UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

DEPARTAMENTO DE BIOLOGIA VEGETAL

# Programmed DNA damage as a new tool to regulate gene activation

Cristiana Marisa de Sousa Morgado

**Mestrado em Biologia Molecular e Genética**

Dissertação orientada por:

Professor Doutor Sérgio Fernandes Almeida

Professora Rita Maria Pulido Garcia Zilhão

2018

# Acknowledgements

O desenvolvimento deste projeto de dissertação trouxe com ele um grande enriquecimento pessoal e científico que, sem a ajuda das várias pessoas que me acompanharam neste percurso não teria sido possível.

Em primeiro lugar, começo por agradecer ao Doutor Sérgio Almeida por me ter recebido no seu laboratório e, assim, me ter dado a oportunidade de conhecer e integrar o mundo da investigação científica. Agradeço igualmente por todos os conhecimentos transmitidos, pela disponibilidade demonstrada para discutir resultados e para resolver os problemas e obstáculos que teimavam em surgir (igualmente, pela calma transmitida nesses mesmos momentos), pela paciência e apoio, e sem dúvida pelo incentivo ao pensamento crítico, ferramenta essencial para a descoberta.

Agradeço à professora Rita Zilhão pela preocupação em garantir que o meu percurso durante este ano fosse o mais enriquecedor possível, pela disponibilidade demonstrada ao longo deste período e que se revelou uma grande ajuda nos momentos mais tortuosos e pela atenção mostrada na resolução de adversidades ou obstáculos impostos para a obtenção de um determinado fim.

Mostro igualmente o meu profundo obrigado à Alexandra, Sílvia e Mafalda pela orientação durante o meu projeto de dissertação e por todo o conhecimento transmitido tanto a nível científico como a nível técnico, familiarizando-me com um vasto leque de técnicas laboratoriais. Pela paciência que demonstraram perante as minhas mil perguntas diárias, que nem assim afetou a vossa disponibilidade, para me ajudar sempre que precisava, e o vosso apoio para ultrapassar obstáculos. Um sincero obrigado!

Aos restantes elementos que constituem o Laboratório SAlmeida! À Margarida e à Doutora Ana Rita Grosso pela ajuda dada na resolução de quaisquer questões bioinformáticas. À Ana, ao João e ao Robert pela disponibilidade, ajuda e apoio que me deram. Um muito obrigado a todos os SAlmeida's pela hospitalidade demonstrada que possibilitou a minha integração e pelo ótimo ambiente de trabalho criado que fez com que o meu percurso no laboratório fosse excelente.

Agradeço igualmente à Aparajita Lahree pela ajuda incansável e pelos conhecimentos transmitidos no que diz respeito ao manuseamento das células HepG2 e ao Michael Oberhuemer por toda a sua ajuda, empenho e dedicação na realização das experiências práticas o que possibilitou a conclusão deste trabalho dentro do tempo definido.

Às minhas amigas, Catarina, Sara, Salomé, Carolina, Luísa e Catarina Campos, pela amizade incondicional e pelo apoio demonstrado, estando sempre presentes em qualquer situação, para partilhar alegrias, mas também para ouvir desabafos. Obrigada pela motivação e por me fazerem acreditar em mim mesma!

Aos meus pais, às pessoas a quem eu devo tudo e que fizerem de mim a pessoa que sou hoje. Pela educação que me deram, por todo o carinho, preocupação, apoio e incentivo em continuar. Sem eles não seria possível chegar onde estou!

E por fim, um agradecimento muito especial à minha irmã, a pessoa que tornou os meus 22 anos de existência mais felizes. Obrigada por teres caminhado sempre comigo e por me ergueres sempre que eu caí. Pela infinita paciência ao aturares os meus devaneios, preocupações e nervosismos (e aqui agradeço igualmente ao meu cunhado que, por arrasto, também é sujeito a tais peripécias), pelos sorrisos partilhados, pelo apoio absoluto e incondicional, por acreditares em mim e por me fazeres acreditar que é possível. Parte do que sou o devo a ti! Obrigada por seres uma segunda mãe!

## Resumo alargado

Sendo considerada a unidade básica da vida, a célula constitui o nível mais baixo de organização a que o organismo pode ser reduzido mantendo as características da vida. Os organismos multicelulares, tal como os conhecemos, são constituídos por diferentes tipos de células que comunicam entre si de forma a assegurar o normal funcionamento dos mesmos. Esta extensa e complexa rede de conexões, quando devidamente regulada, garante a manutenção da homeostasia necessária para a sobrevivência dos organismos. Toda esta organização funcional é ditada pela informação que se encontra codificada na longa molécula de 2 metros que se encontra confinada ao núcleo da célula eucariótica, o DNA. No entanto, para que a célula tenha acesso a essa informação, várias maquinarias moleculares trabalham em conjunto no sentido de copiar e traduzir a informação presente na molécula de DNA para uma "linguagem" percetível à célula. Este processo é definido de expressão génica.

A expressão génica é um processo complexo e altamente regulado que define a identidade das diferentes células e tecidos que constituem o organismo. Sendo constituída por uma sequência ordenada de eventos, a expressão génica consiste, maioritariamente, em duas etapas principais: a transcrição, onde a informação codificada na molécula de DNA é transferida para a molécula de RNA, ou mais especificamente para o RNA mensageiro no caso de genes que codificam proteínas; e a tradução, etapa na qual a informação presente no RNA mensageiro é traduzida para uma sequência de aminoácidos dando origem às proteínas. Para o início da transcrição diversas proteínas regulatórias, nomeadamente fatores de transcrição, têm de ser sequencialmente recrutadas, juntamente com a RNA polimerase (do tipo II, no caso da transcrição de genes codificadores de proteínas), para o promotor do gene alvo. O recrutamento destes componentes, que está dependente da existência de sequências de reconhecimento específicas presentes na região promotora do gene, permite a formação de um complexo de pré-iniciação funcional, sendo este essencial para o início da transcrição. No entanto, alguns constrangimentos estruturais podem impedir a formação do referido complexo. Como referido anteriormente, o DNA é uma molécula bastante longa que, nas células eucarióticas, se encontra empacotada dentro de um pequeno compartimento, o núcleo. De forma a acomodar-se dentro deste, evitando a formação de nós ou emaranhados, várias proteínas, designadas de histonas, interagem com a molécula de DNA, compactando-a sob a forma de cromatina. Dependendo dos diferentes graus de compactação que pode assumir, a cromatina é um importante fator de regulação de vários processos biológicos que requerem acesso ao DNA, tais como a replicação, a reparação do DNA e a transcrição. Por exemplo, durante a transcrição, o reconhecimento, por parte dos fatores de transcrição, das suas sequências específicas de ligação presentes no DNA pode ser impedido pela estrutura compacta da cromatina. Sendo assim, e de forma a facilitar a acessibilidade de diversos fatores ao DNA, as células desenvolveram algumas estratégias que permitem a remodelação da estrutura da cromatina, nomeadamente através de modificações pós-traducionais de histonas e da atividade de complexos remodeladores de cromatina. Desta forma, o controlo da estrutura dinâmica da cromatina permite a montagem do complexo de pré-iniciação que, na ausência de outras adversidades, leva à transcrição produtiva do gene alvo.

Como é possível perceber, a expressão génica é um processo vital à vida, sendo que qualquer perturbação à coordenada cascata de eventos que a constituem pode resultar em consequências devastadoras. Danos no DNA são um dos fatores responsáveis pela alteração da expressão dos genes afetados. Durante o seu período de vida, o genoma de uma célula é sujeito a vários tipos de dano. Na realidade, acredita-se que, por dia, uma célula sofra milhares de lesões que desafiam a integridade do genoma, sendo que a capacidade da mesma em detetar e reparar o dano formado no DNA torna-se crucial para a manutenção das suas funções normais e para a supressão de eventos mutagénicos que podem originar cancro. Diversos fatores endógenos e exógenos originam vários tipos de danos no DNA, o que, consequentemente, levou as células a desenvolveram diversos mecanismos de reparação

especializados para lidar com esses tipos distintos de danos no DNA. O processo de reparação do DNA envolve um amplo conjunto de eventos integrados nos quais diversas proteínas, denominadas proteínas reparadoras, sentem, sinalizam e reparam o dano no sentido de proteger a célula e, consequentemente, o organismo de qualquer ameaça que pode advir.

Danos no DNA podem interferir de diversas maneiras com a transcrição génica. Enquanto que, danos localizados nas cadeias transcritas dos genes que estão a ser expressos podem resultar na obstrução da elongação da RNA polimerase, danos localizados nas regiões promotoras podem alterar e/ou impedir o recrutamento de diversos fatores de transcrição. No seu conjunto, todas estas alterações resultam na inibição da transcrição, sendo esta um dos impactos mais frequentemente associado à interação entre danos no DNA e transcrição. No entanto, esta visão de que a indução de danos no DNA apenas resulta em consequências negativas para a transcrição tem sido contrariada nos últimos anos. Evidências crescentes têm reportado que a indução fisiológica de quebras na molécula de DNA pode resultar na ativação da transcrição de vários genes. Na verdade, vários autores demonstraram que a indução dos programas transcricionais de genes dependentes de estímulos envolve a formação de danos no DNA, nomeadamente quebras da dupla cadeia de DNA (um dos danos mais letais para a células), sendo estes aparentemente necessários para a eficiente ativação transcricional desses genes. Os mecanismos através dos quais a introdução de uma quebra na molécula de DNA medeia a ativação da transcrição ainda não são completamente conhecidos; no entanto, algumas evidências revelam que essa mesma ativação possa resultar da resolução de restrições topológicas que impeçam a RNA polimerase de iniciar o processo de elongação ou então, da remodelação da estrutura da cromatina que se verifica após o dano, e durante a consequente reparação, e que consequentemente pode criar um ambiente permissivo para a transcrição.

Tendo em conta a emergente interação que se verifica ente danos no DNA e transcrição, o presente projeto tem como objetivo investigar se a introdução programada de uma quebra na cadeia dupla de DNA no promotor de um gene regula a sua ativação. Para isso, foram selecionados três genes diferentes, sendo estes o *CDKN2A*, *GSTP1* e o *RASSF1*. Os genes *CDKN2A* e *GSTP1* são genes supressores de tumor que se encontram epigeneticamente silenciados, através da hiper-metilação do seu promotor, no cancro colorretal e no carcinoma hepatocelular, respetivamente, enquanto que o gene *RASSF1*, mais precisamente a sua isoforma C, considerada um oncogene, encontra-se expresso no carcinoma hepatocelular. Para o estudo do impacto da introdução de uma quebra no DNA na transcrição dos genes selecionados, linhas celulares do cancro colorretal e do carcinoma hepatocelular, HCT116 e HepG2, respetivamente, foram infetadas com lentivírus portadores do sistema CRISPR/Cas9, com RNAs guias que têm como alvo a região promotora de cada gene. Os resultados obtidos demonstraram que a introdução de uma quebra no promotor dos genes supressores de tumor, nomeadamente no promotor da isoforma do gene *CDKN2A*, *p14ARF*, e no promotor do gene *GSTP1*, leva à reativação da sua transcrição, o que é consistente com o papel positivo dos danos no DNA na transcrição, como foi reportado em diversos estudos. Por outro lado, danos no DNA na região promotora da isoforma C do gene *RASSF1* resulta numa evidente diminuição da sua transcrição. Resultados semelhantes foram previamente reportados em um estudo recente realizado no nosso laboratório.

Em conjunto, estes resultados demonstram que a indução de uma quebra no promotor de um gene ativamente transcrito ou de um gene silenciado apresenta consequências distintas no processo de transcrição. De facto, os dados revelam que o impacto do dano no DNA na transcrição depende do *status* transcricional da região alvo. O entendimento da organização estrutural da cromatina associada à região promotora, antes e depois do dano, bem como dos componentes moleculares envolvidos na resposta ao dano poderá ajudar a clarificar a razão desta variabilidade.

**Palavras-Chave:** transcrição, danos no DNA, reparação do dano, transcrição induzida por dano

## Abstract

Multicellular organisms are composed of different tissues and cells that interact with each other in a coordinated and synchronized way in order to maintain homeostasis. The identities of the different cells and consequently of the tissues that comprise them are determined by the tightly temporal and spatial control of the transcriptional programs that enable accurate gene expression. In fact, the proportion of genes expressed at a given time defines cell fate and function. Gene transcription in eukaryotes is a complex and layered process accomplished by the interaction of regulatory factors with specific sequences present in the control regions of genes. The coordinated cascade of events involved in this process can be affected, at least transiently, by DNA damage. For instance, transcription repression has been recognized as a major outcome of DNA damage. Nonetheless, this view has been challenged over the last years by studies suggesting that DNA damage can promote gene expression activation. Several studies reported that activation of stimulus-dependent gene expression involves the formation of DNA damage, which in turn seems to be sufficient for transcription initiation. In this project, we aim to investigate whether the programmed introduction of a DNA double strand break (DSB) in the promoter of a gene regulates its activation. For that, we selected three different genes, namely *CDKN2A*, *GSTP1* and *RASSF1*. *CDKN2A* and *GSTP1* are tumour suppressor genes epigenetically silenced by promoter hypermethylation in colorectal cancer and hepatocellular carcinoma, respectively. *RASSF1*, namely its oncogenic isoform C, is expressed in hepatocellular carcinoma. The induction of a locus-specific DSB was accomplished by using the CRISPR/Cas9 system with guide RNAs directed to the promoter region of the genes. The results demonstrated that the introduction of a DSB in the promoter of the tumour suppressor genes, namely in *CDKN2A* and in *GSTP1* leads to transcription activation. On the other hand, the induction of a DSB in the C isoform promoter region of the *RASSF1* gene results in a clear decrease in its transcription. Altogether, these results demonstrate that a DSB in the promoter of an actively transcribed gene or of a silenced gene displays different outcomes. These data reveal that the impact of DNA damage in transcription depends on the transcriptional status of the target region.



**Keywords**: Transcription, DNA damage, DNA damage response; Damage-induced transcription

# Table of contents

# List of figures and tables

## List of figures



## List of tables

## List of abbreviations

| | |
|---|---|
| alt- NHEJ | Alternative non-homologous end joining |
| ATM | Ataxia telangiectasia mutated |
| bp | Base pairs |
| Cas9 | CRISPR associated protein 9 |
| CDK | Cyclin-dependent kinase |
| CDKN1A | Cyclin-dependent kinase Inhibitor 1A |
| CDKN2A | Cyclin-dependent kinase Inhibitor 2A |
| cDNA | Complementary DNA |
| CpG | Cytosine-guanine dinucleotide |
| CpGIs | CpG islands |
| CRISPR | Clustered Regularly Interspaced Short Palindromic Repeats |
| CRISPR/Cas9 | Clustered Regularly Interspaced Short Palindromic Repeats / CRISPR associated protein 9 |
| crRNA | CRISPR RNA |
| dCas9 | Nuclease -null deactivated Cas9 |
| DDR | DNA damage response |
| DMEM | Dulbecco's Modified Eagle Medium |
| DNA-PKcs | DNA-dependent protein kinase catalytic subunit |
| DNMTs | DNA methyltransferases |
| DSBs | DNA double strand breaks |
| dsDNA | Double stranded DNA |
| DSIF | DRB sensitivity inducing factor |
| DTT | Dithiothreitol |
| FBS | Fetal bovine serum |
| GAPDH | Glyceraldehhyde-3-phosphate-dehydrogenase |
| gDNA | Genomic DNA |
| gRNA | Guide RNA |
| GSTs | Glutathione S-transferases |
| GSTP1 | Glutathione S-transferase π 1 |
| h | Hours |

| | |
|---|---|
| HATs | Histone acetyltransferases |
| HCC | Hepatocellular carcinoma |
| HDACs | Histone deacetylases |
| HMGB1/2 | High motility group B1/2 |
| HMTs | Histone methyltransferases |
| HR | Homologous recombination |
| HRP | Horseradish peroxidase |
| kDa | Kilodaltons |
| KDMs | Histone demethylases |
| MDM2 | Murine double minute 2 |
| min | Minutes |
| MNase | Micrococcal nuclease |
| MRN | MRE11-RAD50-NBS1 complex |
| mRNA | Messenger RNA |
| ncRNAs | Noncoding RNAs |
| NDRs | Nucleosome-depleted regions |
| NELF | Negative elongation factor |
| NHEJ | Non-homologous end joining |
| PAGE | Polyacrylamide gel electrophoresis |
| PAM | Protospacer adjacent motif |
| PARPs | PolyADP-ribose polymerases |
| PBS | Phosphate buffered saline |
| PCR | Polymerase chain reaction |
| PIC | Pre-initiation complex |
| Pre-mRNA | Precursor mRNA |
| P-TEFb | Positive transcription elongation factor b |
| qPCR | Quantitative polymerase chain reaction |
| RA | Ras-associated |
| RASSF1 | Ras association domain family 1 |
| RASSF1C | Ras association domain family member 1 isoform C |
| Rb | Retinoblastoma protein |

| | |
|---|---|
| RFLP | Restriction fragment length polymorphism |
| RNAPII | RNA polymerase II |
| RT-PCR | Reverse transcription – polymerase chain reaction |
| RT-qPCR | Reverse transcription quantitative polymerase chain reaction |
| s | Seconds |
| SARAH | Salvador/RASSF/Hippo |
| SDS | Sodium dodecyl sulfate |
| SDS – PAGE | Sodium dodecyl sulfate polyacrylamide gel electrophoresis |
| SEM | Standard error of the mean |
| SSA | Single strand annealing |
| ssDNA | Single stranded DNA |
| TAE | Tris-acetate-EDTA |
| TFs | Transcription factors |
| TGFβ1 | Transforming growth factor β1 |
| TopoIIβ | Topoisomerase IIβ |
| TracrRNA | Transactivating CRISPR RNA |
| TSGs | Tumour suppressor genes |
| TSS | Transcription start site |
| VSV-G | Vesicular Stomatitis Indiana Virus G protein |

# 1. Introduction

## 1.1 Gene expression

Gene expression is a complex and tightly regulated process that, in the case of protein-coding genes, comprises two main steps, transcription, where the information present in the genes is transcribed into a new molecule, the messenger RNA (mRNA), and translation, where the mRNA is translated into an amino acid sequence, giving rise to proteins.[1,2] Eukaryotic transcription involves three distinct events, namely transcription initiation, elongation and termination, and it begins with the formation of a pre-initiation complex (PIC), composed mainly by transcription factors (TFs) along with RNA polymerase II (RNAPII), at the target gene promoters.[1–3] Upon overcoming the obstacle imposed by the organization of the eukaryotic DNA into chromatin, PIC is assembled and specific post-translation modifications on the histone tails and C-terminal domain of RNAPII are induced, which causes a destabilization of the interaction of RNAPII with PIC components and triggers RNAPII escape from the promoter, inducing a productive transcriptional initiation and early elongation. Thus, RNAPII starts to elongate through the coding regions, generating a nascent RNA transcript that, after completely synthetized, is separated from the polymerase and the DNA template, during transcription termination. Once transcribed, the precursor mRNA (pre-mRNA) is processed, which generates a mature molecule that is exported to the cytoplasm to be translated by the ribosomes.[1,2]

### 1.1.1 Chromatin Structure and remodelling

Cell fate and function are determined by the information codified in its vast genome that, in eukaryotic cells, is confined into the nucleus. To guarantee that the DNA, in all of its dimension, fits inside the relatively small compartment that is the nucleus, without knots and tangles, this molecule is packaged along with several proteins, histone and non-histones, to form a structure named chromatin, a stable macromolecular DNA-protein complex that compacts and protects genomic DNA.[1,4,5]

The first level of structural organization and the basic functional unit of chromatin is the nucleosome.[5,6] This unit is composed by a core histone octamer, that comprises two copies of four distinct histones H3, H4, H2A and H2B, wrapped by 147 base pairs (bp) of DNA. Each nucleosome is separated from the other by a linker DNA of 10-80bp commonly associated with the linker histone H1.[5,7] Although being involved mainly in DNA compaction, increasing evidences have demonstrated that nucleosomes functions go beyond that. In fact, the compact organization of the eukaryotic genome into chromatin can constitute a barrier for several vital processes, such as DNA transcription, replication and repair, demonstrating its role in genome regulation.[8] In the case of transcription, the presence of this organized structure hampers the access of several TFs to their binding sequences at the DNA template, during PIC assembly, which interferes transcription initiation. Therefore, chromatin needs to become accessible.[1,8] Chromatin structure is not static and dynamic rearrangements occur in response to many stimuli in order to regulate DNA-dependent cellular processes. Diverse molecular players, namely histone chaperones, histone modifying enzymes and ATP-dependent remodelers, work as chromatin modifiers to reorganize its architecture.[8,9] The interaction of these factors with the chromatin components results in post-translation modifications of histones or in the replacement of the canonical core histone by histones variants, which, together, can lead to the formation of a favourable environment that allow the access of several factors, involved in DNA-related processes, to the hidden DNA. During transcription, it was demonstrated that, before PIC assembly, many chromatin remodelers are recruited to the promoter region, forming an open and permissive environment for the binding of the TFs, through the induction of specific histones modifications and eviction of the existing chromosomes.[1,8,10] Additionally, some studies have revealed the presence of nucleosome-depleted regions (NDRs) or

nucleosome-reduced occupancy at genes promoters. Altogether, this, along with evidences that NDRs are flanked by nucleosomes containing the histone variant H2A.Z and histone tail modifications (trimethylation of histone H3 at lysine 4, H3K4me3) to facilitate nucleosome eviction, enables PIC assembly and transcription initiation. Therefore, the specific position of nucleosomes among distinct genome regions and the capacity of structural reorganization reveal the crucial role of chromatin in genome regulation.[2,11]

### 1.1.2 The role of the epigenome in the regulation of gene expression

The increasing evidences suggesting changes in gene function and consequently in cellular phenotype independently of alterations in DNA sequence, gave rise to the term "epigenetics". Over the last years, the active research focused on epigenetic modifications demonstrated that epigenetic events play a pivotal role in the regulation of many DNA-templated processes, such as replication, transcription and DNA repair.[12] DNA methylation, chromatin modifications and noncoding RNAs (ncRNAs) are some examples of epigenetic mechanisms that can interfere with gene expression, without changing DNA base sequence.[4,12–14] Despite being important in the regulation of molecular processes in normal states, it has become increasingly evident that epigenetic changes are also involve in the development and progression of many diseases, including cancer. DNA and histones modifications are tightly regulated and they are essential in functional epigenetic regulation. Both epigenetic features will be discussed in this project.[4,12]

#### 1.1.2.1 DNA methylation

DNA methylation is a heritable epigenetic mark that consists in the addition of a methyl group ($CH_3$) to the fifth carbon of the cytosine residues (5mC) in the dinucleotide sequence cytosine-guanine (CpG). This modification is mediated by three different DNA methyltransferases (DNMTs): DNMT1, which is involved in the maintenance of the methylation pattern during replication, and DNMT3a and DNMT3b which function as *de novo* methyltransferase. DNA methylation is involved in the regulation of many physiologic processes; however, it was also been associated with pathological conditions. Therefore, despite having an important role during different stages of human development, genomic imprinting, X chromosome inactivation in females and transcriptional regulation, changes in genome methylation were implicated in malignant transformation.[4,12]

Most of the genome is hypomethylated, in contrast to the underrepresented CpG sites that are heavily methylated. However, there are two regions that are maintained in a low-density DNA methylation, the CpG islands (CpGIs) and the enhancers regions. CpGIs are long sequences, associated with gene promoters, that have a high concentration of CpGs in comparison the rest of the genome.[4,15] Since both regions, promoters and enhancers, are pivotal transcriptional elements carrying recognition sequences for several factors, the presence of methylated DNA can strongly hamper the interaction of *trans*-activating factors to their binding sequences, inhibiting directly gene transcription. In fact, CpGIs methylation is a well-established mechanism of gene silencing in cancer, namely at the level of the tumour suppressor genes.[15] The mechanism behind this causality effect is well known. Promoter CpG DNA-methylation, apart from interfering with chromatin accessibility, can be recognized by methylated DNA-binding proteins. These proteins are responsible to induce or stabilize the existent histone repressive marks, such as H3K9me, and to recruit histone deacetylases (HDACs) to the same region, which lead to chromatin condensation.[16] The presence of histone repressive marks leads to subsequent DNA methylation which results in transcription shutdown. Overall, we notice that the crosstalk between DNA methylation and chromatin modifications is a major player in gene expression control.

#### 1.1.2.2 Chromatin modifications

As already referred, chromatin is a very dynamic structure that can assume different conformations depending on the genomic region where it belongs and on the cell context.[8,11,12] Chromatin states are mainly classified into two types: active (or open) chromatin, present in intergenic regions and active regulatory elements, such as promoters and enhancers, and inactive (or condensed/closed) chromatin. These conformations are intimately related with gene transcription as they drive different accessibility of genes to TFs and other regulatory elements. Therefore, active transcription occurs in regions where the chromatin conformation is more open, and inactive genes are found in regions with compact chromatin.[4]

Posttranslational modifications of histone proteins affect chromatin structure and function which, consequently, allows to regulate dynamically all DNA-based processes, such as transcription, replication and DNA repair.[10,17] Many histone modifications have been identified; however, the most common are methylation, acetylation, phosphorylation and ubiquitylation that target, often, several amino acids residues of the histone tails, giving rise to histone marks.[8,10,17] Since it was demonstrated that these modifications can change the net charge of the histones, the DNA-histone interactions may be weakened, which in turn would result in the loss of inter or intra-nucleosomal DNA-protein connections and consequently in a direct change in chromatin. Additionally, it was shown that histone marks can also modify the chromatin in an indirect fashion, by recruiting chromatin remodelers, that recognize the epigenetic modifications.[8,17]

The patterns of histone modifications are established through the activity of many regulatory proteins that write, read and erase histone marks. These three groups of proteins are epigenetic writers, such as histone acetyltransferases (HATs), histone methyltransferases (HMTs) and kinases; epigenetic readers, such as proteins containing bromodomains; and epigenetic erasers, such as HDACs, histone demethylases (KDMs) and phosphatases.[4,18] It was reported that specific posttranslational modifications of histones are often associated with particular chromatin states. For instance, high levels of acetylation of the histone 3 and for 4 (such H3K9ac) and di- or trimethylation of H3 lysine 4 (H3K4me2 and H3K4me3) are found in actively transcribed promoters, while trimethylation of H3 lysine 9 and 27 (H3K9me3 and H3K27me3) are found in repressed genes. Additionally, H3K4me and H3k27ac are marks for active enhancers. Accumulation of marks such H3K4me3 and H3K36me3 has been associated with the act of transcription.[4,8,18]

Given the crucial role of chromatin in several DNA-related processes, chromatin remodelling, by histone modifications or other remodelers, must be tightly regulated, as abnormal widespread of epigenetic marks and consequently aberrant chromatin organization can be linked to several diseases, including cancer.

### 1.1.3 Gene expression and cancer

The precise control of gene expression is essential to all cellular processes, such as cellular differentiation and cell cycle regulation, which, overall, maintains the homeostasis of the organisms; however, misregulation of this process can culminate in many diseases, including cancer.[19]

Cancer is one of the most common human diseases characterized by the accumulation of several molecular abnormalities that are the cornerstone of the known hallmarks of cancer.[4] It is well known that mutations in key regulator genes, such as genes involved in the control of cell proliferation and differentiation, can lead to cancer. All these genes belong to two different categories of genes frequently mutated in cancer, named oncogenes and tumour suppressor genes (TSGs).[19,20] TSGs are defined as genes that code proteins whose function leads to the inhibition of the proliferative and survival capacity of a cell[15], whilst oncogenes are a variant of the normal proto-oncogenes, commonly involved in cellular signalling, that after aberrantly activated encode proteins capable to cause cellular transformation.[20] It was reported that a change in the balance of oncogenes and TSGs expression, namely the overexpression

of oncogenes and the loss of TSGs, leads to a misregulation of normal epistasis being, therefore, a driver of tumorigenesis.[20,21]

For a long time, genetic alterations were considered the main driver of carcinogenesis. However, it became evident that the landscape of genetic alterations was insufficient to explain all the changes in cellular function in cancer. The discovery of the epigenome and its impact in gene expression and in regulation of cell proliferation came to disclose the missing piece in the oncogenic process.[4] In fact, over the past years, growing evidences have depicted that, apart from genetic alterations, epigenetic lesions can play a crucial role in tumour initiation and progression.[4,12,15,19] Currently, it is well known that one of the common aberrations to all cancers and, perhaps, one of the early driving events, is the epigenetic silencing of TSGs.[15] This reprogramming of the epigenetic code is as pervasive as genetic alterations and can be one of the hits proposed in Knudson two-hit hypothesis. According to Knudson, to block the mechanisms, mediated by TSGs, that prevent aberrant cellular growth and tumorigenesis, both alleles of the TSGs need to be affected, classically by genetic alterations. Examination of some cancers revealed that the first "hit" that affects TSGs is a hereditary or somatic mutation in one allele; an acquired mutation or copy number loss affecting the other allele is seen as the second "hit". However, it has been demonstrated that in some cases, epigenetic modifications, instead of genetic changes, affect the remaining wild type allele of TSG, repressing its expression. This second 'hit' is responsible for the loss of function of the TSG.[4,22]

Additionally, misregulation, through gene mutation, of histone modifying proteins or other chromatin remodelers (such as histone mark readers), has been frequently identified in several cancer types. This deregulation leads to the aberrant deposition of histone marks, which can reinforce TSGs silencing, and to several other abnormalities that alter cell cycle regulation and cell homeostasis, an essential requirement for tumorigenesis.[4,12] Understanding the impact of genetic alterations and epigenetic mechanisms in gene expression regulation can reveal new avenues for therapeutic strategies.

### 1.2 DNA damage and the DNA damage response

Over its life, a cell can experience several damages in its genome. In fact, human cells are exposed to thousands of lesions per day that can affect both genome structure and the information encoded, threatening, consequently, the ability of transmit the genetic information to its descendent and cell survival.[23,24] All cellular functions are dictated by the genetic information codified into the genome and, therefore, maintaining its integrity is essential to avoid the development of several pathologies.[25] To safeguard DNA integrity, cells have developed a complex network of signal transduction pathways that detect and signal DNA damage and sets a coordinated response to repair the DNA. This cellular response is named DNA damage response (DDR).[23,24] During DNA repair, several DNA repair proteins are recruited to the damaged site. However, the compact chromatin structure and nucleosome organization that they encounter surrounding the damaged site represent a barrier for efficient DDR. Thus, chromatin must be reorganized upon DNA damage and the minimal steps for this reorganization was described in the "access-repair-restore" model, proposed by *Smerdon* in 1991. According to this model, once DNA damage is recognized, chromatin becomes more accessible to enable the assembly of the DNA repair machinery and consequently the processing and repair of the damage. After repair has been completed, local chromatin organization is restored.[25,26] As previously referred, post-translation modifications of histones and chromatin remodelers are key regulators of chromatin structure, modulating DNA accessibility, being, therefore, fundamental in DNA repair.[27,28] In fact, the activation of DDR induces chromatin-based responses, including the appearance of a broad spectrum of modifications, that are involved in the early reorganization of the chromatin, leading to the formation of an open and relaxed structure for the assembly of the DDR cascade.[25,28,29] Additionally, it is known that chromatin restoration upon completion of DNA repair, relies on histone chaperone and other remodelling factors.[26] Altogether,

this reveal the crucial role of histone modifications and chromatin remodelers in DNA damage response, ensuring its efficiency and accuracy.

Several assaults can affect DNA integrity and result in damaged DNA. Exogenous factors, such as ionizing and ultraviolet radiation, and endogenous factors, namely reactive oxygen species or erroneous DNA replication, are both sources of DNA damage. Considering this panoply of aggressions that can affect DNA, distinct types of DNA damage can be identified, such as DNA single strand breaks, DNA double strand breaks and many others. Depending on the nature of the DNA lesion, specific repair mechanisms are used.[23–25]

### 1.2.1 DNA double strand break repair

DNA double strand breaks (DSBs) are the most catastrophic form of DNA damage as they consist in the physical cleavage of both strands of the sugar backbone that compose the DNA molecule.[25] This type of damage can be repair by, at least, four independent mechanisms: non-homologous end joining (NHEJ), alternative-NHEJ (alt-NHEJ), homologous recombination (HR) and single strand annealing (SSA).[24] However, NHEJ and HR are the major pathways used to repair DSBs.[30] The choice between these two pathways depends mainly on three factors: the cell cycle phase where the damage occurs, the extent of DNA end processing and the presence and activity of specific DDR factors, namely 53BP1 and BRCA1 proteins that seem to play a key role in favouring one repair pathway over the other.[23,24,31] Regarding the cell cycle phase in which the damage occurs, it is known that NHEJ can occur throughout the cell cycle, being the repair mechanism predominant in G1, whilst HR is restricted to S and G2-M phases, where sister chromatids are present to be used as a homology donor for the repair.[24,25,30,32,33]

NHEJ consists in the direct re-joining of the two DNA ends generated upon damage.[32] This mechanism involves the minimal processing of the damaged DNA and, like most DNA repair processes, requires a nuclease to resect damaged DNA and a ligase to restore the DNA strands.[30,32] Therefore, when a DSB occurs, Ku heterodimer, composed by Ku70 and Ku80 proteins, binds rapidly to the DNA and recruits and activates the DNA-dependent protein kinase catalytic subunit (DNA-PKcs). In fact, Ku proteins function as a docking site, since they are capable of interact with several proteins involved in the repair process, such as the nuclease and the ligase. DNA-PKcs complex plays a critical role in stabilizing DSB ends and it has several nucleases activities. Autophosphorylation of this complex provides access to ARTEMIS, an end processing enzyme responsible to cut the damaged DNA overhangs. Upon DNA-PKcs ligation, DNA ligase IV along with its binding partners XRCC4 and XLF, are recruited to religate the broken ends.[24,30,32,34] The flexibility associated with the NHEJ repair enzymes permits them to act on a wider range of possible substrate configuration, in any order. This, consequently, can result in a substantial diversity of junctional outcomes, namely the ligation of incompatible ends.[30,32] Since, during the DNA end processing, insertion or deletion can be formed, NHEJ is considered a low-fidelity repair mechanism, being defined, therefore, as an error-tone pathway.[27]

In contrast, HR is considered a more accurate mechanism as it uses homologous sequences present in the genome to repair the DNA broken ends.[33,34] In this pathway, after the formation of a DSB, a member of the polyADP-ribose polymerases (PARPs) family, PARP1, senses the damage being recruited to the damaged site where it competes with Ku for binding to DNA ends in order to promote HR. PARP1 is then responsible for the recruitment of MRE11-RAD50-NBS1 complex (MRN) to the DSB, and together they promote the recruitment and accumulation of ataxia telangiectasia mutated (ATM).[24] The activation of this kinase leads to the phosphorylation of several proteins; one of its main targets is the phosphorylation of the c-terminal of the histone variant H2AX. Phosphorylated H2AX, known as γH2AX, initiates a cascade of modifications with the recruitment and assembly of many DNA repair factors. In fact, γH2AX creates a binding site for MDC1 protein that functions as a docking spot for additional repair proteins. The interaction between ATM and MDC1 promotes the spreading of

γH2AX to an extend of 100 kilobases around DSBs.[24,25] Following that, CtIP is recruited to the damaged site, in an ATM and MRN dependent manner, and associates with BRCA1 and MRN. The nuclease activity associated to this complex of factors along with the activity of additional nucleases involved in further end processing create single stranded DNA (ssDNA) overhangs, a process named DNA end resection.[24,27] RPA complex binds to 3'-ssDNA, to stabilize this conformation, and then it is replaced by RAD51, through a process mediated by BRCA2 protein. RAD51-coated filament is responsible for searching and invading homologous sequences that will serve as a template for the DNA polymerase. A DNA ligase joins ssDNA and nicks.[24,27,34]

The tightly spatial-temporal regulation of the arsenal of DNA repair factors that cells have is crucial for the efficiency and accuracy of the choreographed response that is DNA damage response, which prevents potentially deleterious alterations that threats DNA integrity and, consequently, protects the organism against pathological consequences.

### 1.3 Interplay between DNA damage and transcription

For many years, DNA damage was considered a detrimental by-product for cells, as it can lead, along with other abnormalities, to mutations, genomic rearrangements and genomic instability, being consequently linked to many pathologic conditions such as cancer.[23,35] Among the several sources that can originate DNA damage, transcription is one of them, as it involves the destabilization of the DNA helical structure.[36] In fact, it was reported that transcription-related stress, associated with the generation of positive supercoiling, or with the susceptibility of the non-transcribed single strand to mutations and to the formation of noncanonical structures, or even with the formation of R-loops, nucleic acid structures formed by the nascent RNA, complementary bound to the template DNA strand, and the displaced non-transcribed strand, can lead to the accumulation of DNA damage, promoting genome instability.[23] However, this interaction between transcription and DNA damage does not occur in one single direction. It is well known that DNA damage leads to alterations in the expression of the affected genes, interfering with the transcriptional programs in multiple ways. The most prevalent and well characterized outcome of DNA damage in transcription is the damage-induced transcriptional repression. This can occur through different mechanisms such as by the forced stalling of the elongating RNAPII complexes or by the modulation of transcription factors recruitment.[23,36] In a recent work develop in our lab, it was demonstrated that the induction of a DSB on a gene results in immediate suppression of the pre-existing transcription (unpublished data), which supports the view that DNA damage on active genes disrupts their expression.[23]

Contrarily to the previous view, emerging outstanding evidences have demonstrated that DNA damage can instead lead to the activation of transcriptional programs. This intriguing relationship was only visualized under certain circumstances, being mainly described in stimulus-response genes.[23,35,36] The mechanisms by which DNA damage mediates transcription activation are not completely understood. However, according to the literature, there are three possible models that can explain DNA damage-dependent transcriptional activation. One of the models postulates that DNA-induced breaks can resolve topological constraints associated with highly transcribed genes, and the class of enzymes, named topoisomerases, seems to be the major players in this mechanism.[23] Favouring this idea, several studies showed that type II topoisomerase, topoisomerase IIβ (TopoIIβ), is frequently associated with multiple nuclear hormone receptors and is also found bonded to promoters and enhancers of newly transcribed genes.[35] In a report published in Cell, *Madabhushi et al.* demonstrated that, upon neuronal activity stimulation, TopoIIβ is recruited or activated, which leads to the formation of a DSB within the promoter of activity-regulated genes. This damage alleviates the high torsional stress proposed to be associated with activity-response genes, allowing the release of RNAPII from its paused state, which results in productive transcriptional elongation. *Madabhushi et al.* also showed that the artificially

introduction of DSBs to the promoter of the study genes, using the CRISPR/Cas9 system exert similar results, which reveals that DNA damage is sufficient for transcription activation.[37,38]

The second model is linked to the previous one and proposes that DNA breaks are involved in the transition between promoter-proximal pausing and productive elongation.[23] RNAPII transcription complexes pausing are often observed at the promoter of several genes, particularly at inducible and developmental regulated genes. This phenomenon is an important mechanism of transcription regulation and is prevalent in higher eukaryotes, where it is mediated by pause-inducing factors namely the DRB sensitivity inducing factor (DSIF) and negative elongation factor (NELF).[1,39] Promoter escape relies on the dissociation of the previous factors, which can be achieved upon phosphorylation by the positive transcription elongation factor b (P-TEFb).[40] TRIM28 is an additional regulator of promoter proximal pausing in mammalian cells and it was found to be associated with RNAPII pause site. *Bunch et. al* demonstrated that the formation of a DNA break, mediated by the TopoIIβ, and the ensuing DDR signalling, which involves the activation of ATM and DNA-PK, both DNA damage proteins responsible for the phosphorylation of TRIM28, is required for effective RNAPII pause release and transcriptional elongation.[40] The present mechanism of transcriptional activation-coupled DDR signalling is also involved in the activation of genes that possess RNAPII paused by DSIF and NELF, since it was showed that ATM activation upon DNA damage can lead to P-TEFb activation.[41]

In the third model, it is proposed that DNA damage induces chromatin modifications that leads to the formation of a transcription-prone chromatin structure.[23] *Rosenfeld et al.* reported that signal-dependent activation of gene transcription by nuclear receptors, in this case estrogen receptors, requires TopoIIβ recruitment and binding to the target promoters, which generates a transient DSB. The establishment of the DDR with the activation of PARP1 and the subsequent recruitment of other DNA repair factors results in local changes of chromatin structure, namely in the nucleosome-specific exchange of H1, a general repressor involved in chromatin compaction, by high motility group B1/2 proteins (HMGB1/2) that is involved in chromatin relaxation being, therefore, considered as a transcription activator.[42,43] A plenty number of other studies provide further evidences supporting this interplay between the molecular machineries involved in DNA repair, and responsible for DNA and chromatin conformational changes, and gene transcription activation. In fact, this connection was also reported in the context of the transforming growth factor β1 (TGFβ1) signalling [42,44] and in the case of the anti-apoptotic *bcl-2* gene estrogen dependent.[42,45] Intriguingly, and in agreement to this view, it was showed that DDR factors can display a dual role, being involved in the RNAPII recruitment. This enzyme has the ability to evict nucleosomes, leading to the formation of nucleosome-free regions that facilitates TFs binding. The seeding of positive chromatin marks for transcription, during DDR signalling, results in the modification of chromatin status, generating a transcription-prone environment.[23]

According to the previous information, cells may have developed a drastic method that relies in the use of programmed DNA breaks to promote transcription activation. Nonetheless, this form of transcription regulation, through the generation of DNA damage, is risky and may represent a critical vulnerability for cells, since it can lead to a significant mutational load.[23,24]

### 1.4 Aim of the study

Recent studies disclosed a puzzling association between transcription activation and DNA damage. They revealed that the induction of transcriptional programs of ligand or signal-dependent genes involves the formation of DSBs and it seems that this induced DNA damage is sufficient to drive transcription initiation.

In the present project we went to investigate whether the introduction of a DSB in the promoter of an inactive tumour suppressor gene drives its activation. In contrast, we reasoned that a DSB at the promoter of an active oncogene would suppress it. For that, we selected three different genes: cyclin-

dependent kinase Inhibitor 2A (*CDKN2A)*, that encodes two different isoforms, *p16INK4a* and *p14ARF*, both subjects of study in this work; ras association domain family member 1 isoform C (*RASSF1C)* and glutathione S-transferase π 1 (*GSTP1)*. *CDKN2A* and *GSTP1* are tumour suppressor genes epigenetically silenced, by promoter hypermethylated, in colorectal cancer and in hepatocellular carcinoma (HCC), respectively. *RASSF1*C functions as an oncogene, is commonly expressed, and is going to be studied here in the context of HCC. HCT116, a colorectal cancer cell line, and HepG2, a HCC cell line, will be infected with virus carrying the CRISPR/Cas9 system, to induce a promoter-specific DSB. We hypothesized that, in the case of TSGs, the induced DNA damage and the ensuing repair will alter methylation pattern of the gene, allowing the reactivation of gene transcription. The chromatin remodelling upon damage can also contribute for this reactivation, as the nucleosome free region generated can facilitate the binding of RNAPII.[23,46] Regarding the oncogene, since it is expressed in the selected cells, we anticipate that its transcription will be suppressed, in agreement with unpublished data obtained in the host lab using reporter genes.

# 2. Methods

## 2.1 Cell Culture

The cells used in this project were HCT116 (human colorectal cancer cell line), HepG2 (hepatocellular carcinoma cell line), HeLa (human cervical cancer cell line) and HEK293 (human embryonic kidney cell line). HepG2 cells were kindly provided by Maria Mota's Laboratory, from Instituto de Medicina Molecular João Lobo Antunes.

All cell lines were grown as monolayers and were maintained in T75 flasks (Thermo Fisher Scientific, product #156499), in Dulbecco's Modified Eagle Medium (DMEM) (Gibco® by life technologies ™, product # LTID 21969-035) supplemented with 10% (v/v) fetal bovine serum (FBS) (Gibco® by life technologies ™, product #LTID 10270-106), 1% (v/v) 200 mM L-glutamine (Gibco® by life technologies ™, product #LTID 25030-024) and 1% (v/v) penicillin-streptomycin (Gibco® by life technologies ™, product #LTI 15070-063), in a 5% $CO_2$ humidified incubator at 37 ºC.

Due to the capacity of the cells of dividing and growing, the proportion of the surface area in the flask covered by cells, defined as confluency, will increase over time, which consequently reduces the area available for cells to grow. Regarding this and to enable further propagation of the cells, when a confluency of 80-90% was achieved, the cell lines used were transferred from the previous flask into a new flask with fresh growth medium, a process called subculturing.[47] First, the medium, from the flask where the cells are, was discarded and the cells were washed with phosphate buffered saline (PBS) 1X. Then, trypsin-EDTA 0.05% (Gibco® by life technologies ™, product LTI 15400-054), which will detach the cells from the surface, was added to cover all the surface of the flask and it was removed before the incubation of the flask 3-5 minutes (min) at 37 ºC. To resuspend the cells, fresh medium was added and after pipetting up and down, a defined volume of the suspension (according to the desired dilution) was transferred to a new flask, already with fresh medium. This standard procedure was applied in all four cell lines used, however in the case of HepG2 cells the procedure was slightly different. Instead of washing the cells one time with PBS, HepG2 cells were washed twice. For the treatment with 0.05% trypsin, the incubation time at 37 ºC were 5 min maximum (should not be more because HepG2 are sensitive to the trypsinization) and the trypsin was not removed before the incubation. These cells are difficult to detach thus, during the incubation time with trypsin the flask should be hit to facilitate dispersal. As HepG2 cells grow in islands (small aggregates), after the resuspension of the cells with medium, the suspension was passed through a 40 µm cell strainer (Falcon®, Product #352340) to get rid of clumps and any debris produced by the cells. Then, the suspension was centrifuged at 1500 rpm, 5 min, the supernatant was discarded, and more medium was added. After that, the cells were counted using a hemocytometer, and according to cell density of the suspension, 1.5 million cells were transfer into the new T75 flask.

To perform all the experiments mentioned below, the cells were seeded and maintained in the same conditions as previously, however in DMEM supplemented with 10% (v/v) FBS and 1% (v/v) 200 mM L-glutamine without antibiotics, to diminish the variability of the experiments.

### 2.1.1 Cell Counting

As previously mentioned, it was used a hemocytometer to count cells. This device is a counting chamber that consists in a big square divide in little squares **(Figure 2.1)**.

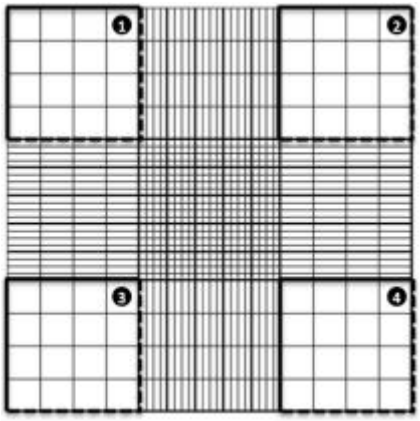

**Figure 2.1: Pattern of squares present in the hemocytometer.** Only the cells present in the numbered squares should be counted. The cells over the solid lines are counted but not the ones over the dashed lines. Adapted from *Fuentes, M.*[49]

To perform a cell count, 10 µL of cell suspension mixed with 10 µL of trypan blue (Sigma, product #T8154) (1:1 dilution) was loaded into the hemocytometer. Trypan blue is a dye used to assess cell viability, that stains blue cells with a damaged cell membrane (dead cells). Healthy viable cells, as they maintain cell membrane integrity, will not be stained because the dye cannot enter. This, therefore, allows to distinguish viable cells from non-viable cells (dye exclusion method).[48] Under a microscope, the viable cells from three quarters ($n_v$) **(Figure 2.1)**, to be representative of the cell density of our culture, were counted and the average was performed to determine the number of viable cells per quarter. As it is known that each quarter of the hemocytometer has a volume of 0.1 $mm^3$, the viable cell density of our suspension (N, cells/µL) can be calculated using the following equation 2.1. Note that, as we used trypan blue in a 1:1 dilution it is necessary to multiplicate by 2.

$$N = \frac{n_v}{3} \times 10 \times 2 \quad (2.1)$$

Once determined the cell concentration, it can be calculated the required volume of cell suspension that contains the desired number of cells to transfer into the new flask.

Something important to be considered during the count is that, as inevitably some cells will be on the boundaries of squares, only the cells present in two sides of the square (for example, the solid lines in figure 2.1) should be counted in order to avoid counting the same cell twice. This practice needs to be consistent throughout the count.[49]

## 2.2 RNA extraction

Gene expression is a complex and dynamic process in which the information encoded in a gene is transmitted to an RNA molecule, through transcription, and then decodified into a protein, after RNA translation. Not all the genes present in a cell are expressed at the same time; depending both on external and internal environmental signals the gene expression pattern changes over the time, which dictates cell function.[50] Without RNA the information from genomic DNA (gDNA) remains inaccessible so, for a gene to be expressed, RNA molecule(s) need to be formed. Therefore, the quantification of RNA (specifically mRNA) in a given time can be used as a measure of gene expression.

To confirm if the cells in study, HCT116 and HepG2, express the TSGs and the oncogene selected (*CDKN2A*, *RASSF1C* and *GSTP1)*, before and after the induction of a promoter-specific DSB, total RNA isolation was performed using TRIzol™ reagent (Life Technologies, Product # 15596018). It was also extracted RNA from HEK293 cells, a positive control for the expression of *CDKN2A* and for *GSTP1*, and from HeLa, a positive control for the expression of *RASSF1C*. The cells were plated on a

35 mm culture dish (Falcon®, Product #353001) and, when confluent, they were washed with PBS X1 and harvested using TRIzol™ reagent. This product is a solution of phenol and guanidine isothiocyanate that allows to perform a sequential precipitation of RNA, DNA and proteins from the same sample.[51] Then, the samples containing TRIzol were homogenized, chloroform was added and a high-speed centrifugation (12000 rcf, 15 min, 4ºC) was performed. At this point, three different phases are formed: an aqueous upper phase which contains the RNA, a middle phase or interphase which contains the DNA and an organic lower phase which contains the proteins.[51] The upper phase was carefully collected and transferred to a new eppendorf tube and isopropanol was added to precipitate the RNA. After an incubation time of 10min at room temperature followed by a high-speed centrifugation (12000 rcf, 10 min, 4 ºC), an RNA pellet was formed. The supernatant was carefully discarded, the precipitated RNA was washed with 75% ethanol (7500 rcf, 10 min, 4 ºC) to remove impurities and then, after air drying, it was resuspended with DNase/RNase free water (at this point the RNA can be stably stored at -80 ºC). To eliminate any residual DNA that may be in the RNA samples, avoiding problems with gDNA contamination in furthers steps, it was performed a treatment with DNase. For this, 5 µg of RNA was added to a new tube along with 5 µL of buffer DNAse I recombinant 10x RNase free, 5 µL of DNase I recombinant 10 U/µL RNase free (Roche, Product # 4716728001), 1 µL of Glycoblue™ coprecipitant (Invitrogen™, Product #LTABAM9516) and DNase/RNase free water until 50 µL. After 2 hours (h) of incubation at 37 ºC, the RNA was precipitated using 100% ethanol and 3 M sodium acetate (-80 ºC at least 30 min or -20 ºC overnight). The RNA pellet formed was then collected through high-speed centrifugation (15000 rpm, 30 min, 4 ºC), washed with 75% ethanol (15000 rpm, 10 min, 4 ºC) and, when air-dried, resuspended in DNase/RNase free water. Before being used in downstream applications, RNA samples were quantified using a spectrophotometer device (NanoDrop 2000, Thermo Fisher Scientific). For long term storage, the RNA samples were stored at -80 ºC and for short term storage at -20 ºC.

### 2.3 Reverse transcription Polymerase Chain Reaction (RT-PCR)

Reverse transcription Polymerase Chain Reaction (RT-PCR) is a common method used in gene expression analysis that takes advantage of the activity of reverse transcriptases, retroviral enzymes that have the ability to synthetize complementary DNA (cDNA) from mRNA transcripts[50]; the DNA obtained can be used subsequently as a template in PCR. In this technique, using DNA polymerases, the number of copies of the starting DNA (gDNA or cDNA) doubles per cycle (denaturation, annealing and extension), leading to an exponential amplification. The RT and the PCR can be performed in a single step (where the techniques are combined) or in two steps (separated techniques).[50,52] In this project it was used the two-steps approach.

To confirm that the cells used, HepG2, don't express the TSG *GSTP1* and express the oncogene *RASSF1C*, as it is referred in the literature[53,54], it was performed a RT-PCR (the epigenetic silencing of *p16INK4a* and *p14ARF* was confirmed by RT-qPCR. See section "2.8 Quantitative PCR"). HEK293 and HeLa cells were used as positive controls. To convert RNA into cDNA, a defined amount of total RNA (500 ng), obtained previously from RNA isolation (see 2.2 RNA Extraction), was mixed with 2 µL of random primers (300 ng/µL, Invitrogen™), 2 µL of dNTPs (10 mM, Nzytech, product # MB08701) and DNase/RNase free water (until 20 µL) and, then, incubated 5 min at 65 ºC. After that, a mix with 8 µL of 5x first strand buffer, 4 µL of dithiothreitol (DTT) (Invitrogen™, product #18064014) and 2 µL of ribonuclease inhibitor (Nzytech, product # MB08402) was added, and the samples were once again incubated, this time 2 min at 25 ºC. Without RNases that could degrading the RNA, the samples were incubated 25 ºC for 10 min, 42 ºC for 50 min and 70 ºC for 15 min with Super Script II Reverse Transcriptase (Invitrogen™, product #18064014) and finally 20 min at 37 ºC with RNAse H (Nzytech, product # MB08501) that degrades the RNA strand hybridized to DNA. A negative control, consisting in a sample with all the components mentioned above but without Super Script II, was also

performed during this protocol, in order to check DNA contamination of the reagents and even of the samples.

The cDNA obtained in reverse transcription was then used in the ensuing PCR. For each PCR reaction, it was used 1 µL of a 1:15 dilution of cDNA, 2,5 µL of reaction buffer 10X, 0,6 µL of dNTPs mix, 0,5 µL of each forward and reverse primers, 1,25 µL of $MgCl_2$, 0,25 µL of NZYTaq II DNA polymerase (Nzytech, Product # MB35401) and DNase/RNase free water up to a total volume of 25 µL. The PCR conditions were one cycle at 95 ºC for 2 min, followed by 35 cycles at 95 ºC for 30 seconds (s), 60 ºC for 30 s and 72 ºC for 1 min and a final cycle at 72 ºC for 5 min. All the primers used were designed using Primer3Plus (https://primer3plus.com/cgi-bin/dev/primer3plus.cgi) or NCBI/Primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/), and their sequences are presented in Annex I **(Table 6.1).** Glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) gene is a housekeeping gene used as a loading control.

The DNA products of the PCR were resolved on a 2% agarose gel, made with Tris-acetate-EDTA (TAE) buffer 1X. The visualization of DNA was achieved using Midori Green (Grisp, Product #MG04), which allowed nucleic acid detection, under UV light, in ChemiDoc™ XRS+ System (Bio-Rad).

### 2.4 Western Blot

The western blot is an analytical technique widely used in cell and molecular biology for protein analysis. It is a qualitative and semi-quantitative method that allows the immunodetection of specific proteins in complex cell homogenates or in other kind of protein mixtures. To accomplish that, after extracting proteins from the sample in study (e.g. cells or tissues), the mixture of proteins is separated by polyacrylamide gel electrophoresis (PAGE), then they are transferred to a membrane and finally the target proteins are selectively detected using specific antibodies.[55,56] In this project, it was used a variant of PAGE, SDS-PAGE, a denaturing polyacrylamide gel system that separates the proteins based on their molecular weight. Sodium dodecyl sulfate (SDS) is a detergent that binds to the proteins destroying their structure (denaturation) and coats them with a negative charge. In this way, the resulting extracts will be constituted by negatively charged linear polypeptide chains, making mobility through the gel unaffected by the amino acid composition of the protein.[57]

The expression of the *CDKN2A* (*p16INKa* and *p14ARF*), *RASSF1C* and *GSTP1*genes was also evaluated by the presence of the encoded proteins of each gene, via western blot. To obtain the cell protein extracts, HCT116, HepG2, HEK293 and HeLa were seeded on a 35 mm culture dish (Falcon®, Product #353001), and when confluent they were washed with PBS 1X. Protein isolation was performed by adding Laemmli Buffer 2X composed by 80 mM Tris-HCl (pH 6,8) , 16% glycerol, 4,5% SDS, 450 mM DTT and 0,01% bromophenol blue, with 1 % (v/v) benzonase (Sigma, Product #E1014-25KU,) and 1 % (v/v) magnesium chloride ($MgCl_2$) directly to the dish. The samples were incubated for 20 min at room temperature and then boiled for 5 min at 100 ºC, which will, along with the anionic detergent SDS, denature the proteins. Equal amounts of total protein extract and NZYColour Protein Marker II (Nzytech, Product # MB09003) were loaded into a 12% polyacrylamide gel (the percentage of polyacrylamide used determines the resolution obtained and is optimized depending on the size of the proteins) and an electric field of 45 mA in a Mini-PROTEAN Tetra Electrophoresis System (Biorad, Product #165-8000) was applied, allowing the migration of the proteins with different mobilities according to their size. Next, the proteins were transferred to a nitrocellulose membrane (GE Healthcare Life Sciences, Product #10600002 and Invitrogen™, Product #IB301001) through either dry transfer, with an iBlot System (Invitrogen™, Product #10042275) in a 7 min program, or wet transfer with a Mini Trans-Blot® Electrophoretic Transfer Cell (Bio-Rad), 100 V for 1 h. After that, the membrane was blocked for 1 hour at room temperature using 5% (m/v) of blocking solution, namely powder milk in PBS-tween 0,05% (PBS 1X containing 0,05% (v/v) Tween 20) to prevent nonspecific binding of the antibodies, and then was incubated overnight, at 4 ºC, with the appropriate dilutions of the respective

primary antibodies in 5% blocking solution. The primary antibodies used were anti-CDKN2A/p14ARF (1:1000, Abcam, ab185620), anti-CDKN2A/p16INK4a (1:2000, Abcam, ab108349), anti-GSTP1 (1:1000, Santa Cruz Biotechnology, sc-66000), anti-RASSF1 (1:1000, Abcam, ab126764) and anti-α-tubulin antibody (1:15000, Sigma T5168). Alpha tubulin is a housekeeping protein that was used as a loading control. In the next day, after washing, the membrane was incubated with the appropriate horseradish peroxidase (HRP) –coupled secondary antibody, diluted in 5% blocking solution, for 1 hour at room temperature. The secondary antibodies used were: goat anti-Mouse-HRP conjugate (1:5000, Biorad, 170-6516) and goat anti-Rabbit-HRP conjugate (1:5000, Biorad, 170-6515). Protein detection was achieved using enhanced chemiluminescence substrates, either less sensitive (GE Healthcare Life Sciences, Product #RPN2209) or more sensitive (Thermo Fisher Scientific, Product #34096) depending on the abundancy of the target proteins.

### 2.5 CRISPR/Cas9 system

CRISPR/Cas9 (Clustered Regularly Interspaced Short Palindromic Repeats / CRISPR associated protein 9) system is an efficient and robust gene editing technology, adapted from the bacterial immune system, that enables precise genome manipulation through deletion, insertion or replacement of DNA. The editing capacity of this tool relies on the generation of a DSB and subsequent cellular repair process. In the natural system, a DNA endonuclease Cas9 binds to a complex formed by a CRISPR RNA (crRNA) and a transactivating crRNA (tracrRNA), that guides the enzyme to the complementary site of the crRNA. At the target site, Cas9 first recognizes a short nucleotide sequence, called protospacer adjacent motif (PAM) which is next to the target sequence, and then cleaves the double-stranded DNA (dsDNA).[58,59]

In the present project, to achieve the genomic modification desired (induction of a DSB in the promoter of the genes) it was used the plasmid LentiCRISPR v2 (Genscript) **(Figure 2.2)**, an all-in-one plasmid that encodes both guide RNA (gRNA) and Cas9 nuclease. It was performed a lentivirus-based transfection to deliver the CRISPR/Cas9 system to the cells.

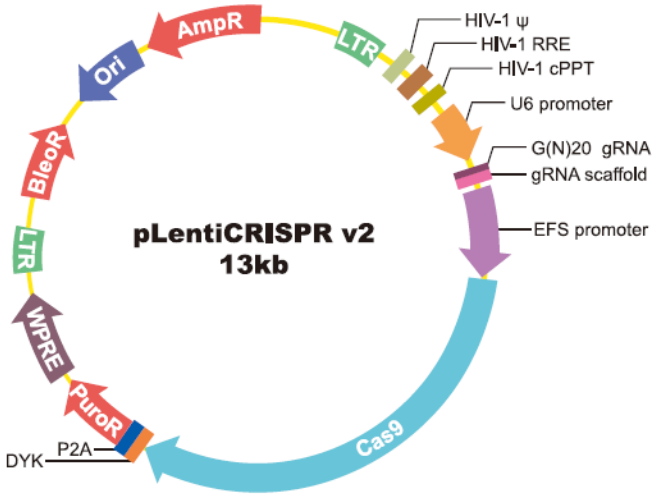


**Figure 2.2: Map of plentiCRISPR v2.** This plasmid contains a site for the introduction of a 20 nucleotides gRNA, which expression is under the control of the U6 promoter. Adjacent to the gRNA site there is a sequence corresponding to gRNA scaffold that is necessary for the Cas-binding (mimics the function of the tracrRNA in the endogenous system). The plasmid has also the Cas9 gene which is under the control of EFS promoter. Adapted from *CRISPR Handbook*.[50]

### 2.5.1 Single Guide RNAs design

To induce a DSB in the promoter of the studied genes using CRISPR/Cas9 system, it was designed single guide RNAs complementary to the region of interest, which will direct Cas9 specifically to the promoter of the candidate genes. Typically, the promoter region is located at the 5' end of genes and it comprises the transcription start site (TSS) of the gene and the surrounding sequences that function mainly as recognition sites for several regulatory factors. Promoters length range among 100 to 1000 bp.[60,61] Many features are used to get some insights about promoter location, namely the presence of "CpG islands", low-density DNA methylation sites enrich in CpG dinucleotides, once they are usually associated with gene promoters, and the presence of histone methylation modifications such as trimethylation of the lysine 4 (Lys4) of histone H3 (H3K4me3) once it is a specific chromatin signature displayed by active promoters.[15,61]

Therefore, using the previous information, and to guarantee that the target region was the promoter, a DNA sequence from a region containing CpG and H3K4me3, upstream or downstream of the first exon, depending on the orientation of each gene, was obtained from UCSC Genome Browser (https://genome.ucsc.edu/). This sequence was then used to design the guide RNA of the corresponding gene using the CRISPR Design tool of Massachusetts Institute of Technology (http://crispr.mit.edu/) **(Figure 2.3)**.

The gRNA sequences for each gene are presented in the Annex I **(Table 6.2)**. For *p14ARF*, it was used two different gRNAs that target different recognition sequences, one of them closer (gRNA1) to the first exon of the gene than the other (gRNA2), which will allow to evaluate if the induction of a DSB in different positions of the promoter can generate different outcomes.


**Figure 2.3: Methodology used to design the gRNAs.** After setting the distribution of H3K4me3 and RNA polymerase II (also used as a criterion to define the promoter region) and the localization of the CpG islands, a sequence from the region containing these three features (red box) was picked. The sequence was then introduced in CRISPR Design tool of MIT, and the best gRNA sequence was chosen from the options provided. Having the gRNA sequence, it was performed a blat to check the target genomic position (red circle). In this example, the design is directed to *p16INK4a* and for that we used the distributions from HepG2, a cell line that expresses the respective gene.

### 2.5.2 Lentivirus production

HEK293T cells, a highly transfectable derivative of HEK293 cell line, were maintained in DMEM supplemented with10% (v/v) FBS, 1% (v/v) 200 mM L-glutamine and 1% (v/v) penicillin-streptomycin,

in a 5% $CO_2$ humidified incubator at 37 ºC. The cells were plated on 100 mm culture dishes (TPP® Product #93100) coated with poly-L-lysine (Sigma, Product #P4832-50ML) and, when an 80-90% confluency was reached, HEK293T were transfected. Cell transfection was carried out using Lipofectamine™ 3000 Transfection Reagent (Invitrogen™, Product # LTI L3000015) that leverages on the lipid nanoparticle technology to delivery plasmid DNA into the cells. According to the transfection protocol, two solutions were prepared: solution A containing Opti-MEM, plasmid DNA and P3000 reagent, and solution B containing Opti-MEM and Lipofectamine 3000. The plasmids and the respective amount used were 6 µg of pLentiCRISPR v2 with the respective gRNA (GenScript), 6 µg of Δ8.9/P22 (from Neuroscience Group, IMM) and 2 µg of VSV-G/P23 (from Luis Moita's Lab, IMM). The plasmid Δ8.9/P22 encodes *gag*, *pol* and *rev* genes and VSV-G/P23 expresses an envelope glycoprotein (Vesicular Stomatitis Indiana Virus G protein, VSV-G) being, therefore, both plasmids essential for virus formation.

After preparing the solutions, they were incubated for 5 min at room temperature. Solution A was gently transferred to solution B and then the mixture was added dropwise to the cells. The cytopathic effect was evident after 48 h of the viral production. The medium, which contains the virus, was removed and filtered through a 0.45 µm syringe filter (VWR, Product #514-0063) into an ultracentrifuge tube (Beckman Coulter Life Sciences, Product #344059). Lentivirus concentration was performed via ultracentrifugation (3 h, 25000 rpm, 4 ºC), using Ultracentrifuge Beckman XL-90. The virus was then resuspended in PBS 1X, transferred to an Eppendorf and left at 4 ºC overnight to allow complete resuspension before transduction. When not immediately used, the viruses were stored at -80 °C.

### 2.5.3 Cells Transduction

The lentiviruses produced carrying the respective CRISPR plasmid with the guide RNA (pLenti*p16INK4*a v2, pLenti*p14ARF* gRNA1 v2, pLenti*p14ARF* gRNA2 v2, pLenti*RASSF1C* v2 or pLenti*GSTP1* v2) were then used to infect HCT116 and HepG2 cells. For that, the cells were plated on a 35 mm culture dish or on a 6-well plate (Corning®, Product # CORN3506) and, when a confluency of 50-60 % was reached, they were infected with 20 µL of virus. Finally, the cells were harvested 12 h, 18 h and 24 h after transduction, for RNA extraction (see section "2.2 RNA extraction"), and 6 h and 12 h post-transduction, for DNA extraction (see section "2.7 DNA extraction").

## 2.6 Bacterial Transformation

In order to have enough quantity of DNA for the transfection experiments, when necessary, the plasmids were re-generated by bacterial transformation. Competent *Escherichia coli* DH5α and *Escherichia coli* Stbl3 was used to amplified Δ8.9/P22 and VSV-G/P23, and pLentiCRISPR v2, respectively. After adding the desired plasmid to the competent bacteria, the samples were incubated 30 min on ice followed by an incubation at 42 ºC (45 s for Stbl3 and 2 min for DH5α). This heat shock enables the plasmids to be assimilated by the bacteria. After that, an outgrowth step (37 ºC in a shaking incubator, 220/220 rpm, for 60 min) was performed to allow bacteria time to generate the antibiotic resistance proteins encoded in the plasmid backbone. Then, the bacteria were plated in LB-agar medium supplemented with ampicillin, the antibiotic to which the plasmid confers bacterial resistance, and the plates were incubated overnight at 37 ºC. Next day, one colony was picked, inoculated in LB medium supplemented with antibiotic and shaken overnight at 37 ºC (220 rpm) to allow bacteria growth and, consequently, plasmid amplification. In the following day, intracellular plasmid DNA was isolated. For that, the bacterial suspension was centrifuged (20 min, 4000 rpm at 4 ºC) and, depending on the size of the pellet, it was used NZYMiniprep kit (Nzytech, Product #MB01001) or Genopure Plasmid Midi Kit (Sigma, Product # 03143414001) for the separation and purification of the plasmid DNA. These kits are based on the lysis of bacterial cells, removal of impurities and adsorption of DNA onto a column.

Therefore, bacterial cells were lysed, the lysate was filtered and subsequently loaded into a column with a material that has affinity to DNA. Then, the column was washed to remove all the impurities and the DNA was eluted. Isopropanol was added to the eluted DNA to precipitate the DNA, that was further collected through centrifugation, washed with 70% ethanol and finally, after air dried, resuspended in DNase/RNase free water. Before being used in transfection assays, the concentration of the plasmid DNA, and its purity, were measured using a spectrophotometer device (NanoDrop 2000, Thermo Fisher Scientific).

### 2.7 DNA Extraction

To evaluate the efficacy of the CRISPR/Cas9 system, in other words the capacity of inducing a DSB, gDNA was extracted from HCT116 and HepG2 cells using phenol/chloroform extraction. This liquid-liquid extraction consists in the separation of molecules based on their different solubilities in two different immiscible allowing, therefore, the elimination of proteins and lipids and the purification of nucleic acids.[62]

The cells were plated and harvested in PBS 1X with the help of a cell scrapper. Cellular suspension was then centrifuged (4000 rpm, 5 min, 4 ºC) and, to the pellet formed, it was added tail lysis buffer composed by 100 mM NaCl, 10 mM Tris-HCl (pH 8.0), 25 mM EDTA (pH 8.0), 0.5 % SDS and proteinase K (Nzytech, product #MB01902), to lyse the cells. The samples were incubated overnight at 56 °C. In the next day, 3 M sodium acetate and phenol:chloroform:isoamyl alcohol were added to the samples which, after a high-speed centrifugation (15000 rpm, 10 min, 4 ºC), allowed the formation of mainly two phases: an upper aqueous phase containing the DNA and a lower organic phase containing the proteins and the lipids. The upper phase was transferred to another tube, chloroform was added and once again the upper phase formed was collected to a new tube, after centrifugation (15000 rpm, 10 min, 4 ºC). Isopropanol (and glycoblue) was added to the tube and the samples were kept at -20 ºC overnight, to precipitate the DNA. Once precipitated, the DNA pellet was collected through high-speed centrifugation (1500 rpm, 30 min, 4 ºC), washed twice with 70% ethanol and, after air dried, resuspended in DNase/RNase free water.

Regarding the viscosity of the samples, associated to the non-fragmented gDNA, the samples were digested using restriction enzymes. A mix with 10x FastDigest Buffer (Thermo Fisher Scientific) and 20U each enzyme: EcoRI (Thermo Fisher Scientific, Product #FD0274), BamHI (Thermo Fisher Scientific, Product # FD0054), HindII (Thermo Fisher Scientific, Product #FD0504), and in some cases XhoI (Thermo Fisher Scientific, Product # FD0694) and Bsp1407I (Thermo Fisher Scientific, Product # FD0934) was added to the samples which were, then, incubated overnight at 37 °C, with rotation. In the next day, phenol:chroloform extraction of the nucleic acids, similar to previously described, was performed (in this new extraction the precipitation step overnight was not performed). Finally, the samples were quantified using NanoDrop 2000.

### 2.8 Quantitative PCR (qPCR)

qPCR is a molecular biological technique that allows to quantify a target DNA molecule. When associated with RT, reverse transcription quantitative polymerase chain reaction (RT-qPCR) allows the quantification of a specific transcript in a sample being, therefore, a widely used method for gene expression analysis.[63] This technique is based on the detection and quantification of fluorescence associated with the accumulation of the newly amplified DNA, after each amplification cycle. The fluorescent signal produced accumulates in a directly proportional manner to amplicon generation, allowing to monitor, in real-time, the complete DNA amplification process.[50] The cycle in which the fluorescent signal overcomes a specific detection threshold above background fluorescence is defined

as threshold cycle ($C_t$), and this value is used to calculate the absolute or relative abundance of the starting template.[64]

In RT-qPCR, the relative quantification assay describes the changes in gene expression in a given sample relative to reference sample.[65] As this assay is a comparative quantification, normalization of the target gene expression against an internal control (a housekeeping gene), to compensate for variations in reaction efficiency, pipetting errors, or differences in the starting quantity of the template, is essential for accurate quantification. To calculate the relative changes in expression, $2^{-\Delta\Delta Ct}$ method, a comparative Ct (between the study sample and the refence sample) method was used because the efficiencies obtained in the standard curves between the different primers used for the target genes and the endogenous gene primers were approximate.[50,65]

Regarding that, in HCT116 cells, *CDKN2A* has one wildtype allele, epigenetically silenced by promoter hypermethylation and one mutant allele, transcriptionally active, that codes for a truncated protein[66,67], to evaluate the expression of *p16INK4a* and *p14ARF* in HCT116 cells before and after the induction of a DSB in the promoter of each gene, a RT-qPCR was performed. This technique was also used to evaluate the expression of *RASSF1C* and *GSTP1* genes after the damage. Additionally, qPCR was used to confirm if the CRISPR/Cas9 system was really inducing a DSB promoter-specific. For this task, it was designed primers flanking the genomic region targeted by the gRNAs **(Figure 6.1, Annex I)**. As a negative control for the expression experiments, HCT116 and HepG2 were transduced with a CRISPR/Cas9 plasmid containing a scrambled gRNA.

Depending on the purpose, cDNA (1:15 dilution) or gDNA (25-100 ng/µL) were used as template for the real time quantitative PCR. The qPCR reactions were performed using ViiA 7 Real-Time PCR System (Thermo Fisher Scientific) and were carried out on a 384-well reaction plate (ThermoFisher Scientific, Product # 4309849) using the iTaq™ Universal SYBR® Green Supermix (Bio-Rad, Product # 1725122) and the respective pair of primers. For each reaction of 10 µL, a mix with 3,6 µL of SYBR® Green Supermix, and 0,7 µL of each forward and reverse primers was added to 5 µL of template. The conditions used were the following: one cycle at 50 ºC for 2 min and 95 ºC for 10 min and 40 cycles (30 for *GSTP1* gDNA) at 95 ºC for 15 s and the respective annealing temperature for each pair of primers for 1 min. It was performed a melt curve stage: 15 s at 95 ºC, 1 min at 60 ºC and 5 min at 95 ºC. All samples were normalized against *GAPDH* and the relative expression was evaluated by $2^{-\Delta\Delta Ct}$ method. Each reaction was carried out in triplicates and for the analysis, when the difference between the technical replicates was higher than 0.5, it was used only the best two replicates.

The primer's sequences used for gene expression experiments and for gDNA are presented in Annex I. **(Table 6.1 and 6.3)**. *GAPDH* primers, for cDNA and gDNA, were used as a loading control. Note that the primers used for *p14ARF* gDNA region were used to evaluate the damage induced by both gRNAs.

### 2.9 Statistical Analysis

To better interpret the changes in gene expression after the induction of a DSB promoter-specific, a statistical analysis was applied. The samples harvested at 12 h, 18 h and 24 h post-transduction were compared with the non-infected samples to check, indeed, if the expression of the genes increase after the damage. For that, when indicated, statistical significance was resolved using a two-tailed Student's t-test. The same test was also applied to assess significant changes in gene transcription of *CDKN2A* between HCT116 and HEK293 cells, and to evaluate changes in DNA levels induced by the CRISPR/Cas9 system.

# 3. Results and discussion

## 3.1 Expression of tumour suppressors and oncogenes: confirmation of *CDKN2A* and *GSTP1* silencing and *RASSF1C* expression

Cancer is a set of diseases associated with the accumulation of numerous and heterogeneous molecular abnormalities, which consequently lead to deregulation of several signalling pathways.[4] An imbalance between the expression of oncogenes and TSGs, the two main types of genes involved in oncogenesis, seems to be one of the main sources for all these functional changes.[21]

Epigenetic transcriptional silencing by aberrant CpGIs methylation is a well-known mechanism of TSGs suppression in cancer.[15] Inactivation of *CDKN2A* is an example of that.[15] Located in the short arm of chromosome 9 at position 21.3, this gene contains two introns and three exons and encodes two different proteins, p16INK4a and p14ARF.[14,68] Although expressed from the same gene, p16INK4a and p14ARF have no amino acid sequence homology since they are synthetized by using distinct first exons, each driven from separated promoters **(Figure 6.2A, Annex II)**.[14,68–70] p16INK4a protein is a negative regulator of the cell cycle and a crucial player in senescence. These functions are mediated through the binding of p16 to the cyclin-dependent kinase (CDK) 4/6, which induces a conformational change in these proteins and, consequently, inhibits the formation of the complex G1-CDK, essential for cell cycle progression **(Figure 6.2B, Annex II)**.[14,67] p14ARF is also involved in the control of cell cycle progression, in this case, via p53 pathway. Its capacity to interact with the p53 regulatory protein, murine double minute 2 (MDM2), an ubiquitin ligase, blocks MDM2-mediated degradation of p53, which consequently stimulates p53-mediated expression of stress-response genes **(Figure 6.2B, Annex II)**.[67,68,70,71] The disruption of this pathway due to the inactivation of *p14ARF* results in uncontrolled cell proliferation.[67]

It has been proved that the loss of *CDKN2A* expression is involved in the pathogenesis of many cancers and it is known that gene inactivation results from several genetic and epigenetic aberrations, such as promoter hypermethylation (DNA or histone methylation), point mutations and homozygous deletion.[14,67] One of the cancers where *CDKN2A* is commonly inactivated is colorectal cancer. In this type of cancer, namely in HCT116, a colon cancer cell line, and in some primary colon carcinomas samples, it was demonstrated that *CDKN2A* inactivation occurs due to a point mutation in one allele, which gives rise to a product with no biological function, along with the methylation of the CpGIs present in the promoter of the wild-type allele.[66,67] Indeed, studies have reported that *p16INKa* and *p14ARF* genes are heterozygously methylated.[67] This is a clear example that satisfied Knudson's hypothesis.

For the sake of this project, we went to confirm if *CDKN2A* gene is inactivated in HCT116 cells, as referred in the literature. First, we evaluated the presence of the respective proteins translated from each isoform, *p16INK4a* and *p14ARF*. For that, total protein extracts were obtained from HCT116 cells, and a western blot was performed using protein-specific antibodies. The results obtained show that p16INK4a and p14ARF proteins are absent from these cells. (**Figure 3.1A**). Then, and to evaluate the transcription of the genes, *p16INK4a* and *p14ARF* cDNA was produced by reverse transcription and was used as a template in qPCR. At this point, it is important to consider that both *p16INK4a* and *p14ARF* genes have one mutant allele that is transcriptionally active, which means that the allele produces a mRNA.[67] Therefore, in the pull of RNAs collected from HCT116, we will always have the contribution of the mutated allele. To assess if the remaining allele is silenced, in agreement to previous works, we compared the expression levels of *p16INK4a* and *p14ARF* between HCT116 and HEK293, a cell line that expresses the genes in study.[72] We are expecting a decrease in the expression levels of the genes in HCT116 cells comparing to HEK293 cells. The results demonstrate the decrease expected (**Figure 3.1B**), which allows us to conclude that the wild-type allele is inactive.

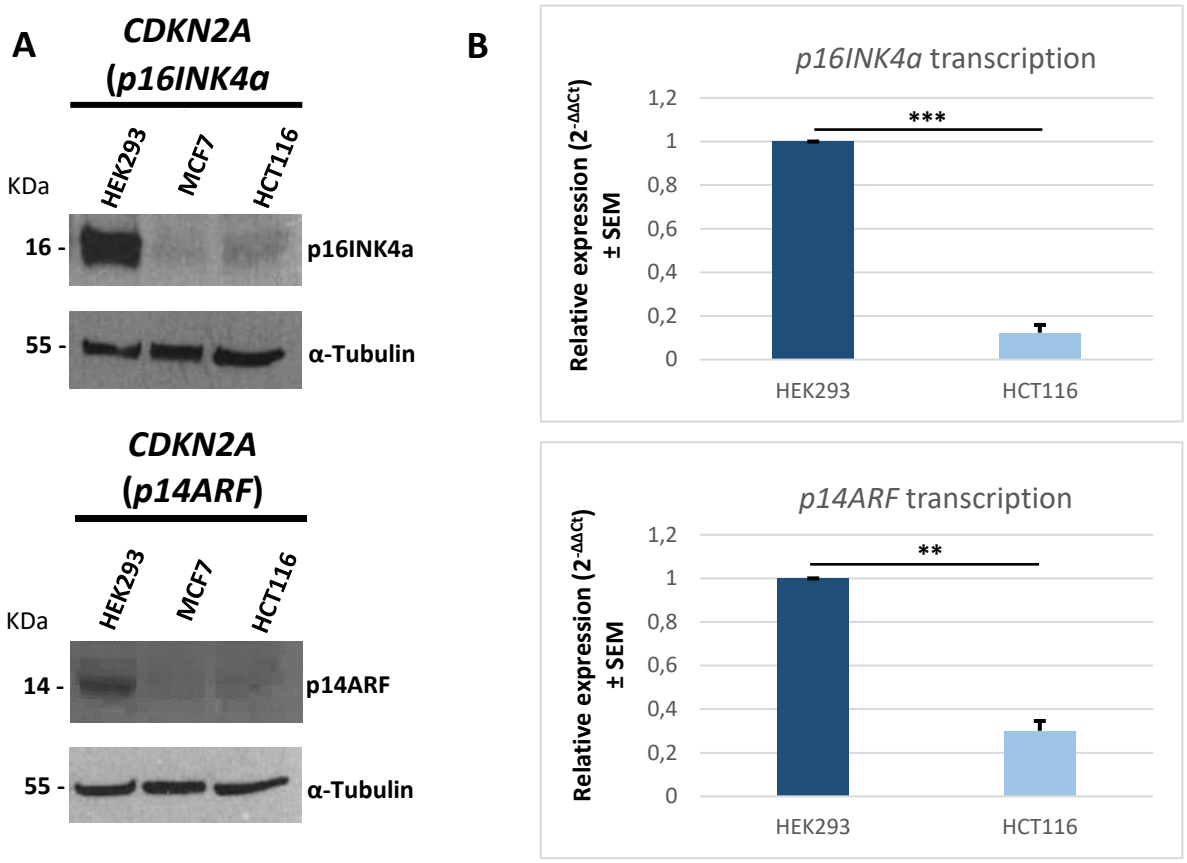


**Figure 3.1: *p16INK4a* and *p14ARF* are silenced in HCT116 cells. (A)** Total protein extracts of HEK293, MCF7 and HCT116 cells were analyzed by western blot, with antibodies against p16INK4a, p14ARF and α-tubulin proteins, as indicated on the right. α-tubulin was used as a loading control. Molecular weight (kDa) markers are shown on the left. The presence of p16INK4a and p14ARF is only detected (strong signal) in HEK293 cells, used as a positive control. In MCF7 and HCT116 cells the referred proteins are absent. MCF7 cells was used as a negative control. Each figure represents the result of one experiment from three independent experiments with similar results. **(B)** Total RNA from HEK293 and HCT116 cells were isolated and converted in cDNA by reverse transcription. The cDNA produced was used as a template in qPCR. The relative expression levels of *p16INK4a* and *p14ARF* were calculated using $2^{-\Delta\Delta Ct}$ method, and these levels were normalized against *GAPDH* (endogenous control). The decrease observed in the transcription of the genes in HCT116 comparing to HEK293 cells indicates the inactivation of *p16INK4a* and *p14ARF* in HCT116. The results shown are the mean of three independent experiments. Error bars denote the standard error of the mean (SEM). *** Student's t-test, p-value = 0,0008. ** Student's t-test, p-value = 0,002.

However, it should be referred that this method is not the best approach to evaluate if *p16INK4a* and *p14ARF* genes are silenced, because different cell lines can have different pathways to control gene expression which, consequently, lead, by itself, to different gene expression patterns. Thus, the decrease verified can be a consequence of different networks of gene expression regulation operating in each cell line and not be related to gene silencing. Therefore, to overcome this situation other techniques should be applied. DNA sequencing, after RT-PCR using specific primers for each gene, and the ensuing analysis allows to verify if the product amplified has the mutations described. The presence of the mutation in all the fragments obtained would allow to conclude that only the mutated allele is producing transcripts, indicating the silencing of the other allele. An alternative method available for the same purpose is RFLP, restriction fragment length polymorphism. This technique is used to investigate differences between homologous DNA sequences based on the presence or absence of restriction enzyme sites.[73,74] Point mutations or polymorphisms occur all over the human genome and these modifications can create or abolish recognition sites for restriction endonucleases. Consequently, after samples digestion, with specific restriction enzymes that recognize and cut a particular site, there will be produced fragments of different lengths. This generates a unique restriction pattern characteristic to a specific genotype, upon separation by gel electrophoresis, that can be detected using probes.[73–75] In

the case of *CDKN2A* gene, if any of the mutations described affects the existent landscape of restriction sites, the restriction patterns that will be obtained if we have a mutated allele and the other silenced or a mutated allele and the other active or even if we do not have any mutation will differ. In other words, if the mutation creates a restriction site, we will see two fragments if one allele is mutated and the other is silenced or if both are mutated, three fragments if one allele is mutated and the other is active and just one fragment if there is no mutation, regardless of whether they have an allele silenced.

The active research developed over the past years revealed that several TSGs are inactivated in the context of tumorigenesis and the number has been growing.[22] One of the TSGs commonly affected in many cancers is *GSTP1*. This gene, located in the long arm of chromosome 11 at position 13.2, encodes a subclass of the glutathione S-transferases (GSTs) superfamily named π class GST, which is involved in the detoxification of endogenous and exogenous substances, protecting cells against damage mediated by cytotoxic and carcinogenic agents and oxidative stress [53,76,77] Besides being involved in phase II metabolism, GSTP1 is considered a major player of cell signalling, as it modulates several pathways and cell proliferation in response to stress, hypoxia and other stimuli.[78,79]

Gene polymorphisms or hypermethylation are the main alterations reported to interfere with *GSTP1* expression in cancer.[77] Indeed, a compelling number of studies demonstrated that the large CpG-rich region present in the promoter region of the gene is frequently hypermethylated in several cancers, including in human hepatocellular carcinoma, thereby leading to transcriptional repression.[53,76,79] In the present work, in order to confirm *GSTP1* inactivation in HCC, a western blot and a RT-qPCR were performed, to check the presence of the protein and the transcript, respectively. The results obtained corroborate with gene inactivation, since neither protein **(Figure 3.2A)** nor transcript **(Figure 3.2B)** is detected in HepG2 cell line.

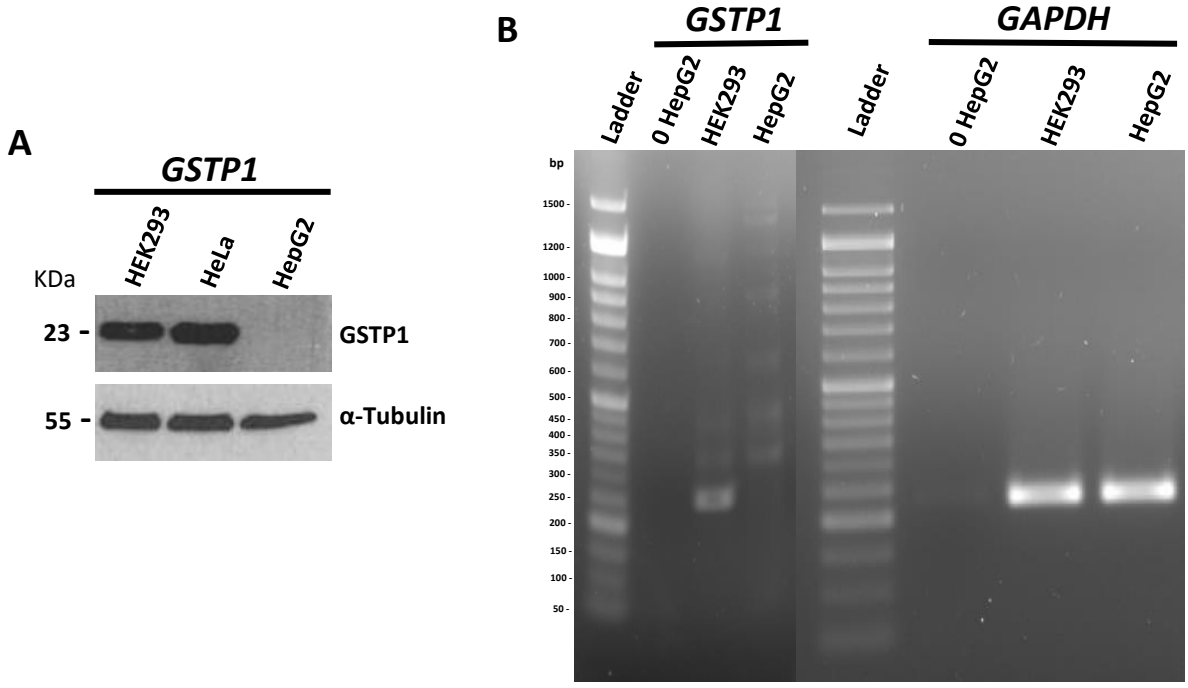


**Figure 3.2: *GSTP1* is silenced in HepG2 cells. (A)** Total protein extracts of HEK293, HeLa and HepG2 cells were analyzed by western blot, with antibodies against GSTP1 and α-tubulin proteins, as indicated on the right. α-tubulin was used as a loading control. Molecular weight (kDa) markers are shown on the left. The presence GSTP1 is only detected (strong signal) in HEK293 and HeLa cells, used as a positive control. In HepG2 cells the referred protein is absent. Each figure represents the result of one experiment from three independent experiments with similar results. **(B)** Total RNA from HEK293 and HepG2 cells were isolated and converted in cDNA by reverse transcription. The cDNA produced was used as a template in PCR, and the product amplified was resolved in a 2% agarose gel. (0) represents the negative control where no Superscript was used. In HEK293, a band with about the expected size of 250bp was detected; however, no band was detected in HepG2 cells which indicates no transcript. *GAPDH* was used as a loading control. The figure represents the result of one experiment from three independent experiments with similar results.

Until now, we have noticed that tumour suppressor loss is associated with cancer susceptibility. However, many other abnormalities can be drivers of tumorigenesis, namely oncogene overexpression.[20] Ras association domain family 1 (RASSF1) is a group of Ras effectors that have a direct role in human cancer.[22,80] Located on the chromosome 3p21.3, the *RASSF1* gene has eight exons that, through alternative splicing and differential promoter usage, produces seven different transcripts (*RASSF1A-G*). Isoform C is one of the major isoforms transcribed, along with isoform A. These two isoforms have different CpGIs promoters and are ubiquitously expressed in tissues.[22,81] All the isoforms encoded are composed by distinct domains and the presence or absence of any of these structural and functional units dictates the function of each isoform **(Figure 6.3, Annex II)**.[22,81,82] It was reported that *RASSF1C* plays a role in tumorigenesis acting as an oncogene, since it promotes cell proliferation and migration and attenuates apoptosis.[80,82–84]

*RASSF1C* is a well express isoform in some human solid tumours, including in HCC.[54,81,85] In this project, HepG2 cells were used to confirm this statement and *RASSF1C* expression was assessed at transcript and protein levels. The results obtained by western blot depict the presence of the protein in the study cells **(Figure 3.3A)**. Additionally, RT-PCR experiments show the presence of *RASSF1C* transcripts, as they are detected in the agarose gel **(Figure 3.3B)**.

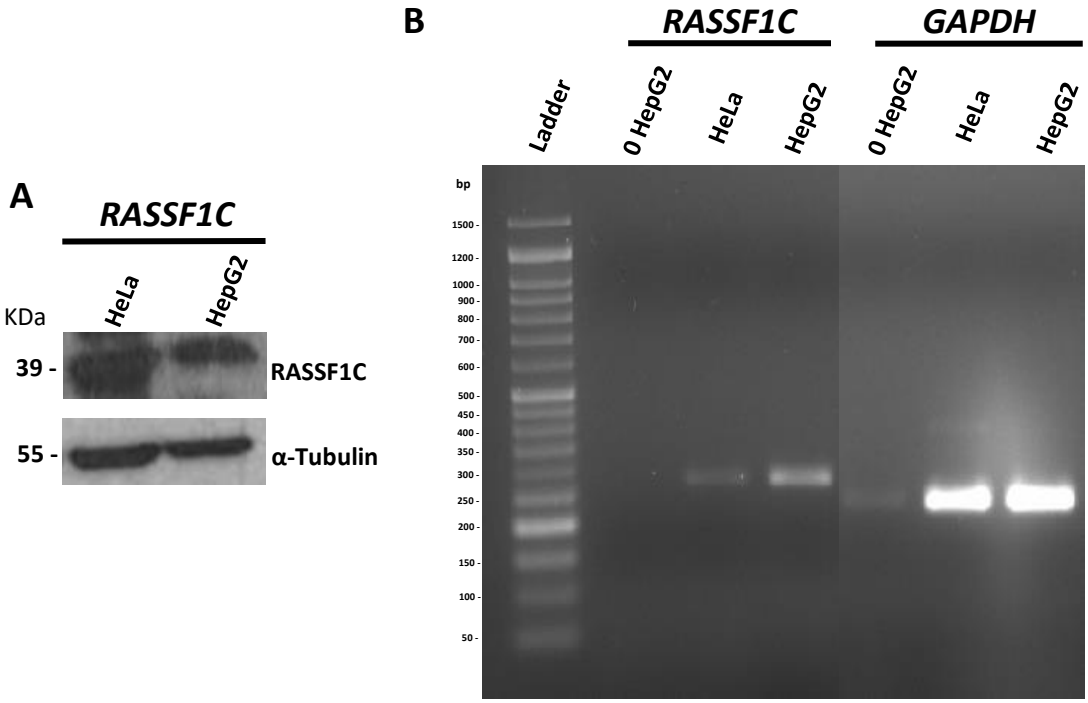


**Figure 3.3:** ***RASSF1C*** **is expressed in HepG2 cells. (A)** Total protein extracts of HeLa and HepG2 cells were analyzed by western blot, with antibodies against RASSF1 and α-tubulin proteins, as indicated on the right. α-tubulin was used as a loading control. Molecular weight (kDa) markers are shown on the left. The presence of RASSF1C is detected HeLa cells, used as a positive control, and in HepG2. Each figure represents the result of one experiment from three independent experiments with similar results. **(B)** Total RNA from HeLa and HepG2 cells were isolated and converted in cDNA by reverse transcription. The cDNA produced was used as a template in PCR, and the product amplified was resolved in a 2% agarose gel. (0) represents the negative control where no Superscript was used. In HeLa and in HepG2, a band with about the expected size of 272bp was detected, which indicates the presence of transcript. *GAPDH* was used as a loading control. The figure represents the result of one experiment from three independent experiments with similar results.

## 3.2 Induction of a DSB locus-specific using CRISPR/Cas9 system

The goal of this project is to understand whether the introduction of a DSB in the promoter of a gene regulates its activation. For that, it was used the CRISPR/Cas9 system, the most widely used tool in genome reprogramming[59], with gRNAs complementary to the promoter region of the selected genes.

One of the crucial tasks that need to be performed, before answering the previous question, is to check if the system is effective. Therefore, to assess if CRISPR/Cas9 system is inducing a locus-specific DSB, it was designed primers that flank the genomic region targeted by each gRNA **(Figure 6.1, Annex I)**. DNA extracts were collected from non-infected and infected cells, the last at specific timepoints after transduction (6 h and 12 h), defined regarding the time that the system takes to be established (entry of the plasmids, expression of Cas9, …). DNA levels were evaluated through qPCR using the previous primers. We anticipate a decrease in DNA levels in the infected samples compared to the uninfected samples, since the induction of a break will prevent the amplification of the fragments by the primers. The results obtained demonstrate that the applied system is working, because we can see a decrease of the DNA levels in all the samples infected with the respective gRNA **(Figure 3.4 and Figure 3.5)**.

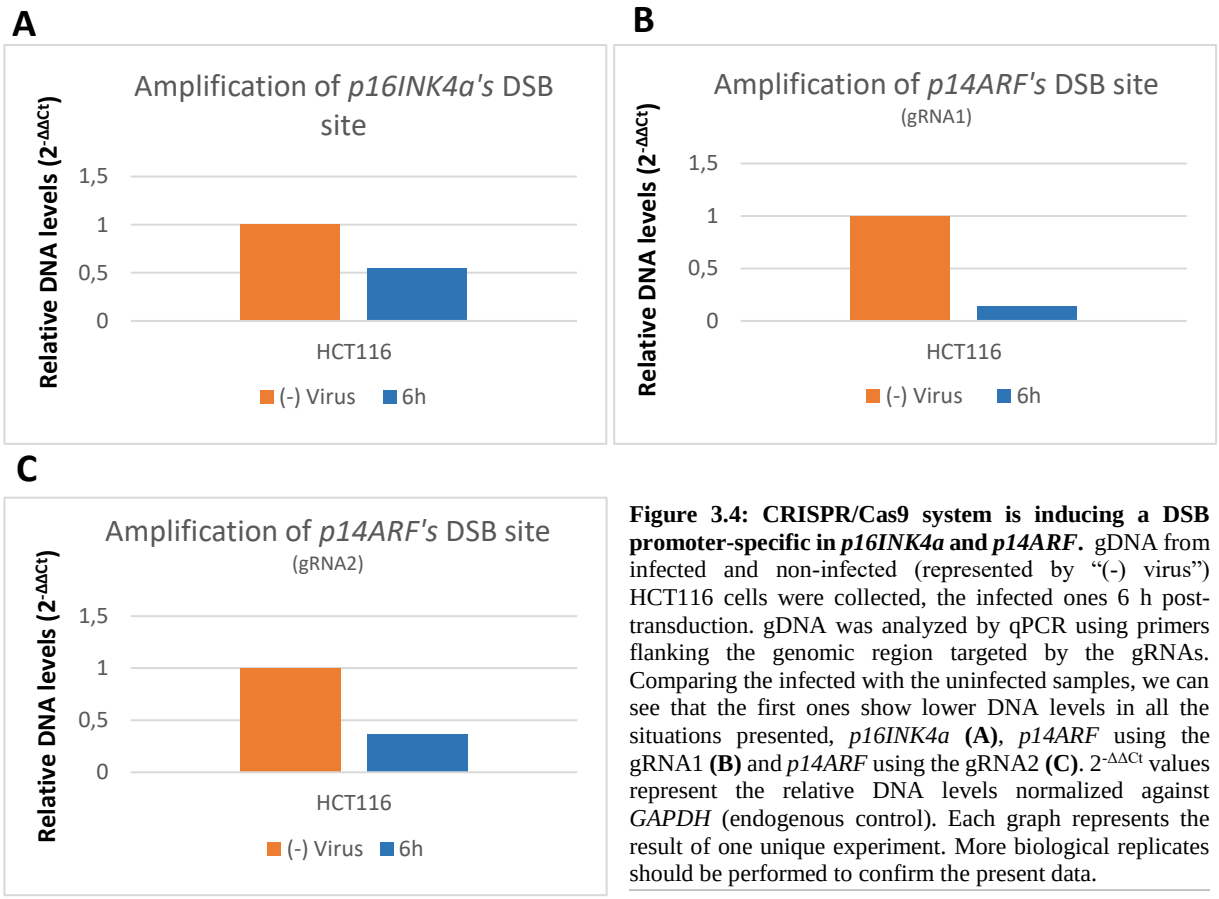


**Figure 3.4: CRISPR/Cas9 system is inducing a DSB promoter-specific in *p16INK4a* and *p14ARF*.** gDNA from infected and non-infected (represented by "(-) virus") HCT116 cells were collected, the infected ones 6 h post-transduction. gDNA was analyzed by qPCR using primers flanking the genomic region targeted by the gRNAs. Comparing the infected with the uninfected samples, we can see that the first ones show lower DNA levels in all the situations presented, *p16INK4a* **(A)**, *p14ARF* using the gRNA1 **(B)** and *p14ARF* using the gRNA2 **(C)**. $2^{-\Delta\Delta Ct}$ values represent the relative DNA levels normalized against *GAPDH* (endogenous control). Each graph represents the result of one unique experiment. More biological replicates should be performed to confirm the present data.

It is possible to notice that the efficiency of the CRISPR/Cas9 system is not the same in all the cases. The cleavage efficiency is higher using the gRNA1 for *p14ARF*, where we see more than a half-fold change in DNA levels between non-infected and infected cells **(Figure 3.4B),** than when it is used the gRNA for *16INK4a* **(Figure 3.4A)** or even the other gRNA for *p14ARF* **(Figure 3.4C)**. This can be due to many factors, namely to the presence of mutations close to the 5' end of the target sequence, and consequently to the number and positions of mismatches tolerated by the gRNA at the target site, and to the genomic and epigenomic landscape present at the genomic targeted segment that can affect site accessibility.[86,87]

The use of primers that flank the genomic target site and the evaluation of the DNA levels by qPCR, is a rapid and easy way to confirm the induction of a DSB promoter-specific; however, it isn't the most efficient and sensitive approach. There are several methods that can be applied to assess on-target DSB and all of them are based on the PCR. Among them, we can refer the mismatch cleavage assays, the high-resolution melting analysis, the heteroduplex mobility assays and others. All these

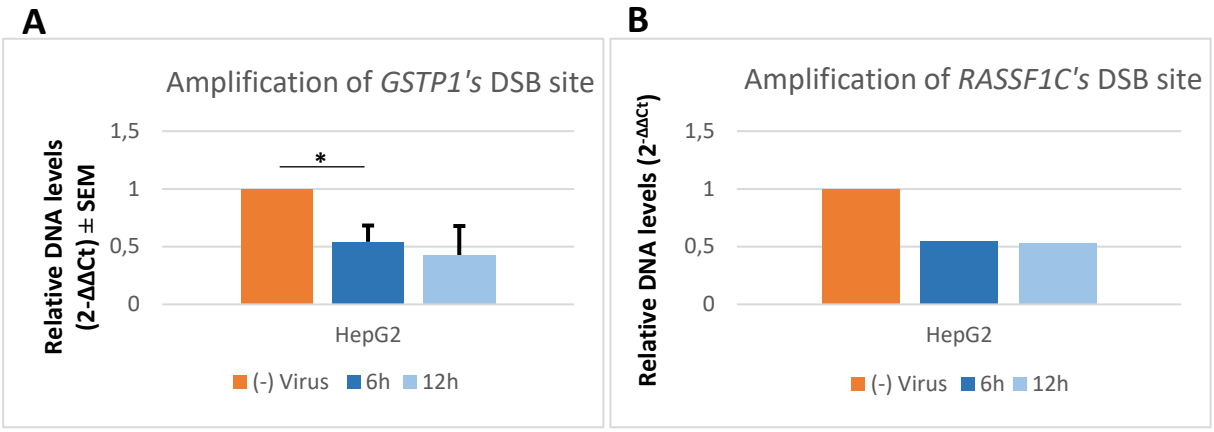


**Figure 3.5: CRISPR/Cas9 system is inducing a DSB promoter-specific in *GSTP1* and *RASSF1C*.** gDNA from infected and non-infected (represented by "(-) virus") HepG2 cells were collected, the infected ones 6 h and 12 h post-transduction. gDNA was analyzed by qPCR using primers flanking the genomic region targeted by the gRNAs. Comparing the infected with the uninfected samples, we can see that the first ones show lower DNA levels in all the situations presented, *GSTP1* **(A)** *RASSF1C* **(B)**. $2^{-\Delta\Delta Ct}$ values represent the relative DNA levels normalized against *GAPDH* (endogenous control). The graph on the right are the mean of three independent experiments and error bars denote the standard error of the mean (SEM). The graph on the left represents the result of one unique experiment and, therefore, more biological replicates should be performed to confirm the present data. * Student's t-test, p-value = 0,04.

methods have strengths but also weaknesses and the ideal method should be chosen depending on the genome editing experiment, namely the type of sample and the size and frequency of the mutations. Sequencing-based detection methods seem to be an efficient approach as they are highly informative about the nature and diversity of the mutations.[88]

### 3.3 Transcription of the genes after the induction of a DSB in their promoters: reactivation of the transcription of *GSTP1* and *p14ARF* and decrease in *RASSF1C* transcription

In the recent years, growing evidences have demonstrated a positive relationship between DNA damage and transcription activation. In fact, it was verified that the induction of some transcriptional programs involves the formation of DSBs, the most catastrophic form of damage.[23,35]

In order to investigate whether the introduction of a DSB in the promoter of a gene regulates its activation, we transduced HCT116 and HepG2 cells with lentivirus carrying the CRISPR/Cas9 system with gRNAs directed to the promoter region of the genes. The selected genes were *p16INK4a* and *p14ARF*, and *GSTP1*, already proved to be silenced in HCT116 and HepG2 cells, respectively, and *RASSF1C*, whose expression was confirmed in HepG2 cells. The expression levels of each gene were evaluated by RT-qPCR, using the total RNA extracts collected from non-infected cells and infected cells, the latter at different timepoints, namely 12 h, 18 h and 24 h.

In the case of *GSTP1*, after inducing a promoter-specific DSB, we see a statistically significant increase in the transcription of the gene, 12 h post-transduction **(Figure 3.6A)**. This increase is maintained in the following timepoint, comparing to the control; however, the expression levels suffer some changes, as we have a decrease between the 18 h and 12 h timepoints **(Figure 3.6A)**. A possible explanation for this decrease can be the induction of a new DSB mediated by CRISPR/Cas9 system, since it is believed that, upon Cas9 expression, the enzyme will constantly cut the target sequence. Thus, a new damage would affect the transcription of the gene. Nonetheless, this explanation has some faults because it is known that the repair of a damage induced by CRISPR/Cas system relies mainly in NHEJ, an error-prone pathway that is associated with the introduction of random insertion/deletions mutations at the site of repair.[59,86,88] Therefore, the sequence of the target region gets modified after damage which consequently abrogates the recognition by the gRNA and the induction of a cut. A deeper analysis, using

additional methods, needs to be performed to understand all the changes verified. For example, the use methylation assays will allow to evaluate the methylation status of *GSTP1* that we know to be silenced through promoter hypermethylation.[77] Understanding how the methylation landscape of the gene changes in the context of a DNA damage can help to explain the previous changes that were visualized. In fact, the increase in *GSTP1* transcription, after the induction of a DSB, may be due to the damage-dependent loss of the hypermethylated pattern associated to the promoter before the damage, as it was already reported in some studies.[21] On the other hand, this increase can be seen as only a boost in transcription mediated by DNA damage, which we know to be associated with chromatin remodelling events and the formation of nucleosome-free region that facilitates RNAPII binding and transcription initiation.[23,46] The possible reestablishment of the methylation pattern upon DNA repair can justify the decrease seen after 18 h **(Figure 3.6A)**; to notice that, if this is the case, it is normal that the expression levels at 18 h timepoint do not reach the same level as the control, because we are working with a pull of cells and therefore we have heterogeneity associated to the cellular population (not all the cells are at the same point in the cell cycle, which affect the course of DNA repair). The infection of the cells with a scrambled gRNA didn't result in any significant changes in *GSTP1* transcription **(Figure 3.6B)** which lead us to conclude that the increased transcription observed was directly due to the DNA damage induced in the promoter and not a result of other factors imposed by system itself.

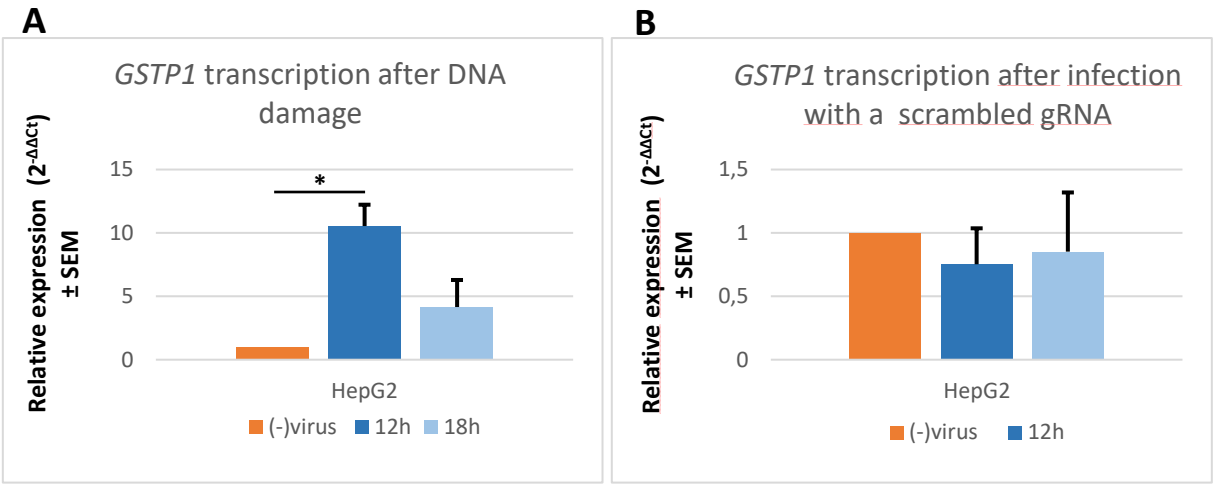


**Figure 3.6: Promoter-specific DNA damage results in the reactivation of *GSTP1* gene.** Total RNA from non-infected (represented by "(-) virus") and infected HepG2 cells, were isolated and converted in cDNA by reverse transcription. The infection was carried out using CRISPR/Cas9 system with the specific gRNA for *GSTP1* promoter region **(A)** or with a scrambled gRNA **(B)** and the RNA from the infected cells was collected 12 h and 18 h post-transduction. The cDNA produced was used as a template in qPCR. The relative expression levels of *GSTP1* were calculated using $2^{-\Delta\Delta Ct}$ method, and these levels were normalized against *GAPDH* (endogenous control). **(A)** It is observed an increase in gene transcription 12 h after the infection, when compared to the non-infected samples, which indicate the reactivation of the gene. **(B)** No relevant changes in gene trancription are observed, at any of the timepoints, using a scrambled gRNA. The results shown are the mean of three independent experiments. Error bars denote the standard error of the mean (SEM). The 24 h timepoint is not presented because the approach didn't work. * Student's t-test, p-value = 0,015.

Regarding *p14ARF*, it is verified that the induction of a DSB in the gene promoter, using the gRNA1, leads to an increase in gene transcription at the timepoints selected, with an almost 2-fold change increase after 18 h, compared to the control. **(Figure 3.7A)** As *p14ARF* is, supposedly, composed by an active and a silent allele, the increase seen indicates the transcription activation of the allele that was previously silenced by hypermethylation. After 24 h, the transcription of the gene decreases **(Figure 3.7A)**, however this decrease is not significant, and it can be explained by the heterogeneity associated with the experiments. The damage in the promoter of the gene was also induced using a different gRNA (gRNA2), but, in this case, the changes observed were not statistically significant (**Figure 3.7B**). Nonetheless, despite the lack of statistical significance, it seems that the expression levels show a

tendency to increase upon damage. More biological replicates should be performed to clarify this situation. The analysis of the results obtained from gRNA1 and gRNA2 demonstrates that using the same approach, but different gRNAs, leads to different outcomes. The differences observed between the gRNA1 and gRNA2 can rely on the location of the target sequences. Therefore, as they are targeting different regions, we supposed that gRNA2 can be interfering with the ligation of some transcription factors, preventing the assembly of the transcription complex. This may explain the differences in the results obtained, namely the fact that the damage induced using gRNA1 results in an evident increase in *p14ARF* transcription, contrarily to the damage induced using gRNA2, where we do not have such clear results.

Until here, the data analysed indicate that the reactivation of *p14ARF* transcription is due directly to the damage induced in the promoter; although many other possibilities need to be considered. One of the tumour suppressive functions of *p14ARF* protein is the induction of senescence mediated by p53.[89] Thus, we may suppose that the increase of *p14ARF* transcription can be the response of the cell to DNA damage. After recognizing the damage induced by CRISPR/Cas9 system, the cell can promote the coordinated transcriptional activation of specific sets of genes and trigger signalling pathways to induce senescence, which is an outcome of *p14ARF* activation. Obviously, it is known that the induction of a single DSB didn't trigger senescence. However, some studies reported that DNA breaks created by CRISPR/Cas system using single gRNAs can result into large indels, over many kilobases, and more

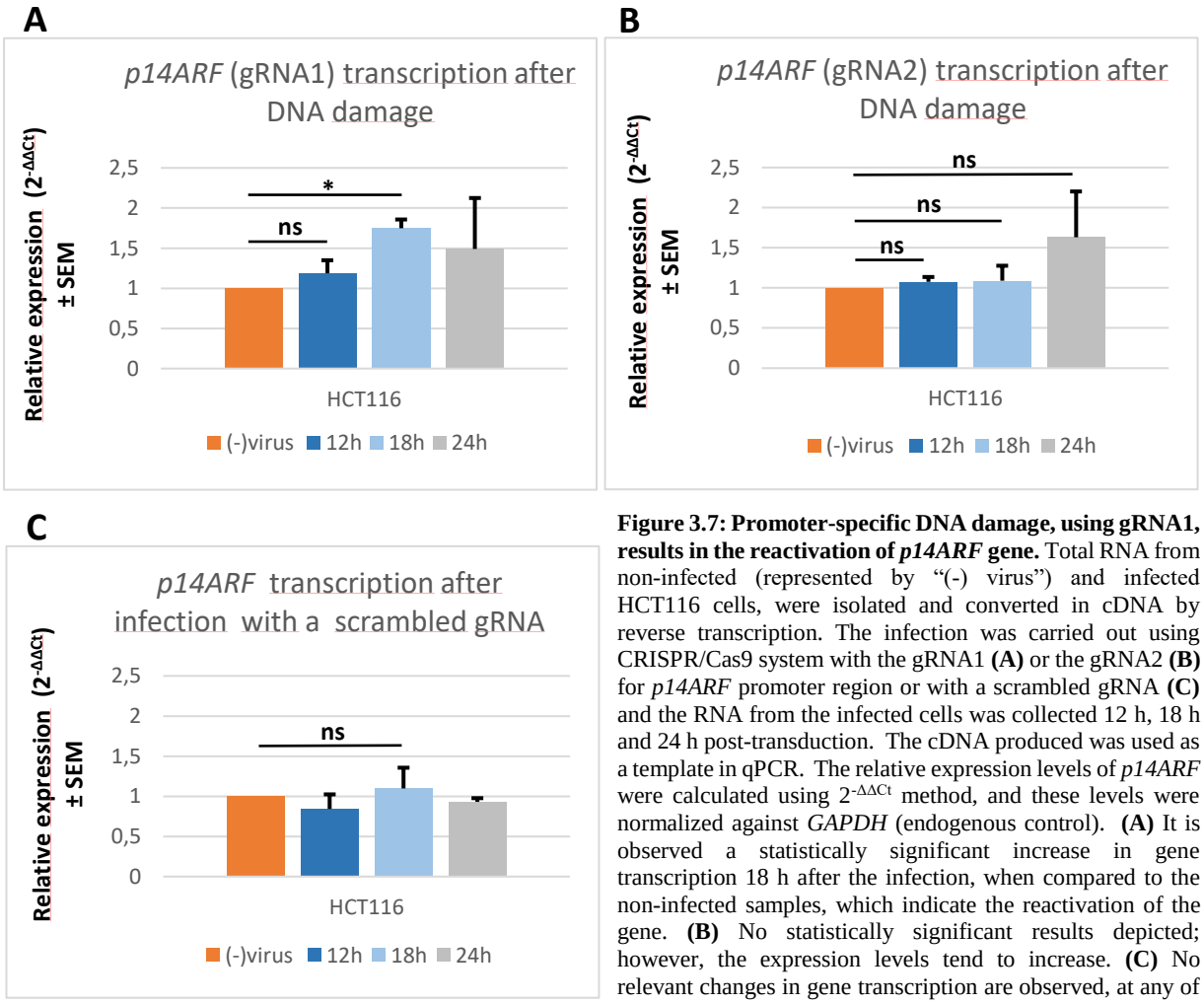


**Figure 3.7: Promoter-specific DNA damage, using gRNA1, results in the reactivation of *p14ARF* gene.** Total RNA from non-infected (represented by "(-) virus") and infected HCT116 cells, were isolated and converted in cDNA by reverse transcription. The infection was carried out using CRISPR/Cas9 system with the gRNA1 **(A)** or the gRNA2 **(B)** for *p14ARF* promoter region or with a scrambled gRNA **(C)** and the RNA from the infected cells was collected 12 h, 18 h and 24 h post-transduction. The cDNA produced was used as a template in qPCR. The relative expression levels of *p14ARF* were calculated using $2^{-\Delta\Delta Ct}$ method, and these levels were normalized against *GAPDH* (endogenous control). **(A)** It is observed a statistically significant increase in gene transcription 18 h after the infection, when compared to the non-infected samples, which indicate the reactivation of the gene. **(B)** No statistically significant results depicted; however, the expression levels tend to increase. **(C)** No relevant changes in gene transcription are observed, at any of the timepoints, using a scrambled gRNA. The results shown are the mean of three independent experiments. Error bars denote the standard error of the mean (SEM). * Student's t-test, p-value = 0,01, ns (non-significant)

complex genomic rearrangements, on-target or far way to the cut site, that lead to the accumulation of several abnormalities in the genome.[90] To protect from the pathogenic consequences that can advent, cells can induce senescence pathways. To clarify if the gene reactivation results indirectly from the damage introduced, it was used a scrambled gRNA. The results obtained allowed to rule out this possibility as it is not detected any significant changes in the expression levels. **(Figure 3.7C)**. Altogether, we can conclude that *p14ARF* transcription reactivation is a direct consequence of the promoter-specific DSB.

For *p16INK4a*, we verified that the differences in gene expression levels, after induction of the damage, at the different timepoints selected, do not show any evident tendency and, also, they are not statistically significant. **(Figure 3.8A)** This indicates that the approach applied did not work to reactivate the transcription of this gene. To explain the present situation, we suppose the existence of some constraints at the cut site, imposed by the system, that can perhaps interfere with the sequential events necessary for transcription initiation. The use of a scrambled gRNA leads to similar results. **(Figure 3.8B)** However, the data obtained for the present gene underpin the conclusion postulated for *p14ARF*. *p16INK4a* is a biomarker of senescence due to its known function to induce cell arrest through to inhibition of CDK4/6.[89,91] As this gene is also linked to senescence and as it is verified that the induction of a damage with a specific gRNA does not result in the reactivation of its transcription, this reinforces that the increase of *p14ARF* transcription is really due to the damage introduced in the promoter and not indirectly due to the activation of a checkpoint or pathway that leads to *p14ARF* - dependent senescence.

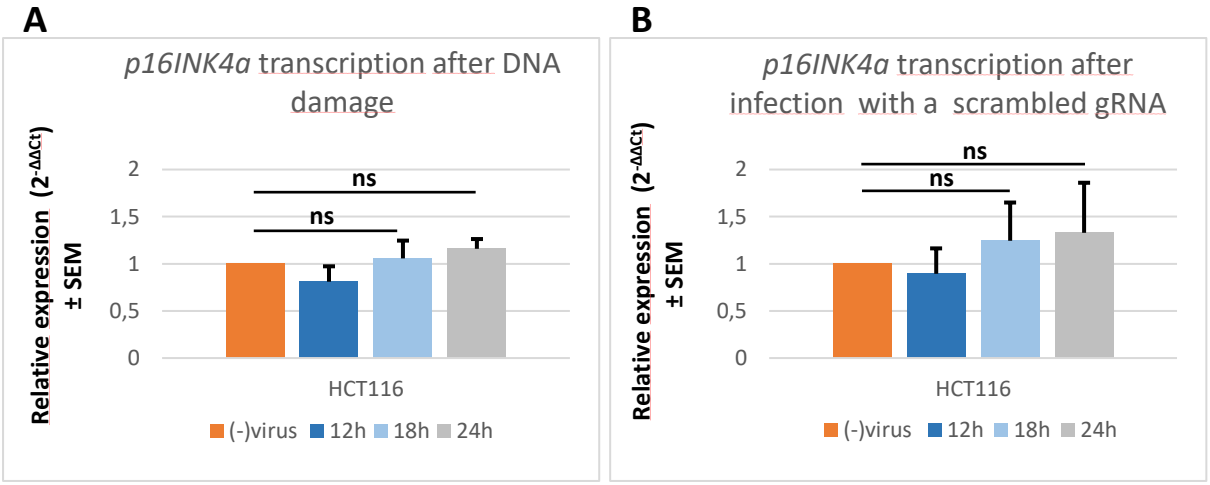


**Figure 3.8: Promoter-specific DNA damage doesn't reactivate *p16INK4a* gene.** Total RNA from non-infected (represented by "(-) virus") and infected HCT116 cells, were isolated and converted in cDNA by reverse transcription. The infection was carried out using CRISPR/Cas9 system with the specific gRNA for *p16INK4a* promoter region **(A)** or with a scrambled gRNA **(B)** and the RNA from the infected cells was collected 12 h, 18 h and 24 h post-transduction. The cDNA produced was used as a template in qPCR. The relative expression levels of *p16INK4a* were calculated using $2^{-\Delta\Delta Ct}$ method, and these levels were normalized against *GAPDH* (endogenous control). No statistically significant changes in gene transcription are reported, at any of the timepoints, using the specific gRNA for the promoter of the gene **(A)** or the scrambled gRNA **(B)**, which indicate the incapability of reactivation of the gene. The results shown are the mean of three independent experiments. Error bars denote the standard error of the mean (SEM). Student's t-test, ns (non-significant)

Altogether, the previous data demonstrated that the introduction of a DSB in the promoter of a gene, until now, in TSGs silenced by epigenetic marks, can give rise to different outcomes, being one of them the reactivation of transcription. The chromatin remodelling associated with DNA damage repair, with the formation of nucleosome-depleted regions along with the accumulation of transcription-prone histone marks, can be one explanation for the damage-induced transcription. Additionally, the alteration of the methylation profile upon DNA repair, reported in some studies[21],can facilitate a rewrite of the epigenetic landscape associated with the genes, making the repaired gene more permissive to transcription initiation.

The previous results elucidate the impact of DNA damage induction in the transcription of genes previously silenced. Moreover, we went to see what happen to the transcription of an actively transcribed gene, in this case the oncogene *RASSF1C*, if a damage is introduced in its promoter. Some previous reports have demonstrated that DNA damage leads to transcription repression.[36] The approach used for *RASSF1C* was the same applied to the other genes and the results obtained showed that the *RASSF1C* transcription is suppressed after the damage. **(Figure 3.9A)** Regarding these results, it would be interesting to perform the same experiments with the same system but with a modified endonuclease, namely a nuclease -null deactivated Cas9 (dCas9), which does not create a DSB, to evaluated if the suppression of transcription is a direct consequence of the damage itself or if it results from the occupancy of Cas9 at the target site that can block the ligation of TFs. However, the hypothesis that is the break the cause of transcription repression is the most plausible one since the present results are consistent with the results obtained in a previous work carried out in our lab, where it was visualized a rapid suppression of pre-existing transcription of active genes, not recovered during the analysis period, upon the induction of promoter-proximal DSB (unpublished data). The use of a scrambled gRNA didn't lead to changes in gene expression levels at the 12h and 18h timepoints; however, a statistically significant decrease is observed 24h post-transduction. **(Figure 3.9B).** This may be explained by the decrease in cell viability dependent on the activation of signalling pathways upon damage recognition. Viability assay should be performed to elucidate this situation.

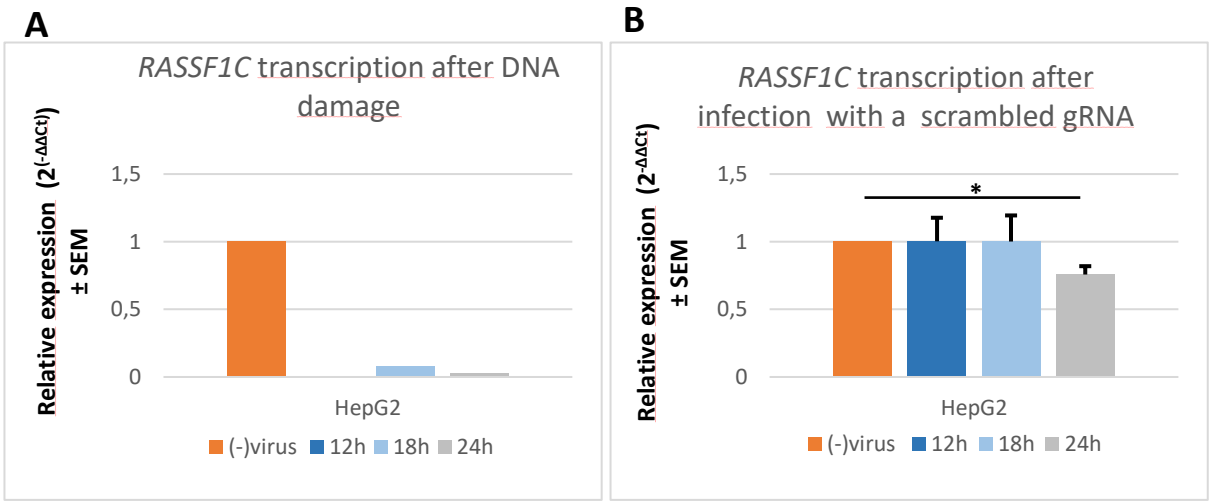


**Figure 3.9: Promoter-specific DNA damage abrogates *RASSF1C* transcription.** Total RNA from non-infected (represented by "(-) virus") and infected HepG2 cells, were isolated and converted in cDNA by reverse transcription. The infection was carried out using CRISPR/Cas9 system with the specific gRNA for *RASSF1C* promoter region **(A)** or with a scrambled gRNA **(B)** and the RNA from the infected cells was collected 12 h, 18 h and 24 h post-transduction. The cDNA produced was used as a template in qPCR. The relative expression levels of *RASSF1C* were calculated using $2^{-\Delta\Delta Ct}$ method, and these levels were normalized against *GAPDH* (endogenous control). **(A)** The transcription of *RASSF1C* stops after the induction of a damage in the promoter region of the gene, as we can see at any timepoint. **(B)** There are no changes in gene transcription 12h and 18h post-transduction, when compared to the non-infected cells. However, 24h post-transduction the levels of gene expression decrease. The results shown are the mean of three independent experiments. Error bars denote the standard error of the mean (SEM). The graph on the right represents the result of one unique experiment and, therefore, more biological replicates should be performed to confirm the present data. The graph on the left are the mean of three independent experiments and error bars denote the standard error of the mean (SEM). * Student's t-test, p-value = 0,03.

# 4. Concluding remarks and future work

Over the past years, the well characterized dogma that postulates that DNA damage is mainly involved in genomic integrity loss, being, therefore, directly linked to several pathological conditions[24] has been challenged. Growing and puzzling evidences have demonstrated that the induction of physiological DNA breaks is instrumental for the initiation of the transcriptional programs of environmental or developmental signal-dependent genes.[23,35] In fact, it was showed that DSBs are frequently detected at the promoter of several genes and the inhibition of DNA damage has been associated with an attenuation of gene expression. [37,43]

In the present project, it was revealed that, indeed, DNA damage plays a critical role in transcription, being sufficient to induce transcription activation. The crosstalk between damage, DDR and transcription is multifaceted and understanding the mechanisms beyond this interaction has become a new and exciting field of research. As previously referred, the initiation of transcriptional programs through damage can result from the resolution of topological barriers to RNAPII movement or from the induction of local changes of chromatin architecture. To disclose the hidden mechanism and players involved in the reactivation of the TSGs used in this project, additional methods need to be performed. The use of micrococcal nuclease (MNase) digestion assays, in order to evaluate the nucleosome occupancy at the promoter region, before and after the induction of the DSB, would allow to investigate chromatin changes upon damage, namely the formation of nucleosome-free regions that can facilitate the assembly of the transcriptional machinery and transcription initiation. Chromatin immunoprecipitation of histone marks would also help to understand how DNA damage mediates transcription activation, as it enables to depict the deposition of positive chromatin marks for transcription and, consequently, the formation of a transcription-prone environment. Additionally, and since we hypothesized that the alteration of the methylation pattern upon damage may be a plausible explanation for damage-mediated transcription, the use of methylation assays to evaluate the methylation landscape of the promoter region before and after the damage can give a huge input in the understanding of the results obtained. In fact, some authors have demonstrated, using a reported system, that the induction of a targeted DSB enables the cells, after a correct repair, to express the affected gene and the differences verified in the expression levels upon damage repair correlate with the change in the methylation pattern, being the formation of hypomethylated state linked to high levels of expression.[21] According to this view, the loss of the hypermethylated landscape, associated to the promoter of the TSGs, upon damage, is ,therefore, viewed as a possible mechanism by which DNA damage can promote transcription. An interesting finding depicted in this thesis is that the impact of the induction of a DSB within an actively transcribed gene or within a silenced gene on transcription is different, resulting in transcription repression or activation, respectively. Elucidating the synchronized events and the molecular players involved after damage, in each circumstance, can unclose the hidden cause of this variability.

In the last years, this emerging model of damage-dependent transcription activation has been exploited in the therapeutic context. For instance, in hormone-dependent neoplasias, such as breast and prostate cancer, that display resistance to hormone ablative therapies but still depend at some extent to estrogen or androgen signalling, the conjugation of a “hormone cycling” strategy, in order to induce DSBs, with inhibitors of DNA repair proteins, overwhelms cancer cells with breaks and leads to cell death.[35] Giving the importance of TSGs in preventing tumorigenesis[20], the reactivation, through damage, of the genes previously silenced opens new avenues for developing therapeutic approaches. Nonetheless, the viability of cancer cell upon TSGs activation should be evaluated to disclose the possible therapeutic potential of the used approach. Needless to say, generation of DSBs may contribute to genomic instability and any therapeutically approach based on this should be study carefully.

# 5. Bibliography


1. Shandilya, J. & Roberts, S. G. E. The transcription cycle in eukaryotes: From productive initiation to RNA polymerase II recycling. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1819,** 391–400 (2012).

2. Danino, Y. M., Even, D., Ideses, D. & Juven-Gershon, T. The core promoter: At the heart of gene expression. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1849,** 1116–1131 (2015).

3. Sainsbury, S., Bernecky, C. & Cramer, P. Structural basis of transcription initiation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **16,** 129–143 (2015).

4. Kagohara, L. T. *et al.* Epigenetic regulation of gene expression in cancer: Techniques, resources and analysis. *Brief. Funct. Genomics* **17,** 49–63 (2018).

5. Chodaparambil, J. V, Edayathumangalam, R. S., Bao, Y., Park, Y. J. & Luger, K. Nucleosome structure and function. *Ernst Schering Res. Found. Workshop* 29–46 (2006).

6. Kornberg, R. Chromatin Structure : A Repeating Unit of Histones and DNA. *Science (80-. ).* **184,** 868–871 (1974).

7. Hübner, M. R., Eckersley-Maslin, M. A. & Spector, D. L. Chromatin organization and transcriptional regulation. *Curr. Opin. Genet. Dev.* **23,** 89–95 (2013).

8. Li, B., Carey, M. & Workman, J. L. The Role of Chromatin during Transcription. *Cell* **128,** 707–719 (2007).

9. Hübner, M. R. & Spector, D. L. Chromatin dynamics. *Annu. Rev. Biophys.* **39,** 471–89 (2010).

10. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res.* **21,** 381–395 (2011).

11. Jiang, C. & Pugh, B. F. Nucleosome positioning and gene regulation: advances through genomics. *Nat. Rev. Genet.* **10,** 161–172 (2009).

12. Dawson, M. A. & Kouzarides, T. Cancer epigenetics: From mechanism to therapy. *Cell* **150,** 12–27 (2012).

13. Pulecio, J., Verma, N., Mejía-Ramírez, E., Huangfu, D. & Raya, A. CRISPR/Cas9-Based Engineering of the Epigenome. *Cell Stem Cell* **21,** 431–447 (2017).

14. Zhao, R., Choi, B. Y., Lee, M. H., Bode, A. M. & Dong, Z. Implications of Genetic and Epigenetic Alterations of CDKN2A (p16INK4a) in Cancer. *EBioMedicine* **8,** 30–39 (2016).

15. Kazanets, A., Shorstova, T., Hilmi, K., Marques, M. & Witcher, M. Epigenetic silencing of tumor suppressor genes: Paradigms, puzzles, and potential. *Biochim. Biophys. Acta - Rev. Cancer* **1865,** 275–288 (2016).

16. Jones, P. A. Functions of DNA methylation: Islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* **13,** 484–492 (2012).

17. Tessarz, P. & Kouzarides, T. Histone core modifications regulating nucleosome structure and dynamics. *Nat. Rev. Mol. Cell Biol.* **15,** 703–708 (2014).

18. Zhang, T., Cooper, S. & Brockdorff, N. The interplay of histone modifications - writers that read.

*EMBO Rep.* **16,** 1467–1481 (2015).

19. Waldmann, T. & Schneider, R. Targeting histone modifications - epigenetics in cancer. *Curr. Opin. Cell Biol.* **25,** 184–189 (2013).

20. Guo, X. E., Ngo, B., Modrek, A. S. & Lee, W.-H. Targeting tumor suppressor networks for cancer therapeutics. *Curr. Drug Targets* **15,** 2–16 (2014).

21. Allen, B. *et al.* Non-homologous end joining induced alterations in DNA methylation: A source of permanent epigenetic change. *Oncotarget* **8,** 40359–40372 (2017).

22. Agathanggelou, A., Cooper, W. N. & Latif, F. Role of the ras-association domain family 1 tumor suppressor gene in human cancers. *Cancer Res.* **65,** 3497–3508 (2005).

23. Vitelli, V. *et al.* Recent Advancements in DNA Damage–Transcription Crosstalk and High-Resolution Mapping of DNA Breaks. *Annu. Rev. Genomics Hum. Genet.* **18,** 87–113 (2017).

24. Ciccia, A. & Elledge, S. J. The DNA Damage Response: Making It Safe to Play with Knives. *Mol. Cell* **40,** 179–204 (2010).

25. Price, B. D. & D'Andrea, A. D. Chromatin Remodeling at DNA Double-Strand Breaks. *Cell* **152,** 1344–1354 (2013).

26. Soria, G., Polo, S. E. & Almouzni, G. Prime, Repair, Restore: The Active Role of Chromatin in the DNA Damage Response. *Mol. Cell* **46,** 722–734 (2012).

27. Gursoy-Yuzugullu, O., House, N. & Price, B. D. Patching Broken DNA: Nucleosome Dynamics and the Repair of DNA Breaks. *J. Mol. Biol.* **428,** 1846–1860 (2016).

28. Deem, A. K., Li, X. & Tyler, J. K. Epigenetic regulation of genomic integrity. *Chromosoma* **121,** 131–151 (2012).

29. Lukas, J., Lukas, C. & Bartek, J. More than just a focus: The chromatin response to DNA damage and its role in genome integrity maintenance. *Nat. Cell Biol.* **13,** 1161–1169 (2011).

30. Lieber, M. R. The Mechanism of Double-Strand DNA Break Repair by the Nonhomologous DNA End-Joining Pathway. *Annu. Rev. Biochem.* **79,** 181–211 (2010).

31. Bunting, S. F. *et al.* 53BP1 Inhibits Homologous Recombination in Brca1-Deficient Cells by Blocking Resection of DNA Breaks. *Cell* **141,** 243–254 (2010).

32. Lieber, M. R. The mechanism of human nonhomologous DNA End joining. *J. Biol. Chem.* **283,** 1–5 (2008).

33. San Filippo, J., Sung, P. & Klein, H. Mechanism of Eukaryotic Homologous Recombination. *Annu. Rev. Biochem.* **77,** 229–257 (2008).

34. Shrivastav, M., De Haro, L. P. & Nickoloff, J. A. Regulation of DNA double-strand break repair pathway choice. *Cell Res.* **18,** 134–147 (2008).

35. Haffner, M. C., De Marzo, A. M., Meeker, A. K., Nelson, W. G. & Yegnasubramanian, S. Transcription-induced DNA double strand breaks: both oncogenic force and potential therapeutic target? *Clin. Cancer Res.* **17,** 3858–64 (2011).

36. Khobta, A. & Epe, B. Interactions between DNA damage, repair, and transcription. *Mutat. Res. - Fundam. Mol. Mech. Mutagen.* **736,** 5–14 (2012).

37. Madabhushi, R. *et al.* Activity-Induced DNA Breaks Govern the Expression of Neuronal Early-Response Genes. *Cell* **161,** 1592–1605 (2015).

38. Sharma, N., Gabel, H. W. & Greenberg, M. E. A Shortcut to Activity-Dependent Transcription. *Cell* **161,** 1496–8 (2015).

39. Adelman, K. & Lis, J. T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.* **13,** 720–731 (2012).

40. Bunch, H. *et al.* Transcriptional elongation requires DNA break-induced signalling. *Nat. Commun.* **6,** 10191 (2015).

41. Gudipaty, S. A., McNamara, R. P., Morton, E. L. & D'Orso, I. PPM1G Binds 7SK RNA and Hexim1 To Block P-TEFb Assembly into the 7SK snRNP and Sustain Transcription Elongation. *Mol. Cell. Biol.* **35,** 3810–28 (2015).

42. Beato, M., Wright, R. H. & Vicent, G. P. DNA damage and gene transcription: accident or necessity? *Cell Res.* **25,** 769–70 (2015).

43. Ju, B.-G. *et al.* A topoisomerase IIbeta-mediated dsDNA break required for regulated transcription. *Science* **312,** 1798–802 (2006).

44. Singh, I. *et al.* High mobility group protein-mediated transcription requires DNA damage marker γ-H2AX. *Cell Res.* **25,** 837–850 (2015).

45. Perillo, B. *et al.* DNA Oxidation as Triggered by H3K9me2 Demethylation Drives Estrogen-Induced Gene Expression. *Science (80-. ).* **319,** 202–206 (2008).

46. Carvalho, S. *et al.* Histone methyltransferase SETD2 coordinates FACT recruitment with nucleosome dynamics during transcription. *Nucleic Acids Res.* **41,** 2881–2893 (2013).

47. Guidelines for Maintaining Cultured Cells. *Thermo Fisher Scientific* Available at: https://www.thermofisher.com/pt/en/home/references/gibco-cell-culture-basics/cell-culture-protocols/maintaining-cultured-cells.html. (Accessed: 7th May 2018)

48. Trypan Blue solution. *Sigma-Aldrich* Available at: https://www.sigmaaldrich.com/catalog/product/sigma/t8154?lang=pt&region=PT. (Accessed: 8th May 2018)

49. Fuentes, M. Hemocytometer protocol. *hemocytometer.org* Available at: https://www.hemocytometer.org/hemocytometer-protocol/. (Accessed: 8th May 2018)

50. Doak, S. H. & Zaïr, Z. M. Real-Time Reverse-Transcription Polymerase Chain Reaction: Technical Considerations for Gene Expression Analysis. in *Parry J., Parry E. (eds) Genetic Toxicology. Methods in Molecular Biology (Methods and Protocols)* 251–270 (Springer, New York, NY, 2012). doi:10.1007/978-1-61779-421-6_13

51. TRIzol Reagent. *Thermo Fisher Scientific* Available at: https://www.thermofisher.com/order/catalog/product/15596018. (Accessed: 10th May 2018)

52. Corthell, J. T. Reverse Transcription (RT) and Polymerase Chain Reaction (PCR). in *Basic Molecular Protocols in Neuroscience: Tips, Tricks, and Pitfalls* 27–40 (Academic Press, 2014). doi:10.1016/B978-0-12-801461-5.00004-6

53. Zhong, S. *et al.* Silencing of GSTP1 gene by CpG island DNA hypermethylation in HBV-

associated hepatocellular carcinomas. *Clin. Cancer Res.* **8,** 1087–1092 (2002).

54. Zhong, S. *et al.* Intensive hypermethylation of the CpG island of Ras association domain family 1A in hepatitis B virus-associated hepatocellular carcinomas. *Clin. Cancer Res.* **9,** 3376–3382 (2003).

55. Hnasko, T. S. & Hnasko, R. M. The Western Blot. in *Hnasko R. (eds) ELISA. Methods in Molecular Biology* 87–96 (Humana Press, New York, NY, 2015). doi:10.1007/978-1-4939-2742-5_9

56. Mahmood, T. & Yang, P.-C. Western blot: technique, theory, and trouble shooting. *N. Am. J. Med. Sci.* **4,** 429–34 (2012).

57. Brunelle, J. L. & Green, R. One-dimensional SDS-Polyacrylamide Gel Electrophoresis (1D SDS-PAGE). *Methods Enzymol.* **541,** 151–159 (2014).

58. Zhang, F., Wen, Y. & Guo, X. CRISPR/Cas9 for genome editing: Progress, implications and challenges. *Hum. Mol. Genet.* **23,** 40–46 (2014).

59. *CRISPR Handbook – Enabling Genome Editing and Transforming Life Science Research*. (GenScript).

60. Felder, Y. Analysis of Biological Networks : Transcriptional Networks - Promoter Sequence Analysis. *Gene* 1–15 (2007).

61. Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* **39,** 311–318 (2007).

62. Zumbo, P. *Phenol-chloroform Extraction*.

63. Wagner, E. M. Monitoring Gene Expression: Quantitative Real-Time RT-PCR. in *Freeman L. (eds) Lipoproteins and Cardiovascular Disease. Methods in Molecular Biology (Methods and Protocols)* 19–45 (Humana Press, Totowa, NJ, 2013). doi:10.1007/978-1-60327-369-5_2

64. Bustin, S. A., Benes, V., Nolan, T. & Pfaffl, M. W. Quantitative real-time RT-PCR - A perspective. *J. Mol. Endocrinol.* **34,** 597–601 (2005).

65. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. *Methods* **25,** 402–408 (2001).

66. Robert, M. F. *et al.* DNMT1 is required to maintain CpG methylation and aberrant gene silencing in human cancer cells. *Nat. Genet.* **33,** 61–65 (2003).

67. Burri, N. *et al.* Methylation silencing and mutations of the p14ARF and p16INK4a genes in colon cancer. *Lab. Invest.* **81,** 217–229 (2001).

68. Badal, V., Menendez, S., Coomber, D. & Lane, D. P. Regulation of the p14ARF promoter by DNA methylation. *Cell Cycle* **7,** 112–119 (2008).

69. Robertson, K. D. & Jones, P. a. The human ARF cell cycle regulatory gene promoter is a CpG island which can be silenced by DNA methylation and down-regulated by wild-type p53. *Mol. Cell. Biol.* **18,** 6457–6473 (1998).

70. Esteller, M. *et al.* Hypermethylation-associated Inactivation of p14 ARF Is Independent of p16 INK4a Methylation and p53 Mutational Status 1. *Cancer Res.* **60,** 129–133 (2000).

71. Ozenne, P., Eymin, B., Brambilla, E. & Gazzeri, S. The ARF tumor suppressor: Structure, functions and status in cancer. *Int. J. Cancer* **127,** 2239–2247 (2010).

72. Cell atlas - CDKN2A - The Human Protein Atlas. *proteinatlas.org* Available at: https://www.proteinatlas.org/ENSG00000147889-CDKN2A/cell. (Accessed: 4th October 2017)

73. Restriction Fragment Length Polymorphism (RFLP). *NCBI* Available at: https://www.ncbi.nlm.nih.gov/probe/docs/techrflp/. (Accessed: 15th August 2018)

74. Saraswathy, N., Ramalingam, P., Saraswathy, N. & Ramalingam, P. Genome mapping. in *Concepts and Techniques in Genomics and Proteomics* 77–93 (Woodhead Publishing, 2011). doi:10.1533/9781908818058.77

75. Forrester, J. V. *et al.* Genetics. in *The Eye - Basic Sciences in Practice* 130–156.e1 (W.B. Saunders, 2016). doi:10.1016/B978-0-7020-5554-6.00003-4

76. Esteller, M. *et al.* Inactivation of Glutathione S -Transferase P1 Gene by Promoter Hypermethylation in Human Neoplasia. *Cancer Res.* **58,** 4515–4518 (1998).

77. Gurioli, G. *et al.* GSTP1 methylation in cancer: A liquid biopsy biomarker? *Clin. Chem. Lab. Med.* **56,** 705–717 (2018).

78. Lo, H.-W. *et al.* Identification and functional characterization of the human glutathione S-transferase P1 gene as a novel transcriptional target of the p53 tumor suppressor gene. *Mol. Cancer Res.* **6,** 843–50 (2008).

79. Kou, X., Chen, N., Feng, Z., Luo, L. & Yin, Z. GSTP1 negatively regulates Stat3 activation in epidermal growth factor signaling. *Oncol. Lett.* **5,** 1053–1057 (2013).

80. Estrabaud, E. *et al.* RASSF1C, an isoform of the tumor suppressor RASSF1A, promotes the accumulation of B-catenin by interacting with BTrCP. *Cancer Res.* **67,** 1054–1061 (2007).

81. van der Weyden, L. & Adams, D. J. The Ras-association domain family (RASSF) members and their role in human tumourigenesis. *Biochim. Biophys. Acta - Rev. Cancer* **1776,** 58–85 (2007).

82. Iwasa, H., Hossain, S. & Hata, Y. Tumor suppressor C-RASSF proteins. *Cell. Mol. Life Sci.* **75,** 1773–1787 (2018).

83. Volodko, N. *et al.* RASSF1A site-specific methylation hotspots in cancer and correlation with RASSF1C and MOAP-1. *Cancers (Basel).* **8,** (2016).

84. Reeves, M. E., Firek, M., Chen, S.-T. & Amaar, Y. The RASSF1 Gene and the Opposing Effects of the RASSF1A and RASSF1C Isoforms on Cell Proliferation and Apoptosis. *Mol. Biol. Int.* **2013,** 145096 (2013).

85. Schagdarsurengin, U. *et al.* Frequent epigenetic inactivation of the RASSF1A gene in hepatocellular carcinoma. *Oncogene* **22,** 1866–1871 (2003).

86. Sander, J. D. & Joung, J. K. CRISPR-Cas systems for genome editing, regulation and targeting. *Nat. Biotechnol.* **32,** 347–355 (2014).

87. Zhao, D. *et al.* CRISPR/Cas9-assisted gRNA-free one-step genome editing with no sequence limitations and improved targeting efficiency. *Sci. Rep.* **7,** 1–9 (2017).

88. Zischewski, J., Fischer, R. & Bortesi, L. Detection of on-target and off-target mutations

generated by CRISPR/Cas9 and other sequence-specific nucleases. *Biotechnol. Adv.* **35,** 95–104 (2017).

89. Martínez-Zamudio, R. I., Robinson, L., Roux, P. F. & Bischof, O. SnapShot: Cellular Senescence Pathways. *Cell* **170,** 816–816.e1 (2017).

90. Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* (2018). doi:10.1038/nbt.4192

91. Sharpless, N. E. & Sherr, C. J. Forging a signature of in vivo senescence. *Nat. Rev. Cancer* **15,** 397–408 (2015).

# 6. Annexes

## 6.1 Annex I

**Table 6.1:** Sequences of the primers used in the gene expression experiments (RT-PCR and RT-qPCR)

| Gene | Primer designation | Sequence (5'-3') | Product length amplified |
|---|---|---|---|
| *p16INK4a* | Forward | CATCGCGATGTCGCACGGTA | 210 bp |
| | Reverse | TACGAAAGCGGGGTGGGTTGTG | |
| *p14ARF* | Forward | GGTTTTCGTGGTTCACATCCCGCG | 254 bp |
| | Reverse | CAGGAAGCCCTCCCGGGCAGC | |
| *GSTP1* | Forward | ATGACTATGTGAAGGCACTG | 250 bp |
| | Reverse | AGGTTCACGTACTCAGGGGA | |
| *RASSF1C* | Forward | CGGAGGCGCCTTCTTTCGAAATGACCT | 272 bp |
| | Reverse | GATGAAGCCTGTGTAAGAACCGTCCT | |
| *GAPDH* | Forward | GAAGGTGGAGGTCGGAGTC | 260 bp |
| | Reverse | GAAGATGGTGATGGGATTT | |

**Table 6.2:** Sequences of the guide RNAs used in this study.

| Gene | gRNA designation | Sequence (5'-3') |
|---|---|---|
| *p16INK4a* | gRNA1 | GCACTCAAACACGCCTTTGC TGG |
| *p14ARF* | gRNA1 | TCAGAGCCGTTCCGAGATCT TGG |
| | gRNA2 | GCGGGAAAGTGGCGGTAGGC GGG |
| *GSTP1* | gRNA1 | AATTTCCCCCCGCGATGTCC CGG |
| *RASSF1C* | gRNA1 | TTGTGCGCTTGCCCGGACGC TGG |

**Table 6.3:** Sequences of the primers used for genomic DNA (qPCR)

| Gene | Primer designation | Sequence (5'-3') |
|---|---|---|
| *p16INK4a* | Forward | GACAGCCGTTTTACACGCAG |
| | Reverse | CAAATCCTCTGGAGGGACCG |
| *p14ARF* | Forward | TAGGGCCGTGTCAGGTGA |
| | Reverse | TCACCAAGAACCTGCGCAC |
| *GSTP1* | Forward | GGC CAC AGC GTG AGA CTA C |
| | Reverse | CAG ACA GCA GGA AGA GGA CC |
| *RASSF1C* | Forward | CGCGCAAGGAGCGTTCTAA |
| | Reverse | GTCCGTAGCCGCCAACC |
| *GAPDH* | Forward | ATAGGCGAGATCCCTCCAA |
| | Reverse | TGAAGACGCCAGTGGAC |

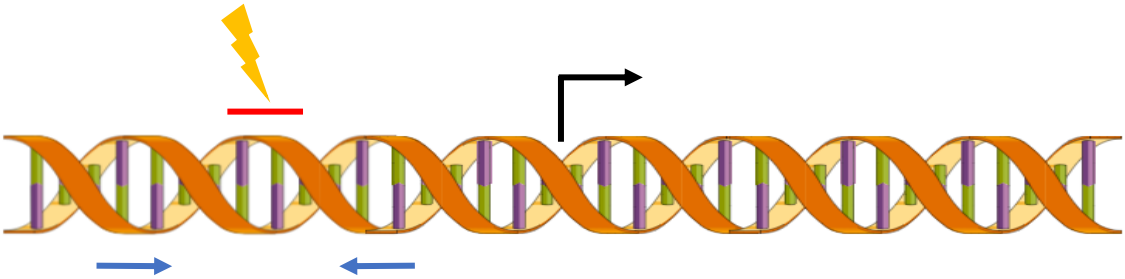

**Figure 6.1: Schematic representation of the methodology used to check if CRISPR/Cas9 system is working**. Each gRNA (indicated by the red line) is complementary to a DNA segment within the promoter region of the respective gene and will target DNA damage for that region. The pairs of primers designed (indicated by the blue arrows) flank the genomic region targeted by the gRNA. Therefore, if a DSB is induced, no amplification product will be detected. The genomic region flanked by the pair of primers designed for *p14ARF* includes the target sites of both the gRNAs used for that gene.

## 6.2 Annex II

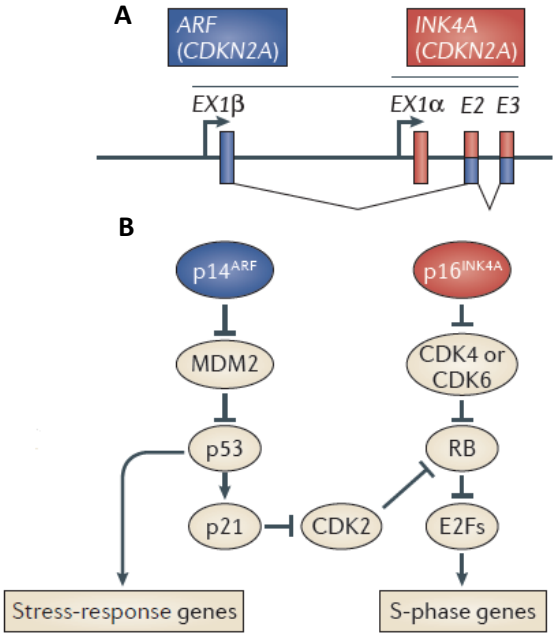


**Figure 6.2: Map of the *CDKN2A* or *INK4a/ARF* locus and the function of the encoded proteins. (A)** Configuration of the *CDKN2A* locus. *CDKN2A* gene encodes two different proteins, p14ARF and p16INK4a. Each protein results from the transcription of three exons (colored vertical bars). The dark blue exons correspond to p14ARF and the three red exons correspond to p16INK4a. Exon 2 and 3 are shared by both proteins, however they have different first exons (Ex1β for p14ARF and Ex1α for p16INK4a). Exon 1β, transcribed from its own promoter, is spliced into exon 2 of *p16INK4a* in an alternative reading frame, which produces a different protein, p14ARF, totally unrelated to p16INK4a. The promoters of each transcript are indicated by arrows. **(B)** Both proteins, p16INK4a and p14ARF, play a pivotal role in cell cycle control, via Rb and p53 pathways, respectively. The p16INK4a binds to CDK4/6 preventing the formation of cyclin D- CDK4/6 complex (also known as G1-CDK complex). This complex is involved in the phosphorylation of the retinoblastoma protein (Rb), which allows the release of the E2F factor from the repressive state imposed by Rb. The lack of G1-CDK maintains Rb in its active state (hypophosphorylated) and the persistence of the Rb/E2Fs-repressive complex blocks the E2F-mediated transcription of S-phase gene, leading to cell cycle arrest in G1. The p14ARF protein prevents p53 degradation, through binding to MDM2, an ubiquitin ligase that mediates the degradation of p53. Therefore, p53 is stabilized and its transcriptional activity is stimulated, triggering the expression of several apoptosis inducers and cell cycle inhibitors genes such as cyclin-dependent kinase inhibitor 1A (*CDKN1A*) that encodes p21, a member of cyclin-CDK inhibitor proteins family, which binds to CDK2 and inactive G1/S-CDK and S-CDK complexes, resulting in cell cycle arrest. Adapted from *Sharpless, N. E. & Sherr, C. J.* [91]

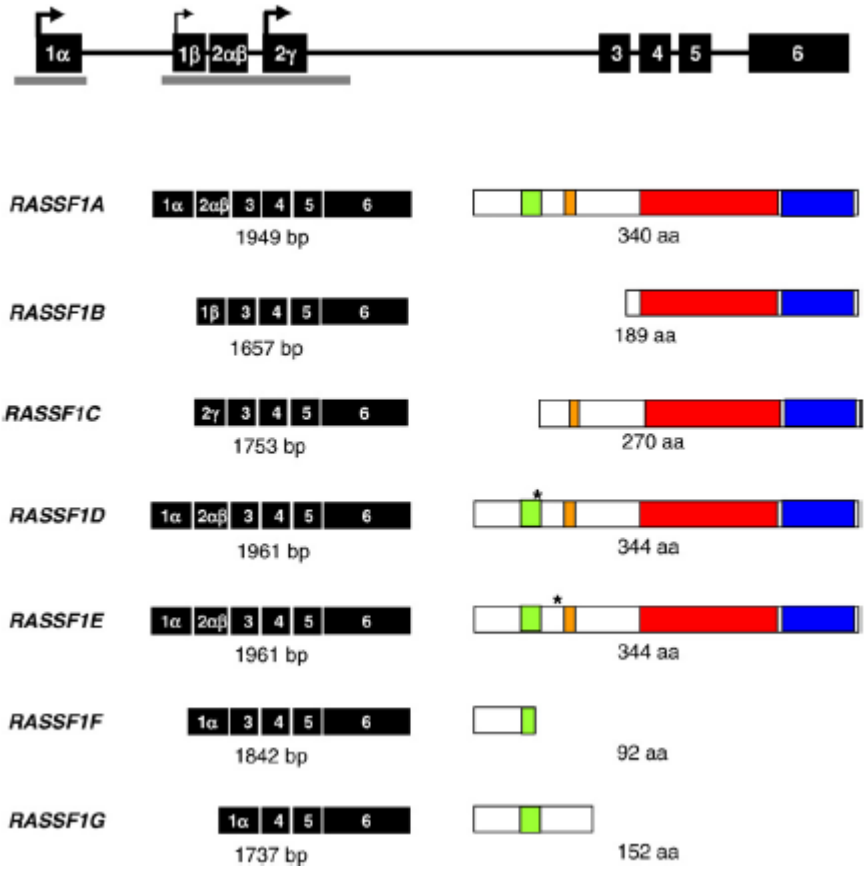


**Figure 6.3: Map of *RASSF1* locus and isoforms/proteins encoded.** *RASSF1* gene encodes seven different transcripts through differential promoter usage, which is associated to two distinct CpGIs (grey bars), and alternative splicing. *RASSF1A* transcription initiates from the promoter located in the first CpGI and the first exon transcribed is the exon 1α followed by exon 2αβ. Contrarily, *RASSF1B* utilizes a different 5'exon, namely the exon 1β. *RASSF1C* transcripts begin at the exon 2γ and use the promoter located in the second CpGI. The remaining isoforms are all splice variants from *RASSF1A*. *RASSF1D* and *RASSF1E* express additional amino acids (black asterisk). *RASSF1E* and *RASSF1F* skip some exons and produce truncated proteins. As the isoforms are originate from different exons, this will be translated into different structural domains in the proteins encode. A coiled-coil motif in the C-terminal region, named Salvador/RASSF/Hippo (SARAH) domain, represented in blue, mediates protein-protein interactions in pathways that induce cell cycle arrest and apoptosis. Ras-association (RA) domain, represented in red, is responsible for the cross-talk with Ras and other small GTPases. The diacylglycerol (DAG) /phorbol ester-binding domain, also known as the protein kinase C conserved region (C1) is represented in green and the ATM-kinase phosphorylation consensus sequence motif, which suggests that RASSF1 might be a substrate of this kinase, is represented in orange. Adapted from *van der Weyden, L. & Adams, D. J.*[81]