUCGGGACAAGCA
		H2A(-14, 11)	CCAAAGGUGACCUAUCGGGACAAGC
		H2A(-13, 12)	GCCAAAGGUGACCUAUCGGGACAAG
		H2A(-12, 13)	AGCCAAAGGUGACCUAUCGGGACAA
		H2D(13, -12)	AAACGUCAGUACCUGGCAGUGUGAU
		H2D(12, -13)	AAAACGUCAGUACCUGGCAGUGUGA
		H2D(11, -14)	UAAAACGUCAGUACCUGGCAGUGUG
		H2D(10, -15)	GUAAAACGUCAGUACCUGGCAGUGU
		H2D(9, -16)	AGUAAAACGUCAGUACCUGGCAGUG
		H2D(8, -17)	AAGUAAAACGUCAGUACCUGGCAGU
		H2D(7, -18)	AAAGUAAAACGUCAGUACCUGGCAG

3.3.2 Harnessing the potential of our pipeline on well-studied cancer genes: BRAF and TP53

As a second set of case studies, we applied our computational pipeline to analyse one oncogene (BRAF) and one tumor suppressor (TP53), among those most frequently mutated in cancer (Figure 3.3, inset). For each of the two selected genes, the pipeline identified exons with the potential for being targeted by ES-based therapeutic strategies. The pipeline prioritised these exons based on their frequency of mutation in cancer patients. Table 3.3

displays the top ten exons that exhibit the highest mutation frequencies for each selected gene. The table also provides the number of candidate ASOs designed for inducing ES at the splice junctions (ASO-J) and within the target exon (ASO-E). Additionally, a flag (IN/OUT) is included to indicate the correctness of the transcript frame following targeted ES, according to the predicted role of the gene in cancer. In addition, Table E1 in Appendix section lists the best candidate ASO-J sequences for both the BRAF and TP53 genes.

Table 3.3: Top ten mutated exons of the BRAF and TP53 genes. The table specifically focuses on BRAF as an oncogene and TP53 as a tumor suppressor. For each chosen gene, the table includes the top ten mutated exons and indicates whether the transcript frame following targeted exon skipping is in-frame or out-of-frame (flagged "IN"/"OUT" in the table). The last two columns display the number of candidate ASOs designed to induce exon skipping, that satisfy all the physicochemical parameters thresholds and with a unique match in the genome. Columns ASO-J and ASO-E refer to ASOs targeting splice junctions and ESEs within the corresponding exon, respectively. Exons that have at least one ASO sequence available for both the ASO-J and ASO-E design strategies, and result in the desired frame in the shortened transcript, are shaded in grey to indicate their significance.

Gene	Classification	Top 10 mutated exons	IN/OUT frame	ASO-E	ASO-J
BRAF	Oncogene	ENSE00003485507.1	OUT	4	5
		ENSE00003559218.1	OUT	13	4
		ENSE00003569635.1	OUT	16	0
		ENSE00003587655.1	IN	18	14
		ENSE00001035295.1	IN	34	4
		ENSE00001907699.1	OUT	3	4
		ENSE00003527888.1	IN	12	7
		ENSE00003521664.1	OUT	15	12
		ENSE00003487759.1	IN	13	13
		ENSE00003687908.1	OUT	21	0
TP53	Tumor Suppressor	ENSE00003518480.1	OUT	59	4
		ENSE00003725258.1	OUT	57	13
		ENSE00003712342.1	OUT	30	6
		ENSE00002048269.1	OUT	25	5
		ENSE00003723991.1	OUT	56	10
		ENSE00002073243.1	OUT	25	7
		ENSE00003625790.1	IN	97	5
		ENSE00003670707.1	IN	10	3
		ENSE00003545950.1	OUT	46	5
		ENSE00003786593.1	OUT	18	7

3.4 IL-8 gene: further in-depth case study

As a third case study, our computational pipeline was utilised to analyse in depth one oncogene, the C-X-C motif chemokine ligand 8 (CXCL8), also known as Interleukin 8 (IL-8), as part of a work in collaboration with the University of Verona and the IRCCS Regina Elena National Cancer Institute of Rome. IL-8 is a pro-angiogenic and pro-inflammatory factor that acts by binding to its cognate receptors (CXC receptor 1 and 2). Both of these are G protein-coupled receptors expressed by both immune/stromal and cancer cells. Upon binding to its membrane-bound receptor on target cells, IL-8 activates specific downstream signaling pathways, such as the phosphoinositide 3-kinase (PI3K) and the mitogen-activated protein kinase (MAPK) cascades. This activation leads to the promotion of different pro-tumoral phenotypes [76]. We conducted a comprehensive investigation of IL-8 gene expression profile within the context of colorectal cancer (CRC). Initially, we analysed the expression levels of IL-8 to unveil its distinct behavior in colon normal compared to colorectal cancer tissues. Then, we integrated this expression analysis with our computational procedure to identify potential exon-skipping (ES)-events and design ASO candidate sequences optimised to achieve our desired therapeutic objectives.

3.4.1 IL-8 gene expression analysis in colorectal cancer

To gain insights into the expression of IL-8 gene in CRC, we conducted an extensive analysis of publicly available gene expression data for both cancer and normal human tissues through the PCAWG [70] and GTEx resources[71]. The study was performed at the transcript level, focusing on two specific protein-coding transcripts annotated for the IL-8 gene. These transcripts,

namely ENST00000307407 and ENST00000401931, were present in the RNA-sequencing data repositories utilised for our analysis. Both IL-8 transcripts revealed a significant increase in their expression levels (Wilcoxon rank-sum test p-values equal to 4.6e-19 and 3.4e-13 for ENST00000307407 and ENST00000401931, respectively) in colorectal cancer compared to normal colon tissues (Figure 3.8 and Table 3.4).

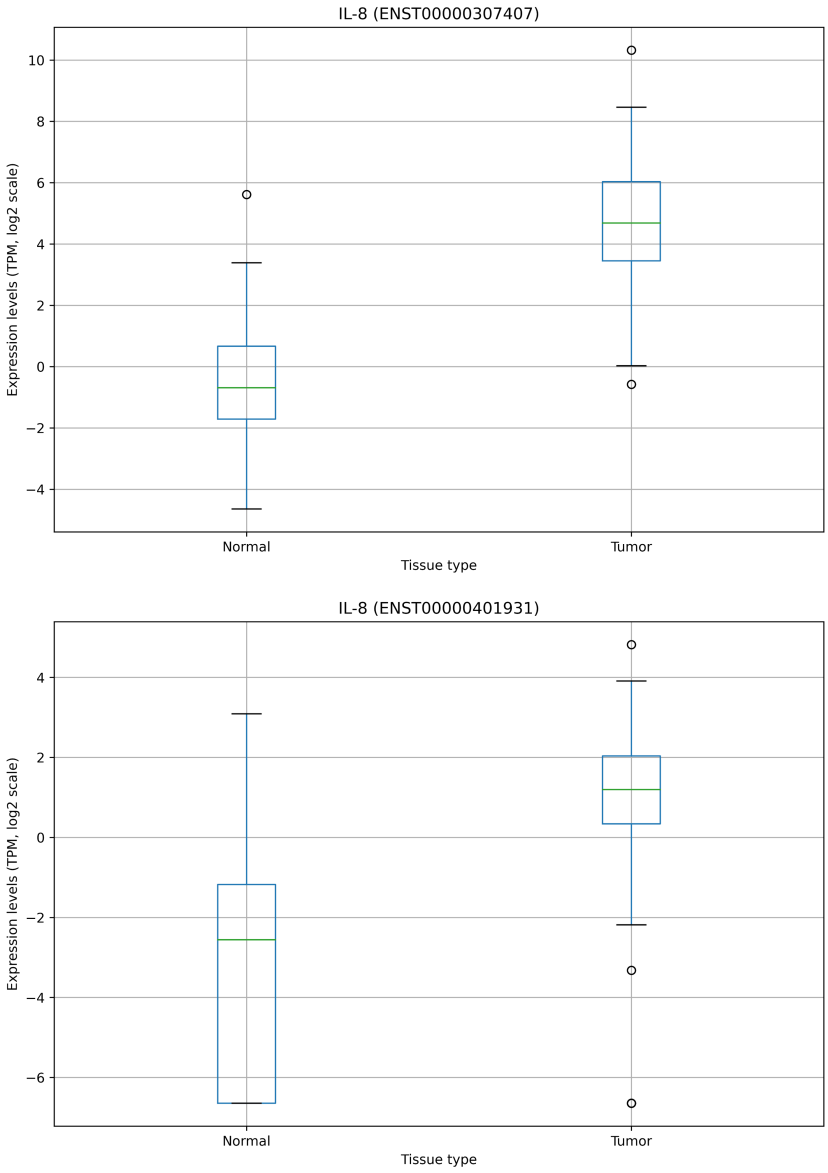

Figure 3.8: IL-8 gene expression in colorectal cancer and normal colon tissues. Boxplots show expression values for two protein-coding transcripts annotated for the IL-8 gene, namely the ENST00000307407 (top panel) and the ENST00000401931 (bottom panel) transcripts, across 51 colorectal cancer samples from the PCAWG resource ("tumor") and 88 normal colon samples from the GTEx resource ("normal"). Expression values are given in log2-transformed TPM counts. **Abbreviations**: **GTEx** = Genotype-Tissue Expression; **PCAWG** = Pan-Cancer Analysis of Whole Genomes; **TPM** = Transcripts per Million

Gene	Transcript (Ensembl Identifier)	Mean_normal (TPM)	Mean_tumor (TPM)
IL-8 (ENSG00000169429)	ENST00000307407	1.9	80.6
IL-8 (ENSG00000169429)	ENST00000401931	0.4	3.6

Table 3.4: IL-8 gene mean expression values in normal colon tissues and colorectal cancer samples. The table lists two protein-coding transcripts for the IL-8 gene (Ensembl gene identifier: ENSG00000169429). Expression values were obtained from the GTEx (Genotype-Tissue Expression) and PCAWG (Pan-Cancer Analysis of Whole Genomes) resources for normal colon tissues and colorectal cancer samples, respectively. **Abbreviations**: **TPM** = Transcripts per Million.

3.4.2 Targeting IL-8 with antisense oligonucleotides for therapeutic intervention

Based on the results of the gene expression analysis, the therapeutic goal of an ES-based approach targeting the IL-8 oncogene would be to induce protein degradation in order to mitigate its overexpression and the associated gain-of-function effects.

From our computational analysis, one particular exon, exon 2, emerged as a potential target for ES. Skipping this exon would result in out-of-frame transcripts, although our computational procedure predicts that they would not undergo degradation via NMD following ES. We then utilised our computational procedure to design customized ASO candidates, resulting in a total of 125 ASOs sequences designed to induce selected ES variants of IL-8. Among these ASOs, 111 were specifically tailored to target regulatory splice sites within the exon, while 14 ASOs were designed to target splice sites at the junctions with flanking introns. After filtering based on both specificity and physicochemical requirements, the procedure selected a subset of 58 ASOs, all of which targeting internal ESEs.

Chapter 4

Materials and Methods

The entire computational procedure described in this work was implemented using the Python programming language (v.3.10).

4.1 Selection of test-case genes based on cancer mutations

To collect the set of test-case genes, we employed mutation data obtained from the publicly available COSMIC database (v.96) (`https://cancer.sanger.ac.uk/cosmic`) [14]. Specifically, we retrieved the "CosmicMutantExportCensus.tsv" file, which can be accessed after logging in from the main menu of the COSMIC webpage through the following path: Data → Downloads → All Mutations in Census Genes → "CosmicMutantExportCensus.tsv" (Download Date: March 26, 2023). From this extensive dataset, a cohort of 72 genes was selected by identifying the top 10% most frequently mutated genes in cancer patients.

4.2 In-silico classification of genes as oncogenes or tumor suppressors

Mutation annotation data obtained from the COSMIC database was subsequently used to accurately classify *in silico* the selected cancer genes (N=72 genes) into the distinct categories of oncogenes or tumor suppressors by applying the well-established "20/20 rule" [4]. Based on this rule, for a gene to be classified as an oncogene, it must exhibit recurrent missense mutations that account for more than 20% of all reported mutations, indicating gain-of-function alterations. On the other hand, to be classified as a tumor suppressor, a minimum of 20% of annotated mutations within the gene should be inactivating. Thus, this rule captures the two main categories of mutations, namely gain-of-function and loss-of-function mutations, and accounts for their respective frequencies. Specifically, gain-of-function mutations encompass the following description type labels from COSMIC: substitution_missense, deletion_in-frame, insertion_in-frame, complex_deletion_in-frame. Conversely, loss-of-function mutations include the following labels: substitution_non-sense, deletion_frameshift, insertion_frameshift. To implement the rule in our computational procedure, we followed the methodology outlined by Pavel *et al.,* 2016 [11]. This involves the calculation, for any gene of interest, of two distinct scores based on annotated mutation data: the oncogene (ONG) and the tumor suppressor gene (TSG) score. To determine these scores, we started by assessing the total number of annotated variants within a given gene. Subsequently, we computed the frequencies of gain-of-function and loss-of-function mutations in relation to the total number of annotated variants for the same gene. In particular, the ONG score denotes the frequency of recurrent gain-of-function mutations,

whereas the TSG score corresponds to the frequency of loss-of-function mutations. Lastly, based on the criteria summarised in Table 4.1, the gene was assigned a specific label. Specifically, if the ONG score is greater than 20% and the TSG score is less than or equal to 5%, the gene is classified as an oncogene. Alternatively, if the ONG score exceeds 20% and the TSG score exceeds 5%, or if the ONG score is less than 20% and the TSG score is greater than 20%, the gene is labeled as a tumor suppressor. Finally, genes that do not meet either of these criteria are designated as "unclassified."

Table 4.1: Criteria implemented for the *in-silico* classification of genes as oncogenes or tumor suppressors. The table presents the threshold values, corresponding to the indicated types of mutations, employed to categorize a given gene into different groups (namely, oncogene, tumor suppressor, or unclassified) based on available mutation annotations. These threshold values, which are implemented in our computational procedure, are derived from Pavel *et al.,* 2016 [11].

ONG score (x/total_mutations $\times$ 100)	**TSG score** (y/total_mutations $\times$ 100)	**Classification**
> 20%	$\leq 5\%$	Oncogene
> 20%	> 5%	Tumor suppressor
< 20%	> 20%	Tumor suppressor
< 20%	< 20%	Unclassified

x = substitution_missense + deletion_in-frame + insertion_in-frame + complex_deletion_in-frame

y = substitution_non-sense + deletion_frameshift + insertion_frameshift

To evaluate the accuracy of our classification, we compared results with two publicly available compendia of oncogenes and tumor suppressor genes. The first was taken from the Molecular Signatures Database (MSigDB), which provides a comprehensive collection of annotated gene sets for human and mouse genomes. Specifically, a set consisting of genes documented in the literature as being mutated and implicated in cancer development

were retrieved from MSigDB (https://www.gsea-msigdb.org/gsea/msigdb). Specifically, we retrieved the lists of oncogenes and tumor suppressors, which can be accessed after logging in from the main menu of the MSigDB webpage through the following path: [from left panel] Gene Families → then select "oncogenes" and "tumor suppressors" from the table listing all the available gene families. The list was last updated in 2004 [67]. The second compendium was obtained from the study conducted by Tokheim *et al.,* 2016 [68]. In their work, the authors provide a selection of genes classified as oncogenes or tumor suppressors based on the integration of results obtained by applying multiple prediction methods.

4.3 Identification of targets for exon skipping

To identify potential exons suitable for therapeutic ES approaches, the following steps were performed for each gene under investigation. Firstly, all alternative transcript isoforms annotated were collected from GENCODE (`www.gencodegenes.org/`) (Release 43, GRCh38.p13). Specifically, we retrieved the GFF3 file "gencode.v43.annotation.gff3" (content description: "Comprehensive gene annotation"; regions: CHR) which contains comprehensive gene annotation concerning reference chromosomes of the human genome. Additionally, the corresponding sequences of each transcript and its respective exons were downloaded from Ensembl (Release 109) (`www.ensembl.org/`).

4.4 Prediction of degradation via Nonsense Mediated Decay

We performed Nonsense Mediated Decay (NMD) degradation prediction exclusively to the shortened transcripts labeled as out-of-frame resulting

from *in-silico* simulation of ES events. This is because out-of-frame transcripts have the potential to generate premature termination codons (PTCs), which are relevant for NMD analysis. To determine whether NMD would be triggered for these transcripts, we applied a set of well-established rules associated with NMD evasion. These include two canonical rules, which are known as the '50-55 nt rule' and 'last-exon rule', and a noncanonical rule called the 'start-proximal rule' (illustrated in Figure 4.1). Under the '50-55 nt rule', a PTC located less than 50-55 nucleotides upstream of the last exon-exon junction typically does not activate NMD [77, 78]. Specifically, in our analysis, we adopted the more stringent threshold of 55 nucleotides, in line with previous research findings [79, 80]. The 'last-exon rule' of NMD evasion states that PTCs within the last exon are usually not recognised as premature codons because normal termination codons are predominantly found in this exon. Our procedure, therefore, considered PTCs in the last exon not to be subject to NMD [81, 82]. Finally, the 'start-proximal rule' of NMD evasion, which is a noncanonical rule discovered in cancer data, states that the efficiency of NMD decreases within the 5'-most nucleotides of the coding region of a transcript. Specifically, PTCs located approximately between 150 nucleotides from the start codon typically do not trigger NMD [77, 82, 83]. These rules collectively guided our assessment of the potential for NMD activation or evasion in the transcripts under analysis.

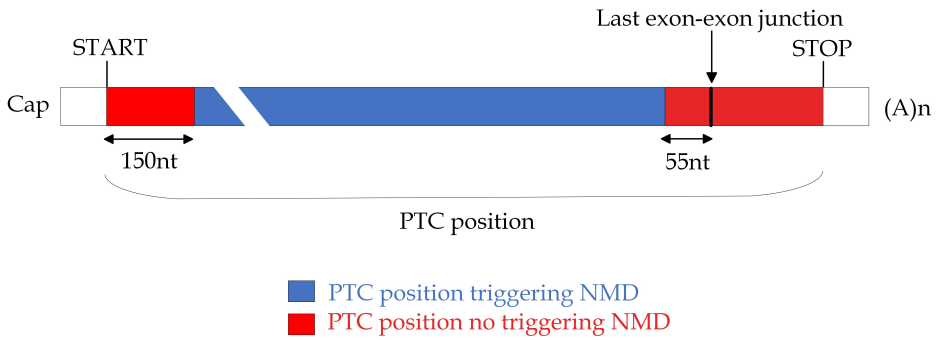

Figure 4.1: Rules to predict the occurrence of Nonsense Mediated Decay (NMD) based on the position of a premature termination codon (PTC). This figure shows a schematic representation of an mRNA, indicating the positions of translation initiation (START) and termination (STOP) codons that define the main open reading frame. The figure illustrates the application of three rules associated with NMD evasion: the '50-55 nt rule', the 'last-exon rule', and the 'start-proximal rule'. Distinct colors are used to highlight regions where PTCs may be located and the predicted consequences. Specifically, blue shading denotes PTC positions that are more prone to trigger NMD, while red shading indicates positions less likely to elicit NMD. Figure modified from Carrier *et al.,* 2010 [79].

4.5 Antisense oligonucleotide design

The implemented procedure employs two distinct approaches to generate 25-nucleotide-long ASO sequences aimed at inducing ES. The chosen length (25-mer) is indeed recommended for designing Morpholino-type ASOs [64]. The first approach involves designing ASOs that target splicing regulatory sites at the splice junctions (*donor* and/or *acceptor* splice sites), while the second approach focuses on splicing enhancer sites located within the exon, also known as ESEs. For each exon, a fixed number of 14 ASOs (i.e., a set of 7 ASOs each for both donor and acceptor splice sites) are designed to target splicing regulatory sites at the splice junctions, while the number of exonic ASOs designed depends on both the exon length and number of annotated ESEs therein. To generate the ASO sequences, a sliding window

of 25 nucleotides is used. Specifically, for ASOs designed within exonic regions, the window is shifted by one nucleotide at each step, starting from the first position of the exon and ending with the last window within the exon region. Subsequently, only the 25-mer sequences that completely overlap with at least one annotated ESE site are retained for further evaluation of physicochemical parameters. In the case of ASOs designed to target the *acceptor* splice site, a set of 7 ASO sequences that overlap the intron-exon border is collected. The first ASO in this set begins at intronic position -18 (i.e., ASO sequence -18 to +7 from the intron-exon border, where negative and positive numbers indicate intronic and exonic nucleotides, respectively). Starting from there, the next ASO sequences are drawn by sliding a 25-mer window by one nucleotide at the time until reaching the last ASO sequence in the set, which begins at intronic position -12 (i.e., ASO sequence -12 to +13). Similarly, in case of the set of 7 ASOs designed to target the *donor* site, the first 25-mer window starts at position +12 from the exon-intron border within the exon (i.e., ASO sequence +12 to -13), and this window is shifted by one nucleotide until reaching the beginning of the last ASO in the set, which is exonic position +7 from the exon-intron border (i.e., ASO sequence +7 to -18). Annotated ESE regions were taken from the SpliceAid database (`www.introni.it/splicing.html`), a database of experimentally validated target RNA sequences that are bound by splicing regulatory proteins in humans [84]. In particular, each binding site in the database is assigned a score ranging from 1 to 10. In our analysis, we focused only on sites annotated with positive scores (i.e., splicing enhancers), and mapped them within target sequences. Concerning ASO nomenclature, we followed the conventions described by Mann *et al.,* 2002 [75].

4.6 Candidate-antisense oligonucleotide evaluation and selection

Designed ASOs undergo evaluation and filtering based on reference physicochemical parameters (including CG percentage, G percentage, presence of tetra G, self-complementarity, temperature of melting (Tm)), and specificity. The ViennaRNA Package was used to assess self-complementarity, resulting in a dot-bracket string notation that indicates both paired and unpaired bases in the predicted ASO secondary structure. Only ASO sequences with a maximum of 16 contiguous base pairs of H-bonds are selected [57]. Overall, the design of ASOs follows the guidelines provided by Moulton *et al.,* 2008 [64], including optimal values for physicochemical parameters used to select the best-candidate sequences (which are summarised in Table C1 for easy reference). Tm, a crucial parameter that significantly affects the specificity and effectiveness of ASOs, was calculated using three different methods, with reference to the OligoCalc web tool [56]:

- Melting temperature Basic (Tm) (°C)

$$\text{Tm} = 64.9 + 41(yG + zC - 16.4)/(wA + xT + yG + zC)$$

- Melting temperature Salt Adjusted (Tm) (°C)

$$\text{Tm} = 100.5 + 41(yG + zC)/(wA + xT + yG + zC) - \frac{820}{wA + xT + yG + zC} + 16.6 \log_{10}\left(\left[\text{Na}^+\right]\right)$$

- Melting Temperature Nearest Neighbor (TmNN) (°C)

$$\text{TmNN} = \left(\Delta H - 3.4 \frac{\text{kcal}}{\text{Kmole}}\right)/\Delta S + R \ln\left(\frac{1}{[\text{primer}]}\right) + 16.6 \log\left(\left[\text{Na}^+\right]\right)$$

The thermodynamic parameters were calculated assuming standard conditions (namely, 1M NaCl, pH=7, and temperature of 25°C). The nearest-neighbor and thermodynamic calculations were performed as described by Breslauer *et al.,* 1986 [85] but using the values published by Sugimoto *et al.,* 1996 [86]. RNA thermodynamic properties were taken from Xia *et al.,* 1998 [87]. The specificity was evaluated by assessing the absence of other potential binding sites in the genome with up to 2 nucleotide mismatches. The Bowtie [88] short-read aligner was used to map ASO sequences against the human genome (release GRCh38) with the following parameters: -v 2 (i.e., map allowing up to 2 mismatches per read alignment) -a (i.e., report all possible alignments). Only ASO sequences that met both specificity requirement (i.e. unique match in the human genome considering up to 2 mismatches) and physicochemical requirements (i.e. recommended threshold for selected parameters taken from Moulton *et al.,* 2008 [64]) were reported for any gene of interest as candidates to induce the selected ES products.

4.7 Expression analysis for selected genes in cancer and normal samples

Expression data from RNA-sequencing experiments [24, 69] were obtained interrogating resources made available from large genomic initiatives, specifically the Pan-Cancer Analysis of Whole Genomes (PCAWG) for cancer samples [70] and the Genotype-Tissue Expression (GTEx, v.4) resource [71], for normal tissues. Expression data matrices summarized at the transcript level (Transcripts Per Million (TPM) counts) as well as sample metadata for both datasets (i.e. PCAWG and GTEx) were downloaded from the International Cancer Genome Consortium data portal (`https://dcc.`

icgc.org/releases/PCAWG/, and folders transcriptome/ and metadata therein) [89]. Sample cohorts included in the PCAWG pan-cancer expression data matrix (N=1,359 samples) and the GTEx expression data matrix of normal human tissues (N=3,247 samples) were filtered on samples by selecting only those cancer samples with available reference normal tissues and viceversa (based on cancer *vs.* normal tissue pairing provided in Kahraman *et al.*, 2020 [72]). Specifically for the NRAS gene, our analysis covered 1,359 cancer samples across 19 distinct tissue types, and 2,231 corresponding normal samples (Table A3 in Appendix D). Expression levels, measured in TPM counts, for the NRAS gene were retrieved by focusing on its protein-coding transcript ENST00000369535 annotated by Ensembl. Regarding the IL-8 gene, we examined 51 colorectal cancer samples with genomic data labeled as of optimal quality (histology abbreviation equal to "ColoRect-AdenoCA" and genomic data label equal to "Whitelist" in the PCAWG metadata file) from the PCAWG expression matrix. Additionally, we considered 88 colon samples from the GTEx expression matrix (i.e. histological type equal to "Colon" in the GTEx metadata file). Expression levels, measured in TPM counts, for the IL-8 gene were obtained considering two alternative transcripts annotated by Ensembl and included in the expression dataset, specifically ENST00000401931 and ENST00000307407.

For each selected transcript, both for the NRAS and the IL-8 gene, the abundance measurements were tested for differential expression between the two sets of samples (i.e. cancer and normal samples) by using the Wilcoxon rank-sum test. A test p-value lower than 0.05 was taken as indicative of statistically significant expression difference.

Chapter 5

Discussion

The development of personalised therapeutic approaches based on molecular tumor profiling has made significant advancements in cancer treatment. However, many cancer driver genes, including frequently altered genes like the RAS family, MYC, and TP53, remain challenging to target with conventional approaches, leaving them as "undruggable." This poses a major hurdle to oncology drug development.

This thesis work aims to present an integrated computational procedure that facilitates the exploration and implementation of exon skipping (ES)-based therapeutic strategies for cancer treatment. The developed computational procedure leverages existing knowledge of annotated transcripts and disease-causing mutations for a specific gene of interest. It guides the selection of target exons and the design of antisense oligonucleotides (ASOs) to induce selected ES events. The procedure also provides insights into the consequences of exon exclusion, including the potential degradation of transcripts through Nonsense Mediated Decay (NMD). This enables the evaluation of the impact of ES on protein expression and functionality, leading

to a deeper understanding of the case of interest. The approach presented is versatile and can support strategies aimed at producing different desired protein variants, such as variants of inactive oncogenes or partially functional restored variants of tumor suppressors.

As a proof-of-concept, the study focused on the top 10% most mutated genes in cancer and ranked them based on their suitability for ES-targeted interventions. This resulted in a list of the most promising candidate genes for ES-based therapies that contain the highest percentage of exons meeting the following criteria: a) The exon, when excluded, should result in desired protein variants aligning with the intended therapeutic effect; b) The ASOs designed for the exon skipping should meet recommended physicochemical parameters and have a unique match in the genome, ensuring their effectiveness in inducing exon exclusion; c) The target exon should rank among the top ten most mutated exons in cancer patients, indicating its clinical relevance. Based on these criteria, NRAS, a member of the RAS protein family, emerged as one of the most promising candidates, with all of its exons being eligible for skipping. Similarly, other genes such as KRAS, VHL, CALR, GRIN2A, JAK2, FLT3, IDH1, and IDH2 also exhibited a significant percentage of their exons meeting the criteria for optimal skipping targets. In particular, the percentage of optimal exons ranged from over 20% for GRIN2A, JAK2, FLT3, IDH1, and IDH2 to over 60% for KRAS (another member of the RAS family). These findings highlight the potential of these genes as promising candidates for ES-based therapeutic interventions in cancer.

In order to provide more detailed examples, the study specifically investigated NRAS and VHL, the top-ranking oncogene and tumor suppressor, respectively, in our list of best candidates for the ASO-induced exon

skipping procedure. Oncogenic NRAS mutations occur in several cancer types, notably melanoma, acute myeloid leukemia, colon and thyroid cancers, and other hematologic malignancies. While attempts have been made to action NRAS for therapeutic purposes by targeting either downstream effectors (e.g., MEK, CDK4/6; [90]) or upstream activators (e.g., STK19; [91]), the NRAS oncogene itself remains currently undruggable and could be theoretically targeted by ES-inducing strategies, as proposed here. For this oncogene, a preliminary evaluation of its expression was undertaken in the tumor context compared to physiological context. This investigation aimed to verify the presence of its protein-coding transcript, which could be potentially targeted, and confirm its overexpression in cancer tissues relative to normal ones. These findings provide insights into the feasibility of a strategy aimed at reducing its expression. Similarly, mutations that inactivate the tumor suppressor VHL are a major genetic driver of both hereditary and sporadic renal cell carcinoma [92]. Although VHL-defective cancers can be targeted with clinical success by inhibiting its downstream effector HIF1a or VEGF-driven angiogenesis [93], we hypothesise that VHL function could be restored, at least in part, by ASO-mediated transcript modification. The comprehensive *in-silico* approach successfully identified the highest-scoring exon for each gene of interest (NRAS and VHL), and designed corresponding ASO sequences to induce its exclusion.

Expanding the scope to include the most frequently mutated genes in cancer, as notable cases with clinical relevance, the computational pipeline was applied to the extensively studied oncogene BRAF and tumor suppressor TP53. Small-molecule inhibitors of oncogenic BRAF have proven highly successful for the clinical treatment of several tumor types. However, atypical (non-V600E) BRAF mutations and BRAF-dependent acquired re-

sistance both remain significant challenges and unmet medical needs for which ES-inducing strategies could provide a targeting approach [94]. On the other hand, TP53, one of the most frequently mutated genes in cancer, remains the prototype of an undruggable tumor suppressor, for which no successful therapeutic strategies have been devised yet [95]. The computational procedure successfully predicted four potential ES events in the BRAF gene. These ES events would result in the generation of out-of-frame transcripts by skipping frequently mutated exons in cancer. The shortened transcripts would likely be targeted for degradation through NMD, thereby reducing the corresponding protein expression level. For each identified exon, the procedure designed an average of six ASO sequences at the exon–intron junctions and nine sequences overlapping exonic splicing enhancers (ESEs) within the exon. Regarding the TP53 gene, two exon candidates for ASO-mediated therapeutic ES events were identified. The aim of these events would be to maintain the transcript frame after exon deletion and potentially restore, at least partially, the biological protein function. An average of four ASOs to induce skipping were designed at the exon-intron junctions, while approximately 53 ASO sequences that cover the ESE regions were designed within the exon.

Subsequently, we applied our computational procedure to the oncogene IL-8 and investigated the expression of its annotated transcript isoforms specifically in the context of colorectal cancer (CRC). This exploration was prompted by a collaboration with the University of Verona and the IRCCS Regina Elena National Cancer Institute of Rome, focusing on CRC. IL-8 is a pro-angiogenic chemokine that, through activation of specific signaling pathways (such as the PI3K and the MAPK cascades), may promote CRC invasion and metastasis [96, 97]. Additionally, circulating IL-8 represents

a strong prognostic factor in CRC, and this gene emerged as a possible determinant of response to immunotherapy and targeted treatment in several cancer types [96]. This evidence supports the investigation of IL-8 as a candidate druggable target for CRC treatments. Within this framework, we initially investigated expression levels of IL-8 at the transcript level by interrogating public atlases of cancer and normal expression data. Our analysis confirmed significant overexpression of both annotated isoforms of the IL-8 gene in CRC. Subsequently, utilising our *in-silico* approach, we successfully identified the highest-scoring IL-8 exon for targeted ES strategies and designed corresponding ASO sequences to induce its exclusion.

The results obtained from applying the entire procedure to these cancer-relevant case studies demonstrate suitability of the strategy in supporting the design of ASOs that can be experimentally validated as candidates for innovative therapeutic interventions. This approach aims to induce targeted exon skipping in oncogenes and tumor suppressors, even in cases with significant mutation burdens that are difficult to target with conventional pharmacological methods. It is important to note that the criteria employed in this procedure are independent of tissue and cancer type, aligning with an agnostic approach. However, the same computational method can be tailored to specific cases by utilising mutation profiles obtained from individual patient screenings. Such a personalised approach has the potential to enhance targeting accuracy and minimise off-target or side effects. In fact, while the systematic approach presented in this study allows for wide-ranging applicability across diverse contexts, offering several advantages, we recognise the value of tailoring this computational method to specific cases, developing further in-depth analysis methods. Among others, the assessment of whether a target splicing isoform is actually expressed could be more effec-

tively conducted by analysing expression profile data related to the specific tissue or tumor under investigation [72–74].

Importantly, the proposed procedure has certain limitations that should be taken into account. For instance, the current approach relies on genomic and transcriptomic data available in annotation databases like COSMIC and GENCODE, which may have limitations in terms of coverage and accuracy. In addition, it might be advisable to integrate the proposed procedure with a more detailed examination of the consequences of various mutations that can perturb splicing regulatory regions [98–100], including synonymous and intronic mutations, by utilising ad-hoc existing tools [101, 102]. Likewise, the potential role of cryptic regulatory sites in influencing splicing outcomes would deserve further study [103]. Moreover, it should be noted that the adopted classification of oncogenes and tumor suppressor genes, which in turn determines the type of molecular effect sought through ES, can be debatable from several perspectives: classification methods and available compendia might not cover all potentially relevant genes; a small but non-negligible percentage of genes may be classified differently using various procedures; in some cases, the same gene can function as an oncogene or a tumor suppressor, depending on the type of mutation it undergoes (loss vs gain of function), such as in the case of TP53 [95]. Furthermore, the delivery of ASOs to target tissues or cells poses unique challenges that must be overcome as a prerequisite for their effective therapeutic application. Several efforts are currently invested in improving the delivery of ES-based therapeutic drugs through innovative chemical modifications and conjugation with delivery-enhancing agents, such as fatty acids or peptides [26, 104–107]. Finally, experimental validation is crucial to confirming the actual impact of predicted ASO-induced exon skipping on protein expression and

function. For instance, it has been reported that the inhibition of a splice site may trigger the activation of a cryptic splice site, leading to the formation of a transcript with an unexpected architecture [64].

Of note the current approach, which prioritises as ES target the most frequently mutated exon in cancer patients, is flexible and can be adapted to different scenarios and therapeutic goals. For instance, in cases where the therapeutic strategy involves inactivating a mutated oncogene, a safer approach to ASO design may entail targeting frame-shifting exons that are possibly not mutated and are located upstream of the most frequently mutated one. This approach would result in the desired outcome of an out-of-frame transcript, reducing the risk that patient-specific mutations might decrease the affinity for the designed ASO. On the other hand, in more complex scenarios related to the rescuing of function of tumor suppressors, where it is critical to safeguard the healthy allele, the existence of patient-specific mutations may be the key to enable the selectivity needed in the design of allele-specific therapeutic strategies. A notable example illustrating the relevance of envisioning allele-specific therapeutic strategies is the haploinsufficiency of the tumor suppressor PTEN [108], frequently mutated in human cancer. Here, similar to the strategy pursued with Duchenne muscular dystrophy (DMD) drugs, even subtle changes in PTEN expression levels seem to have the potential to alter tumor cell behavior [108], motivating attempts to pursue even limited functional rescue obtained by allele-selective interventions.

In conclusion, the integrated computational procedure developed in this study presents a strategy and provides valuable tools for investigating ES-based therapeutic approaches in oncology. ASOs show promise as innovative and personalised therapeutic interventions, particularly for targeting

currently undruggable driver genes. Further experimental validation, optimisation, and technological advancements are necessary to fully harness the potential of ASOs as clinically effective therapies. These efforts could pave the way for their effective utilisation in cancer treatment, making a substantial contribution to the field of personalised medicine in oncology.

Chapter 6

Appendix

6.1 Appendix A

To validate our classification, available for 48 genes, the results were compared with two publicly available compendia of oncogenes and tumor suppressors: 1) Futreal *et al.,* 2004 [67], hereafter *MSigDB*; 2) Tokheim *et al.,* 2016 [68], hereafter *Tokheim*. When comparing our gene classification results with the *MSigDB* compendium (N=410, comprising 328 oncogenes and 82 tumor suppressors), we found 35 shared genes available for the purpose of classification comparison, while 13 (i.e., 6 tumor suppressors and 7 oncogenes) out of the 48 genes composing our classified list were not present in the *MSigDB* compendium. Among the shared genes (N=35), 30 genes were consistently classified, of which 14 out of 19 tumor suppressors (74%) and all the oncogenes (16/16). However, 5 genes (CREBBP, NOTCH1, NOTCH2, NPM1, RUNX1) were classified differently (i.e., *MSigDB* classified them as oncogenes, while our procedure as tumor suppressors) (Figure A1, panel A). Next, we compare our gene classification results to the *Tokheim* compendium (N=290 genes, including 79

oncogenes and 211 tumor suppressors) (Figure A1, panel B). Eleven out of the 48 genes in our classified list were not included in the *Tokheim* compendium. As for the remaining 37 genes of our classified list (i.e., 24 tumor suppressors and 13 oncogenes) shared with the *Tokheim* compendium, all the tumor suppressor genes (24/24, 100%) and most of the oncogenes (12/13, 92%) were consistently classified. There was only one gene (JAK2) with conflicting assignments as an oncogene (our procedure) and tumor suppressor (Tokheim).

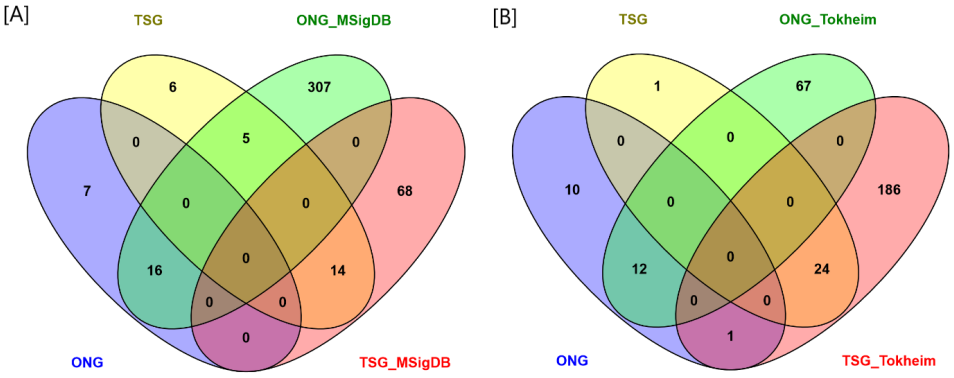

Figure A1: Cancer-genes classification comparison. Comparison of our classification of the 72 selected genes as oncogenes (ONG) or tumor suppressors (TSG) with two compendia of ONGs and TSGs. [A] In this panel, the Venn diagram shows the overlap between our (TSG, ONG) and the MSigDB (ONG_MSigDB, TSG_MSigDB) results. [B] In this panel, the Venn diagram shows the overlap between our (TSG, ONG) and the Tokheim (ONG_Tokheim, TSG_Tokheim) results. **Abbreviations**: **ONG** = oncogene; **TSG** = tumor suppressor gene; **MSigDB** = list of oncogenes and tumor suppressors from Futreal *et al.,* 2004 [67]; **Tokheim** = list of oncogenes and tumor suppressors from Tokheim *et al.,* 2016 [68]

6.2 Appendix B

As of today, there are four ASO-based drugs approved by the FDA: Eteplirsen, Golodirsen, Viltolarsen, Casimersen [38–41]. All four drugs have been developed to modulate the splicing of the dystrophin gene for therapeutic purposes in patients affected by Duchenne Muscular Dystrophy.

The Casimersen case, showing the largest deviation from the ASO sequence identified as a potential drug through our computational approach, underscores how crucial the need for experimental validation of ASO sequences is for tangible improvements in the design phase. In fact, our procedure initially designs other ASO sequences with better alignment to the Casimersen drug; however, they are excluded during filtering due to non-compliance with some recommended physicochemical parameter thresholds (Table C1) [64].

In light of this, cases like Casimersen's highlight the importance of aligning computational predictions with an increasing body of experimental results to optimize the design process.

Table B1: Comparison between results from ASO designed from our pipeline and currently approved ASO drugs. This table provides a comparison between the released sequence of the approved drug (column 3, top line) and the best match with ASOs designed using our procedure (column 3, bottom line) to target the corresponding exon. Blue color highlights nucleotides identical between the drug sequence and candidate ASO sequence designed by our procedure.

ASO drug	**Target Exon** **Ensembl ID**	**Drug sequence (length)** **Pipeline matching hit (length)**
Eteplirsen	Exon 51 (ENSE00003669071.1)	CTCCAACATCAAGGAAGATGGCATTTCT (28nt) CTCCAACATCAAGGAAGATGGCATT (25nt)
Golodirsen	Exon 53 (ENSE00001258577.1)	GTTGCCTCCGGTTCTGAAGGTGTTC (25nt) GTTGCCTCCGGTTCTGAAGGTGTTC (25nt)
Viltolarsen	Exon 53 (ENSE00001258577.1)	CCTCCGGTTCTGAAGGTGTTC (21nt) CCTCCGGTTCTGAAGGTGTTCTTGT (25nt)
Casimersen	Exon 45 (ENSE00003988228.1)	CAATGCCATCCTGGAGTTCCTG (22nt) GCTGCCCAATGCCATCCTGGAGTTC (25nt)

6.3 Appendix C

Table C1: Threshold values for physicochemical parameters used to evaluate the goodness of designed ASOs. The reported values are extracted from guidelines published by Moulton *et al.,* 2008 (see Table 2 therein entitled: "Targeting Recommendations for 37°C Systems") [64].

Parameter	Recommendation	Comments
CG range	40%-60%	At lower GC, affinity may be too low to inhibit processes; higher GC favors nonspecific binding of subsequences. Count$(G + C)$/Count$(A + T + G + C) \times 100\%$ (`https://www.biologicscorp.com/tools/GCContent/`)
G content	Up to 36% G	Higher G causes loss of water solubility; avoid upper end of acceptable range, if possible.
Self-complementarity	16 contiguous H-bonds maximum	For intermolecular (complementary palindrome) and intramolecular (stem loop) binding. Example: AGCGCT has 16 H-bonds (2+3+3+3+3+2 = 16). Check for non-Watson-Crick G-T pairing, which can participate in self-complementarities.
Consecutive G	3 consecutive Gs maximum	Runs of $\geq$4 G can be associated through Hoogsteen bonding to form oligo tetramers.
Oligo length	25 bases or shorter by only a few bases	Using shorter oligos can decrease the chance of off-target interaction for high CG oligos.

6.4 Appendix D

Information regarding the NRAS gene expression in various tissues was examined to investigate its specific behavior in cancers and assess the potential for therapeutic interventions aimed at mitigating its recurrent upregulation. Table D1 provides details on the number of samples analyzed for each tissue, whereas Figure D2 shows distribution of NRAS expression levels across cancer tissues and their normal tissues as a reference.

Table D1: Cancer and normal tissues datasets used to assay NRAS expression. This table lists numbers of cancer and normal tissue samples utilised in the NRAS expression analysis which sum up to a total of 1,359 cancer samples retrieved from the PCAWG resource [70] and 2,231 normal samples from the GTEx project [71].

Normal tissues (GTEx)	Sample number	Tumor tissues (PCAWG)	Sample number
Bladder	11	Bladder	27
Blood	269	Blood	173
Brain	420	Brain	46
Breast	66	Breast	97
Cervix Uteri	10	Cervix Uteri	20
Colon	88	Colon	51
Esophagus	249	Esophagus	7
Kidney	8	Kidney	211
Liver	35	Liver	187
Lung	152	Lung	95
Muscle	175	Muscle	34
Ovary	39	Ovary	110
Pancreas	70	Pancreas	75
Prostate	42	Prostate	20
Salivary Gland	6	Salivary Gland	43
Skin	343	Skin	36
Stomach	80	Stomach	31
Thyroid	132	Thyroid	51
Uterus	36	Uterus	45
	(tot) 2231		(tot) 1359

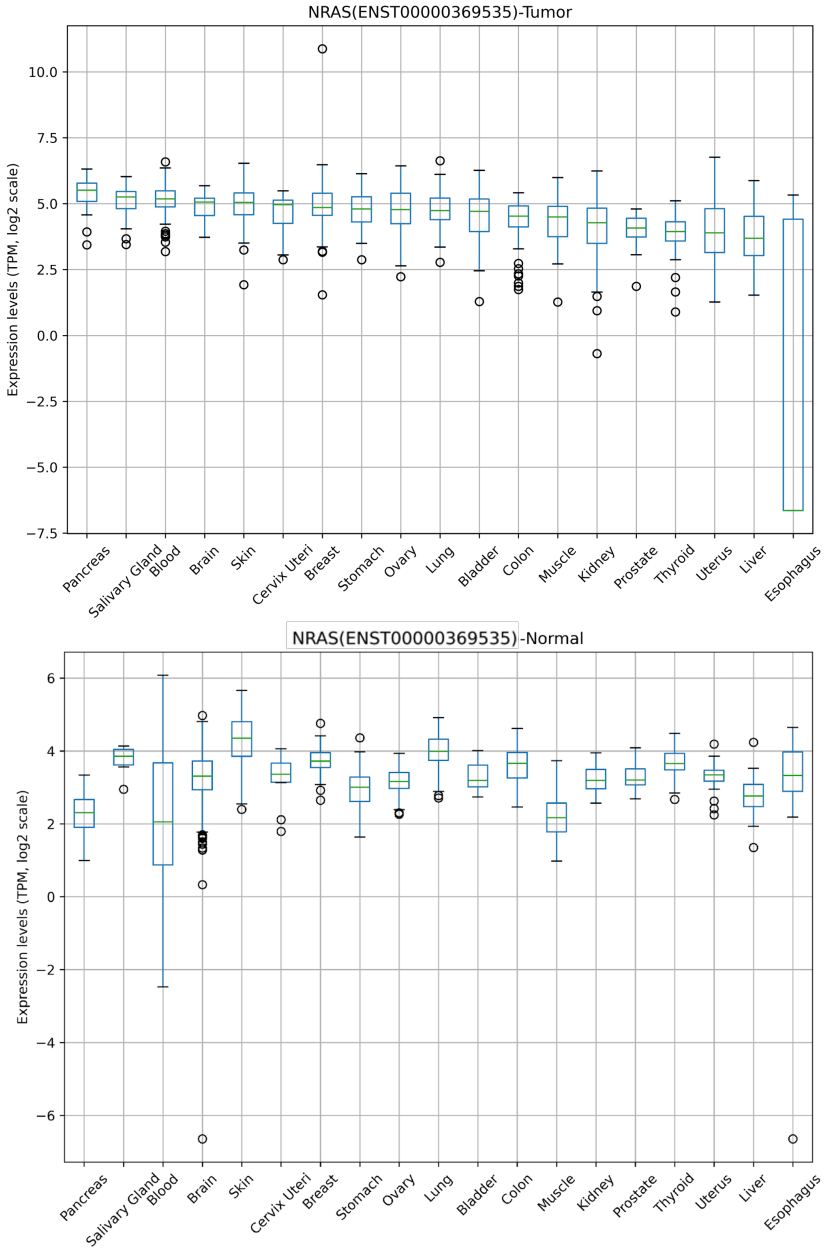

Figure D2: NRAS expression across various tumor and normal tissues. Boxplots show the distribution of expression values (log2-transformed TPM) of the NRAS gene across 19 tumor tissues from PCAWG (top panel) and corresponding normal tissues from GTEX (bottom panel). Each boxplot corresponds to a distinct tissue type (X-axis). Expression values refer to the protein-coding transcript annotated for this gene: ENST00000369535. **Abbreviations**: **GTEx** = Genotype-Tissue Expression; **PCAWG** = Pan-Cancer Analysis of Whole Genomes; **TPM** = Transcripts per Million

6.5 Appendix E

Table E1: Best candidate ASO sequences designed at the selected exons splice junctions of the BRAF and TP53 genes. The table provides the ASOs that were specifically designed at the junctions of the selected exons of BRAF and TP53 to yield desired ES outcomes (out-of-frame transcripts for the oncogene BRAF and in-frame transcripts for the tumor suppressor TP53). The listed ASOs, along with their corresponding sequences indicated in the table, have successfully passed all physicochemical filters and have a unique match in the genome. The nomenclature used for the ASOs follows the conventions described by Mann *et al.,* 2002 [75].

Gene	Exon-id	ASO-id	ASO sequence
BRAF	ENSE00001907699.1	H18D(10, -15)	GGACAGGAAACGCACCAUAUCCCCC
		H18D(9, -16)	UGGACAGGAAACGCACCAUAUCCCC
		H18D(8, -17)	GUGGACAGGAAACGCACCAUAUCCC
		H18D(7, -18)	AGUGGACAGGAAACGCACCAUAUCC
	ENSE00003485507.1	H15D(11, -14)	GCCUCAAUUCUUACCAUCCACAAAA
		H15D(10, -15)	AGCCUCAAUUCUUACCAUCCACAAA
		H15D(9, -16)	UAGCCUCAAUUCUUACCAUCCACAA
		H15D(8, -17)	AUAGCCUCAAUUCUUACCAUCCACA
		H15D(7, -18)	AAUAGCCUCAAUUCUUACCAUCCAC
	ENSE00003521664.1	H12A(-16, 9)	CCACAUCACCUAAAAGGCAAUUGUU
		H12A(-15, 10)	GCCACAUCACCUAAAAGGCAAUUGU
		H12A(-14, 11)	UGCCACAUCACCUAAAAGGCAAUUG
		H12A(-13, 12)	CUGCCACAUCACCUAAAAGGCAAUU
		H12A(-12, 13)	ACUGCCACAUCACCUAAAAGGCAAU
		H12D(13, -12)	ACACAAGCUCACCUGAGUACUCCUA
		H12D(12, -13)	CACACAAGCUCACCUGAGUACUCCU
		H12D(11, -14)	UCACACAAGCUCACCUGAGUACUCC
		H12D(10, -15)	UUCACACAAGCUCACCUGAGUACUC
		H12D(9, -16)	AUUCACACAAGCUCACCUGAGUACU
		H12D(8, -17)	AAUUCACACAAGCUCACCUGAGUAC
		H12D(7, -18)	UAAUUCACACAAGCUCACCUGAGUA
	ENSE00003559218.1	H11D(13, -12)	UUACAUACUUACCAUGCCACUUUCC
		H11D(9, -16)	CACAUUACAUACUUACCAUGCCACU
		H11D(8, -17)	CCACAUUACAUACUUACCAUGCCAC
		H11D(7, -18)	ACCACAUUACAUACUUACCAUGCCA
TP53	ENSE00003670707.1	H5A(-16, 9)	GCAAAACAUCUUGUUGAGGGCAGGG
		H5A(-15, 10)	GGCAAAACAUCUUGUUGAGGGCAGG
		H5A(-14, 11)	UGGCAAAACAUCUUGUUGAGGGCAG
	ENSE00003625790.1	H3D(13, -12)	AGGGCAACUGACCGUGCAAGUCACA
		H3D(12, -13)	CAGGGCAACUGACCGUGCAAGUCAC
		H3D(11, -14)	UCAGGGCAACUGACCGUGCAAGUCA
		H3D(10, -15)	CUCAGGGCAACUGACCGUGCAAGUC
		H3D(9, -16)	CCUCAGGGCAACUGACCGUGCAAGU

List of Figures

List of Tables

List of Publications

Papers:

- **Pacelli C**, Rossi A, Milella M, Colombo T, Le Pera L. RNA-Based Strategies for Cancer Therapy: In Silico Design and Evaluation of ASOs for Targeted Exon Skipping. International Journal of Molecular Sciences. **2023**; 24(19):14862. https://doi.org/10.3390/ijms241914862

- Bazzichetto C, Milella M, Zampiva I, Simionato F, Amoreo CA, Buglioni S, **Pacelli C**, Le Pera L, Colombo T, Bria E, et al. Interleukin-8 in Colorectal Cancer: A Systematic Review and Meta-Analysis of Its Potential Role as a Prognostic Biomarker. Biomedicines. **2022**; 10(10):2631. https://doi.org/10.3390/biomedicines10102631

Conference abstracts/posters:

- December 16th, 2022 – On-line workshop Società Italiana di Virologia-Italian Society for Virology (SIV-ISV) Workshop: "Young Minds at Work: Blending Biology and Bioinformatics" Abstract:"In silico design and evaluation of exon-skipping strategies induced by anti-sense oligonucleotides for therapeutic intervention in cancer" Link: `https://rb.gy/6392e`

- April 4th -5th, 2022 – National Research Council (CNR), Aula Convegni, Rome Department of Biomedical Sciences (DSB)-CNR Conference: “Target discovery for unmet medical needs and precision/personalized medicine.” Abstract:”In silico design and evaluation of exon skipping-inducing antisense oligonucleotides for therapeutic intervention in rare cancers” Link: `https://rb.gy/rzldu`

Bibliography

[1] Weiming Sun, Qianling Shi, Huiyun Zhang, Kexin Yang, Yuxia Ke, Yuping Wang, and Liang Qiao. Advances in the techniques and methodologies of cancer gene therapy. *Discovery medicine*, 27:45–55, jan 2019.

[2] Bhupender S. Chhikara and Keykavous Parang. Global Cancer Statistics 2022: the trends projection analysis. *Chemical Biology Letters*, 10(1):451–451, 2023.

[3] Apostolia M. Tsimberidou, Elena Fountzilas, Mina Nikanjam, and Razelle Kurzrock. Review of precision cancer medicine: Evolution of the treatment paradigm. *Cancer treatment reviews*, 86:102019, jun 2020.

[4] Bert Vogelstein, Nickolas Papadopoulos, Victor E. Velculescu, Shibin Zhou, Luis A. Diaz, and Kenneth W. Kinzler. Cancer genome landscapes. *Science (New York, N.Y.)*, 339:1546–1558, mar 2013.

[5] Emmanuel N. Kontomanolis, Antonios Koutras, Athanasios Syllaios, Dimitrios Schizas, Aikaterini Mastoraki, Nikolaos Garmpis, Michail Diakosavvas, Kyveli Angelou, Georgios Tsatsaris, Athanasios Pagkalos, Thomas Ntounis, and Zacharias Fasoulakis. Role of

Oncogenes and Tumor-suppressor Genes in Carcinogenesis: A Review. *Anticancer Research*, 40(11):6009–6015, nov 2020.

[6] E. Y. H. P. Lee and W. J. Muller. Oncogenes and Tumor Suppressor Genes. *Cold Spring Harbor Perspectives in Biology*, 2(10):a003236–a003236, oct 2010.

[7] Zeev Waks, Omer Weissbrod, Boaz Carmeli, Raquel Norel, Filippo Utro, and Yaara Goldschmidt. Driver gene classification reveals a substantial overrepresentation of tumor suppressors among very large chromatin-regulating proteins. *Scientific Reports*, 6(1):38988, dec 2016.

[8] Alvin Telser. Molecular biology of the cell, 4th edition. *New York: Garland Science*, 18:289, 2002.

[9] Pramod Chandrashekar, Navid Ahmadinejad, Junwen Wang, Aleksandar Sekulic, Jan B. Egan, Yan W. Asmann, Sudhir Kumar, Carlo Maley, and Li Liu. Somatic selection distinguishes oncogenes and tumor suppressor genes. *Bioinform.*, 36(6):1712–1717, 2020.

[10] Jie Lyu, Jingyi Jessica Li, Jianzhong Su, Fanglue Peng, Yiling Elaine Chen, Xinzhou Ge, and Wei Li. Dorge: Discovery of oncogenes and tumor suppressor genes using genetic and epigenetic features. *Science Advances*, 6(46):eaba6784, nov 2020.

[11] Ana Brândusa Pavel and Cristian Ioan Vasile. Identifying cancer type specific oncogenes and tumor suppressors using limited size data. *J. Bioinform. Comput. Biol.*, 14(6):1650031:1–1650031:16, 2016.

[12] Simon A. Forbes, David Beare, Harry Boutselakis, Sally Bamford, Nidhi Bindal, John G. Tate, Charlotte Cole, Sari Ward, Elisabeth Dawson, Laura Ponting, Raymund Stefancsik, Bhavana Harsha, Chai Yin Kok, Mingming Jia, Harry Jubb, Zbyslaw Sondka, Sam L. Thompson, Tisham De, and Peter J. Campbell. COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res.*, 45(Database-Issue):D777–D783, 2017.

[13] Zbyslaw Sondka, Sally Bamford, Charlotte G. Cole, Sari A. Ward, Ian Dunham, and Simon A. Forbes. The cosmic cancer gene census: describing genetic dysfunction across all human cancers. *Nat Rev Cancer*, 18:696–705, 2018.

[14] John G. Tate, Sally Bamford, Harry Jubb, Zbyslaw Sondka, David Beare, Nidhi Bindal, Harry Boutselakis, Charlotte Cole, Celestino Creatore, Elisabeth Dawson, Peter Fish, Bhavana Harsha, Charlie Hathaway, Steve C. Jupe, Chai Yin Kok, Kate Noble, Laura Ponting, Christopher C. Ramshaw, Claire E. Rye, Helen E. Speedy, Raymund Stefancsik, Sam L. Thompson, Shicai Wang, Sari Ward, Peter J. Campbell, and Simon A. Forbes. COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res.*, 47(Database-Issue):D941–D947, 2019.

[15] Chandan Kumar-Sinha and Arul M. Chinnaiyan. Precision oncology in the age of integrative genomics. *Nat Biotechnol*, 36:46–60, 2018.

[16] Michael J. Duffy and John Crown. Drugging "undruggable" genes for cancer treatment: Are we making progress? *Int. J. Cancer.*, 148:8–17, 2021.

[17] Adrienne D. Cox, Stephen W. Fesik, Alec C. Kimmelman, Ji Luo, and Channing J. Der. Drugging the undruggable ras: Mission possible? *Nature reviews. Drug discovery*, 13:828–851, 2014.

[18] Rameen Beroukhim, Craig H. Mermel, Dale Porter, Guo Wei, Soumya Raychaudhuri, Jerry Donovan, Jordi Barretina, Jesse S. Boehm, Jennifer Dobson, Mitsuyoshi Urashima, Kevin T. Mc Henry, Reid M. Pinchback, Azra H. Ligon, Yoon-Jae Cho, Leila Haery, Heidi Greulich, Michael Reich, Wendy Winckler, Michael S. Lawrence, Barbara A. Weir, Kumiko E. Tanaka, Derek Y. Chiang, Adam J. Bass, Alice Loo, Carter Hoffman, John Prensner, Ted Liefeld, Qing Gao, Derek Yecies, Sabina Signoretti, Elizabeth Maher, Frederic J. Kaye, Hidefumi Sasaki, Joel E. Tepper, Jonathan A. Fletcher, Josep Tabernero, José Baselga, Ming-Sound Tsao, Francesca Demichelis, Mark A. Rubin, Pasi A. Janne, Mark J. Daly, Carmelo Nucera, Ross L. Levine, Benjamin L. Ebert, Stacey Gabriel, Anil K. Rustgi, Cristina R. Antonescu, Marc Ladanyi, Anthony Letai, Levi A. Garraway, Massimo Loda, David G. Beer, Lawrence D. True, Aikou Okamoto, Scott L. Pomeroy, Samuel Singer, Todd R. Golub, Eric S. Lander, Gad Getz, William R. Sellers, and Matthew Meyerson. The landscape of somatic copy-number alteration across human cancers. *Nature*, 463(7283):899–905, feb 2010.

[19] Peter Priestley, Jonathan Baber, Martijn P. Lolkema, Neeltje Steeghs, Ewart de Bruijn, Charles Shale, Korneel Duyvesteyn, Susan Haidari, Arne van Hoeck, Wendy Onstenk, Paul Roepman, Mircea Voda, Haiko J. Bloemendal, Vivianne C. G. Tjan-Heijnen, Carla M. L.

van Herpen, Mariette Labots, Petronella O. Witteveen, Egbert F. Smit, Stefan Sleijfer, Emile E. Voest, and Edwin Cuppen. Pan-cancer whole-genome analyses of metastatic solid tumours. *Nature*, 575(7781):210–216, oct 2019.

[20] Bernard Leroy, Martha Anderson, and Thierry Soussi. Tp53 mutations in human cancer: Database reassessment and prospects for the next decade. *Human mutation*, 35:672–688, 2014.

[21] Niamh Coleman and Jordi Rodon. Taking aim at the undruggable. *American Society of Clinical Oncology educational book. American Society of Clinical Oncology. Annual Meeting*, pages e145–e152, 2021.

[22] Chi V. Dang, E. Premkumar Reddy, Kevan M. Shokat, and Laura Soucek. Drugging the 'undruggable' cancer targets. *Nature Reviews Cancer*, 17(8):502–508, aug 2017.

[23] Stanley T. Crooke, Joseph L. Witztum, C. Frank Bennett, and Brenda F. Baker. RNA-Targeted Therapeutics. *Cell Metabolism*, 27(4):714–739, apr 2018.

[24] Zhong Wang, Mark Gerstein, and Michael Snyder. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 10(1):57–63, jan 2009.

[25] Maria Gagliardi and Ana Tari Ashizawa. The Challenges and Strategies of Antisense Oligonucleotide Drug Delivery. *Biomedicines*, 9(4):433, apr 2021.

[26] Thomas C. Roberts, Robert Langer, and Matthew J. A. Wood. Advances in oligonucleotide drug delivery. *Nat Rev Drug Discov*, 19:673–694, 2020.

[27] Ryan L. Setten, John J. Rossi, and Si-ping Han. The current state and future directions of RNAi-based therapeutics. *Nature Reviews Drug Discovery*, 18(6):421–446, jun 2019.

[28] Anaïs M. Quemener, Laura Bachelot, Anne Forestier, Emmanuelle Donnou-Fournet, David Gilot, and Marie-Dominique Galibert. The powerful world of antisense oligonucleotides: From bench to bedside. *WIREs RNA*, 11(5):e1594, sep 2020.

[29] Mallory A. Havens and Michelle L. Hastings. Splice-switching antisense oligonucleotides as therapeutic drugs. *Nucleic Acids Research*, 44(14):6549–6563, aug 2016.

[30] Ryszard Kole, Adrian R. Krainer, and Sidney Altman. RNA therapeutics: beyond RNA interference and antisense oligonucleotides. *Nature Reviews Drug Discovery*, 11(2):125–140, feb 2012.

[31] C. L. Will and R. Luhrmann. Spliceosome Structure and Function. *Cold Spring Harbor Perspectives in Biology*, 3(7):a003707–a003707, jul 2011.

[32] Marina M. Scotti and Maurice S. Swanson. RNA mis-splicing in disease. *Nature Reviews Genetics*, 17(1):19–32, jan 2016.

[33] Annemieke Aartsma-Rus and Gert-Jan B. van Ommen. Antisense-mediated exon skipping: A versatile tool with therapeutic and research applications. *RNA (New York, N.Y.)*, 13:1609–1624, 2007.

[34] Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter. From RNA to Protein. *Molecular Biology of the Cell. 4th edition*, 2002.

[35] Guido Leoni, Loredana Le Pera, Fabrizio Ferrè, Domenico Raimondo, and Anna Tramontano. Coding potential of the products of alternative splicing in human. *Genome Biology*, 12(1):R9, jan 2011.

[36] Takenori Shimo, Rika Maruyama, and Toshifumi Yokota. Designing effective antisense oligonucleotides for exon skipping, 2018.

[37] Yeonjoon Kim. Drug discovery perspectives of antisense oligonucleotides. *Biomolecules & therapeutics*, 31:241–252, 2023.

[38] Lim KRQ, Maruyama R, and Yokota T. Eteplirsen in the treatment of duchenne muscular dystrophy. *Drug Design, Development and Therapy*, Volume11:533–545, 2 2017.

[39] Roshmi, R,R, and Yokota, T. Viltolarsen for the treatment of Duchenne muscular dystrophy. *Drugs of Today*, 55(10):627, 2019.

[40] Anwar, S. and Yokota, T. Golodirsen for Duchenne muscular dystrophy. *Drugs of Today*, 56(8):491, 2020.

[41] Matt Shirley. Casimersen: First Approval. *Drugs*, 81(7):875–879, may 2021.

[42] John C. Carter, Daniel W. Sheehan, Andre Prochoroff, and David J. Birnkrant. Muscular Dystrophies. *Clinics in Chest Medicine*, 39(2):377–389, jun 2018.

[43] Catherine L. Bladen, David Salgado, Soledad Monges, Maria E. Foncuberta, Kyriaki Kekou, Konstantina Kosma, Hugh Dawkins, Leanne Lamont, Anna J. Roy, Teodora Chamova, Velina Guergueltcheva, Sophelia Chan, Lawrence Korngut, Craig Campbell, Yi Dai, Jen Wang, Nina Barišić, Petr Brabec, Jaana Lahdetie, Maggie C. Walter, Olivia Schreiber-Katz, Veronika Karcagi, Marta Garami, Venkatarman Viswanathan, Farhad Bayat, Filippo Buccella, En Kimura, Zaïda Koeks, Janneke C. Van Den Bergen, Miriam Rodrigues, Richard Roxburgh, Anna Lusakowska, Anna Kostera-Pruszczyk, Janusz Zimowski, Rosário Santos, Elena Neagu, Svetlana Artemieva, Vedrana Milic Rasic, Dina Vojinovic, Manuel Posada, Clemens Bloetzer, Pierre-Yves Jeannet, Franziska Joncourt, Jordi Díaz-Manera, Eduard Gallardo, A. Ayşe Karaduman, Haluk Topaloğlu, Rasha El Sherif, Angela Stringer, Andriy V. Shatillo, Ann S. Martin, Holly L. Peay, Matthew I. Bellgard, Jan Kirschner, Kevin M. Flanigan, Volker Straub, Kate Bushby, Jan Verschuuren, Annemieke Aartsma-Rus, Christophe Béroud, and Hanns Lochmüller. The TREAT-NMD DMD Global Database: Analysis of More than 7,000 Duchenne Muscular Dystrophy Mutations. *Human Mutation*, 36(4):395–402, apr 2015.

[44] M. Koenig, E. P. Hoffman, C. J. Bertelson, A. P. Monaco, C. Feener, and L. M. Kunkel. Complete cloning of the duchenne muscular dystrophy (dmd) cdna and preliminary genomic organization of the dmd gene in normal and affected individuals. *Cell*, 50:509–517, 1987.

[45] Annemieke Aartsma-Rus, Judith C. T. Van Deutekom, Ivo F. Fokkema, Gert-Jan B. Van Ommen, and Johan T. Den Dunnen. En-

tries in the leiden duchenne muscular dystrophy mutation database: An overview of mutation types and paradoxical cases that confirm the reading-frame rule. *Muscle Nerve*, 34:135–144, 2006.

[46] Chengmei Sun, Luoan Shen, Zheng Zhang, and Xin Xie. Therapeutic Strategies for Duchenne Muscular Dystrophy: An Update. *Genes*, 11(8):837, jul 2020.

[47] Anthony P. Monaco, Corlee J. Bertelson, Sabina Liechti-Gallati, Hans Moser, and Louis M. Kunkel. An explanation for the phenotypic differences between patients bearing partial deletions of the dmd locus. *Genomics*, 2:90–95, 1988.

[48] Matsuo M. Antisense Oligonucleotide-Mediated Exon-skipping Therapies: Precision Medicine Spreading from Duchenne Muscular Dystrophy. *JMA Journal*, 4(3):232–240, 2021.

[49] Lucía Echevarría, Philippine Aupy, and Aurélie Goyenvalle. Exon-skipping advances for Duchenne muscular dystrophy. *Human Molecular Genetics*, 27(R2):R163–R172, aug 2018.

[50] Veronica Verdile, Gloria Guizzo, Gabriele Ferrante, and Maria Paola Paronetto. RNA Targeting in Inherited Neuromuscular Disorders: Novel Therapeutic Strategies to Counteract Mis-Splicing. *Cells*, 10(11):2850, oct 2021.

[51] P Adamo and M R Ladomery. The oncogene ERG: a key factor in prostate cancer. *Oncogene*, 35(4):403–414, jan 2016.

[52] Ling Li, Lisa Hobson, Laura Perry, Bethany Clark, Susan Heavey, Aiman Haider, Ashwin Sridhar, Greg Shaw, John Kelly, Alex Free-

man, Ian Wilson, Hayley Whitaker, Elmar Nurmemmedov, Sebastian Oltean, Sean Porazinski, and Michael Ladomery. Targeting the erg oncogene with splice-switching oligonucleotides as a novel therapeutic strategy in prostate cancer. *British journal of cancer*, 123:1024–1032, 2020.

[53] Wai Kit Ma, Dillon M. Voss, Juergen Scharner, Ana S. H. Costa, Kuan-Ting Lin, Hyun Yong Jeon, John E. Wilkinson, Michaela Jackson, Frank Rigo, C. Frank Bennett, and Adrian R. Krainer. Aso-based pkm splice-switching therapy inhibits hepatocellular carcinoma growth. *Cancer research*, 82:900–915, 2022.

[54] Volkher Scharnhorst, Alex J. van der Eb, and Aart G. Jochemsen. Wt1 proteins: functions in growth and differentiation. *Gene*, 273:141–161, 2001.

[55] Jane Renshaw, Rosanne M. Orr, Michael I. Walton, Robert te Poele, Richard D. Williams, Edward V. Wancewicz, Brett P. Monia, Paul Workman, and Kathryn Pritchard-Jones. Disruption of wt1 gene expression and exon 5 splicing following cytotoxic drug treatment: Antisense down-regulation of exon 5 alters target gene expression and inhibits cell survival. *Molecular cancer therapeutics*, 3:1467–1484, 2004.

[56] Warren A. Kibbe. Oligocalc: an online oligonucleotide properties calculator. *Nucleic Acids Res.*, 35(Web-Server-Issue):43–46, 2007.

[57] Ronny Lorenz, Stephan H. Bernhart, Christian Höner zu Siederdissen, Hakim Tafer, Christoph Flamm, Peter F. Stadler, and Ivo L. Hofacker. Viennarna package 2.0. *Algorithms Mol. Biol.*, 6:26, 2011.

[58] Jessica S. Reuter and David H. Mathews. Rnastructure: software for RNA secondary structure prediction and analysis. *BMC Bioinform.*, 11:129, 2010.

[59] Francesco Piva, Matteo Giulietti, Alessandra Ballone Burini, and Giovanni Principato. Spliceaid 2: A database of human splicing factors expression data and rna target motifs. *Human mutation*, 33:81–85, 2012.

[60] Simone Sciabola, Hualin Xi, Dario Cruz, Qing Cao, Christine Lawrence, Tianhong Zhang, Sergio Rotstein, Jason D. Hughes, Daniel R. Caffrey, and Robert V. Stanton. Pfred: A computational platform for sirna and antisense oligonucleotides design. *PLoS ONE*, 16(1):e0238753, jan 2021.

[61] Shuntaro Chiba, Kenji Rowel Q Lim, Narin Sheri, Saeed Anwar, Esra Erkut, Md Nur Ahad Shah, Tejal Aslesh, Stanley Woo, Omar Sheikh, Rika Maruyama, Hiroaki Takano, Katsuhiko Kunitake, William Duddy, Yasushi Okuno, Yoshitsugu Aoki, and Toshifumi Yokota. eSkip-Finder: a machine learning-based web application and database to identify the optimal sequences of antisense oligonucleotides for exon skipping. *Nucleic Acids Research*, 49(W1):W193–W198, jul 2021.

[62] Adam Frankish, Mark Diekhans, Irwin Jungreis, Julien Lagarde, Jane E. Loveland, Jonathan M. Mudge, Cristina Sisu, James C. Wright, Joel Armstrong, If Barnes, Andrew E. Berry, Alexandra Bignell, Carles Boix, Silvia Carbonell Sala, Fiona Cunningham, Tomás Di Domenico, Sarah M. Donaldson, Ian T. Fiddes, Car-

los García-Girón, Jose M. Gonzalez, Tiago Grego, Matthew Hardy, Thibaut Hourlier, Kevin L. Howe, Toby Hunt, Osagie G. Izuogu, Rory Johnson, Fergal J. Martin, Laura Martínez, Shamika Mohanan, Paul Muir, Fabio C. P. Navarro, Anne Parker, Baikang Pei, Fernando Pozo, Ferriol Calvet Riera, Magali Ruffier, Bianca M. Schmitt, Eloise Stapleton, Marie-Marthe Suner, Irina Sycheva, Barbara Uszczynska-Ratajczak, Maxim Y. Wolf, Jinrui Xu, Yucheng T. Yang, Andrew D. Yates, Daniel R. Zerbino, Yan Zhang, Jyoti Choudhary, Mark Gerstein, Roderic Guigó, Tim J. P. Hubbard, Manolis Kellis, Benedict Paten, Michael L. Tress, and Paul Flicek. GENCODE 2021. *Nucleic Acids Res.*, 49(Database-Issue):D916–D923, 2021.

[63] Paul A. Morcos. Achieving targeted and quantifiable alteration of mrna splicing with morpholino oligos. *Biochemical and biophysical research communications*, 358:521–527, 2007.

[64] Jon D. Moulton and Yi-Lin Yan. Using Morpholinos to Control Gene Expression. *Current Protocols in Molecular Biology*, 83(1), jul 2008.

[65] Nele Hug, Dasa Longman, and Javier F. Cáceres. Mechanism and regulation of the nonsense-mediated decay pathway. *Nucleic Acids Research*, 44(4):1483–1495, feb 2016.

[66] Loredana Le Pera, Paolo Marcatili, and Anna Tramontano. PICMI: mapping point mutations on genomes. *Bioinformatics*, 26(22):2904–2905, nov 2010.

[67] P. Andrew Futreal, Lachlan Coin, Mhairi Marshall, Thomas Down, Timothy Hubbard, Richard Wooster, Nazneen Rahman, and

Michael R. Stratton. A census of human cancer genes. *Nature Reviews Cancer*, 4(3):177–183, mar 2004.

[68] Collin J. Tokheim, Nickolas Papadopoulos, Kenneth W. Kinzler, Bert Vogelstein, and Rachel Karchin. Evaluating the evaluation of cancer driver genes. *Proceedings of the National Academy of Science*, 113(50):14330–14335, dec 2016.

[69] Ali Mortazavi, Brian A Williams, Kenneth McCue, Lorian Schaeffer, and Barbara Wold. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*, 5(7):621–628, jul 2008.

[70] ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. *Nature*, 578(7793):82–93, feb 2020.

[71] The GTEx Consortium, Kristin G. Ardlie, David S. Deluca, Ayellet V. Segrè, Timothy J. Sullivan, Taylor R. Young, Ellen T. Gelfand, Casandra A. Trowbridge, Julian B. Maller, Taru Tukiainen, Monkol Lek, Lucas D. Ward, Pouya Kheradpour, Benjamin Iriarte, Yan Meng, Cameron D. Palmer, Tõnu Esko, Wendy Winckler, Joel N. Hirschhorn, Manolis Kellis, Daniel G. MacArthur, Gad Getz, Andrey A. Shabalin, Gen Li, Yi-Hui Zhou, Andrew B. Nobel, Ivan Rusyn, Fred A. Wright, Tuuli Lappalainen, Pedro G. Ferreira, Halit Ongen, Manuel A. Rivas, Alexis Battle, Sara Mostafavi, Jean Monlong, Michael Sammeth, Marta Mele, Ferran Reverter, Jakob M. Goldmann, Daphne Koller, Roderic Guigó, Mark I. McCarthy, Emmanouil T. Dermitzakis, Eric R. Gamazon, Hae Kyung Im, Anuar Konkashbaev, Dan L. Nicolae, Nancy J. Cox, Timothée Flutre, Xi-

aoquan Wen, Matthew Stephens, Jonathan K. Pritchard, Zhidong Tu, Bin Zhang, Tao Huang, Quan Long, Luan Lin, Jialiang Yang, Jun Zhu, Jun Liu, Amanda Brown, Bernadette Mestichelli, Denee Tidwell, Edmund Lo, Mike Salvatore, Saboor Shad, Jeffrey A. Thomas, John T. Lonsdale, Michael T. Moser, Bryan M. Gillard, Ellen Karasik, Kimberly Ramsey, Christopher Choi, Barbara A. Foster, John Syron, Johnell Fleming, Harold Magazine, Rick Hasz, Gary D. Walters, Jason P. Bridge, Mark Miklos, Susan Sullivan, Laura K. Barker, Heather M. Traino, Maghboeba Mosavel, Laura A. Siminoff, Dana R. Valley, Daniel C. Rohrer, Scott D. Jewell, Philip A. Branton, Leslie H. Sobin, Mary Barcus, Liqun Qi, Jeffrey McLean, Pushpa Hariharan, Ki Sung Um, Shenpei Wu, David Tabor, Charles Shive, Anna M. Smith, Stephen A. Buia, Anita H. Undale, Karna L. Robinson, Nancy Roche, Kimberly M. Valentino, Angela Britton, Robin Burges, Debra Bradbury, Kenneth W. Hambright, John Seleski, Greg E. Korzeniewski, Kenyon Erickson, Yvonne Marcus, Jorge Tejada, Mehran Taherian, Chunrong Lu, Margaret Basile, Deborah C. Mash, Simona Volpi, Jeffery P. Struewing, Gary F. Temple, Joy Boyer, Deborah Colantuoni, Roger Little, Susan Koester, Latarsha J. Carithers, Helen M. Moore, Ping Guan, Carolyn Compton, Sherilyn J. Sawyer, Joanne P. Demchok, Jimmie B. Vaught, Chana A. Rabiner, Nicole C. Lockhart, Kristin G. Ardlie, Gad Getz, Fred A. Wright, Manolis Kellis, Simona Volpi, and Emmanouil T. Dermitzakis. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science*, 348(6235):648–660, may 2015.

[72] Abdullah Kahraman, Tülay Karakulak, Damian Szklarczyk, and Christian von Mering. Pathogenic impact of transcript isoform switching in 1,209 cancer samples covering 27 cancer types using an isoform-specific interaction network. *Sci Rep*, 10, 2020.

[73] Hu Zhuhong, Bai Zhenyu, Chen Xiangyuan, Xu Tingzhen, and Song Libin. Genome-wide isoform-level analysis reveals tumor-specific isoforms for lung adenocarcinoma diagnosis and prognosis. *Cancer Genet*, 230:58–65, 2019.

[74] Christian L. Barrett, Christopher DeBoever, Kristen Jepsen, Cheryl C. Saenz, Dennis A. Carson, and Kelly A. Frazer. Systematic transcriptome analysis reveals tumor-specific isoforms for ovarian cancer diagnosis and therapy. *Proc Natl Acad Sci U S A*, 112, 2015.

[75] Christopher J. Mann, Kaite Honeyman, Graham McClorey, Sue Fletcher, and Stephen D. Wilton. Improved antisense oligonucleotide induced exon skipping in themdx mouse model of muscular dystrophy. *The Journal of Gene Medicine*, 4(6):644–654, nov 2002.

[76] Soroush Najdaghi, Sepideh Razi, and Nima Rezaei. An overview of the role of interleukin-8 in colorectal cancer. *Cytokine*, 135:155205, 2020.

[77] Rik G. H. Lindeboom, Michiel Vermeulen, Ben Lehner, and Fran Supek. The impact of nonsense-mediated mrna decay on genetic disease, gene editing and cancer immunotherapy. *Nature Genetics*, 51:1645–1651, 2019.

[78] Eszter Nagy and Lynne E. Maquat. A rule for termination-codon position within intron-containing genes: when nonsense affects rna abundance. *Trends in biochemical sciences*, 23:198–199, 1998.

[79] Lucie Carrier, Saskia Schlossarek, Monte S. Willis, and Thomas Eschenhagen. The ubiquitin-proteasome system and nonsense-mediated mrna decay in hypertrophic cardiomyopathy. *Cardiovascular research*, 85:330–338, January 2010.

[80] Min-Kung Hsu, Hsuan-Yu Lin, and Feng-Chi Chen. Nmd classifier: A reliable and systematic classification tool for nonsense-mediated decay events. *PloS one*, 12:e0174798, 2017.

[81] Maximilian W. Popp and Lynne E. Maquat. Leveraging rules of nonsense-mediated mrna decay for genome engineering and personalized medicine. *Cell*, 165:1319–1322, 2016.

[82] Fran Supek, Ben Lehner, and Rik G. H. Lindeboom. To nmd or not to nmd: Nonsense-mediated mrna decay in cancer and other genetic diseases. *Trends Genet.*, 37:657–668, 2021.

[83] Kun Tan, Dwayne G. Stupack, and Miles F. Wilkinson. Nonsense-mediated rna decay: an emerging modulator of malignancy. *Nat Rev Cancer*, 22:437–451, 2022.

[84] Francesco Piva, Matteo Giulietti, Linda Nocchi, and Giovanni Principato. Spliceaid: a database of experimental RNA target motifs bound by splicing proteins in humans. *Bioinform.*, 25(9):1211–1213, 2009.

[85] K J Breslauer, R Frank, H Blöcker, and L A Marky. Predicting DNA

duplex stability from the base sequence. *Proceedings of the National Academy of Sciences*, 83(11):3746–3750, jun 1986.

[86] Naoki Sugimoto, Shu-ich Nakano, Mari Yoneyama, and Kei-ich Honda. Improved Thermodynamic Parameters and Helix Initiation Factor to Predict Stability of DNA Duplexes. *Nucleic Acids Research*, 24(22):4501–4505, nov 1996.

[87] Tianbing Xia, John SantaLucia, Mark E. Burkard, Ryszard Kierzek, Susan J. Schroeder, Xiaoqi Jiao, Christopher Cox, and Douglas H. Turner. Thermodynamic Parameters for an Expanded Nearest-Neighbor Model for Formation of RNA Duplexes with Watson-Crick Base Pairs. *Biochemistry*, 37(42):14719–14735, oct 1998.

[88] Ben Langmead, Cole Trapnell, Mihai Pop, and Steven L. Salzberg. Ultrafast and memory-efficient alignment of short dna sequences to the human genome. *Genome Biol*, 10:R25, 2009.

[89] The International Cancer Genome Consortium. International network of cancer genome projects. *Nature*, 464(7291):993–998, apr 2010.

[90] Tijana Randic, Ines Kozar, Christiane Margue, Jochen Utikal, and Stephanie Kreis. NRAS mutant melanoma: Towards better therapies. *Cancer Treatment Reviews*, 99:102238, sep 2021.

[91] Ling Qian, Kun Chen, Changhong Wang, Zhen Chen, Zhiqiang Meng, and Peng Wang. Targeting NRAS-Mutant Cancers with the Selective STK19 Kinase Inhibitor Chelidonine. *Clinical Cancer Research*, 26(13):3408–3419, jul 2020.

[92] Lucy Gossage, Tim Eisen, and Eamonn R. Maher. VHL, the story of a tumour suppressor gene. *Nature Reviews Cancer*, 15(1):55–64, jan 2015.

[93] Hajime Takamori, Toshinari Yamasaki, Rui Kitadai, Yoji Andrew Minamishima, and Eijiro Nakamura. Development of drugs targeting hypoxia-inducible factor against tumor cells with *VHL* mutation: Story of 127 years. *Cancer Science*, 114(4):1208–1217, apr 2023.

[94] Zaman, Wu, and Bivona. Targeting Oncogenic BRAF: Past, Present, and Future. *Cancers*, 11(8):1197, aug 2019.

[95] Ori Hassin and Moshe Oren. Drugging p53 in cancer: one protein, many targets. *Nature Reviews Drug Discovery*, 22(2):127–144, feb 2023.

[96] Chiara Bazzichetto, Michele Milella, Ilaria Zampiva, Francesca Simionato, Carla Azzurra Amoreo, Simonetta Buglioni, Chiara Pacelli, Loredana Le Pera, Teresa Colombo, Emilio Bria, Massimo Zeuli, Donatella Del Bufalo, Isabella Sperduti, and Fabiana Conciatori. Interleukin-8 in colorectal cancer: A systematic review and meta-analysis of its potential role as a prognostic biomarker. *Biomedicines*, 10:2631, 2022.

[97] Yoshiro Itatani, Kenji Kawada, Susumu Inamoto, Takamasa Yamamoto, Ryotaro Ogawa, Makoto Taketo, and Yoshiharu Sakai. The Role of Chemokines in Promoting Colorectal Cancer Invasion/Metastasis. *International Journal of Molecular Sciences*, 17(5):643, apr 2016.

[98] Timothy Sterne-Weiler, Jonathan Howard, Matthew Mort, David N. Cooper, and Jeremy R. Sanford. Loss of exon identity is a common mechanism of human inherited disease. *Genome Res*, 21:1563–1571, 2011.

[99] I. Garcia-Consuegra, J. C. Rubio, G. Nogales-Gadea, J. Bautista, S. Jimenez, A. Cabello, A. Lucia, A. L. Andreu, J. Arenas, and M. A. Martin. Novel mutations in patients with mcardle disease by analysis of skeletal muscle mrna. *J Med Genet*, 46:198–202, 2009.

[100] Yuichi Shiraishi, Ai Okada, Kenichi Chiba, Asuka Kawachi, Ikuko Omori, Raúl Nicolás Mateos, Naoko Iida, Hirofumi Yamauchi, Kenjiro Kosaki, and Akihide Yoshimi. Systematic identification of intron retention associated variants from massive publicly available transcriptome sequencing data. *Nature Communications*, 13(1):1–13, 9 2022.

[101] Ruebena Dawes, Adam M. Bournazos, Samantha J. Bryen, Shobhana Bommireddipalli, Rhett G. Marchant, Himanshu Joshi, and Sandra T. Cooper. Splicevault predicts the precise nature of variant-associated mis-splicing. *Nat Genet*, 55:324–332, 2023.

[102] François-Olivier Desmet, Dalil Hamroun, Marine Lalande, Gwenaëlle Collod-Béroud, Mireille Claustres, and Christophe Béroud. Human splicing finder: an online bioinformatics tool to predict splicing signals. *Nucleic acids research*, 37:e67–e67, 2009.

[103] K. K. Nelson and M. R. Green. Mechanism for cryptic splice site activation during pre-mrna splicing. *Proc Natl Acad Sci U S A*, 87:6253–6257, 1990.

[104] Thomas C. Roberts, Matthew J. A. Wood, and Kay E. Davies. Therapeutic approaches for Duchenne muscular dystrophy. *Nature Reviews Drug Discovery*, aug 2023.

[105] Steven F Dowdy. Overcoming cellular barriers for RNA therapeutics. *Nature Biotechnology*, 35(3):222–229, mar 2017.

[106] Thazha P Prakash, Adam E Mullick, Richard G Lee, Jinghua Yu, Steve T Yeh, Audrey Low, Alfred E Chappell, Michael E Østergaard, Sue Murray, Hans J Gaus, Eric E Swayze, and Punit P Seth. Fatty acid conjugation enhances potency of antisense oligonucleotides in muscle. *Nucleic Acids Research*, 47(12):6029–6044, jul 2019.

[107] Mohammad Shadid, Mohamed Badawi, and Abedelnasser Abulrob. Antisense oligonucleotides: absorption, distribution, metabolism, and excretion. *Expert Opinion on Drug Metabolism & Toxicology*, 17(11):1281–1292, nov 2021.

[108] Andrea Alimonti, Arkaitz Carracedo, John G Clohessy, Lloyd C Trotman, Caterina Nardella, Ainara Egia, Leonardo Salmena, Katia Sampieri, William J Haveman, Edi Brogi, Andrea L Richardson, Jiangwen Zhang, and Pier Paolo Pandolfi. Subtle variations in Pten dose determine cancer susceptibility. *Nature Genetics*, 42(5):454–458, may 2010.

Acknowledgements

Special thanks to my thesis advisors Teresa Colombo and Loredana Le Pera for teaching, supporting and helping me all the time, making this possible. Thanks also to Prof. Malatesta who always had a guardian look on my path, rewarding my every effort. Thanks to bioinformatic colleagues Veronica Morea, Allegra Via and Gianmarco. I would like to thank Prof. Michele Milella from the University of Verona for his collaboration and contribution and the group of the Regina Elena National Cancer Institute of Rome for their collaboration.