U.PORTO

MESTRADO EM ONCOLOGIA

ONCOLOGIA LABORATORIAL

# NOVEL PREDICTIVE MARKERS OF THERAPY RESPONSE IN GASTRIC CANCER

SILVANA PATRÍCIA LOBO FERREIRA DE SOUSA

**M**

**2019**

U.PORTO

Silvana Lobo. NOVEL PREDICTIVE MARKERS OF THERAPY RESPONSE IN GASTRIC CANCER

**M**.ICBAS **2019**

NOVEL PREDICTIVE MARKERS OF THERAPY RESPONSE IN GASTRIC CANCER

SILVANA PATRÍCIA LOBO FERREIRA DE SOUSA

INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR

Silvana Patrícia Lobo Ferreira de Sousa

# Novel predictive markers of therapy response in Gastric Cancer

Tese de candidatura ao grau de Mestre em Oncologia Molecular submetida ao Instituto de Ciências Biomédicas Abel Salazar da Universidade do Porto


| | |
|---|---|
| Orientador | Doutora Gabriela Martinho de Almeida |
| Categoria | Investigadora Auxiliar e Professora Afiliada |
| Afiliação | Instituto de Investigação e Inovação em Saúde / Instituto de Patologia e Imunologia Molecular da Universidade do Porto<br>Faculdade de Medicina da Universidade do Porto |
| Coorientador | Doutora Carla Oliveira |
| Categoria | Investigadora Principal e Professora Afiliada |
| Afiliação | Instituto de Investigação e Inovação em Saúde / Instituto de Patologia e Imunologia Molecular da Universidade do Porto<br>Faculdade de Medicina da Universidade do Porto |

## Financial Support

The project NORTE-01-0145-FEDER-000029 “Mapping genetic and phenotypic heterogeneity in HER2 positive cancers to anticipate and counteract resistance phenotypes”, supported by Norte Portugal Regional Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF).

***"The ultimate measure of a man is not where he stands in moments of comfort and convenience, but where he stands at times of challenges and controversy"***

***Martin Luther King***

# Acknowledgements/Agradecimentos

Este último ano foi repleto de novas experiências e aquisição de conhecimentos que levarei para o resto da minha vida pessoal e profissional. E isto deveu-se em grande parte devido a certas pessoas que me ajudaram e às quais muito agradeço.

Em primeiro lugar quer de agradecer à Dra. Carla Oliveira por me ter aceite no seu grupo de investigação e me permitir realizar a minha dissertação de mestrado no mesmo. Agradeço imenso a oportunidade dada de desenvolver uma dissertação tão ambiciosa e desafiante e que me permitiu adquirir tantos conhecimentos que me vão ser extremamente úteis na minha carreira. Agradeço também à Dra. Gabriela M. Almeida por me ter orientado durante toda a minha dissertação. Obrigada por todos os ensinamentos, por todos os conselhos, todas as sugestões e todas as críticas dadas, que me ajudaram imenso na concretização desta dissertação.

Queria agradecer também à Carla Pereira por durante toda esta dissertação se ter mostrado 100% disponível para me ensinar e me ajudar com tudo o que fosse necessário. Mesmo estando super ocupada a acabar a sua própria tese de doutoramento, sempre que eu precisava ela estava lá para me ajudar.

Agradeço imenso a todos os membros do grupo ERiC por me terem integrado tão bem no grupo e por me terem ensinado tanto durante este ano. Convosco aprendi como trabalhar em equipa num laboratório e muito mais. Agradeço-vos todas as memórias criadas durante este ano (tanto no laboratório como fora dele).

Não posso deixar de agradecer ao meu grupo de amigos que me acompanha mesmo antes de ter entrado na faculdade. Joana, Teresa, Bruno, Ruben, Vítor, Paulo e Pacheco, apesar de todos termos seguido caminhos diferentes na faculdade, estivemos sempre lá uns para os outros.

Às minhas “oncogirls”, Sara, Rute, Rita e Márcia, estes dois anos de mestrado definitivamente não teriam sido tão bons se não vos tivesse ao meu lado para todas as alegrias e principalmente para todo o stress que foi este ano! Não posso deixar de agradecer ao madeirense, Augusto, pela alegria e boa disposição durante este ano no i3s.

Às minhas deprimidas e deprimido, Ana, Marisa, Ricardina e Eduardo, estes cinco anos na faculdade foram muito melhores do que poderia esperar, em grande parte graças a vós. Obrigada por todos os jantares, serões de “Catacumbas”, conversas, festas… Não conseguiria imaginar o meu percurso sem vós.

Não posso deixar de agradecer à pessoa que mais me apoiou durante todo esta dissertação, ao Jorge Diogo, que desde o início esteve lá para mim sempre que necessário. Por me ajudar, me aconselhar, me chamar à atenção quando necessário e principalmente por me aturar nos momentos de stress sempre com um sorriso na cara e por não me deixar ir abaixo quando tudo corre mal. Tenho a certeza que não teria chegado ao fim desta dissertação sem ti.

Por último, não poderia deixar de agradecer aos meus pais. Sem os ensinamentos que me passaram e a educação e moral que me deram, não seria a pessoa que sou hoje. É graças a vós que tive a oportunidade de estudar Biologia, e é graças a vós que termino esta dissertação. Em 23 anos de vida sempre fizeram os possíveis e impossíveis para que eu atingisse os meus sonhos, e agradeço-vos do fundo do coração por tudo.

# Abstract

Gastric cancer (GC) is the 5th most common cancer and the 3rd with the highest mortality, worldwide. Almost 80% of GC patients are only diagnosed at advanced stage of disease, where tumors are usually inoperable and the standard of care is conventional chemotherapy, with an overall survival of ~1 year. Therefore, it is imperative to improve both GC stratification and therapy. CD44 is a cell adhesion molecule codified by *CD44* gene that undergoes extensive alternative splicing, resulting in the production of many different isoforms. CD44v6-containing isoforms are all the isoforms that contain the variant exon 6 and have been often associated with cancer development and aggressiveness. A previous study conducted by our group have shown that *de novo* expression of CD44v6 is a poor prognosis factor in GC, but also a likely modulator of response to conventional chemotherapy in GC cell lines. Relying on the concept of exon skipping, we aimed to develop models of GC cell lines where exon v6 is skipped from CD44v6-containing isoforms using PMOs and the CRISPR/Cas9 system, to further explore the role of CD44 exon v6, by itself, in the response to chemotherapy in GC cell lines.

We used two main approaches, PMOs and CRISPR/Cas9, to specifically remove exon v6 from CD44 gene, transiently and permanently, in two GC cell lines that express *CD44v6* endogenously, whilst maintaining the reading frame. We genotyped and characterized the total *CD44* and *CD44v6* expression of the edited cell lines and *wild-type* controls. Afterwards, we used the permanently edited cell lines by CRISPR/Cas9 for treatments with chemotherapeutic agents. Cell survival was evaluated in short and long-term treatments.

Using PMOs, we obtained transient GC cells that skipped exon v6, splicing together exon v5 and exon v7. Regarding CRISPR/Cas9 approach, we obtained various homozygous clones lacking exon v6 only. To understand the role of exon v6 in response to chemotherapy, we tested cell survival of GC cells expressing CD4v6- isoforms, the same cell lines either submitted to CD44 exon v6 RNAi leading to depletion of all CD44v6-containing isoforms or lacking specifically exon v6 due to CRISPR/Cas9 editing. Results suggest that depletion of all CD44v6-containing isoforms by RNAi increases cell survival in short-term treatments, while in long-term treatments, cell survival is decreased, when comparing to CD44v6 expressing cells. Among edited clones, no consistent differences were observed in cell survival between clones with or without exon v6. These data support that depletion of all CD44v6-containing isoforms modulate cell response to chemotherapy, while exon v6 removal from CD44v6-containing isoforms has no effect in response to chemotherapy in GC cell lines.

In conclusion, results obtained throughout this project suggest that exon v6 is not responsible for the modulation of chemotherapy response associated with CD44v6-containing isoforms in GC cells.

# Resumo

O cancro gástrico (GC) é o 5º tipo de cancro mais comum e o 3º com maior mortalidade, globalmente. Aproximadamente 80% dos doentes com GC são diagnosticados apenas em estadio avançado da doença, onde os tumores são geralmente inoperáveis e o tratamento de primeira linha é a quimioterapia convencional, com uma sobrevida de ~1 ano. Portanto, é imperativo melhorar a estratificação e a terapia dos doentes com GC. O CD44 é uma molécula de adesão celular codificada pelo gene *CD44*, que é sujeito a um extensivo *splicing* alternativo, resultando na produção de diversas isoformas. As isoformas CD44v6 são todas as isoformas que contêm o exão variante 6 e têm sido frequentemente associadas ao desenvolvimento e agressividade de vários tipos de cancro. Um estudo anteriormente realizado pelo nosso grupo mostrou que o CD44v6 é expresso *de novo* em lesões pré-malignas e malignas de GC e que é um fator de mau prognóstico em doentes com GC, no entanto, é também um provável modelador de resposta à quimioterapia convencional em linhas celulares de GC. Recorrendo ao conceito de *skipping* de exão, o nosso objetivo era desenvolver modelos de linhas celulares de GC onde é feito o *skipping* do exão v6 das isoformas CD44v6, usando as tecnologias de PMOs e CRISPR / Cas9, para explorar qual é o papel do exão v6 do gene CD44, por si só, na modulação da resposta à quimioterapia em linhas celulares de GC.

Para alcançar o nosso objetivo, usamos duas abordagens principais, PMOs e CRISPR/Cas9, para remover especificamente o exão v6 do gene *CD44*, de forma transiente e permanente, em duas linhas celulares de GC que expressam *CD44v6* endogenamente, mantendo a grelha de leitura. Caracterizamos a expressão do *CD44* total e *CD44v6* nas células editadas e linhas parentais. Posteriormente, usamos as linhas celulares editadas permanentemente por CRISPR/Cas9 para tratar com agentes quimioterapêuticos. A sobrevivência celular foi avaliada em tratamentos de curta e longa duração.

Usando PMOs, obtivemos linhas celulares de GC que transientemente inibem o exão v6, fazendo o *splicing* conjunto do exão v5 e v7. Em relação à técnica de CRISPR/Cas9, obtivemos vários clones homozigóticos sem o exão v6. De modo a investigar o papel do exão v6 na resposta à quimioterapia, foi testada a sobrevivência celular de células de GC que expressam isoformas CD44v6 e as mesmas linhas celulares submetidas a RNAi para o exão v6 do *CD44*, levando a que todas as isoformas CD44v6 fossem depletadas, ou especificamente sem o exão v6 editado por CRISPR/Cas9. Os resultados sugerem que a depleção de todas as isoformas CD44v6 por RNAi aumenta a sobrevivência celular em tratamentos de curta duração, enquanto que em tratamentos de longa duração, a

sobrevivência celular diminui quando comparada às células que expressam CD44v6. Entre os clones editados, não foram observadas diferenças consistentes na alteração da sobrevivência celular entre os clones e células com e sem CD44v6. Estes dados suportam que a depleção das isoformas CD44v6 modula a resposta celular à quimioterapia, enquanto que a remoção do exão v6 das isoformas CD44v6 não afeta a resposta à quimioterapia em linhas celulares de GC.

Em conclusão, os resultados obtidos ao longo deste projeto sugerem que o exão v6 não é responsável pela modulação da resposta à quimioterapia associada às isoformas CD44v6 em células de GC.

# Table of Contents

# List of abbreviations

| | |
|---|---|
| 5-FU | 5-Fluorouracil |
| Abs | Absorbance |
| Acp | Acceptor |
| ANOVA | Analysis of variance |
| AONs | Antisense oligonucleotides |
| BMD | Becker Muscular Dystrophy |
| BP | Branch Point |
| bp | base pair |
| BSA | Bovine Serum Albumin |
| Cas | CRISPR associated |
| CD44 | Cluster of Differentiation 44 |
| CD44e | CD44 (epithelial isoform) |
| CD44s | CD44 (standard isoform) |
| CD44v | CD44 (variant isoforms) |
| CD44v3 | CD44 (isoforms containing exon v3) |
| CD44v4-v7 | CD44 (isoforms containing exon v4 to v7) |
| CD44v6 | CD44 (isoforms containing exon v6) |
| CD44v9 | CD44 (isoforms containing exon v9) |
| cDNA | complementary DNA |
| cMET | tyrosine kinase Met |
| CRISPR | Clustered Regularly Interspaced Short Palindromic Repeats |
| crRNA | CRISPR RNA |
| CSCs | Cancer Stem Cells |
| DMD | Duchenne Muscular Dystrophy |
| DNA | Deoxyribonucleic acid |
| Don | Donor |
| DSB | Double Strand Break |
| DSS | Disease Specific Survival |
| ELISA | Enzyme-linked immunosorbent assay |
| EP | Endo-Porter |
| ERM | Ezrin, Radixin and Moesin |
| FBS | Fetal Bovine Serum |
| FDA | Food and Drug Administration |
| Fluo | Fluorescence |
| GC | Gastric Cancer |

| | |
|---|---|
| gDNA | genomic DNA |
| GNS | Gold Nanostars |
| h | hours |
| HA | Hyaluronan |
| HDR | Homology Directed Repair |
| HER2 | Human Epidermal growth factor Receptor 2 |
| HGF | Hepatocyte Growth Factor |
| LB | Lysogeny Broth |
| mAb | monoclonal Antibody |
| min | minutes |
| mRNA | Messenger RNA |
| NA | Not Available |
| NHEJ | Non-Homologous End Joining |
| NSCLC | Non-Small Cell Lung Cancer |
| ON | Overnight |
| OS | Overall Survival |
| PAM | Protospacer Adjacent Motif |
| PB | PrestoBlue |
| PBS | Phosphate-Buffered Saline |
| PCR | Polymerase Chain Reaction |
| PD-1 | Programmed cell Death-1 |
| PEG | Polyethylene Glycol |
| PFA | Paraformaldehyde |
| PMOs | Phosphorodiamidate Morpholino oligomers |
| qRT-PCR | quantitative Real-Time PCR |
| RNA | Ribonucleic Acid |
| RNAi | RNA interference |
| RPMI | Roswell Park Memorial Institute |
| RT | Reverse Transcriptase |
| SD | Standard Deviation |
| sec | seconds |
| sgRNA | single guide RNA |
| siRNA | short interference RNA |
| SMA | Spinal Muscular Atrophy |
| SMN | Survival Motor Neuron |
| SOC | Super Optimal broth with Catabolite repression medium |

| | |
|---|---|
| SRB | Sulforhodamine B |
| TCA | Trichloroacetic Acid |
| TGF-β2 | Transforming Growth Factor-Beta 2 |
| TNM | Tumor, Node, Metastasis |
| tracrRNA | Transactivating CRISPR RNA |
| VEGF | Vascular Endothelial Growth Factor |
| VEGFR-2 | Vascular Endothelial Growth Factor Receptor 2 |
| *wt* | wild type |

# Figure Index

# Table Index

# Chapter I | Introduction

# 1. Gastric Cancer (GC)

Cancer is considered a major health problem around the globe. The World Health Organization estimated, in 2015, that cancer was the first or second leading cause of death in people with less than 70 years in more than half of the countries in the world. A complex combination of factors, such as aging and growth of the world's population and changes in prevalence and distribution of the main cancer risk factors contribute to the rising cancer incidence and mortality worldwide (1).

One of the greatest contributors to this health burden is GC. Even though GC incidence has declined in the past fifty years (2), according to the International Agency for Research in Cancer, in 2018, more than one million people were diagnosed with GC and 780 000 people perished from this disease, making GC the 5$^{th}$ most common type of cancer and the 3$^{rd}$ leading cause of cancer-related deaths, globally (3).

GC has a characteristic distribution around the world, with Eastern Asia and South and Eastern Europe having the highest rates regarding both GC incidence and mortality. Concerning gender, men are more susceptible to GC when compared to women, as GC incidence and mortality rates in men are 7.5% and 9.2%, respectively, while in women these rates decrease to 4.1% and 6.5%, respectively, regarding all types of cancers (3).

## 1.1. GC diagnosis and staging

GC is a silent disease at initial stages due to the non-specific symptoms that most patients exhibit, with merely vague stomach aches or abdominal pain and, consequently, approximately 80% of GC patients are only diagnosed at advanced stages of disease. At these stages, the ambiguous symptoms may evolve to vomits, anorexia, weight loss, difficulty in swallowing, excessive belching and increased tiredness (4, 5).

The diagnosis of gastric tumors is performed by endoscopy and biopsy (6) and, afterwards, staging is an imperative step to properly stratify GC patients in order to deliver the best personalized treatment to each patient. Tumor, nodes and metastases (TNM) classification system is the method of choice regarding staging of GC patients. The TNM system focus on tumor size and extent (T), presence or absence of lymph node invasion (N) and on presence or absence of distant metastases (M) (7). To classify a gastric tumor according to their stage, the techniques of choice are endoscopic ultrasonography and compute tomography scan of the chest and abdomen (6). The combination of the three TNM components results in a final staging ranging from 0 to IV (7), as shown in Table 1. Gastric tumor development is schematically represented in Figure 1.

Table 1 – GC TNM staging overview (Modified after reference (7)).

| Stage Grouping | T stage | N stage | M stage |
|---|---|---|---|
| **Stage 0** | Tis | N0 | |
| **Stage 1A** | T1 | N0 | |
| **Stage 1B** | T2 | N0 | |
| | T1 | N1 | |
| **Stage IIA** | T3 | N0 | |
| | T2 | N1 | |
| | T1 | N2 | |
| **Stage IIB** | T4a | N0 | |
| | T3 | N1 | |
| | T2 | N2 | M0 |
| | T1 | N3 | |
| **Stage IIIA** | T4a | N1 | |
| | T3 | N2 | |
| | T2 | N3 | |
| **Stage IIIB** | T4b | N0 – 1 | |
| | T4a | N2 | |
| | T3a | N3 | |
| **Stage IIIC** | T4b | N2 – 3 | |
| | T4a | N3 | |
| **Stage IV** | Any T | Any N | M1 |

T1 - tumor growth into the inner layers of the stomach wall, T2 - tumor growth into the muscle layer of the stomach wall, T3 - tumor growth into the outer lining of the stomach wall, T4 - tumor growth beyond the outer lining of the stomach wall. N0 – No cancer cells are encountered in lymph nodes, N1 – There are cancer cells in 1-2 lymph nodes close to the stomach, N2 - There are cancer cells in 3-6 lymph nodes close to the stomach, N3 - There are cancer cells in more than 7 lymph nodes close to the stomach. M0 – GC has not spread into distant organs, M1 - GC has spread into distant organs.

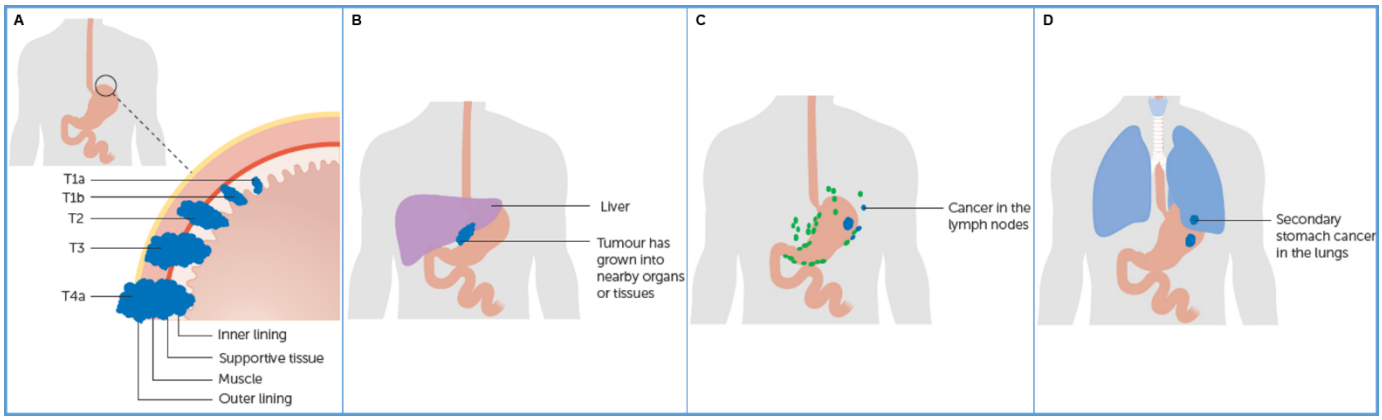


Figure 1 - GC TNM staging (A) As the tumor grows and invades the outer layers of the stomach, the T increases; (B) When the tumor has grown so big that invades the nearby organs, it is classified as T4b; (C) The N increases with the increase of invaded lymph nodes; (D) The M0 changes to M1 when distant metastases are discovered (Modified from Cancer Research UK: https://www.cancerresearchuk.org/about-cancer/stomach-cancer/stages/tnm-staging).

### 1.2. GC patient management

The management of gastric tumors relies on their staging and on the fitness of the patients. The practical recommendations for management of local and locoregional disease, which comprises tumors from stage I to stage III, is surgery. For early diagnosed tumors (stage IA), endoscopic resection of the tumor is the best approach. For patients with stage IB – III tumors, the indication if for a radical gastrectomy in combination with pre- and postoperative chemotherapy. Regarding both advanced and metastatic GC, the recommendations for first-line treatments are doublet or triplet platinum/fluoropyrimidine based combinations. These are very aggressive treatments and, therefore, are only indicated for fit patients. These treatments demonstrated to improve both survival and quality of life when compared to the best supportive care alone, leading to a median overall patient survival of ~1 year (8). In recent years, targeted therapies have been introduced for the treatment of GC patients. The first targeted therapy approved by the Food and Drug Administration (FDA) for the treatment of GC was the anti-HER2 Trastuzumab monoclonal antibody for patients with HER2 overexpression, in 2010. In the ToGa phase 3 trial, it was shown that combining Trastuzumab with chemotherapy improved survival in advanced GC patients overexpressing HER2, when comparing to patients treated with chemotherapy alone (9). Afterwards, in 2014 the anti-VEGFR-2 monoclonal antibody Ramucirumab was also approved by the FDA for the treatment of unresectable and metastatic GC. The RAINBOW phase 3 trial proved that Ramucirumab combined with Paclitaxel improved both overall and progression-free survival when compared to Paclitaxel alone (10). These are the approved target therapies for GC management. However, these GC targeted therapies have showed a limited overall survival (OS) improvement, of only 2 to 4 months (9, 10)

#### 1.2.1. Resistance to GC therapy

The success of GC treatment mainly relies on the stage at diagnosis. Radical surgery is the only curative treatment, however, approximately 80% of GC patients are diagnosed at an advanced stage of disease, where tumors are usually unresectable and chemotherapy is the main treatment of choice. Nevertheless, the success rate of chemotherapy is small and eventually most GC patients relapse or do not even respond (11). This therapy failure can occur due to pre-existent resistance factors present in the tumor (intrinsic resistance) or due to an acquired resistance to the drugs used in the chemotherapy regimens (12, 13). The intrinsic resistance might be a consequence of host factors, such as poor absorption or rapid metabolism that lead to the excretion of the drug, or specific alterations in cancer cells at the genetic and epigenetic level. On the other hand, the most common reasons for

acquired cancer drug resistance are the overexpression of energy-dependent transporters (that are capable of removing cancer therapy agents from cells), overexpression of DNA repair enzymes (that allow cancer cells to repair the drug-induced DNA damage more efficiently) or overexpression of anti-apoptotic molecules, that reduce the levels of drug-induced apoptosis in cancer cells (12).

### 1.3. Biomarkers in GC: State of art

A biomarker can be defined as a measurable characteristic that designates a certain biological state as normal or irregular and can be many types of biomolecules, such as DNA, RNA, proteins, peptides, among others (14). In the past years, biomarkers have gained a relevant role in cancer. There are biomarkers that can be used to properly define specific cancer diagnosis (diagnostic biomarkers), provide important information about the course of the disease (prognostic biomarkers) and predict the most likely response to therapy of certain patients (predictive biomarkers) (4).

Regarding GC, few are the biomarkers already used in the clinical practice. The use of TNM system at diagnosis is still the best way to determine patient's prognosis and decide the best treatment for GC patients (4, 15). The histologic type is also a prognostic factor for GC patients, where intestinal type or well differentiated is associated with a better survival after 5-years when compared to diffuse type or poorly differentiated in advanced GC, while in early GC the intestinal type presents the worst outcome (16).

## 2. CD44 in GC

### 2.1. *CD44* gene and protein

Cluster of Differentiation 44 (CD44) is a family of glycoproteins that are encoded by one single gene, the *CD44* gene, which is constituted by 20 exons in mice and 19 exons in humans. The human *CD44* gene has been mapped and it is located in the chromosomal locus 11p13, while in mice, it is located in chromosome 2 (17, 18). As represented in Figure 2, this gene is constituted by a constant part from exon 1-5 and exon 16-20 and a variant part from exon 6-15 (also known as v1-v10), that is regulated by an intensive alternative splicing, resulting in many different transcripts. The different formed transcripts result in different protein isoforms that are constituted by an identical cytoplasmic and transmembrane domain, however, its extracellular domain differs between isoforms, since the translation of the variant exons result in the addition of amino acids to a specific part of the extracellular domain (17, 19, 20), as shown in Figure 2.

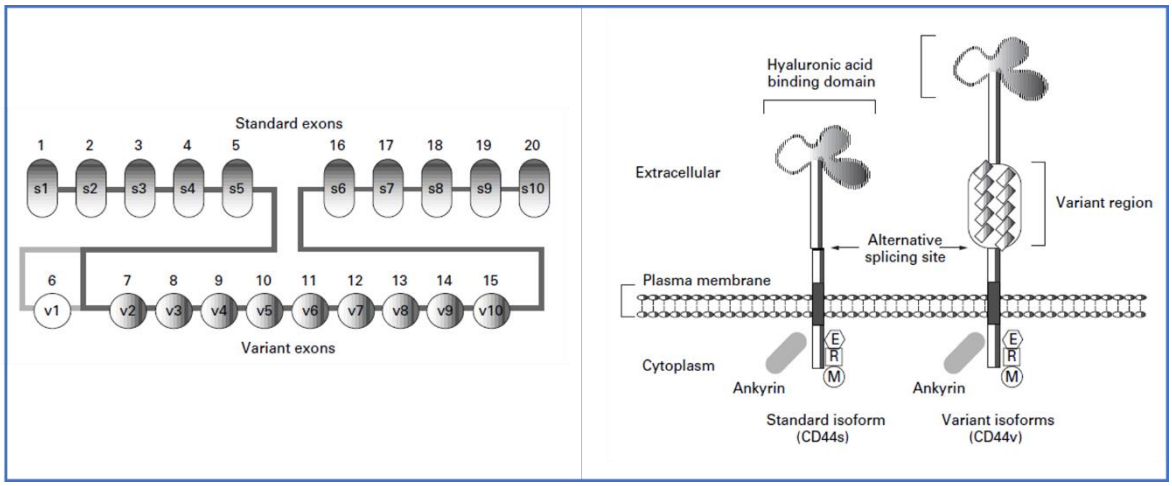


Figure 2 - The CD44 gene composition and its different isoforms. Human CD44 gene is composed of 10 contitutively transcribed exon (S1 to S10) and 9 variant exons (V2 to V10), which are inserted between exon S5 and S6 through alternative splicing (Modified after reference (17)).

This molecule has been widely studied over the years. It was first described as a lymphocyte homing receptor (21), however, nowadays many are the functions to which *CD44* is associated. This cell adhesion molecule controls cell behavior by mediating cell-cell and cell-matrix contact, being essential for the tissue integrity and maintenance. It is also included in lymphocyte activation, circulation and homing, and hematopoiesis. CD44 binds primarily to hyaluronan (HA) through its extracellular domain, which is important to the maintenance of the three-dimensional structure of the tissues, to the proliferation of epithelial cells, and to cell repair (17). CD44 also binds to other ligands, such as ankyrin and ezrin-radixin-moesin (ERM) through its cytoplasmic tail (18).

### 2.1.1. Alternative splicing and variant isoforms

As mentioned above, *CD44* gene is subjected to an intensive alternative splicing (17-20), a mechanism of genetic regulation and variation in higher eukaryotes, which results in a greater source of protein diversity from the genetic code (22). The alternative splicing patterns are responsible for including only certain portions of the coding sequence in the mRNA, after transcription. This process results in many protein isoforms that vary in peptide sequence and, consequently, in chemical and biological activity (23), as it is schematically described in Figure 3.

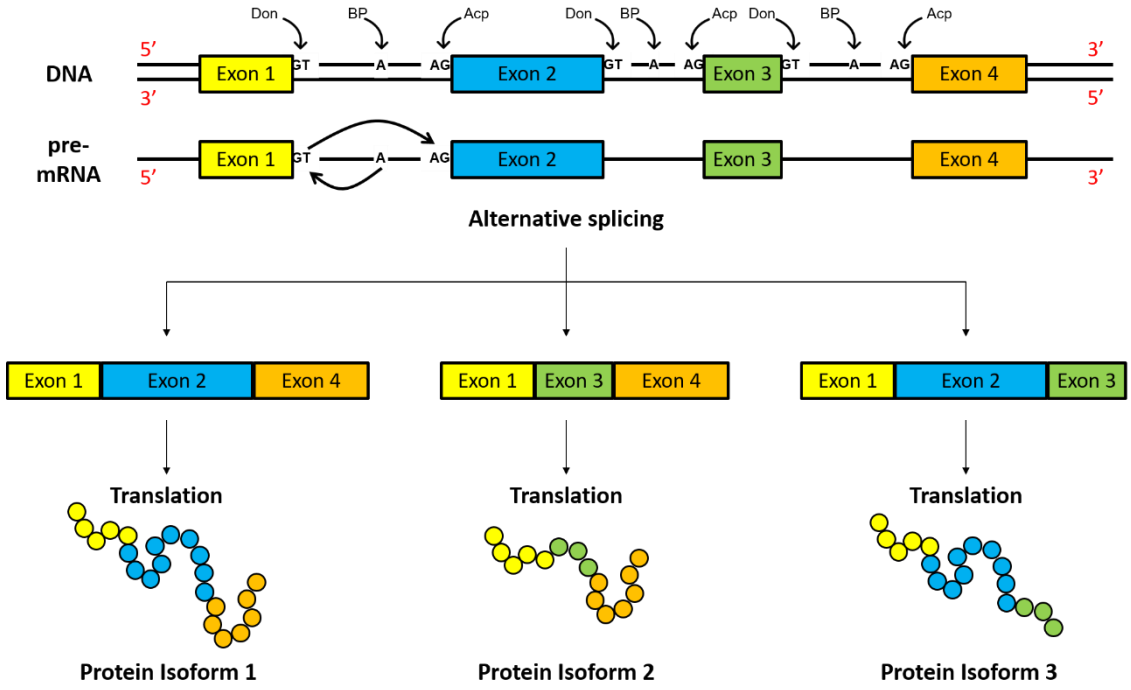


Figure 3 – Alternative splicing machinery. The 2'OH of a specific branch point within the intron attacks the donor site at the same intron, forming a loop. Hereafter, the donor site 3'OH attacks the acceptor site in the same intron, removing the intronic area. Afterwards, exons connect and are translated. By alternative splicing, not all exons might be translated and different protein isoforms are produced, all originated from the same gene. Don-Donor; BP-Branch Point; Acp-Acceptor (Based on reference (22)).

Regarding *CD44* gene, many CD44 variant isoforms are produced. The most abundant isoforms are the standard (CD44s) and the epithelial isoforms (CD44e). The first one is only constituted by the constant part of the gene and the second one has the insertion of exons v8-v10 by alternative splicing. Besides these isoforms, other CD44 variant isoforms (CD44v) are originated by the insertion of certain variant exons by alternative splicing between exons 5 and 16 of the constant part of the gene (19). Both CD44s and CD44v play important roles in tumor initiation, progression, invasion and metastization of certain types of cancer (20).

### 2.1.2. Expression pattern and tissue specificity

The different CD44 protein isoforms do not display the same expression pattern and usually vary in function (24). CD44s is ubiquitously expressed in almost all mammalian cells. However, CD44v are expressed in specific epithelial cells, particularly during embryonic development, lymphocyte maturation and activation. Specific CD44v isoforms have also been reported to be aberrantly expressed in several types of cancer (25), such as CD44v3, which has been reported to be overexpressed in GC, uterine cervix squamous cell carcinoma, nasopharyngeal carcinoma and colorectal cancer and is associated with poor

cellular differentiation, advanced stage, lymph node metastases, worse overall survival (OS) and disease-specific survival (DSS) (26-29). CD44v9 has also been reported to be highly expressed in early GC, esophageal squamous cell carcinoma, triple-negative breast cancer, and head and neck carcinoma and has been associated with poor prognosis, worse survival rate and higher number of lymph node metastases (30-33). CD44v6 isoforms have also been associated with cancer, as it will be mentioned bellow.

### 2.2. CD44v6 isoforms

The CD44v6-containing isoforms are composed by a cluster of isoforms that contain the variant exon 6 (18, 34). These isoforms have been widely investigated in the past decades due to its additional binding motifs that connect specific ligands and that relate these proteins with some features commonly associated with cancer (18). The portion of the protein translated from exon v6 contains extra binding sites that ligate CD44 to hepatocyte growth factor (HGF) and vascular endothelial growth factor (VEGF) (19). The HGF is the ligand to the receptor tyrosine kinase MET (c-MET), a well-known proto-oncogene and, the binding of CD44v6 to HGF results in the activation of c-MET by phosphorylation, which ultimately promotes cancer cell invasion. The recruitment of ERM proteins by the cytoplasmic tail, results in the interaction of ERM with VEGF receptor 2 (VEGFR-2), which contributes to angiogenesis, cancer cell division, proliferation and invasion (35). Figure 4 details the signaling pathways of CD44v6.

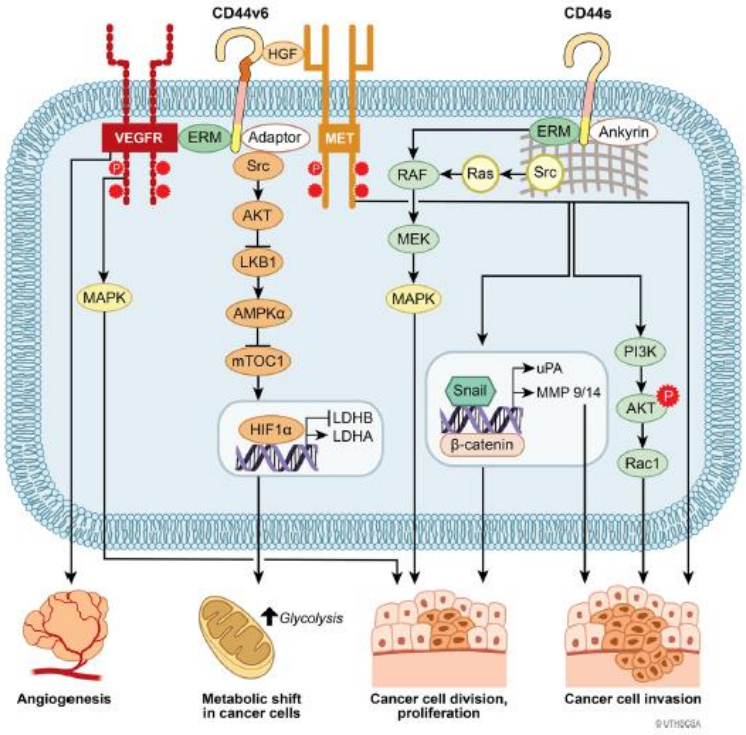


Figure 4 - CD44v6 and CD44s downstream signaling pathways. CD44v6 possesses specific binding sites for c-MET and VEGFR-2, leading to its activation and the activation of specific pathways, which ultimately results in cancer cell proliferation, invasion, angiogenesis and shifts in metabolism (Modified after (35)).

### 2.2.1. CD44v6 in normal tissues

Only a few specific types of cells express CD44v6-containing isoforms in a non-pathologic manner. Soukka and colleagues (1997) discovered that some skin cells expressed CD44v6 and CD44s likewise. According to them, all the cell layers from the epidermis, except the keratinized upper layer, express both CD44s and CD44v6. Hair follicles and sweat glands express both isoforms as well (36). Kooy and colleagues (1999) study concluded that CD44v6 is also expressed in adherens junctions, desmosomes and complex interdigitating membrane structures of the epidermis (37). Another research, from Afify and colleagues (2005), investigated the expression of CD44v6 in endometrium cells and concluded that the up-regulation of CD44v6 in the secretory glands of the endometrium were coincident with the time in which endometrium is most accessible to embryo implantation (38). In a study by Mack and Gires (2008) it was possible to understand that CD44v6 is also expressed in normal epithelial cells of the head and neck (39). Nevertheless, as most CD44v, CD44v6 is majorly expressed in malignant cells when compared to normal tissues (40).

### 2.2.2. CD44v6 in cancer

Studies on the association between CD44v6 and cancer began in the early 1990's after Gunthert and colleagues (1991) discovered a new CD44 glycoprotein (CD44v4-v7 isoform) in a metastatic rat pancreatic carcinoma cell line (BSp73ASML) and that, upon transfecting that specific variant, into a related non-metastasizing cell line (BSp73AS), conferred a metastatic potential to those cells (41). Afterwards, Seiter and colleagues (1993) investigated the use of a specific antibody for CD44v6 in the transfected cells. They treated mice with the transfected cells and with the CD44v6 specific antibody, resulting in the blockade of lymph node and lung metastases, hence proving the metastatic potential conferred to these cells by CD44v6-containing isoforms (42).

Since then, many researchers have studied the expression of CD44v6 in various types of cancer and its aberrant expression has been associated with poor prognosis in gastric cancer (43, 44), esophageal cancer (45), colorectal cancer (46), non-small cell lung cancer (NSCLC) (47) and breast cancer (48). Nevertheless, it has also been associated with lymph node or/and tumor metastases in gastric cancer (44), pancreatic cancer (49), esophageal cancer (45), breast cancer (48, 50) and colorectal cancer (46). Moreover, some studies have associated CD44v6 with a worse survival rate at the 5-year marker in gastric cancer (44), pancreatic cancer (49) and in colorectal cancer (46) as well. CD44v6 is also associated with tumor differentiation in gastric cancer (44), colon cancer (45) and breast cancer (48)

and with advanced tumor staging in gastric (44) and breast cancer (50). Since CD44v6 is involved in tumor initiation, progression and invasion in various types of cancer, this protein might work as a cancer biomarker.

#### 2.2.2.1. CD44v6 *de novo* expression in premalignant and malignant lesions

As mentioned above, the expression of the variant isoforms is a rare event in the organism. However, its expression in premalignant and malignant lesions is frequent (18, 19). In a previous study from our group, it was discovered that CD44v6 expression changed between normal gastric tissue and gastric tumor tissue. Analyzing samples from normal gastric mucosa, hyperplasia, metaplasia, dysplasia and advanced gastric carcinomas, it was observed that there was no expression of CD44v6 in normal gastric mucosa, however, as the lesion evolved to hyperplasia, transdifferentiated into metaplasia, dysplasia and finally GC, CD44v6 became gradually overexpressed, indicating that CD44v6 may be involved in the malignant transformation of the gastric mucosa (51).

#### 2.2.2.2. CD44v6 expression in cancer stem cells

Cancer stem cells (CSCs) have been defined as a small sub-population of tumor cells with increased tumor-initiating capacities and self-renewal potential. Over the years, it has been discovered that CSCs possess the ability to self-renew and that tumors who derived from purified CSCs can recapitulate the heterogeneous phenotype from the original tumor, hence their differentiation capacity. These cells are, often, highly resistant to apoptosis and are essential in the process of metastases formation after a latency period (25). Both CD44s and CD44v are considered CSC markers and they have been shown to be engaged in performing functions shared by CSCs and normal stem cells (19). As reported in the literature, CD44v6 is highly expressed in CSCs from brain (52), colon (53), pancreas (54) and head and neck tumors (39).

#### 2.2.2.3. CD44v6 and associated therapy resistance

A high percentage of GC patients treated with chemotherapy often do not respond to it or relapse upon an initial favorable response (11). Therefore, intrinsic and acquired resistance to chemotherapeutic agents is one of the main reasons for therapy failure. It has already been reported in the literature the association between CD44v6 expression and chemotherapy resistance in cancer patients (55). Lin Lv and colleagues demonstrated that CD44v6 overexpression is implicated in acquired resistance in colon cancer through a

process of autophagy flux (56). Nevertheless, little is known about the implication of CD44v6 in GC therapy resistance.

#### 2.2.2.4. Anti-CD44v6 targeted therapies and clinical trials

As CD44v6 began to be studied in the cancer field and its potential use as a biomarker or a therapeutic target in the clinic investigated, the use of monoclonal antibodies (mAb) against CD44v6 started being tested. A mAb anti-CD44v6 was developed in combination with $^{99m}$Technetium, namely BIWA 1, and was tested for head and neck cancer diagnosis However, BIWA 1 was proved to be immunogenic and to have a heterogeneously accumulation in the tumor cells (57). Afterwards, the covalent combination of the same mAb anti-CD44v6 with the chemotherapeutic cytotoxic drug mertansine (*i.e*. bivatuzumab mertansine), was tested in patients with incurable squamous cell carcinoma of the head and neck or esophagus. This phase 1 clinical trial resulted in the death of one of the study participants through grade 4 skin toxicity, where the patient suffered a large loss of epidermis in the body, with the separation of the skin from the body on touch. At the microscopic level, there was an extensive apoptosis of the keratinocytes from the basal and suprabasal layers, where CD44v6 is known to be expressed (58).

Recently, gold nanostars-based PEGylated multifunctional nanoprobes conjugated with CD44v6 mAb (*i.e.* CD44v6-GNS) have been developed for the treatment of GC. These conjugated nanoparticles display a high affinity towards spheroids of GC stem cells and are capable of destroying those spheroids with a low power laser treatment. These nanoprobes were injected in xenograft nude mice models as well, and it was demonstrated that CD44v6-GNS were capable of targeting the GC vascular system and inhibit tumor growth after laser irradiation. Likewise, CD44v6-GNS nanoprobes revealed great potential for GC therapy in the near future (59). Nevertheless, anti-CD44v6 cancer therapy is yet to be implemented into clinical practice.

## 3. Genetic manipulation strategies for ablation of exon v6 from *CD44* gene

Technologies for manipulating the human genome have enabled several advances in many biological fields in the past decades (60). The techniques developed to genetically manipulate both DNA and RNA have become extremely precise over the course of years of investigation and can alter the genome either transiently or permanently (60, 61). These techniques have been able to change gene expression and, therefore, its usage is being considered for certain pathologies as gene therapy (62).

As CD44v6 is implicated in tumor initiation, progression and invasion in many cancer pathologies (19), the application of genetic manipulation technology to successfully remove exon v6 from the *CD44v6*-containing transcripts would be an interesting approach to investigate for cancer therapy.

### 3.1. Antisense oligonucleotide strategy

In recent years, there has been a renewed interest in antisense oligonucleotides (AONs) usage for gene therapy (63). AONs are synthetic strings of nucleic acid in single stranded form, which possess 8 to 50 nucleotides in length and can bind to mRNA through Watson–Crick base pairing hybridization. These oligonucleotides can interfere with the expression of certain genes by altering their RNA function (63-65). Stephenson and Zamecnik (1978) were the first to develop a DNA molecule of 13 nucleotides with modifications at the 3′ and 5′ OH moieties, which was capable of inhibiting replication and cell transformation of the Rous sarcoma virus (66). Since then, many modifications in AONs chemistry were developed to improve AONs stability, binding strength and specificity which allowed the use of AONs for therapy applications. For instance, giving rise to several different tools to interfere with gene expression, allowing their use to restore protein expression, to reduce the expression of certain toxic proteins or even modify mutant proteins (63, 64).

The modifications developed to improve AONs efficacy over the years relate with their backbone, sugar component and some alterations in ribose and nucleoside components. Figure 5 displays in detail AONs types of modifications.

According to their function, AONs can be divided in dependent or non-dependent of RNase H. The first ones eventually lead to mRNA degradation, while the RNase H non-dependent AONs prevent or inhibit the splicing machinery (64). One application of the RNase H non-dependent ANOs is to cause the skipping of exons in certain genes. The concept of exon skipping relies on the splice out of a specific exon of multiple exons by alternative splicing. This mechanism occurs naturally in cells as part of the splicing machinery and is the main event occurring in vertebrates and invertebrates in terms of alternative splicing (67). However, the concept of exon skipping is a very useful tool for gene manipulation, since skipping an out-of-frame mutated exon that causes knockout of an important protein can lead to restauration of the reading frame, hence resulting in a smaller but functional protein (68).

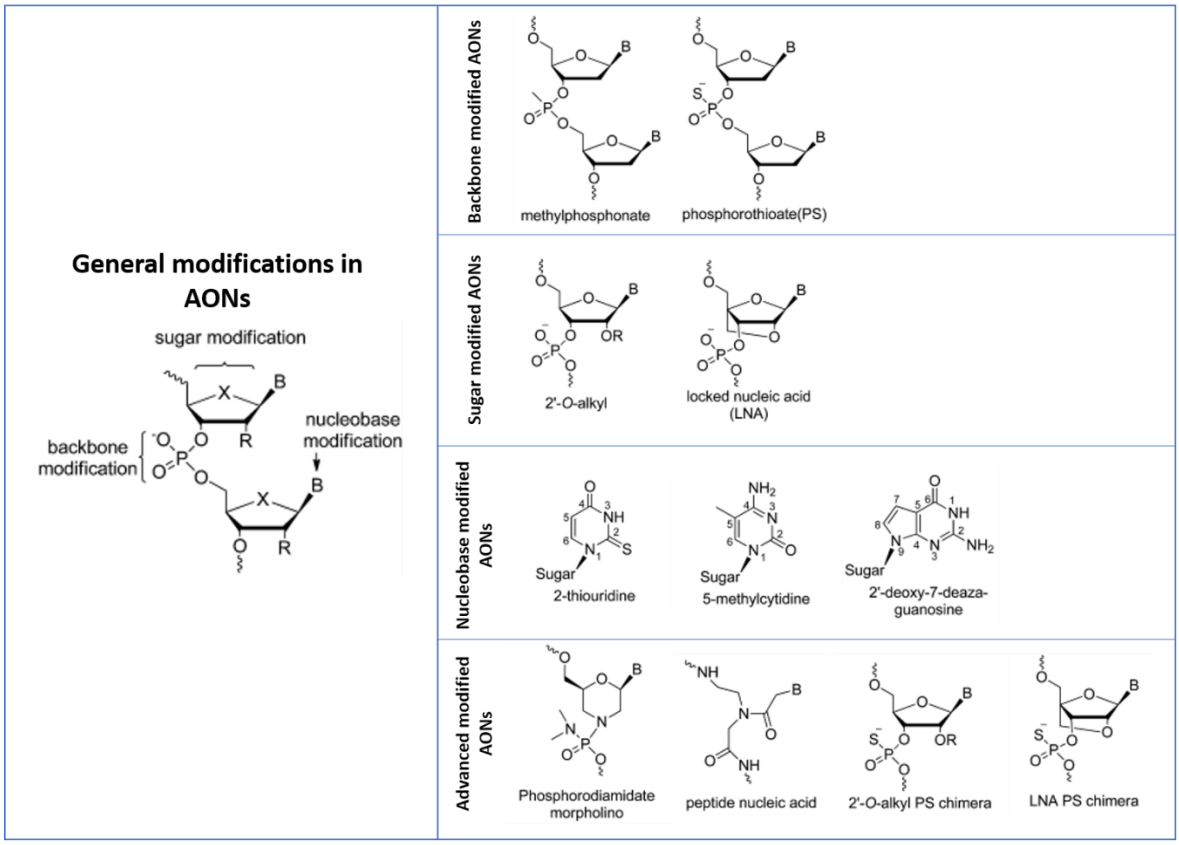


Figure 5 – General modifications developed in AON strategy. The main modifications occur at the backbone, sugar and nucleobase. Combined modifications occur at the "Advanced modified AONs" (Modified after (69)). AONs – Antisense Oligonucleotides.

One of the most prone AONs for exon skipping is the Phosphorodiamidate Morpholino Oligomer (PMO). PMOs are synthetic and uncharged analogs of nucleic acids that are usually 25 subunits linked together, with each one bearing a nucleotide. Its phosphorodiamidate backbone consists of morpholine rings with methylene groups which are bound to modified phosphates with a nonionic dimethylamine group. One standard DNA nucleotide is bound to each morpholine ring. Antisense PMOs can block macromolecule interactions with mRNA by paring through Watson–Crick base pairing hybridization. This prevents the initiation complex read-through or modifies the splicing of many cell types, including human cells (70, 71).

Recently, a PMO was approved for the treatment of Duchenne Muscular Dystrophy (DMD), namely Eteplirsen. Also known as Exondys 51 or AVI-4658, this 30-mer drug developed by Sarepta Therapeutics hybridizes with exon 51 of the *DMD* gene, blocking its splicing into the mRNA. In DMD patients carrying an out-of-frame mutation near this exon, Eteplirsen can block the splicing of exon 51, which restores the reading frame. This leads to a shortened functional DMD protein, resulting in milder symptoms, similar to the ones observed in Becker Muscular Dystrophy (BMD) (72). These two disorders result from

mutations on the same gene, however, DMD is caused by out of-of-frame mutations that result in no production of DMD protein, while BMD is caused by in-frame mutations that result in a truncated but functional protein. They only differ in severity, age of onset and progression rate. In DMD, muscle weakness commonly appears in early childhood and rapidly gets worse. DMD patients usually are wheelchair-dependent when they reach adolescence and, due to rapid disease progression, normally live only into their twenties. On the other hand, BMD patients present the same symptoms, however, these symptoms begin later, and evolve much slower, hence BMD patients can survive into their forties and beyond (73). The mechanism of gene therapy for DMD patients is underlined in Figure 6.

Another AON has been approved, by the FDA, for the treatment of Spinal Muscular Atrophy (SMA). Nusinersen, also known as Spiranza, was developed by Biogen and is an 18-mer phosphorothioate 20 -O-methoxyethoxy AON with all cytidine's methyl-modified at the 5-position. SMA is caused by a mutation in the *SNM1* gene, which leads to deficient SNM protein. The *SMN1* gene is the principal responsible for SMN protein production, however, there is a *SMN2* gene, which produces very small quantities of protein. This AON can include exon 7 in *SMN2* mRNA by targeting and blocking an intron 7 internal splice site, resulting in SMN2 producing a SMN protein similar to functional SMN1 (74).

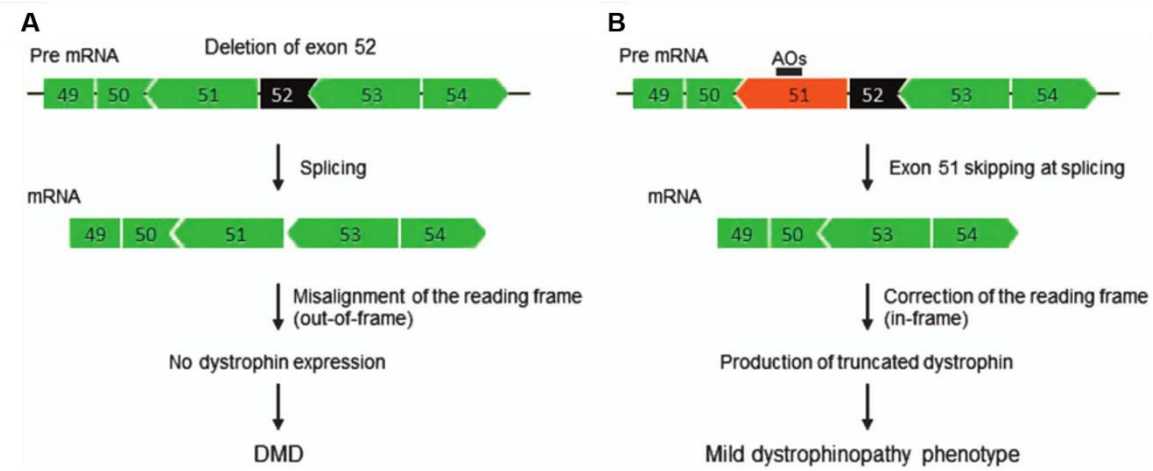


Figure 6 – Gene therapy for DMD disease. (A) a mutation on exon 52 lead to the misalignment of the reading frame, resulting in no dystrophin protein expression. (B) Targeting an AON to exon 51 results in skipping of exon 51, correction of the reading frame and consequently in the production of a smaller but functional dystrophin protein (Modified after reference (75)).

### 3.2. CRISPR/Cas9 genome editing

The genome editing techniques started to be developed in the 1970s with the discovery of restriction enzymes that protect bacteria against phages. Since then, gene editing has evolved, being CRISPR/Cas9 system the most promising genomic engineering technology presently available (60).

CRISPR stands for Clustered Regularly Interspaced Short Palindromic Repeats and is essential in the adaptive immunity of several microorganisms (60, 76, 77). Various CRISPR systems have been identified from a wide range of bacterial and archaeal hosts, all comprising a cluster of CRISPR-associated genes (*Cas*), noncoding RNAs and repetitive elements that are interspaced by short sequences of exogenous DNA targets, namely protospacers that are associated with a protospacer adjacent motif (PAM). The type II CRISPR system derived from *Streptococcus pyogenes* is the best-characterized one. The CRISPR system was modified from these microorganisms to be used as an engineering genomic tool to precisely cut DNA. A single guide RNA (sgRNA), which is constituted by a CRISPR RNA (crRNA) and a trans-activating crRNA (tracrRNA), is joined to the Cas9 protein, and this complex is transfected into cells. The sgRNA targets a DNA portion via Watson-Crick base paring and the Cas9 protein recognizes the PAM sequence, causing a double strand break (DSB) approximately 3bp upstream the PAM sequence, in the target DNA. The DSBs can then be repaired by the error-prone non-homologous end-joining (NHEJ) pathway or the highly precise homology-directed repair (HDR). NHEJ leads to insertions or deletions that cover for the DSB site, while the HDR requires a DNA template with homolog regions overlapping each side of the cut DNA to be precisely inserted into the region of the DSB (60, 76). The CRISPR/Cas9 system is described in detail in Figure 7.

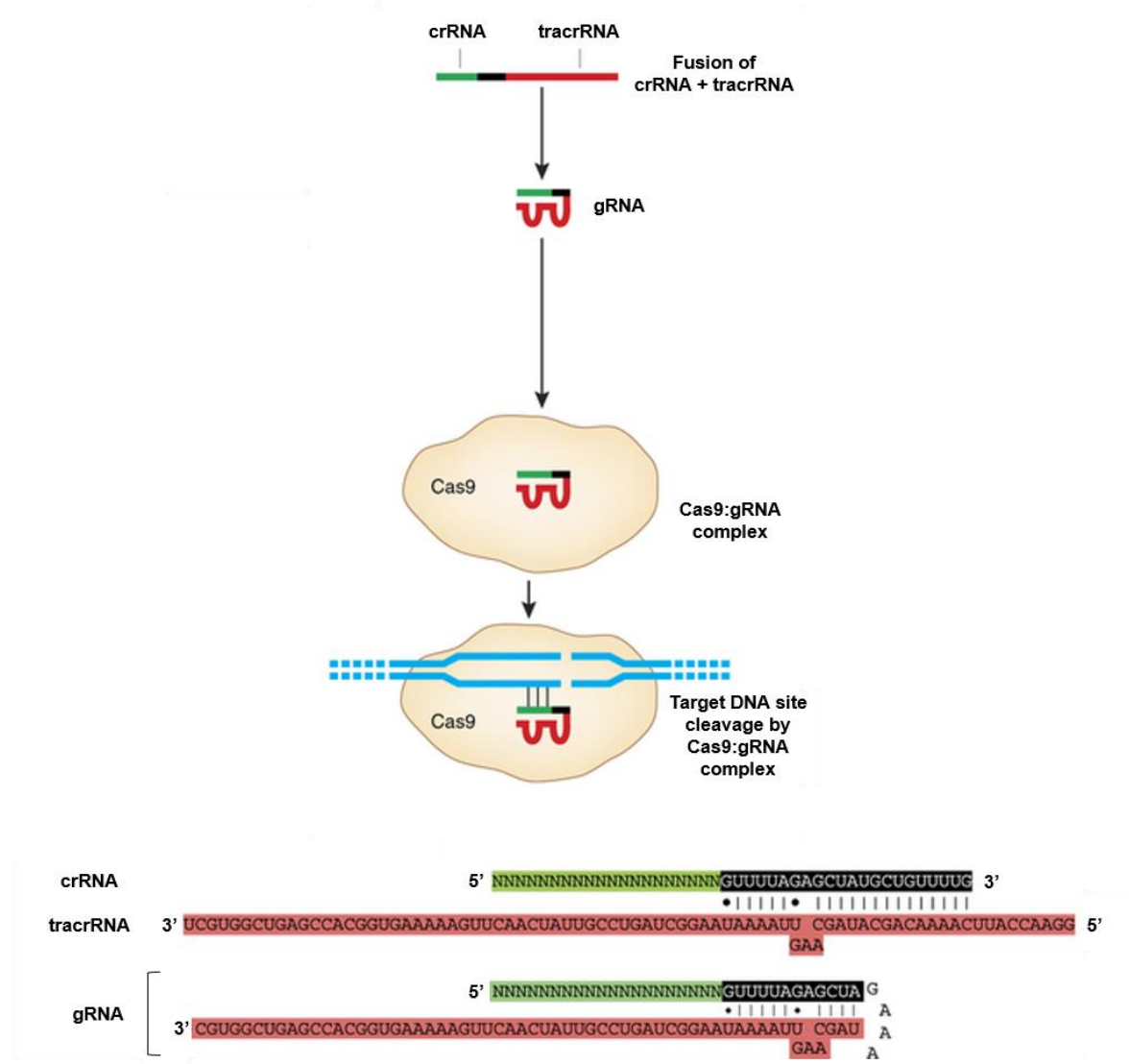


Figure 7 - Engineered CRISPR-Cas9 system. Fusion between a crRNA and a tracrRNA sequence, results in a sgRNA. Afterwards, the sgRNA complexes with Cas9, and this complex is capable of breaking the DNA 3-4bp upstream the PAM sequence, which is recognized by the Cas9 (Modified after (78)). crRNA - CRISPR RNA; tracrRNA - transactivated CRISPR RNA; sgRNA - single guide RNA; PAM - Protospacer Adjacent Motif.

Even though CRISPR/Cas9 is a gene editing technique with enormous possibilities, its use for gene therapy for human diseases is still controversial. Despite its potential, how CRISPR/Cas9 editing tools are efficiently delivered to targeted cells *in vivo* and how to avoid or reduce unintended off-target effects remain major challenges, which are crucial for its clinical applications (79). Nevertheless, this technology is a good approach to mimic the results obtained with AONs for the treatment of several disorders in a permanent fashion. Only few clinical trials have started testing the use of CRISPR/Cas9 technology in human disorders, however, some researchers are studying it *in vitro* and *in vivo*, in several animal models (80-83).

### 3.3. Potential of genetic manipulation for cancer therapy

Cancer is a disease mainly caused by the accumulation of genetic and epigenetic alterations. Likewise, it is feasible that genetic manipulation may have great potential for cancer therapy. Cancer cells often present aberrant expression of certain genes, namely oncogenes or tumor suppressor genes. Using gene manipulation to restore the normal expression of such genes would be a proper mechanism to treat cancer (79).

While this idea seems promising, the truth is that no gene manipulation therapy has yet been approved for the treatment of oncologic pathologies (84, 85). Regarding AONs strategies, there are currently approximately 90 clinical trials evaluating their potential for cancer therapy (website: https://clinicaltrials.gov/ct2/results?term=antisense+and+cancer&Search=Search). Nevertheless, none of these clinical trials have passed phase 3 stage with satisfying results (84). Some were promising, such as Custirsen (developed by Isis and Teva/Oncogenex), an AON developed to target clusterin, which is implicated in prostate cancer cell survival and drug resistance (86), or Lucanix (developed by Nova Rx), an AON that targets TGF-β2, for the treatment of NSCLC (NCT00676507). Imetelslat (developed by Geron) is another AON developed to target telomerase by binding to the RNA at the active site of the enzyme (NCT01731951). However, none of these AONs have shown improvement in patients survival comparing to standard therapies (84).

Concerning CRISPR/Cas9 technology, there are significantly less clinical trials ongoing, and only still in initial states of investigation, nevertheless, a few show promising results. One example is a phase 1 clinical trial that genetically engineers autologous lymphocytes *ex vivo* to knockout PD-1. The genetically engineered T-cells are infused back into the patient for the treatment of advanced NSCLC (NCT02793856) (87).

Although gene manipulation has entered the clinics as therapy for some genetic disorders, to this date there is still no approved gene manipulation for cancer therapy (88). Nevertheless, in the future, manipulation of genes engaged in cancer initiation and development will be an important field of investigation due to its potential as an anticancer therapy.

# Chapter II | Rational and Aims

GC is one of the deadliest cancers worldwide. Most GC patients are diagnosed at stage IV, where usually tumors are inoperable and where chemotherapy is the first line of treatment. Usually GC patients respond poorly to chemotherapy, therefore, both GC patient stratification and therapy must be improved. CD44 has been widely investigated in various types of cancer, including GC. This protein is important in cell-cell- and cell-matrix junctions. *CD44* gene endures extensive alternative splicing, which ultimately results in the production of various variant isoforms. CD44v6-containing isoforms are mainly expressed in cells from the gastric mucosa that undergo malignant transformation. CD44v6 has been implicated in cancer development, and it was discovered that CD44v6-containing isoforms were capable of modulating GC cell lines response to chemotherapy. CD44v6 has also been investigated for targeted therapy however, the clinical trial to test CD44v6 inhibition did not produce conclusive results.

Keeping all of this in mind, our general aim was to design and establish GC cell line models where *CD44* exon v6 is skipped from *CD44v6*-containing transcripts, maintaining the reading frame. These models would then be pivotal to dissect the role of exon v6, itself, in the modulation of chemotherapy response associated to CD44v6-containing isoforms in GC cell lines. To accomplish this main goal, three specific goals must be achieved:

- **Specific aim 1:** Design strategies to obtain GC cell line models lacking only exon v6, in a transient and permanent manner, using PMOs and CRISPR/Cas9 techniques, respectively.
- **Specific aim 2:** Characterize the GC cell lines obtained from the editing techniques.
- **Specific aim 3:** Use the permanent GC cell line model, obtained with CRISPR/Cas9, to assess whether the removal of exon v6 from *CD44v6*-containing transcripts modulates the response of GC cells to chemotherapy.

With the proposed aims, we will generate two exon skipping models for GC investigation and through those models, it will be possible to study the specific role of exon v6, namely in the modulation of therapy response in GC cells.

# Chapter III | Material and Methods

## 1. Experimental design

The workflow for this project involved two main approaches for genetic manipulation of *CD44* gene, as observed in Figure 8. PMOs and CRISPR/Cas9 system were used to either transiently or permanently delete exon v6 from *CD44v6* expressing GC cell lines, without knocking down the *CD44* gene. To achieve this goal, two GC cell lines that endogenously express *CD44v6*, GP-202 and MKN-45, were used.

Initially, PMOs were designed to target the splice sites at the beginning and at the end of the *CD44* exon v6. Prior to the use of the designed PMOs, transfection conditions were optimized in order to achieve the best transfection efficacy. After transfecting PMOs into GP-202 and MKN-45 cell lines, cells were analyzed regarding their RNA and protein modifications.

While designing the PMOs, sgRNAs sequences were designed to target the adjacent exon v6 introns, the exon-intron boundaries from exon v6, and the exon v6 itself. Afterwards, the designed sgRNAs were cloned into a plasmid containing the Cas9 gene and the recombinant plasmid was transfected into GP-202 and MKN-45 cell lines. After cell selection, the obtained clones for each cell line were characterized and only the ones with a homozygous deletion and no *CD44v6* expression, at the DNA, RNA and protein level, were selected to carry out the experiments.

For each cell line, three clones permanently edited by CRISPR/Cas9 and mock cell lines were treated with chemotherapeutic agents used in the treatment of GC patients. Two types of treatments were performed, short-term and long-term treatments, in order to evaluate if exon-v6 is involved in the modulation of the response to chemotherapy observed in GC cells expressing CD44v6-containing isoforms. Initially, we aimed to perform chemotherapy treatments using cells lines from both PMOs and CRISPR/Cas9 approaches. However, later we decided to pursue the treatments using only the clones obtained with the CRISPR/Cas9 technology.

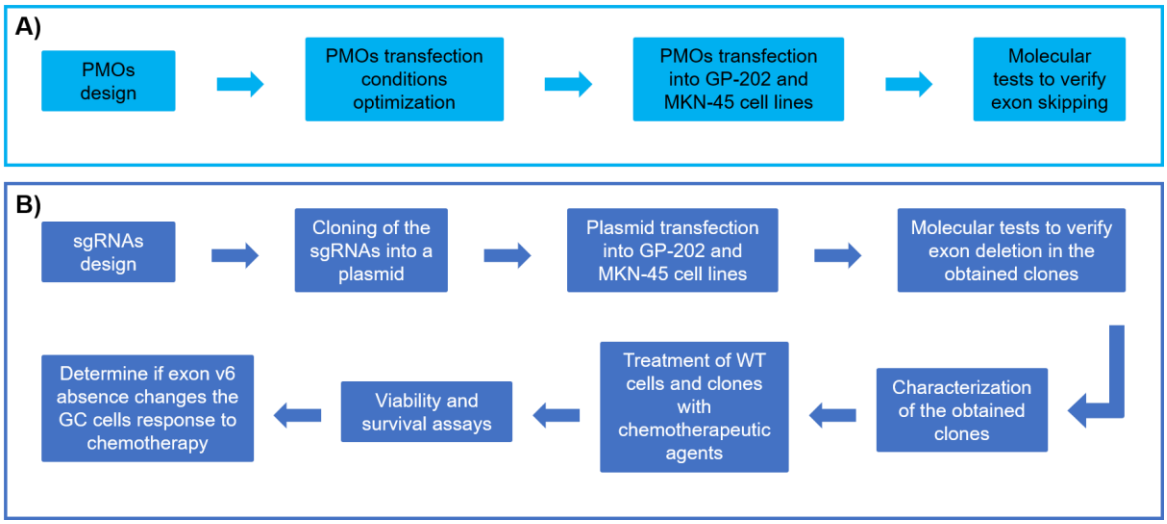


Figure 8 – Workflow for this project regarding the (A) PMOs and (B) CRISPR/Cas9 experiments.

## 2. Cell culture

Human GC cell line MKN-45 was purchased from the JCRB Cell Bank (Japanese Collection of Research Bioresources Cell Bank), and the non-commercial GP-202 cell line was established at Instituto de Patologia e Imunologia Molecular da Universidade do Porto (Ipatimup). Both cell lines were routinely cultured in Roswell Park Memorial Institute medium (RPMI) from ThermoFisher Scientific (Waltham, MA, USA) containing 10% (v/v) of Fetal Bovine Serum (FBS) from Biowest (Nuaillé, France). Cell cultures were maintained at 37 °C under a 5% $CO_2$ humidified atmosphere. Cells were sub-cultured every 4-5 days, for a maximum of 5 months. Cells were routinely confirmed to be free of mycoplasma contamination.

## 3. PMO approach

Three PMOs were used in the project: one fluorescent PMO (labelled with Fluorescein) to assess transfection efficiency; two PMOs specifically designed to target splice sites near exon v6. All PMOs were purchased from Gene Tools (Philomath, OR, USA).

### 3.1. PMO design

PMOs used in this project were designed to target the splice sites at the beginning and at the end of exon v6 of the *CD44* gene, as observed in Figure 9. *CD44* gene transcripts were analyzed using *Ensembl genome browser 97* (https://www.ensembl.org/index.html).

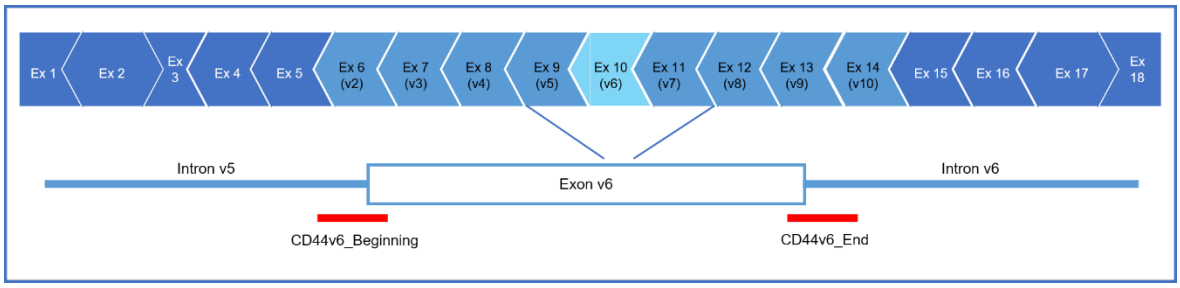


Figure 9 – PMOs location in the *CD44* gene. The CD44v6_Beginning PMO targets the donor site and the CD44v6_End PMO targets the acceptor site adjacent to exon v6.

To search for the possible splice sites near exon v6, *Human Splicing Finder* online tool (http://www.umd.be/HSF3/HSF.shtml) was exploited and the PMOs were designed to target the specific splice acceptor and donor at the 5' and 3' sites of exon v6, respectively. Two PMOs were designed: "CD44v6_Begining" (5' CTGGACTGTGAGAAGAATATCAGTT 3'), at the 5' site of exon v6 and "CD44v6_End" (5' CTTGTTAAACCATCCATTACCAGCT 3'), at the 3' site of exon v6.

### 3.2. Transfection optimization of PMOs

PMOs were delivered into the cells by transfecting with Endo-Porter (EP) solution from Gene Tools (Philomath, OR, USA). EP is an aqueous polyethylene glycol (PEG) formulation reagent (1 mM of EP peptide and 10% polyethylene glycol in water) capable of forming endosomes that contain the PMOs and enter the cells by endocytosis, hence delivering the PMOs into the cytosol, as shown in Figure 10.

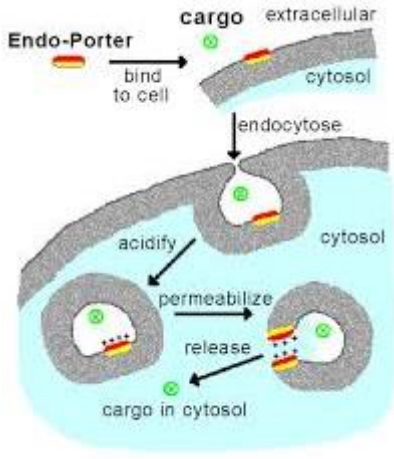


Figure 10 - Endo-Porter delivery mechanism of PMOs into cell cytosol. EP forms endosomes that involve the PMOs and enter the cell by endocytosis. In the cytosol, the endosomes became permeable and PMOs are release into the cytosol Adapted from Gene Tools: https://www.gene-tools.com/endo_porter.

Before transfecting the PMOs specific to the intron^exon/exon^intron boundaries, MKN-45 and GP-202 cells were transfected with a standard fluorescent control and EP solution. Different concentrations of both reagents were tested to assess transfection efficacy and, 48 h after transfection, cells were washed 4 times with 1x Phosphate-Buffered Saline (PBS), trypsinized and centrifuged (304 x *g*; 5 min). After discarding supernatant, pellet was resuspended and cells were fixed in 200 µL of 2% (v/v) Paraformaldehyde (PFA) from Merck (Darmstadt, Germany). Cells fluorescence was analyzed by FACS Canto II (BD Biosciences, Franklin Lakes, NJ, USA). Data analysis was performed using FlowJo version 10 software (FlowJo LLC, Ashland, OR, USA).

### 3.2. Generation of transient cell line models lacking exon v6 using PMOs

Cells were seeded in 12 well plates, $7x10^4$ MKN-45 and $1x10^5$ GP-202 cells per well and allowed to adhere overnight. In the following day, RPMI medium was replaced (750 µL per well), PMOs were added to the medium and subsequently EP solution was added at the desired concentrations (2 µM of EP for both cell lines; 4 µM of both CD44v6_Beginning and CD44v6_End PMOs for GP-202 cell line; 4 µM CD44v6_Beginning and 6 µM CD44v6_End for MKN-45 cell line). Cells were incubated for 48 h and then collected either for RNA extraction, expression analysis by reverse transcription quantitative real-time polymerase chain reaction (RT qRT-PCR), or immunofluorescence analysis of total CD44 and CD44v6 expression. Both PMOs (5' UTR and 3' UTR) were individually transfected.

## 4. CRISPR/Cas9 system

### 4.1. Single guided RNA (sgRNA) design

All sgRNAs were purchased from Sigma-Aldrich® (Poole, UK) and were designed to target the adjacent exon v6 introns, the exon-intron boundaries from exon v6, and the exon v6 itself from *CD44* gene. To locate the best sgRNA target regions, *Benchling online tool* (https://benchling.com/) was assessed to identify and score the target regions in the flanking introns and within the exon v6. Six sgRNAs were selected according to their strategic location, as shown in Figure 11.

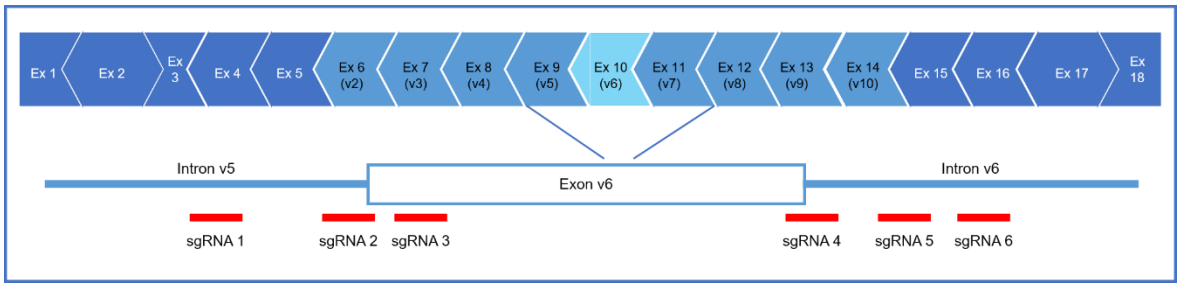


Figure 11 – sgRNA location in the *CD44* gene. sgRNA 1 targets intron v5, sgRNAs 2 and 4 target exon v6 boundaries, sgRNA 3 targets exon v6 and sgRNAs 5 and 6 target intron v6.

### 4.2. Cloning sgRNA into pSpCas9(BB)-2A-Puro (Px459) V2.0 plasmid

The sgRNAs were separately cloned into the pSpCas9(BB)-2A-Puro (PX459) V2.0 plasmid (Addgene plasmid #62988; http://n2t.net/addgene:62988; RRID:Addgene_62988) from Addgene (Watertown, MA, USA), which contains the Cas9 gene from *S. pyogenes* and the PX459 vector backbone, as shown in Figure 12.

The PX459 vector contains a *Bbs I* restriction site downstream the U6 promoter, where sgRNA were inserted. Using *Bbs I* restriction enzyme, the vector was digested, resulting in cohesive ends. The sgRNAs were inserted into the open vector after being annealed to form a double chain, as demonstrated in Figure 13.

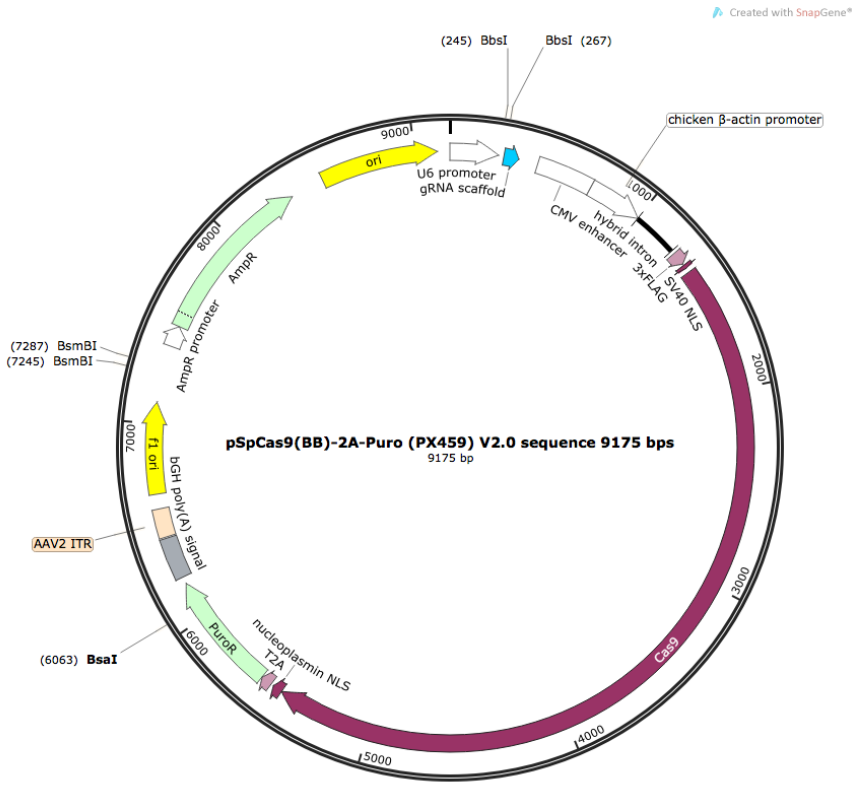


Figure 12 - pSpCas9(BB)-2A-Puro (PX459) V2.0 plasmid. It contains a *Bbs* I restriction site, the Cas9 gene, the puromycin gene and the ampicillin gene (Modified from reference (76)).

Cloning of the six sgRNAs into the pSpCas9 plasmid was performed according to "Morrisey Lab Protocol: Generating Large (>1 kb) Genomic Deletions using CRISPRs" (https://www.med.upenn.edu/genetics/tcmf/documents/Protocol2.pdf), with few adaptations. 2 µg of PX459 vector was digested using *Bbs I* restriction enzyme from New England Biolabs (Ipswich, MA, USA) (37 °C; 1 h incubation with enzyme → 65 °C; 15 min for enzyme inactivation). To dephosphorylate the vector, Shrimp Alkaline Phosphatase (rSAP) from New England Biolabs (Ipswich, MA, USA) was added to the reaction (37 °C; 30 min incubation with enzyme → 65 °C; 5 min for enzyme inactivation).

The sgRNAs are customized as single chain oligomers and, in order to ligate the sgRNAs into the vector, complementary single chain sgRNAs annealing is necessary, to produce a double chain form. To link to the dephosphorylated vector, the sgRNAs were phosphorylated using T4 PNK enzyme from New England Biolabs (Ipswich, MA, USA). The annealing and phosphorylation were performed in a thermocycler (37 °C; 30 min → 95 °C; 5 min → ramp down -5 °C/min until 25 °C).

The ligation reaction works in a proportion of *insert:vector* of 1:1 and was composed of 150 ng of vector to 150 ng of insert. T4 ligase from Invitrogen (Carlsbad, CA, USA) was used to ligate the insert to the vector.

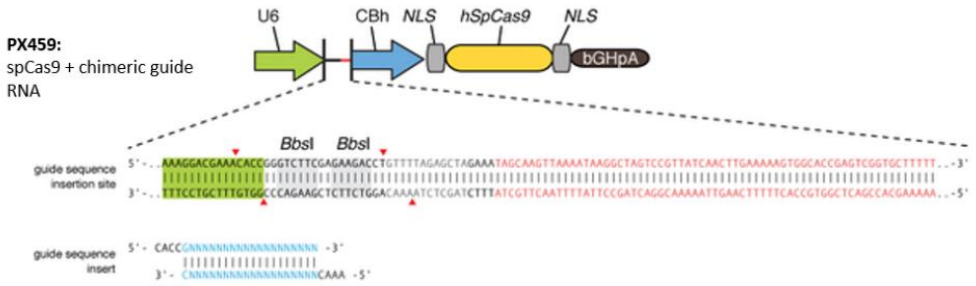


Figure 13 - Scheme for Bbs I digestion site. An extra "CACCG" sequence is placed in the beginning of all sgRNAs to match the Bbs I digestion site (Modified from reference (76)).

One Shot™ Stbl3™ Chemically Competent *Escherichia coli* from ThermoFisher Scientific (Waltham, MA, USA) was utilized to expand PX459 vector with the desired insert. pUC19 DNA was added to Stbl3 cells as a positive transformation control. After incubating the ligation product and Stbl3 mix (30 min on ice), a heat shock was performed (42 °C; 45 sec), followed by recovery on ice to allow the vector to enter the cells. Super Optimal broth with Catabolite repression medium (SOC) was used for cell recovery. Stbl3 cells were incubated at 37 °C for 1 h with agitation. Afterwards, the entire reaction was plated in lysogeny broth (LB) plates with ampicillin for bacteria selection overnight (ON).

#### 4.2.1. Positive colony screening

Isolated colonies were harvested for screening by PCR amplification of the insertion region. For that, primers binding to the promotor region were used (Primers A and E from Table S1). The multiplex PCR kit from Qiagen (Venlo, Netherlands) was used for colony screening, according to manufacturers' protocol. Samples were placed in a thermocycler with the following program: (95 °C, 15 min) once → (95 °C, 30 sec; 58 °C, 90 sec; 72 °C, 60 sec) 34 cycles → (72 °C, 10 min) once → hold at 12 °C.

Afterwards, PCR-product sizes were assessed through a 2% (w/v) agarose gel electrophoresis to validate sgRNA insertion in the vector. Gel Doc™ XR+ Gel Documentation System from Bio-Rad (Hercules, CA, USA) was used to reveal the agarose gel. PCR products were sequenced by Sanger Sequencing.

#### 4.2.2. Plasmid DNA extraction

Extraction of recombinant plasmid DNA was performed using Plasmid maxi kit from Qiagen (Venlo, Netherlands) according to manufacturers' protocol. Briefly, after PCR validation, one positive colony was selected and placed in 100 mL of LB broth with ampicillin to grow ON. After DNA extraction, the eluted DNA was precipitated by addition of 10.5 mL of 2-propanol and centrifuged 17,000 x *g*. DNA pellet was washed using 70% (v/v) ethanol and allowed to air-dry. DNA was dissolved in preheated sterile water. DNA quantity and quality were measured using NanoDrop® ND-1000 UV-Vis Spectrophotometer from ThermoFisher Scientific (Waltham, MA, USA). Quality measured by Abs260/Abs280 ratio (above 1.80).

### 4.3. Generation of stable isogenic cell lines lacking exon v6 using CRISPR/Cas9

MKN-45 e GP-202 cell lines were transfected using Lipofectamine™ 3000 transfection reagent from Invitrogen (Carlsbad, CA, USA). Cells were transfected with several combinations of sgRNAs previously cloned into the vector or with the empty vector. 1 µg of each sgRNA was used in the transfection. Successfully transfected cells were selected using 5 µg/mL and 7.5 µg/mL of puromycin in MKN-45 and GP-202 cell lines, respectively. Puromycin selection started 48 h after transfection and was renewed every 72 h, until colony formation. Three independent transfections were performed.

## 5. Characterization of the generated transient and stable isogenic cell lines

### 5.1. Genomic DNA extraction

Genomic DNA (gDNA) was extracted from cell pellet, using the NZY Tissue gDNA Isolation kit from NZYTech (Lisbon, Portugal), according to manufacturers' protocol. The extracted DNA was eluted in a small amount of preheated to 70 °C sterile water and NanoDrop® ND-1000 UV-Vis Spectrophotometer from ThermoFisher Scientific (Waltham, MA, USA) was used to determine the gDNA quantity and quality. Quality measured by Abs260/Abs280 ratio (above 1.80).

### 5.2. RNA extraction

For RNA isolation, medium was removed and cells were washed twice with 1x PBS. 1 mL of TriPure Isolation Reagent from Sigma-Aldrich (Poole, UK) was added to a 6-well plate, in order to disrupt cell membrane. Afterwards, the cellular components were separated using chloroform, followed by centrifugation (12,000 x *g*; 15 min; 4 °C). The aqueous phase containing the RNA was transferred to a new tube and 2-propanol was added to precipitate RNA, followed by centrifugation (12,000 x *g*; 10 min; 4 °C). RNA pellet was washed with 75% (v/v) ethanol and resuspended in a small amount of sterile water preheated to 55 °C - 60 °C. NanoDrop® ND-1000 UV-Vis Spectrophotometer from ThermoFisher Scientific (Waltham, MA, USA) was accessed to determine RNA concentration and quality. Quality measured by Abs260/Abs280 ratio (above 1.80).

### 5.3. cDNA synthesis

To synthesize complementary DNA (cDNA), 1 µg of template RNA was used. SuperScript® II reverse transcriptase from ThermoFisher Scientific (Waltham, MA, USA) was used, according to the manufacturers' protocol. Briefly, random primers were added to RNA and incubated at 70 °C for 10 min. Afterwards, 0.75 µL of SuperScript® II reverse transcriptase, 4 µL of SuperScript® II buffer, 2 µL of 0.1 M dithiothreitol (DTT), 1 µL of 10 mM deoxynucleotide solution (dNTPs), 0.2 µL of RNAsin (40 U/µL) and 0.75 µL of sterile water was mixed and added to RNA. This mixture was incubated at 37 °C for 1 h for cDNA synthesis and the cDNA was then stored at -20 °C. Non-RT negative controls were also produced by substituting the SuperScript® II reverse transcriptase for sterile water.

### 5.4. gDNA and cDNA genotyping by Sanger Sequencing

gDNA and cDNA from CRISPR/Cas9 clones and cDNA from cells treated with PMOs were sequenced using Sanger Sequencing. Briefly, gDNA and cDNA were amplified by PCR using multiplex PCR kit from Qiagen (Venlo, Netherlands) with the following program: (95 °C, 15 min) once → (95 °C, 30 sec; 56 °C (gDNA) / 60 °C (cDNA), 90 sec; 72 °C, 60 sec) 34 cycles → (72 °C, 10 min) once → hold at 12 °C. The primers used for this reaction were designed to flank exon v6, either in the flanking introns (for gDNA) or the flanking exons (for cDNA). Primer sequences are described in Table S1.

Afterwards, 1 µl of PCR product was purified using ExoSAP-IT Express PCR Cleanup Reagent from ThermoFisher Scientific (Waltham, MA, USA), followed by a thermal cycle (37 °C; 4 min → 80 °C; 1 min). The primers used for the following reaction were the same from the previous reaction. Products were sequenced in forward and reverse manner. To sequence the PCR product, BigDye™ Terminator v3.1 Cycle Sequencing Kit from Applied Biosystems (Foster City, CA, EUA) was utilized according to manufacturers' protocol. A sequencing PCR was carried out (96 °C, 3 min) once → (96 °C, 30 sec; 54 °C, 45 sec; 60 °C, 3 min) 34 cycles → (60 °C, 10 min) once → hold at 12 °C. Afterwards, PCR product was purified using illustra Sephadex G-50 DNA Grade SF from GE Healthcare (Chicago, IL, USA) according to manufacturers' protocol. PCR products were short sequenced, with each primer used to sequence the exon v6 and adjacent sites. The used primers are described in Table S1 (Primers B / F for cDNA sequencing, C / G for gDNA sequencing, D / H for total *CD44* cDNA sequencing).

### 5.5. qRT-PCR

*CD44v6* and total *CD44* gene expression were assessed by qRT-PCR, using probes specific for *CD44v6* (exon span v5-v6, Hs.PT.58.45400024) and *CD44* total (exon span 2-3, Hs.PT.58.4880087) from iDT (Coralville, IA, USA). qRT-PCR was performed in triplicate using 1 µL of template cDNA per well. Briefly, a mix containing a 5 µL of TaqMan Master Mix, 0.2 µL of Rox Low reagent both from ThermoFisher Scientific (Waltham, MA, USA), 3.3 µL of sterile water and 0.5 µL of the desired probe was added to 1 µL of template cDNA and placed on an ABI Prism 7500 Sequence Detection System.

*CD44v6* and total *CD44* relative expression was calculated by comparative $2^{-\Delta\Delta C_T}$ method using the housekeeping gene *18s* (custom assay) from iDT (Coralville, IA, USA) in order to normalize and obtain the expression results between samples ($RQ = 2^{-\Delta\Delta C_T}$, where $\Delta\Delta C_T = \Delta C_{T\,Sample} - \Delta C_{T\,Control\ sample}$ and $\Delta C_T = C_{T\,gene\ of\ interest} - C_{T\,endogenous\ gene\ control}$).

### 5.6. Flow cytometry

Flow cytometry was used to determine the expression of total CD44 and CD44v6 in GC cell lines. Cells were seeded in 6-well plates and allowed to grow until cells were approximately 90% confluent. Cells were then washed once with 1x PBS, detached using Versene (ThermoFisher Scientific, Waltham, MA, USA) for approximately 30 min and washed in ice-cold 3% (w/v) BSA – PBS. Cells were then blocked in 3% (w/v) BSA – PBS for 30 min and incubated with the primary antibody. The primary antibodies used were: mouse monoclonal antibody against CD44 (clone 156-3C11; 1:100 dilution; 60 min incubation; Cell Signaling Technology, Beverly, MA, USA) and against CD44v6 (MA54; 1:100; 60 min incubation; ThermoFisher Scientific, Waltham, MA, USA). Cells were subsequently washed with 3% (w/v) BSA – PBS and incubated with the anti-mouse Alexa Fluor 647 (1:500; 30 min incubation; ThermoFisher Scientific, Waltham, MA, USA) secondary antibody.

Fluorescence was analyzed using FACS Canto II from BD Biosciences (Franklin Lakes, NJ, USA). Mean fluorescence intensity was measured for at least 15,000 gated events per sample and Flow Jo version 10 software was used to analyze the data.

### 5.7. Immunofluorescence

Immunofluorescence was used to determine total CD44 and CD44v6 expression and protein location in GC cell lines. Cells were seeded in 6-well plates with glass cover slips in the bottom and allowed to grow until cells were 80% - 90% confluent. Afterwards, cells were washed three times with 1x PBS and fixed with 4% (v/v) PFA (Merck, Darmstadt, Germany) for 15 min. After cell fixation, cells were washed again with PBS and incubated with 50 mM $NH_4Cl$ for 10 min. Afterwards, cells were incubated in 0.2% TritonX-100 for 5 min and subsequently blocked with 5% (w/v) BSA – PBS for 30 min. After blocking, cells were incubated ON with the primary antibodies. The primary antibodies utilized were: mouse mAb anti-CD44v6 (clone MA541; 1:100 dilution; ThermoFisher Scientific, Waltham, MA, USA) and mouse mAb anti-CD44 antibody (156-3C11; 1:100 dilution; Cell Signaling Technology, Beverly, MA, USA). Cells were then washed three times with 1x PBS and incubated in the dark with secondary antibody Alexa Fluor 488 Donkey Anti-Mouse secondary antibody (1:500; 2 h incubation; Life Technologies, Carlsbad, CA, USA). Cells were washed, incubated with DAPI (CAS Number 28718-90-3; 1:1000; 5 min incubation; Sigma-Aldrich, Poole, UK), washed and mounted in Vectashield mounting media (Vector Laboratories, Burlingame, CA, USA). Cells were analyzed by fluorescence microscopy (Imager.Z1,

AxioCam fluorescence microscope or Eclipse TE-2000, Zeiss, Gottingen, Germany) using AxioVision software (Rockville, MD, USA).

## 6. Cell treatment with chemotherapeutic agents

GP-202 and MKN-45 stable isogenic cell lines lacking exon v6, and respective Mock controls, were treated with cisplatin and 5-Fluorouracil (5-FU), two chemotherapeutic agents frequently used in GC therapy. Drug response in the different cells/conditions was assessed using three different cell survival assays as described below.

Depending on the experiments being performed, an additional control was also used, where GP-202 and MKN-45 cells were transfected with a short interference RNA (siRNA) targeting CD44v6 in order to destroy all CD44v6-containing isoforms (as detailed in 6.4.).

### 6.1. Optimization of the short-term and long-term experiments

To test cell survival in the presence of chemotherapeutic drugs compared to vehicle control, two techniques were used: PrestoBlue (PB) and Sulforhodamine B (SRB). Both PB and SRB assays can provide an indirect estimation of cell number, however, PB is a resazurin-based assay that is reduced by metabolically active cells, while SRB is a protein-binding dye that stains protein content.

To evaluate the number of cells to use for each cell line in the short-term treatments, a dilution series ranging from 40,000 cells to 625 cells of MKN-45 and GP-202 cell lines were evaluated by PB and SRB on the day that cells were meant to be treated (T0 h) and 48 h after (T48 h), as shown in Figure 14.

In order to choose a number of cells that, after 48 h, cells would still be growing exponentially without reaching absorbance levels where the SRB assay is known to saturates (absorbance above 2.0), 2500 cells and 4000 cells for MKN-45 and GP-202 were selected, respectively.

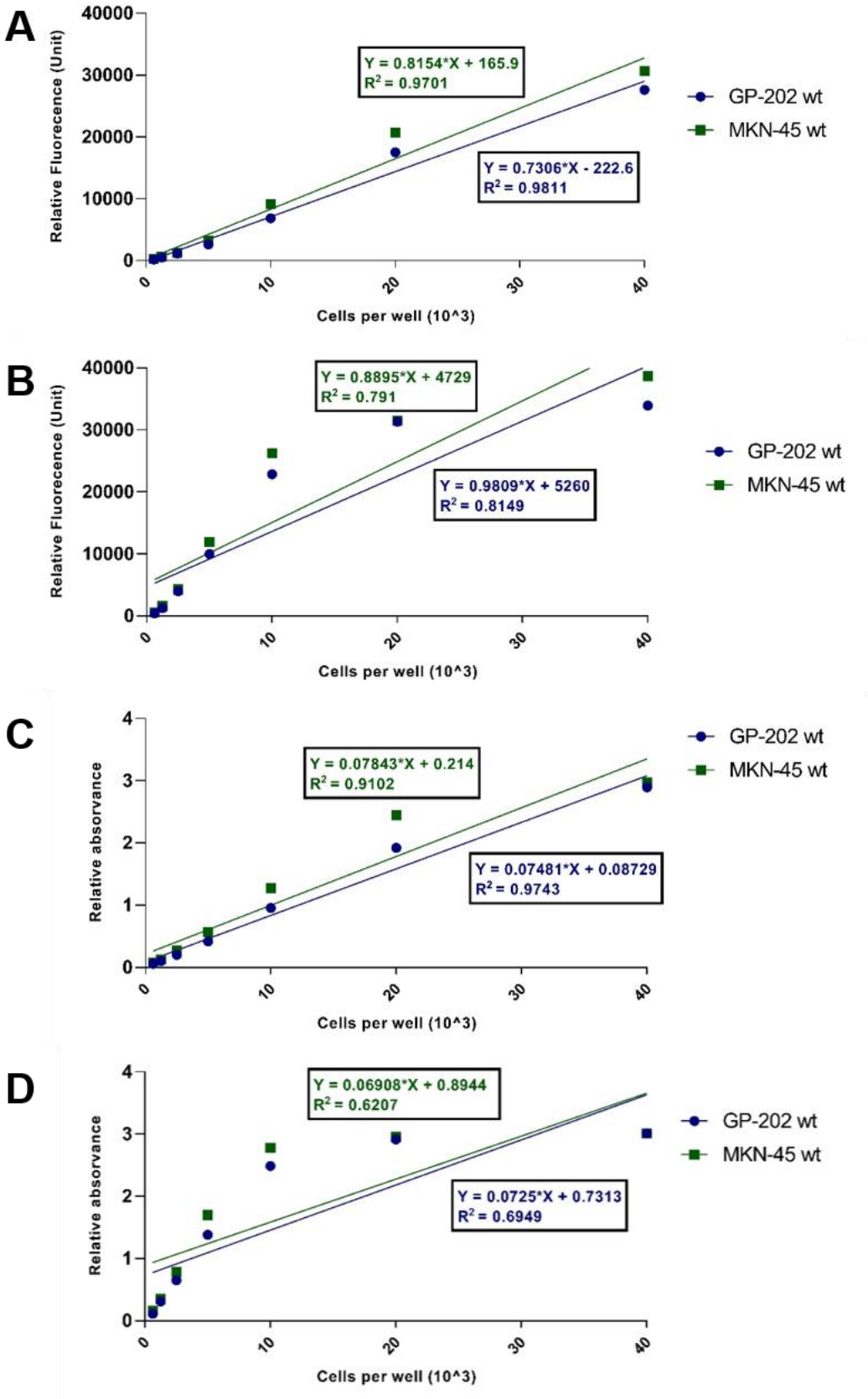


Figure 14 – Analysis of cell growth 24 h and 72 h after seeding for GP-202 and MKN-45 *wt* cell lines. (A) 24 h and (B) 72 h after seeding, analysis by PB assay; (C) 24 h and (D) 72 h after seeding, analysis by SRB assay.

For the long-term treatments, the number of cells to use for each cell line was assessed. MKN-45 and GP-202 *wt* cells were placed in 6-well plated with, 300, 1000 and 5000 cells per well and allowed to grow for several days until colonies were approximately 50 cells each. After 10 days, for MKN-45, 1000 cells were the optimal number of cells to carry out the treatments, as for GP-202, an intermediate concentration of cells was chosen, therefore 3000 cells were selected to plate in subsequent experiments (data not shown).

Regarding the concentrations of chemotherapeutic agents to use, MKN-45 and GP-202 *wt* and edited clones were subjected to a range of concentrations of cisplatin and 5-FU. Cisplatin was tested from 0.01 µM to 20 µM for GP-202 and from 0.001 µM to 2.5 µM for MKN-45, while 5-FU was tested from 0.01 µM to 5 µM for both cell lines (data not shown). The concentrations chosen to continue further experiments were 0.1 µM of cisplatin and 0.5 µM of 5-FU, for GP-202 cell lines, and 0.5 µM of cisplatin and 2.5 µM of 5-FU, for MKN-45 cell lines.

### 6.2. Short-term cell survival assessment: PB and SRB assays

PB and SRB assays were used to assess cell survival after 48 h treatments. MKN-45 and GP-202 mock cells transfected with siRNA Scramble and CD44v6, and the selected edited clones were plated in 96-well plates, with $2.5x10^3$ and $4x10^3$ cells per well, respectively. Cells adhered for approximately 24 h, under normal conditions, were either treated with cisplatin, 5-FU or vehicle. Cisplatin was reconstituted in 0.9% NaCl and 5-FU in sterile water, and both diluted in RPMI medium to the desired concentrations. Vehicle used had the equivalent concentration to the higher concentration of drug used, but with NaCl in RPMI medium. After 48 h, cell medium was removed and cells were washed once with 1x PBS.

10x PB from Invitrogen (Carlsbad, CA, USA) was diluted in RPMI medium to a final concentration of 1x and added to cells. Cells were incubated for 45 min under normal conditions and, afterwards, fluorescence was measured at 560 nm using a microplate reader PowerWave HT Microplate Spectrophotometer from BioTek (Bad Friedrichshall, Germany).

Afterwards, cells were fixed in 10% trichloroacetic acid (TCA) for 1 h on ice and proteins were stained with 0.4% SRB solution from Sigma-Aldrich (Poole, UK) for 30 min. Cells were subsequently washed with 1% (v/v) acetic acid to remove any unbound dye, and protein stain was solubilized with 10 mM Tris solution. Using a microplate reader PowerWave HT Microplate Spectrophotometer from BioTek (Bad Friedrichshall, Germany), SRB

absorbance was measured at 560 nm, with background correction at 655 nm. Three independent experiments were performed, each in triplicate.

To calculate % of cell survival, fluorescence/absorbance of treatment conditions was normalized to vehicle (% cell survival = ($Treatment_{Fluo/Abs}$ * 100) / $Vehicle_{Fluo/Abs}$).

### 6.3. Long-term cell survival assessment: Clonogenic assay

Clonogenic assay was used to access the cell lines colony formation ability when treated with cisplatin, 5-FU or vehicle. Cisplatin was reconstituted in 0.9% NaCl and 5-FU in sterile water, and diluted in RPMI medium to the desired concentrations. Vehicle used had the equivalent concentration to the higher concentration of drug used, but with 0.9% NaCl in RPMI medium. Mock cells transfected with siRNA Scramble, siRNA for CD44v6 and the selected edited clones were plated in 6-well plates, with $1x10^3$ and $3x10^3$ cells per well for MKN-45 and GP-202 cell lines, respectively. Cells were allowed to adhere and 24 h and then treated with 0.1 µM of cisplatin and 0.5 µM of 5-FU (GP-202 cells), or 0.5 µM of cisplatin and 2.5 µM of 5-FU (MKN-45 cells). After 48 h, RPMI medium containing the chemotherapeutic agents was removed, cells were washed with PBS and 2 mL of new RPMI medium was added to cells. Cells were then incubated for several days until colonies had approximately 50 cells each (10 days). Afterwards, cells were fixed with cold methanol and stained with 0.5% (v/v) crystal violet from Sigma-Aldrich® (Poole, UK). The number of colonies formed upon drug treatments were compared to vehicle. Three independent experiments were performed, each in triplicate.

To calculate % of cell survival, number of colonies of each treatment conditions was normalized to vehicle (% cell survival = ($Treatment_{Nº\ of\ colonies}$ * 100) / $Vehicle_{Nº\ of\ colonies}$).

### 6.4. RNA interference (RNAi)

GP-202 and MKN-45 cell lines were transfected with a siRNA targeting *CD44v6*. The transfection reagent used was Lipofectamine® RNAiMax from ThermoFisher Scientific (Waltham, MA, USA), according to manufacturers' instructions. Briefly, cells were seeded in 6-well plates with $2x10^5$ cells and $1.5x10^5$ cells per well for GP-202 and MKN-45, respectively. After 24 h, lipid based conjugates were prepared by mixing either 20 nM of siRNA Scramble (negative control DS NC1; iDT, Leuven, Belgium), or 20 nM of human siRNA CD44v6 (Sense strand: 5'- GCGUCAGGUUCCAUAGGAAUCCUTT - 3'; Antisense strand: 5'- AAAGGAUUCCUAUGGAACCUGACGCAG - 3', custom made from iDT, Leuven, Belgium), to diluted Lipofectamine® RNAiMax Reagent (1:1 ratio). Cell medium was

replaced with fresh medium and conjugates were incubated for 5 min at room temperature and added dropwise to cells. Cells were incubated for 24 h, after which they were detached, counted and plated for different assays.

### 6.5. Assessment of chemotherapy response in isogenic cell lines lacking exon v6 or exon v6 containing isoforms

MKN-45 and GP-202 mock cells and corresponding edited clones were seeded in 6-well plates and allowed to adhere ON. Afterwards, cells were transfected with either siRNA Scramble or CD44v6 siRNA. After 24 h incubation, cells were detached, counted and replated for up to three distinct assays: i) short-term (PB and SRB assay) and ii) long-term (clonogenic assay) cell survival assessment, and for CD44v6 expression assessment by immunofluorescence (in the case of mock cells transfected with siRNA Scramble or CD44v6 siRNA). 24 h after being replated, both short-term and long-term cells were treated either with cisplatin, 5-FU or vehicle. After 48 h, short-term treatment was assessed by PB and SRB assays (as detailed in section 6.2) and long-term treatment cells were washed and medium was renewed. After 10 days, long-term treatment was assessed (as detailed in 6.3).

## 7. Statistical analysis

Statistical analysis was performed using GraphPad Prism version 7.00 software (GraphPad Software Inc., CA, USA). Normality test was performed using Shapiro-Wilk test, to see the distributions of each condition. Two-way ANOVA, with Tukey's Post Hoc Test for multiple comparison analysis was used, with the confidence interval of 95%. Significant differences were considered at $p < 0.05$, $< 0.01$, $< 0.001$ and $< 0.0001$.

# Chapter IV | Results

# 1. Experimental design for induction of transient and permanent skipping of exon v6 from *CD44v6*-containing transcripts

## 1.1. Concept and design of PMOs targeting exon v6 boundaries

To develop a transient model for the skipping of exon v6 from the *CD44v6*-containing transcripts, two different PMOs targeting two splicing sites in the boundaries of exon v6 were used. To design these PMOs, the region to target was assessed to properly obtain the skipping of exon v6. The best regions for that are the splice sites at the exon boundaries, splicing acceptors and donors. To underline the splicing acceptors and donors near exon v6, *Human Splicing Finder* (http://www.umd.be/HSF3/HSF.shtml) website was used. The DNA sequence from the *CD44* gene (ENST00000415148.6) was analyzed and several possible splicing acceptors and donors were identified. The canonical splicing acceptor and donor were localized in the intron/exon and exon/intron boundaries of the exon v6, respectively. The PMOs were designed to target these regions, hence resulting in the skipping of exon v6 and consequently in the splicing of exon v5 and exon v7 together, as demonstrated in Figure 15. The specific sequences of the acceptor and donor targeted by the PMOs are stated in Table 2. The PMO sequences are detailed in Table 3.

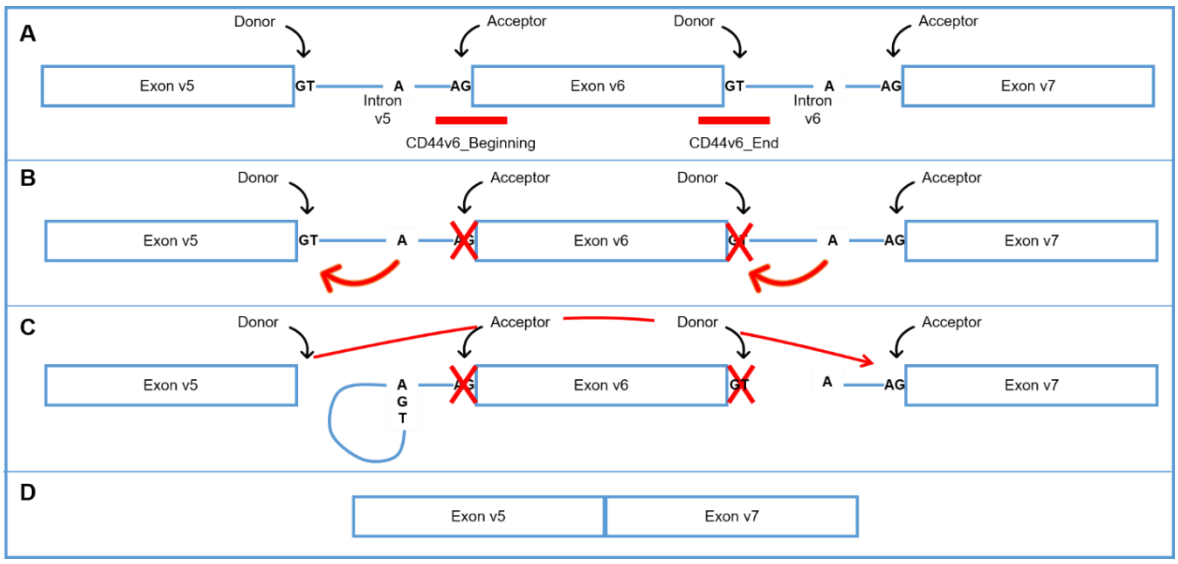


Figure 15 - Skipping of exon v6 using PMOs. (A) PMOs target either the acceptor or the donor site near exon v6, disguising it. (B) The splicing machinery starts, and the 2'OH of a specific branch point within the intron attacks the donor site right after, forming a loop. (C) Afterwards, the donor site 3'OH attacks the acceptor site right after, however, since the acceptor is masked, the donor attacks the next acceptor site, (D) leading to the skipping of the exon, and the splicing together of the previous and subsequent exons.

Table 2 – Potential splice sites near exon v6 of the CD44 gene detected by Human Splicing Finder

| Splice site type | Potential splice site | Score | Location |
|---|---|---|---|
| **Acceptor** | ttcttctcacagTC | 91.44 | Int v5^Exo v6 |
| **Donor** | CTGgtaatg | 81 | Exo v6^Int v6 |

Table 3 – PMOs designed to target the splice sites near exon v6 of the CD44 gene

| PMO ID | Sequence | Location |
|---|---|---|
| **CD44v6_beginning** | 5' CTGGACTGTGAGAAGAATATCAGTT 3' | Int v5^Exo v6 |
| **CD44v6_end** | 5' CTTGTTAAACCATCCATTACCAGCT 3' | Exo v6^Int v6 |

Both PMO sequences were blasted using the NCBI Basic Local Alignment Search Tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi) to search for possible off targets in different genes and no significant off target locations appeared for any sequence.

### 1.2. Concept and design of sgRNAs for CRISPR/cas9 genome editing targeting exon v6 boundaries and adjacent intronic regions

To develop a stable model for the skipping of exon v6 from the *CD44v6*-containing transcripts, sgRNAs were designed to either cause a DSB in the splice site near exon v6 (by using only one sgRNA), disrupting it and consequently leading to splicing out of exon v6, or to cause a large deletion (by using two sgRNAs), resulting in the deletion of the entire exon without disrupting the reading frame, as demonstrated in Figures 16 and 17, respectively. Taking these two approaches into account, 6 different sgRNA were designed to target either the splice acceptor and donor near exon v6 or to target the adjacent intronic regions.

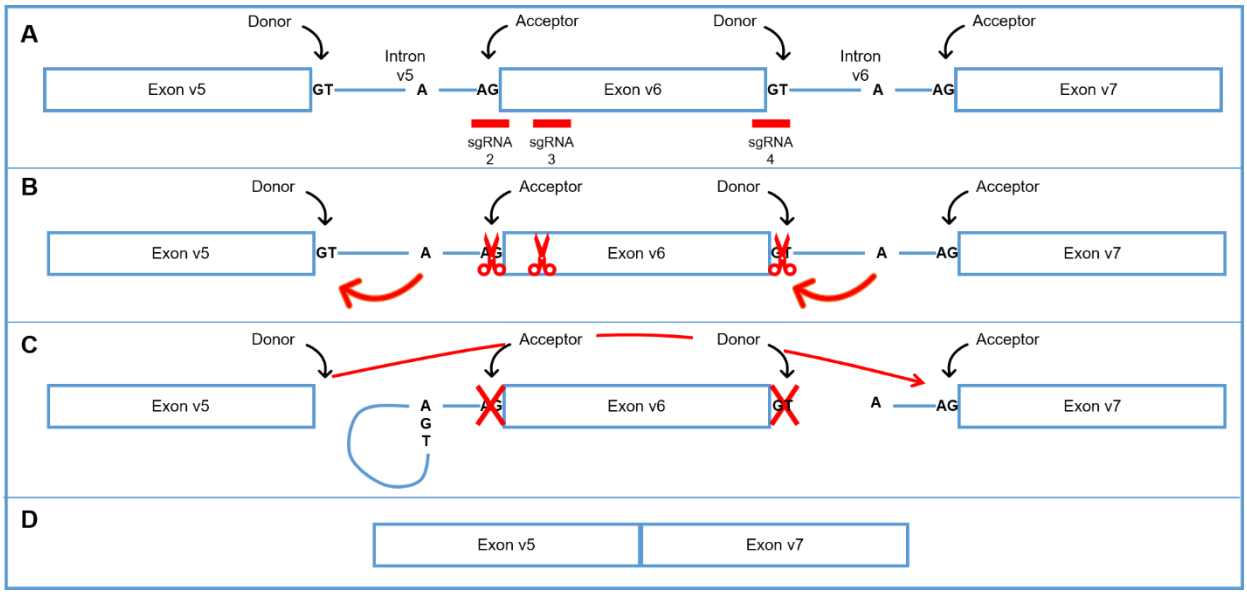


Figure 16 - Skipping of exon v6 using CRISPR/Cas9. (A) One sgRNA targets the acceptor or the donor site near exon v6, disguising it. (B) The splicing machinery starts, and the 2'OH of a specific branch point within the intron attacks the donor site right after, forming a loop. (C) Afterwards, the donor site 3'OH attacks the acceptor site right after, however, since the acceptor is masked, the donor attacks the next acceptor site, (D) leading to the skipping of the exon, and the splicing together of the previous and subsequent exons.

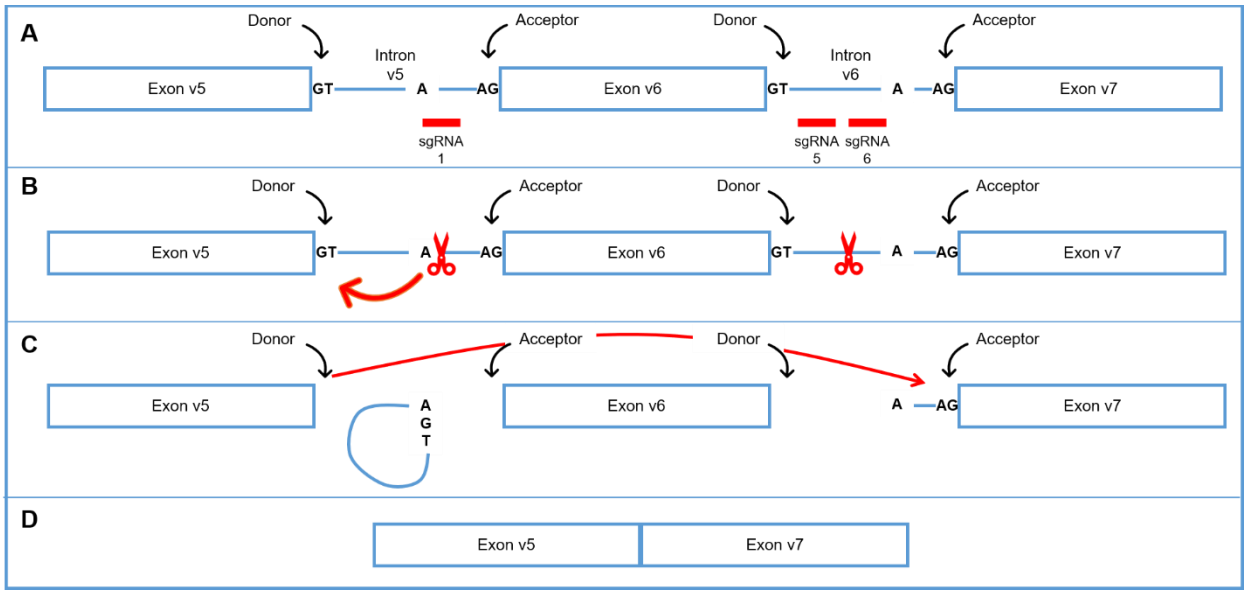


Figure 17 – Mimicking the skipping of exon v6 using CRISPR/Cas9. (A) Using two sgRNAs that target the adjacent introns of exon v6. (B) and (C) The deletion caused is so large that results in the deletion of the whole exon v6, including it splice sites, mimicking the skipping of exon v6, (D) leading to the skipping of the exon, and the splicing together of the previous and subsequent exons.

To locate the possible sgRNA target regions in the location of interest, *Benchling online tool* (https://benchling.com/) was assessed to identify and score the target regions in the flanking introns and within exon v6. These target regions must possess a protospacer adjacent motif with an NGG sequence downstream to be recognized by the *Cas9* protein.

Six sgRNAs sequences were selected according to their score and strategic target location, as shown in Table 4. An extra "CACCG" sequence was placed in the beginning of all sgRNAs to insert the sgRNAs into the pSpCas9(BB)-2A-Puro (PX459) V2.0 plasmid, through the *Bbs I* digestion site.

Table 4 – Specifications of the six sgRNAs design to target exon v6, exon v6 boundaries and adjacent introns.

| sgRNA | Sequence | PAM | Score | Strand | Location |
|---|---|---|---|---|---|
| **1** | 5' CACCGTATGTTGACAGCTATTGGTG 3' | AGG | 76 | + | Intron v5 |
| **2** | 5' CACCGTGATATTCTTCTCACAGTCC 3' | AGG | 59 | + | Int v5^Exo v6 |
| **3** | 5' CACCGGGCAACTCCTAGTAGTACAA 3' | CGG | 83 | + | Exon v6 |
| **4** | 5' CACCGCAGGGACAGCTGGTAATGGA 3' | TGG | 61 | + | Exo v6^Int v6 |
| **5** | 5' CACCGGGGTTAAACGGTAGACATTG 3' | AGG | 83 | + | Intron v6 |
| **6** | 5' CACCGAGGAATTGTCACGAGATGTT 3' | AGG | 81 | + | Intron v6 |

All sgRNA sequences were blasted using the NCBI Basic Local Alignment Search Tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi) to search for possible off targets in different genes and no significant off target locations appeared for any of the sequences.

## 2. Skipping of exon v6 from *CD44v6*-containing transcripts is effective using PMOs and CRISPR/cas9 technologies

### 2.1. The transient approach – PMOs

PMOs were used as a transient antisense strategy, and are constituted by 25 subunits that targets RNA without its degradation. Regarding the concept of exon skipping, PMOs are a suitable strategy since the mRNA is maintained and translated into a functional protein (71). It has been reported that PMOs can maintain their function until approximately 9 days (89).

#### 2.1.1. Optimization of *CD44* exon v6 skipping using PMOs

For each cell line, GP-202 and MKN-45, the concentration of transfection reagent required to internalize PMOs (with the minimal toxicity to cells) was optimized, using a fluorescent PMO, as shown in Figure 18. Two concentrations of EP transfection reagent were tested and results show that, for both cell lines, 2 µM of transfection reagent resulted in PMO internalization in a greater percentage of cells, when comparing to 6 µM. The 2 µM caused no significant changes in cell viability in both cell lines (data not shown).

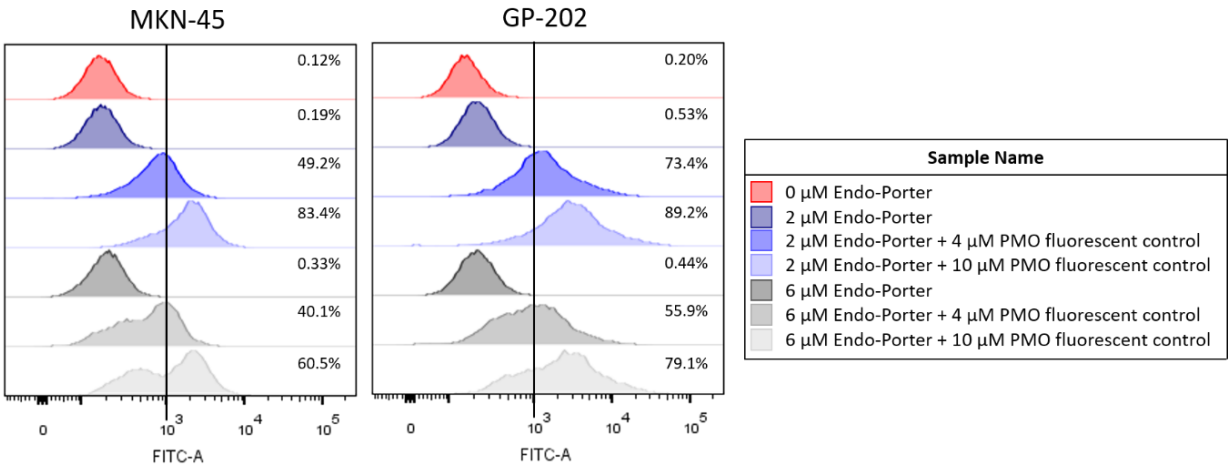


Figure 18 – Optimization of PMO transfection into MKN-45 (left) and GP-202 (right) cells. Cells were transfected with different concentrations of transfection reagent (Endo-Porter) and fluorescence labelled PMO. Transfection efficiency was assessed by quantifying cell fluorescence by Flow cytometry. Results are representative of one experiment.

To evaluate the CD44v6_Beginning and CD44v6_End PMOs efficiency to inhibit exon v6, two different concentrations were selected for each cell line, based on the experiments performed with PMO fluorescent control (1 µM and 4 µM for GP-202 and 4 µM and 6 µM for MKN-45 cells). To assess the efficiency of the designed CD44v6 PMOs in inhibiting exon v6, the total CD44 and exon v6 expression were analyzed by qRT-PCR, as shown in Figure 19.

Results obtained by qRT-PCR show that both CD44v6_Beginning and CD44v6_End PMOs are more efficient in inhibiting exon v6 in GP-202 than in MKN-45 cells. For subsequent experiments, 4 µM of CD44v6_Beginning and CD44v6_End was used to transfect GP-202 cells. For the MKN-45 cell line, 4 µM of CD44v6_Beginning and 6 µM of CD44v6_End were used.

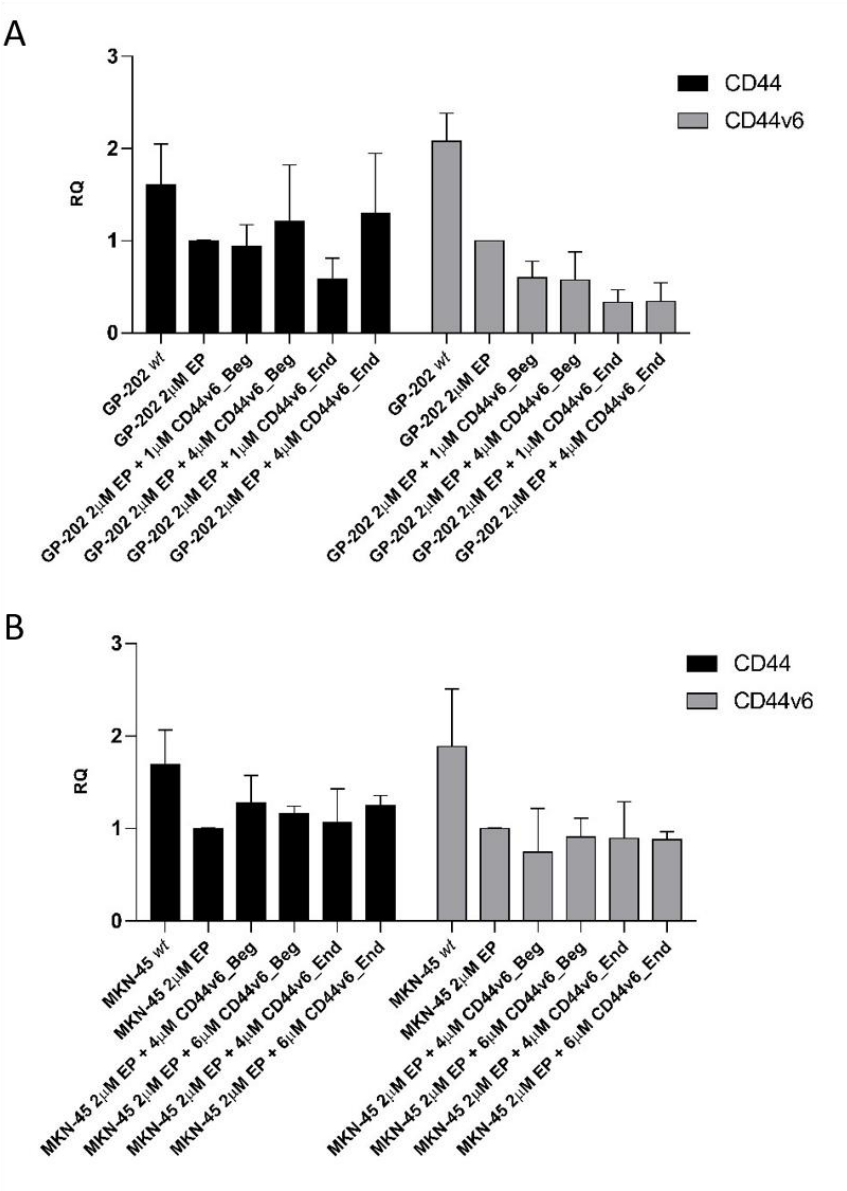


Figure 19 – CD44 total and CD44v6 expression in GP-202 and MKN-45. qRT-PCR results for the optimization of the concentration of CD44v6_Beginning and CD44v6_End for (A) GP-202 and (B) MKN-45 cell lines.

### 2.1.2. Characterization of PMOs-induced *CD44* exon v6 skipping

To understand if PMOs caused the skipping of exon v6 from *CD44v6*-containing transcripts, cDNA from parental and transfected cells were amplified from exon v5 to exon v7 and analyzed by gel electrophoresis, as shown in Figure 20.

Results show that the untreated parental cells (lanes 1 and 5) and cells incubated with transfection reagent alone (lanes 2 and 6) possess only one transcript (B), whereas cells transfected with CD44v6_Beginning (lanes 3 and 7) and CD44v6_End (lanes 4 and 8) present two new transcripts (A and C).

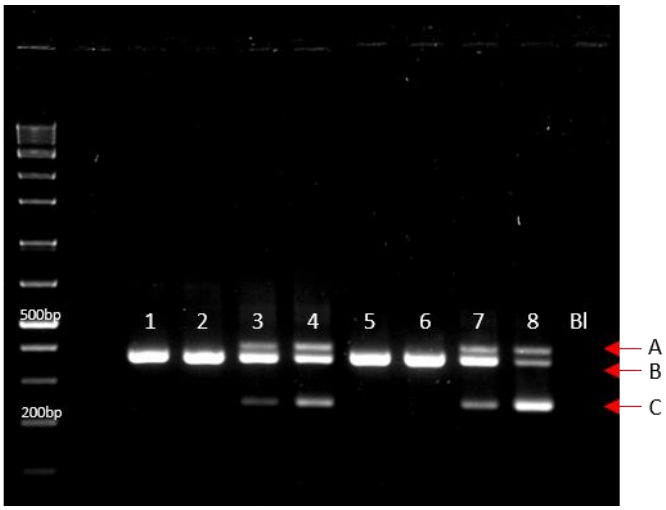


Figure 20 – Transcript analysis following PMOs transfection in MKN-45 and GP-202 cells. MKN-45: (1) untreated control, (2) EP, (3) EP+CD44v6_Beginning PMO, (4) EP+CD44v6_End PMO; GP202: (5) untreated control, (6) EP, (7) EP+CD44v6_Beginning PMO, (8) EP+CD44v6_End PMO; (Bl) Negative control. EP – Endo-Porter. PCR products correspond to (A) artefact originated by a heterodimer of PCR products B and C; (B) splicing of exon v5, v6 and v7; (C) skipping of exon v6.

To confirm that these PCR products encompass the sequence of the expected transcripts, the cDNA from each condition was sequenced (Figure 21). The sequences obtained show that the PCR product B includes variant exons 5, 6 and 7 from the *CD44* gene, which is the transcript present in parental cells. On the other hand, PCR product C reveals the skipping of exon v6, resulting in exon v5 and v7 being spliced together.

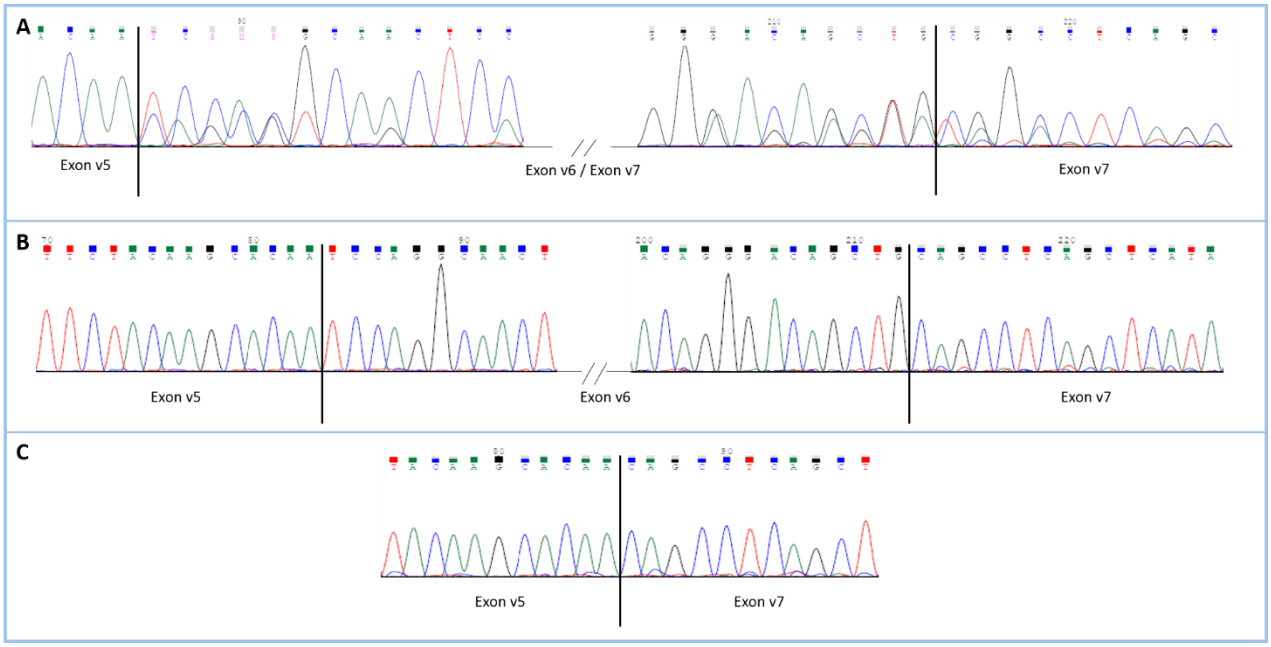


Figure 21 - Analysis of the new transcripts formed through PMO transfection by Sanger sequencing technique (representative of both MKN-45 and GP-202 cell lines). Resulting sequences from (A) the artefact originated by a heterodimer of sequences B and C; (B) the splicing of exons v5, v6 and v7; (C) the skipping of exon v6.

PCR product A represents an artefact originated by a heterodimer of the two PCR products, confirmed by the presence of two different sequences after exon v5: the sequence of exon v6 and the sequence of exon v7.

The RNA expression of CD44v6 and total CD44 in parental and transfected cells was then assessed, as shown in Figure 22. In GP-202 cell line, CD44v6 expression is significantly lower in cells transfected with either the CD44v6_Beginning or CD44v6_End PMO. However, in MKN-45 cell line, no significantly decrease in CD44v6 expression was observed using either CD44v6_Beginning or CD44v6_End.

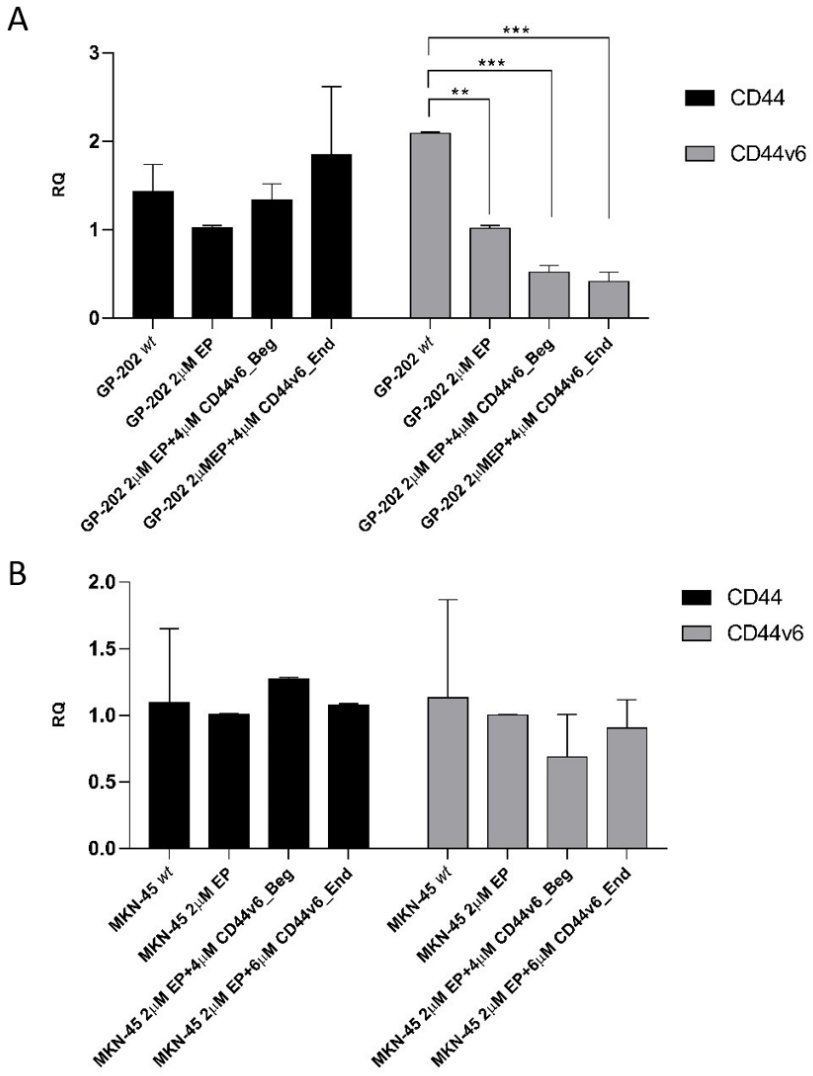


Figure 22 – Total *CD44* and *CD44v6* gene expression in (A) GP-202 and (B) MKN-45 cell lines. All conditions were normalized to transfection reagent (EP) control. Results represent the average + SD of 3 independent experiments. p= * <0.05; ** <0.01; *** <0.001; **** <0.0001.

These results indicate that both PMOs are efficient in achieving exon v6 skipping from *CD44v6*-containing transcripts in GP-202 cells, which was not consistently shown in MKN-45 cells.

Immunofluorescence, performed in GP-202 cells, confirms that, CD44v6 expression is inhibited by PMOs, while expression of other CD44 isoforms is maintained (assessed by using an antibody targeting a canonical region of CD44), as shown in Figure 23.

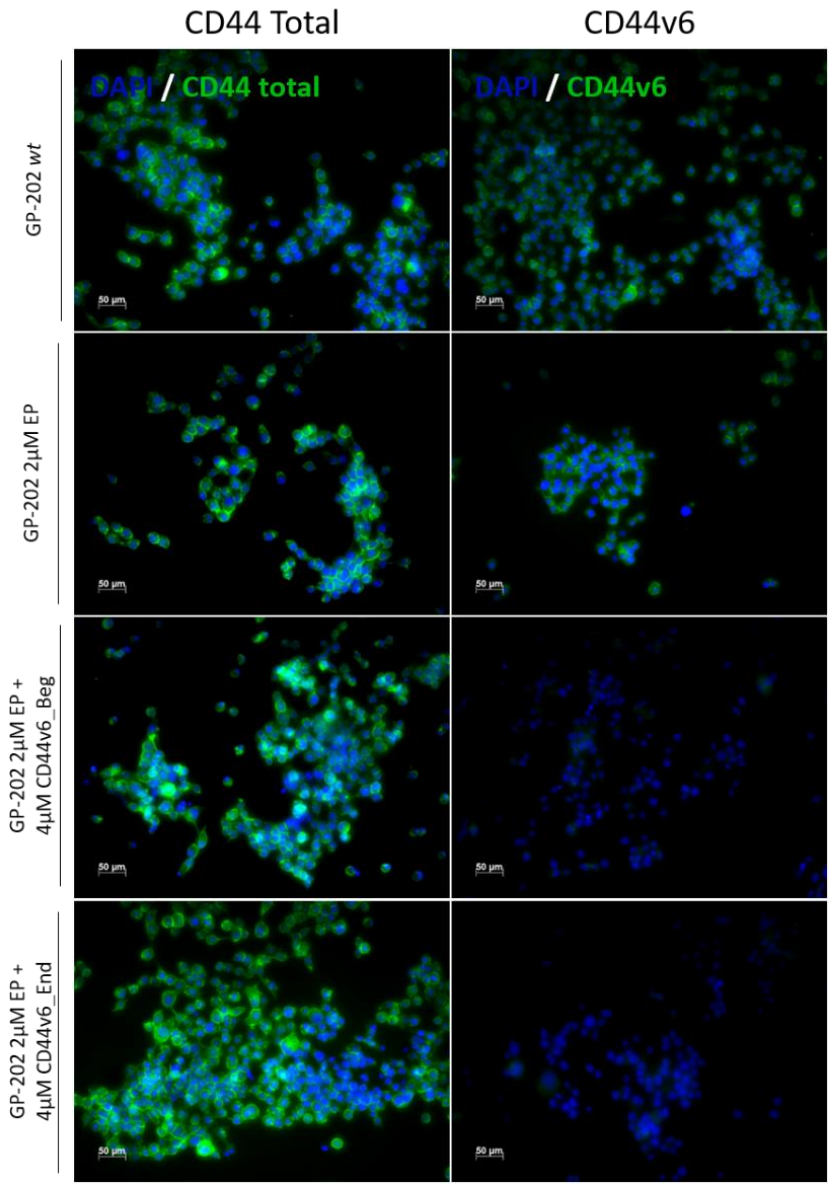


Figure 23 - Immunofluorescence for total expression of CD44 and specific expression of exon v6 from CD44 in GP-202 cells transfected with PMOs. Nuclei are stained with DAPI (represented in blue) and white scale bars represent a distance of 50 µm.

### 2.2. The permanent approach – CRISPR/cas9 system

CRISPR/Cas9 is the most recent approach using target nucleases. This system can be used to efficiently engineer the genome from eukaryotic cells by targeting approximately 20 nucleotides from a DNA of interest. The double-strand break caused by the *Cas9* nucleases can be repaired via NHEJ or HDR, resulting in a permanent mutation in the target genome (76).

#### 2.2.1. Cloning the sgRNAs into a Cas9 containing vector

In order to successfully edit GP-202 and MKN-45 cell lines, each sgRNA was cloned into the pSpCas9(BB)-2A-Puro (PX459) V2.0 plasmid, which contains the *Cas9* gene, the ampicillin gene, for positive colony selection, and the puromycin gene, for edited cell selection.

The PX459 vector contains a *Bbs I* restriction site downstream the U6 promoter. Using *Bbs I* restriction enzyme, the vector was digested and the sgRNAs were inserted into the open vector. Afterwards, the plasmid was inserted into Stbl3™ and the DNA obtained from positive colonies was sequenced in order to verify the correct insertion of the sgRNAs into the vector, as shown in Figure 24.

After confirming the sequences of the vector containing the sgRNAs, bacteria containing the vector was expanded and bacterial DNA was extracted. Sequencing was also performed following maxiprep DNA extraction and it was confirmed that bacteria massive expansion did not cause any mutations in the vectors' DNA (data not shown).

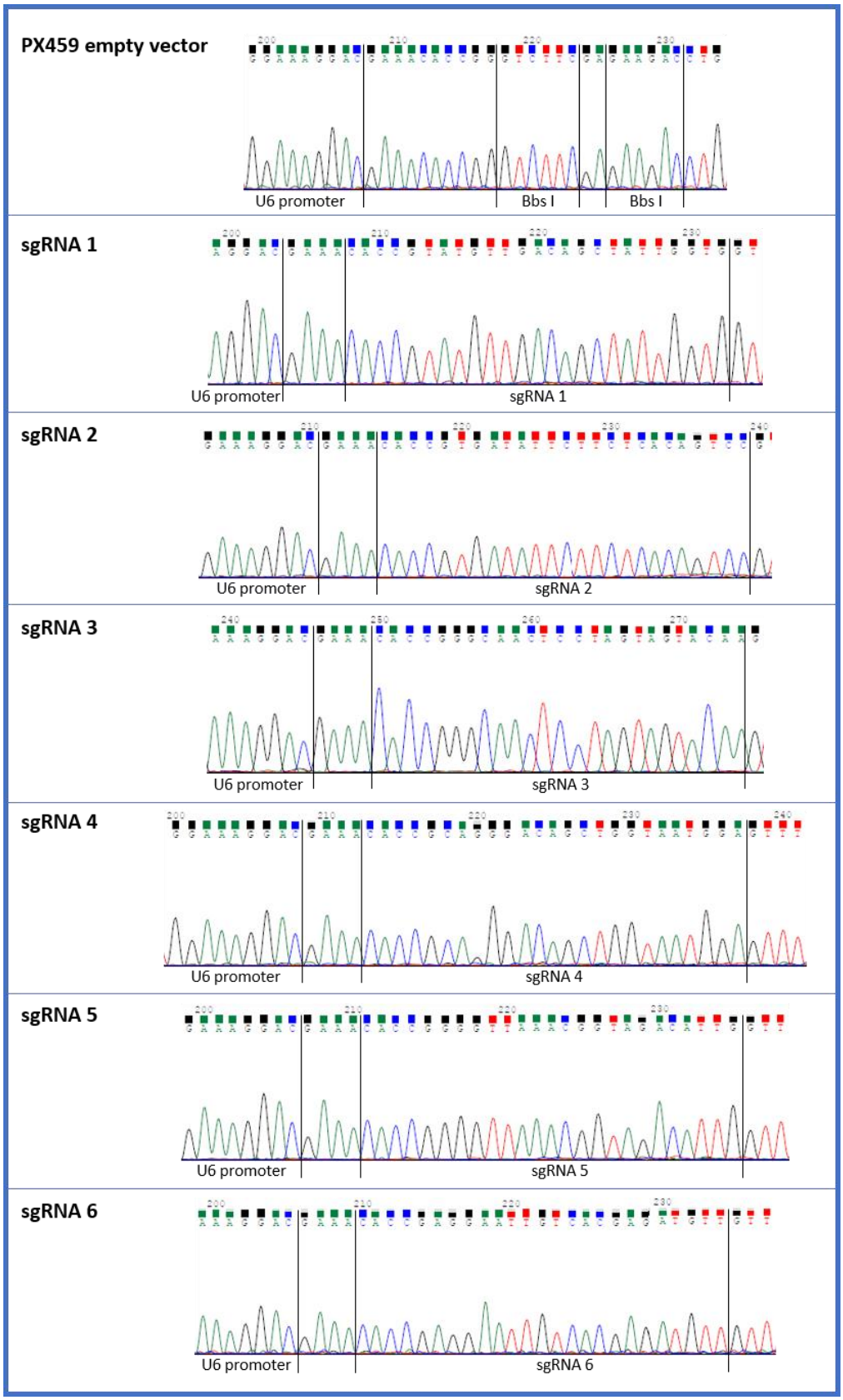


Figure 24 – Sanger sequencing for analysis of the insertion of each sgRNA into Bbs I restriction site. The sequence for the empty vector is also shown.

### 2.2.2. Establishment and characterization of the *CD44* exon v6 skipping permanent models

The six sgRNAs were single or paired transfected into MKN-45 and GP-202 cells to test both exon skipping by disruption of the splice sites or by the deletion of the whole exon, as shown in Table 5. The empty vector was also transfected into cells (MKN-45 mock and GP-202 mock).

Table 5 – Identification of MKN-45 and GP-202 possible edited clones according to the sgRNAs transfection.

| MKN-45 cell line | | | GP-202 cell line | | |
|---|---|---|---|---|---|
| **Clone** | **sgRNA 5'** | **sgRNA 3'** | **Clone** | **sgRNA 5'** | **sgRNA 3'** |
| **MKN-45_O15** | 1 | 5 | **GP-202_O15** | 1 | 5 |
| **MKN-45_O16** | 1 | 6 | **GP-202_O16** | 1 | 6 |
| **MKN-45_O2** | 2 | ----- | **GP-202_O2** | 2 | ----- |
| **MKN-45_O3** | 3 | ----- | **GP-202_O3** | 3 | ----- |
| **MKN-45_O4** | 4 | ----- | **GP-202_O4** | 4 | ----- |
| **MKN-45_O24** | 2 | 4 | **GP-202_O24** | 2 | 4 |
| **MKN-45_O14** | 1 | 4 | **GP-202_O14** | 1 | 4 |
| **MKN-45_O25** | 2 | 5 | **GP-202_O25** | 2 | 5 |
| **MKN-45_O26** | 2 | 6 | **GP-202_O26** | 2 | 6 |

After puromycin selection, only certain sgRNA combinations resulted in edited GP-202 and MKN-45 cells. To evaluate the sgRNAs efficiency in removing exon v6, DNA was extracted and amplified for exon v6 and adjacent intronic regions, as shown in Figure 25.

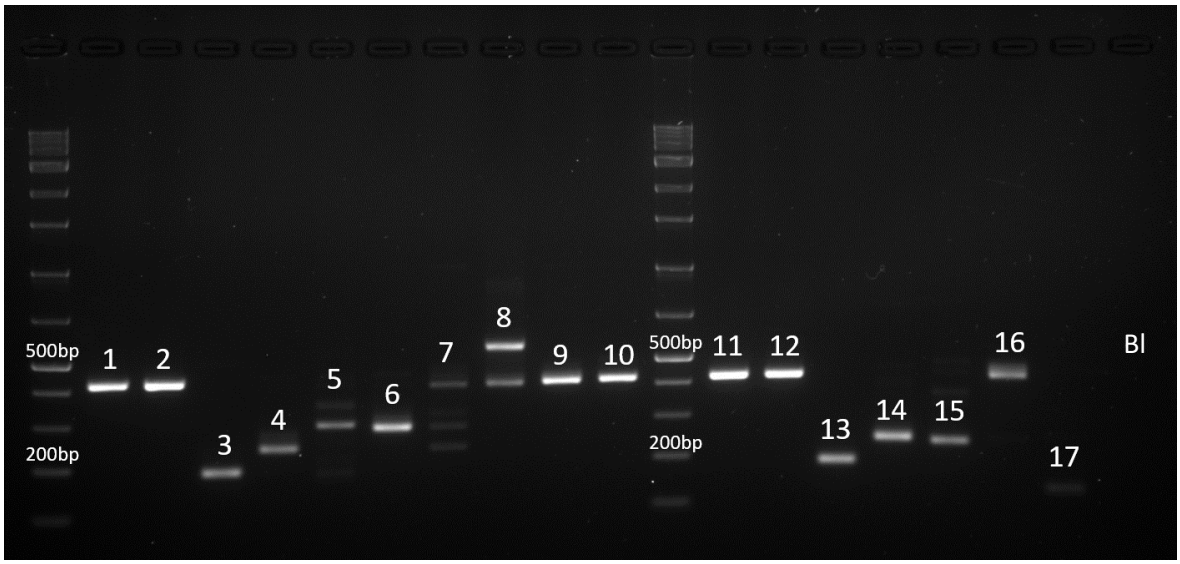


Figure 25 - Genomic mutations caused by CRISPR/Cas9 in comparison to the *wt* DNA of MKN-45 and GP-202 cell lines. (1) MKN-45 wt, (2) MKN-45 mock, (3) MKN-45_O26, (4) MKN-45_O14, (5) MKN-45_O15, (6) MKN-45_O24, (7) MKN-45_O25, (8) MKN-45_O2, (9) MKN-45_O3, (10) MKN-45_O4; (11) GP-202 wt, (12) GP-202 mock, (13) GP-202_O26, (14) GP-202_O14, (15) GP-202_O25; (16) GP-202_O4, (17) GP-202_O16; (Bl) Negative control.

It was observed that most clones present only one specific edition, while few clones present various types of editions. Figure 26 schematically represents the expected large deletion for each sgRNA mix. Nevertheless, it is not possible to predict the type of editing for the sgRNAs individually transfected.

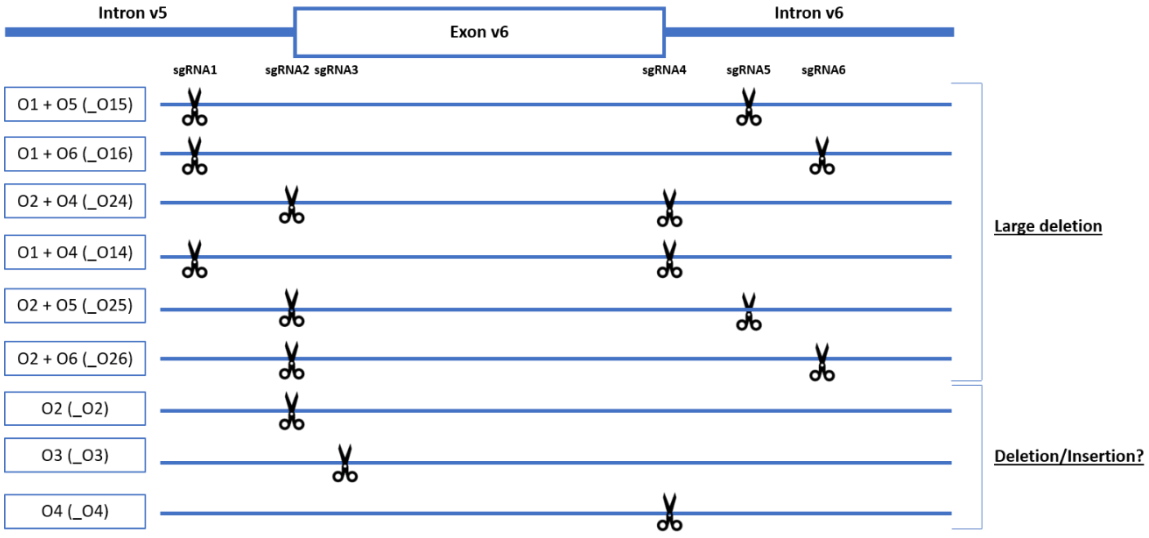


Figure 26 - Schematic representation of the expected editions caused by each sgRNA combination. In the paired sgRNA transfection it is expected a large deletion, however, in the single sgRNA transfected it is not possible to predict the caused mutation.

In order to determine if the mutations of the obtained clones matched the expected editions, the DNA extracted from the clones was sequenced. Figures 27 and 28 show the break points for the clones obtained for the GP-202 and the MKN-45 cell lines, respectively (Figure S1 displays the target sites of each sgRNA in the *wt* CD44 gene).

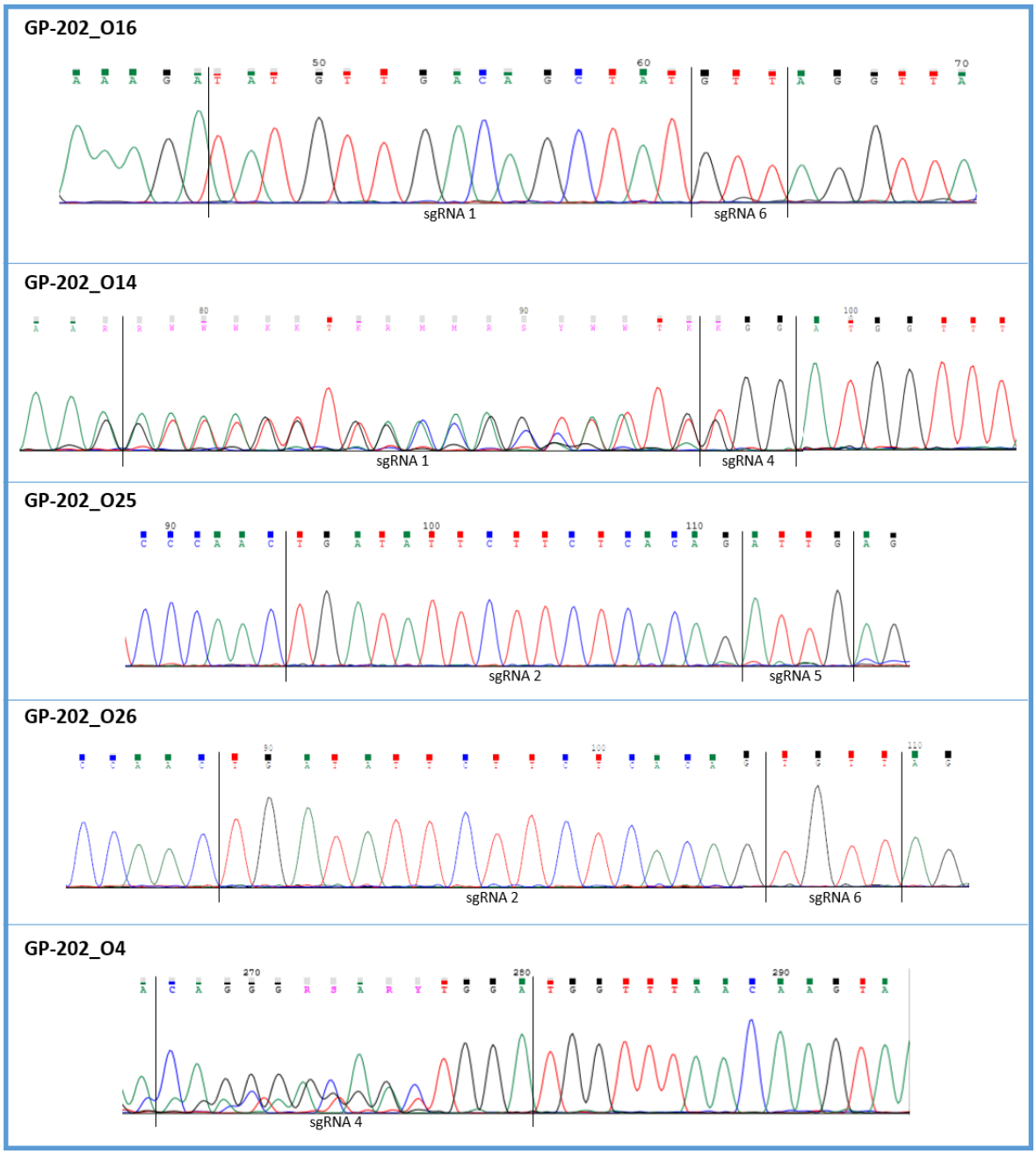


Figure 27 – Representation of the break points for the obtained clones by CRISPR/Cas9 for the GP-202 cell line by Sanger sequencing.

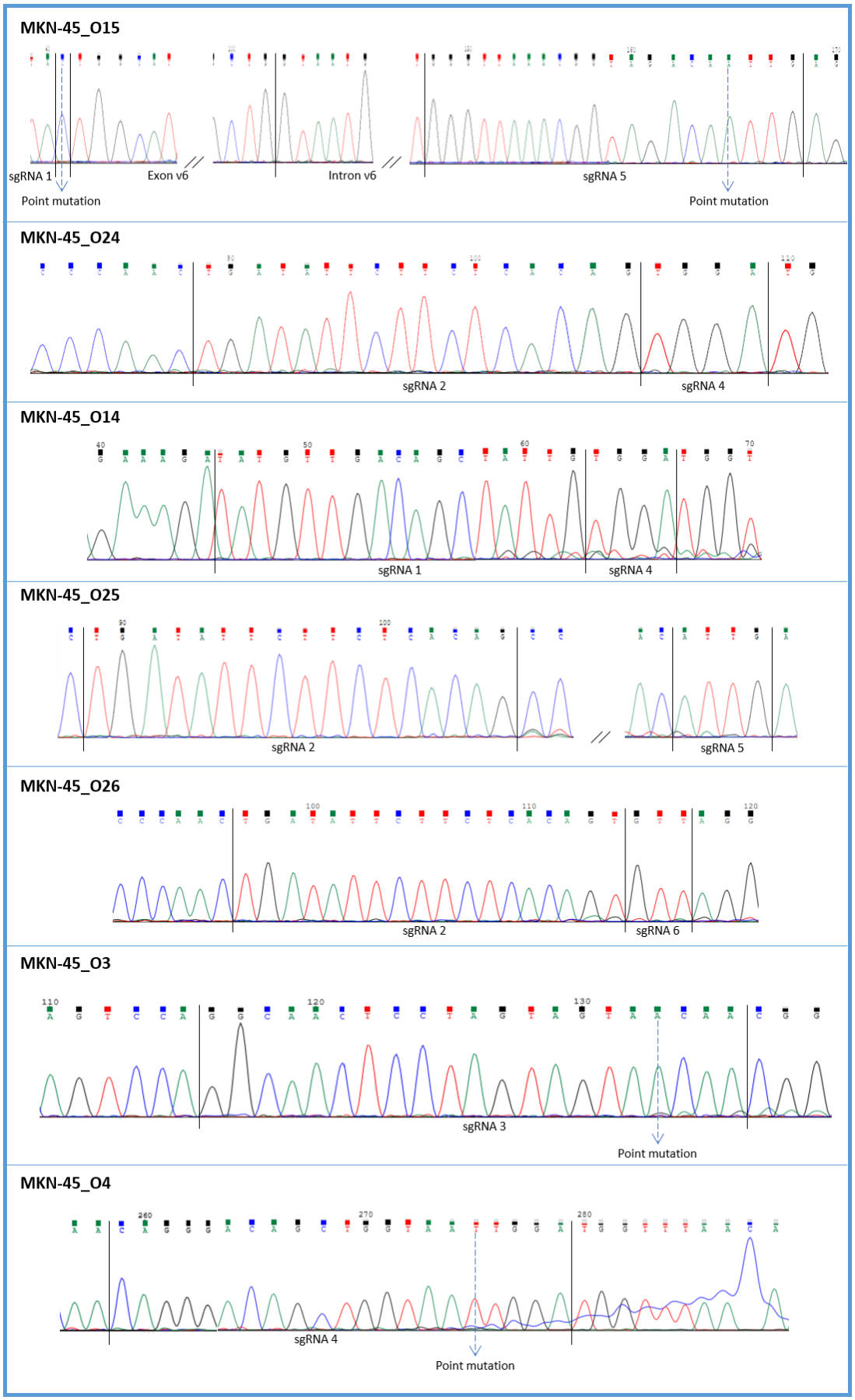


Figure 28 - Representation of the break points for the obtained clones by CRISPR/Cas9 for the MKN-45 cell line by Sanger sequencing.

DNA sequencing results show that the clones GP-202_O16, GP-202_O25, GP-202_O26, MKN-45_O15, MKN-45_O24, MKN-45_O14, MKN-45_O26, MKN-45_O3 and MKN-45_O4 have a homozygous edition, either deletion or insertion. Nevertheless, after analyzing the GP-202_O14 sequence, it was perceived that this clone has two different types of deletions with the difference of only 1bp. The MKN-45_O25 clone displays various editions, however, the sequence presented corresponds to the most representative edition, as shown in Figure 25. Table 6 summarizes the expected editing for each clone according to their target site, and their actual edition after performing CRISPR/Cas9.

Table 6 – Expected editing vs Real editing of CRISPR/Cas9 obtained clones from MKN-45 and GP-202 cell lines. NA means no clones were obtained using that sgRNA combination.

| **MKN-45 cell line** | | | **GP-202 cell line** | | |
|---|---|---|---|---|---|
| **Name** | **Expected editing** | **Real editing** | **Name** | **Expected editing** | **Real editing** |
| **MKN-45_O15** | Deletion of 239bp | Deletion of 131bp<br>Insertion of 2bp | **GP-202_O15** | Deletion of 239bp | NA |
| **MKN-45_O16** | Deletion of 284bp | Insertion/deletion | **GP-202_O16** | Deletion of 284bp | Deletion of 286bp |
| **MKN-45_O24** | Deletion of 134bp | Deletion of 133bp | **GP-202_O24** | Deletion of 134bp | NA |
| **MKN-45_O14** | Deletion of 182bp | Deletion of 182bp | **GP-202_O14** | Deletion of 182bp | Deletion of 181bp |
| **MKN-45_O25** | Deletion of 191bp | Inversion of 178bp | **GP-202_O25** | Deletion of 191bp | Deletion of 190bp |
| **MKN-45_O26** | Deletion of 236bp | Deletion of 235bp | **GP-202_O26** | Deletion of 236bp | Deletion of 236bp |
| **MKN-45_O2** | Insertion/Deletion? | Insertion | **GP-202_O2** | Insertion/Deletion? | No editing |
| **MKN-45_O3** | | Insertion of 1bp | **GP-202_O3** | | NA |
| **MKN-45_O4** | | Insertion of 1bp | **GP-202_O4** | | No editing |

To understand if there was skipping of exon v6 from mRNA and splicing together of exons v5 and v7, cDNA was amplified for the region from exon v5 to exon v7, as shown in Figure 29. Two major PCR products are observed, one corresponding to the PCR product observed in *wt* (lanes 1 and 11) and mock (lanes 2 and 12) cells, and a lower molecular weight PCR product that appears to be newly formed. In order to understand what these two PCR products encompass, cDNA was sequenced (Figure 30). Sequencing results show that the PCR product present in *wt* and mock cells corresponds, as expected, to the transcript where splicing together of exons v5, v6 and v7 (Figure 30A). Sequencing of the lower molecular weight PCR product reveals a transcript where splicing together v5 and v7 occurred in the edited clones (Figure 30B).

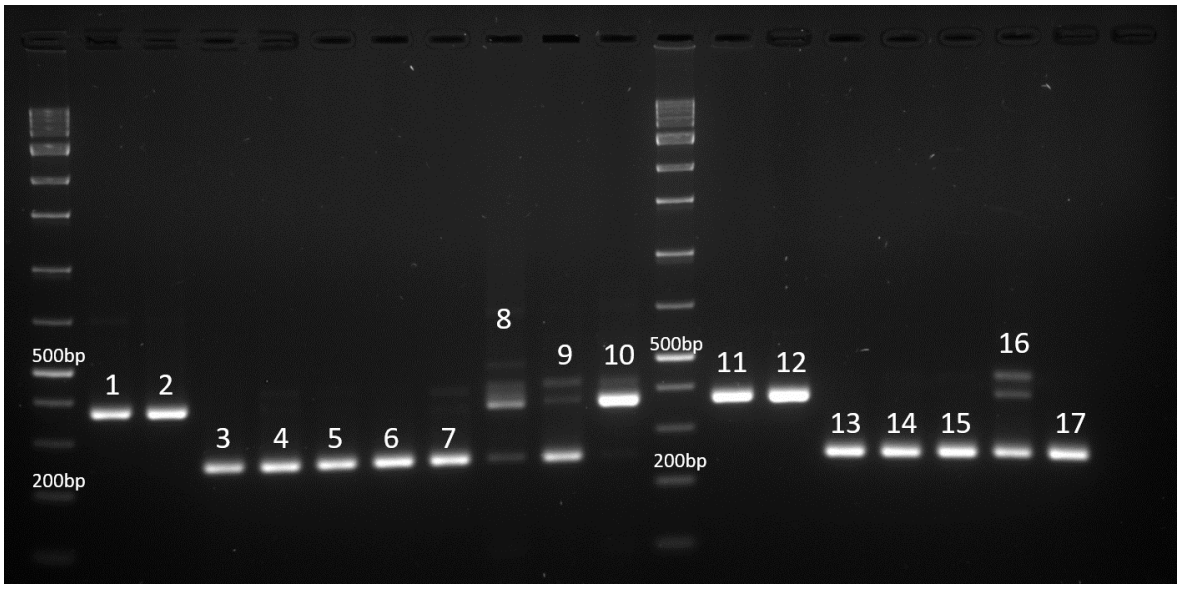


Figure 29 – Formation of new transcripts in the RNA of the clones obtained by CRISPR/Cas9 in comparison to the wt RNA of MKN-45 and GP-202 cell lines. cDNA of: (1) MKN-45 wt, (2) MKN-45 Mock, (3) MKN-45_O26, (4) MKN-45_O14, (5) MKN-45_O15, (6) MKN-45_O24, (7) MKN-45_O25, (8) MKN-45_O2, (9) MKN-45_O3, (10) MKN-45_O4; (11) GP-202 wt, (12) GP-202 mock, (13) GP-202_O26, (14) GP-202_O14, (15) GP-202_O25; (16) GP-202_O4, (17) GP-202_O16; (Bl) Negative control.

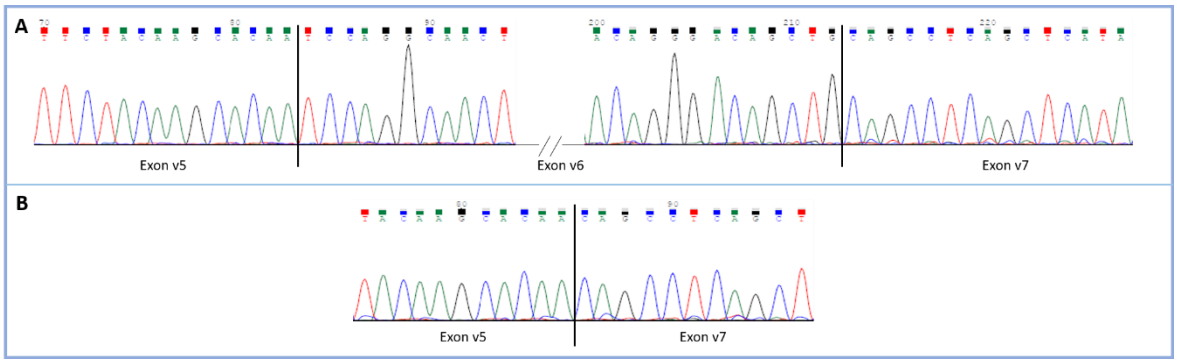


Figure 30 – Transcript analysis following CRISPR/Cas9 edition by Sanger sequencing (representative of both MKN-45 and GP-202 cell lines). Resulting sequences from (A) the splicing of exons v5, v6 and v7; (C) the skipping of exon v6.

Therefore, clones GP-202_O25, GP-202_O26, GP-202_O14, GP-202_O16 (lanes 13 to 15 and 17 of Figure 29, respectively), MKN-45_O26, MKN-45_O14, MKN-45_O15, MKN-45_O24 and MKN-45_O25 (lanes 3 to 7 of Figure 29, respectively) have a large deletion of exon v6 and splicing together of v5 with v7, mimicking exon v6 skipping. When single sgRNAs were transfected, the induced edition did not lead to a total skipping of exon v6 from RNA, as can be observed for clones MKN-45_O2, MKN-45_O3, MKN-45_O4 and GP-202_O4 (lanes 8 to 10 and 16 of Figure 29, respectively).

The edited clones, presenting mimicking of exon v6 skipping, where further characterized for *CD44v6* and total *CD44* RNA expression by qRT-PCR. Only clone GP-

202_O16 was not tested, because preliminary flow cytometry analysis showed that 50% of cells were still positive for CD44v6 (data not shown). The majority of edited clones presented normal *CD44* total gene expression. Clones MKN-45_O25 and MKN-45_O24 still possessed residual *CD44v6* expression (Figure 31).

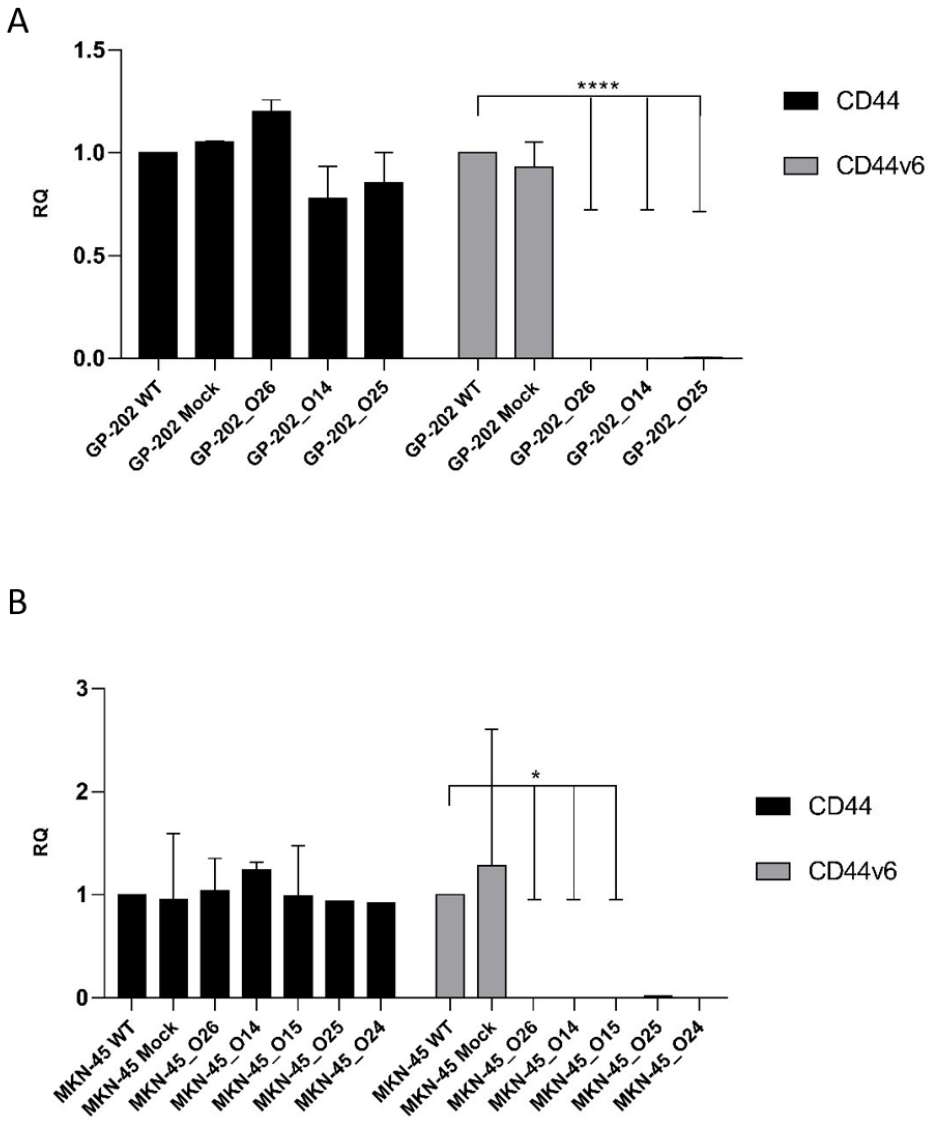


Figure 31 – qRT-PCR analysis of *CD44* total and *CD44v6* gene expression in (A) GP-202 and (B) MKN-45 cell lines. Results represent the average + SD of at least 3 independent experiments (except MKN-45_O25 and MKN-45_O24, which are representative of 1 experiment). All conditions are normalized to *wt* control. p= * <0.05; ** <0.01; *** <0.001; **** <0.0001.

For subsequent characterization and experiments, three clones were selected per cell line. Only clones with homozygous v6 DNA deletion and mimicking the skipping of exon v6 at the RNA level and, consequently, with no *CD44v6* expression were selected.

GP-202_O26, GP-202_O14 and GP-202_O25 and MKN-45_O26, MKN-45_O14 and MKN-45_O15 clones were further characterized at the protein level, by flow cytometry and

immunofluorescence. Flow cytometry results show that all six clones analyzed have significantly decreased CD44v6 expression when comparing to *wt* and mock cells, while no significant differences are observed in terms of total CD44 expression (Figures 32 and 33). This was corroborated by immunofluorescence analysis, where no CD44v6 expression was observed in the edited clones, while total CD44 expression is maintained (Figure 34).

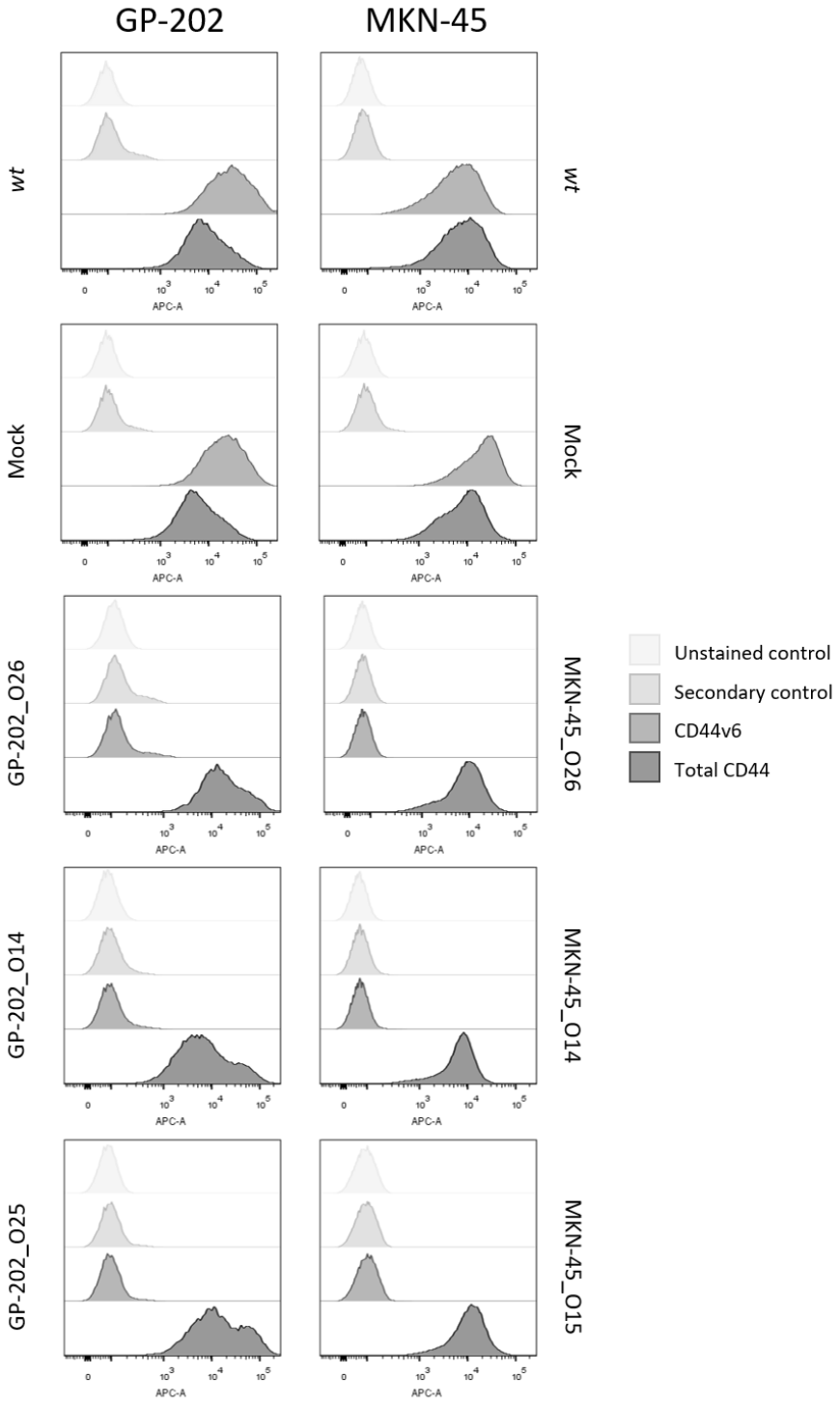


Figure 32 – Total CD44 and CD44v6 expression in the *wt*, mock and edited cells of GP-202 and MKN-45 cell lines analyzed by Flow cytometry. Results shown are representative from those obtained from at least three independent experiments.

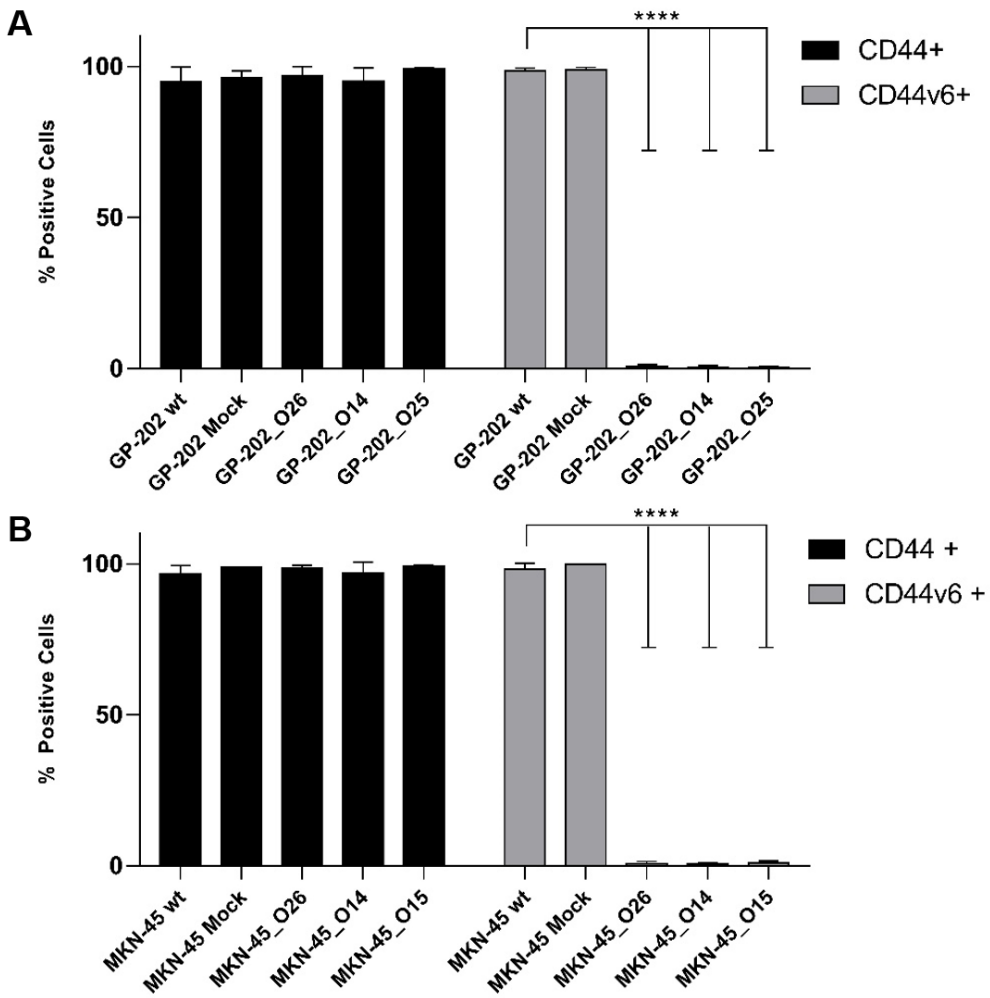


Figure 33 - Flow cytometry against total CD44 and CD44v6 for assessing expression at the post-translational level in the *wt*, mock and edited GP-202 (Panel A) and MKN-45 (Panel B) cells. Results are the average + SD of at least 3 independent experiments. p= * <0.05; ** <0.01; *** <0.001; **** <0.0001.

At the protein level, it was verified that all GP-202 and MKN-45 selected clones present total inhibition of CD44v6, while maintaining the expression of other CD44 isoforms. From these results, it is not possible to assess whether new isoforms lacking only the portion corresponding to exon v6 are formed. However, using the Expasy online tool (https://web.expasy.org) to translate the nucleotide sequence of CD44 without exon v6 into a protein sequence, the whole sequence maintains an open reading frame, suggesting the entire sequence can be translated to form a functional protein.

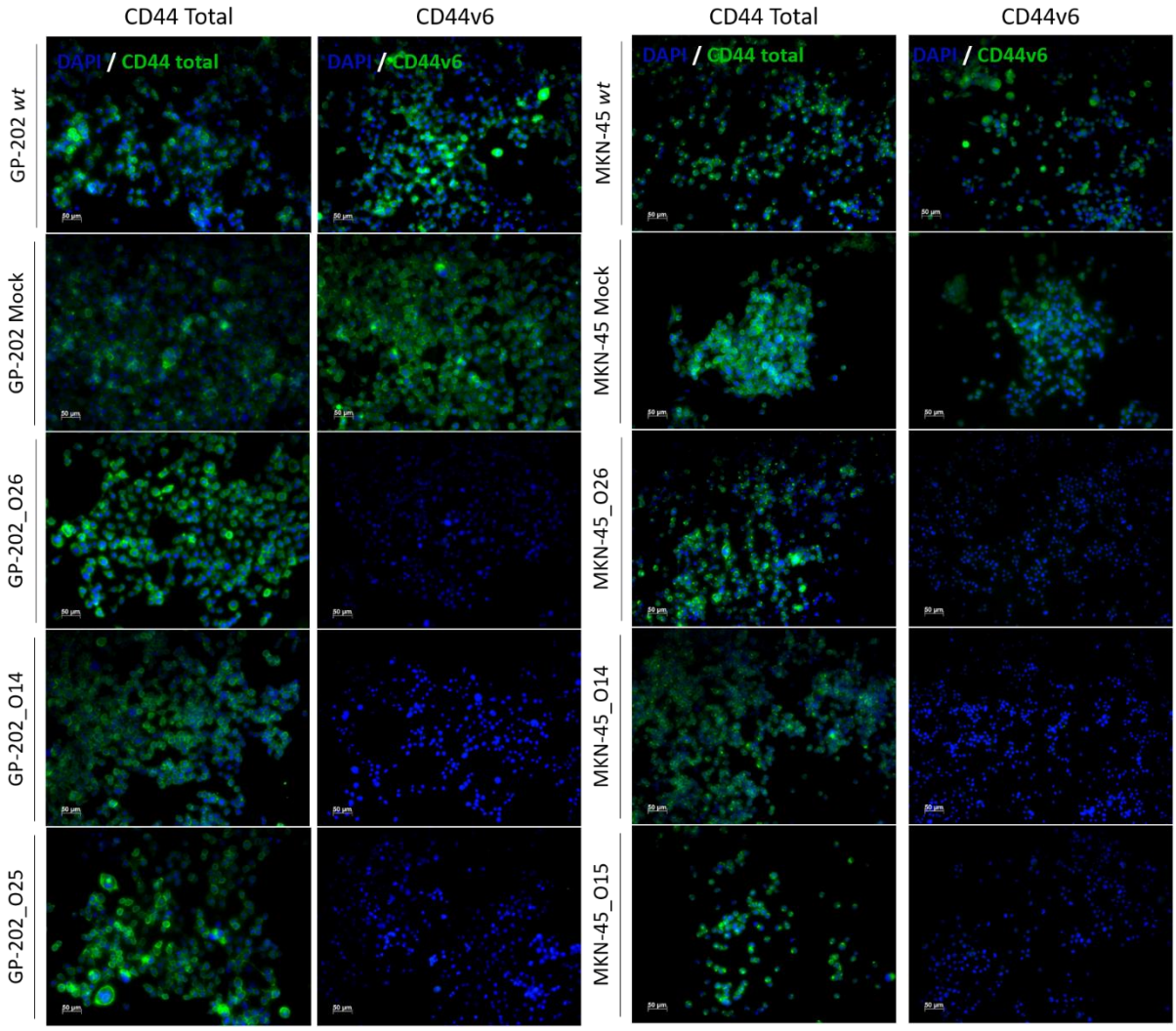


Figure 34 – Immunofluorescence for total CD44 and CD44v6 protein expression in wt, mock and edited clones from GP-202 and MKN-45 cell lines. Nuclei are stained with DAPI (represented in blue) and white scale bars represent a distance of 50 µm.

## 3. Functional consequences of stable induction of skipping of *CD44* exon v6: effects in chemotherapy response

In order to evaluate the response of GC cells to chemotherapy, three of the obtained CRISPR/Cas9 edited clones of MKN-45 (MKN-45_O26, MKN-45_O14 and MKN-45_O15) and GP-202 (GP-202_O26, GP-202_O14 and GP-202_O25) that present a homozygous deletion and perform the skipping of exon v6 were used, as well as the mock cell lines. The mock cell lines were also transfected with a short interference RNA (siRNA) targeting exon v6 in order to compare the response to chemotherapy between the parental cell line without any CD44v6-containing isoforms (induced by siRNA) and the specific exon v6 lacking clones.

For this part of the project, two different chemotherapeutic agents were used, cisplatin and 5-FU, since they are often used for the treatment of GC patients and due to their different mechanisms of actions (8). Cisplatin is a cytotoxic platinum-based drug that binds to DNA, affecting DNA replication and inhibiting cancer cell division. Due to its poor selectivity for tumor cells over normal tissues, cisplatin presents a wide range of severe side effects (90, 91). 5-FU is an antimetabolite and pyrimidine analog that undergoes ribosylation and phosphorylation, resulting in a nucleotide like product, that binds to thymidylate synthase, blocking the synthesis of the pyrimidine thymidine (a nucleoside required for DNA replication) and consequently inhibiting DNA synthesis, leading to cell death in rapidly growing cells (92, 93).

### 3.1. Optimization of treatment conditions

Preliminary tests using *wt* cells and the selected edited clones were performed to estimate the cisplatin and 5-FU concentrations that inhibit cell survival in 50% ($IC_{50}$). For MKN-45 cells the $IC_{50}$ for cisplatin was approximately 2.5 µM and 5 µM for 5-FU. In GP-202 cells the $IC_{50}$ for cisplatin was approximately 20 µM and 5 µM for 5-FU (Figure 35). These concentrations were selected for subsequent experiments.

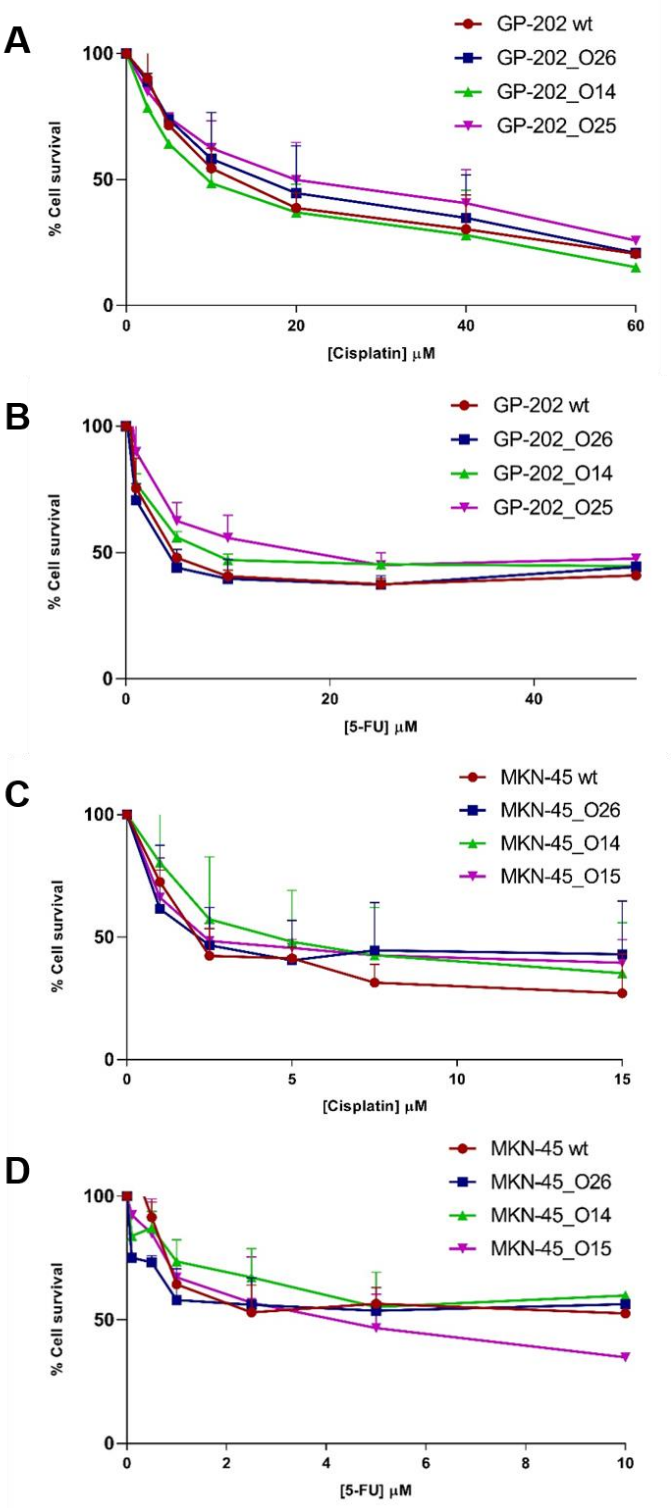


Figure 35 - Cell survival in response to cisplatin (A and C) and 5-FU (B and D) in MKN-45 and GP-202 *wt* cells and edited clones, analyzed by PB assay. Results represent the average + SD of at least two independent experiments.

### 3.2. Skipping of *CD44* exon v6 does not affect cell survival in two GC cell lines in the presence of chemotherapeutic agents

Cell survival in response to cisplatin and 5-FU was assessed in mock cells and selected clones without exon v6, using two short-term assays (lasting for 48 h) and one long term assay (lasting ~10 days). In order to also assess whether drug response varies between cells that specifically lack exon v6 and cells that lack all *CD44v6* transcripts (as induced by treating cells with siRNA) one of the conditions used in these experiments was to incubate mock cells with a siRNA for *CD44v6*. Mock cells and edited clones were also transfected with a siRNA scramble for maintenance of experiment conditions. Inhibition of CD44v6 containing isoforms upon incubation with siRNA for *CD44v6* was confirmed by immunofluorescence (Figure 36). Immunofluorescence results also showed that this CD44v6 expression inhibition lasts throughout the duration of the short-term drug treatments.

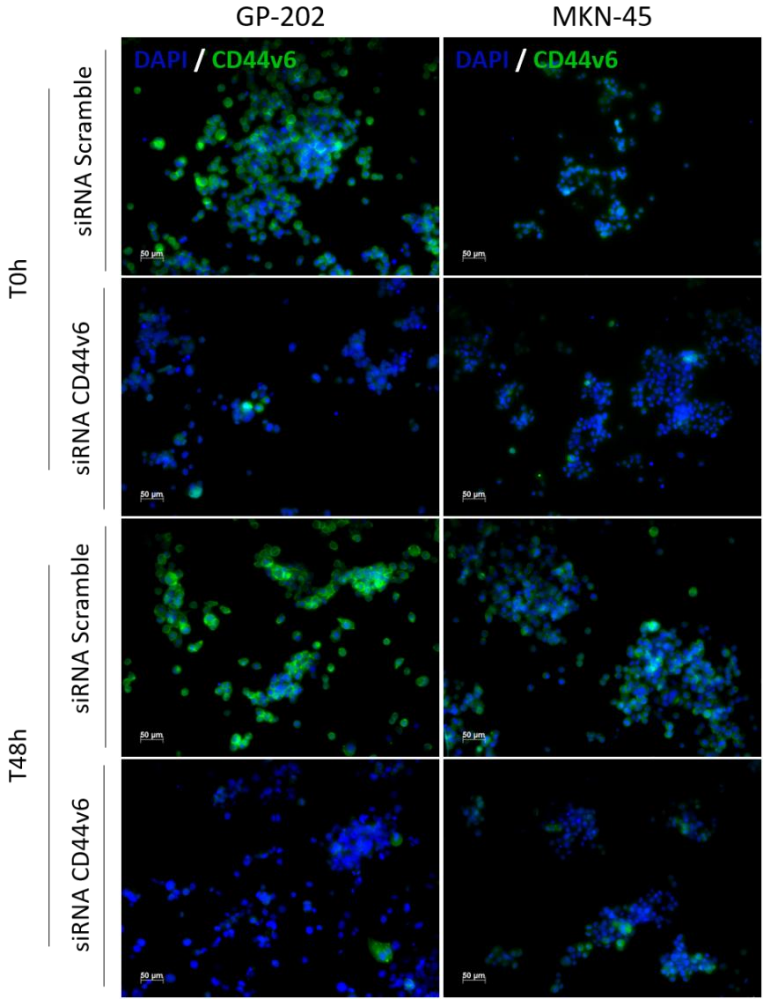


Figure 36 - CD44v6 expression levels through immunofluorescence in GP-202 and MKN-45 mock cells treated with siRNA scramble or siRNA CD44v6, at the time when cells were treated with chemotherapy and 48 h after. Nuclei are stained with DAPI (represented in blue) and white scale bars represent a distance of 50 µm.

To test cell survival in a short-term manner, GP-202 and MKN-45 mock cells transfected with siRNA Scramble, siRNA CD44v6, and edited clones were analyzed by PB assay (Figure 37) and SRB assay (Figure 38).

Results show that cells with CD44v6 expression, cells with no CD44v6 expression and the edited clones respond differently to chemotherapy in short-term treatments. Regarding results obtained in the resazurin-based assay (Figure 37), it is possible to observe that cells expressing no CD44v6 isoforms (*i.e.* those incubated with CD44v6 siRNA) have a significantly increased cell survival when compared to cells expressing CD44v6, in both cell lines in response to cisplatin and 5-FU. GP-202_O26 and GP-202_O14 clones present similar cell survival compared to CD44v6 expressing control cells and have a significant decrease in survival when compared to GP-202 mock cells transfected with CD44v6 siRNA, both for cisplatin and 5-FU. Cisplatin response in the GP-202_O25 clone is similar to that obtained in cells expressing no CD44v6 isoforms, hence having a significant increase in cell survival when compared to control cells (Figure 37A). Upon cisplatin treatment, MKN45_O26 and MKN-45_O14 clones showed a significant increase in cell survival when compared to mock cells, and MKN-45_O15 has a significant decrease in cell survival comparing to cells expressing no CD44v6 isoforms. When treated with 5-FU, all MKN-45 edited clones present a similar response compared to CD44v6 expressing cells and have a significant decrease in cell survival when compared with cells expressing no CD44v6 isoforms (Figure 37B).

When analyzing the results obtained from the short-term treatments in GP-202 cells by the SRB assay (Figure 38A) it was observed that all experimental conditions presented similar levels of cell survival. The only exception being the GP-202_O26 edited clone that had a significant decrease in cell survival when compared to cells expressing no CD44v6 isoforms. MKN-45 cells expressing no CD44v6 isoforms presented a significant increase in cell survival when compared to CD44v6 expressing cells, in response to both cisplatin and 5-FU (Figure 38B). MKN-45_O26 and MKN-45_O14 clones also presented a significant increase in cell survival to cisplatin when compared to CD44v6 expressing cells. In addition, MKN-45_O15 clone had a significant decrease in cisplatin cell survival when compared to cells expressing no CD44v6 isoforms. Regarding 5-FU treatment, MKN-45_O15 has a significant decrease in cell survival when compared to cells without all CD44v6 isoforms and to MKN-45_O26 clone. MKN-45_O14 clone has a significant decrease in cell survival comparing to cells expressing no CD44v6 isoforms.

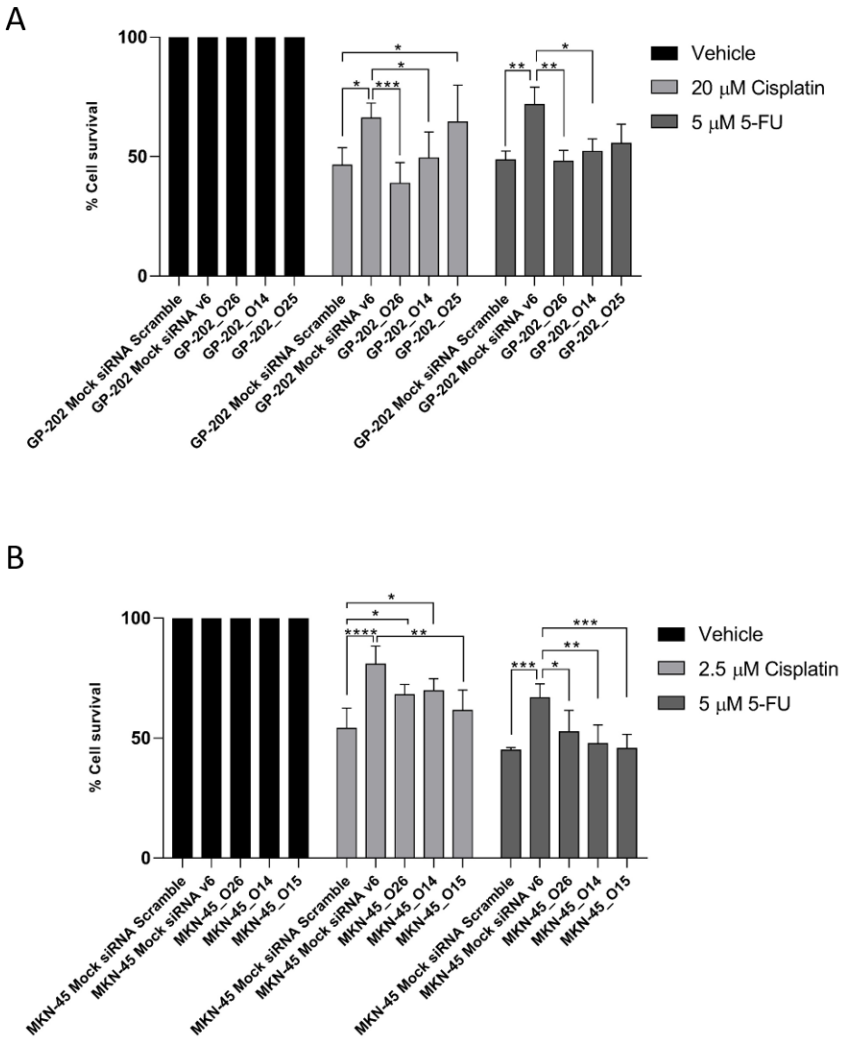


Figure 37 – Response of GP-202 and MKN-45 mock cells, siRNA transfected cells and edited clones when treated with cisplatin, 5-FU and vehicle for 48 h. % Cell Survival of (A) GP-202 and (B) MKN-45 cell lines analyzed by PrestoBlue assay. All conditions are normalized to the Vehicle condition. Results represent the average + SD of at least 3 independent experiments. p= * <0.05; ** <0.01; *** <0.001; **** <0.0001.

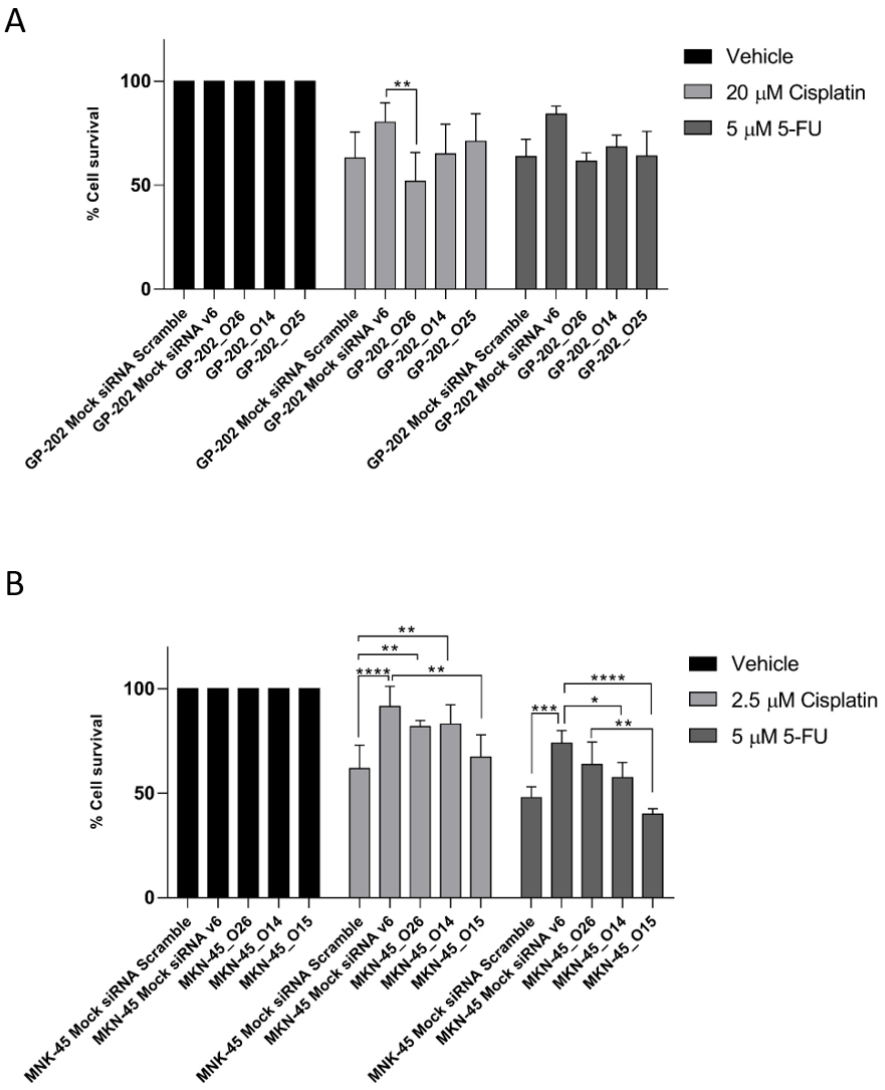


Figure 38 - Response of GP-202 and MKN-45 mock cells, siRNA transfected cells and edited clones when treated with cisplatin, 5-FU and vehicle for 48 h. % Cell Survival of (A) GP-202 and (B) MKN-45 cell lines analyzed by Sulforhodamine B assay. All conditions are normalized to the Vehicle condition. Results represent the average + SD of at least 3 independent experiments. p= * <0.05; ** <0.01; *** <0.001; **** <0.0001.

Throughout the duration of these experiments, it was observed that cells incubated with CD44v6 siRNA presented a much lower cell growth rate (in the vehicle condition) when comparing to cells treated with siRNA scramble and the edited clones, for both GP-202 and MKN-45 cell lines (data not shown).

To evaluate if exon v6 modulates long-term cell survival in response to cisplatin and 5-FU, the clonogenic assay was performed (Figure 39). No differences in cell response were observed in GP-202 cells. Regarding MKN-45 cells, cells expressing no CD44v6 isoforms presented a significant decrease in cell survival when compared to CD44v6 expressing cells. In addition, all three MKN-45 edited clones presented similar survival levels than CD44v6 expressing cells, whereas a significant increase in cell survival was observed when compared to cells expressing no CD44v6 isoforms.

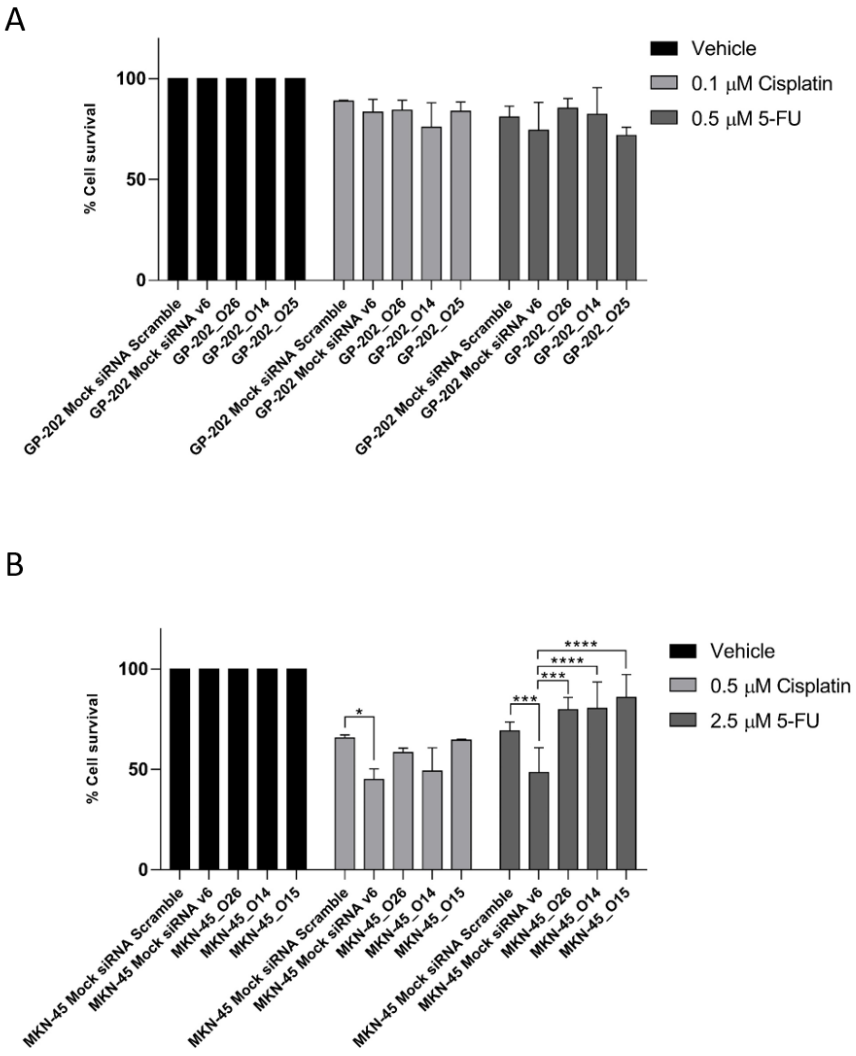


Figure 39 - Cell survival analysis by clonogenic assay in GP-202 and MKN-45 cells. All conditions are normalized to the vehicle. Results represent the average + SD of up to 3 independent replicates. p= * <0.05; ** <0.01; *** <0.001; **** <0.0001.

When performing the clonogenic assay, it was observed that the number of colonies obtained in the vehicle control of MKN-45 and GP-202 cells that had been previously incubated with siRNA for *CD44v6* (*i.e.* that possessed no *CD44v6* transcripts) was consistently lower than those obtained for the other isogenic cells, even though the same number of cells was plated for all conditions (Figure 40). Two edited clones, GP-202_O25 and MKN-45 O26 also present a lower growth rate in vehicle conditions, when comparing to the respective control.

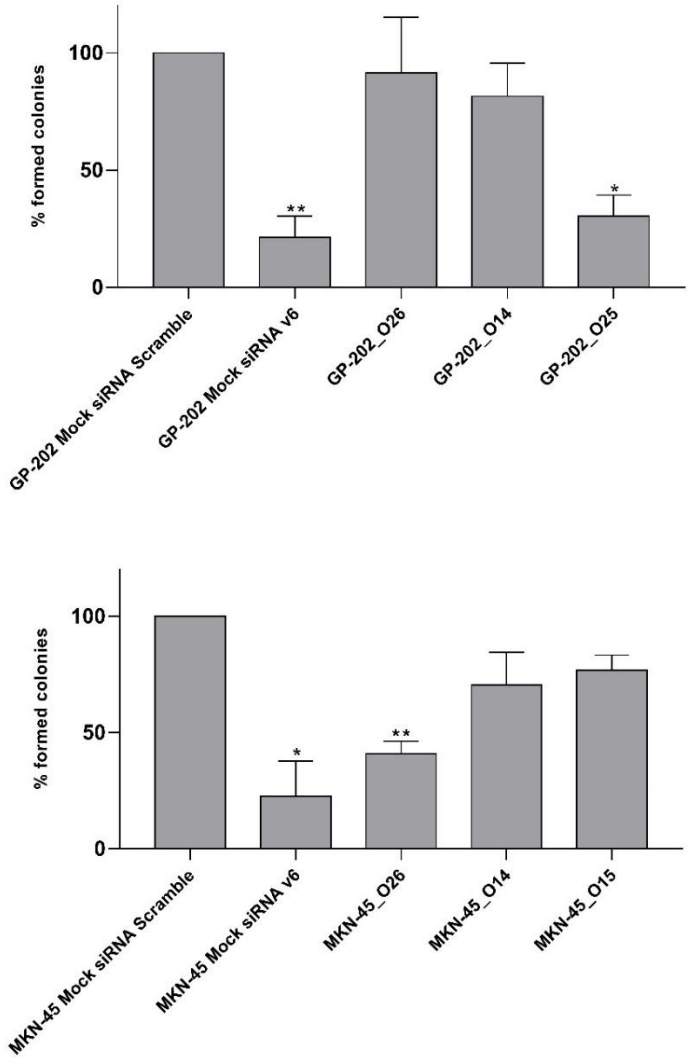


Figure 40 - Colony formation analysis in the vehicle for all conditions. All conditions are normalized to mock siRNA scramble. Results represent the average + SD of 3 independent replicates. p= * <0.05; ** <0.01; *** <0.001; **** <0.0001.

# Chapter V | Discussion

## 1. Transient and permanent study models of exon v6 skipping were successfully generated

Our first goal was to obtain GC cell lines lacking only exon v6 from the *CD44* gene. Through the results obtained in this project, we can confirm that we successfully obtained two different types of exon skipping models, one transient with the exon v6 skipping from *CD44v6*-containing transcripts, using PMOs, and one permanent, mimicking the skipping of exon v6 by deleting the whole exon from DNA, using CRISPR/Cas9.

Regarding PMOs, these were more efficient to cause exon skipping in GP-202 cells rather than MKN-45 cell line. However, in both cell lines, a higher inhibition at the RNA level was expected, since the preliminary results performed with a PMO fluorescent control showed that approximately 90% of cells internalized PMOs with Endo-Porter, which transports PMOs into the cytosol. Still, each specific PMO may differ in efficiency, therefore, even though PMOs might be internalized, they may not be effective in causing the skipping of exon v6. It is recommended to use Endo-Porter as a transfection reagent, however, PMOs work at the nuclei level to target the pre-mRNA and Endo-Porter reagent releases PMOs in the cytosol. Therefore, it is possible that PMOs are not entering the nucleus as efficiently as expected. One way to allow PMOs to properly enter the nucleus would be to use a transfection reagent that would release PMOs into the nuclei, such as Nucleofector® reagent (94). Nevertheless, the results obtained at the protein level demonstrate that GP-202 cells practically have no CD44v6 protein expression. This was not expected since at the mRNA level, total exon v6 inhibition was not achieved, however, immunofluorescence is a technique that does not analyze the entire population of cells within a condition. Therefore, studies that allow the evaluation of the total cell population would be required to further validate these results, such as flow cytometry, western blot or ELISA for CD44v6 and total CD44 protein expression.

Regarding CRISPR/Cas9, we successfully obtained homozygous clones that mimic the skipping of exon v6 in a permanent manner. Our results suggest that causing a large deletion in exon v6 is more effective to cause the removal of exon v6 at the mRNA level, mimicking exon v6 skipping, rather that causing only one DSB to disrupt the splice site. When designing the sgRNA, we hypothesized that disrupting the splice sites would lead to the non-recognition of the splice site by the splicing machinery, resulting in the skipping of exon v6 of *CD44v6*-containing transcripts. However, none of the clones transfected with only one sgRNA resulted in total exon v6 skipping at the mRNA. The transcripts formed in these clones mimic the pattern observed in PMOs, meaning that the sequence near exon v6 or in exon v6 itself may possess *cis*-regulatory elements that enhance the splicing of

new transcripts, resulting not only in the skipping of exon v6 but in the formation of longer transcripts (95, 96).

We have proven that a new transcript is formed, splicing together exon v5 and exon v7 and at the protein level, CD44v6 expression is undetected. Using an online tool to predict protein formation through the analysis of cDNA sequence, it is demonstrated that removing exon v6, a modified protein is produced without the amino acids corresponding to the v6 portion but maintaining the open reading frame, leading to a shorter but functional protein. However, further studies are required to show how these modified proteins work. CD44v6-containing isoforms possess specific binding sites to VEGFR-2 and to HGF, the last one being a binding molecule to cMET. CD44v6 leads to these oncogenes activation (18), and it has been associated with cancer metastization, angiogenesis and aggressiveness in various types of cancer (97-100), therefore, it would be interesting to understand if the loss of the v6 domain on CD44v6 protein upon exon v6 skipping will affect the ligation to these two oncogenes. If so, it would be interesting to evaluate how affected would be the pathways associated to these two oncogenes after inactivation and how this would impact the development and progression of GC (101).

Both technologies, PMOs and CRISPR/Cas9, arose as good techniques for modulating splicing, since both can cause exon skipping of the specific target. Nevertheless, mutations caused by CRISPR/Cas9 are stable and, in a clinical context, it would mean that no further maintenance therapy would be required, at least in non-proliferative or low-proliferating tissues (102). However, CRISPR/Cas9 in biomedical investigations is still being tested only in animal models, and when tested in human embryos, off target effects were found, meaning that CRISPR/Cas9 technology is yet associated to negative secondary effects and needs to be improved (103) before introducing it into clinical practice (104). Besides, a study conducted by Charlesworth and colleagues suggests that more than half of the human population may present humoral and cell-mediated adaptive immune response to the Cas9 proteins used for CRISPR/Cas9, since the bacteria from where they are extracted (*S. pyogenes*) frequently cause infectious diseases in humans (105). This means that the human organism would attack Cas9 proteins, destroying them. On the other hand, PMOs are already used in clinical practice, being integrated in the treatment of DMD patients and successfully improving their quality of life (72, 74). Although the PMO model would be more adequate to be applied in clinical practice, if required, for the present study and for the sake of simplicity, we aimed to understand how exon v6 modulated chemotherapy response by studying the stable CRISPR/Cas9 created models.

With this project, we understood that exon skipping might be a good model to study gene alterations that cause specific diseases, since they provide information regarding the functional role of a specific exon. It would be a good form of gene therapy to remove a mutation in a specific exon, without knocking out the whole gene, hence maintaining protein production. For example, DMD disorder has a gene therapy relying on the concept of exon skipping, resulting in the production of a smaller but functional DMD protein (72). However, the question remains if exon skipping gene manipulation would be a god approach for cancer therapy. Nevertheless, no studies regarding this subject were found in the literature. Hence, our study proves to be pioneer in the research area of exon skipping gene manipulation in cancer, since we are the first to develop exon skipping models for the study of cancer cells therapy response with the manipulation of *CD44* gene.

## 2. *CD44* exon v6 does not modulate chemotherapy response in GC cells

In order to understand whether exon v6 was responsible, itself, by the modulation of the response of GC cell lines to chemotherapy, cell survival was analyzed in response to short and long-term treatments with chemotherapeutic agents. Our results suggest that the absence of all CD44v6 isoforms gives GC cells a certain advantage in the acute phase after treatment (48 h after), when comparing to CD44v6 endogenous expression. Nevertheless, this advantage is lost after several days of treatment, since cells endogenously expressing CD44v6 are capable of surviving better after 10 days of treatment. This is consistently observed in most conditions, particularly in the MKN-45 cell line. For GP-202 cells, this is not observed in SRB and clonogenic assays. However, we verified that the assays used to assess cell survival presented some degree of variability within independent experiments, likely leading to loss of statistical power when differences are small between experimental conditions. As for the clonogenic assay, we believe that the concentrations used for each drug in GP-202 cells were too low and due to that, cells did not show any differences in cell survival after 10 days of treatment with both cisplatin and 5-FU. Therefore, additional studies on cell survival, to increase the statistical power, and with higher concentrations of both drugs are required.

Regarding the response to chemotherapy of the MKN-45 and GP-202 selected clones, some of those presented a significant differential cell survival comparing to either CD44v6 expressing or non-expressing cells. Regarding the GP-202 cells, the results suggest that GP-202_O26 and GP-202_O14 clones have a pattern of response to chemotherapy similar to CD44v6 expressing cells, while GP-202_O25 clone presents a response to chemotherapy similar to cells that express no CD44v6 isoforms, in the acute phase after

treatment. As for the MKN-45 cells, MKN-45_O26 and MKN-45_O14 appear to respond to chemotherapy in a similar manner to cells that do not express CD44v6 isoforms, while MKN-45_O15 responds to cisplatin in a similar way than CD44v6 expressing cells. However, all MKN-45 clones' respond to 5-FU in a similar manner to CD44v6 expressing cells, in the acute phase after treatment. Since all GP-202 and MKN-45 clones presented the mimicking of exon v6 skipping in the mRNA, this differential response to chemotherapy was not expected. However, all clones present different deletions, meaning that some important regulatory elements might be lost within the different editing's by CRISPR/Cas9, such as *cis* regulatory elements (106), epigenetic modifications (107) and changes in chromatin structure (108), which will ultimately result in differences in gene expression. Therefore, studies to identify important genetic and epigenetic regulatory elements would be required, such as reporter gene assays, chromatin immunoprecipitation microarrays (108, 109) and bisulfite modification analysis (108).

A previous study conducted by our group unveiled that GC cell lines expressing CD44v6-containing isoforms were more resistant to cisplatin when compared to cells without any CD44v6 isoforms (110), which differs from the results obtained in the present study. However, in that study, cell response to treatment was evaluated through apoptosis assays, while in the current study cell response to treatment was evaluated through cell survival assays. The end points analyzed for each technique are different, since apoptosis assay measures the quantity of cell in a pre-apoptotic or apoptotic state, and cell survival assays measure the number of live cells. Nevertheless, increasing apoptosis resistance associated with CD44v6 has also been found in a study conducted by Jung and colleagues (2011), where they discovered that CD44v6 is capable of coordinating tumor matrix leading to increased apoptosis resistance to cisplatin through the activation of anti-apoptotic molecules, in pancreatic adenocarcinoma cell lines (111). Other studies found association of CD44v6 with increased apoptosis resistance to chemotherapy (56, 112). Our results suggest that cancer cells may be responding to the treatment without dying, entering in a pre-apoptotic state where cells are only prone to dye. Likewise, further short-term treatment studies evaluating apoptosis and cell cycle in cells subjected to chemotherapy would be interesting results to add to the already obtained results.

Regarding the long-term treatments, results suggest that all MKN-45 selected clones respond in a similar manner to 5-FU, and equally to cells expressing CD44v6, meaning that in long-term treatment, which better mimics what may happen during GC patients chemotherapy, cells lacking only the exon v6 do not decrease cell survival, whereas cells with no CD44v6 expression do. These results support the hypothesis formed before for the short-term treatments, meaning that cells may need more time to either complete the

apoptotic process or to recover from a senescent state and continue to proliferate under drug stress conditions. In fact, some researchers have discovered that apoptosis induced by a certain compound can be reversed after removing such compound, leading to cell recovering from a senescent state (113-115). Therefore, we hypothesize that cells expressing no CD44v6 isoforms at the acute phase of treatment might be in a senescent state due to chemotherapy, causing cell cycle arrest, however without dying. Nevertheless, with the removal of chemotherapeutic agents and increased treatment duration, cells with no CD44v6-containing isoforms are not capable of recovering from the senescent state, and eventually die. On the other hand, cells expressing CD44v6 and some of the edited clones, appear to respond well to chemotherapy in the acute phase of treatment, nevertheless, the cells that are capable of surviving eventually regain their normal proliferative state, surviving better after several days of treatment. Still, further studies are required to validate this hypothesis, such as functional assays to evaluate the apoptosis pathway, like caspases activation or organoid fragmentation. However, if this hypothesis is correct, it would mean that exon v6 from the *CD44* gene is not, itself, responsible for the modulation of chemotherapy response associated with CD44v6-containing isoforms observed in GC cells, however, removing all CD44v6-containing isoforms modulates therapy response. This means that the responsible could be another variant exon of the gene, multiple exons or even one specific isoform. To prove this, functional studies like the present one would be required, applying exon skipping for each variant exon or to a group of exons, to understand whether the skipping of other variant portions of the *CD44* gene would cause differential responses to chemotherapy. Ultimately, transfecting each specific CD44v6-containing isoform separately to a non-expressing CD44v6 GC cell line, would allow to understand if the one specific isoform was responsible for the modulation of chemotherapy response in GC cells. Thus, many and more complex studies must be performed to validate the results obtained throughout the present study and to identify which variant exons are responsible for modulating therapy response in CD44v6 containing isoforms.

Moreover, our results also suggest that cells expressing no CD44v6-containing isoforms loose growing capacity, since in both cell lines, mock cells treated with CD44v6 siRNA have a significant decreased growth rate, when comparing to mock cells treated with siRNA scramble, meaning that CD44v6-containing isoforms may be important for tumor cell growth and self-renewal, and that destroying these isoforms may be a good approach to decrease tumor cell division.

Results obtain throughout this project suggest that inhibiting all CD44v6-containing isoforms from GC cells modulates response to chemotherapy, right after treatment and

several days after that. However, exon v6 itself, does not modulates chemotherapy response in GC cells.

# Chapter VI | Conclusion and Future Perspectives

We can state that, using PMOs, we successfully generated transient isogenic cell lines that perform skipping of exon v6 from the *CD44v6*-containing transcripts. Nevertheless, we did not obtain complete skipping of exon v6 on GP-202 cell line and PMOs barely cause exon v6 skipping on MKN-45 cells. Therefore, in the future, more studies are required to improve PMOs transfection efficiency and test new transfection approaches, like Nucleofactor® or transfect PMOs together.

Similarly, using CRISPR/Cas9, we successfully generated permanent isogenic cell lines that either perform exon skipping through disruption of the splice site or exon deletion that mimics exon skipping through deleting the whole exon v6. However, disrupting the splice site was not effective in all clones, meaning that, in the future, it would be important to deepen the knowledge about the splicing machinery in exon v6 and in the adjacent intronic regions. On the other hand, causing a large deletion on the genome was more successful in removing exon v6. This type of model appears to be more reliable as an exon skipping model, however, in the future, it would important to prove the specific removal of the v6 domain at the protein level. Since CD44v6 isoforms contain specific binding sites to VEGFR and HGF, that activates cMET, it would be interesting to observe what occurs with these two bindings with the removal of the v6 domain from the CD44v6-containing isoforms and if these two oncogenes (*VEGFR-2* and *cMET*) remain activated or not.

The results obtained in this project with the treatment of GC cells with cisplatin and 5-FU show that GP-202 and MKN-45 cells expressing no CD44v6-containing isoforms generally survive better than CD44v6 expressing cells within 48 h of treatment, however, with longer treatments the opposite occurs, at least for MKN-45 cells. Our results suggest that certain edited clones present a response pattern similar to CD44v6 expressing cells while other clones present a therapy response similar to cells expressing no CD44v6-containing isoforms, in the short-term treatments, suggesting that the different editions of the clones might interfere in cell survival due to loss of certain genetic or epigenetic regulatory elements. Nevertheless, MKN-45 clones respond similarly to 5-FU in long-term treatments, suggesting that in the long-term, cells lacking only exon v6 regain their proliferative capacity while cells without CD44v6 isoforms do not.

In conclusion, our results suggest that inhibiting CD44v6-containing isoforms modulated GC cell lines response to chemotherapy in both short and long-term treatments, however, the exon v6 from CD44v6-containing isoforms is not, by itself, capable of modulating cell response to chemotherapy in GC cells. Nevertheless, further studies are required to validate the obtained results.

# Chapter VII | Bibliography

1. Bray, F., et al., 2018, Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. **68**(6): p. 394-424.
2. Balakrishnan, M., et al., 2017, Changing Trends in Stomach Cancer Throughout the World. Curr Gastroenterol Rep. **19**(8): p. 36.
3. Chang, A.B., A. Bush, and G.K. Bronchiectasis, 2018, GLOBOCAN 2018: counting the toll of cancer (vol 392, pg 985, 2018). Lancet. **392**(10154): p. 1196-1196.
4. Duraes, C., et al., 2014, Biomarkers for gastric cancer: prognostic, predictive or targets of therapy? Virchows Arch. **464**(3): p. 367-78.
5. Abbas, M., et al., 2018, Current and future biomarkers in gastric cancer. Biomed Pharmacother. **103**: p. 1688-1700.
6. Van Cutsem, E., et al., 2016, Gastric cancer. Lancet. **388**(10060): p. 2654-2664.
7. Edge, S.B. and C.C. Compton, 2010, The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. Ann Surg Oncol. **17**(6): p. 1471-4.
8. Smyth, E.C., et al., 2016, Gastric cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. **27**(suppl 5): p. v38-v49.
9. Van Cutsem, E., et al., 2015, HER2 screening data from ToGA: targeting HER2 in gastric and gastroesophageal junction cancer. Gastric Cancer. **18**(3): p. 476-84.
10. Wilke, H., et al., 2014, Ramucirumab plus paclitaxel versus placebo plus paclitaxel in patients with previously treated advanced gastric or gastro-oesophageal junction adenocarcinoma (RAINBOW): a double-blind, randomised phase 3 trial. Lancet Oncol. **15**(11): p. 1224-35.
11. Zhang, D. and D. Fan, 2010, New insights into the mechanisms of gastric cancer multidrug resistance and future perspectives. Future Oncol. **6**(4): p. 527-37.
12. Gottesman, M.M., 2002, Mechanisms of cancer drug resistance. Annu Rev Med. **53**: p. 615-27.
13. Lippert, T.H., H.J. Ruoff, and M. Volm, 2011, Current status of methods to assess cancer drug resistance. Int J Med Sci. **8**(3): p. 245-53.
14. Goossens, N., et al., 2015, Cancer biomarker discovery and validation. Transl Cancer Res. **4**(3): p. 256-269.
15. Pietrantonio, F., et al., 2013, A review on biomarkers for prediction of treatment outcome in gastric cancer. Anticancer Res. **33**(4): p. 1257-66.
16. Adachi, Y., et al., 2000, Pathology and prognosis of gastric carcinoma: well versus poorly differentiated type. Cancer. **89**(7): p. 1418-24.
17. Goodison, S., V. Urquidi, and D. Tarin, 1999, CD44 cell adhesion molecules. Mol Pathol. **52**(4): p. 189-96.
18. Orian-Rousseau, V., 2010, CD44, a therapeutic target for metastasising tumours. Eur J Cancer. **46**(7): p. 1271-7.

19. Zoller, M., 2015, CD44, Hyaluronan, the Hematopoietic Stem Cell, and Leukemia-Initiating Cells. Front Immunol. **6**: p. 235.

20. Senbanjo, L.T. and M.A. Chellaiah, 2017, CD44: A Multifunctional Cell Surface Adhesion Receptor Is a Regulator of Progression and Metastasis of Cancer Cells. Front Cell Dev Biol. **5**: p. 18.

21. Gallatin, W.M., I.L. Weissman, and E.C. Butcher, 1983, A cell-surface molecule involved in organ-specific homing of lymphocytes. Nature. **304**(5921): p. 30-4.

22. Black, D.L., 2003, Mechanisms of alternative pre-messenger RNA splicing. Annu Rev Biochem. **72**: p. 291-336.

23. Grabowski, P.J. and D.L. Black, 2001, Alternative RNA splicing in the nervous system. Prog Neurobiol. **65**(3): p. 289-308.

24. Han, S., et al., 2019, Prognostic Value of CD44 and Its Isoforms in Advanced Cancer: A Systematic Meta-Analysis With Trial Sequential Analysis. Front Oncol. **9**: p. 39.

25. Zoller, M., 2011, CD44: can a cancer-initiating cell profit from an abundantly expressed molecule? Nat Rev Cancer. **11**(4): p. 254-67.

26. Kopp, R., et al., 2009, Frequent expression of the high molecular, 673-bp CD44v3,v8-10 variant in colorectal adenomas and carcinomas. Int J Mol Med. **24**(5): p. 677-83.

27. Wang, L., et al., 2014, Expression of CD44v3, erythropoietin and VEGF-C in gastric adenocarcinomas: correlations with clinicopathological features. Tumori. **100**(3): p. 321-7.

28. Liu, Y.Q., et al., 2014, CD44v3 and VEGF-C expression and its relationship with lymph node metastasis in squamous cell carcinomas of the uterine cervix. Asian Pac J Cancer Prev. **15**(12): p. 5049-53.

29. Sagawa, K., et al., 2016, Expression of CD44 variant isoforms, CD44v3 and CD44v6, are associated with prognosis in nasopharyngeal carcinoma. J Laryngol Otol. **130**(9): p. 843-9.

30. Aso, T., et al., 2015, Induction of CD44 variant 9-expressing cancer stem cells might attenuate the efficacy of chemoradioselection and Worsens the prognosis of patients with advanced head and neck cancer. PLoS One. **10**(3): p. e0116596.

31. Go, S.I., et al., 2016, CD44 Variant 9 Serves as a Poor Prognostic Marker in Early Gastric Cancer, But Not in Advanced Gastric Cancer. Cancer Res Treat. **48**(1): p. 142-52.

32. Taniguchi, D., et al., 2018, CD44v9 is associated with epithelial-mesenchymal transition and poor outcomes in esophageal squamous cell carcinoma. Cancer Med. **7**(12): p. 6258-6268.

33. Tokunaga, E., et al., 2019, CD44v9 as a poor prognostic factor of triple-negative breast cancer treated with neoadjuvant chemotherapy. Breast Cancer. **26**(1): p. 47-57.

34. Hasenauer, S., et al., 2013, Internalization of Met requires the co-receptor CD44v6 and its link to ERM proteins. PLoS One. **8**(4): p. e62357.

35. Chen, C., et al., 2018, The biology and role of CD44 in cancer progression: therapeutic implications. Journal of Hematology & Oncology. **11**.

36. Soukka, T., et al., 1997, Regulation of CD44v6-containing isoforms during proliferation of normal and malignant epithelial cells. Cancer Res. **57**(11): p. 2281-9.

37. Kooy, A.J., et al., 1999, Expression of E-cadherin, alpha- & beta-catenin, and CD44V6 and the subcellular localization of E-cadherin and CD44V6 in normal epidermis and basal cell carcinoma. Hum Pathol. **30**(11): p. 1328-35.

38. Afify, A.M., et al., 2005, Expression of hyaluronic acid and its receptors, CD44s and CD44v6, in normal, hyperplastic, and neoplastic endometrium. Ann Diagn Pathol. **9**(6): p. 312-8.

39. Mack, B. and O. Gires, 2008, CD44s and CD44v6 expression in head and neck epithelia. PLoS One. **3**(10): p. e3360.

40. Wang, Z., et al., 2018, CD44/CD44v6 a Reliable Companion in Cancer-Initiating Cell Maintenance and Tumor Progression. Front Cell Dev Biol. **6**: p. 97.

41. Gunthert, U., et al., 1991, A new variant of glycoprotein CD44 confers metastatic potential to rat carcinoma cells. Cell. **65**(1): p. 13-24.

42. Seiter, S., et al., 1993, Prevention of tumor metastasis formation by anti-variant CD44. J Exp Med. **177**(2): p. 443-55.

43. Xie, J.W., et al., 2015, Evaluation of the prognostic value and functional roles of CD44v6 in gastric cancer. J Cancer Res Clin Oncol. **141**(10): p. 1809-17.

44. Xu, Y.Y., et al., 2017, Regulation of CD44v6 expression in gastric carcinoma by the IL-6/STAT3 signaling pathway and its clinical significance. Oncotarget. **8**(28): p. 45848-45861.

45. Hu, B., et al., 2015, Meta-Analysis of Prognostic and Clinical Significance of CD44v6 in Esophageal Cancer. Medicine (Baltimore). **94**(31): p. e1238.

46. Wang, J.L., et al., 2017, CD44v6 overexpression related to metastasis and poor prognosis of colorectal cancer: A meta-analysis. Oncotarget. **8**(8): p. 12866-12876.

47. Jiang, H., W. Zhao, and W. Shao, 2014, Prognostic value of CD44 and CD44v6 expression in patients with non-small cell lung cancer: meta-analysis. Tumour Biol. **35**(8): p. 7383-9.

48. Qiao, G.L., et al., 2018, Prognostic value of CD44v6 expression in breast cancer: a meta-analysis. Onco Targets Ther. **11**: p. 5451-5457.

49. Matzke-Ogi, A., et al., 2016, Inhibition of Tumor Growth and Metastasis in Pancreatic Cancer Models by Interference With CD44v6 Signaling. Gastroenterology. **150**(2): p. 513-25 e10.

50. Wu, X.J., et al., 2015, Clinical significance of CD44s, CD44v3 and CD44v6 in breast cancer. J Int Med Res. **43**(2): p. 173-9.

51. da Cunha, C.B., et al., 2010, De novo expression of CD44 variants in sporadic and hereditary gastric cancer. Lab Invest. **90**(11): p. 1604-14.

52. Jijiwa, M., et al., 2011, CD44v6 regulates growth of brain tumor stem cells partially through the AKT-mediated pathway. PLoS One. **6**(9): p. e24217.

53. Lee, H.J., et al., 2013, Colorectal micropapillary carcinomas are associated with poor prognosis and enriched in markers of stem cells. Mod Pathol. **26**(8): p. 1123-31.

54. Gaviraghi, M., et al., 2011, Pancreatic cancer spheres are more than just aggregates of stem marker-positive cells. Biosci Rep. **31**(1): p. 45-55.

55. Bourguignon, L.Y., M. Shiina, and J.J. Li, 2014, Hyaluronan-CD44 interaction promotes oncogenic signaling, microRNA functions, chemoresistance, and radiation resistance in cancer stem cells leading to tumor progression. Adv Cancer Res. **123**: p. 255-75.

56. Lv, L., et al., 2016, Upregulation of CD44v6 contributes to acquired chemoresistance via the modulation of autophagy in colon cancer SW480 cells. Tumour Biol. **37**(7): p. 8811-24.

57. Stroomer, J.W., et al., 2000, Safety and biodistribution of 99mTechnetium-labeled anti-CD44v6 monoclonal antibody BIWA 1 in head and neck cancer patients. Clin Cancer Res. **6**(8): p. 3046-55.

58. Tijink, B.M., et al., 2006, A phase I dose escalation study with anti-CD44v6 bivatuzumab mertansine in patients with incurable squamous cell carcinoma of the head and neck or esophagus. Clin Cancer Res. **12**(20 Pt 1): p. 6064-72.

59. Liang, S., et al., 2015, CD44v6 Monoclonal Antibody-Conjugated Gold Nanostars for Targeted Photoacoustic Imaging and Plasmonic Photothermal Therapy of Gastric Cancer Stem-like Cells. Theranostics. **5**(9): p. 970-84.

60. Doudna, J.A. and E. Charpentier, 2014, Genome editing. The new frontier of genome engineering with CRISPR-Cas9. Science. **346**(6213): p. 1258096.

61. Crooke, S.T., 2004, Antisense strategies. Curr Mol Med. **4**(5): p. 465-87.

62. Montefiori, L.E. and M.A. Nobrega, 2019, Gene therapy for pathologic gene expression. Science. **363**(6424): p. 231-232.

63. Evers, M.M., L.J. Toonen, and W.M. van Roon-Mom, 2015, Antisense oligonucleotides in therapy for neurodegenerative disorders. Adv Drug Deliv Rev. **87**: p. 90-103.

64. Dias, N. and C.A. Stein, 2002, Antisense oligonucleotides: basic concepts and mechanisms. Mol Cancer Ther. **1**(5): p. 347-55.

65. Siva, K., G. Covello, and M.A. Denti, 2014, Exon-skipping antisense oligonucleotides to correct missplicing in neurogenetic diseases. Nucleic Acid Ther. **24**(1): p. 69-86.

66. Zamecnik, P.C. and M.L. Stephenson, 1978, Inhibition of Rous sarcoma virus replication and cell transformation by a specific oligodeoxynucleotide. Proc Natl Acad Sci U S A. **75**(1): p. 280-4.

67. Wang, Y., et al., 2015, Mechanism of alternative splicing and its regulation. Biomed Rep. **3**(2): p. 152-158.

68. Havens, M.A. and M.L. Hastings, 2016, Splice-switching antisense oligonucleotides as therapeutic drugs. Nucleic Acids Res. **44**(14): p. 6549-63.

69. Sharma, V.K., R.K. Sharma, and S.K. Singh, 2014, Antisense oligonucleotides: modifications and clinical trials. Medchemcomm. **5**(10): p. 1454-1471.

70. Suwanmanee, T., et al., 2002, Repair of a splicing defect in erythroid cells from patients with beta-thalassemia/HbE disorder. Mol Ther. **6**(6): p. 718-26.

71. Moulton, J.D., 2017, Using Morpholinos to Control Gene Expression. Curr Protoc Nucleic Acid Chem. **68**: p. 4 30 1-4 30 29.

72. Lim, K.R., R. Maruyama, and T. Yokota, 2017, Eteplirsen in the treatment of Duchenne muscular dystrophy. Drug Des Devel Ther. **11**: p. 533-545.

73. Flanigan, K.M., 2014, Duchenne and Becker muscular dystrophies. Neurol Clin. **32**(3): p. 671-88, viii.

74. Stein, C.A. and D. Castanotto, 2017, FDA-Approved Oligonucleotide Therapies in 2017. Mol Ther. **25**(5): p. 1069-1075.

75. Nakamura, A., 2017, Moving towards successful exon-skipping therapy for Duchenne muscular dystrophy. J Hum Genet. **62**(10): p. 871-876.

76. Ran, F.A., et al., 2013, Genome engineering using the CRISPR-Cas9 system. Nat Protoc. **8**(11): p. 2281-2308.

77. Mali, P., et al., 2013, RNA-guided human genome engineering via Cas9. Science. **339**(6121): p. 823-6.

78. Sander, J.D. and J.K. Joung, 2014, CRISPR-Cas systems for editing, regulating and targeting genomes. Nat Biotechnol. **32**(4): p. 347-55.

79. Chen, M., et al., 2019, CRISPR-Cas9 for cancer therapy: Opportunities and challenges. Cancer Lett. **447**: p. 48-55.

80. Bella, R., et al., 2018, Removal of HIV DNA by CRISPR from Patient Blood Engrafts in Humanized Mice. Mol Ther Nucleic Acids. **12**: p. 275-282.

81. Young, C.S., et al., 2017, Creation of a Novel Humanized Dystrophic Mouse Model of Duchenne Muscular Dystrophy and Application of a CRISPR/Cas9 Gene Editing Therapy. J Neuromuscul Dis. **4**(2): p. 139-145.

82. Ifuku, M., et al., 2018, Restoration of Dystrophin Protein Expression by Exon Skipping Utilizing CRISPR-Cas9 in Myoblasts Derived from DMD Patient iPS Cells. Methods Mol Biol. **1828**: p. 191-217.

83. Maruyama, R. and T. Yokota, 2018, Creation of DMD Muscle Cell Model Using CRISPR-Cas9 Genome Editing to Test the Efficacy of Antisense-Mediated Exon Skipping. Methods Mol Biol. **1828**: p. 165-171.

84. Castanotto, D. and C.A. Stein, 2014, Antisense oligonucleotides in cancer. Curr Opin Oncol. **26**(6): p. 584-9.

85. Yi, L. and J. Li, 2016, CRISPR-Cas9 therapeutics in cancer: promising strategies and present challenges. Biochim Biophys Acta. **1866**(2): p. 197-207.

86. Saad, F., et al., 2011, Randomized phase II trial of Custirsen (OGX-011) in combination with docetaxel or mitoxantrone as second-line therapy in patients with metastatic castrate-resistant prostate cancer progressing after first-line docetaxel: CUOG trial P-06c. Clin Cancer Res. **17**(17): p. 5765-73.

87. Lu, Y., et al., 2018, A phase I trial of PD-1 deficient engineered T cells with CRISPR/Cas9 in patients with advanced non-small cell lung cancer. Journal of Clinical Oncology. **36**(15).

88. Cornu, T.I., C. Mussolino, and T. Cathomen, 2017, Refining strategies to translate genome editing to the clinic. Nat Med. **23**(4): p. 415-423.

89. Zammarchi, F., et al., 2011, Antitumorigenic potential of STAT3 alternative splicing modulation. Proc Natl Acad Sci U S A. **108**(43): p. 17779-84.

90. Apps, M.G., E.H. Choi, and N.J. Wheate, 2015, The state-of-play and future of platinum drugs. Endocr Relat Cancer. **22**(4): p. R219-33.

91. Oun, R., Y.E. Moussa, and N.J. Wheate, 2018, The side effects of platinum-based chemotherapy drugs: a review for chemists. Dalton Trans. **47**(19): p. 6645-6653.

92. Longley, D.B., D.P. Harkin, and P.G. Johnston, 2003, 5-fluorouracil: mechanisms of action and clinical strategies. Nat Rev Cancer. **3**(5): p. 330-8.

93. Moore, A.Y., 2009, Clinical applications for topical 5-fluorouracil in the treatment of dermatological disorders. J Dermatolog Treat. **20**(6): p. 328-35.

94. Knuth, S.T., J.D. Moulton, and P.A. Morcos, 2008, Achieving Efficient Delivery of Morpholino Oligos with Nucleofection®. Amaxa Inc. Report.

95. Wang, Z. and C.B. Burge, 2008, Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. RNA. **14**(5): p. 802-13.

96. Lee, Y. and D.C. Rio, 2015, Mechanisms and Regulation of Alternative Pre-mRNA Splicing. Annu Rev Biochem. **84**: p. 291-323.

97. Tremmel, M., et al., 2009, A CD44v6 peptide reveals a role of CD44 in VEGFR-2 signaling and angiogenesis. Blood. **114**(25): p. 5236-44.

98. Damm, S., et al., 2010, HGF-promoted motility in primary human melanocytes depends on CD44v6 regulated via NF-kappa B, Egr-1, and C/EBP-beta. J Invest Dermatol. **130**(7): p. 1893-903.

99. Ghatak, S., et al., 2014, Overexpression of c-Met and CD44v6 receptors contributes to autocrine TGF-beta1 signaling in interstitial lung disease. J Biol Chem. **289**(11): p. 7856-72.

100. Hara, T., et al., 2019, Effect of c-Met and CD44v6 Expression in Resistance to Chemotherapy in Esophageal Squamous Cell Carcinoma. Ann Surg Oncol. **26**(3): p. 899-906.

101. Thiel, A. and A. Ristimaki, 2015, Targeted therapy in gastric cancer. APMIS. **123**(5): p. 365-72.

102. Adli, M., 2018, The CRISPR tool kit for genome editing and beyond. Nat Commun. **9**(1): p. 1911.

103. Liang, P., et al., 2015, CRISPR/Cas9-mediated gene editing in human tripronuclear zygotes. Protein Cell. **6**(5): p. 363-372.

104. Huang, J., Y. Wang, and J. Zhao, 2018, CRISPR editing in biological and biomedical investigation. J Cell Physiol. **233**(5): p. 3875-3891.

105. Charlesworth, C.T., et al., 2019, Identification of preexisting adaptive immunity to Cas9 proteins in humans. Nat Med. **25**(2): p. 249-254.

106. Riethoven, J.J., 2010, Regulatory regions in DNA: promoters, enhancers, silencers, and insulators. Methods Mol Biol. **674**: p. 33-42.

107. Perri, F., et al., 2017, Epigenetic control of gene expression: Potential implications for cancer treatment. Crit Rev Oncol Hematol. **111**: p. 166-172.

108. DeAngelis, J.T., W.J. Farrington, and T.O. Tollefsbol, 2008, An overview of epigenetic assays. Mol Biotechnol. **38**(2): p. 179-83.

109. Narlikar, L. and I. Ovcharenko, 2009, Identifying regulatory elements in eukaryotic genomes. Brief Funct Genomic Proteomic. **8**(4): p. 215-30.

110. Pereira, C., et al., 2018, CD44v6 expression is a novel predictive marker of therapy response and poor prognosis in gastric cancer patients. bioRxiv: p. 468934.

111. Jung, T., W. Gross, and M. Zoller, 2011, CD44v6 coordinates tumor matrix-triggered motility and apoptosis resistance. J Biol Chem. **286**(18): p. 15862-74.

112. Wang, Q.X., et al., 2019, Downregulation of CD44v6 Enhances Chemosensitivity by Promoting Apoptosis and Inhibiting Autophagy in Colorectal Cancer HT29 Cells. Ann Clin Lab Sci. **49**(4): p. 481-487.

113. Tang, H.L., et al., 2009, Reversibility of apoptosis in cancer cells. Br J Cancer. **100**(1): p. 118-22.

114. Tang, H.L., et al., 2012, Cell survival, DNA damage, and oncogenic transformation after a transient and reversible apoptotic response. Mol Biol Cell. **23**(12): p. 2240-52.

115. Wang, K., et al., 2005, Reversibility of caspase activation and its role during glycochenodeoxycholate-induced hepatocyte apoptosis. J Biol Chem. **280**(25): p. 23490-5.

# Supplementary material

## 1. Primers design

The primers used in this project were purchased from Sigma-Aldrich® (Poole, UK) unless stated otherwise. Table S1 describes in detail the primers utilized.

Table S1 – Description of the primers used throughout this project

| Orientation | In-text Name | Binding Site | Primer Sequence (5' – 3') | Melting Temperature |
|---|---|---|---|---|
| **Forward** | A | pSpCas9(BB)_Bbs I | GGGCCTATTTCCCATGATTCCTT | Tm = 68.5°C |
| | B | CD44 Exon 8 (v5) | ATGTAGACAGAAATGGCACCAC | Tm = 62.7°C |
| | C | CD44 Intron 8 (v5) | ATCAGTGGCCTGTTTCCTTG | Tm = 64.0°C |
| | D | CD44 Exon 5 | GAAGACATCTACCCCAGCAACC | Tm = 68.0°C |
| **Reverse** | E | pSpCas9(BB)_Bbs I | GACTCGGTGCCACTTTTTCAA | Tm = 66.2°C |
| | F | CD44 Exon 10 (v7) | CCATCCTTCTTCCTGCTTGATG | Tm = 66.8°C |
| | G | CD44 Intron 9 (v6) | TTTGGCTCTGTGTGAACTGC | Tm = 64.1°C |
| | H | CD44 Exon 17/18 | TGATCAGCCATTCTGGAATTTG | Tm = 65.9°C |

The set of primers B / F and C / G were designed specifically for this project. Firstly, *Primer3 software* (http://bioinfo.ut.ee/primer3-0.4.0) was assessed in order to understand if the primers designed were good in terms of primer length, melting temperature, GC content, self-complementarity. The following Table S2 shows these primers properties. For each primer was also performed a Blat (https://genome.ucsc.edu/cgi-bin/hgBlat) to search for possible primer mismatches in other places of the human genome. No significant mismatches were found.

Table S2 - Discriminated properties of the primers designed during this project

| Orientation | Primer | Primer length | GC % | Any self-complementarity | 3' self-complementarity |
|---|---|---|---|---|---|
| **Forward** | B | 22 | 45.45 | 4.00 | 2.00 |
| **Reverse** | F | 20 | 50.00 | 4.00 | 0.00 |
| **Forward** | C | 22 | 50.00 | 4.00 | 2.00 |
| **Reverse** | G | 20 | 50.00 | 2.00 | 2.00 |

## 2. DNA sequence of the target regions for CRISPR/Cas9

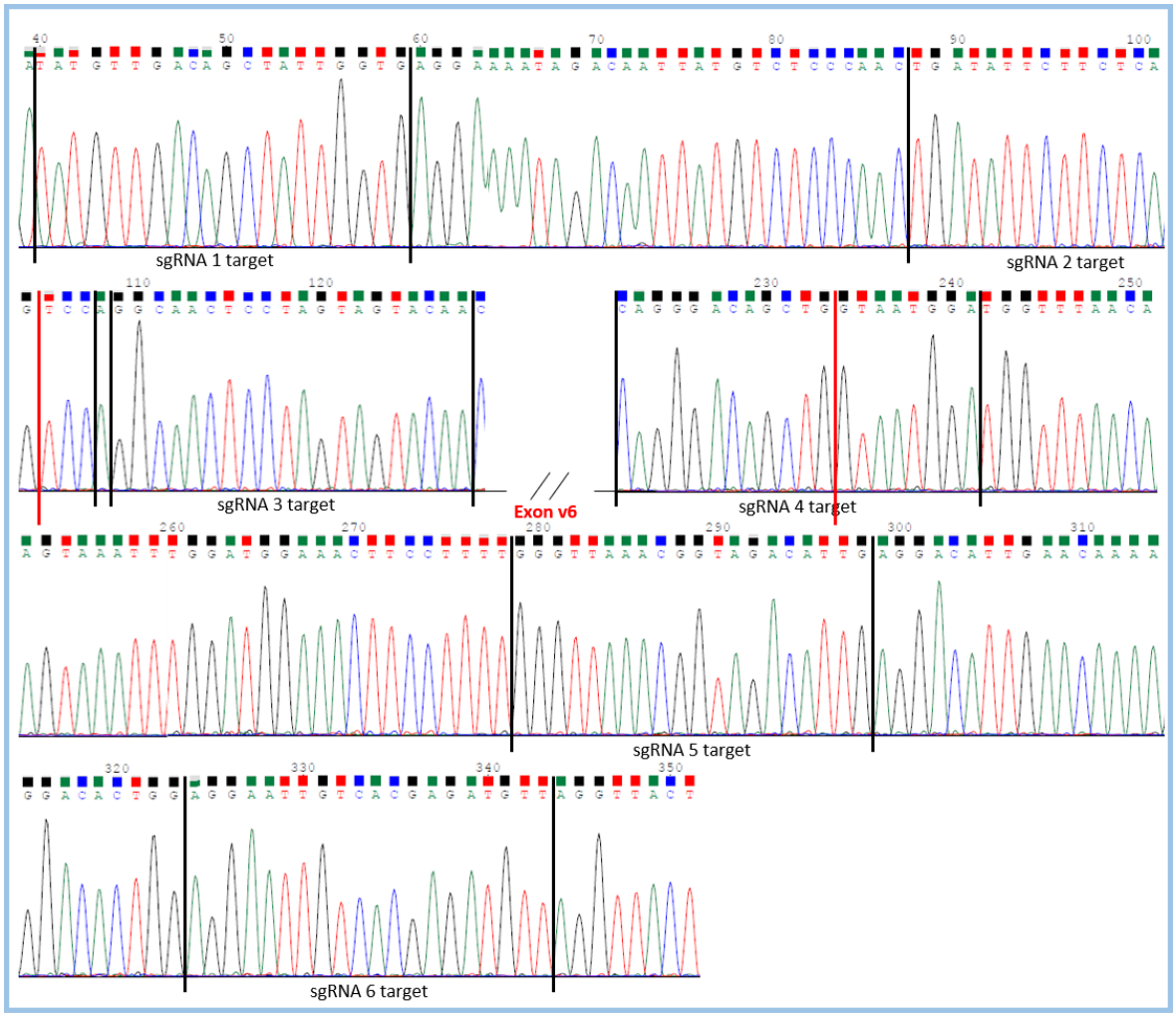


Figure S1 - Genomic DNA sequence of the v6 locus demonstrating the target regions for each sgRNA (Representative of both GP-202 and MKN-45 cell lines).