Universidade de Lisboa
Faculdade de Farmácia

Cancer Epigenetics

Luís Miguel Terrinca da Silva

Monografia orientada pelo Professor Doutor Carolino José Nunes Monteiro, Professor Associado com Agregação.

Mestrado Integrado em Ciências Farmacêuticas

2021

Universidade de Lisboa
Faculdade de Farmácia

Cancer Epigenetics

Luís Miguel Terrinca da Silva

Trabalho Final de Mestrado Integrado em Ciências Farmacêuticas apresentado à Universidade de Lisboa através da Faculdade de Farmácia

Monografia orientada pelo Professor Doutor Carolino José Nunes Monteiro, Professor Associado com Agregação.

2021

Resumo

A Epigenética traz um largo percurso histórico, desde a sua descoberta até á sua aplicação em práticas laboratoriais da vida real. Fundamentalmente, Epigenética traduz-se nas modificações químicas que o DNA sofre, mas que não alteram diretamente o código genético, alterando a regulação dos genes através de uma grande variedade de mecanismos capazes de modificar o estado da cromatina, bem como o das proteínas responsáveis pela sua transcrição. A metilação no DNA, as modificações nas histonas e os *non-coding* RNAs são os principais mecanismos através dos quais os traços epigenéticos são passados entre as gerações. Estas marcas epigenéticas são transmitidas de acordo com a sua "impressão genómica", o processo pelo qual os genes "impressos" mantêm a sua informação genética e são capazes de a transmitir às novas linhagens de células, apesar de estas sofrerem um *reset* epigenético durante o seu desenvolvimento embrionário, de maneira a adquirir o seu estado de pluripotência.

Como o Cancro mantém a sua posição no pódio, em segundo lugar entre as doenças que causam mais mortes anualmente, as variações epigenéticas nos cancros têm sido um campo de estudo muito abordado recentemente nesta área, de maneira a potenciar as terapêuticas e os procedimentos de diagnóstico. Através de cuidadosas investigações, foi descoberto que os tecidos tumorais apresentavam frequentemente marcas epigenéticas nas suas células, e os recentes avanços tecnológicos permitiram o estudo destes traços epigenéticos mutados em diferentes tipos de cancros. Foi ainda provado que as células tumorais sofrem um fenómeno idêntico ao *reset* epigenético, mas que é controlado pelas próprias células tumorais, o que lhes permite a continua renovação do seu estado de pluripotência, concedendo-lhes liberdade para se de-diferenciarem e trans-diferenciarem. Todos os traços epigenéticos previamente mencionados já foram encontrados alterados em diversas maneiras em diferentes tecidos cancerosos, tendo sido levadas a cabo diversas investigações no sentido de desenvolver procedimentos de diagnóstico e terapias com o âmbito de tratar estes cancros. Apesar de se considerar um campo relativamente recente, o estudo de variações epigenéticas no cancro já conta com bastantes procedimentos de diagnóstico e terapias que são aplicadas por todo o mundo.

Palavras-chave: epigenética; variações epigenéticas no cancro; metilação no DNA; ncRNA; reset epigenético.

Abstract

Epigenetics have come a long way since its discovery, through its application in real world practices. Epigenetics in its fundamental meaning, translates in chemical modifications in the DNA sequence that does not directly alter the DNA code, but instead alters its gene regulation through a variety of mechanisms that change the chromatins state, as well as the proteins responsible for its transcription. DNA methylation, histone modifications and non-coding RNAs are the main mechanisms through epigenetic traits are passed down from generation to generation. These epigenetic marks are passed down through generations according to its genomic imprinting, the process through which the imprinted genes maintain their epigenetic information and are able to transmit them to the new lineages of cells, although they go through an epigenetic reset process during embryonic development, in order to gain its pluripotent cell state.

As Cancer still maintains its podium position as number two amongst the diseases which cause the most deaths annually, cancer epigenetics has been a recent field of study amongst cancer subjects, in order to enhance cancer therapies and diagnostic procedures. Through thoroughtfully cared investigations, it was found that tumorous tissues often present epigenetic marks in their cells, and recent advances in technical methods have allowed the study of these mutated expressed epigenetic traits along the different cancer types. It was also proven that tumorous cells undergo a similar process as epigenetic reset, but this one is controlled by the tumorous cell and allows it to continually regain its pluripotent state, granting it full liberty to de-differentiate and transdifferentiate. Each of the epigenetic traits have already been found to be present in numerous altered ways in various cancerous tissues, and therefore further investigations have already developed diagnostic and therapeutical procedures in order to treat this epigenetically modified cancers. Although a considerably recent field of study, cancer epigenetics already counts with multiple diagnostic and therapeutical methods being applied all over the world, and it only hopes to grow even further.

Keywords: epigenetics; cancer epigenetics; DNA methylation; ncRNA; epigenetic reset.

Agradecimentos

Findo o meu percurso académico na Faculdade de Farmácia da Universidade de Lisboa enquanto aluno do Mestrado Integrado em Ciências Farmacêuticas, gostaria de agradecer a todas as pessoas que fizeram parte do mesmo e que o tornaram irrefutavelmente inesquecível.

Em primeiro lugar, quero agradecer à minha família, que nestes 5 anos de curso nunca deixaram que os 200 e muitos quilómetros que nos separavam se tornassem um pesar e que, apesar da distância, nunca deixaram de me acompanhar nesta caminhada e que me foram constantemente e incansavelmente apoiando em todos os projetos de que eu tomei parte. A vocês, o meu maior obrigado.

Aos meus mais queridos colegas, que se tornaram a minha família Lisboeta, pela qual fui capaz de continuar a querer sempre mais e melhor de mim, e ao lado de quem persisti e lutei desde o início até ao fim do curso, um enorme obrigado. Aos meus três mosqueteiros, Carlos, João e Simone, à minha conterrânea, Marta, aos meus melhores amigos da Tuna, Bernardo e Guto e aos que Falam e dizem tudo, este agradecimento é para vocês. Às minhas madrinhas e aos meus afilhados, obrigado por me continuarem a fazer crescer e a crescer comigo a meu lado. Aos que apesar de não estudarem, ainda assim fizeram parte do meu percurso, Pedro e Susana, um grande obrigado a vocês por terem sido os meus Orientadores da vida na Faculdade. Ainda aos amigos que já trazia de casa, a vocês em particular, Manuel, Miguel Demétrio e Miguel Ribeirinho, obrigado por todo o desagrado por não conseguir passar tanto tempo convosco e por mesmo assim não desistirem de mim, também eu não penso desistir de vocês. A ti, Maria, obrigado por continuares a ver sempre o melhor de mim e por continuares a caminhar a meu lado rumo ao que ainda está por vir.

Às Associações em que tive o prazer de me cultivar nos mais variados sentidos, a vocês AEFFUL, TAFUL e MNAP, obrigado por todos os Saraus, Festivais e Queimas (mais ou menos) que me proporcionaram. Sem vocês não seria metade da pessoa e jovem Farmacêutico que me estou a tornar e ainda ambiciono ser.

Por fim, mas não menos importante, um grande obrigado ao Professor Carolino Monteiro que sempre e incansavelmente me acompanhou na construção desta monografia. Obrigado por nunca desistir de mim e por continuar a ser para mim um Exemplo como Professor e como Orientador.

A vós, o meu mais sincero e sentido obrigado.

Abreviaturas

5-aza – 5-Azacitadine

5caC – 5-carboxylcytosine

5fC – 5-formyl-cytosine

5hmC - 5-hydroxymethylcytosine

5mC – 5-methylcytosine

ALL – Acute Lymphoid Leukemias

AML – Acute Myeloid Leukaemia

ANRIL – Anti-sense Non-coding RNA in the INK locus lncRNA

BER – base excision repair

Bp – base pairs

CDE – Cancer-derived exomes

ceRNA – competing endogenous RNAs

cfDNA – Cell-free DNA

CH – Clonal Hematopoiesis

CRC – Colorectal Cancer

CSC – Cancer stem cell

DC – Dendritic cell

DIPG – Diffused Intrinsic Pontine Glioma

DMR – Differentially Methylated Region

DNA – Desoxyribonucleic Acid

DNMT – DNA methyltransferases

dsRNA – Double-stranded RNA

EAC – Esophageal Adenocarcinoma

ECV – Extracelular Vesicles

EMT – Epithelial-Mesenchymal Transition

ESCC – Esophageal Squamous Cell Carcinoma

GC – Gastric Cancer

HAT – Histone acetyltransferases

HCC – Hepatocellular carcinoma

HDAC – Histone deacetylases

HDACi – Histone deacetylases inhibitor

HDM – Histone demethylases

HER2 – Human epidermal growth factor receptor 2

HMT – Histone methylases

HNSCC – Head and Neck Squamous Cell Carcinoma

HOTAIR - Homeobox transcript antisense RNA

HSC – Haematopoietic stem cells

ICR – Imprint Control Region

IDH – Isocitrate Dehydrogenase

IGF2 – insulin-like growth factor 2

KMT – Lysine methylatransferases

lncRNA – Long non-coding RNA

LOI – Loss of Imprinting

LPS – Lipopolysaccharide

MDS – Myelodysplastic syndromes

MHC – Major Histocompatibility Complex

MLL – Mixed-lineage Leukemia

mRNA – Messenger RNA

miRNA – Micro RNA

ncRNA – non-coding RNA

NEAT1 – Nuclear Paraspeckle Assembly Transcript 1

onco-miRNA – oncogenic miRNAs

PBMC – Peripheral Blood Mononuclear Cell

PDAC – Pancreatic Ductal Adenocarcinoma

PGC – Primordial germ cell

PHD – Plant Homeodomain gene cluster

piRNA – Piwi-interacting RNA

PPI – Protein-protein Interactions

PRC – Polycomb Repressive Complexes

Pri-miRNA – Primary miRNA

PTGS – Post-transcriptional Gene Silencing

PTM – post-translational modification

RISC – RNA-induced silencing complex

RNA – Ribonucleic acid

RNAi – RNA interference

RT – Reverse Transcriptase

rRNA – Ribosomal RNA

SAH – S-adenosylhomocysteine

SAM – S-adenosylmethionine

SC – Stem Cell

SFRP2 – Secreted frizzled-related protein 2

SGI-110 – Guadecitabine

shRNA – Short Harpin RNA

siRNA – Small interference RNA

snRNA – Small Nuclear RNA

snRNP – Small Nuclear ribonucleoprotein

sncRNA – Short non-coding RNA

snoRNA – Small Nucleolar RNA

ssRNA – Single-stranded RNA

SWI/SNF – SWItch/Sucrose Non-Fermentable complex

TAA – Tumour-associated Antigens

TCR – T-cell Receptor

TDG – DNA glycosylase

TET – Ten-eleven translocation enzymes

TIL – Tumour-infiltrating lymphocytes

TME – Tumour Microenvironment

tRNA – Transfer RNA

ts-RNA – Tumour-suppressive miRNAs

UCA1 – Urothelial Carcinoma Associated 1

WHO – World Health Organization

Index:

Figure Index:

1 Preamble

Nowadays, cancer continues to be one of the major causes of death in the world. As so, it has been targeted as one of the most exhaustive investigation studies in what comes to its physiological pathways of both initiation and development, what still remains a bit unclear.

With some of its physiological routes being found, cancer therapies have been developed in order to fight against the disease but in a non-global way, as some of the patients would still not respond to the therapies. With technological advances, genomic sequencing was enabled, and some genetic therapies have started to be develop in combination with the previous ones, in order to strike the cells genetically, promoting its apoptosis and the regression of the tumour.

Nevertheless, some cancer cases continue to show non-responsive to the concomitant therapies, and therefore a new field of genetics has started to emerge as one viable option for cancer treatments- cancer epigenetics.

Epigenetic information is based in one individual's gene expression, which its specific mutations along the DNA structure may change, without modifying the DNA sequence. DNA methylation, histone post-translational modifications and chromatin structure regulators are the main responsibles for epigenetic mutations mechanisms.

In tumorous cells, these epigenetic mechanisms are taken over by the cells malignant state, being aberrantly expressed and regulated, in order to avoid immune restriction and contribute to the cancerous cells' growth.

Therefore, there has been studied a pharmacological pathway to modulate these epigenetic mechanisms in its various presentations, as 'writers', 'readers', 'erasers', or 'remodelers', in order to cease its antitumorous action and/or trigger immune responses.

1.1 Objectives

The main objective of this monography is to understand the main mechanisms behind epigenetic traits and to be able to connect them with the events that have been described among the literature with various cancer episodes and pathways. Additionally, this monography also shows some of the most recent work being made among epigenetic diagnostic and therapeutic findings across the world.

The highlights of this monography are to understand:

- How epigenetic traits are passed down through different generations;
- How epigenetic marks can alter the expression of gene clusters and, therefore, the regulation of its proteins, taking part, ultimately, in cancer pathways;
- How genomic imprinting works and how it can affect cancer prognosis through loss-of-imprinting;
- How current advances in diagnostic and therapeutic cancer epigenetics are being used among the different types of cancers.

1.2 Materials and methods

The main source of information in this research was PubMed, Google Scholar, the International Agency for Research for Cancer (IARC), the New South Wales Government Health Centre for Genetics Education and the New South Wales Government eviQ. The papers referenced in this monography were from scientific journals related to genetics, epigenetics, oncology, among others.

The research was exclusively done in English. Some examples of keywords used in this search were: cancer epigenetics, epigenetic reset, DNA methylation, histone modifications, noncoding RNAs, and cancer epigenetic therapies. The utilised papers were mostly published between 2018 and 2021, nevertheless some articles were less recent in order to make historical background or to give context to the subject.

2 Introduction

2.1 Cancer 101

2.1.1 Cancer nowadays

Right after Cardiovascular-related diseases, Cancer leads as the 2nd major cause of death worldwide. Accounting almost 10 million deaths in 2020, the most prevalent cancer types were breast and lung, each accounting more than 2 million cases, colon and rectum with more than 1.9 million cases, prostate with 1.4 million cases, and skin and stomach cancers, averaging 1.2 and 1.1 million cases, respectively, according to the World Health Organization (WHO). The most fatal ones were lung with 1.8 million deaths, colon and rectum with 935 thousand deaths, liver with 830 thousand deaths, stomach and breast, each with 769 and 685 thousand deaths, respectively. (1,2)

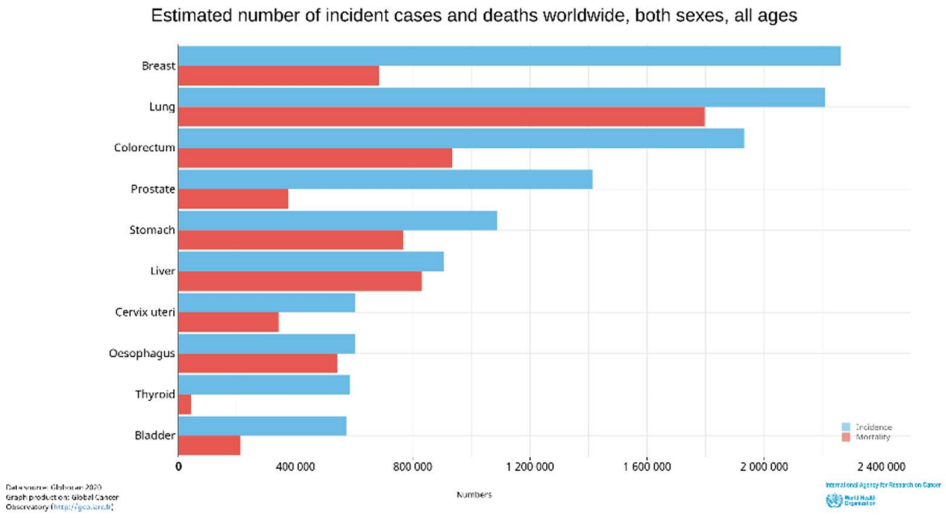

Figure 1 - WHO data on cancer cases and mortality, from IARC 2021(2)

Carcinogenesis is the process in which a normal cell transforms into a tumorous one. It is a multi-staged process that can arise from physical, chemical and/or biological agents, known as carcinogens. Its prognostic worsens with age and in some cases it is gender-associated. Among the most common and day-to-day risk factors are tobacco and alcohol use, abuse of unhealthy diet and the lack of physical activity. Some chronic infections may act as catalysts for cancer, being classified as carcinogenic infections.

Nowadays, the most crucial step in treating cancer is an early detection and diagnosis. The sooner the cancer is detected and evaluated, the brighter the prognosis is and the less invasive and belligerent it is to the organism. A useful tool that has only

recently began to be exploited is cancer screening, allowing genetic cancer-characterisation and adapting therapies through it. Cancer's genetic screening allowed the development of a data-base of genomic information along cancer types and specific mutations that turn the cell more vulnerable to oncogenesis, thus called oncogenes.

2.1.2 Cancer immunity-cycle

Cancer cells do not act as normal cells, as they do not seek its prosperity or its host's. Cancer cells metabolism is altered in a way that they only seek unbridled growth, seeing no means to its end. As so, they are seen as 'foreign bodies' by the host, and the host's immune system should, theoretically, act against the threat.

A theoretical cancer-immunity cycle was proposed as described on Figure 2 - Cancer Immunity-Cycle, from Cao et al., 2020(3), at first on 2013, by Chen, *et al*., and most recently revised in 2020 by Cao, *et al*. Theoretically, as a foreign body, upon the cancer-cell's death, (I) tumour-associated antigens that are released on the tumour microenvironment (TME) are later (II) presented to dendritic cells (DCs) which will be processed and later presented in the DCs surface of the Major Histocompatibility Complex (MHC) traveling to the lymphoid organs, in which (III) naïve T cells would recognize selected peptide-MHC complexes through T-cell receptors (TCRs), triggering the priming and activation of effector T-cells. Later on, (IV) differentiated effector T-cells would leave the lymphoid organs, traveling downstream blood vessels until they find the same tumour-antigens, infiltrating (V) the tumour's bed and migrating to the TME, becoming tumour-infiltrating lymphocytes (TILs), which would (VI) recognize and kill the cancer cells by direct or indirect immune attack, leading to the release of additional antigens, triggering a new round of the cancer-immunity cycle.(3,4)

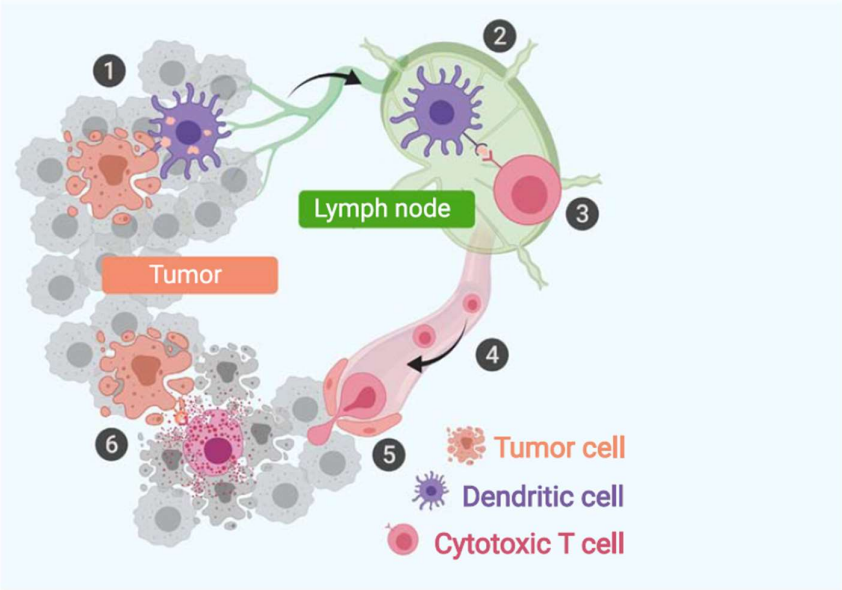

Figure 2 - Cancer Immunity-Cycle, from Cao et al., 2020(3)

2.1.3 Current cancer therapies

As spoken earlier, an early and correct diagnosis is a crucial part of any cancer's treatment, as it defines the therapeutic regimen according to the cancer's type and stage. Among nowadays cancer therapy's treatments are chemotherapy, radiotherapy, surgery, and most recently, immunotherapy.

Chemotherapy consists of using cytotoxic agents to induce apoptosis in the tumorous cells, although it is not cell-specific, as it devastates every cell type that is susceptible to the cytotoxic agents, leading to the frailty of the immune system and one of the most common and noticed secondary effects, loss of partial or total hair, according to the intensity of the treatment.

Radiotherapy is based in the destruction of the tumour cells by applying ionizing radiation directly in the tumour, which due to its constant state of replication, is more susceptible to radiation than normal healthy cells. Radiotherapy aims to damage the tumorous cells DNA, preventing its replication and causing eventual apoptosis of the cell. It targets both single and double-stranded DNA, as double-stranded DNA damage is much tougher to repair, causing the cell's apoptosis, and single-stranded DNA damage aims to hamper replication and cell growth.

Surgery is the most invasive therapy that is currently applied in cancer, as it consists of the physical removal of the tumorous-mass. It is the safest therapy when

regarding side-effects, but as it consists in an open-body surgery, it has all the other possible secondary complications of every surgery. It also depends on the location of the tumour, some of them not being able to undergo surgery as it may be located in an extremely delicate area.

Immunotherapy involves the usage of components of the immune system to counter the evasion strategies used by tumour cells to boosts the host's immune system, actively or passively. Active immunotherapy enhances the host's immune system, in order to boost its response to the tumour foreign mass, whereas passive immunotherapy consists in the transfer of monoclonal antibodies along with adoptive cell transfer. As immunotherapies we have T-cell based therapies as tumour-infiltrating lymphocytes (TILs), TCR-T cells, CAR-T cells and γδ T-cells. A downfall on the usage of T-cell therapies is that the cells must be sensitised to produce toxicity against the cancerous mass presenting tumour-associated antigens (TAAs), although this is not experienced when using NK cells therapy, as they act in a TAA-independent MHC-unrestricted approach thus not needing to be sensitised as they present significant toxicity against tumours. There is also dendritic cells (DCs) therapy, based on exposing DCs obtained *ex vivo* by previously collecting peripheral blood mononuclear cells (PBMCs), to tumour-derived antigens and then injecting them as a vaccine once again in the host, triggering the immune response in the host.(5)

2.2 Epigenetics 101

2.2.1 A brief history

It was late in the eighteenth century that it first started been debated acquired tracks provenance, through natural selection or heritability. Entering the nineteenth century, the observation of highly dense structures in the cell's nucleus began to awaken the insight of scientists as the definition of chromosome and chromatin were created. In 1910, Thomas Hunt Morgan was the first to experience heredity with his *Drosophila* studies, implying that chromosomes must encode genetic information, which would then lead to proteins, being asserted by Watson and Crick, in 1953, when presenting the double-stranded helix of DNA.(6)

Since Edwin Southern's article publication in 1975, regarding a gel electrophoresis method for separating and detection of specific DNA sequencies – known nowadays as the Southern Blot – there was a large amount of fieldworks developed around the premise known as the central dogma of molecular biology – DNA begets RNA, which then begets proteins – that enabled a considerable advance in molecular biology science, gaining a wild turnaround with the discovery of the reverse transcriptase (RT) which refuted this unidirectional idea of gene expression.(6)

Three centuries and a considerable amount of approved and revised models later, it is now acknowledged that heritable tracks begin with the DNA sequence and can then be affected at the transcriptional, translational, and post-translational levels, by both external and internal agents and factors.

2.2.2 Epigenetic mechanisms landscape

2.2.2.1 Genetic coding and epigenetic regulation of transcription mechanisms

Genetic cell information is transmitted via DNA sequence mutations, which translate into individual-specific structural modifications in the proteins and its expression on the organism. On other hand, epigenetic information is based on an individual's gene expression, according to its epigenetic landscape among the DNA sequence, which mutations may lead to structural changes in the DNA sequence's chromatin, without modifying the DNA, or regulatory alterations, as up- or downregulation of genes, silencing of enhancers and/or promoters are some of the most common epigenetic marks found among epigenetic mutations.

2.2.2.1.1 Chromatin-associated mechanisms

Altering the chromatin involves a whole process of dynamic changes of the chromatin structure and organization which may influence gene transcription regulation, DNA replication and repair, chromosome condensation and segregation and, ultimately, apoptosis. By default, chromatin remodelling is perceived as the sum of the events in the 3D chromatin configuration, such as histone modifications and DNA methylation, but it is also influenced by numerous intrinsic factors that contribute to its landscape, such as non-coding RNAs, and a great variety of enzymatic complexes.(7)

2.2.2.1.1.1 DNA methylation

DNA methylation consists in post-replication modifications based on the addition of a methyl group at the carbon-5 position of the pyrimidine ring of cytosine, 5-methilcytosine (5mC), out of the action of DNA methyltransferases enzymes (DNMTs), and can occur in any type of DNA sequence, coding or non-coding. Methylation plays a fundamental part in several processes such as regulation of transcription, genomic imprinting, and cellular reprogramming and differentiation. It occurs preferentially in CpG dinucleotide sites clusters, forming the so-called "CpG islands", this latest taking place mostly on promoters, been registered more than 28 million CpG dinucleotides in the human genome, of which 70% is normally methylated in somatic cells. (7–11)

DNA methyltransferases enzymes (DNMTs) operate using S-adenosylmethionine (SAM) as a methyl donor, converting it to S-adenosylhomocysteine (SAH). There are three DNMTs that are essentially and directly responsible for the DNA methylation and methylation patterns, which are DNMT1, DNMT3A and DNMT3B, the last two being responsible for *de novo* methylation and the DNMT1 assuming a maintenance and control role in DNA methylation during replication. (7–11)

DNA methylation is an extremely sensitive process in which the slightest mutation can result in aberrant modifications in regulation patterns and procedures. The two main dysregulations regarding DNA methylation are both the lack and the excess of methylation, namely DNA hypo- and hypermethylation. DNA hypomethylation can be found genome-wide and is mostly related with gene activation and chromosomal instabilities, whilst DNA hypermethylation is most commonly found in gene promoters

regions, concerning genes associated with tumour suppression, decreasing this latest's expression, leading to a more prone disposition to cancer. (7–11)

2.2.2.1.1.1.1 DNA-repair Epigenetics

One of the most recent discoveries in DNA methylation is the presence of DNA methylation reshaping enzymes such as the ten-eleven translocation (TET) enzymes which can oxidise 5mC to 5-hydroxymethylcytosine (5hmC), and further oxidise it to 5-formyl-cytosine (5fC) and 5-carboxylcytosine (5caC), the last two being capable of getting enzymatically excised by the action of the thymine DNA glycosylase (TDG) enzyme in the base excision repair (BER) pathway and replaced with an normal unmodified cytosine, contributing to DNA methylation dynamics.(11–13)

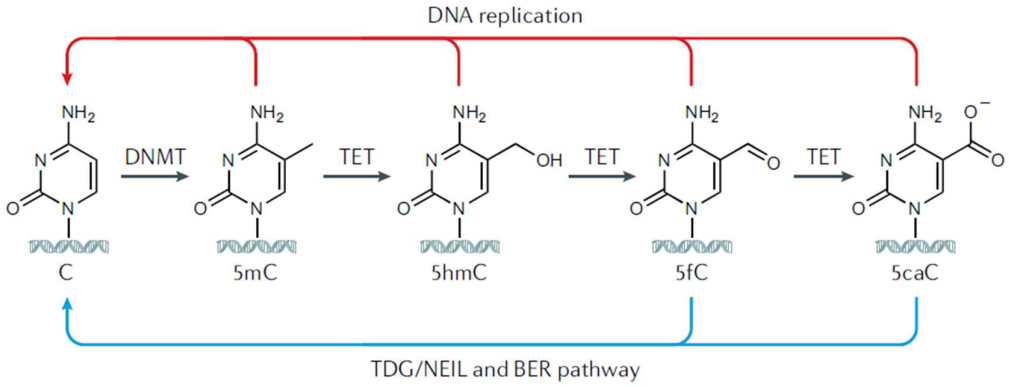

Figure 3 - TET enzymes MoA, from Parry et al., 2021(12)

There are three TET proteins present in the mammalian cell lineage, which are TET1 (mostly expressed in primordial germ cells (PGCs)), TET2 and TET3 (being mostly expressed during the rest of the development).(13)

The oxidative derivates of the action of TET proteins, such as 5hmC, 5fC and 5caC, can also be considered reliable epigenetic marks which can be precisely recognized and bound to "readers" such as transcriptional regulators, thus impacting genomic integrity, transcriptional elongation and even DNA repair. For instance, it is documented that there are a greater number of proteins which have more specific interactions with 5fC and 5caC than with 5hmC, and, likewise, during the BER pathway 5fC and 5caC are able to gather a higher number of DNA repair proteins when in comparison to 5hmC. On the other hand, modified cytosines can also inhibit the binding of transcriptional enhancers, and therefore reduce gene expression, whilst oxi-mCs can

prevent the binding of transcriptional silencers and hence stimulate gene expression.(13)

Additionally, both TET1 and TET2 can synthesise 5hmC in mRNAs that are essential to pluripotency, which can lead to reduced mRNA stability. As a result, pluripotency genes do not attain the expression levels they should and therefore repress the expression of lineage-specifying factors.

2.2.2.1.1.2 Histone modifications

The primary unit of chromatin is the nucleosome, and it consists of two copies of each one of four types of histone proteins, making a histone octamer (H2A, H2B, H3 and H4), each nucleosome containing 147 base pairs (bp) of DNA swathed around it.

Chromatin structure regulation and, consequentially, DNA transcriptional activity are both affected by histone modifications. Histones are composed by a globular C-terminal domain and N-terminal tails, conferring them a basic trait. The last ones can suffer different posttranslational modifications (PTMs). PTMs control the accessibility of the DNA to the interacting proteins by adding or removing the chemical group on the amino acids arginine, serine, or lysine, modulating the interaction between DNA and the histone octamers, such as histone acetylation, methylation, phosphorylation or ubiquitination, as further described below.(7)

Histone acetyltransferases (HATs) and histones deacetylases (HDACs) are, correspondingly, responsible for histone acetylation and deacetylation, and occur on the amino acid lysine. While histone acetylation causes chromatin looseness, by neutralizing the positively charged lysines thus distressing the interactions with the negatively charged DNA, it creates a more auspicious environment for gene transcription by having a better access to DNA content, histone deacetylation otherwise causes gene silencing by chromatin constriction, as it decreases transcription factors accessibility to DNA. Deregulations in any of the previously mentioned acetylation processes can lead to both excessive transcription and silencing of genes, respectively, which can lead to future complications, such as tumorous malignancies.(7)

Histone methylation consists in the addition of a methyl group by histone methylases (HMTs), while histone demethylation is performed by histone demethylases (HDMs), and occur on lysine, histidine and arginine amino acids. The effect in chromatin regulation differs within the affected histone amino acid's location, as

Histone H3 possesses different lysine methylation locations, which can be mono-, di-, or trimethylated, such as K4, K9, K27, K36, and K79, and each one can have different effects in the chromatin's structure: methylation of H3K4 results in chromatin relaxation, whilst methylation of H3K9 results in chromatin condensation. Further studies demonstrated that histone methylation can act as markers for equally active or silenced gene transcription, as in H3K4me3, H3K36me3 or H3K79me3 for transcriptionally active genes and H3K9me2, H3K9me3 or H3K27me3 for transcriptionally silenced genes.(7)

Histone phosphorylation befalls principally on threonine, serine, and tyrosine amino acids located in the N-terminal histone tails, by the action of kinases and phosphatases, resulting in the addition of a negative charge, a phosphate group, that comes from an ATP molecule, subsequentially causing a lower positive charge in the histone tails, therefore interfering with the chromatin structure and, consequentially, affecting its interactions with transcription factors. A well-documented histone modification is the phosphorylation of H3S10 executed by the Aurora kinases which leads to chromosome condensation during mitosis and the dephosphorylation of the same site by PP2A are both associated with gene silencing, whereas the same histone phosphorylation carried out by the MSK/Jil1 family results in chromatin relaxation, acting as a positive transcription regulator.(7,9)

Histone ubiquitination is primarily found on H2A and H2B, and is described as a signal for ensuing histone modifications as it recruits several other histone modification machineries. Histone ubiquitination is also found to be able to have both a positive or a negative role in transcription and gene activation. A well-documented case is the monoubiquitination of H2AK119 which is executed by RING1A/B in the PRC1 complex, being associated with chromatin condensation and, thus, gene transcription silencing, and its removal by the BAP1 deubiquitinase complex. On the other hand, H2BK120 monoubiquitination is correlated with positive gene transcription and is highly associated with elevated levels of methylation on H3K4 and H3K79, both histone methylations correlated with gene transcription activation, as described earlier in this chapter. This phenomenon comes described as a crosstalk among epigenetic histone modification factors, such as the just describe H2BK120ub1 comes as a precondition for both H3K4 and H3K79 methylation, but H3K4 methylation automatically inhibits H3R2 methylation, and vice-versa, precluding the two traits of

coexisting. Howbeit, there are phenomena which are dependent of each other to prevent malignancies, as, for example, the BAP1 H2A deubiquitinase enhances monomethylation of gene enhancers by calling up H3K4 monomethylase MLL3, whilst the lack of this interaction between BAP1 and MLL3 is described to contribute to several cancers' pathogenesis. Other associations between histone-modifying complexes are also found to have interactions within transcription factors' regulation and gene stability and activity, such as lysine methyltransferases (KMTs) and HDACs. (7,14)

Through further investigation of these histone-modifying processes, a large number of other modifications have been found to influence chromatin organization and, therefore, gene transcription. Some of these histone modifications are lysine crotonylation, butyrylation, propionylation, tyrosine hydroxylation, biotinylation, neddylation, sumoylation, O-GlcNAc, ADP ribosylation, N-formylation, proline isomerization, and citrullination.(14) TET proteins have also been found to regulate gene expression through interactions with other proteins that interfere in chromatin structure, such as the TET1-SIN3A complex, which together with HDACs 1 and 2, have been found to have the capacity to supress transcription by histone deacetylation mediation.(13)

2.2.2.1.2 Non-Coding RNAs

Non-coding RNAs (ncRNAs) are functional, transcribed, but untranslated RNA molecules which can regulate the cellular epigenetic landscape post-transcriptionally, through gene expression activation or silencing. ncRNAs are present in the majority of the biological processes happening in one's organism, such as DNA synthesis and genome rearrangement, gene expression regulation in both transcription and translation, and RNA processing. ncRNAs can act by targeting chromatin in order to directly cause gene silencing through transcription factors blockage, as they can act by direct interference with transcription factors to either induce or inhibit the expression of certain targeted genes, and can also promote gene silencing by binding to mRNA directly, causing the so-called RNA-induced silencing.(9,15)

ncRNAs can be divided into two groups, according to its length: the short non-coding RNAs (sncRNAs) which are no longer than 30 nucleotides (nts) and the long non-coding RNAs (lncRNAs) which are longer than 200 nts.(9)

2.2.2.1.2.1 Small ncRNAs

Micro RNAs (miRNAs) are a class of short (17-25 nts) single-stranded RNAs (ssRNAs), formed by the cleavage of a double-stranded RNA (dsRNA), that are partially complementary to the 3'- untranslated region of mRNAs. miRNAs are responsible for modulating nearly 30% of the human genes' expression, taking part in various major cellular processes such as cell differentiation, cell cycle, immune function, and finally apoptosis, through direct interaction with the transcript mRNAs, binding to them and causing their degradation or inhibiting their translation, repressing their action, exerting the previously mentioned RNA-induced silencing at the post-translational level.(7,9,15)

miRNAs biogenesis, as further illustrated in Figure 4 - miRNAs biogenesis pathway, from Kumar et al., 2020(15), starts in the nucleus, forming a double-stranded, stem-loop structure, designated primary miRNA (pri-miRNA), which is later on processed by the Drosha-Pasha complex (RNAse III enzyme and its cofactor, respectively) into pre-miRNA. Pre-miRNA is them transported to the cytoplasm where it is cleaved by the ATP-dependent protein/enzyme Dicer into two single-stranded miRNAs creating a miRNA duplex, where the functional strand is captured by an ARGONAUTE protein family member, forming the RNA-induced silencing complex (RISC), which then delivers the miRNA to its targeted mRNA, causing its post-translational silencing, preventing its translation or triggering its degradation, while the non-functional strand proceeds directly to degradation.(15)

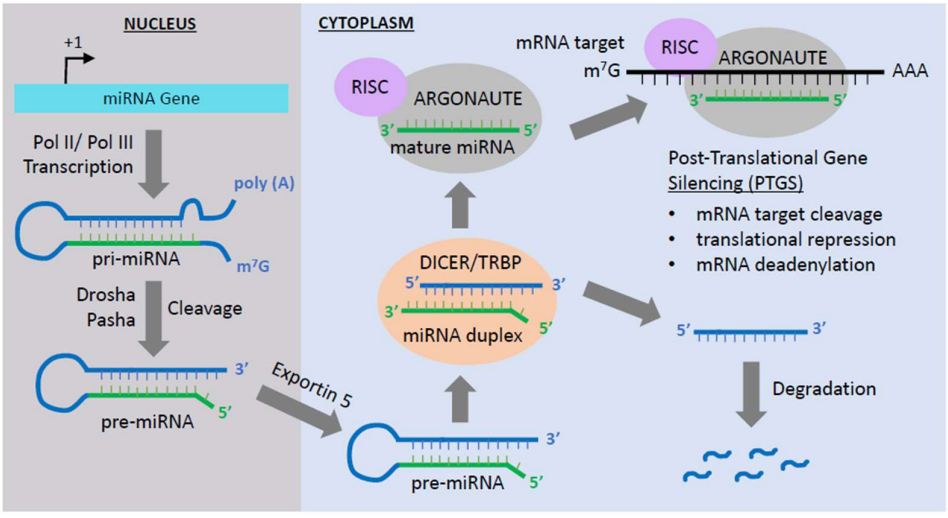

Figure 4 - miRNAs biogenesis pathway, from Kumar et al., 2020(15)

Another class of sncRNAs are small interference RNAs (siRNAs), used mainly as gene silencers, as it has been proved to contribute to the reduction of the expression of specific genes. Its use has been refined by the implementation of these RNA interference (RNAi) in the shape of short hairpin RNAs (shRNAs) in plasmid vectors, as shRNAs were capable of incorporating in the DNA.(15)

miRNAs can regulate the epigenome and alter the chromatin's landscape through the repression of key epigenetic modifiers by post-transcriptional gene silencing (PTGS) and mRNA degradation. Such epigenetic factors may include HDACs, HMTs and DNMTs, which traits can be specifically targeted and repressed by the upregulation of certain miRNAs, causing the reactivation of genes that were initially silenced by those modifiers. As examples, it is known that miR-29 expression causes DNMT3A and DNMT3B repression at the PTGS level, while miR-148 only act against DNMT3B but can be affected by both mRNA degradation or translation repression, and, finally, multiple miRNAs are known to target DNMT1, such as miR-148a, miR-152, miR-185 and miR-342. (15)

miRNAs expression can be equally regulated by other epigenetic traits, such as DNA methylation, for instance: in approximately half of all miRNA genes there were found CpG islands, which are usually found in gene promoters, meaning they can equally suffer modified DNA methylation, and thus lead to either miRNA expression upregulation or downregulation..(15)

2.2.2.1.2.2 Long ncRNAs

lncRNAs are a class of ncRNAs about 200 nts long and are mostly associated with gene repression, although they can also act in chromatin remodelling to induce or inhibit transcription, modulating pre-mRNA splicing and mRNA translation repression. lncRNAs can also have different molecular functions by which we are able to distinguish them, as they can act as indicators of gene activation or repression, guides to direct chromatin modifiers to targeted genomic loci, distractions by dislocating transcriptional inhibitors, scaffold for various protein complexes and competing endogenous RNAs (ceRNAs). (15)

lncRNAs have numerous physiological roles, the most valuable including X-chromosome inactivation, chromatin landscape remodelling and imprinting. As an example, lncRNA XIST is a ncRNA about 15-17 kb long and encountered solely in the

nucleus. which controls X-chromosome inactivation. This process occurs in females as a gene dosage compensatory system to equalize gene expression with males, which have only one X-chromosome. XIST mechanism of action consists in recruiting inhibitive epigenetic factors such as polycomb repressive complexes (PRC) 1 and 2, and HDACs, which result in gene suppression through chromatin compaction. As XIST, numerous lncRNAs can interplay with epigenetic modifiers and regulate gene expression though chromatin modification.(15)

There are other ncRNAs which play specific roles, such as small nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs), transfer RNAs (tRNAs), ribosomal RNAs (rRNAs), and piwi-interacting RNAs (piRNAs). (15)

snRNAs are mainly encountered within the splicing regions of nuclei, about 150 nts long, and are mainly correlated with pre-mRNA splicing regulation, specifically with the spliceosome formation, as it is a dynamic complex formed of the five snRNAs (U1, U2, U4, U5, and U6) and other splicing protein components, creating a small nuclear ribonucleoprotein (snRNP) complex. (15)

snoRNAs are ncRNAs essentially located in de nucleoli of eukaryotic cells, functioning as modifiers and promoters to rRNA processing, specifically during the ribosomal subunits' synthesis. snoRNAs play a key role in the methylation process of the ribosomal subunits, known as 20O-ribose-methylations, and in the conversion of uridine to pseudouridine, allowing the maturation of rRNAs, process known as pseudouridylation. (15)

rRNAs and tRNAs are long-known and studied ncRNAs, with well-defined functions in mRNA translation. rRNAs act in the attachment of the ribosomal subunits and tRNAs in the allocation of amino acid subunits into the ribosome throughout mRNA translation, both being essential in protein synthesis. (15)

piRNAs are ncRNAs about 24-32 nts long, largely linked to transposons, acting in DNA arrangements regulation, mRNA turnover and epigenetic programming, and playing a key role in protecting the genome from intrusive transposable elements in the germlines of animals, via gene repression. piRNAs are synthesised after the transcription of genome regions known as piRNA clusters, and depend on the interaction with PIWI proteins (a subfamily of ARGONAUTE proteins) to accurately regulate their targets. (15)

As we can see, both sncRNAs and lncRNAs can play key roles in epigenetic modulation, specifically through gene transcription modifications as an answer to both intracellular and extracellular stimuli.

2.2.3 Genomic imprinting

Imprinted genes are genes that are reversibly marked on one of the two homologous loci during their development, resulting in a functional non-equivalent gene expression. This mark consists in a parent-to-offspring transmission of epigenetic traits which regulate gene expression to a specific allele.(16–18)

It is well-known that DNA methylation is the main mechanism behind genomic imprinting, as its ability to methylate CpG islands is essential in supressing promoters and, thus, repressing transcription. Imprinted genes can be found in clusters of more than three or four genes, and they are able to code to at least one ncRNA. Genes that are in between these clusters are most commonly imprinted and therefore supressed to a monoallelic expression, with some occasional exceptions in which they may evade imprinting, and there can also be imprinted genes out of these clusters, as they are designated as micro-imprinted loci.(16–19)

Even though, gene transcription regulation within these loci depends on multiple epigenetic marks. The two imprinted alleles need to be marked differently in order to correctly regulate the transcription process. There are differentially methylated regions (DMRs), known as Imprint Control Regions (ICRs), which present specific patterns of parental methylation, being the regulatory elements responsible by regulating the monoallelic expression of the loci, regulating the chromatin structure via DNA methylation and histone alterations. ICRs are of great importance since its deletion implies a process known as loss of imprinting (LOI), in which the parental-origin of the gene is lost and hence its expression is compromised and, in some cases, may lead to cancer. Genomic imprinting occurs in different occasions in males and female gametes, paternal imprint of genes occurs prenatally in pro-spermatogonia before meiosis, while maternal methylation of ICRs occurs postnatally in growing oocytes.

There are also specific cell types that can selectively “switch off” or lose its genomic imprint in specific checkpoints in one’s development, in order activate a commonly repressed allele.(17,18)

2.2.3.1 Epigenetic Reset

During embryonic development, there exists both gains and losses in the DNA methylation levels and, therefore, epigenetic marks. The first demethylation process occurs after fertilization where both parental methylation patterns are erased, with the exception of regulatory epigenetic regions that endure this global demethylation, such as ICRs. This demethylation process occurs initially at the male pronucleus, as active demethylation, and then as both an active and passive demethylation, in the maternal pronucleus, and is regulated by the TET proteins. At the end of this first demethylation phase, the methylation levels are reduced from nearly 70% in the zygote to approximately 25% in the pre-implantation epiblast.(12,17–20)

Then, after implantation, occurs *de novo* methylation, stabilizing the overall DNA methylation levels, with the exception of germ cell lineage. Since primordial germ cells (PGCs) possess the methylation patterns of somatic cells, they undergo a second demethylation process, allowing the removal of parental epigenetic traits in order to re-establish new parental-specific marks at ICRs in both oocytes and sperm as the development proceeds, according to the sex of the offspring, ensuring the inheritance of paternal or maternal imprinted gene allelic expression, respectively, throughout the generations. This germline ICR methylation is carried out by DNMT3A with its co-factor DNMT3L, as TET1 and TET2 are downregulated, and also takes place at different times in females and males. In somatic cells, the latest acquired methylation patterns must be protected against the extensive demethylation processes in order to be kept and transmitted to new somatic cells. (12,17–20)

3 Cancer Epigenetics

3.1 Epigenome-modulating agents/factors

Among the epigenetic changes that may occur at any stage of the cancer's lifespan, there is a distinguish classification between the genes which inactivation would be beneficial to the tumour's opulence – the tumour supressing genes – and the ones which activation would benefit the tumours growth and development - oncogenes. Feinberg *et al.* proposed a functional classification of these genes, according to its role within the epigenetic modification.

As so, Feinberg *et al.* classified the genes that participate in cancer epigenetic modification in three classes.

3.1.1 Epigenetic Modifiers

Genes which mutation directly modify the DNA chain structurally, as with DNA methylation, chromatin structural modifications or post-translational chromatin modification. Haematological malignancies such as Acute Myeloid Leukaemia (AML) and T-cell Lymphoma are commonly found modified at the DNA-methylation level, being regularly found mutations in the DNA methyltransferase3α (DNMT3A) affecting haematopoietic stem cells (HSCs) differentiation, whilst chromatin-remodelling mutations are most commonly found in solid tumours, having exposed the SWI/SNF chromatin remodelling complexes as one of the most prominent mutations sites. As histone-modifying enzymes mutations, increased H3K27 trimethylation (H3K27me3) results in gain-of-function and amplification mutations in non-Hodgkin lymphomas and some solid tumours.(21)

3.1.2 Epigenetic Mediators

The targets of epigenetic modification, genes that turn the cells more prone to carcinogenesis, giving 'birth' to the so-called cancer stem cells (CSCs). They are responsible by stem cell reprogramming, genes mutations and regulatory growth factors such as the insulin-like growth factor 2 (IGF2), in which its loss of imprinting (LOI) results in its doubling of production and consequentially downstream regulation in various tumours.(21)

3.1.3 Epigenetic Modulators

Placing them upstream of the modifiers and mediators, these genes are responsible for initiating the carcinogenic cascade of events and can result of both internal as/and external stimuli, such as environmental agents, nutrition, stress, and even natural ageing. Natural inflammatory processes in the organism can trigger an exacerbated response that may cause epigenetic mutation the same way tumour-supressing genes can be viewed as epigenetic modulators. While natural ageing can modulate epigenetic alterations via age-related DNA-hypomethylation and heterochromatin loss, environmental factors such as methyl-deficient diet or contact with carcinogenic agents, less natural agents, can also result in epigenetic modulation causing hypo- or hypermethylation. (21)

Table 1 | **Three classification systems for cancer genes**

Class	Definition	Examples
Genetic classification		
Oncogene	A gene whose activation by mutation is advantageous to the cancer cell. Acts as dominant	*MYC, KRAS, PIK3CA, ABL1, BRAF*
Tumour suppressor gene	A gene whose inactivation by mutation is advantageous to the cancer cell. Generally acts as recessive	*RB1, TP53, WT1, NF1, NF2, VHL, APC, CDKN2A*
Selection classification		
Driver gene	A gene whose mutation or aberrant expression is subject to selection during tumorigenesis	*MYC, KRAS, PIK3CA, ABL1, RB1, TP53, WT1*
Passenger gene	A gene mutated in cancer that is not a driver	Estimated as 99.9% of all mutational changes in cancer
Epigenetic functional classification		
Epigenetic modulator	A gene, mutated or not, that activates or represses the epigenetic machinery in cancer	*IDH1/2, KRAS, APC, TP53, STAT1/3, YAP1, CTCF*
Epigenetic modifier	A gene, mutated or not, that modifies DNA methylation or chromatin structure or its interpretation in cancer	*SMARCA4, PBRM1, ARID1A, ARID2, ARID1B, DNMT3A, TET2, MLL1/2/3, NSD1/2, SETD2, EZH2, BRD4*
Epigenetic mediator	A gene regulated by an epigenetic modifier in cancer (mutations rare or absent) that increases pluripotency or survival	*OCT4, NANOG, LIN28, SOX2, KLF4*

Figure 5 - Cancer genes classification, from Feinberg et al., 2016(21)

More recently, Cao *et al.* described the epigenetics mechanisms as a highly organized arrangement providing stable and durable gene regulation, which is driven by epigenetic regulators, allowing and coordinating the epigenetic alterations. With roughly more than 1000 epigenetic regulators been described among various articles, he categorized them in four classes: Epigenetic 'Writers', which are responsible by the addition of epigenetic marks; Epigenetic 'Erasers', being responsible by the removal of the epigenetic marks; Epigenetic 'Readers', recognizing the epigenetic traits and

coordinating the downstream effects accordingly to the specific traits; and Epigenetic 'Remodelers' – controlling the main-status of the chromatin.

Epigenetic features	Regulator	Gene family
DNA methylation	Writer	DNA methyltransferases (DNMTs)
	Reader	5-methylcytosine-binding domain proteins (MeCP2 and MBDs)
	Eraser	Ten-eleven translocation dioxygenases (TETs), ALKBH1
Histone modifications	Writer	Lysine methyltransferases (KMTs), protein arginine methyltransferases (PRMTs), lysine acetyltransferases (KATs or HATs), histone ubiquitin ligases, histone kinases, and others
	Reader	Chromodomain, Tudor domain, MBT domain, PhD finger, bromodomain-containing proteins
	Eraser	Lysine demethylases (KDMs), histone deacetylases (HDACs and SIRTs), histone deubiquitinating enzymes, histone phosphatases, and others
Chromatin structure	Remodeler	SWI/SNF, ISWI, CHD, and INO80/SWR complexes

Figure 6 - Epigenetic regulators according to Cao et al., 2020(3)

3.1.3.1 Extracellular Vesicles (ECVs)

Recent work from Kok *et al.,* who pursued the research of Reilley *et al*., Nixon *et al*., and Shah *et al*., describes extracellular vesicles (ECVs) as potent vehicles of epigenetic modification.(22) Extracellular vesicles are membrane-bound organelles that are found in circulation and contain different types of cell-derived biomolecules such as DNA, RNA, proteins, lipids and metabolites.(22,23) ECVs can have various sizes, depending on their place of origin and molecular constitution, and different functions, being able to act as biomarkers for disease, catalysts for cancer and vehicles for epigenetic inheritance. (22–24)

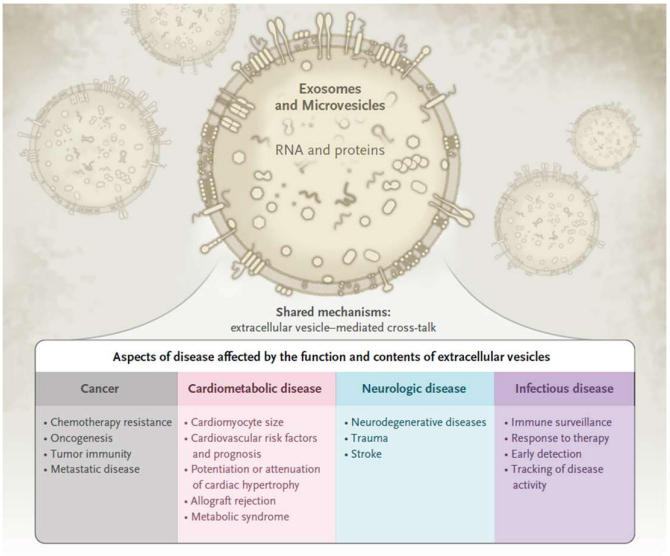

Figure 7 - ECVs possible role in prognosis and diagnosis of various diseases, from Shah et al., 2018(22)

ECVs can be found in most of the biological fluids of the organism (e.g. blood, breast milk, saliva, urine, and cerebrospinal fluid).(22) The main component of ECVs that can have regulatory effects are micro RNAs (miRNAs) and small RNAs derived from small regulatory tRNAs, which will play a role in intercellular communication with the transport of these proteins and RNA to one cell to the others, via blood vessels or neighbouring tissue by paracrine transportation.(22,24)

ECVs may act as cancer catalysts by funnelling chemotherapeutic agents out of the cell and/or by the transport of carcinogenic agents that can redirect immune response away from cancer cells, such as the human epidermal growth factor receptor 2 (HER2) in breast cancer, for example.(22) Recent research by Kok *et al.* also describes ECVs not only as vectors of angiogenesis recruiting agents as they carry such receptors, but also as axonogenesis agents being able to recruit nerves via the release of neurotrophic factors and axonal guidance molecules, being considered a new possible feature of cancer growth.(23) Cancer-derived exomes (CDEs) present specific exosomal markers such as tetraspanin proteins CD63, CD9, and CD81 that allow sorting, selective recruitment, capturing, or profiling of CDEs, being a powerful tool for prognosis and diagnosis and for future therapeutics using these ECVs as vehicles to selectively attack the cancer cell.(22,23)

In what concerns epigenetic inheritance, Sharma *et al*., Reilley *et al*. and Nixon *et al*., found evidence of the assimilation of the epigenetic content of the ECVs in the maturing of spermatozoa in the epididymis. They discovered an alteration in the epigenetic background of the spermatozoa during its voyage along the epididymis, having different permeability and susceptibility to such ECVs in the different sections of the epididymis. From the *caput* to the *cauda* of the epididymis, spermatozoa showed an increase in miRNA signals corresponding to those marked and supplied by the investigators, particularly by the ones containing cell growth and proliferation information.(25–27) More recently, Spadafora proposed a model for this epigenetic reprogramming of early embryos by spermatozoa, in which (1) RNA-based information from tissues are packed in ECVs due to environmental stress or intrinsic processes of the cell which are, later on, released to the blood stream and (2) eventually taken up by the epididymal spermatozoa which internalize the information in its nuclei, generating an ‘RNA storage’ formed by a large amount of small regulatory RNAs that come from

different cell types and tissues as a result of different stimuli. Finally, (3) as sperm's RNA is known to persist in early-stage embryos, it will start remodelling the epigenetic background of the zygote soon after the fertilization proceeding until one- and two-cell embryonic nuclei, therefore been able to induce epigenetic alterations that may result in phenotypic variations.(24)

3.2 Cancer-related epigenetic dysregulations

During normal cell development, cells take different pathways of differentiation according to its progenitor cell, pathways that are irreversible under physiological conditions. Nonetheless, during carcinogenesis, cells are able to de-differentiate or transdifferentiate, leading to aberrant cell types that come as a combination of several lineages pathways.(13)

The LOI and, therefore, the dysregulation of the imprinting patterns have been found in several tumour cases, having being noticed as one of the most commonly and earliest events to take place in human cancers. LOI can lead to physiological disorders as a consequence of modifications in the genomically imprinted genes, since a young age to adulthood. Some of these physiological disorders result in human imprinting syndromes such as Angelman, Prader-Willi or Beckwith-Wiedemann, which result in significant neurological damage. As stated, carcinogenesis can be induced by dysregulation of the imprinted genes, as LOI is found to be one of the most common and early events that take place in a great variety of human cancers, such as meningiomas, gliomas, colorectal or esophageal cancer, and chronic myeloid leukemia and so on.(18)

3.2.1.1 DNA-modified methylation - Chain of events and specific cancer dysregulations

DNA hypomethylation was one of the first epigenetic modifications in tumours to be noted when compared to normal cell tissue. As it has been described as a normal event during ageing and senescence, the answer to the debate on how cancer might be an age-related disease lays on DNA methylation processes.(28) During carcinogenesis, the loss of methylation of DNA sequences, such as LINE and SINE sequences and Alu elements, and coding regions and introns, increases with the progression of the lesion from a benign cell proliferation to an invasive tumour. On the other hand,

hypermethylation of promoter regions of tumour suppressor genes happens in specific CpG islands in the DNA and inhibits its expression, contributing to the proliferation of cancers. In addition to gene supressing via CpG islands hypermethylation, and gene activation by DNA hypomethylation, hypermethylation of gene bodies promotes oncogene overexpression.(13,29)

TET activity is also modified during tumorigenesis, as it is reduced by tumour hypoxia, which leads to diminished TET enzymes activity by influencing increased promoter methylation. In Head and Neck Squamous Cell Carcinoma (HNSCC) there have been registered low levels of expression in TET1 and TET3 when their promoter methylation levels were reportedly higher in cancer cells, suggesting a correlation between the two events.(7) TET1 has also been found active in Acute Myeloid (AML) and Lymphoid Leukemias (ALL) as it was found to have a role in the translocation process of the mixed-lineage Leukemia (MLL) gene. On the other hand, TET2 has been associated with tumorous evens as its loss of function have been found to affect mast cell differentiation and proliferation as well as cytokine production. It has been strongly correlated with myeloid leukemias, myeloproliferative neoplasms and myelodysplastic syndromes (MDS), as TET2's loss of function in HSCs causes a state of increased self-renewal in stem cells, boosts the progenitor cell count and biases the differentiation of the stem cells towards the monocyte/macrophage lineage.(13) This event is described by Challen *et al.* as Clonal Hematopoiesis (CH), a physiological state where a certain lineage of cells grows at a far superior rhythm than the others, and can be affected by both defects in TET2 and in DNMT3A.(30)

In Colorectal Cancer (CRC) there has been documented solid evidence that transcriptional silencing of tumour-supressing genes in tumorous cells is correlated with promoter CpG hypermethylation, being some of the well-documented promoter CpG regions the ones in the CDKN2A, MLH1 and APC genes. On the other hand, genome-wide hypomethylation is also an early indicator of colorectal carcinogenesis. It is also documented that the grade of hypomethylation in CRC is directly correlated with the development stage of the disease, as it is observed along the different disease stages, since early adenomas to metastases and more complex adenocarcinomas.(28)

3.2.1.2 Chromatin modifications - Chain of events and specific cancer dysregulations

As described above, there exists a strict but frail regulation of the chromatin structure that controls gene transcription activation and repression, essentially out of the action of histones. Epigenetic modifiers and modulators can shift the histones state and cause modifications in the chromatin's structure that lead to the repression or overexpression of certain genes, and may unlock environments propitious to carcinogenesis.

Methylation of histone H3K4 is executed by methyltransferases of the COMPASS family, namely SET1A and SET1B, being responsible for tri-methylating H3K4 at promoter sites, and MLL1-4, at enhancers and promoter sites. SET1A has already been linked with several cases of cancer tumorigenesis through its action, specifically, breast, lung and colorectal cancer. MLL proteins have also been associated with tumorous events as MLL1 has been found to be frequently mutated in nearly 80% of childhood cases and from 5% to 10% in adult cases of AML and ALL. Likewise, MLL3 and MLL4 are also observed extensively mutated in tumours, as MLL3 is found more actively mutated at the plant homeodomain (PHD) gene cluster, causing interference within its interaction with the BAP1 tumour suppressor and therefore being associated with poor patient prognosis, while MLL4 mutations are more equally distributed along the protein.(14,31) H3K4 methylation levels are also found aberrantly elevated in CRC and in adenomas. (29)

DOT1L is the sole enzyme responsible for the methylation of H3K79. To date, there has not been found a single demethylase to take action in H3K79. DOT1L is the only lysine methylation enzyme that does not possess a SET domain and works jointly with MLL translocation partners in a complex named DotCom. Its action is associated with tumorous activity in breast cancer cases, namely with aberrant cell proliferation and metastasis, and in leukemias, as H3K79 is found oddly upregulated.(14)

Histone tails mutations, such as H3K27M, H3K36M and H3G34V/R, have also been found regular in multiple cancer types, as one common characteristic to all histones actively mutated in tumour cases is that they all block the correct deposit of the histone modification at the mutation site, causing transcriptional reprogramming and, in these cases, tumorigenesis. In diffused intrinsic pontine gliomas (DIPGs) there

have been found a specific set of modifications that are common to the vast majority of cases: global hypomethylation of H3K27me3 but with elevated levels of acetylation in the same H3K27, reduced PRC2 catalytic activity, and a specific single-nucleotide substitution culminating in H3.3K27M. Meanwhile, H3K36M mutations are usually found in chondroblastoma, HNSCC and CRC, and H3G34V/R mutations are a constant feature in glioma and bone cancers.(14,31)

Histone acetylation has already been linked to tumoral progression as high levels of acetylated H3K18 and low expression of H3K4ac has been associated with advanced stages of oral cancer.(7) Histone deacetylation by HDAC activity also constitutes a major factor in tumorous cell survivability, as HDACs overexpression claims very strong evidence on poor patient prognosis and low patient overall survival rate.(14) Histone acetylation levels in H3K27 and H4K12 are also found aberrantly elevated in CRC, when compared with normal mucosa.(29)

As described earlier, histone modifiers have interactions themselves and can harness different outcomes and different cell proliferation states, in both normal cells and malignant ones. One example is the synergy between BAP1 and MLL3, in which the first recruits the second in order to monomethylate gene enhancers, whilst the lack of this interaction leads to the pathogenesis of a great variety of tumours.(9)

Regarding the already reviewed TET proteins, besides its demethylating activity, TET2 also has a regulatory role in the recruitment of HDAC2 in order to supress IL6 expression. Zhang *et al.* studied this phenomenon in mice and came to the realisation that TET2-deficient mice would suffer an aggravated inflammation process due to the administration of lipopolysaccharide (LPS), and be more susceptible to an endotoxin-induced septic shock and colitis than the normal mice.(13,32)

3.2.1.3 ncRNA dysregulations - Cancer-induced susceptibility events (chain of events)

As discussed previously, ncRNAs act post-transcriptionally in the regulation of the cellular epigenetic landscape, by the up- or downregulation of gene expression.

MiRNA expression can be easily dysregulated due to a great variety of genetic modification, such as mutations, deletions, amplifications and translocations, as they are regularly found in delicate sites in the genome. Therefore, miRNAs can be up- or downregulated in cancer tissue, although they are most commonly upregulated, due to

both hyper- or hypomethylation, since both can modify miRNA expression, as a great variety of studies have already found aberrant miRNA expression levels in both tumorous tissue and tumour-adjacent tissues. MiRNAs can thus be described as oncogenic miRNAs (onco-miRNAs) or tumour-suppressive miRNAs (ts-miRNAs), as they can act by inhibiting the tumour-suppressive gene expression or the oncogene expression, respectively. (29)

When miRNAs that function as suppressors of DNMTs, such as DNMT1, DNMT3A, DNMT3B and EZH2, have their expression downregulated, the DNA methylation landscape can be abnormally changed, resulting in a more tumour propitious environment out of DNA hypermethylation. This environment ought to result in the abnormal regulation of certain gene targets, namely the reactivation of oncogenes or the downregulation of tumour suppressor genes, simplifying the tumorigenic process, as in its formation, evolution and metastasis.(15) In Figure 8 - MiRNAs targeting epigenetic regulators adapted from Kumar et al., 2020(15) we have a list of miRNAs that have been found to regulate DNMTs and HDACs, such as the ones we previously referred to.(15)

miRNA	*Targets*	*Function*	*References*
miR-29 a, b, c	DNMT3A and DNMT3B	Tumor suppression by repression of de novo DNA methylation. Protects tumor-suppressor genes from been silenced by DNA methylation.	Fabbri et al., 2007 [44] Suzuki et al., 2013 [47]
miR-148	DNMT3B DNMT1	Negative feedback loop between DNMT1 and miR-148 in AML. Inhibition of cell proliferation and increase of apoptosis.	Duursma et al., 2008 [45] Wang et al., 2019 [48]
miR-449a	HDAC1	Inhibition of tumor growth, invasion and metastasis. Promotes apoptosis and differentiation.	Noonan et al., 2009 [49] Yong-Ming et al., 2017 [50]
miR-152 *miR-185* *miR-342*	DNMT1	DNA hypomethylation. Promotes the expression of tumor-suppressor genes.	Suzuki et al., 2013 [47]
miR-26a *miR-98* *miR-124* *miR-214* *let-7* *miR-101* *miR-137*	EZH2	Prevents the progression of prostate cancer and metastasis.	Suzuki et al., 2013 [47]

Figure 8 - MiRNAs targeting epigenetic regulators adapted from Kumar et al., 2020(15)

There have also been found a substantial number of epigenetic aberrations due to the interactions between epigenetic modifiers and lncRNAs, when in the context of cancer. A well-studied example of lncRNA action in multiple cancers is the Homeobox transcript antisense RNA (HOTAIR), which takes part in the regulation of the epigenome when having its expression dysregulated, as it alters histone methylation at gene promoters, binding to histone modifiers, such as the REST/CoREST/LSD1 complex and PRC2. In these specific cases, HOTAIR's interaction with the REST/CoREST/LSD1 complex and the PRC2, leads to the methylation of H3K4 and H3K27, respectively, which we already found to be bound to gene activation, at H3K4, and suppression, at H3K27. When dysregulated, HOTAIR's expression levels can result in aberrant epigenetic traits which can ultimately lead to tumorous phenotypes: when HOTAIRs expression levels get upregulated, both H3K4 and H3K27 are susceptible of getting upregulated, following its corresponding gene targets' activation or repression, respectively; in the same way, when HOTAIRs expression levels is found downregulated, there can exist hypomethylation in H3K4 and/or H3K27, leading to its aberrant gene expression. (15)

On the other hand, there are lncRNAs that synergize with its epigenetic factors, such as Second Chromosome Locus Associated with Prostate-1 (SCHLAP1), which is able to regulate the recruitment of the ATP-dependent chromatin-remodelling complex SWItch/Sucrose Non-Fermentable (SWI/SNF), which is responsible by simplifying the transcription factor recruitment by loosening of the chromatin structure. Mutations and/or the aberrant expression of SCHLAP1 will, consequentially, affect the recruitment of SWI/SNF and thus lead to deregulated gene expression. The Urothelial Carcinoma Associated 1 (UCA1) and the Nuclear Paraspeckle Assembly Transcript 1 (NEAT1) lncRNAs, are also able to control SWI/SNF's regulation and recruitment, as they can bind to BRG1, a subunit of SWI/SNF. The Anti-sense Non-coding RNA in the INK locus (ANRIL) lncRNA is capable of supressing gene transcription out of the interaction with PRC1 and PRC2 complexes, being able to supress genes clusters such as p15/CDKN2B, p16/CDKN2B and p14ARF, which have been found to have tumour-supressing activity. The Metastasis-associated Lung Adenocarcinoma Transcript 1 (MALAT1) is able to create a complex with BRG1 and HDAC9, resulting in smooth muscle tissue dysfunctions and thoracic aortic aneurysms.(15) In CRC, the mostly

found dysregulated lncRNAs were HOTAIR and CCAT1. (29) In 53, there can be found a more exhaustive list of lncRNAs interactions with protein complexes, such as the ones we already described.

3.2.1.4 Epigenetic Reset in cancer

As discussed previously, the methylation landscape of the DNA is dynamically altered on a constant basis in cancer. As so, to acquire its pluripotent state, CSCs undergo a process of demethylation of their own, in order to be able to further differentiate into different parent-of-origin cells. The global hypomethylated landscape that is obtained functions as an activator of oncogenes, resulting in genomic instability, and creating a more prone environment to cancer development.(33,34)

Thus, CSCs formation and tumorigenesis are heavily dependent in the loss of methylation in the DNA, as global LOI events support cancer initiation and proliferation, even including loss of growth inhibition by TGFβ. DNMTs inhibition has also been proven a fundamental mechanism in CSCs generation, as DNMT1 inhibition was found to be the process through which loss of methylation happened in the Nanog promoter, and loss-of-function mutations in DNMT3A also have been proved to result in the expansion of pre-leukemic stem cells (SCs).(33)

Nonetheless, demethylation events continue throughout the tumorigenic process, as there has been found data indicating that the CpG region of the CD133 gene promoter was found hypomethylated in multiple cancer types, such as CRC, glioblastoma, breast, and ovarian cancer. Likewise, there exist hypomethylation events exclusive of CSCs, as in the pancreatic ductal adenocarcinoma (PDAC), as there have been found genes that take part in CSCs pathways, such as GATA6, SOX9 and BMP4, that were found hypomethylated.(33)

Somatic cell reprogramming can be attained by the expression of the Yamanaka factors (Oct3/4, Sox2, Klf4, c-Myc), and results in various epigenetic changes, among them, demethylation. This reprogramming, as we have already discussed, is regulated by TET proteins, namely TET1, which influences Nanog levels and could even take functions as a replacement for Oct4. Although, recent studies have shown that TET1 is fundamental in the induction of the pluripotent state, but not in its maintenance, as there is data which leads to believe that the suppression of TET1 in embryonic stem cells (ESCs) did not affect the pluripotent state.(33)

Although, depending on the tumour type, there can also be divergences on the functions of the proteins. As of DNMT1, in colon cancer it was found to have significant importance in its initiation, and in PDAC, breast and lung cancer, and in established leukaemia it was shown that it had a crucial role in CSC function, while in prostate cancer it was found that its deregulation would promote CSCs and Epithelial-Mesenchymal Transition (EMT). (33,34)

The shared relevance of both methylation and demethylation processes, led by DNMTs and TET proteins, respectively, indicates that the particular localization of 5mC and 5hmC marks is probably an even more relevant detail in the determination of the cellular identity than the overall global landscape. As for hypermethylation at tumour supressing genes, as for hypomethylation at pluripotent loci, both are of great relevance to the generation of CSCs.(33)

3.3 Epigenetic diagnostic techniques for cancer

With epigenetic modifications growing more notable in various types of cancers, its detection has become of great importance and value as cancer diagnostic techniques.

Although, highly frequent in cancer cells, CpG methylation mutations patterns are highly heterogeneous in individual patients when analysing somatic cell mutations, restricting its ability to manifest early stages of cancers, thus not being considered the best markers for early diagnostic.(29) As for DNA methylation biomarkers for diagnosis of CRC, the methylation of the SEPT9 gene in plasma has become one of the most studied non-invasive biomarkers. This biomarker was the first approved in the United States Food and Drugs Administration (US FDA) as a molecular blood-based assay for CRC diagnosis, showing high overall rates of both sensitivity and specificity. Nevertheless, it was found statistically most effective in patients with advanced stage CRC (stages III and IV) than in early staged CRC (stages I and II). This test required only a small blood sample, and the methylated SEPT9 DNA within the plasma would be amplificated via PCR, being very appealing to patients and while having elevated compliance rates.(29)

The VIM gene, which encodes to the intermediate filament protein Vimentin, a key structural part of the cytoskeleton along with the microtubules and the actin microfilaments, has also been used as a biomarker for CRC. Methylation of the VIM gene appears to have greater clinical value in stool samples than in blood samples,

contributing to greater diagnostic precision. In adenomas and hyperplastic polyps, the methylation of the secreted frizzled-related protein 2 (SFRP2), has also shown great clinical relevance for the diagnosis of precancerous lesions.(29)

Still among the DNA methylation markers, recently, Luo *et al.* compared CRC tissue's methylation signatures with the ones from normal blood leukocytes and identified specific methylation signatures. That along with an algorithm based on the methylation patterns obtained from cell-free DNA (cfDNA) samples of a study featuring over 800 patients with CRC and over 1000 controls, allowed the authors to develop a predictive diagnostic model with very high sensitivity and specificity levels.(29,35) Applying the same method, there are ongoing studies trying to identify methylation-based biomarkers for specific gastrointestinal cancers, such as esophageal gastric cancer (GC), squamous cell and adenocarcinoma (ESCC and EAC), pancreatic ductal adenocarcinoma (PDAC), hepatocellular carcinoma (HCC), and including CRC.(29)

Likewise, Chen et al. recently announced a new blood-based cancer diagnostic technique that also utilizes cfDNA methylation analysis. The authors assembled a panel of nearly 600 genomic regions to further study in plasma samples. As in early-staged cancers tumour DNA tends to be less available and the standard methods imply a heavy DNA loss rate, the authors used a different method to build its library, the Singlera library construction method. Using semi-targeted PCR, only a single ligation event and a single PCR primer per amplicon is required, allowing a greater molecular recovery rate.(29,36)

There have also been developments towards the usage of histone alterations in circulating nucleosomes as non-invasive cancer diagnosis biomarkers. Namely, in CRC, there have been found evidence that the levels of trimethylation of H3K9 and H4K20 in circulating nucleosomes were diminished in CRC's patients when compared with healthy controls, through the use of chromatin immunoprecipitation studies. Although, as histone mutations have a key part in cancer pathogenesis, they show very little cancer specificity, and it therefore constitutes an obstacle to its use as non-invasive biomarkers. (29)

On the other hand, miRNAs potential as cancer diagnosis biomarkers remains in its small size and their limited number when compared to other far larger protein-coding

genes or mRNAs. Its stability in a great variety of biological specimens, such as blood, stool, and tissue, allows them to be easily identified and quantified in a large variety of laboratory techniques, further strengthening its application as biomarkers. In fact, miRNA panels are being increasingly more applied in cancer diagnostic, as they are being produced aiming to be more and more cancer specific, utilizing the mostly known and present miRNAs in each cancer for its detection. In fact, there is being carried out investigation using saliva as the fluid test in some miRNA panels, but they are still in validation trials as they need to be further developed. (29)

3.4 Epigenetic therapies for Cancer

With its primary epigenetic traits being further and further exploited and described among authors, it is only natural that epigenetic therapies have come to be of discussion among scientists.

In what concerns aberrant DNA methylation therapies, direct inhibitors of the COMPASS family of H3K4 methyltransferases are not yet available, as the development of these molecules would not only contribute vastly to the clinical treatment of numerous cancers, but also to further comprehend its detailed functions. Having the catalytic activity of the COMPASS methyltransferases family, the protein-protein interactions (PPI) among key COMPASS subunits, or the protein binding to methylated H3K4, pharmacologically targeted could all contribute to further improve the knowledge of the downstream events and the therapeutic opportunities for cancer treatment.(14)

Hypomethylating agents, such as DNMT inhibitors, are the older and most commonly used epigenetic therapies regarding epigenetic cancer treatments. Principally in the treatment of AML and MDS. 5-Azacitadine (5-Aza), 5-aza-2′-deoxycytidine (decitabine) and SGI-110 (guadecitabine), are all analogues of the nucleoside cytidine and they all bind irreversibly to DNMT proteins, resulting in global DNA hypomethylation.(37,38)

When R132 in isocitrate dehydrogenase 1 (IDH1) and at R172 in isocitrate dehydrogenase 2 (IDH2) have their enzymatic activity altered through intrinsic mutations, it results in a favoured production of the oncometabolite 2-Hydroxyglutarate (2-HG) over the normal α-ketoglutarate. Mutations in IDH1 and IDH2 are frequently

found in AML, angioimmunoblastic T cell lymphoma and glioma patients. 2-HG is responsible by the suppression of α-ketoglutarate-dependent dioxygenases, such as the TET proteins family and the Jumanji family of histone demethylases which lead to global DNA hypermethylation and result in increased histone lysine methylation, respectively. As so, AML patients that harbour IDH1 or IDH2 mutation are often characterized by a globally hypermethylated DNA and genomic instability, factors that heavily foster tumorigenic development.(37,38)

Regarding MLL rearranged leukemia, there were advances concerning its treatment while stabilizing the wild-type copy of MLL, mitigating its aberrant transcription by MLL fusion proteins and the super elongation complex (SEC), its oncogenic co-factor. However, targeting the oncogenic fusion proteins completely is still a hard target to reach in MLL-rearranged leukemia. There have been developed PPI disruptors of the Menin-MLL interaction, such as MI-463, MI- 503, and M-525 and OICR-9429, aiming for the WDR5- MLL interaction, expecting to treat MLL-rearranged and CEBPA mutant leukemia. Since the same domain of LEDGF binds to both the MLL1 and HIV integrase, a cyclic peptide utilized in the inhibition of HIV viral replication, the CP65, has been granted limited success in the treatment of MLL-rearranged leukemia through the targeting LEDGF. An inhibitor of DOT1L, EPZ-5676, was developed for the treatment of MLL-rearranged leukemia, aiming to the aberrant upregulation of methylation in H3K79 that is found in leukemia cases, being still under clinical scrutiny. EZH2 aberrant expression in cancer cells has been targeted as EZH2 inhibitors are usually used to inhibit undesirable histone methylation of cancer suppressor genes. Zhao *et al.* showed that tumorous cell that carry the MLL3 PHD mutations are more vulnerable to the depletion of EZH2, SUZ12, and EED in the PRC2 complex. The use of EZH2 inhibitors thus shows as a promising therapeutic option while exploiting the MLL3-UTX-PRC2 axis dependence and synthetic lethality - the process through which the simultaneous perturbation of two genes results in cellular death, whilst the perturbation of each gene independently does not.(14,39)

As of the dynamic nature of histone modifications, namely histone acetylation, HDACs inhibitors (HDACi), HATs inhibitors and bromodomain proteins inhibitors, such as vorinostat, belinostat, panobinostat and romidepsin, have been developed as potential cancer therapies and have gotten their approval in the US FDA, and a fifth, chidamide, has received regulatory approval in China. Vorinostat had its approval for

the treatment of cutaneous T-cell lymphomas, while belinostat had its approval for the treatment of peripheral T-cell lymphomas, and romidepsin had its approval to both of the lasts referred treatments; on the other hand, panobinostat had its approval for the concomitant treatment of multiple myeloma along with dexamethasone. (3,37,38)

In order to widen the response rates amongst patients, epigenetic therapies have been combined with a great variety of other chemotherapies, epigenetic factors, targeted therapies, and immune checkpoint inhibitors. Nonetheless, its combined effect is easily demonstrated in vitro but it is hard to translate into in vivo cases, as its clinical results have been discouraging, since the only combination approved by the FDA was the combination of panobinostat, bortezomib, and dexamethasone. DNMT-HDAC inhibitor combinations were early identified as possibilities for synergic interactions in 1983, but regardless of multiple studies, there has not been consistent evidence of a beneficial interaction in AML. At last, immune checkpoint combinations have also been in development with DNMT, HDAC inhibitors, and most recently, EZH2 inhibitors, in order to induce the expression of genes that encode proteins that are directly involved in the immune response reaction.(37)

Table 2 Examples of inhibitors for chromatin-related proteins

Mode of action	Target	Compound name	Types of cancer	Reference
Enzymatic inhibition	DOT1L	EPZ-5676	MLL-rearranged leukemia	[85, 86]
	EZH2	EPZ6438, GSK126, CPI-1205	Lymphoma, malignant rhabdoid tumor	[9, 77, 87]
	p300	C646, A-485	hematological malignancies and androgen receptor-positive prostate cancer	[88]
	HDACs	Vorinostat, romidepsin	CTCL	[89, 90]
	CARM1	EZM2302	Multiple myeloma	[91]
PPI disruption	Menin-MLL	MI-503, MI-463, M-525	MLL-rearranged leukemia	[82, 83]
	WDR5-MLL	OICR-9429	C/EBPα N-terminal leukemia	[84]
	LEDGF-MLL	CP65	MLL-rearranged leukemia	[92]
Competitive binding	BET family of BRD proteins	JQ1, I-BET, I-BET151	NUT midline carcinoma, MLL-rearranged leukemia	[93–95]
Protein degradation	BRD4	dBET1, dBET6, ARV-825, ARV-771, BETd-246	AML, T-ALL, Burkitt's lymphoma, castration-resistant prostate cancer, TNBC	[96, 97]

Figure 9 - Inhibitors of chromatin-related proteins, from Kumar et al., 2020(14)

Synthetic miR-143 in vitro experiments have proven its antiproliferative activity by targeting several major molecules in the RAS-RAF-MEK-ERK pathway, such as KRAS, AKT, ERK, and SOS1. As of this reason, miR-143 clinical relevance grew not only as a potential biomarker for early diagnosis, but also as a promising anti-cancer drug.(29) In another in vitro study, it was described that the replenishment with miR-143 was able to re-sensitize KRAS-mutant LoVo CRC cells to paclitaxel treatment.(28)

3.5 Future Directions

DNA methylation profiling appears to have various advantages over somatic mutation analysis for cancer diagnosis, in which a greater dynamic range and clinical sensitivity, larger predominance in precancerous lesions and early-staged tumours, and presence of numerous altered CpG sites in each targeted genomic region.(29)

Through eager research of the enzymatic and non-enzymatic activity of TET proteins, there is the possibility of being able to manipulate cells and epigenetically reprogram them, preventing hyperproliferation and malignant states that could ultimately lead to carcinogenesis.(13) TET enzymes can also have an effect on the response to chemotherapy in HNSCC, as Song et al. noted that 5-hmC formation were upregulated by TET-2 when there was administration of chemotherapeutic agents like doxorubicin.(7,40)

In what concerns CSCs, there needs to be a greater deepening in the knowledge of the epigenetic mechanisms involved in CSCs formation in each different population of cells, as each different population of cells has its different epigenetic traits, being from a normal cell population, a normal stem cell, pre-CSC, or cancerous cell. As each different population has its specific epigenetic traits, there also needs to be a greater understanding on which mechanisms contribute to carcinogenesis, and which lead to its initiation.(33)

Histone modifications have had a great input of effort in what attempts to understand the role of this alterations in the enzymatic machinery during development and disease, in particular in cancer cases. There are currently being developed specific techniques to map the localization and function of these histone modifications in the genome, hoping to narrow its whereabouts to a few single cells, as it is essential to describe exactly what functions PTMs and their modifiers have in cancers. As important is the correct and precise identification of non-histone substrates, in order to define histone modifiers functions for the development of more specific inhibitors for the chosen pathway. For the improvement of small molecule inhibitors, the deepening of the understanding of key enzymes work with other subunits will allow improvements in the ability to design these small molecule disruptors, as the MLL-Menin and MLL-WDR5 complex inhibitors, for example. The latest targets along with other histone modification targets, such as enzymatic activity suppression and small molecule

degraders by PROTACs, are thought to be the most encouraging drug targets, as a great number of clinical trials are currently undergoing its experimentation with other chemotherapies already available. Finally, histone modification can be hypothetically used as biomarkers for both diagnosis and prognosis of cancers.(14)

Using the synthetic lethal approach, there are experiments being taken in order to try to reduce the toxic off-target side effects and to avoid therapy resistance, as multiple genes are being targeted at the same time using a combination of therapies. The final objective is to turn theoretical epigenetic therapies into reliable daily practices to fight different cancer types each time more specifically and accurately.(14)

In the Figure 10 - Targeted epigenetic axis for the creation of new epigenetic therapies, from Bates et al., 2020 (37) we can get the idea of what epigenetic axis are being currently targeted in the development of new epigenetic therapies and in what stage they are currently.

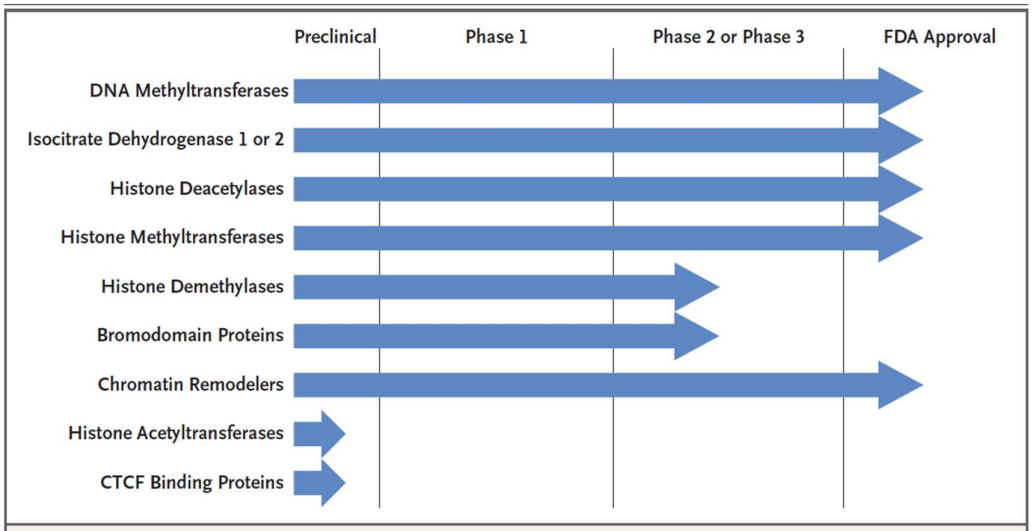

Figure 5. Epigenetic Regulators under Investigation or Approved by the Food and Drug Administration (FDA), According to Protein Family.
Shown is the stage of drug development for key protein families of epigenetic regulators.

Figure 10 - Targeted epigenetic axis for the creation of new epigenetic therapies, from Bates et al., 2020 (37)

4 Conclusions

Cancer epigenetics has appeared as one of the most prone fields for the development of cancer therapies and diagnosis procedures. Its mechanisms allows an eager investigation on the cancer pathways in which they take part.

DNA-modified methylation is a vastly know epigenetic trait among various cancer types, as in both DNA hyper- and hypomethylation mechanisms, as in DNA hydroxymethylation regarding TET proteins. These epigenetic traits may take place all over the genome, specifically in clusters distal from promoters, affecting its expression and regulation, and also the chromatin's structure.(11)

In what regards TET proteins, there have been improvements on the proteins' functions knowledge, as its function does not lay anymore only on DNA demethylation, but also has a regulatory function in gene expression and cell identity maintenance. By further studying them, scientists hope to achieve cell epigenetic reprograming procedures, wishing to avoid its hyperproliferation and, ultimately, its malignant progression into tumorigenesis.(13)

Histone modifications are still the less known field of epigenetic marks, as they will require an approach requiring genetics and gene editing procedures, chemical and structural knowledges, as well as protein biochemistry studies and next-generation-sequencing. Although, their mechanistic pathways are well described among histone methylation, acetylation and phosphorylation, along with many others, and have also been found present in various cancer types, having tumorigenic functions towards its expression,(14,29)

Non-coding RNAs have also found their way amongst epigenetic pathways as their function ends up being mainly regulatory among post-transcriptional modifications, controlling its up- or downregulation, but they are also very sensitive against structural changes, such as mutations, deletions, amplifications or translocations, which may very well be present in tumorous tissue.(15,29)

Albeit its initial differentiation status, cancer cells possess the ability to epigenetically reset to its immature-form, allowing itselves to avoid immune checkpoints and modulate its response to the immune system possessing phenotypic flexibility.(21) CSCs formation have some epigenetic mechanisms involved in it that

have some serious implications in tumour development, metastasis and relapse, and therefore, a great variety of studies have been drawn in order to further explain the links between them, having been found some appealing targets to develop new cancer diagnosis and therapies procedures.(33)

There already exists a high number of epigenetic therapies and diagnostic procedures being utilized across the world with high sensitivity and specificity values, leading us to believe that its clinical value will only tend to grow. Given the multiple targets still able for studying, we can have reasons to believe that for the years to come there will always be novel therapies and diagnostic procedures on the verge of discovery.(37,38)

Bibliographic References

1. WHO. World Health Organization [Internet]. 2021. Available from: https://www.who.int/news-room/fact-sheets/detail/cancer

2. IARC. International Agency for Research on Cancer [Internet]. 2021. Available from: https://gco.iarc.fr/today/online-analysis-multi-bars?v=2020&mode=cancer&mode_population=countries&population=900&populations=900&key=total&sex=0&cancer=39&type=0&statistic=5&prevalence=0&population_group=0&ages_group%5B%5D=0&ages_group%5B%5D=17&nb_items=10&group_cancer=1&include_nmsc=1&include_nmsc_other=1&type_multiple=%257B%2522inc%2522%253Atrue%252C%2522mort%2522%253Atrue%252C%2522prev%2522%253Afalse%257D&orientation=horizontal&type_sort=0&type_nb_items=%257B%2522top%2522%253Atrue%252C%2522bottom%2522%253Afalse%257D

3. Cao J, Yan Q. Cancer Epigenetics, Tumor Immunity, and Immunotherapy. Trends in Cancer [Internet]. 2020;6(7):580–92. Available from: https://doi.org/10.1016/j.trecan.2020.02.003

4. Chen DS, Mellman I. Oncology meets immunology: The cancer-immunity cycle. Immunity. 2013;39(1):1–10.

5. Hayes C. Cellular immunotherapies for cancer. Ir J Med Sci. 2021;190(1):41–57.

6. Weber W. Cancer epigenetics. Prog Mol Biol Transl Sci. 2010;95(C):299–349.

7. Gaździcka J, Gołąbek K, Strzelczyk JK, Ostrowska Z. Epigenetic Modifications in Head and Neck Cancer [Internet]. Vol. 58, Biochemical Genetics. Springer US; 2020. 213–244 p. Available from: https://doi.org/10.1007/s10528-019-09941-1

8. Ebrahimi V, Soleimanian A, Ebrahimi T, Azargun R, Yazdani P, Eyvazi S, et al. Epigenetic modifications in gastric cancer: Focus on DNA methylation. Gene. 2020;742(February).

9. Shao S, Gudjonsson JE. Epigenetics of Psoriasis. Vol. 1253, Advances in Experimental Medicine and Biology. 2020. 209–221 p.

10. Shanak S, Helms V. DNA methylation and the core pluripotency network. Dev Biol [Internet]. 2020;464(2):145–60. Available from: https://doi.org/10.1016/j.ydbio.2020.06.001

11. Skvortsova K, Stirzaker C, Taberlay P. The DNA methylation landscape in cancer. Essays Biochem. 2019;63(6):797–811.

12. Parry A, Rulands S, Reik W. Active turnover of DNA methylation during cell fate decisions. Nat Rev Genet [Internet]. 2021;22(1):59–66. Available from: http://dx.doi.org/10.1038/s41576-020-00287-8

13. Tsiouplis NJ, Bailey DW, Chiou LF, Wissink FJ, Tsagaratou A. TET-Mediated Epigenetic Regulation in Immune Cell Development and Disease. Front Cell Dev Biol. 2021;8(January).

14. Zhao Z, Shilatifard A. Epigenetic modifications of histones in cancer. Genome Biol. 2019;20(1):1–16.

15. Kumar S, Gonzalez EA, Rameshwar P, Etchegaray JP. Non-coding RNAs as mediators of epigenetic changes in malignancies. Cancers (Basel). 2020;12(12):1–32.

16. Barchetta I, Arvastsson J, Sarmiento L. Epigenetic Changes Induced by Maternal Factors during Fetal Life : Implication for Type 1 Diabetes. 2021;1–15.

17. Thamban T, Agarwaal V, Khosla S. Review Role of genomic imprinting in mammalian development. 2020;0123456789:1–21.

18. Lozano-Ureña A, Jiménez-Villalba E, Pinedo-Serrano A, Jordán-Pla A, Kirstein M, Ferrón SR. Aberrations of Genomic Imprinting in Glioblastoma Formation. Front Oncol. 2021;11(March):1–10.

19. Tucci V, Isles AR, Kelsey G, Ferguson-smith AC, Group I. Review Genomic Imprinting and Physiological Processes in Mammals. 2019;7.

20. Wray S. Developmental Neuroendocrinology.

21. Feinberg AP, Koldobskiy MA, Göndör A. Epigenetic modulators, modifiers and mediators in cancer aetiology and progression. Nat Rev Genet. 2016;17(5):284–99.

22. Shah R, Patel T, Freedman JE. Circulating Extracellular Vesicles in Human

Disease. N Engl J Med. 2018;379(10):958–66.

23. Kok VC, Yu CC. Cancer-derived exosomes: Their role in cancer biology and biomarker development. Int J Nanomedicine. 2020;15:8019–36.

24. Spadafora C. Transgenerational epigenetic reprogramming of early embryos: a mechanistic model. Environ Epigenetics. 2020;6(1):1–9.

25. Nixon B, Stanger SJ, Mihalas BP, Reilly JN, Anderson AL, Tyagi S, et al. The MicroRNA Signature of Mouse Spermatozoa Is Substantially Modified During Epididymal Maturation. Biol Reprod. 2015;93(4):1–20.

26. Reilly JN, McLaughlin EA, Stanger SJ, Anderson AL, Hutcheon K, Church K, et al. Characterisation of mouse epididymosomes reveals a complex profile of microRNAs and a potential mechanism for modification of the sperm epigenome. Sci Rep. 2016;6(July):1–15.

27. Sharma U, Sun F, Conine CC, Reichholf B, Kukreja S, Herzog VA, et al. Small RNAs Are Trafficked from the Epididymis to Developing Mammalian Sperm. Dev Cell. 2018;46(4):481-494.e6.

28. Jung G, Hernández-Illán E, Moreira L, Balaguer F, Goel A. Epigenetics of colorectal cancer: biomarker and therapeutic potential. Nat Rev Gastroenterol Hepatol. 2020;17(2):111–30.

29. Ruiz-bañobre J, Goel A. Genomic and epigenomic biomarkers in colorectal cancer : From diagnosis to therapy [Internet]. 1st ed. Vol. 151, Novel Approaches to Colorectal Cancer. Elsevier Inc.; 2021. 231–304 p. Available from: http://dx.doi.org/10.1016/bs.acr.2021.02.008

30. Goodell M, States U, Challen G. Clonal Hematopoiesis : Mechanisms Driving Dominance of Stem Cell Clones. 2021;

31. Morgan MAJ, Shilatifard A. Reevaluating the roles of histone-modifying enzymes and their associated chromatin modifications in transcriptional regulation. Nat Genet [Internet]. 2020;52(12):1271–81. Available from: http://dx.doi.org/10.1038/s41588-020-00736-4

32. Zhang Q, Zhao K, Shen Q, Han Y, Gu Y, Li X, et al. Tet2 is required to resolve inflammation by recruiting Hdac2 to specifically repress IL-6. Nature. 2015;525(7569):389–93.

33. French R, Pauklin S. Epigenetic regulation of cancer stem cell formation and maintenance. Int J Cancer. 2021;148(12):2884–97.

34. Poli V, Fagnocchi L, Zippo A. Tumorigenic cell reprogramming and cancer plasticity: Interplay between signaling, microenvironment, and epigenetics. Stem Cells Int. 2018;2018.

35. Luo H, Zhao Q, Wei W, Zheng L, Yi S, Li G, et al. Circulating tumor DNA methylation profiles enable early diagnosis, prognosis prediction, and screening for colorectal cancer. Sci Transl Med. 2020;12(524):1–12.

36. Chen X, Gole J, Gore A, He Q, Lu M, Min J, et al. Non-invasive early detection of cancer four years before conventional diagnosis using a blood test. Nat Commun [Internet]. 2020;11(1):1–10. Available from: http://dx.doi.org/10.1038/s41467-020-17316-z

37. Bates SE. Epigenetic Therapies for Cancer. N Engl J Med. 2020;383(7):650–63.

38. Hogg SJ, Beavis PA, Dawson MA, Johnstone RW. Targeting the epigenetic regulation of antitumour immunity. Nat Rev Drug Discov [Internet]. 2020;19(11):776–800. Available from: http://dx.doi.org/10.1038/s41573-020-0077-5

39. Jones PA, Ohtani H, Chakravarthy A, De Carvalho DD. Epigenetic therapy in immune-oncology. Nat Rev Cancer [Internet]. 2019;19(3):151–61. Available from: http://dx.doi.org/10.1038/s41568-019-0109-9

40. Kim SY, Han YK, Song JM, Lee CH, Kang K, Yi JM, et al. Aberrantly hypermethylated tumor suppressor genes were identified in oral squamous cell carcinoma (OSCC). Clin Epigenetics. 2019;11(1):1–12.

Annexes

A1. Chromatin modifiers, binding factors and functions of PTMs on histones, from Zhao et al., 2019(14)

Table 1 Chromatin modifiers, binding factors and functions of selected PTMs on histones

Histones	Modification	Modifiers		Binding factors	Functions	Reference
H2A	K119Ub1	RING1A/B	BAP1, USP16, USP21, 2A-DUB, USP3, USP22		Transcriptional repression	[12, 13]
H2AX	S139p (γH2AX)	ATM, ATR, DNA-PK	PP2A, Wip1, PP6 and PP4	MDC1, CRB2	DNA repair	[14]
H2B	K120Ub1	RNF20, RNF40	USP3, USP7, USP22		Transcriptional activation, DNA damage response	[15]
H3	K4me1/2/3	SET1A/B, MLL1-4	KDM1A/B, KDM2A/B, KDM5A/B/C/D	BRWD2/PHIP, MLL,TAF3, CHD1,RAG2, BPTF, PHF2/6/8, JMJD2	Transcriptional activation	[16–18]
	K9me1/2/3	Suv39H1/2, G9a, GLP, SETDB1	KDM1A, KDM3A/B, JMJD1C, KMD4A/B/C/D	HP1, ATRX	Transcriptional activation (K9me1), repression (K9me2/3), X-inactivation and imprinting (K9me2)	[19]
	S10p	Aurora B, MSK/RSK/Jil-1	PP2A, PP1	14–3-3ζ	Mitosis, meiosis, transcriptional activation	[20]
	R26me2	CARM1	PADI4		Transcriptional activation	[21]
	K27 ac	CBP/p300	HDACs	BRDs	Transcriptional activation	[22]
	K27me1/2/3	EZH1/2	KDM6A/B, KDM7A, PHF8	EED, PC, CBX7	Transcriptional activation (K27me1); Transcriptional silencing, X-inactivation, bivalent genes/gene poising (K27me2/3)	[23]
	K36me1/2/3	NSD1–3, SETD2/3, ASH1L, SETMAR, SMYD2	KDM2A/B, KDM4A/B/C/D, JHDM1A	ZYMND11, PHF19, LEDGF	Transcriptional elongation, repression, DNA repair	[19]
	K79me1/2/3	DOT1L	?	p53BP1	Transcriptional activation	[24]
H4	K20me1	PR-Set7	LSD1n	CRB2, p53BP1	Transcriptional activation	[25, 26]
	K20me2/3	SUV4-20H1/2	LSD1n, DPY-21	CRB2, p53BP1, JMJD2	Transcriptional silencing, Heterochromatin	[25, 27]
	K16 ac	MOF	HDACs, Sirt2	BRDs	Transcriptional activation, DNA repair	[28, 29]

A2. LncRNA interactions with epigenetic regulatory complexes, from Kumar et al., 2020 (15)

lncRNA	*Origin/Location*	*Interactions with Epigenetic Regulators*	*Function*	*References*
HOTAIR (HOX transcript antisense RNA)	Transcribed from antisense strand of homeobox C gene in chromosome 12	PRC2 LSD1/CoREST	Gene silencing by methylation of H3K27me3 and demethylation of H3K4me2	Cai et al., 2014 [65]
SCHLAP1 (second chromosome locus associated with prostate-1)	From chromosome 2	SWI/SNF	Partially antagonizes location and function of SWI/SNF	Raab et al., 2019 [67]
NEAT1 (nuclear paraspeckle assembly transcript 1)	Transcribed from the multiple endocrine neoplasia locus in chromosome 11	Subpopulation of SWI/SNF complexes	Nuclear paraspeckle (nuclear bodies) assembly	Neve et al., 2018 [69]
XIST (X-inactive specific transcript)	Chromosome X	PRC1	Silencing one pair of X chromosomes	Pintacuda et al., 2017 [63]
ANRIL (antisense non-coding RNA in the INK4 locus)	Transcribed from the CDKN2A/B gene cluster at chromosome 9 in the antisense direction	PRC1 (CBX7), PRC2 (SUZ12)	Transcriptional repression	Chi et al., 2017 [71]
GAS5 Growth arrest-specific 5)	From chromosome 1	PRC2	Repression of glucocorticoids receptors, IRF4 (interferon regulatory factor 4)	Wang et al., 2018 [73]
MEG3 (maternally expressed 3)	Maternally expressed, generates multiple isoforms by alternative splicing, from chromosome 14	JARID2, EZH2	Transcriptional repression	Wang et al., 2018 [73]
PVT1 (plasmacytoma variant translocation 1)	From chromosome 8	PRC2 (EZH2)	Oncogene	Yu et al., 2018 [75]
MALAT1 (metastasis associated lung adenocarcinoma transcript 1)	Also known as NEAT2 (non-coding nuclear-enriched abundant transcript 2). Infrequently spliced ncRNA, from chromosome 11	PRC2 (EZH2), HDAC9, BRG1	Tumorigenesis Vascular disease	Wang et al., 2018 [73] Cardenas et al., 2018 [77]
KCNQ1OT1 (KCNQ1 overlapping transcript 1)	Part of an imprinting control region in chromosome 11	G9a, PRC2 (EZH2)	Gene silencing by H3K9me2 H3K27me3	Wang et al., 2018 [73]
H19 (H19 imprinted maternally expressed transcript)	From imprinted region in chromosome 11	SAHH, PRC2 (EZH2)	Tumor-suppressor Oncogene	Zhou et al., 2015 [76]
UCA1 (urothelial cancer associated 1)	From chromosome 19	PRC2 (EZH2), SWI/SNF	Tumorigenesis	Neve et al., 2018 [69]
PANDAR (promoter of CDKN1A antisense DNA damage activated RNA)	From chromosome 6	PRC1 PRC2	Tumorigenesis	Puvvula et al., 2014 [78]