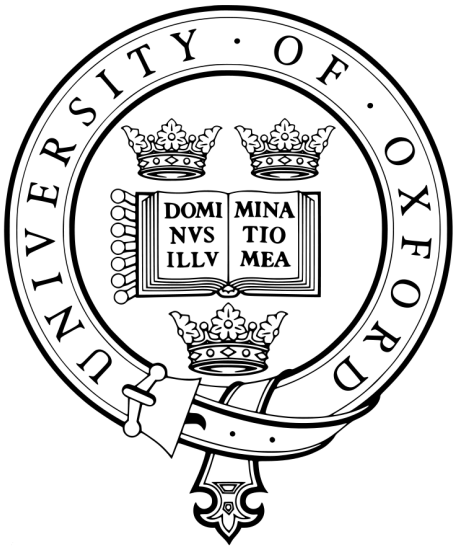
Tet2 function and relevant potential intervention in cancer

A thesis submitted in fulfillment of the requirements for the degree of

Doctor of Philosophy



Melania Zauri
St Peter's College

Supervisor: Dr. Skirmantas Kriaucionis
Ludwig Institute for Cancer Research
Nuffield Department of Clinical Medicine
University of Oxford

Trinity 2014

Melania Zauri
St Peter's College

Doctor of Philosophy
October 2014

Tet2 function and relevant potential intervention in cancer

Abstract

5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxycytosine (5caC) are products of TET enzyme activity, which initiates at 5-methylcytosine (5mC) in DNA. 5hmC has been found to be globally depleted in cancer, while *TET2* is highly mutated. The first objective of this work was the investigation of the fate of the associated nucleosides in the context of nucleotide metabolism. Moreover, since many nucleoside analogs are used in cancer therapy to induce replication arrest and DNA demethylation, the effects on proliferation induced by the administration of nucleosides with bases, which are implicated in DNA demethylation, were evaluated. Finally, the TET2 protein interaction network was investigated to elucidate its lost function in cancer and shed light on possible means of intervention. It is demonstrated that DNA polymerase is capable of accepting modified cytidine triphosphates for polymerization. However triphosphate generation is impaired by substrate selectivity of CMPK1. 5hmdC and 5fdC given as media supplements to a number of cancer cell lines caused selective inhibition of proliferation without obvious defects in DNA methylation. By analyzing gene expression datasets, cytidine deaminase (*CDA*) overexpression was identified in sensitive cell lines. Overexpression and knockdown experiments demonstrated that *CDA* is necessary and sufficient to confer sensitivity to 5hmdC. Furthermore, *CDA* was able to deaminate *in vitro* 5hmdC and 5fdC, but not 5cadC, and a deaminated version (5hmU) was detected in the DNA of treated cells. 5hmU and 5fU in the DNA correlated with S/G2 phase arrest and increased levels of γH2AX at 72 h, suggesting that defects in proliferation relate to DNA damage. 5hmdC and 5fdC administration to mice (IP route) resulted in no overt toxicity and some degree of antitumor activity. Furthermore, *TET2* was found to interact with *WDR61*, a member of the PAF complex (involved in transcriptional elongation), possibly explaining its role in 5hmC deposition over gene bodies and in oncogenesis, given the documented role of PAF in leukemia. In conclusion, this study identifies the salvage route of new biological cytidine variants, a new avenue with which to target *CDA* overexpressing cancers, and a possible mechanism to explain the role of *TET2* in leukemogenesis.

This thesis is submitted to the Department of Medical Science, University of Oxford, in fulfillment of the requirements for the degree of Doctor of Philosophy. This thesis is entirely my own work, and except where otherwise stated, describes my own research.

Word count: approx. 49,550

Melania Zauri, St Peter's College

Ai miei genitori...

I have a friend who is an artist, and he sometimes takes a view which I don't agree with. He'll hold up a flower and say, "Look how beautiful it is", and I'll agree. But then he'll say, "I, as an artist, can see how beautiful a flower is. But you, as a scientist, take it all apart and it becomes dull." I think he's kind of nutty. There are all kinds of interesting questions that come from a knowledge of science, which only adds to the excitement and mystery and awe of a flower. It only adds. I don't understand how it subtracts [Feynman and Leighton, 1988].

Acknowledgements

I am thankful to my supervisor, Dr. Skirmantas Kriaucionis, for having hosted me in his laboratory for the past 4 years. Having a young supervisor has made me stronger through many difficulties and made me learn to be a scientist.
I thank the Ludwig Institute for Cancer Research for support and funding.
I would like to thank John Christianson for his help as a second supervisor and Sarah De Val for her energy and the example of a woman in science with family.
I would also like to convey a big thanks to Mark Shipman, for his technical assistance on the microscopes and image analysis and for all the invaluable things that he has thought me, including a bit more of passion for microscopy analysis; Indrika Ratnayaka for her precious help with the sectioning of tissue samples. A special acknowledgement goes to Claire Beveridge for her critical reading and many useful suggestions on the thesis.
Next, a big thanks goes to all my laboratory colleagues, for having shared with me time in the laboratory. I thank as well other members of the institute, in particular Sofia Koch, Richard Lisle for having thought me many things about histology. Zhe Zhao, Martin Fritzsche and Philipp Becker for creating an excellent working environment at the LICR and for their friendship also outside the lab. I wish them all the best in life and hope to keep in touch. I thank all the present and past Italians in the institute for having shared the sorrows of Italians abroad. I thank Caroline Sanvitale for her help in going to hospital after my bike accident and all the J. Radcliffe hospital Trauma Unit for having taken care of me. Many thanks go to the Oxford tube for having facilitated my frequent escapes to London.
Deepest gratitude is also due to all the collaborators mentioned in the thesis who provided expertise. In particular, I am grateful to Georgina Berridge and Marie Laetitia Thezenas for the MS data and Prof. Robert Goldin for his invaluable help with histology analysis. I also thank Stephen Laird and Jordan Tanner for their help with mouse work. Last, but not least, Rapolas Zilionis, a master student who spent his internships at the LICR collaborating with me on my research project. His commitment, enthusiasm and achievements have helped my project and me on becoming a supervisor in small scale.
Most importantly, I want to thank my parents Anna and Luciano, my brother Daniele and my grandparents Rosina and Berardo, whom passed during my third year. My thesis is dedicated to them who shared with me joy and pains of my PhD and renewed my enthusiasm from far away.

Contents

List of Figures

List of Tables

Abbreviations

CCLE Cancer Cell Line Encyclopedia

NCI60 National Cancer Institute 60 cancer cell lines panel

qPCR Real time PCR

ChIP Chromatin immunoprecipitation

kb kilobase (1000 bases)

bp base pair

h hour

μM micro molar

s seconds

°C degrees Celsius

kDa kilo Dalton

HPLC High Performance Liquid Chromatography

EC_{50} Effective concentration of 50%

K_m Michaelis Menten's constant

v_{max} Maximum velocity

IVRA *In vitro* replication assay

FDA Food and Drug Administration

sh-RNA small hairpin RNA

AML Acute myeloid leukemia

CML Chronic myeloid leukemia

Genes and protein convention

CDA human/mouse protein

CDA human gene or DNA or RNA

Cda mouse gene or DNA or RNA

DNA conventions

5mC 5-methylcytosine

5hmC 5-hydroxymethylcytosine

5fC 5-formylcytosine

dC deoxycytidine

5mdC 5-methyldeoxycytidine

5hmdC 5-hydroxymethyldeoxycytidine

5fdC 5-formyldeoxycytidine

5cadC 5-carboxyldeoxycytidine

5azadC 5-aza-deoxycytidine

dU deoxyuridine

dT deoxythymidine

5hmdU 5-hydroxymethyldeoxyuridine

5fdU 5-formyldeoxyuridine

5mdCMP 5-methyldeoxycytidine-monophosphate

5hmdCMP 5-hydroxymethyldeoxycytidine-monophosphate

5hmdCTP 5-hydroxymethyldeoxycytidine-triphosphate

Yeast two hybrid screen

AD Gal4 activation domain

BD Gal4 DNA-binding domain

AbA antibiotic Aureobasidin A

Ade- auxotrophic for adenine

His- auxotrophic for histidine

Leu- auxotrophic for leucine

Trp- auxotrophic for tryptophan

SD minimal, synthetically defined medium for yeast

DO dropout supplement or solution

DDO double dropout medium: SD/–Leu/–Trp

DDO/X/A double dropout medium supplemented with X-α-Gal and AbA

QDO quadruple dropout medium: SD/–Ade/–His/–Leu/–Trp

QDO/X/A quadruple dropout medium supplemented with X-α-Gal and AbA

YPDA complete yeast medium.

Chapter 1

Introduction

Nucleotides are key components of life. Many roles, aside from their participation in the structure of DNA and RNA have been attributed to nucleotides. In particular, they are: sources of energy for metabolic processes including nucleotide and histidine synthesis, facilitate the transfer of particular groups in transfer reactions (e.g. glycosyl groups, alcohol phosphate transfer, electron transfer, alkyl and acyl-group transfer) and as cyclic forms (cAMP) get involved in a number of physiological responses (hormone action). Their involvement in these processes, together with the potential intervention points in their bio-synthetic pathways that linked them closely to cellular proliferation, has led to significant research efforts being directed toward them with the aim of developing therapeutic agents. Currently, many of the widely used cancer drugs that are nucleoside analogs target DNA nucleotide bio-synthesis precursors or DNA signaling pathways, such as DNA methylation. This introduction will first focus on nucleotide production inside the cell, later focus on modified forms of cytosine that are present in DNA (production and biological role) and on relevant nucleoside analogs that the FDA (Food and Drug Administration) has approved for cancer therapy.

1.1 Nucleotide metabolism

Nucleotides were first discovered as monomeric constituents of nucleic acids. For practical reasons, a nomenclature was established by Levene in which bases with an attached ribose were named "nucleosides", and the combination of the nucleoside and phosphoric acid was named "nucleotide" [Levene and Jacobs, 1909]. Nucleotides are the monomers that are linked to form nucleic acids. The sugar that is incorporated into deoxyribonucleic acid (DNA) monomers is a deoxyribose, with an oxygen missing at the 2' position of the carbon ring, whereas ribose is present in ribonucleic acid (RNA). In the DNA, the sugars are linked through phosphodiester bridges, which join the 3' carbon of a deoxyribose to a phosphate group that is linked to the 5' carbon of sugar in the next monomer. This chain of sugars linked by phosphodiester bridges is called the DNA backbone. Although the backbone is uniform in a nucleic acid, the bases vary from one monomer to the next. DNA bases are derivatives of purines and pyrimidines. The purines bases are adenine (A) and guanine (G), while the pyrimidines are thymine (T) and cytosine (C) (Fig. 1.1). N-9 of a purine, or N-1 of a pyrimidine, attaches to C-1' of the sugar in each monomer through a N-glycosidic linkage. Once individual monomers have been joined together, two anti-parallel chains assemble into a double helical structure to form DNA, which is stabilized through the formation of hydrogen bonds between guanine-cytosine and adenine-thymine base pairs. The canonical structure of DNA is referred to as "B-form" [Watson and Crick, 1953a] [Watson and Crick, 1953b] [Franklin and Gosling, 1953].

1.1.1 *De novo* nucleotide biosynthesis

The synthesis of nucleotides proceeds in ordered steps, which see the formation of the purine and pyrimidine rings and their coupling to a ribose-5-phosphate. Two main routes can be undertaken by the cell to produce DNA building blocks: *de novo* synthesis and

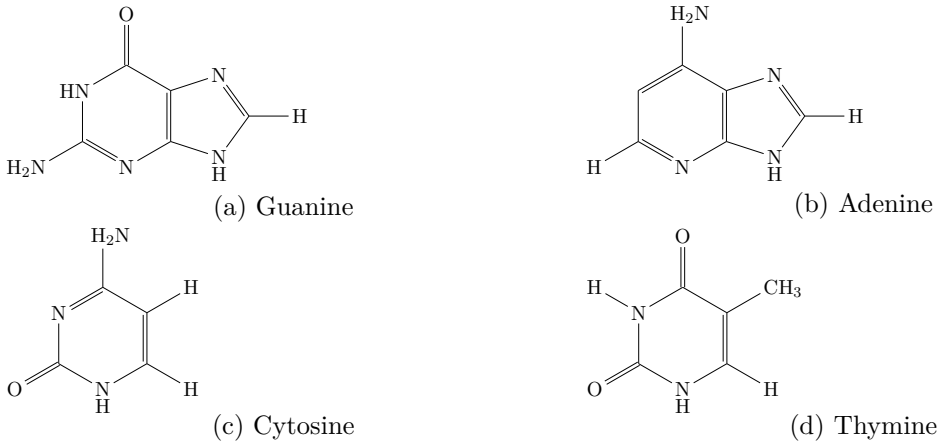


Figure 1.1: The DNA bases that are the major components of the mammalian DNA. a and b: purines; c and d: pyrimidines

the salvage pathway. In the *de novo* bio-synthesis pathway, pyrimidine bases are first assembled and then attached to a ribose, while purine nucleotides are synthesized directly from a ribose-based structure. While purine synthesis takes place entirely in the cytoplasm, a step in pyrimidine bio-synthesis takes place in the mitochondria. Furthermore, *de novo* pathways lead to the synthesis of ribonucleotides, which needs to be reduced to form deoxyribonucleotides.

Purine bio-synthesis

In mammals, the existence of a *de novo* purine bio-synthetic pathway was established through the use of purine-free diets in mice. Subsequently, numerous lines of investigation led to the identification of the steps by which synthesis occurs.

Purines are synthesized starting from ribose-5-phosphate, which is converted to 5-phosphoribosyl-1-pyrophosphate (PRPP) in a reaction catalyzed by phosphoribosyl pyrophosphate synthetase (PRPS) [Fox and Kelley, 1971]. PRPP is subsequently converted to 5-phosphorybosylamine (PRA) by glutamine phosphorybosyl amidotransferase (GPAT activity encoded by the *PPAT* gene), which displaces the pyrophosphate utilizing glutamine-derived ammonia [Hartman, 1963]. Next glycinamide ribonucleotide synthetase (GARS

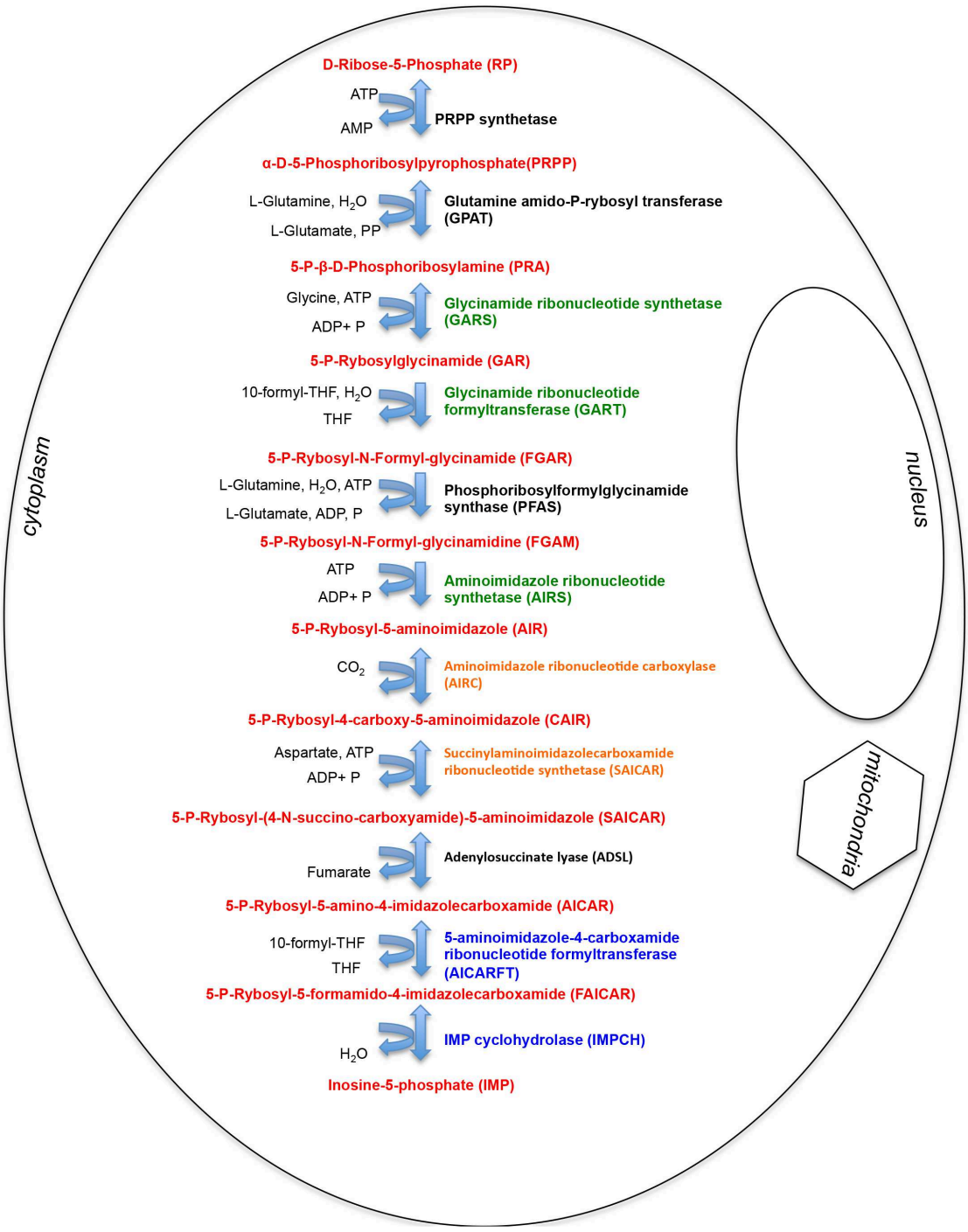


Figure 1.2: *De novo* purine synthesis (a): steps leading to inosine-5-phosphate (IMP). The enzymes activities encoded by the same gene (*GART*, *PAICS*, *ATIC*) are shown in green, orange and blue, respectively.

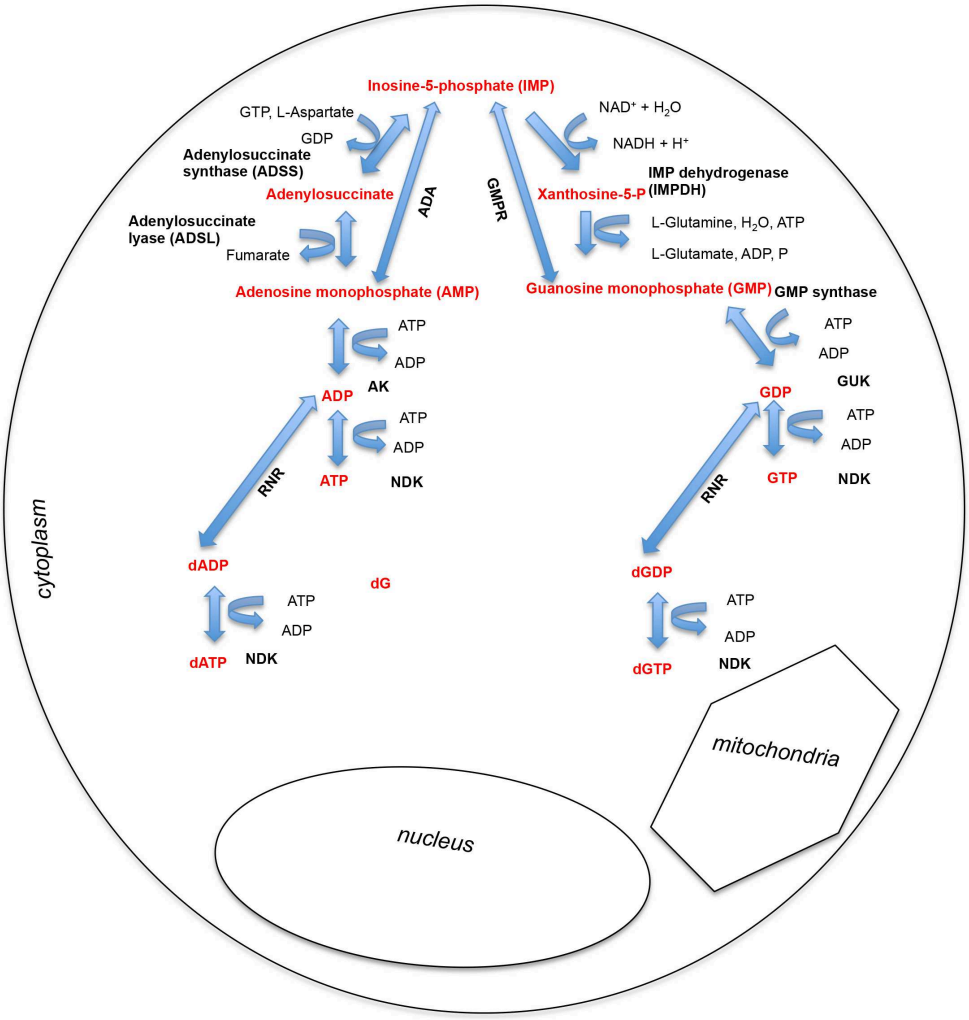


Figure 1.3: *De novo* purine synthesis (b): steps leading to GTP and ATP formation.

activity encoded by the *GART* gene) couples a glycine to the amino group of 5-phosphorybosylamine forming glycinamide ribonucleotide (GRA), through an ATP-dependent reaction [Aimi et al., 1990]. The reaction is followed by the addition of formate to form formylglycineamide phospho-ribonucleotide (FGAR), catalyzed by GAR transformylase (GART activity encoded by the *GART* gene) [Aimi et al., 1990]. A further ammonia derived from glutamine is added by formylglycineamidine synthetase (PFAS activity encoded by the *PFAS* gene) to form formylglycineamidine ribonucleotide (FGAM) [Barnes et al., 1994],

which is then converted into a 5-aminoimidazole ribonucleotide (AIR) by aminoimidazole ribonucleotide synthase (AIRS activity encoded by the *GART* gene) through an ATP-dependent reaction [Aimi et al., 1990]. Subsequently, bicarbonate is activated by ATP and attached to the imidazole ring to form carboxyaminoimidazole ribonucleotide (CAIR) in a step catalyzed by carboxyaminoimidazole ribonucleotide synthase (AIRC activity encoded by the *PAICS* gene). Successive catalysis by succinylaminoimidazolecarboxamide ribonucleotide synthetase (SAICAR activity encoded by the *PAICS* gene) attaches an aspartate to the imidazole carboxylate, generating 5-aminoimidazole-4-succinylcarboxamide ribonucleotide (SAICAR) [Schild et al., 1990]. SAICAR is then cleaved by adenylosuccinate lyase (ADSL activity encoded by the *ADSL* gene) to form fumarate and 5-aminoimidazole-4 carboxamide ribonucleotide (AICAR) [Stone et al., 1993]. This product is converted to 5-formaminoimidazole-4 carboxamide ribonucleotide (FAICAR) by 5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase (AICARFT activity encoded by the *ATIC* gene) and from it, finally, inositol monophosphate (IMP) is formed by IMP cyclohydrolase (IMPCH activity encoded by the *ATIC* gene) [Stone et al., 1993] (Fig. 1.2).

Studies on the interconversion of purines began with the discovery that, by administering radioactive glycine to the cell, both adenine and guanine could be labeled. Therefore, these two nucleotides had to have the same origin. Subsequently the discovery was made that the synthesis pathway splits into two branches after the generation of IMP, tightly interdependent for the inverse energy requirements (GTP for AMP and ATP for GMP) (Fig. 1.3). In the first branch, which leads to the generation of AMP, adenylsuccinate synthetase (*ADSS* gene) converts IMP to adenylsuccinate using GTP as cofactor [Van der Weyden and Kelly, 1974], after which adenylsuccinate lyase (*ADSL* gene) produces AMP by releasing fumarate [Stone et al., 1993]. The reaction that reverses these steps, converting AMP to IMP, has been traced to adenylate deaminase (ADA) [Lee, 1957]. In the other branch, IMP is first reduced to xanthosine monophosphate by IMP dehydrogenase (IMPDH) [McFall and Magasanik, 1960], before being aminated by guanine monphos-

phate synthetase to produce GMP, using ATP as source of energy and glutamine as an amine donor [Abrams and Bentley, 1959]. In the same way as AMP, GMP can be deaminated back to IMP by GMP reductase (GMPR) [Mackenzie and Sorensen, 1973]. GMP and AMP are subsequently diphosphorylated by guanylate kinase (GUK1) [Agarwal et al., 1978] and by adenylate kinase (AK) [Panayiotou et al., 2014] to form GDP and ADP. They can then be reduced by ribonucleotide reductase to their respective deoxyribonucleotides [Moore and Reichard, 1964] or further phosphorylated by nucleoside diphosphate kinases (NDK encoded by *NME1*), which generates guanine and adenine triphosphate [Mourad and Parks, 1966].

Pyrimidine bio-synthesis

The first step in pyrimidine bio-synthesis is the production of carbamoyl phosphate from a bicarbonate, an ammonia derived from glutamine and 2 molecules of ATP (Fig. 1.4). This reaction is catalyzed by carbamoyl phosphate synthetase II (CPSII) [Jones et al., 1955] [Tatibana and Ito, 1969]. This enzyme is able to hydrolyze glutamine to obtain the ammonia needed for the reaction. Subsequently, carbamoyl phosphate (CP) is used as substrate of aspartate transcarbamoylase to form carbamoylaspartate [Reichard, 1954], which then gets converted into dihydroorotate by dihydroorotase. In mammals these enzymatic activities are all encoded by a unique gene called *CAD* (carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase) [Coleman et al., 1977]. The next step is catalyzed by dihydroorotate dehydrogenase (DPYD) which oxidizes dihydroorotate, through a NAD^+ dependent reaction to produce orotate [Lu et al., 1992]. This step is carried out in the mitochondria. In another branch of the pathway, ribose-5-phosphate reacts with ATP to generate 5-phosphoribosyl-1-pyrophosphate (PRPP), in a reaction catalyzed by phosphoribosyl pyrophosphate synthetase (PRPS) [Fox and Kelley, 1971]. PRPP is the most important ribose phosphate donor for purine and pyrimidine metabolism. Finally, the products of these two branches, orotate and PRPP, are substrates

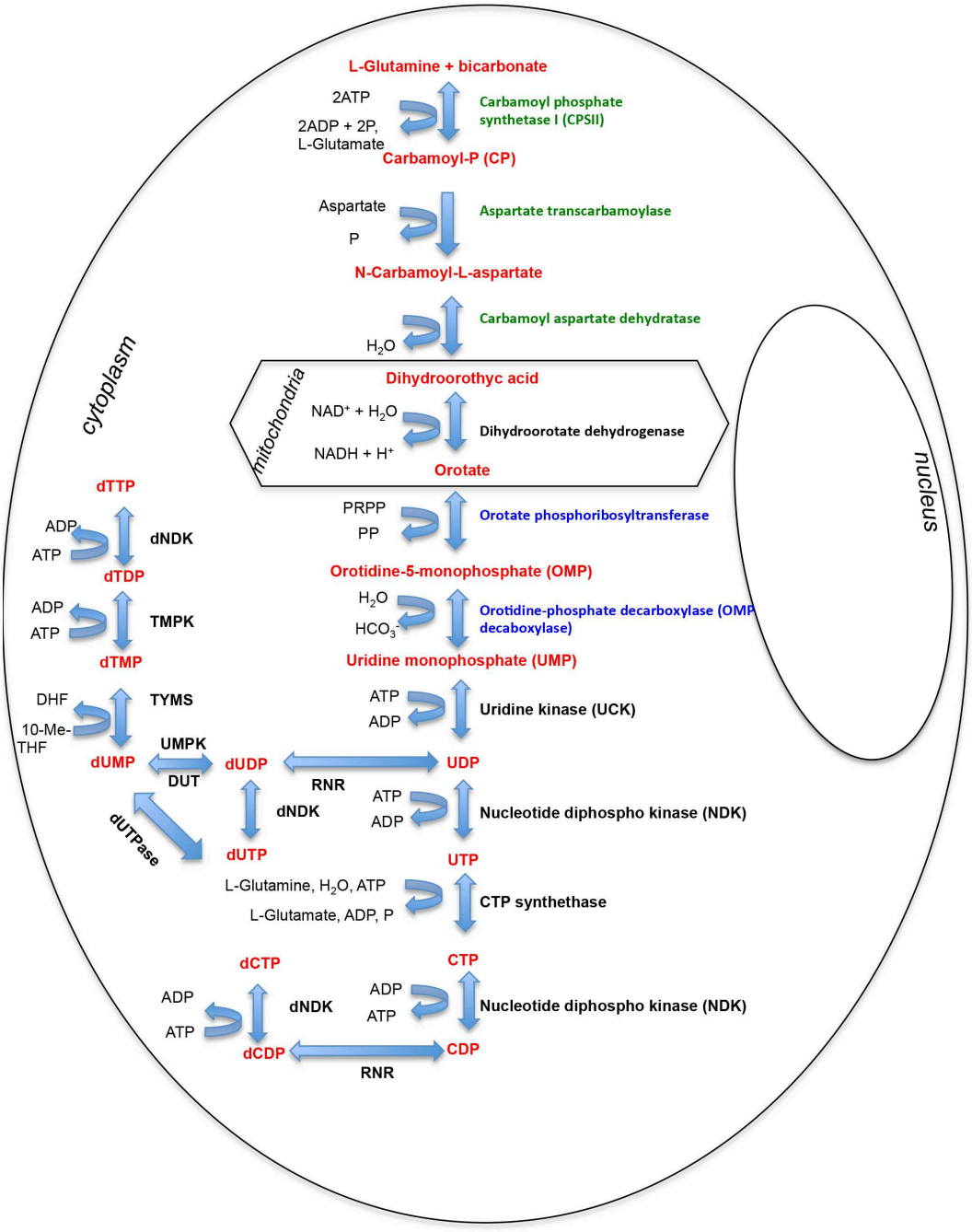


Figure 1.4: *De novo* pyrimidine synthesis. The enzyme activities shown in green and blue are encoded by the same genes in mammals (CAD and UMPS, respectively).

of orotate phosphoribosyl transferase (OPRT), which hydrolyses the pyrophosphate to produce orotidylate (OMP). Orotidylate decarboxylase subsequently converts it to uridylate (UMP). In mammals, both these activities are encoded by the uridine monophosphate synthetase (*UMPS*) gene (orotate phosphoribosyl transferase and orotidine-5'-decarboxylase) [McClard et al., 1980]. UMP-CMP kinases (UCK) then converts it into uridine diphosphate (UDP) [Van Rompay et al., 2001], with the subsequent phosphorylation step carried out by nucleoside diphosphate kinase (NDK encoded by *NME1*) which generates uridine triphosphate (UTP)[Mourad and Parks, 1966]. NDK has broad specificity over both phosphate donor and acceptor. Cytidine triphosphates (CTP) do not have an independent origin *de novo*, they are formed through the ammination of the carbamyl group of UTP, catalyzed by CTP synthethase (CTPS), which utilizes ATP and one glutamine as cofactors [Hurlbert and Kammen, 1960].

Deoxynucleotides are synthesized from the corresponding ribonucleotide diphosphates (CDP, UDP) by iron-dependent reductions catalyzed by ribonucleotide reductase (RNR) which, in mammals, is encoded by the *RNR1* and *RNR2* genes [Moore and Reichard, 1964]. This enzyme has been shown to be regulated by proliferation, being mostly active during the S phase of the cell cycle to ensure the maintenance of correct deoxynucleoside diphosphate pools [Turner et al., 1968].

Phyrophosporolysis of dUTP by deoxyuridine triphosphatase (DUT, dUTPase) leads to dUMP, protecting the cell from the dangerous incorporation of dUTP into the DNA [Ladner et al., 1996] and allowing the bio-synthesis of thymidine triphosphate (dTTP). Thymidylate (dTMP) is produced through the methylation of dUMP by thymidylate synthetase (TYMS) in the presence of methylene tetrahydrofolate [Davisson et al., 1989]. Subsequently, thymidylate kinase (TMPK) catalyzes the formation of dTDP [Lee and Cheng, 1977], which is triphosphorylated by nucleoside diphosphate kinase (NDK) to give TTP [Mourad and Parks, 1966].

It has been shown that bacterial polymerases cannot distinguish between dUTP and dTTP

in vitro [Bessman et al., 1958], therefore, synthesis enzymes like TYMS and TMPK, together with DUT, play the major role in the maintenance of cellular nucleotide pools [Reichard, 1988]. Further supporting this line of evidence, it has been reported that the inhibition of TMPK is able to sensitize tumor cells to DNA damaging agents, following dUTP incorporation [Hu et al., 2012].

1.1.2 The salvage pathway

Salvage pathways have primarily been studied in yeast and bacteria, where single gene deletions have helped to elucidate the chain of events that leads to nucleotide production. The salvage route for nucleotide production is energy efficient, as only one mole of ATP is necessary for one mole equivalent of each purine, whereas *de novo* synthesis would require six. Evidence for the presence of a nucleotide salvage pathway in mammals came from early studies in which radio-labeled bases were fed directly to animals in their diets, and then traced back into nucleic acids. The functional relevance of the pathway is manifold, and not only related to energy efficiency. In resting or G1 cells where *de novo* synthesis of DNA precursors is absent, the salvage pathway is the sole provider of deoxyribonucleotides to be used in DNA repair or mitochondrial DNA replication. Some enzymes (TK and dGK1) of the pathway localize to the mitochondria and provide precursors for mitochondrial DNA replication. Moreover, together with the enzymes of the *de novo* pathway, the enzymes of the salvage pathway strive to maintain balanced nucleotide pools, fundamental for genome stability [Mathews, 2014].

Substrates for the salvage pathway originate from both intracellular and extracellular sources. Among the intracellular sources of bases, nucleosides and nucleotides there are catabolic processes such as: the degradation of RNA; nucleotide sanitation mechanisms responsible for maintaining a balanced cellular nucleotide pool [Nagy et al., 2014]; DNA damage repair pathways, such as base excision repair; and mitophagy, which facilitates the selective degradation of mitochondria, documented in some cell types [Youle and Naren-

dra, 2011]. Diet-derived nucleotides, particularly from foods with a high cellular density such as legumes and meat, provide an extracellular source of nucleosides and bases. These foods are digested into nucleotides by pancreatic nucleases and phosphodiesterases, and to nucleosides by intestinal alkaline phosphatase (ALP) and nucleotidases [Bianchi and Spychala, 2003]. Tracer studies in animals indicate that 2 to 5% of dietary nucleotides are incorporated into tissue pools, primarily within the small intestine, which is known to lack *de novo* nucleotide synthesis [Carver and Allan Walker, 1995]. These processing events are necessary since only bases or nucleosides can gain entry into the mammalian cell through the use of dedicated channels.

Nucleoside transporters fall into two classes: concentrative (CNT) and equilibrative (ENT), encoded by the *SLC28* and *SLC29* gene families, respectively. Members of these protein families are broadly expressed among human tissues. CNTs mediate the unidirectional flow of nucleosides in an active process coupled to a transmembrane sodium gradient. SLC28 only contains three members: CNT1 (*SLC28A1*), CNT2 (*SLC28A2*), and CNT3 (*SLC28A3*). All three of these proteins can accept uridine as substrate, but differ with respect to selectivity for other substrates. Thus, CNT1 prefers pyrimidine nucleosides, CNT2 prefers purine nucleosides and CNT3 transports both pyrimidine and purine nucleosides. The sodium/nucleoside coupling ratio of CNT1 and CNT2 seems to be 1:1, while CNT3 shows a 2:1 stoichiometry [Pastor-Anglada et al., 2008]. The SLC29 gene family has four members: *SLC29A1* (ENT1), *SLC29A2* (ENT2), *SLC29A3* (ENT3) and *SLC29A4* (ENT4). These proteins, except ENT4, mediate the diffusion of natural nucleosides with broad selectivity but relatively lower affinity than their CNT-counterparts. ENT3 has been found to be localized on the mitochondrial membrane mediating intracellular nucleoside transport. ENT4 is a cation transporter that has poor affinity for nucleosides [Young et al., 2008]. Studies with nucleoside analog have shown that non-canonical routes can be responsible for the import of nucleosides into the cell, like peptide transporters of the *SLC15* family, which includes four members, and organic cation and anion transporters

of the *SLC22* family which has 25 members. The expression and localization of these channels is mainly limited to the kidney, the liver and the intestine [Cano Soldado and Pastor-Anglada, 2012].

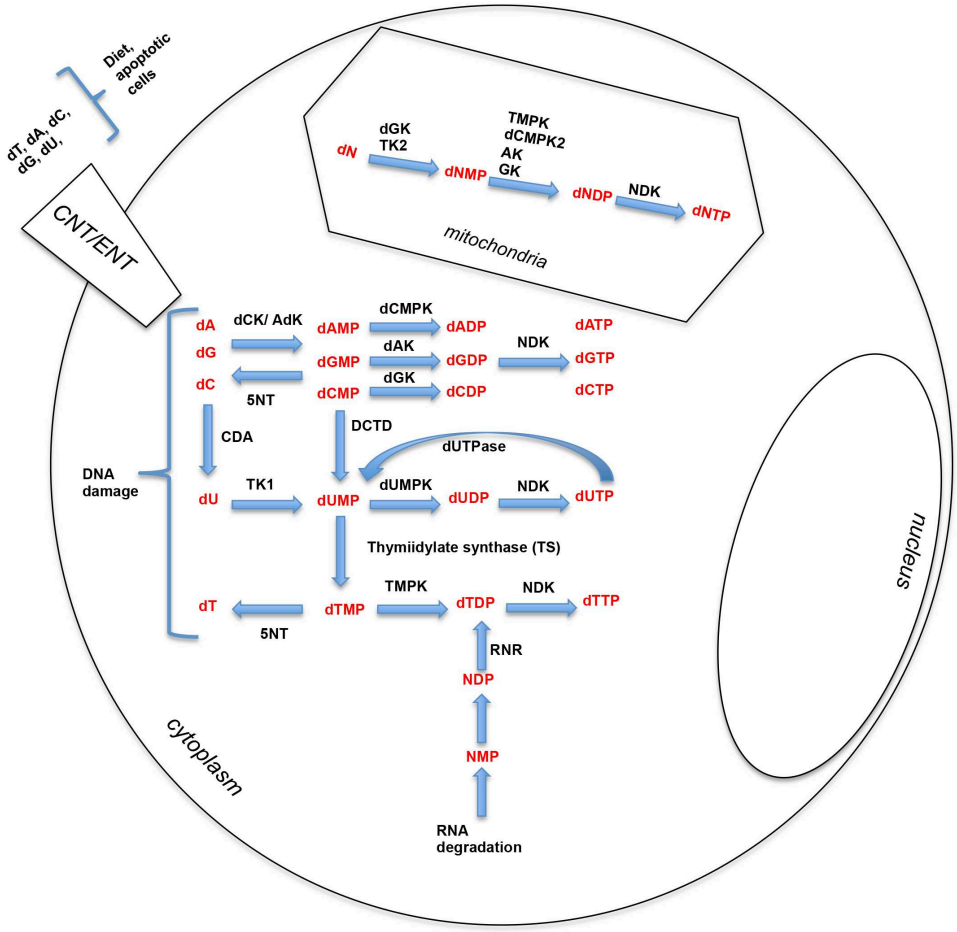


Figure 1.5: The deoxyribonucleosides salvage pathway.

Purine nucleosides salvage

The first proof of the existence of a purine salvage pathway came from studies showing the incorporation of labeled adenine and guanine from diets into nucleic acids [Brown and Roll, 1948]. Free purines can be converted to ribonucleosides together with PRPP, in a

reaction catalyzed by phosphoribosyltransferases and then salvaged following the *de novo* synthesis route. In particular, adenine is converted to adenylate (dA) by adenine phosphoribosyltransferase (APRT) [Holden et al., 1979], with the highest activity observed in the brain [Rosenbloom, 1967]. Guanylate and inosinate, precursor for AMP and GMP are formed from guanine or inosine plus PRPP by the action of hypoxanthine-guanine phosphoribosyltransferase (HGPRT), which shows its greatest activity in brain [Rosenbloom, 1967]. Deoxyribonucleosides can be converted by deoxynucleoside kinases to their monophosphate forms. Adenine is converted by adenosine kinase (AdK) and deoxycytidine kinase (dCK) and guanine by deoxyguanine kinase (dGK), which resides in the mitochondria [Arnér and Eriksson, 1995]. The subsequent diphosphorylation step is catalyzed by adenine kinase (AK), of which 9 members are known [Panayiotou et al., 2014], and guanylate kinases (GUK) [Van Rompay et al., 2000]. The last phosphorylation step is carried out by nucleotide diphospho kinase (Fig. 1.5).

Pyrimidine nucleoside salvage

Given the ability of cells to synthesize pyrimidine bases *de novo*, pyrimidine salvage was not elucidated until isotopically-labeled pyrimidines became available. Early studies showed that uracil and cytosine could be incorporated into nucleic acids if administered to cells or fed to mice [Lagerkvist et al., 1955]. In particular, a direct link between uracil administration to cells and the inhibition of the rate of *de novo* nucleotide synthesis was established by the discovery of an inhibitory role of dU towards OPRT [Hoogenraad and Lee, 1974]. Uracil can be salvaged to form UMP through the concerted action of uridine phosphorylase (UPP) which catalyzes the conversion of uracil to uridine [Pontis et al., 1961] and uridine kinase (UCK) which produces UMP [Van Rompay et al., 2001]. dU can be monophosphorylated by TK1 and subsequently by CMPK1-UMPK, which has a variant in the mitochondria. The salvage of thymine first involves its conversion to thymidine through a thymidine phosphorylase (TYMP)-catalyzed reaction [Friedkin and Roberts,

1954a] [Friedkin and Roberts, 1954b]. Subsequently, monophosphorylation is caused by thymidine kinase (TK) which produces TMP partly in the mitochondria [Van Rompay et al., 2000] (Fig. 1.5).

The salvage of deoxycytidine is catalyzed by deoxycytidine kinase (DCK) that converts it into dCMP [Durham and Ives, 1970]. This enzyme is present mainly in lymphoid tissues and it is greatly inhibited by its own products [Durham and Ives, 1969]. At the ribonucleotide level, CMP can be formed by UCK which is able to use cytidine as a substrate, as well as uridine [Van Rompay et al., 2001]. Another route that can be undertaken for the salvage of dC or dCMP is their respective deamination by cytidine deaminase (CDA) and deoxycytidylate deaminase (DCTD), which contributes to the formation of DNA by providing dU and dUMP as substrates for thymidine kinase (TK) and thymidylate synthase (TYMS). A role for the involvement of CDA in DNA synthesis was suggested from early studies that noted elevated levels of it in rapidly dividing tissues such as tumors, embryos and the regenerating liver [Maley and Maley, 1959][Maley and Maley, 1960][Camiener and Smith, 1965][Laliberté and Momparler, 1994]. Additionally, elevated levels of DCTD were found in the serum of patients with various disease states associated with cellular proliferation [Miller and Ressler, 1969]. Further emphasizing its potential role in cell division and the link with thymidine, dC/dCMP deaminases from most sources have been found to be allosterically regulated by the end-products of their metabolic pathways, dCTP and dTTP, with the former acting as an activator and the latter as an inhibitor [McFerran et al., 1969][Maley and Maley, 1972]. This fine degree of control, not only provides a regulated flow of dCTP and dTTP for DNA synthesis, but also prevents the introduction of mutations into the DNA, which can result as a consequence of restricted or defective levels of dCMP deaminase [de Saint Vincent et al., 1980] [Sargent and Mathews, 1987]. CMP/dCMP can be then phosphorylated by UCK/CMPK1-UMPK to produce CDP/dCDP, which can be converted to triphosphates by nucleoside diphospho-kinases. Phosphorolysis for dC has never been documented in mammals [De Verdier and Potter,

1960], therefore dC cannot became part of sugar metabolic sources, but is instead forced to enter the nucleotide route (Fig. 1.5).

1.1.3 Nucleotide sanitation pathways

De novo nucleotide synthesis has been shown to be mainly coupled to replication and thus to the synthesis phase (S) of the cell cycle. Nucleotide pools are therefore established in a critical phase of mitosis. Maintenance of a correct balance of nucleosides triphosphates is therefore of fundamental importance for the cell. Responsible of this task in mammalian cells are the enzymes: SAM domain and HD domain-containing protein 1 (SAMHD1) and dCTP pyrophosphatase (DTTP1) which have been shown to hydrolize respectively dATP, dCTP, TTP and dCTP [Goldstone et al., 2011][Requena et al., 2014]. Moreover, it is crucial that canonical nucleoside triphosphates are maintained in the pool to prevent the erroneous incorporation of non canonical nucleotides into the DNA by DNA polymerases. Non-canonical nucleotides can originate from reactive oxygen species (ROS) induced oxidation of nucleotides (generating 8-oxo-GTP or 2-OH-dATP from guanidine or adenosine), base excision repair of DNA damage byproducts, or from *de novo* synthesis of nucleotides resulting in small amounts of dUTP and dITP. Moreover, it has been shown that different variants of DNA polymerase, which incorporate dATP, dGTP, dCTP and dTTP into the nascent DNA, are highly selective against ribonucleotides, but have only a limited selectivity with respect to nitrogenous bases, even using substituted benzimidazoles as substrates [Kincaid, 2005]. Therefore, the interception and hydrolysis of non-canonical dNTPs is a way of directly preventing DNA damage [Galperin et al., 2006]. To perform this task, the cellular machinery has evolved "house-cleaning" NTP pyrophosphatases that target non-canonical NTPs. These enzymes belong to at least four structural superfamilies: MutT- related (Nudix) hydrolases, dUTPase, ITPase (Maf/ HAM1) and all-α NTP pyrophosphatases (MazG). One of the best studied is MTH1, which in humans catalyzes the hydrolysis of 8-oxo-dGTP, as well as two more oxidized NTPs, 2-oxo-dATP

and 8-oxo-dATP, and prevents their incorporation into the DNA. Another well studied member is dUTPase, which is able to dephosphorylate dUTP, preventing its incorporation into DNA. In this enzyme, tight contact between the buried dU C5 atom and the protein atoms ensures specificity towards U and discrimination against the extra 5-methyl group of thymine, as well as variants of uridine substituted at the fifth carbon position [Galperin et al., 2006]. The newly discovered Maf proteins display high specificity towards modifications at the fifth carbon position of cytidine triphosphates [Tchigvintsev et al., 2013]. Most of the NTPs have been shown to be non-essential if inactivated, but display a synthetic lethal phenotype if inactivated with other proteins [Galperin et al., 2006].

1.1.4 Erroneous incorporation induces a DNA damage response

The erroneous incorporation of non-canonical nucleotides into DNA results in the induction of a DNA damage response that facilitates the enzymatic removal of these bases from the DNA, and the induction of cell cycle arrest to give the cell time to repair it [Branzei and Foiani, 2008]. Evidence for enzymes catalyzing such active removal from the DNA came with the discovery of base excision repair activities associated with the aberrant presence of non-canonical nucleotides in the DNA of glycosylase knockout mice. In Mth1-null mice, an increase in 8-oxoG content was detected in nuclear DNA, suggesting that MTH1 plays a role in avoiding errors in the nuclear genome caused by the aberrant incorporation of oxidized purine nucleoside triphosphates such as 8-oxo-dGTP and 2-OH-dATP [Tsuzuki et al., 2001]. Once incorporated, these bases are recognized by DNA glycosylases, which excise them, creating abasic sites and initiating the base excision repair (BER) pathway. 8-oxoG DNA glycosylase, encoded by the *OGG1* gene, is another glycosylase that is able to excise 8-oxoG as a free base from DNA. Furthermore, during replication, DNA polymerases can insert adenine opposite 8-oxoG in the nascent strand. Therefore, another glycosylase activity encoded by the *MUTYH* gene is necessary to excise 8-OxoG in this context. MUTYH protein also has the ability to excise 2-OH-A incorporated opposite to

guanine. 2-OH-A in the DNA can only be derived from the incorporation of 2-OH-dATP at replication, since oxidation of adenine in the DNA is barely detectable [Nakabeppu et al., 2006].

Uracil, when erroneously incorporated into the DNA or resulting from enzymatic deamination of cytosine, is excised from the genome by uracil-DNA glycosylases (UDGs). Mammalian cells contain four classes of UDGs: uracil-DNA glycosylase (UNG1, UNG2), single-strand-selective mono-functional uracil-DNA glycosylase (SMUG1), T/U mismatch DNA glycosylase (TDG) and methyl-binding domain 4 protein (MBD4). UNG2 (a nuclear form of uracil-DNA glycosylase) is the main enzyme involved in the repair of uracil incorporated into the DNA (as A:U pair), while all UDGs contribute to the repair of U:G pairs. It was found that, while UNG2 is adapted to the fast and highly coordinated excision of uracil from U:G and U:A pairs in replicating DNA, the less efficient SMUG1 may be more important in the repair of deaminated cytosine in non-proliferating cells or proliferating cells outside S phase. In all these cases an apyrimidinic site is created until the base excision repair pathway is initiated [Visnes et al., 2009][Olinski et al., 2010][Hashimoto et al., 2012b][Hashimoto et al., 2012c]. Mbd4 deficient mice show a mutator phenotype with increased C:G to T:A transversions [Millar et al., 2002], while Tdg null mice, which are embryonic lethal, do not display a mutator phenotype in mouse embryonic fibroblast (MEF)-derived cells, highlighting possible alternative roles of this DNA glycosylase [Cortázar et al., 2012].

1.1.5 Erroneous nucleotide incorporation, mutagenicity and cancer

If an increase in dUTP levels overwhelms cellular dUTPase activity, dUTP may accumulate. Such abnormal accumulation may result in extensive incorporation into DNA, which may be followed by glycosylase mediated repair. As a consequence AP sites may be generated, resulting in detrimental DNA damage and increased mutagenesis [Olinski et al., 2010].

Most uracil DNA glycosylase knockout mouse models are viable given their redundant functions. Although, a combined deficiency of Smug1 and Ung exacerbated the cancer predisposition of Mismatch repair 2 (Msh2) null mice, suggesting that when both base excision and mismatch repair pathways are defective, the mutagenic effects of dU incorporation are sufficient to increase cancer incidence but not to preclude mouse development. It also was shown, that triple knockout mice develop lymphoid tumors [Kemmerich et al., 2012], while Mbd4-deficient mice develop tumors in the gastrointestinal tract if combined with the oncogenic adenomatous polyposis coli (Apc) I638N mutation [Wong et al., 2002]. Given the Tdg-null mice embryonic lethality, its role in carcinogenesis has not yet been fully explored [Cortázar et al., 2012].

As for uracil incorporation, the oxidation of DNA appears to result in either spontaneous mutagenesis or cell death and, as a result, has been implicated in various diseases such as cancer. Among the different types of oxidative damage that affect DNA, 8-oxoguanine (8-oxoG) and 2-hydroxyadenine (2-OH-A), namely the oxidized forms of guanine and adenine, can form relatively stable base pairs with adenine or guanine in DNA (respectively) during DNA replication. Thus, this is considered to be a spontaneous cause of mutagenesis, such as A:T to C:G and G:C to T:A transversion mutations. Most of the enzymes responsible for their removal show an increase in tumor development, when ablated in mice. Mth1 efficiently hydrolyzes two forms of oxidized dATP, 2-hydroxy(OH)-dATP and 8-oxo-dATP, as well as 8-oxo-dGTP and 8-oxo-GTP. Null mice, despite their weak mutator phenotype, exhibit a several-fold increased incidence of spontaneous tumorigenesis in the liver, in which 8-oxoG content in the nuclear DNA is increased, in comparison to the wild-type [Tsuzuki et al., 2001]. Ogg1-null mice were found to develop spontaneous lung adenocarcinomas, accompanied by an increased accumulation of 8-oxoG in their nuclear genomes in comparison to wild-type mice [Nakabeppu et al., 2006]. Lastly, Mutyh null mice show an increased occurrence of intestinal adenoma or adenocarcinoma [Nakabeppu et al., 2006].

Given the primary role of DNA glycosylases in the DNA damage response, their therapeutic inhibition might sensitize cancer cells, whose nucleotide pools contain high levels of ROS-induced oxidative lesions, to cell death. Two recent reports have shown that the inhibition of MTH1 in cancer increases its susceptibility to incorporation into the DNA of oxidized nucleotides, resulting in induction of the DNA damage response and cell death [Gad et al., 2014][Huber et al., 2014].

1.1.6 Approved nucleoside and nucleotide analogs in cancer therapy

A variety of nucleoside and nucleotide analogs are currently FDA approved for the treatment of different kinds of cancer. They usually mimic the structure of canonical nucleotides (Fig. 1.6), thus interfering with normal DNA synthesis (Table 1.1).

2-Fluoroadenosine phosphate is dephosphorylated by extracellular phosphatases and then

Nucleoside analog	Target	Diseases
Purine analogs		
2-Chlorodeoxyadenosine	DNA Replication	HCL, NHL
2-Fluoroadenosine phosphate	DNA Replication	CLL
Clofarabine	DNA Replication	ALL
Nelarabine	DNA Replication	T-cell ALL
2-deoxycoformycin	DNA Replication	HCL, CLL
Pyrimidine analogs		
Cytosine β-D-arabinofuranoside	DNA Replication	AML, ALL
Gemcitabine	DNA Replication	PC, LC, BC, BlC
Fluoropyrimidines		
Fluorouracil	DNA Replication	GIC, HNC, RC, SC , PrC, BC
Capecitabine	DNA Replication	Relapsed BC and CRC

Table 1.1: FDA-approved nucleoside analogs for cancer treatment. Hairy cell leukemia (HCL), Chronic lymphocytic leukemia (CLL), Acute myelogenous leukemia (AML), Acute lymphoblastic leukemia (ALL), non-Hodgkin lymphoma (NHL) Pancreatic cancer (PC), Lung cancer (LC), Breast cancer (BC), Bladder cancer (BlC), Gastrointestinal cancer (GIC), Head and neck (HNC), Renal cancer (RC), Skin cancer (SC), Prostate cancer (PrC), Colorectal cancers (CRC) [Jordheim et al., 2013].

enters the cell. All purine and pyrimidine analogs are monophosphorylated by DCK or

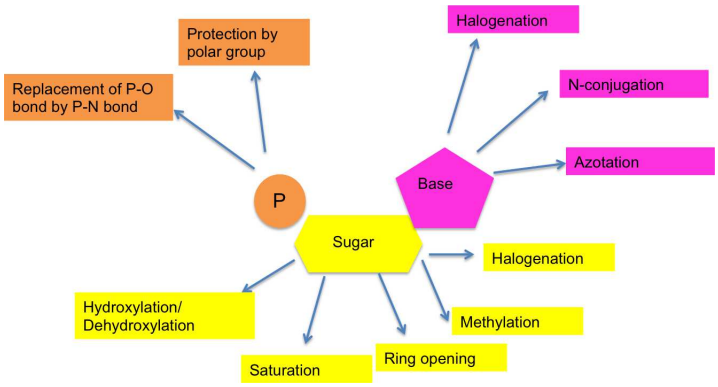


Figure 1.6: Chemical characteristics of nucleoside and nucleotide analogs. Nucleoside and nucleotide analogs consist of a nucleobase (a purine or pyrimidine derivative) linked to a sugar moiety and in case of nucleotides, a phosphate group (P) linked to the sugar moiety. The chemical diversity of these compounds is based on the modifications shown [Jordheim et al., 2013].

mitochondrial dGK and converted to triphosphate through the canonical purine route. Once they are incorporated into DNA, chain elongation mediated by DNA polymerases is terminated, inducing apoptosis in replicating cells. These compounds also indirectly impair DNA replication by inhibiting ribonucleotide reductase, consequently reducing the pool of deoxynucleotide triphosphates (dNTPs) available for DNA synthesis [Galmarini et al., 2002] [Bonate et al., 2006] [Jordheim et al., 2013].

Flourouridine acts as an inhibitor for thymidylate synthase (TS), which is expressed in highly proliferative cells where it converts dUMP into dTMP. TS inhibition creates an imbalance in the nucleotide pools toward dUTP, leading to its increased incorporation, DNA damage and a mutator phenotype [Longley et al., 2003].

Resistance mechanisms in cancer cells are due to the downregulation of nucleoside transporters or intracellular nucleoside kinases such as deoxycytidine kinase (DCK), as well as increased activity of ribonucleotide reductase and the expression of nucleotidases. The levels of possible metabolizing enzymes must be carefully checked, because they can mediate the increased toxicity of these drugs. This is the case for decreased cytidine deaminase activity in clinical blood samples of patients treated with gemcitabine, which results in increased drug toxicity [Jordheim and Dumontet, 2007].

1.2 Cytosine biological DNA modifications

After the initial discovery of the canonical base components of nucleic acids [Levene and Jacobs, 1909] [Levene, 1910] at the beginning of the twentieth century, Hotchkiss reported the discovery of a variant of cytosine in calf thymus DNA in 1948, which he called "epicytosine" since it migrated in close proximity to cytosine in paper chromatography preparations. In the paper, he speculated that it might have been 5-methylcytosine (5mC) (Fig. 1.7 a), given its previously documented presence in *Bacillus tuberculosis* DNA [Johnson and Coghill, 1925]. In addition, its spectral properties related to cytosine, as the ones of uracil to thymine, where thymine differs from uracil for the presence of a methyl group at the fifth carbon position [Hotchkiss, 1948]. Subsequent studies confirmed the finding in plant, in bacterial and mammalian DNA [Wyatt, 1951]. Nowadays we know that 5mC is present throughout the kingdom of life in: the fruit fly *Drosophila melanogaster* [Jaenisch et al., 2000][Gowher, 2000][Capuano et al., 2014]; the fish *Danio rerio*[Yamakoshi and Shimoda, 2003]; the sea squirt *Ciona intestinalis*[Simmen et al., 1999], which is the closest invertebrate to vertebrates; the fungi *Neurospara crassa*[Bull and Wootton, 1984], *Sporotrichum dimorphosporum* and *Phycomyces blakesleeanus* [Antequera et al., 1984]; in plants like the algae *Chlamydomonas reinhardii*[Hattman et al., 1978]; the amoeba *Dictyostelium discoidium* [Katoh et al., 2006]; the parasitic worm *Schistosoma mansoni* [Geyer et al., 2011] and the bacteria *Escherichia coli* [Capuano et al., 2014]. Both the nematode worm *Caenorhabditis elegans* and various species of yeast lack DNA methylation.

Another modified form of cytosine, 5-hydroxymethyldeoxycytosine (5hmC) (Fig. 1.7 b) was first published in 1972 as being present in brain and liver rat DNA [N W Penn, 1972], but no paper could reproduce the finding until 2009, when two reports showed the presence of 5hmC in Purkinje cells, brain [Kriaucionis and Heintz, 2009] and in embryonic stem cells [Tahiliani et al., 2009]. To date, 5hmC has been detected in mice, humans [Lister et al., 2013], *Danio rerio* [Jiang et al., 2013], *Arabidposis thailiana* [Liu et al., 2013c], *Xenopus*

(a) 5-methlcytosine (b) 5-hydroxymethlcytosine (c) 5-formylcytosine (d) 5-carboxylcytosine

Figure 1.7: Cytosine variants in mammalian DNA

laevis [Xu et al., 2012] and chicken [Liu et al., 2013c]. In mouse DNA, while 5mC content is relatively constant across tissues, the highest levels of 5hmC are detected in the brain cerebral cortex, while medium levels are found in kidney, lung, bladder, and muscle and low levels in the liver and spleen [Globisch et al., 2010].

Subsequent studies have discovered further oxidative modifications of 5mC such as 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) [Ito et al., 2011] (Fig. 1.7 c, d), bringing the total number of cytosine variants detected in DNA to date to four (Fig. 1.7). 5fC and 5caC are present at very low levels in mouse and human DNA [Liu et al., 2013b] and have also been detected in *Arabidposis thailiana* DNA [Liu et al., 2013c].

1.2.1 Enzymatic production of cytosine variants

Soon after the discovery of 5mC, research efforts moved towards understanding whether this widespread base originated from enzymatic synthesis of the nucleotide, or deposition of the methyl mark onto the genome. The first hints of a possible solution for this dilemma, came with the discovery that despite 5mdCTP could be incorporated into the DNA, there was no kinase in *E. coli* that could produce a triphosphate from 5mdC [Bessman et al., 1958]. Therefore, the focus was soon moved towards identifying the enzymatic activity responsible for 5mC production [Kornberg et al., 1959]. In the meantime, accumulating evidence linked 5mC to a variety of functions. Following the purification of a DNA/RNA methylating enzyme [Gold and Hurwitz, 1964] from *E. coli*, the first functional implication for 5mC came with the discovery of a DNA methylation-dependent host restriction mechanism in *Bacteria* upon phage infection [Dussoix and Arber, 1962][Urs Kühnlein, 1969].

The path towards the identification of the first mammalian enzymes was tightly linked to functional studies given the unavailability of genomic technologies. Therefore, following their activity, S-adenosylmethionine (SAM)-dependent DNA methyltransferases (DNMT) were isolated in mammalian systems [Shied et al., 1968][Bestor and Ingram, 1983]. With the advent of restriction enzymes and recombination technologies, the sequence encoding mouse DNMT1 was found through Edman degradation of the corresponding purified enzyme, oligonucleotide construction and annealing [Bestor et al., 1988].

1.2.1.1 The mammalian DNA methyltransferase family

Three DNMT activities were initially found in mammalian cells (mouse eritroleukemia) [Bestor and Ingram, 1985], and were found to be strictly linked with cell cycle phases. In particular, log-phase cells contained DNA methylase III (Mr 190,000), cells approaching stationary phase contained DNA methylase II (Mr 175,000), and stationary-phase cells contain DNA methylase T (Mr 150,000) [Bestor and Ingram, 1985]. It is now known that the DNMT family is composed of 3 active members: DNMT1, DNMT3A and DNMT3B (Fig. 1.8). During the catalytic reaction, DNMTs binds covalently to position 6 of the cytosine ring and after transfer of the methyl group, coming from S-adensyl methionine (SAM), a β-elimination reaction frees the enzyme.

DNMT1 has a preference for hemimethylated DNA and, for this reason, it was assigned the function of maintaining DNA methylation across each cellular generation [Yoder et al., 1997a]. It possesses a large N terminal regulatory domain containing a nuclear localization signal (NLS), replication *foci* targeting sequences (RFTS) [Leonhardt et al., 1992], a DNA binding domain (CXXC) and protein interaction domains (BAH). The C-terminal region possesses the methyltransferase (MTase) domain. It has been found that DNMT1 forms discrete nuclear *foci* during replication coherent with its *de novo* activity tightly linked to the replication machinery [Leonhardt et al., 1992]. Given its fundamental role, Dnmt1 knockout mice were found to be not viable [Li et al., 1992] and embryonic stem cells de-

rived from them failed to undergo differentiation. A transcriptional product initiated from an ATG in exon 4 produces DNMT1O [B E Hayward, 2003], a variant only expressed in oocytes, where it becomes excluded from the nucleus at fertilization and translocates into the cytoplasm until re-enters the nucleus transiently at the 8-cell stage. Dnmt1o knockout oocytes, fail to establish imprinting at the 8-cell stage after fertilization [Howell et al., 2001].

DNMT3 activity is encoded by 3 different genes. *DNMT3A* and *DNMT3B* were identified with a homology search in EST databases using full-length bacterial type II cytosine-5 methyltransferase as query [Li et al., 1998]. Later, it was found that both Dnmt3a and Dnmt3b single knockout mice are embryonic lethal [Okano et al., 1999]. Moreover, regions that are found unmethylated in ICF patients, with heterozygous mutations in *DNMT3B*, are satellite DNA and regions subject to X chromosome inactivation, which are newly methylated after birth. This group of DNMTs was therefore assigned the role of *de novo* DNA methyltransferases. Additional proof for their *de novo* methylating activity came from double knockout cell lines, which failed to methylate a retroviral genome after infection, as opposite to wild type cells [Okano et al., 1999]. A catalytic inactive isoform was later identified through bioinformatic analysis: DNMT3L [Aapola et al., 2000]. Knockout mice for Dnmt3l are lethal, due to a failure to develop extraembryonic tissues [Bourc'his et al., 2001]. Moreover, DNMT3L has been shown to be required for the establishment of methylation patterns during gametogenesis [Bourc'his and Bestor, 2004]. Furthermore, it was shown, that DNMT3L binds the tail of histone 3 when K4 is unmethylated [Ooi et al., 2007]. DNMT3s are quite different from DNMT1, possessing an ATRX-DNMT3-DNMT3L (ADD) domain which is Cys rich and a Pro-Trp-Trp-Pro motif (PWWP), together with their methyltransferase domain (MTase). The exception is DNMT3L, which contains only the ADD domain. The PWWP motif has been shown to be required for the DNA binding [Qiu et al., 2002] and the ADD domain for the interaction with the H3 histone tail which is abolished when methylated at lysine 4 [Zhang et al., 2010b].

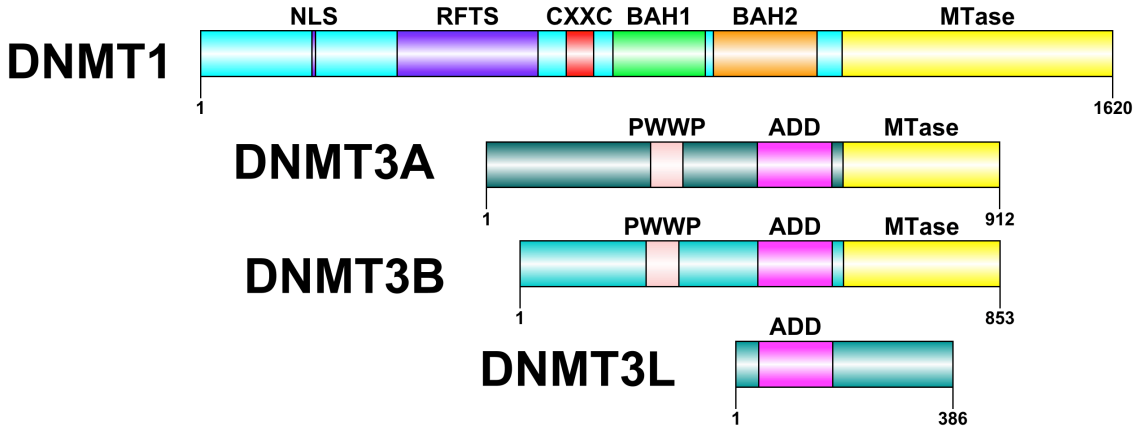


Figure 1.8: Schematic representation of mammalian DNA methyltransferases. Nuclear localization signal (NLS), replication foci targeting sequences (RFTS), DNA binding domain (CXXC) and bromo adjacent homology (BAH), Pro-Trp-Trp-Pro motif (PWWP), ATRX-DNMT3-DNMT3L (ADD), DNA methyltransferase (MTase).

1.2.1.2 Mammalian DNA oxygenases: the TET family

The TET family of proteins was named after the ten-eleven translocation found in leukemia, where the C-terminal part of TET1 on chromosome 11, was found fused to the N-terminal part of MLL on chromosome 10 [Lorsbach et al., 2003]. More than five years passed until this protein was assigned, through bioinformatic studies, to the family of 2-oxoglutarate and iron (II) dependent oxygenases (2OGFeDO), which catalyze the oxidation of 5mC to 5hmC [Iyer et al., 2009]. Previously, this modification had been detected and studied in the DNA of T-even phages, where, together with 5-hydroxymethyluracil (5hmU) dedicated synthases exist, derived from the classical thymidylate synthases, that catalyze their production as nucleoside triphosphates. Thus, phage 5-hydroxymethylpyrimidines are not derived by direct DNA modifications but by the incorporation of pre-modified base during viral DNA synthesis. Furthermore, it has been shown that phages are able to modify them, once they are incorporated into the DNA, via the action of DNA base glycosyltransferases. The starting point for the assignment of a function to the TET proteins, was a bioinformatic search for mammalian homologs of enzymes that had been

shown to produce another similar modification, base J (oxidation of the methyl group on thymine to generate 5hmU) in trypanosomes. This reaction occurs *in situ* on DNA and it is catalyzed by JBP1 and JBP2, enzymes of the 2OGFeDO family. Therefore, through the alignment of sequences to search for homologs of JBP1/2, the Tet subfamily of protein was found. During the evolution of the gnathostome lineage of vertebrates, the Tet subfamily underwent a triplication to generate the *TET1*, *TET2* and *TET3* genes, which are conserved in all gnathostomes [Iyer et al., 2009]. These proteins possess the canonical double-stranded β-helix dioxygenase domain (DSBH), which contains three signature motifs conserved among 2OG- and iron (II)-dependent dioxygenases: histidine, any aminoacid, aspartic acid (HxD); histidine, any aminoacid, small residue (Hxs); arginine, 5 aromatic residues, arginine (Rx5a/R). These motifs mediate their catalytic activity by binding, respectively, the first, divalent iron and the last α-ketoglutarate [Iyer et al., 2009]. Upstream of the DSBH domain there is a cysteine-rich region (Cys R), which has been proposed to serve as a redox center for oxidoreductase activity [Zhang et al., 2010a]. Moreover, TET1 and TET3 have been shown to possess CG-containing DNA binding activity encoded by a CXXC domain [Zhang et al., 2010a][Xu et al., 2011][Xu et al., 2012], characterized by two CGXCXXC repeats. TET2, which lacks this domain, has been shown to interact with IDAX (CXXC4). IDAX is encoded by a neighboring gene that is thought to have undergone inversion, and thus separation from the coding region of TET2 following the triplication of the TET genes [Ko et al., 2013]. The CXXC domain of the TET family is non-canonical, since it lacks the typical KFGG motif that is found in the majority of characterized CXXC domains. Furthermore, the CXXC domain of TET1 exhibits greater binding to CG-methylated DNA oligonucleotides than other CXXC domains, which normally favor unmethylated CpGs [Zhang et al., 2010a] (Fig. 1.9).

The TET enzymes were experimentally validated, as catalyzing the formation of 5hmC from 5mC in a reaction dependent on iron and α-ketoglutarate in embryonic stem cells [Tahiliani et al., 2009]. After this initial validation, many studies followed that tried to

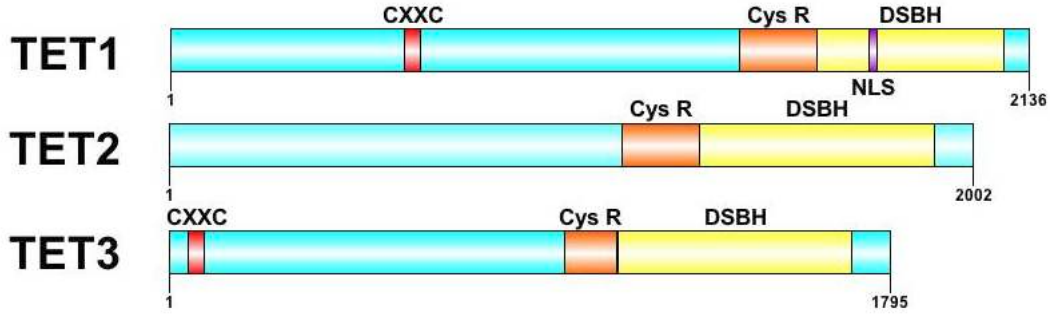


Figure 1.9: The TET family of proteins. DNA binding domain (CXXC); the cysteine-rich region (Cys r); double-stranded β-helix domain (DSBH); nuclear localization sequences (NLS).

dissect their involvement in processes where DNA methylation plays a major role, such as transcription, development and cancer. In particular, the first papers that came out proposed a role for TET-mediated 5hmC production in DNA demethylation, since a global loss of 5mC was detected when TET proteins were overexpressed [Tahiliani et al., 2009]. Furthermore, it was soon discovered that these proteins catalyzed the additional conversion of 5mC to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) [Ito et al., 2011]. Recently, another report came out documenting their role in the oxidation of thymine to produce 5hmU [Pfaffeneder et al., 2014]. The deposition of 5hmC and 5fC has been documented genome-wide in embryonic stem cells and in the brain where they have been mainly found across gene bodies, promoters and enhancers [Williams et al., 2011] [Putiri et al., 2014]. Tet family members expression profiles vary upon stem cell differentiation; whereas Tet1 and Tet2 expression is downregulated, the one of Tet3 is upregulated [Koh et al., 2011]. In mouse embryonic stem cells, TET protein stability has been shown to be controlled by the calpains, a class of calcium dependent proteases which, if inhibited, extend the half-life of TET [Wang and Zhang, 2014]. Several groups have reported a direct interaction between the catalytic domain of TET proteins and nuclear O-linked N-acetylglucosamine (GlcNAc) transferase (OGT) [Chen et al., 2012b][Deplus et al., 2013][Vella et al., 2013], which has been shown to O-GlcNacylate these proteins promoting their chromatin stability [Ito et al., 2014]. This interaction might promote the O-GlcNAc modification of histones, a

process important for nucleosome assembly and entry into mitosis [Fong et al., 2012], as well as transcription [Deplus et al., 2013]. Furthermore, given the documented presence of 5hmC in the brain, TET proteins are here expressed and have been found to mediate a variety of phenotypes in this context such as: behavioral adaptation and learning following fear conditioning with 5hmC deposition mediated by Tet3 [Li et al., 2014]; memory formation and extinction together with synaptic plasticity and hypermethylation detected at genes regulating neuronal activity in Tet1 knockout mice [Kaas et al., 2013][Rudenko et al., 2013]; neuroneogenesis with a reduction in the pool of neuronal progenitors in the adult brains of mice deleted for Tet1 [Zhang et al., 2013]; and active demethylation in the adult brain of mice overexpressing Tet1 [Guo et al., 2011].

In mESCs, Tet1 has been shown to participate in the maintenance of self-renewal through direct regulation of Nanog expression, an embryonic stem cell transcription factor implied in the maintenance of pluripotency [Ito et al., 2010]. Furthermore, mapping experiments have shown that it is present mainly at gene promoters where it interacts with the Sin3A co-repressor complex and is able to both repress and activate transcription [Williams et al., 2011] [Wu et al., 2011b]. Consistent with the finding that showed Tet1 downstream of Oct4 in embryonic stem cells (ESC) [Koh et al., 2011], Tet1 was implicated in pluripotency. In ESC it can substitute for Oct4 in the combination of transcription factors (Sox2, Klf4, c-Myc, Oct-4) that is necessary to induce pluripotent stem cells from fibroblasts [Gao et al., 2013], mainly through erasure of DNA methylation at imprinted regions [Piccolo et al., 2013]. Tet1 downregulation and knockout in embryos also lead to a bias towards trophectoderm differentiation, despite giving rise to viable mice that do not show obvious defects in development [Ito et al., 2010][Dawlaty et al., 2011]. However, a decreased number of germ cells was observed in knockout females that failed to activate the genes required for meiotic progression and gonad development [Yamaguchi et al., 2012]. In connection to cancer, Tet1 downregulation has been linked to increased tumorigenesis and metastasis through the inhibition of TIMPs (Tissue Inhibitors of Metalloproteinases), which are neg-

ative regulators of MMPs (metalloproteinases) that are known to promote cancer invasion and metastasis [Hsu et al., 2012][Liu et al., 2013a].

Tet3 is mainly expressed in oocytes, and in zygotes within the male pronucleus, where it converts 5mC to 5hmC. Consequently, knockouts of Tet3 show neonatal lethality, while conditional knockouts in the female germ line show reduced fertility and failure to convert 5mC to 5hmC [Gu et al., 2011][Guo et al., 2014]. Additionally, Tet3 has been shown to be implicated in eye development in *Xenopus laevis* and mouse. In *Xenopus*, partial rescue of this phenotype could be obtained with expression of the catalytically inactive Tet3, indicating 5hmC independent effects [Xu et al., 2012].

Double knockouts for Tet1 and Tet2 are viable, but most die perinatally within the first 2 days. The majority of homozygous Tet1/Tet2 mutants display a variety of malformations such as exencephaly, hemorrhage in the head and profound growth retardation. The mice that survive and become adults can be crossed, with females displaying reduced fertility, concurrent with the observed phenotype in Tet1 knockout ovaries [Dawlaty et al., 2013]. Their organs as DKO-derived embryonic stem cells display aberrant DNA methylation and gene expression profiles and a global reduction in 5hmC levels [Dawlaty et al., 2013]. Triple knockout of the Tet genes (TKO) allows the formation of embryoid bodies, but results in improper differentiation of the endoderm and mesoderm lineages as compared to the wild type. TKO ESCs did not contribute to chimera formation when injected into the blastocyst, indicating defects in pluripotency [Dawlaty et al., 2014].

For 5mC oxidation, the TET proteins use iron (II) and oxygen as cofactors along with α-ketoglutarate (α-KG), a metabolite of the Krebs cycle [Tahiliani et al., 2009]. α-KG can be produced and consumed through different metabolic pathways: as a key intermediate in the TCA cycle for energy metabolism, as an entry point for several 5-carbon amino acids to enter the TCA cycle after being converted into glutamate; by being reduced back to isocitrate and then citrate for the eventual synthesis of acetyl-CoA; and by being used as a co-substrate for multiple α-KG-dependent dioxygenases. Tet enzymes have been

found inhibited by 2-hydroxyglutarate (2-HG), which is produced by mutant isocitrate dehydrogenase (IDH) enzymes and has been shown to occupy the same site as α-KG. Succinate and fumarate can mimic α-KG and mutant fumarate hydrogenase and succinate dehydrogenase have been shown to inhibit Tet function [Xiao et al., 2012]. Another layer of regulation is provided by the antioxidant ascorbic acid (vitamin C) which induces rapid formation of 5hmC, 5fC, and 5caC, probably by reducing Fe(III) back to Fe(II) after enzymatic catalysis [Blaschke et al., 2013]. Lastly, in mice it was shown that injections of glucose, glutamine or glutamate induced 5hmC and 5fC formation in the DNA [Yang et al., 2014], without significantly affecting protein levels, indicating that regulation of these proteins can be tightly dependent on the bio-availability of their cofactors.

TET2

Before being assigned a function, TET2 had been discovered as the member of the TET family most frequently mutated in cancer, where more than a thousand mutations have been detected, especially in blood cancers such as acute myeloid leukemia (AML), with an overall mutation percentage of 15% [Forbes et al., 2001]. The high mutation frequency is an indication of the primary nature of this mutation in AML [Delhommeau et al., 2009]. Subsequent studies discovered the function of the related protein in the production of 5hmC [Tahiliani et al., 2009], 5fC and 5caC [Ito et al., 2010].

A crystal structure for a portion of Tet2 that comprises its cysteine rich region and the DSBH domain, in complex with DNA, has been solved. The DSBH domain has a central core, that is comprised of a double-stranded β-helix, packed between flanking-segments from both the DSBH and Cys-rich domains. The Cys-rich domain separates into N (Cys-N) and C-terminal (Cys-C) subdomains. The DNA is located above the DSBH core and two loops of the Cys-C subdomain, which form a groove for DNA interaction, with a methylated cytosine (mC6) flipped out and inserted into the catalytic cavity. Three zinc cations are coordinated by Tet2 residues and help the catalysis in both the interaction

with DNA and in the assembly of the structure. The Fe(II) is coordinated by conserved residues H1382, D1384 (HxD motif) and H1881 (Hxs motif) of the DSBH core, whereas α-ketoglutarate is stabilized by residues R1261, H1416, R1896 and S1898 (R5a motif) of TET2. All these residues involved in Fe(II) coordination, α-ketoglutarate interaction and Zn coordination are highly conserved in the TET subfamily members and frequently mutated in cancer [Hu et al., 2013] (Fig. 1.10).

TET2 stability has been documented to be 10 hours in mouse embryonic stem cells [Wang and Zhang, 2014] where its expression has been shown to be regulated by Oct4 [Wu et al., 2013]. Tet2 knockdown in embryonic stem cells affect 5hmC deposition mainly over gene bodies implying a regulatory role for this protein in transcription [Huang et al., 2014]. Different mouse models deficient for Tet2 displayed an increased number of hematopoietic progenitor cells in the bone marrow, and their skewed differentiation toward the myelomonocytic lineage [Ko et al., 2011][Moran-Crusio et al., 2011][Quivoron et al., 2011].

Tet2 has been found to interact with a number of proteins in the cell. The first interaction partner found is OGT [Chen et al., 2012b][Vella et al., 2013][Deplus et al., 2013], an O-linked N-acetylglucosamine transferase which is able to add N-acetylglucosamine to histone 2B on S112 (H2BS112) and to mediate the ubiquitination by BER1 enzymes at H2BK120 [Fujiki et al., 2011]. Tet2 also promotes the N-GlcNacylation of host cell factor (HCF1) through this interaction, a modification linked to the recruitment of the SET/COMPASS complex, which includes the MLL methyltransferase responsible for the production of H3K4me3 that is also a marker of active promoters [Lee et al., 2007] [Deplus et al., 2013]. Following OGT, a number of other interacting partners have been found. The first one is IDAX, a CXXC domain-harboring protein regulated by Wnt signaling, which is responsible for the binding of TET2 to unmethylated CG-containing DNA and has been found able to induce TET2 caspase-dependent degradation upon stem cell differentiation [Ko et al., 2013]. EBF1 [Guilhamon et al., 2013] is a transcription factor involved in B-cell differentiation [Treiber et al., 2010], with which TET2 can contribute

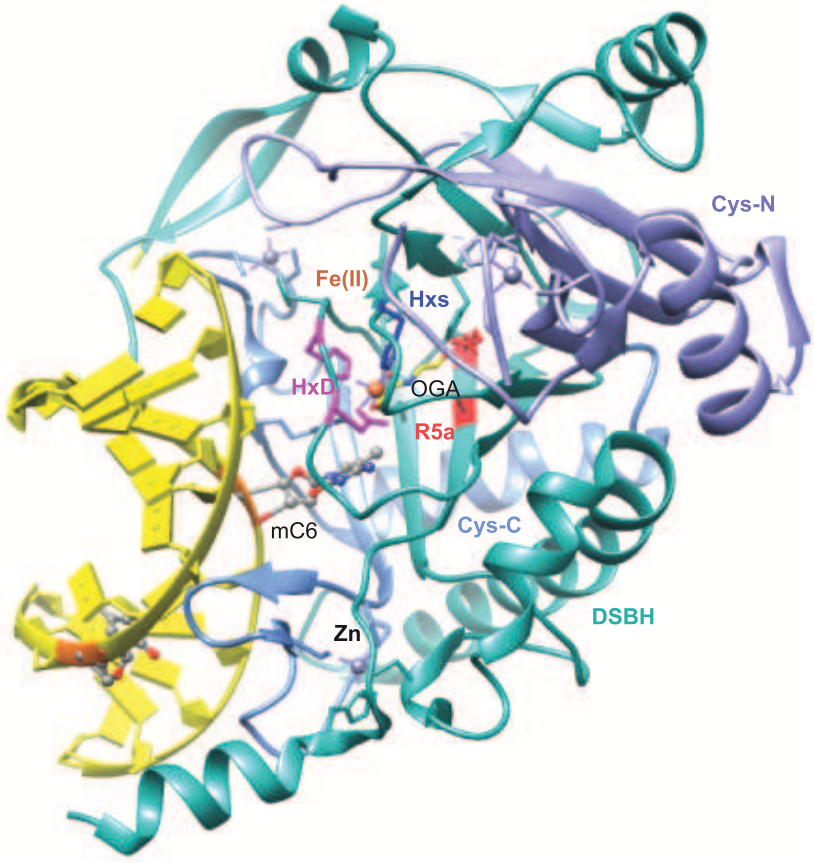


Figure 1.10: Crystal structure of the catalytic domain of TET2 (PDB:4NM6) [Hu et al., 2013]. Highlighted the domains: Cys-N (1129-1232), Cys-C (1233-1312) and DSBH (1313-1481, 1843-1936). The Fe(II) is coordinated by conserved residues H1382, D1384 (HxD motif) and H1881 (Hxs motif) of the DSBH core, whereas α-ketoglutarate (OGA) is stabilized by residues R1896 and S1898 (R5a motif) of TET2.

to DNA demethylation at certain gene promoters [Li et al., 2010]. With NANOG, TET2 mediates the establishment of pluripotency in embryonic stem cells [Costa et al., 2013]. TET2 promotes differentiation with PU.1, which is a transcription factor involved in osteoclast [de la Rica et al., 2013], myeloid and lymphoid differentiation and alternative splicing [Guillouf, 2006]. With PRDM14 TET2 participates in the establishment of DNA demethylation at pluripotency related genes [Okashita et al., 2014]. Lastly, TET2 also interacts with proteins of the BER pathway (TDG, MLH, NEIL1,2,3) with which it pro-

motes DNA demethylation. Among them, it has been reported that Neil3 knockout mice are viable and NEIL3 expression is elevated in hematopoietic tissues, suggesting a function in hematopoiesis [Müller et al., 2014]. TET2 interaction partners, among which the strongest evidences are for OGT, suggest a functional link between 5hmC deposition and transcription.

1.2.2 The biological functions of cytosine variants

Since the discovery of 5mC great progress has been made in determining its function in DNA. After early studies in bacteria, which identified 5mC as a component of the host restriction system against invading phages, attention was directed toward the identification of the context in which 5mC is deposited. Early studies reported a significant deviation from the random distribution of 5mC dinucleotides obtained from DNA preparation from mammalian DNA. In particular, the neighboring base of 5mC was found to be mainly guanine. The hypothesis followed that this must have been related to an enzymatic activity able to recognize this context [Sinsheimer, 1955][Doskočil and Šorm, 1962]. Later studies confirmed that 5mC is deposited mainly in the CpG dinucleotide context, with minor prevalence in the CpH context (especially CpA) [Doskočil and Šorm, 1962][Salomon and Kaye, 1970]. This finding was subsequently further validated, first through nearest neighbor analysis [Woodcock et al., 1987][Ramsahoye et al., 2000] and later by whole genome bisulfite sequencing [Guo et al., 2013][Lister et al., 2013]. Furthermore, it was soon shown that other than in the CG context, 5mC was deposited in clusters of DNA (200 bp) that varied for DNA methylation percentage across the different tissues studied (satellite DNA) [Sano and Sager, 1982]. Moreover it was found that some segments of DNA rich in CpG dinucleotides, spanning 1-2 kilobases [D N Cooper, 1983] and present at gene promoters [Antequera et al., 1989] were largely unmethylated and they were given the name of CpG islands [Bird, 1987].

With the advent of restriction endonucleases able to differentially cut unmethylated or

methylated DNA, more studies confirmed the context of DNA methylation and the presence of symmetric DNA methylation at palyndromic sequences across double stranded DNA (at CG sites) in *Xenopus laevis*, which indicated that DNA methylation must be replicated upon cell division since hemimethylated DNA (e.g. DNA that is methylated only on one strand) could not be detected [Bird and Southern, 1978][Bird, 1978]. These observations were soon confirmed in mammalian cells, reinforcing the concept of heritability of DNA methylation [Wigler et al., 1981], together with the purification of mammalian DNA methyltransferases that were shown to be able to methylate hemimethylated sequences [Bestor and Ingram, 1983].

Such heritability of CpG methylation suggests a role for 5mC in long-term regulation required for diverse biological processes, such as the establishment of genomic imprinting, stable silencing of gene expression and maintenance of genome stability.

Imprinting is defined as the conditioning of the maternal and paternal genomes during gametogenesis, such that a specific parental allele is more abundantly (or exclusively) expressed in the offspring. The role of DNA methylation in imprinting soon became clear after studies investigating early embryogenesis and gametogenesis found that newly methylated regions appeared in the DNA of sperm suggesting the existence of a mechanism that allowed the distinction between maternal and paternal alleles [Groudine and Conkin, 1985]. It was later found that these modifications could be stably propagated and could contribute to different allelic expression patterns. Only with the use of an inserted transgene that became selectively silenced, and with methyltransferases-deficient embryos that showed a loss of imprinting, it was demonstrated that this phenomenon is dependent on DNA methylation [Mohandas et al., 1981][Reik et al., 1987][Li et al., 1993]. An example of imprinted gene is Igrf2, which is expressed from the maternal allele. It has been found unmethylated in the maternal genome promoter and methylated on the imprinted region (intron 2), which results in absent expression of an antisense RNA that would block transcription of the sense if present. Paternally inherited alleles showed expression of antisense

RNA and lack of methylation at the imprinted region, associated with methylation of the Igf2r promoter and therefore absence of Igf2r mRNA [Wutz et al., 1997].

Having discovered that only one X chromosome is necessary for normal development, Lyon proposed the hypothesis that one of the copies, which had already been detected as silent in somatic cells, must be switched off during development by genetic mechanisms [Lyon, 1961]. Studies with the demethylating drug 5-azacytidine pointed to a role for DNA methylation in silencing gene expression, consequent to the reactivation of genes on the inactive X chromosome that was achieved by the induction of demethylation [Mohandas et al., 1981].

The role of DNA methylation in gene silencing was later confirmed in hypomorphs for DNMT1, where imprinted gene re-expression could be obtained [Li et al., 1993] together with the disruption of monoallelic expression of imprinted genes. It was later discovered that DNA methylation represses transcription in a manner that depends on the location and density of the methyl-CpGs relative to the promoter. A high percentage of DNA methylation was shown to be generally stronger than a low percentage, and to be present to silence genomic locations stably [Boyes and Bird, 1992].

Another well-documented consequence of DNA methylation deficiency is the activation of transposable element-derived promoters. Transposable element-related sequences are heavily methylated and transcriptionally silent in somatic cells. In mouse embryos lacking Dnmt1, the normally repressed transcription of intracisternal A particle (IAP) elements, which constitute a homogeneous and transpositionally active family of elements, is highly induced [Walsh et al., 1998]. The biological significance of transposable-element re-expression is uncertain; and it has been proposed that it may facilitate genetic instability by increasing the number of insertional mutagenesis, DNA breaks and chromosome translocations [Yoder et al., 1997b].

1.2.2.1 Interplay between DNA modifications and transcription

Approximately 70% of mammalian promoters are associated with CpG islands [Saxonov et al., 2006], which remain largely unmethylated in somatic cells. When these promoters are active they are usually characterized by nucleosome-depleted regions at the transcription start site, often flanked by nucleosomes containing the histone variant H2A.Z and marked with trimethylation of histone H3 at lysine 4 (H3K4me3). These have been shown to be able to exclude DNMTs and to correlate with transcription [Ooi et al., 2007][Zilberman et al., 2008]. Due to low levels of 5mC at CpG islands, 5hmC/5fC and 5caC are generally depleted around transcription start site [Szulwach et al., 2011][Wu et al., 2011a][Shen et al., 2013]. Conversely, TET has been shown to be enriched at these promoters, thus might contribute to transcription through a catalytic independent function [Williams et al., 2011][Deplus et al., 2013][Vella et al., 2013][Huang et al., 2014].

5fC and 5caC have been mapped at inactive promoters, in the absence of TDG [Shen et al., 2013][Song et al., 2013a]. Indeed, TET proteins might contribute both to activation and repression mechanisms. They have been shown to interact with ODG and to recruit it to CpG-rich promoters [Deplus et al., 2013] [Vella et al., 2013]. TET depletion reduces OGT binding to chromatin and its activity to N-GlcNacylate histone H2B at S112, which has been shown to promote transcription through recruitment of the SET/COMPASS-complex, which possesses H3K4 methyltransferase activity. A repressive role might be promoted via its interaction with the SIN3A co-repressor complex [Williams et al., 2011]. Readers for 5hmC/5fC belonging to transcriptional complexes (NuRD) have been detected in pull-down assays [Spruijt et al., 2013], further highlighting possible roles of these modifications in transcription.

Another putative role for cytosine modification in transcription might come from gene bodies. They are mostly CpG-poor, extensively methylated and contain multiple repetitive and transposable elements. It has long been known that gene body methylation is a feature of transcribed genes [Wolf et al., 1984]. Extensive positive correlations between

active transcription and gene body methylation have recently been confirmed on the active X chromosome [Hellman and Chess, 2007]. Thus, even though a gene body can become extensively methylated, this does not block transcription elongation. One of the functions of gene body methylation might be regulation of splicing, as enrichment of 5mC has been detected on exon/intron boundaries [Laurent et al., 2010]. 5hmC and 5fC have also been detected over gene bodies in mouse ESCs and the brain, as well as in the human frontal cortex and ESCs [Pastor et al., 2011][Szulwach et al., 2011][Wu et al., 2011a][Mellén et al., 2012][Shen et al., 2013][Lister et al., 2013]. Moreover, high 5hmC levels over gene bodies have been shown to correlate with high levels of transcription [Mellén et al., 2012], possibly causing enhanced rates of transcription. Conversely, when 5fC or 5caC are present over gene bodies, they are able to reduce the processivity of RNA polymerase [Kellinger et al., 2012].

Enhancers are situated at variable distances from promoters and are key in the control of gene expression. They are mostly CpG-poor regions and, in the mouse genome, are not 100% methylated or unmethylated and therefore are termed 'low-methylated regions' (LMRs) [Stadler et al., 2011]; these regions also show DNaseI hypersensitivity, H3K4me1 and H3K27ac. At these sites 5hmC could be detected [Pastor et al., 2011][Szulwach et al., 2011][Stroud et al., 2011][Lister et al., 2013] and in the case of poised enhancers, defined as H3K4me1 positive and H3K27ac negative, it was shown enriched together with 5fC, indicating a possible role for DNA demethylation in their activation [Shen et al., 2013][Song et al., 2013a].

1.2.3 Possible routes to DNA demethylation

DNA methylation accounts for about 1% of the total bases in our genome [Ehrlich et al., 1982]. The methylation level characteristic for the mammalian genome is established around gastrulation. Before gastrulation, there is a wave of genome-wide demethylation that removes the 5mC present in the zygote, by replication dependent mechanisms in the

maternal genome and, before replication, in the paternal genome, so that the DNA of the blastocyst is reduced. Between implantation and gastrulation, a wave of global *de novo* methylation re-establishes the overall methylation pattern, which is then maintained throughout life in the somatic cells of the organism (reviewed in [Wu and Zhang, 2014]). While mechanisms for the enzymatic establishment of DNA methylation, both maintenance and *de novo*, have been largely elucidated, DNA demethylation processes were initially only attributed to replication dependent mechanisms. When 5hmC, 5fC and 5caC were discovered [Kriaucionis and Heintz, 2009][Tahiliani et al., 2009][Ito et al., 2011], it was realized that replication independent mechanism might also occur. Replication dependent mechanisms of 5mC depletion are mainly explained by a failure of DNMT1, which normally restores the symmetrical CpG methylation pattern upon DNA replication by methylating the unmodified cytosine in the nascent DNA strand (red arrows in Fig. 1.11) [Bird and Southern, 1978][Bird, 1978][Bestor and Ingram, 1983]. One example of replication-dependent loss of 5mC is the global erasure of 5mC in the maternal genome during mouse preimplantation development. A mechanism that could possibly explain the failure of DNMT1 could be one through which it fails to recognize hemi-hydroxymethylated DNA [Hashimoto et al., 2012a].

Different hypotheses still exist to explain the active mechanisms of 5mC removal. Overexpression of TET proteins has been shown to induce a global decrease in 5mC levels [Tahiliani et al., 2009]. Therefore, active demethylation mechanisms have been attributed to an initial oxidative step of 5mC to 5hmC, 5fC and 5caC followed by their enzymatic removal (blue line Fig. 1.11). 5hmC has been shown to be possibly dehydroxymethylated by the *de novo* DNMTs (3A and 3B) [Chen et al., 2012a]. Other ways of reversion to C include the putative decarboxylation of 5cadC [Schiesser et al., 2012] or the DNA-glycosylase mediated excision of 5fC, 5caC or 5hmU (deamination product of 5hmC by AID/APOBEC) followed by base excision repair mechanisms [Guo et al., 2011][He et al., 2011]. Thymine DNA glycosylase (TDG), normally responsible for the removal of the

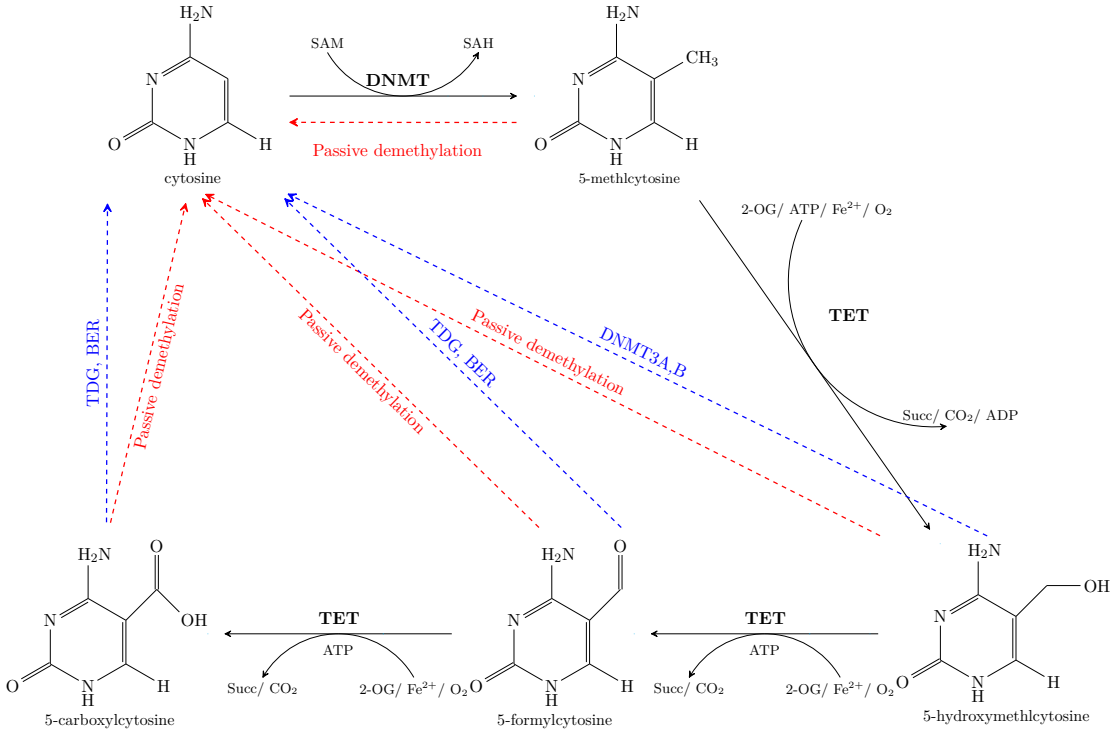


Figure 1.11: Schematic representation of major DNA methylation and demethylation pathways in mammals. DNA methyltransferases (DNMTs) methylate cytosine to yield 5-methylcytosine (5mC; 1% of all bases) by transferring the methyl group from S-adenosylmethionine (SAM) to cytosine. Ten-eleven translocation (TET) enzymes oxidize 5mC to 5-hydroxymethylcytosine (5hmC; 0.1% of all bases), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC). Multiple pathways can lead to DNA demethylation, including thymine DNA glycosylase (TDG)-mediated base excision repair (BER) of 5fC:G and 5caC:G base pairs, and replication-dependent passive demethylation or activation-induced deaminase (AID)/APOBEC that can mediate the deamination of 5hmC to 5hmU followed by TDG-mediated BER [Wu and Zhang, 2014].

pyrimidine base from T:G or U:G mismatches, has been shown to have great activity on 5fC:G and 5caC:G base pairs [Maiti and Drohat, 2011]. TDG-deficient mice are not viable and display embryonic lethality, while derived MEFs show DNA hypermethylation [Cortellino et al., 2011][Cortázar et al., 2012]. Recent reports indicate that TDG might be dispensable for the enzymatic removal of 5fC and 5caC as no difference in their amounts is detected in TDG knockout zygotes relative to wild type zygotes [Guo et al., 2014]. Other data point to the possibility that other DNA glycosylases might compensate for TDG loss, as NEIL glycosylases, which are responsible for the repair of oxidative cytosine

lesions [Hazra et al., 2002], have been shown to be able to interact with TETs [Müller et al., 2014].

1.2.4 Nucleotide salvage of cytidine derivatives

Following the discovery of 5mC in the DNA [Hotchkiss, 1948], another branch of research investigated whether this modified cytosine could be produced by nucleotide-metabolism enzymes. It was discovered that, in *E. coli*, there was no enzyme able to produce 5mdCTP [Bessman et al., 1958], therefore, the focus of research was moved towards studies that looked at the salvage routes for 5mdC. The first reports showed that 5mdC *per se* could not be incorporated into the DNA if administered to cells, but instead could be detected in the form of thymine [Jekunen et al., 1983]. Soon, it was proposed that the barrier for the incorporation of 5mdC lays in the absence of an enzymatic ability capable of diphosphate production [Vilpo and Vilpo, 1991]. Therefore, it was concluded that salvage of 5mdC proceeded through its conversion into thymidine by deamination catalyzed by cytidine deaminase (CDA) [Vilpo and Vilpo, 1991]. Further support for this idea, came from *in vitro* evidence that cytotoxicity induced by 5mdC administration (through conversion to thymidine that has been shown to be toxic when administered to cells [Akman et al., 1981]) could be reversed by co-treatment with the CDA inhibitor tetrahydrouridine (THU) [Jekunen and Vilpo, 1984]. Later it was discovered with purified enzymes, that DCK was able to produce 5mdCMP [Eriksson et al., 1991] and CMPK1 was unable to catalyze the subsequent step of 5mdCDP production.

1.3 Cytosine DNA modifications and producing enzymes in cancer

DNA methylation can be considered one of the hallmarks of cancer, since it has been found to be generally deregulated across all cancer types. The first studies that linked cancer to DNA methylation used Southern blotting to analyze DNA that had been digested with methylation-sensitive restriction enzymes and found that a substantial proportion of CpGs that were methylated in normal tissues were unmethylated in primary cancer tissues [Feinberg and Vogelstein, 1983]. Similar investigations using high-performance liquid chromatography (HPLC) were able to confirm global losses of 5mC across multiple cancer types [Gama-Sosa et al., 1983]. Following these findings, the oncosuppressor Rb was found to be silenced by promoter hypermethylation in cancer [Ohtani-Fujita et al., 1993]. Later, more gene promoters were found to be hypermethylated, such as: the oncosuppressors p16 [Gonzalez-Zulueta et al., 1995] and von Hippel-Lindau (VHL) [Herman et al., 1994], and the base excision repair MutL homolog 1 (MLH1) [Cunningham et al., 1998]. Subsequently, it was discovered that promoter CpG-islands associated silencing was widespread in different cancers, and this general phenomena was given the name of CpG-island methylator phenotype (CIMP) [Toyota et al., 1999]. Other genes that have been found to be deregulated in cancer through hypermethylation are imprinted genes in their expressed unmethylated allele such as H19, which has been found aberrantly silenced in Wilms' tumor [Steenman et al., 1994].

Other than gene silencing, another role that can be attributed to DNA methylation in the promotion of oncogenesis is the increased mutagenic potential of 5mC compared to unmethylated C. Evolutionarily a CpG depletion due to methylation of CpG in the germline, has been suggested. This is thought to be due to the more frequent spontaneous hydrolytic deamination of 5-methylcytosine (5mC) compared to cytosine, which gives rise to thymine rather than uracil, generating a T:G mismatch that is less efficiently repaired compared

to the U:G one [Ehrlich et al., 1986][Lutsenko and Bhagwat, 1999]. As an example of this, it was found that most mutations in p53 occur in the coding region at CG sites, causing hotspots in somatic cells that are thought to be cancer causing [Rideout et al., 1990]. Generally, mutation frequency has been found to be significantly higher at CpG dinucleotides [Alexandrov et al., 2013].

Following large sequencing projects of cancer genomes, DNA methylation enzymes have been found to be mutated. *DNMT3A* is mutated in about 20% of AML cases [Ley et al., 2010]. *DNMT1* has been found mutated in colon cancer [Kanai et al., 2003] and other solid cancers (breast, cervix and lung) [Forbes et al., 2001]. Concordant with these reports, Dnmt3a knockout mice in the hematopoietic compartment results in impaired differentiation over serial transplantation, and expansion of stem cells numbers in the bone marrow [Challen et al., 2012], while Dnmt1-deficient hematopoietic stem cells (HSC) showed a marked increase in apoptosis and defective self-renewal and repopulating capacity. Furthermore, Dnmt1 hypomorph expression in the hematopoietic compartment leads to a normal ability to form myeloid and erythroid progeny, but impairment in the ability to commit to lymphoid differentiation [Bröske et al., 2009]. Recently, it has been shown that the prevailing mutation of DNMT3A R882H, lying in the catalytic domain, disturbs gene expression and promotes the proliferation of HSC transplanted in a mouse model, resulting in expansion of the myeloid compartment and monocytic lineage in a way that resembles chronic myelomonocytic leukemia (CMML) [Xu et al., 2014].

5hmC has been found to be generally depleted in cancer (leukemias, breast, lung, pancreas, liver, colorectal), sometimes in association with Tet2 mutations [Ko et al., 2010], but other times in association with decreased expression of Tet genes [Yang et al., 2012] and sometimes without any of the two [Kudo et al., 2012]. The decrease of 5hmC has been proposed to be a biomarker for melanoma [Lian et al., 2012]. Recently, a causative relationship between tumor progression and 5hmC amount has been established through a report that documented the loss of 5hmC and the insurgence of tumors in the liver of

rats exposed to the genotoxic agents riddelliine and aristolochic acid, which was paralleled by decreased Tet2 expression [Lian et al., 2014]. Tet1 expression is decreased in a cellular model of transformation [Kudo et al., 2012] and *in vivo* during prostate cancer progression, where it promotes invasion and metastasis [Hsu et al., 2012]. In breast cancer Tet1 upregulation gives rise to smaller tumors and suppresses invasion, while its downregulation correlates with poor prognosis [Hsu et al., 2012]. Microarray analysis found that genes responsible for this phenotype are the inhibitors of the metalloproteinases (TIMP) [Hsu et al., 2012], which are negative regulators of MMPs (metalloproteinases) that are known to promote cancer invasion and metastasis. In another case, the high mobility group AT-hook (HMGA)2 was shown to inhibit TET1 and homeobox A (HOXA) expression in breast cancer cells and breast cancer tumors, where high levels of HMGA2 correlated with poor survival [Sun et al., 2013]. MLL-rearranged tumors display induction of TET1, which cooperates with MLL activity to activate target genes such as HoxA9 and Meis1, which have been shown to be sufficient for the development of leukemia [Kroon et al., 1998] [Huang et al., 2013].

TET (especially TET2) levels, other than being regulated at the transcriptional level, have been shown to be diminished by several miRNAs (including miR-125b, miR-29b, miR-29c, miR-101, and miR-7) that disrupted normal hematopoiesis and were overexpressed in AML patients harboring wild type TET2 [Huang and Rao, 2014]. The pro-metastatic microRNA miR-22 suppressed TET expression and thus demethylation of the promoter of the anti-metastatic microRNA miR-200 in a mouse breast cancer model. This correlated with high-grade cancers and the expression of genes involved in metastasis and poor survival in human patients [Song et al., 2013c]. miR-22 is upregulated in MDS and its overexpression leads to hematopoietic cancers in mice in which decreased levels of Tet2 have been shown to be paralleled by a decrease in 5hmC. Tet2 re-expression can rescue the phenotype and reduce the repopulating ability of miR-22 overexpressing HSCs [Song et al., 2013b]. Other than at the translational level, TET proteins can be regulated by metabolites that

can interfere with the role of the cofactor α-ketoglutarate (α-KG). IDH mutant cancers (lymphomas and gliomas), which show the accumulation of 2-hydroxyglutarate (2-HG), a competitive inhibitor for Tet2 that occupies the same site as α-KG, have been shown to display an hypermethylator phenotype [Sasaki et al., 2012][Lu et al., 2012][Turcan et al., 2012] correlating with a hypothetical disruption of TET2 function [Figueroa et al., 2010]. Aberrant DNA methylation is also observed in succinate dehydrogenase (SDH) mutant cancers [Letouzé et al., 2013], in which succinate accumulates and competes with α-KG [Xiao et al., 2012]. *In vivo* evidence for FH mutant cancers that display accumulation of fumarate, behaving as a competitive inhibitor of α-KG for TET, is yet to be obtained.

TET family genes mutations in cancer

The *TET1* catalytic domain has been found translocated with the N terminal part of MLL in t(10;11) [Lorsbach et al., 2003] and *TET2* has been found extensively mutated in AML [Delhommeau et al., 2009], CMLL, myelodysplastic syndromes (MDS) and T-cell lymphoma [Langemeijer et al., 2009][Tefferi, 2010] [Bacher et al., 2010][Rocquain et al., 2010][Busque et al., 2012][Brecqueville et al., 2012][Couronné et al., 2012]. Mouse models carrying HSC deleted for Tet2 show characteristics resembling AML, with expansion of HSCs and their skewed differentiation toward the myelomonocytic lineage [Quivoron et al., 2011][Yang et al., 2011]. So far, more than a thousand TET2 mutations have been identified in cancer, mainly of hematopoietic origin (Fig. 1.12). Some missense mutations impair the enzymatic activity of TET2, with a resultant decrease in 5hmC levels and aberrant DNA methylation [Ko et al., 2010]. Human cord blood-derived stem cells depleted of TET2, or stem cells isolated from leukemia patients bearing TET2 mutations, showed a consistent phenotype in *in vitro* differentiation assays, with an increased myeloid-lineage and decreased erythroid-lineage cells [Pronier et al., 2011]. The missense mutations tend to be clustered in two highly conserved regions of the human TET2 protein

(amino acids 1104–1478 and 1845–2002) that correspond to the Cys rich region and the catalytic DSBH core (Fig. 1.12). Based on the crystal structure, many of the residues affected by the missense mutations are located on the surface of the TET2 catalytic domain. These residues might be important for protein–protein interactions and/or may be subject to post-translational modifications [Hu et al., 2013][Huang and Rao, 2014]. Analysis of the clonal architecture and mutation frequency of myeloid disorders indicates that TET2 mutations often occur as an early oncogenic event and lead to the clonal expansion of leukemic cells. A second hit is then necessary for the development of leukemia. Genes frequently found mutated with Tet2 include the polycomb protein enhancer of zeste homolog 2 (EZH2), which represents the catalytic component of polycomb repressive complex 2 (PRC2) [Muto et al., 2013]; additional sex combs like 1 (ASXL1), which is the regulatory subunit of the ASXL1-BAP1 complex [Rocquain et al., 2010], a deubiquitinase for H2AK119Ub [Brecqueville et al., 2012]; Dnmt3A in T-cell lymphoma [Couronné et al., 2012]; and Janus Kinase 2 (JAK2) implicated in erythropoiesis [Rocquain et al., 2010]. Less frequently, Tet genes have been found mutated in colon [Seshagiri et al., 2012], kidney and thyroid cancer [Forbes et al., 2001].

1.4 FDA-approved DNA methylation inhibitors for cancer therapy

Aberrant DNA methylation has been targeted in cancer therapy by DNA methylation inhibitors. 5-azacytidine (5azaC) and 5-azadeoxycytidine (5azadC) are DNMT inhibitors that are currently FDA approved for the treatment of MDS, AML and CMML (Fig. 1.13). They were synthesized in the Czech Republic at the beginning of the 1960s [Pliml and Šorm, 1964] as anticancer agents that, like other nucleoside analogs, could interfere with nucleotide metabolism processes. These compounds are cytidine analogs that bear nitrogen instead of carbon at the fifth position of the pyrimidine ring. They are salvaged

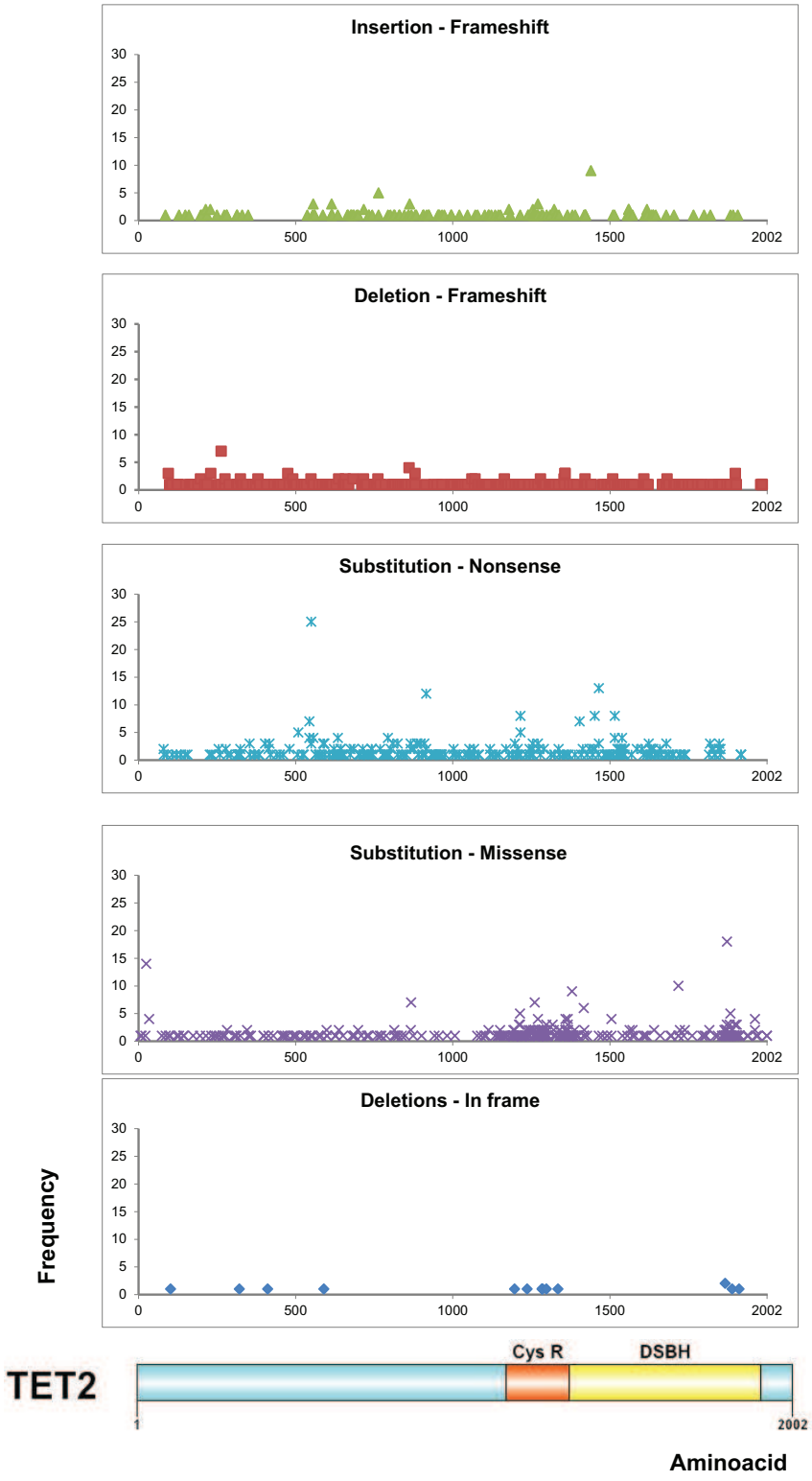
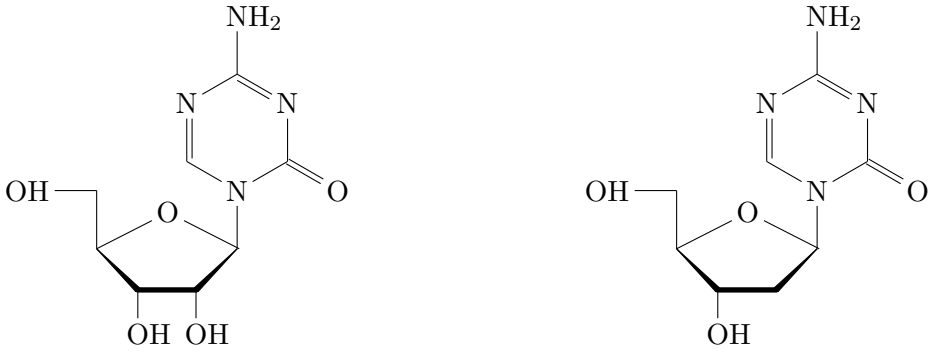


Figure 1.12: TET2 mutations observed in cancer. Most of the missense mutations cluster in the functional Cys-rich and DSBH domains of Tet2 [Forbes et al., 2001]. Frequency is indicated on the y-axis as total number of cases detected. X-axes indicate protein aminoacids.

through the canonical cytidine route to produce the respective nucleosides triphosphates. They can then be incorporated into DNA, or RNA in the case of 5-azacytidine, being indistinguishable from cytosine for the DNA polymerase [Patel et al., 2010]. When maintenance DNMTs initiate the methylation reaction on hemimethylated cytosines following replication, they establish a covalent bond between the carbon-6 atom of the cytosine ring and the enzyme [Santi et al., 1984][Song et al., 2012]. Normally this is resolved through a β-elimination reaction, which cannot happen for the presence of a nitrogen instead of a carbon at position 5 of the ring (i.e. lack of a proton), resulting in the covalent trapping of DNMT1 [Schermelleh et al., 2005] and the formation of protein-DNA complexes that are resolved through proteasomal degradation [Patel et al., 2010]. In addition, the covalent protein bound to the DNA also compromises its functionality by inducing double strand breaks and the DNA damage response [Palii et al., 2008].

Various studies led to the discovery of the mechanism of action of 5azaC and 5azadC. It was first shown that the compounds are able to transform and kill cultured cells by inducing their differentiation [Taylor and Jones, 1979], and later revealed that the effect was dependent on DNA demethylation [Jones and Taylor, 1980][Jones, 1984]. Subsequent studies with Dnmt1 knockout cells showed that the mechanism of action of 5azadC might be due to the trapping, and thus the depletion of Dnmt1 since resistance to the compounds was detected in cells depleted for Dnmts [Jüttermann et al., 1994]. It has been reported that the loss of imprinting observed in cancer can be reversed by 5azadC treatment, which reactivates the expression of imprinted genes from the silenced allele [Barletta et al., 1997]. Numerous clinical trials led to the FDA approval of these drugs for clinical use in patients with myelodysplastic syndrome and leukemia [Fenaux et al., 2009][Garcia-Manero et al., 2013]. Crucial in this process was the recognition that low dosages of these drugs showed better efficacy and increased tolerability [Fenaux et al., 2009][Garcia-Manero et al., 2013].



(a) 5-Azacytidine (b) 5-Azadeoxycytidine

Figure 1.13: DNA methylation inhibitors FDA-approved in the clinic for treatment of AML, MDS and CMLL.

Common resistance mechanisms

The applicability of these drugs has been largely restricted to hematopoietic cancers because of the expression of cytidine deaminase (*CDA*) in the kidney and liver [Camiener and Smith, 1965] which quickly converts them into the inactive DNA methylation inhibitors 5-azauridine (5-azaU) and 5-azadeoxyuridine (5-azadU) [Chabner et al., 1973]. Another drawback limiting their application, is the inherent chemical instability of these molecules in the aqueous environment, resulting in a half-life of less than 24 hours [Lin et al., 1981]. Therefore, strategies to overcome their enzymatic inactivation by CDA have been [Camiener, 1968][Driscoll et al., 1991], or are being attempted [Ferraris et al., 2014] with the development of CDA inhibitors (tetrahydrouridine (THU) or zebularine derived). To provide a solution to their chemical instability, different formulations might prove useful in the clinic, like the conjugation of 5azadC to another nucleoside in the form of oligonucleosides [Stresemann and Lyko, 2008].

Other than inactivation by deamination in the liver and kidney, *CDA* overexpression in cancer cells has been shown to decrease their sensitivity to the cytotoxic effects of these nucleoside analogues [Eliopoulos et al., 1998][Kanno et al., 2007]. CDA overexpressing cells display decreased sensitivity to the toxicity induced by 5azadC, and only co-treatment

with CDA inhibitors, like THU [Eliopoulos et al., 1998] or zebularine, is able to increase cellular death following 5azadC treatment [Lemaire et al., 2008]. In the clinic, a number of reports link CDA overexpression, to poor clinical response upon 5azadC treatment, and poor survival following treatment [Qin et al., 2011][Mercier et al., 2013]. However, CDA downregulation can induce increased toxicity of 5azadC treatment, until documented death [Mercier et al., 2007]. Moreover, another factor to consider is the genetic variants of *CDA*. Two variants are known and have been characterized enzymatically: K27Q and A70T (associated with the polymorphism 79A>C and 208G>A) [Kühn et al., 1993][Laliberté and Momparler, 1994][Yue et al., 2003]. A70T displays a generally low catalytic activity; while K27Q has enhanced catalytic activity towards dC and 5azaC, but not in respect to 5azadC, compared to the K27 variant. Moreover, K27Q displayed enhanced inhibition by THU and zebularine [Micozzi et al., 2014]. In agreement with these findings, CDA A70T has been associated with higher sensitivity to 5azadC together with enhanced toxicity in the clinic [Mercier et al., 2007][Giovannetti et al., 2008]; while CDA K27Q treated patients normally display lower efficacy of 5azaC treatment [Bhatla et al., 2009][Falk et al., 2013]. Interestingly, Down's syndrome AML patients display a high cure rate compared to non-Down's syndrome patients. This has been proposed to be linked to the consequences of low *CDA* expression, due to inactivating mutations in the transcription factor GATA1, which has been shown to bind the *CDA* promoter. Moreover, the downregulation of *CDA* is paralleled by a high expression of *DCK* in AML cases in Down's syndrome patients [Ge et al., 2004][Ge et al., 2005]. In addition, a SNP on the promoter (-451C>T) of CDA that results in loss of the transcription factor ETS1 binding site and the gain of a binding site for the transcriptional repressor LYF1, has been found to promote increased efficacy for cytarabine treatment, another analog of cytidine, linked to decreased expression of *CDA* [Fitzgerald et al., 2006][Mahlknecht et al., 2009].

The success of these therapies against hematologic malignancies has prompted at desire to employ these compounds in the treatment of solid tumors. Historically these drugs have

shown limited efficacy in this context, but new developments might come with the use of combinations of DNA methylation and HDAC inhibitors, both mechanisms that govern gene silencing [Cameron et al., 1999][Gore, 2006]. Moreover, other possible avenues might come from combination therapy with immunomodulators, as it has been shown the 5azadC can positively boost immune function by activating the expression of immunomodulators such as interferon [Li et al., 2015].

Chapter 2

Aims of the study

5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxycytosine (5caC) are products of TET enzyme activity, which initiates at 5-methylcytosine (5mC) in DNA. 5hmC, has been found to be globally depleted in cancer while *TET2* is highly mutated, particularly in blood cancers. In this setting, the aims of this study were:

1. the investigation of the fate of 5hmdC, 5fdC and 5cadC in the context of nucleotide metabolism, with a particular focus on the salvage pathway (Chapter 4);
2. the assessment of the effects on proliferation induced by the administration of nucleosides with bases implicated in DNA demethylation (Chapter 5 and 6);
3. the description of the molecular function of TET2 in haematopoiesis by identification of interacting proteins (Chapter 7).

Chapter 3

Materials and methods

3.1 Nucleic acid procedures

3.1.1 Plasmid cloning

The plasmids that were utilized in this study are listed in Table 3.1, with sources outlined in each case. For the ones for which cloning was necessary, we mainly utilized a cloning strategy from PCR products. Whenever it was necessary to switch vector or protein tag, subcloning strategies were adopted.

3.1.2 Polymerase chain reaction (PCR)

The primers utilized for the amplification of target sequences, with their respective annealing temperatures, are listed in Table 3.2. *TET2A* and *TET2B* were amplified from hematopoietic stem cell cDNA (kindly provided by Dr. Higgs' lab) with the oligonucleotides indicated in Table 3.2. Reaction and polymerase conditions were set up according to manufacturers' instructions (FailSafe™PCR PreMix Selection Kit, FS99060, Cambio) with a Piko™Thermal Cycler (Thermo Scientific).

Name	Vector	Insert	Host	Selection	Source
pCMV-Tet2h	pCDNA3	*TET2*	h,m	Neo	Schofield lab
pGBKT7 DNA-BD	pGBKT7	-	y	Trp	Clontech
pGBKT7-53	pGBKT7	p53	y	Trp	Clontech
PGADT7-T	pGADT7	TAg SV40	y	Leu	Clontech
pGBKT7-Lam	PGBKT7	Lam	y	Trp	Clontech
pGBKT7-Tet2A	pGBKT7	*TET2*(1-732AA)	y	Trp	cloned
pGBKT7-Tet2B	pGBKT7	*TET2*(733-2002AA)	y	Trp	cloned
pGADT7-Lib	pGADT7	Library	y	Leu	Clontech
pET28-CDA	PET28a	*CDA*	b	Kan	cloned
pET28-DCK	PET28a	*DCK*	b	Kan	cloned
pET28-CMPK1	PET28a	*CMPK1*	b	Kan	cloned
IMAGE-CDA	pCMV-SPORT6	BC054036	b	Amp	SGC
IMAGE-DCK	pCMV-SPORT6	BC103764	b	Kan	SGC
IMAGE-CMPK1	pOTB7	BC014961	b	Cam	SGC
IMAGE-WDR61	pOTB7	BC010080.2	b	Cam	SGC
pCMV-FlTet2h	pCDNA3	Fl-TET2	m,h	Neo	cloned
TRCN0000051288	pLKO.1	sh-CDA_8	h	puro	Sigma
TRCN0000051290	pLKO.1	sh-CDA_0	h	puro	Sigma
pLKO.1-Luc	pLKO.1	sh-Luc	-	puro	Lu lab
pCMV-V-SVG	pCMV	V-SVG	-	-	Amati lab
pCMV-dR8.9	pCMV	dR8.9	-	-	Amati lab
pEFIresdsRed	pEF	dsRed	h,m	puro	Goding lab
pLdsRed_CDA	pLenti	dsRed_CDA	h, m	puro	cloned
pGFP	pCMV	GFP	h, m	-	Christianson lab
39481	pLentiCMV	-	h, m	puro	Addgene
pET-28a(+)	pET-28-a	-	b	Kan	Novagen
pcDNA5/FRT	pcDNA5/FRT	-	Flp IN	hygro	Christianson lab
pC3xFlNTet2	pcDNA5/FRT	3xFlTET2(1-732AA)	h	hygro	cloned
pC3xFlF1Tet2	pcDNA5/FRT	TET2(1-188AA)	h	hygro	cloned
pC3xFlF2Tet2	pcDNA5/FRT	TET2(189-406AA)	h	hygro	cloned
pC3xFlF3Tet2	pcDNA5/FRT	TET2(407-552AA)	h	hygro	cloned
pC3xFlF4Tet2	pcDNA5/FRT	TET2(553-732AA)	h	hygro	cloned
pOG44	pOG44	Flippase	-	-	Christianson lab
pL3xFlNTet2	pLenti	TET2(1-732AA)	h	puro	cloned
pLGFP-S-WDR61	pLenti	GFP-S-WDR61	h	puro	cloned
pIC113	pBabe	GFP-S	h, m	Neo	Cheeseman lab
pL3xFlTet2	pLenti	3xFl-TET2	h, m	puro	cloned

Table 3.1: List of plasmids used in the study

For all the other constructs, PCRs were done with proofreading DNA polymerase (Phu-

PP	Sequence	T_{ann}	Target	Template	Plasmid cloned	RE
SK13	GCGTCGACCTGAACAGGATAGAACCAACC	56°C	TET2(1-2196 bp)	cDNA)	pGBKT7Tet2A	SalI
SK14	GCGCGGCCGCTTATGGTTGTGTTTGTGCTG					NotI
SK15	GCGTCGACCTTCCCAGAGTTCACATCTCC	56°C	TET2(2196-6009 bp)	cDNA)	pGBKT7Tet2B	SalI
SK16	GCGCGGCCGCTCATATATATCTGTTGTAAGG					NotI
SK243	GCCATATGATGGCCCAGAAGCGT	48 °C	CDA	IMAGE-CDA)	pET28-CDA	XhoI
SK244	GCCTCGAGTCACTGAGTCTTCTGC					NdeI
SK363	CGGATTCATGGCCACCCCGC	52 °C	DCK	IMAGE-DCK)	pET28-DCK	XhoI
SK315	CTCGAGTCACAAAGTACTCAAAAAC					BamHI
SK560	GCGGATCCATGCTGAGCCGCTGCC	54 °C	CMPK1	IMAGE-CMPK1	pET28-CMPK1	XhoI
SK561	GCCTCGAGTTAGCCTTCCTTGTCAAAAATC					BamHI
SK273	GCGTTAACATGGACTACAAGGATGACG-ATGACAAAATGGAACAGGATAGAAC	50°C	TET2(1-1040 bp)	pCMV-Tet2	pCMV-FlTet2	HpaI
SK274	CGGCTAGCTTTGTGGTTC					NheI
SK637	GCGGATCCGATTACAAGGATGACG-ACGATAAGATGGAACAGGATAGAACCAAC	50°C	TET2(1-564 bp)	pCMV-Tet2	pC1xFlF1Tet2	BamHI
SK638	GCCTCGAGTTATTTCCCCTCCTGCTCATTC					XhoI
SK639	GCGGATCCGATTACAAGGATGACGACG-ATAAGATGAGTGCTAATTACCATGACAAGAA	50°C	TET2(565-1218 bp)	pCMV-Tet2	pC1xFlF2Tet2	BamHI
SK640	GCCTCGAGTTAAAGAAGCAATTGTGATGGTGGT					XhoI
SK641	GCGGATCCGATTACAAGGATGACG-ACGATAAGATGTCTCCCCCTCCTCCTC	50°C	TET2(1219-1656 bp)	pCMV-Tet2	pC1xFlF3Tet2	BamHI
SK642	GCCTCGAGTTAAAGATCTCGTGTTTGCTCC					XhoI
SK643	GCGGATCCGATTACAAGGATGACG-ACGATAAGATGGTGCCCCCAACACAGC	50°C	TET2(1657-2196 bp)	pCMV-Tet2	pC1xFlF4Tet2	BamHI
SK644	GCCTCGAGTTAGGATGGTTGTGTTTGTGCT					XhoI
SK637	GCGGATCCGATTACAAGGATGACG-ACGATAAGATGGAACAGGATAGAACCAAC	50°C	TET2(1-2196bp)	pCMV-Tet2	pC1xFlFNTet2	BamHI
SK644	GCCTCGAGTTAGGATGGTTGTGTTTGTGCT				pL1xFlFNTet2	XhoI
SK118	ACTCACCCATCGCATACCTC	54°C	Tet2h (1035-6009 bp)	pCMV-Tet2	pL3xFlTet2	NheI
SK503	GCGGGCCCTCATATATATCTGTTGTAAGG					ApaI
SK277	GCCTCGAGATGGCCCAGAAGCGT	50°C	CDA	pET28-CDA	pEFCDA_IresdsRed	XhoI
SK278	GCGAATTCTCACTGGGTCTTCTGC					EcoRI
SK728	GCCTCGAGATGACCAACCAGTACGGTATT	54°C	WDR61	pIC116-WDR61	pIC113-WDR61	XhoI
SK729	GCGGATCCTTAAATTGGACAATCATAGATGTG					BamHI

Table 3.2: Primer pairs (PP) used to clone plasmids utilized in the study. Sequences, annealing temperatures (T_{ann}), plasmids cloned, templates and restriction enzymes (RE) used are indicated.

sion High Fidelity DNA polymerase, F-530S) according to the manufacturer's instructions for cycling parameters, with 10 ng of template and 0.5 μM of each primer. The reactions were performed in a Piko™Thermal Cycler (Thermo Scientific). After PCR, the size of one tenth of the products was assessed via agarose gel electrophoresis by adding 1X of DNA loading buffer (6 X: 2.5 g Ficoll 400, 400 μL 0.5 M EDTA, 18 μL 10% SDS, H_2O to 10 mL).

3.1.3 Restriction digest of DNA and dephosphorylation

For cloning, 1 μg of plasmid and the product of a PCR reaction, were digested with 1 μL of one of the restriction enzymes listed in Table 3.3 and the appropriate buffer, as suggested in the manufacturer's instructions. The reaction were carried at 37 °C for 1 h in an incubator (Heratherm, Thermo Scientific).

The digested plasmid was then dephosphorylated with 1 μL alkaline phosphatase (CIP,

Restriction enzymes	Manufacturer	Temperature
ApaI	NEB, R0114S	37 °C
BamHI	NEB, R0136S	37 °C
EcoRI	NEB, R3101S	37 °C
HpaI	NEB, R0105S	37 °C
KpnI	NEB, R0142S	37 °C
NdeI	NEB, R0111S	37 °C
NheI	NEB, R0131S	37 °C
SalI	NEB, R0138S	37 °C
SpeI	NEB, R0133S	37 °C
XhoI	NEB, R0146S	37 °C

Table 3.3: Restriction endonucleases utilized in the study

M0290S, NEB) for 1 h at 37 °C. The products were then checked on an agarose gel and purified via a PCR purification kit (Qiagen) according to manufacturer's instructions, prior to ligation.

3.1.4 Ligation and transformation

The digested vector and inserts were then ligated overnight with T4 Ligase (Thermo Scientific) in a 16 °C water bath (Thermo scientific), according to the manufacturer's instructions. Reactions were assembled with 10 ng of vector, and insert amount determined using the following formula:

$$\text{insert amount in ng} = (\frac{\text{insert}}{\text{vector}})_{MR} \times \frac{\text{insert length (bp)}}{\text{vector length (bp)}} \times \text{vector mass (ng)} \qquad (3.1)$$

with a insert vector molar ratio (MR) of 5. A vector only sample, was always included as a negative control. Subsequently, the ligation was transformed into chemocompetent cells (XL-10 Gold) by keeping them on ice for 20 minutes, before transferring them to a 42 °C water bath (Thermo Scientific) for 1 minute, following incubation again on ice for 2 minutes. After transformation, the bacteria were resuspended in 1 mL prewarmed SOC media (0.5% (w/v) yeast extract, 2% (w/v) tryptone, 10 mM NaCl, 2.5 mM KCl, 20 mM $MgSO_4$ and 20 mM glucose) and left to grow for 1 h at 37 °C in a shaking incubator (Stuart). After this, the bacteria were plated on LB agar plates with the appropriate antibiotic and left to grow overnight at 37 °C in an incubator (Heratherm, Thermo Scientific). The next morning, colonies were checked for the presence of the insert by colony PCR.

3.1.5 Colony PCR

A PCR reaction was assembled with 0.5 μM primers and reaction mix as indicated in the DreamTaq (Thermo Scientific) manufacturer's instructions. Each colony was picked with a pipette tip that was first streaked onto an LB agarose plate with the appropriate antibiotic, then swirled and soaked in the reaction mix previously assembled. The PCR reactions were then run in a Piko thermo cycler (Thermo scientific). 10-20 colonies were screened each time. PCR products were then loaded on an agarose gel to assess the presence of the product of the desired size.

3.1.6 Sequence validation of the insert and plasmid amplification

Positive colonies, from colony PCR, were grown overnight in 5 mL LB and the plasmid isolated via Miniprep kit (Qiagen) according to the manufacturer's instructions. All constructs were sent for sequence verification following the sample submission guidelines of the sequencing service (Source bioscience). Sequences were aligned to reference ones downloaded from UCSC Genome Browser, using ApE software. Subsequently, the cor-

rect plasmids were amplified and purified using a MaxiPrep kit (Qiagen) according to the manufacturer's instructions. The purified plasmid was then quantified with Nanodrop (Thermo Scientific) and stored in T buffer (10 mM Tris-HCl pH 8) at -20 °C at a working dilution of 1 μg/μL.

3.1.7 Annealing of oligonucleotides for cloning

In order to add a double Flag tag to single Flag vectors (p1xFl(N, F1, F2, F3, F4, empty) Tet2) obtained in table 3.2, we used annealed oligos in Table 3.4 with respective restriction site compatible ends: KpnI-BamHI for pcDNA5-based vectors and SpeI-BamHI for pLenti-based vectors. Annealing was performed by resuspending equimolar amounts of each oligo in annealing buffer (10 mM Tris, pH 7.5; 50 mM NaCl; 1 mM EDTA) and putting them in a heat block at 95°C, which was then slowly left to cool down slowly at room temperature. The annealed oligos were then ligated to the pLenti and pCDNA5 vectors produced, which were screened and sent for sequencing, to verify the absence of mutations.

Oligo pair	Sequence	Target	RE
SK676 SK677	CATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACG GATCCGTCATGATCTTTATAATCACCGTCATGGTCTTTGTAGTCCATGGTAC	2xFlag	KpnI-BamHI
SK678 SK679	CTAGTATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACG GATCCGTCATGATCTTTATAATCACCGTCATGGTCTTTGTAGTCCATA	2xFlag	SpeI-BamHI

Table 3.4: Oligos used to insert 2xFlag into pcDNA5- and pLenti-based vectors. Restriction endonucleases (RE) compatible ends are indicated.

3.1.8 Subcloning strategies for the insertion of a tag or for vector switch

To obtain pL3xFlTet2h, we took advantage of pL3xFlTet2N. A portion of Tet2 was amplified from pCMVTet2h with the primers listed in Table 3.2, before double digesting this and pL3xFlTet2N with the unique restriction enzymes NheI and ApaI. Subsequently, the digested products were ligated, and positive colonies screened, with one clone sent for verification via sequencing (Source Bioscience).

To clone pLdsRed_CDA, CDA was first PCR amplified and digested using the primers and strategy indicated in Table 3.2. Subsequently it was ligated in pEFIresdsRed upstream of Ires dsRed. The tagged version of CDA was double digested with XhoI and XbaI. The product was then ligated into pLenti to obtain pLdsRed_CDA.
WDR61 was first cloned in pIC113 (Cheeseman lab), which contains a GFP-S tag, as indicated in Table 3.2. The vector was sent for sequencing (Source Bioscience) to verify the absence of mutations. Subsequently, the vector was switched by digesting pIC113-WDR61 with NdeI-BamHI, then ligating the product into pLenti to obtain pLGFP-S-WDR61.

3.1.9 Genomic DNA extraction

After transfection or nucleoporation, DNA was extracted with a Gene Jet Genomic DNA extraction Kit (Thermo Fisher), following the manufacturer's instructions. RNase A/T1 treatment was done on a column following the kit manufacturer's instructions.
After 5hmdC, dC and 5fdC treatments, DNA was extracted from cells with TRI®Reagent (T9424 Sigma Aldrich), according to the manufacturer's instructions. 5 μg of DNA were subsequently treated with 20 μL RNase A/T1 (Thermo Fisher) in digestion buffer (10 mM Tris pH 7.5, 300mM NaCl) for 1 hour at 37 °C and phenol-chloroform extracted via the following procedure. The sample volume was brought up to 300 μL, an equal volume of TE (10mM Tris pH 8, 1mM EDTA) saturated Phenol:Chloroform:Isoamyl Alcohol 25:24:1 (77617, Sigma Aldrich) was added and the sample centrifuged at 15,871 rcf (Eppendorf, 5424, 24 tubes) for 5 minutes. The aqueous phase was subsequently transferred to a new tube, with care taken not to transfer any of the protein at the interphase, and the DNA precipitated by adding 1/10 volume of 3M sodium acetate pH 5.2 and 2.5 volumes of 100% ethanol. The mixture was incubated at -80 °C for an hour. The sample was then centrifuged at 4 °C at 20,000 rcf and the supernatant aspirated. The pellet was washed once with 75% ethanol and then air dried.
DNA from organs or tumor samples was extracted with TRI Reagent (Sigma Aldrich),

according to the manufacturer's instructions. DNA was subsequently treated twice with RNase A/T1 (Thermo Fisher) and phenol-chloroform extracted as indicated previously. The purified DNA was always resuspended in T buffer (10mM Tris pH 8) and stored at -20°C, with concentrations ranging from: 100-1000 ng/μL.

3.1.10 DNA hydrolysis for HPLC analysis

DNA was quantified spectrophotometrically in triplicate using Nanodrop (Thermo Scientific), which allowed additional assessment of the homogeneity of the solution and the quality of the DNA. 3-5 μg were digested for HPLC (high performance liquid chromatography) analysis. DNA was hydrolyzed as described [Kriaucionis and Heintz, 2009]. Briefly, it was incubated overnight in the hydrolysis solution (100 mM NaCl, 20 mM $MgCl_2$, 20 mM Tris pH 7.9, 1000 U/mL Benzonase, 600 mU/mL Phosphodiesterase I, 80 U/mL Alkaline phosphatase, 36 μg/mL EHNA hydrochloride, 2.7 mM deferoxamine). Protein components were removed by centrifugation through an Amicon Microcoon centrifugal filter unit (3 kDa cut-off, Millipore). Hydrolysates were lyophilized (SpeedVac), re-suspended in 30 μL of 100 mM ammonium acetate pH 6 and automatically injected into HPLC. HPLC was done as described [Kriaucionis and Heintz, 2009], using an Agilent UHPLC 1290 instrument fitted with an Eclipse Plus C18 RRHD 1.8 μm, 2.1x150 mm column and detected with an Agilent 1290 DAD fitted with a Max-Light 60 mm cell. Buffer A was 100 mM ammonium acetate pH 6.5, and buffer B 40% acetonitrile. The run conditions were: 100mM ammonium acetate and a gradient between 1.8-100% of 40% acetonitrile (1-2 min, 100% A; 2-16 min 98.2% A - 1.8% B; 16-18 min 70% A - 30% B; 18-20 min 50% A - 50% B; 20-21.5 min 25% A - 75% B; 21.5-24.5 min 100% B) with a flow rate of 0.4 mL/min. The signal was detected at 260 nm, 280 nm and 254 nm and the data analyzed at 260 nm.

3.1.10.1 Determination of nucleoside extinction coefficients ϵ for our HPLC system

Nucleosides retention times, that is the amount of time spent by the nucleoside inside the column from the time of injection, were previously determined, by separately running each one of them. A representative chromatogram with a nucleotide mixture is shown in Figure 3.1.

The linear range of the instrument detector was assessed by injecting serial dilutions of

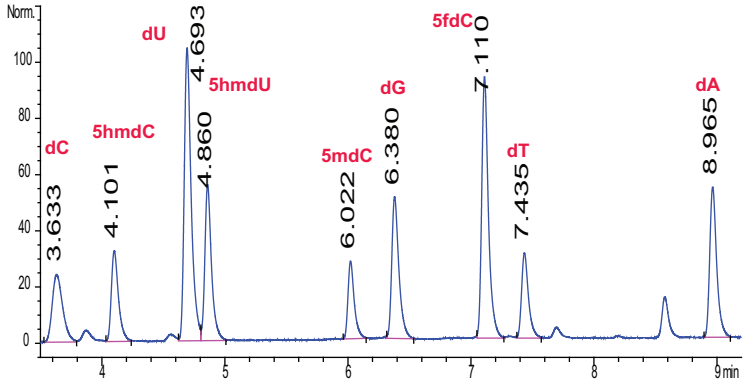


Figure 3.1: Representative HPLC chromatogram, from a C_{18} column, of all the nucleosides used in the study: names are indicated close to the respective peaks. Norm: absorbance unit. Min: minutes.

each nucleoside. It was determined to be up to 2000 AU (absorbance units) in peak height. Moreover, to be able to calculate the concentration (c) of nucleosides in DNA from the optic measurements given by the peak integration of an HPLC chromatogram, we needed to refer to the Lambert-Beer law, which relates the absorbance (A) of a certain molecule to the path length (b) that it crosses and to its relative extinction coefficient (ϵ) according to the following equation:

$$A = \epsilon \times b \times c \tag{3.2}$$

where absorbance is equal to the integrated peak measure (peak area) [Moldoveanu and David, 2002]. We made use of the equation in 3.2 and determined ϵ as:

$$\epsilon = \frac{\text{Peak area}}{b \times c} \tag{3.3}$$

where the path length of the instrument is 0.6 cm (b). Therefore, by injecting a known amount of each nucleoside, we determined their ϵ. The ones for the nucleosides for the HPLC instrument (Agilent 1290 Infinity) used in this study are outlined in Table 3.5.

For quantification, the peaks corresponding to 5-methyldeoxycytidine, 5-hydroxymethyl-

Nucleoside	$\epsilon[1/\mu\text{M/cm}]$
dC	8.87
5hmdC	7.49
5dU	43.8
5hmdU	16.9
5fdC	18.49
5mdC	6.61
dG	13.3
T	9.36
dA	18.2

Table 3.5: Extinction coefficients for the nucleosides used in the study, determined by HPLC

deoxycytidine, deoxyguanidine, deoxythymidine and 5-hydroxymethyldeoxyuridine were automatically or manually defined, when the amount of them was closer to detection limit, and software integrated. The levels of 5mdC were calculated with the following equation:

$$\%_{5mdC} = \frac{A_{5mdC} * \epsilon_{dG}}{A_{dG} * \epsilon_{5mdC}} * 100 \tag{3.4}$$

where A is the area of the peaks and the percentage of 5mdC is calculated relative to the percentage of the pairing base dG.

The amount of 5hmdU was calculated instead with the following equation:

$$\%_{5hmdU} = \frac{A_{5hmdU} * \epsilon_T}{A_{dT} * \epsilon_{5hmdU}} * 100 \tag{3.5}$$

relative to the base that substitutes when incorporated [Miyashita et al., 2002].
The amount of 5hmdC was calculated using the following formula:

$$\%_{5hmdC} = \frac{A_{5hmdC} * \epsilon_{dG}}{A_{dG} * \epsilon_{5hmdC}} * 100 \tag{3.6}$$

3.1.11 RNA extraction and cDNA analysis

Total RNA was purified onto RNeasy columns (Qiagen) and treated on-column with DNase (Qiagen) according to the manufacturer's instructions. RNA concentration was determined at Nanodrop (Thermo Scientific) and RNA stored in RNAse free Water at -80°C. Complementary DNA (cDNA) was produced using the reverse-transcriptase MuLV (Thermo Scientific, EP0352). The first step of the reaction was assembled with 1-2 μg RNA, 1 μL of 0.5 μg/μL oligo dT (Thermo Scientific, SO131), 0.25 μL of 300 ng/μL random hexamers (Thermo Scientific, SO141) and RNA and RNase free water (Qiagen) up to 11.5 μL.
Samples were incubated at 65°C for 15 minutes following which, with the tubes kept on ice, 9.5 μL of the second mixture was added, which was composed of: 4 μL MuLV buffer 5X (Thermo Scientific); 2 μL of 10 mM each dNTP Mix (Thermo Scientific); 0.5 μL RiboLock RNase Inhibitor (Thermo Scientific, EO0381), 2μL MuLV retro-transcriptase (Thermo Scientific). The samples were incubated at 42°C for 50 minutes and subsequently at 70°C for 15 minutes. Finally, the cDNA was diluted in T Buffer (TrisHCl pH 8 10mM) and kept at -20°C.
A total of 20 ng of cDNA was used for real-time PCR reactions with FAST SYBR Green Master Mix (Qiagen) in an Applied Biosystems 7500 Real-Time PCR System, as in the

SYBR manufacturer's instructions. The primers utilized for QPCR in gene expression studies are detailed in Table 3.6. Primers were designed with Primer 3 [Untergasser et al., 2012] to have a melting temperature of 60°C and a 50-60% GC content, to obtain an amplicon of 75-150 bp, spanning an exon-intron boundary. Data analysis was performed according to the $\Delta\Delta C_T$ method which relies on the relative quantification of the sample tested by direct comparison of its value both to the control of the experiment and to an internal standard, which in our case was GAPDH. In order to apply this method, it was checked that primers were amplifying with linear efficiency, by using them with a 6 points serial dilution of cDNA. By plotting the amount of cDNA (Log(cDNA)) on the x axis and the C_t (threshold cycle) of amplification, that is the intersection between an amplification curve and a threshold line, on the y axis, a linear relation ($R^2 = 0.99$) could be established, from which a slope was calculated. The slope (s) of this line relates to the efficiency (e) of amplification as:

$$e = 10^{-1/s} - 1 \tag{3.7}$$

If e was between 90 and 105 %, i.e. there would be an approximate two-fold increase in the number of copies with each cycle, therefore primers were selected for further steps. Only with this condition valid, could the relative quantification be performed with the $\Delta\Delta C_T$ method, which assumes an efficiency of amplification of 100% of primers for both the housekeeping gene and test gene. The first step determines the ΔC_T with the following equations:

$$\Delta C_T(\text{control, housekeeping}) = C_T(\text{control}) - C_T(\text{housekeeping}) \tag{3.8}$$

$$\Delta C_T(\text{sample, housekeeping}) = C_T(\text{sample}) - C_T(\text{housekeeping}) \tag{3.9}$$

for the control and the test sample (e.g. treated and untreated) relative to a housekeeping gene. Subsequently the $\Delta\Delta C_T$ is calculated with:

$$\Delta\Delta C_T(\text{sample, control}) = \Delta C_T(\text{sample}) - \Delta C_T(\text{control}) \tag{3.10}$$

Finally, the change in gene expression is calculated with:

$$\text{Fold change} = 2^{\Delta\Delta C_T} \tag{3.11}$$

that expresses the fold change of gene expression of the target gene in the test sample relative to the same gene in a control sample, both normalized to the expression of a housekeeping gene.

Primer Pair	Sequence	Target
SK70	GGAAAACACCGAGTCGGAATAC	TGF3β
SK71	GCGGAAAACCTTGGAGGTAAT	
SK60	AGACATTATGACACCGCCAAAT	PTEN
SK61	GGTGGGTTATGGTCTTCAAAAGG	
SK66	TTATCCAGCGGCCAGCTAATG	MLH1
SK67	GCCTCCCTCTTTAACAATCACTT	
SK64	AAGGTGAAGGTCGGAGTCAAC	GAPDH
SK65	GGGGTCATTGATGGCAACAATA	
SK68	CGCCAGGAGGAGATTCTGA	DERMO1
SK69	GTAGCAGAAGGTCTGGCAATG	
SK260	CTGGGCATCTGTGCTGAAC	CDA
SK261	GCCCCACATGGAGAGATAAA	
SK262	GCAAGTCATGAGAGAGAGATGATT	CDA
SK263	CATCCGGCTTGGTCATGTA	

Table 3.6: Primers used for QPCR.

3.2 Yeast two-hybrid assay

For the yeast two hybrid (Y2H) assay, we used the Matchmaker Gold Yeast Two-Hybrid System (Clontech, 630489), following the manufacturer's instructions. Our bait was generated by cloning into pGBKT7 the cDNA correspondent of TET2, which was divided into

two parts: TET2 (1-732AA) and TET2(733-2002AA) (see Table 3.1). The exact breakage position (see Appendix 8.2, Fig. 6, 7) was determined considering sequence conservation among species and secondary structure predictions (BLAST, ClustalW and Dompred), in order to avoid breaking the protein on conserved, and thus potentially functional, residues involved in secondary structure formation. Our prey came from a pre-transformed library derived from human bone marrow (Clontech, 630477), having the following characteristics:

- number of independent clones: $8.8 * 10^6$;
- insert size range: 0.5 to ≥ 3 kb;
- titer: $\geq 5 * 10^7$ cfu/mL.

The genotypes of the yeast strains used are indicated in Table 3.7.

Y2HGold was transformed with pGBKT7TET2A and TET2B as in the kit instructions

Strain	**Reporters**	**Transformation markers**	**Genotype**
Y187	MEL1, LacZ	trp1, leu2	MATα, ura3-52, his3-200, ade2-101, trp1-901, leu2-3, 112,gal4Δ, gal80Δ, met–, URA3 :: $GAL1_{UAS} - GAL1_{TATA} - LacZ$, MEL1
Y2HGOLD	AbA^r, HIS3, ADE2, MEL1	trp1, leu2	MATa, trp1-901, leu2-3, 112,ura3-52, his3-200, gal4Δ, gal80Δ,LYS2 :: $GAL1_{UAS} - GAL1_{TATA} - His3$, $GAL2_{UAS} - GAL2_{TATA} - Ade2$, URA3 : : $MEL1_{UAS} - Mel1_{TATA} - LacZ$AUR1-C MEL1

Table 3.7: Genotype of the yeast strains used

and their expression checked by PCR on their complementary cDNA. The reverse transcription was performed as previously described, and PCR on the cDNA performed with the primers indicated in Table 3.8. Control and mating experiments were performed according to the kit manual. Positive clones, able to grow on QDO/X/AbA media, were iso-

PP	Sequence	Target	T_{ann}
SK25	GCCATATGATGGAACAGGATAGAACC	TET2A(1-900 bp)	52°C
SK26	GCGAATTCTTAATCATCAGCATCACAGG		
SK72	TCAATTCTCATTGAGTGTGC	TET2B(5600-6006 bp)	52°C
SK16	GCGCGGCCGCTCATATATATCTGTTGTAAGG		

Table 3.8: Primers to check expression of constructs TET2A and TET2B in transformed yeast.

lated and DNA extracted with QIAprep Spin Miniprep kit (Qiagen, 27104) as in the manufacturer's instructions, with the following modifications: the colony was grown overnight in 5 mL of selective yeast media (SD-Leu). Subsequently, cells were harvested by centrifugation at 5,000 rcf and re-suspended in 250 μL of P1 buffer containing 0.1 mg/mL RNase A. 50 μL of acid-washed glass beads (Sigma Aldrich G-8772) were added and the sample vortexed for 5 min. The supernatant was transferred to a fresh 1.5 mL microcentrifuge tube and the isolation carried on as in the kit manual. Once isolated, the plasmids were sent for sequencing. A subsequent screen was carried out to check for autoactivation of the positive clones, by mating them with empty bait vector containing the DNA binding domain and screening them on DDO/X and QDO/X/AbA media.

3.3 Cell culture materials and methods

3.3.1 Cell lines utilized in the study

The cell lines utilized in the study are listed in Table 3.9. The following ones belong to the NCI60 panel, a collection of cell lines with different cancer origins [Shoemaker et al., 1988]: HOP-92, MDA-MB-231, DU-145, HCC-116, PC-3, SN12C, HCC-2998, HS578T, MDA-MB-435, BT-549, SW-620, NCI-H522, MCF-7, OVCAR-5. This is a useful public resource which associates cell lines with profiles of gene [Pfister et al., 2009][Scherf et al., 2000] and protein expression [Gholami et al., 2013]. The other cancer cell lines belong to the Cancer Cell Line Encyclopedia (CCLE) [Barretina et al., 2012], a more recent and comprehensive set of cancer cell lines.

Cell line	Source	Splitting ratio	Cancer/tissue origin
Capan-2	S. Lunardi, Grey Institute	1:2, 1:5	Pancreas
BT-549	Dr. Bond lab, Ludwig Institute	1:2, 1:6	Breast
MDA-MB-231	Dr. Bond lab , Ludwig Institute	1:2, 1:5	Breast
MCF7	Dr. Bond lab, Ludwig Institute	1:3, 1:6	Breast
HS578T	Dr. Bond lab, Ludwig Institute	1:2, 1:5	Breast
HCC-116	Dr. Bond lab, Ludwig Institute	1:3, 1:8	Colon
HCC-2998	Dr. Bond lab, Ludwig Institute	1:2, 1:4	Colon
SW-620	Dr. Bond lab, Ludwig Institute	1:3, 1:8	Colon
Colo-320	Dr. Lu lab, Ludwig Institute	1:3, 1:5	Colon
DU-145	Dr. Bond lab, Ludwig Institute	1:4, 1:6	Prostate
PC-3	Dr. Bond lab, Ludwig Institute	1:3, 1:6	Prostate
SN12C	Dr. Bond lab, Ludwig Institute	1:2, 1:5	Kidney
MDA-MB-435	Dr. Bond lab, Ludwig Institute	1:3, 1:6	Melanoma
A375	Dr. Lu lab, Ludwig Institute	1:3, 1:8	Melanoma
HOP-92	Dr. Bond lab, Ludwig Institute	1:2, 1:4	Lung
NCI-H522	Dr. Bond lab, Ludwig Institute	1:3, 1:6	Lung
NCI-H1299	Dr. Lu lab, Ludwig Institute	1:3, 1:10	Lung
OVCAR-5	Dr. Bond lab, Ludwig Institute	1:3, 1:6	Ovarian
HeLa	Dr. Lu lab, Ludwig Institute	1:5, 1:10	Ovarian
LN18	Dr. Lu lab, Ludwig Institute	1:4, 1:6	Glioma
BL-70	Dr. Lu lab, Ludwig Institute	1:3, 1:5	Lymphoma
THP-1	Dr. Lu lab, Ludwig Institute	1:3, 1:5	Leukemia
293T	Dr. Klose lab, Biochemistry Department	1:5, 1:10	Embryonic kidney
FLP-IN™-T-Rex™-293	Dr. Christianson lab, Ludwig institute	1:5, 1:10	Embryonic kidney

Table 3.9: Cell lines utilized in the study.

3.3.2 Growth conditions of cell lines

Capan-2, MDA-MB-231, DU-145, HCC-116, HCC-2998, HS-578T, MDA-MB-435, BT-549, SW-620, NCI-H522, MCF-7, OVCAR-5, LN18, A375, HeLa and 293T cells were cultured in DMEM medium (Lonza). PC-3, SN12C, Colo-320, BL-70 and THP-1 were grown in RPMI medium (Lonza). Both media were supplemented with 10% fetal bovine serum (Biosera), 2 mM L-Gln (Lonza), 1% penicillin/streptomycin (Lonza) (Complete media). All cell lines were grown at 37°C in 5% CO_2 atmosphere. For adherent cells, passaging was performed by removing the media, washing the plate once with PBS and adding a sufficient amount of Trypsin/EDTA (Lonza, Life Technologies) to cover the plate surface. The plate was then incubated at 37 °C until the cells were detached. Trypsin was subsequently inactivated by the addition of culture media and cells centrifuged at 2,000 rcf for 3 minutes. The supernatant was then aspirated and cells resuspended in

fresh media, prior to seeding according to the splitting ratios in Table 3.9. To thaw cells from liquid nitrogen stock, vials were placed in a 37 °C water bath (Thermo Scientific) for 2 minutes. Addition of pre-warmed growth medium and centrifugation for 5 minutes at 2,000 rcf enabled the removal of the freezing medium. The pellet of cells was then re-suspended in pre-warmed growth medium and seeded in a 10 cm dish. The cells were passaged at least once, prior to any manipulation.

For freezing, cells were allowed to grow in a 10 cm dish to about 80% confluency before collection by trypsinization (as described above). Cells were centrifuged for 5 min at 2000 rcf and the supernatant discarded. The cell pellet was re-suspended in 1 mL of freezing media (Complete media, 10% DMSO) and aliquoted in cryovials (Corning). The vials were labeled and cooled, at a rate of 1 °C per minute, in a Mr Frosty (Nalgene) freezing container placed in a -80 °C freezer for at least 24 hours, before being transferred to a liquid nitrogen tank for long term storage.

3.3.3 Cell culture treatment and proliferation analysis

An appropriate number of cells were seeded in a p60 cell plate or in a T25 flask and treated with the indicated concentration of 5hmdC, 5fdC or dC at day 0, day 1, day 3, day 5, day 7 and day 9. Cells were passaged and pelleted at day 3, day 5, day 7 and day 10 with splitting ratios as outlined in Table 3.9. At the same days (3, 5, 7, 10) 1 Volume of Trypan blue staining solution (Lonza) was added to an aliquot of live cells in a 50% ratio, to exclude dead cells, and the cells counted with a TC-20 Cell Counter (Biorad). The cells treated at day 3 were used for further assays such as HPLC, immunofluorescence and cell cycle analysis (Figure 3.2).

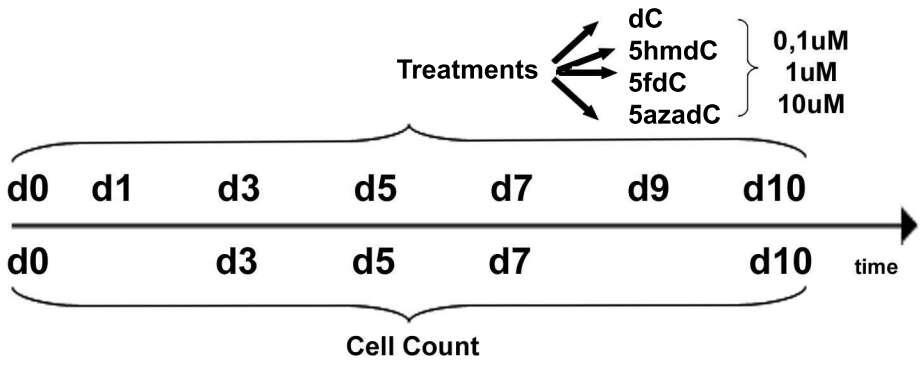


Figure 3.2: Cell culture treatment scheme (d indicates day).

3.3.4 EC_{50} determination

The half maximal effective concentration (EC_{50}) of a compound is defined as the amount of it that inhibits cellular growth by half, between the baseline and maximum, at a defined time point. To determine the EC_{50} for 5hmdC and 5fdC in SN12C, MDA-MB-231 and H1299 cells, a CellTox Green (CTG) (Promega) kit was used following the manufacturer's instructions. To apply the method, it was first necessary to establish the number of cells which, when plated, would be able to grow for the treatment period (3 days) without being confluent and would be sufficient for a signal to be detected. Therefore, a 2-fold serial dilution of cells was made and plated on a 96-well plate (Corning). Growth was monitored until day 3, after which, 6 ranges of serial dilution were selected with cell numbers giving 80% confluency. Subsequently, cytotoxicity assay were performed, as in the kit manual, at day zero and day three, to check whether a signal was detected and if it was linear at the beginning and end points. All cytotoxicity measurements were performed with black tissue culture grade 96-well plates (Greiner). The experiment was performed, as in the kit manual, for 10 serial dilutions of 5hmdC, 5fdC and dC (negative control). Fluorescent signals were detected with a GloMax microplate reader (Promega). Measurements were taken at day 3 after treatment and at day 0 to quantify the starting number of cells. CTG values obtained after 3 days of treatment with 5hmdC and 5fdC were normalized to the

values obtained for dC treatment, were expressed as a percentage of the day 0 value and plotted against compounds concentrations. Data were fitted with a non-linear sigmoidal relation and the concentration of drugs necessary to inhibit 50% of growth (EC_{50}) was determined with Prism (Graph Pad).

3.3.5 Transfection of adherent cells

10^6 cells (MDAMB231) were nucleofected with Amaxa nucleofector kit (Lonza), using GFP plasmid as a positive control (2 μg) and 2 mM dNTPs (5hmdCTP) in a 100 μL volume, following the manufacturer's instructions. Soon after transfection, the cells were seeded in a 6-well plate. After 24 h, cells were washed twice with PBS to remove excess triphosphates and collected for DNA extraction for HPLC analysis at 48 h.
10^6 cells (H1299) were transfected with Lipofectamine 2000 (Life Technologies) according to the manufacturer's instructions, using GFP plasmid as a positive control (2 μg) and 50mM dNTPs (5hmdCTP) in a 100 μL volume. After 24 h, cells were washed twice with PBS to remove excess triphosphates and, after 48 h, were collected and DNA extracted for HPLC analysis.
293T were transfected via the calcium phosphate method. Sixteen hours prior to transfection, $2 * 10^6$ cells were seeded in a 10 cm dish. A first mixture containing 1-20 μg of DNA, 428 μL water, 62 μL calcium chloride (2 M $CaCl_2$) was assembled. Slowly (drop-wise) 500 μL of 2x HBS (50 mM HEPES, pH 7.05; 10 mM KCl; 12 mM dextrose; 280 mM NaCl; 1.5 mM Na_2PO_4. pH adjusted to 7.05 with 1 N HCl. Filtered using .22 micron filter. Aliquoted in 50 mL Falcon tubes. Stored at -20 °C) were added, while vortexing. The resulting mixture was added slowly to the cells. Medium was changed after 16 h incubation.

3.3.6 Generation of stable cell lines: FLP-N system

Stable cell lines were generated by making use of FLP-IN^{TM}-T-Rex^{TM}-293 (R780-07, Life Technologies) cells, which carry a FRT recombination site and a Tet repressor to silence the integrated transgene in the absence of doxycycline. The transgenes inserted were from the pcDNA5/FRT vectors generated (p3xFl(N,F1,F2,F3,F4,empty)Tet2) (see Table 3.1). The plasmids were cotransfected with pOG44, which expresses the Flippase gene, in the host cell lines, according to the manufacturers' instructions. After hygromycin selection (100 μg/mL) and induction of transgene expression with doxycycline (1 μg/mL), the expression of the target was checked via western blot for the FLAG tag. This project was carried out with the help of a summer student, Rapolas Zilionis.

3.3.7 Lentiviral infection

Lentiviral infection was utilized to generate stable cell lines for:

1. CDA knock down;
2. CDA overexpression;
3. 3xFlag TET2 overexpression;
4. GFP-S-WDR61 overexpression.

A general procedure was utilized for all the cell lines generated. In brief, 293T cells were used as a packaging cell line for lentiviral particle production with lentiviral plasmid and second generation packaging plasmids (pCMV-VSVG, pCMV-dR8.9: a generous gift from the Dr. Amati lab, IIT, Milan) [MacKenzie and Shioda, 2011]. At 24 h post-transfection with calcium phosphate precipitation, the medium was changed. The lentiviral particles were collected at 48 h post-transfection for infection. 24 h post-infection (p.i.) the medium was changed, and at 36 h p.i. cells were put under appropriate selection, until cells in the

non-infected plate were all dead.

MDA-MB-231 and SN12C were infected for CDA knockdown with five different hairpins in pLKO.1 vectors (SHCLND-NM_001785, Sigma-Aldrich) plus a control vector against Luciferase (a generous gift of the Dr. Lu lab, Oxford Ludwig Institute for Cancer Research) and selected with 1.5 μg/mL puromycin (Sigma) for 48h. The 2 that gave the best efficiencies (TRCN0000051290, TRCN0000051288), according to QPCR data, were selected for the experiment. The knock down was further assessed by immunoblot.

For overexpression of *CDA*, a lentiviral strategy was used with pLdsRed_CDA. H1299 and MCF7 were infected with the same procedure described for knockdown infections. A puromycin selection of 2 μg/ml was applied until all the cells in the control plate were dead. The overexpression was assessed by immunoblot.

Furthermore, a stable 293T cell line overexpressing 3xFlag-Tet2 was generated through lentiviral infection with pL3xFlTet2, which was confirmed by immunoblot.

GFP-S-WDR61 was stably overexpressed in FLP-N cells with pLGFP-S-WDR61 by adopting the same procedure described above.

3.3.8 FACS

FACS was utilized to study the cell cycle and to count fluorescent populations of cells.

3.3.8.1 FACS cell cycle analysis

At 3 days of treatment, $5 * 10^5$ cells were trypsinized (Gibco, Life Technologies) and re-suspended in 1 mL of PBS. They were then pelleted and re-suspended in 70% ethanol, and fixed for 1 h on ice. After centrifugation, the pellet was re-suspended in 250 μL staining solution (50 μg/mL propidium iodide (Sigma), 0.1 mg/mL RNaseA (Qiagen) and 0.05% Triton-X100) and incubated at 37° C for 40 min. Controls were used for G1 (serum

starvation overnight) and G2 (0.1 μg/μL nocodazole overnight). Data for 10,000 cells were collected with a FACS Canto flow cytometer (BD Biosciences) equipped with 630 nm laser and then analyzed using FlowJo software (TreeStar) with gates set with the use of the controls.

3.3.8.2 Competitive cell growth assay

An equal number of H1299 and H1299_dsRedCDA cells were seeded in wells of a 24-well plate and treated with 5hmdC and 5fdC for 10 days at days 0, 1, 3, 5, 7, 9. Cells were passaged at days 3, 5 and 7. Upon passaging, a fraction of cells was subjected to FACS measurements to count the red and non fluorescent populations. Therefore, data were collected at days 0, 3, 5, 7 and 10, with a FACS Canto flow cytometer (BD Biosciences) equipped with 630 nm laser for 10,000 cells and analyzed with FlowJo software (TreeStar).

3.4 *In vitro* assay materials and methods

3.4.1 Nucleosides and nucleotides utilized in the study

The following nucleotides and nucleosides were utilized in the study: 5hmdC (PY-7588, Berry & Associates), 5fdC (PY-7589, Berry & Associates), 5cadC (PY-7593, Berry & Associates), 5azadC (A3656, Sigma Aldrich), ATP solution (Thermo Fisher), [γ-32P]-ATP (Perkin Elmer), dC (D3897, Sigma Aldrich), dCMP (D7625, Sigma Aldrich), 5hmdCTP (BIO-39046, Bioline).

3.4.2 Nucleoside stability assay

100 μM solutions of 5hmdC, 5fdC and 5azadC were prepared in HPLC grade water (Thermo Fisher) or in DMEM medium (Lonza). The solutions were incubated at 37°C

for 10 days and a sample taken every 24 h and subjected to HPLC-UV analysis.

3.4.3 Enzyme purification

Human DCK was cloned in pET28a(+) and expressed in *E. coli* BL21-(DE3)RIPL (Life Technologies) for 4 h at 37° C following induction with 1 mM IPTG. The resulting bacterial pellet was re-suspended in 50 mM sodium phosphate pH 8, 300 mM NaCl and protease inhibitors (Complete Mini, Roche) and lyzed by sonication with 5 cycles of 1 minute on, 1 minute off at 60 amplitude settings (Soncis Vibra Cell™) [Johansson and Karlsson, 1995][Hazra et al., 2010]. The supernatant was cleared by centrifugation at 10,000 rcf for 20 minutes at 4°C and loaded onto a HisTRAP HP 5 mL Column (GE Healthcare) on an AKTA purifier. After washing for 20 column volumes (CV), the protein was eluted with a linear gradient of imidazole (0-500 mM: 30 CV) in 50 mM sodium phosphate pH 6, 300 mM NaCl and 10% glycerol. The fractions were assessed by SDS gel electrophoresis and subsequent staining with Gel Code Blue reagent (Pierce) according to the manufacturer's instructions. The fractions containing the protein were pooled, concentrated with Amicon 3 kDa centrifugal filter units (UFC900324, Millipore) and further purified on a HiPrep 16/60 Sephacryl S-200 gel filtration column (GE Healthcare) with 2 CV of equilibration and 1.5 CV of elution in 50 mM sodium phosphate pH 6, 300 mM NaCl and 10% glycerol. The fractions containing proteins were again concentrated using Amicon columns, supplemented with final 10 mM DTT and 40% glycerol, snap frozen and stored in aliquots at -80°C. The purity was assessed by SDS-PAGE electrophoresis followed by Gel Code Blue staining (Figure 3.3 A).

CMPK1 was cloned into pET28a(+) and expressed in *E. coli* BL21-(DE3)RIPL for 4 h at 37°C through 0.5 mM IPTG induction [Johansson and Karlsson, 1997]. Bacteria were lyzed by sonication, as previously described, in 50 mM Tris pH 7.5, 10 mM NaCl and protease inhibitors (Complete Mini, Roche) [Liou et al., 2002]. The supernatant was cleared by centrifugation at 10,000 rcf for 20 minutes at 4°C, and loaded onto a HisTRAP HP

5 mL column. The protein was eluted with an increasing concentration of imidazole, as indicated above. A subsequent step was gel filtration with HiPrep 16/60 Sephacryl S-200 Column as described before. Lastly followed an anion exchange column step with a HiTrap Q HP 5 mL (GE Healthcare), from which the protein was washed with 10 CV and eluted with a 20 CV linear gradient of 0-1 M NaCl. The salt was removed by dialysis in 50 mM Tris pH 8, the protein concentrated and 10 mM DTT added to the final preparation, prior to storage in 40% glycerol at -80°C. The purity was assessed by SDS-PAGE followed by gel Code Blue staining (Figure 3.3 B).

CDA was cloned into pET28a(+) and expressed in *E coli* BL21-(DE3)RIPL at 37°C

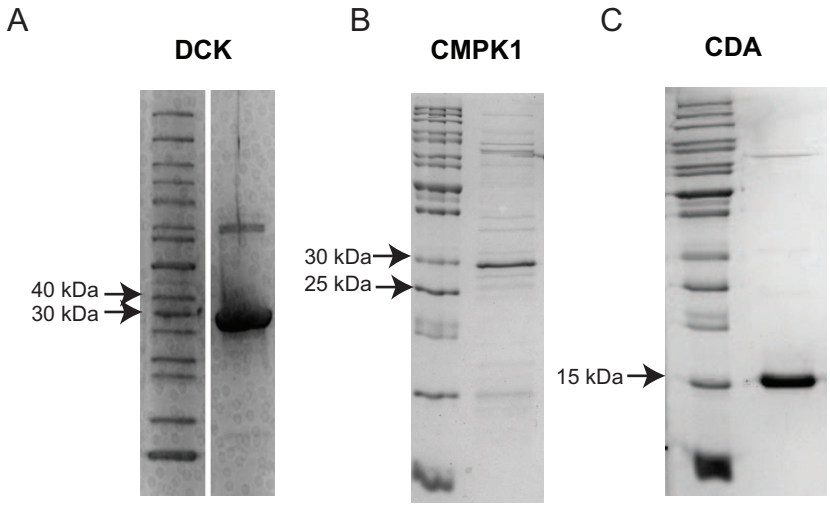


Figure 3.3: Purified enzymes: A. DCK; B. CMPK1; C. CDA

through 0.6 mM IPTG induction at late exponential growth ($A_{435} = 1$) for 19 h [Vincenzetti, 1996]. Cells were harvested by centrifugation at 5,000 rcf and washed with 0.9% NaCl. They were then re-suspended in buffer A (Tris–HCl 50 mM, pH 7.5; 1 mM DTT; 1 mM EDTA) supplemented with protease inhibitors (Complete Mini, Roche) and lysed by sonication, as described before [Vincenzetti, 1996]. Supernatant was cleared by centrifugation at 10,000 rcf for 20 min and loaded onto a HisTRAP HP 5 mL Column (GE Healthcare). After washing, the protein was eluted with a linear gradient of imidazole (0-500 mM) in 50 mM Tris-HCl pH 7.5, 300 mM NaCl and 10% glycerol. Imidazole was

removed by dialyzing the protein overnight in 20 mM Tris-HCl pH 8, 150 mM NaCl. His Tag was cut by the addition of 10 U of thrombin (Sigma, T513) per mg of protein in cleavage buffer (20 mM Tris-HCl pH 8, 150 mM NaCl, 2.5 mM $CaCl_2$ and 0.5 mM TCEP) and overnight incubation at 4°C. This was followed by HisTRAP HP 5 mL Column purification, to remove the cleaved tag. A second step of purification was performed by gel filtration with a HiPrep 16/60 Sephacryl S-200 Column, as previously described. 10 mM DTT was added to the final preparation and the protein solution stored in 40% glycerol at -80°C. The purity was assessed by SDS-PAGE followed by gel Code Blue staining (Figure 3.3 C) and identity by mass spectrometry with a collaboration done by G. Berridge (TDI, Oxford).

3.4.4 Enzyme assay

The purified enzymes were assessed for activity via different methods. Kinase (DCK, CMPK1) activities were assayed by thin layer chromatography (TLC). Deamination by CDA was assayed spectrophotometrically.

3.4.4.1 Kinase activity assays and TLC separation

The activity of 1μg DCK was determined at 37°C for 2 h via a radiochemical method in a buffer containing 100 mM Tris pH 7.5, 100 mM KCl, 10 mM $MgCl_2$, 1 mM [γ-32P]-ATP and a nucleoside concentration of 200 μM in 50 μL. 1 μL of products was separated on 2D TLC on a glass backed AVICEL cellulose plate (Analtech) by following the protocol in [Ramsahoye et al., 2000]. Briefly, 1 μL of product was spotted onto 20 x 20 cm plates, 1.5 cm from each corner, and allowed to let dry. A fresh solution of 33 volumes (1 volume is equal to 1 mL) of isobutyric acid, 9 volumes of water and 1.5 volumes of 30% ammonium hydroxide solution was prepared and added to the TLC tank (Analtech). Plates were placed, by keeping the same inclination angle, in a TLC tank and the lid closed. The

first dimension was allowed to run until the mobile phase reaches the top of the plate, which was then dried overnight in a fume hood. The next morning, a solution to run the second dimension was prepared with: 40 volumes of saturated ammonium sulphate, 9 volumes of 1 M sodium acetate and 1 volume of isopropanol. The plates were oriented with the first dimension at the bottom and the second dimension run, fully, for around 4 h. Subsequently, they were dried for 2 h in a fume hood, wrapped in saran wrap and incubated with a phosphor screen overnight. The TLC sheets were auto radiographed using a Phosphoimager (Biorad) and images analyzed via ImageLab software (Biorad).
CMPK1 was assayed through a coupled assay with DCK following conditions outlined in [Van Rompay et al., 1999b] with 1 μg DCK, 1 μg CMPK1 and 1 mM substrate in 50 μL for 2 h at 37°C. In this case, the products were analyzed via 1D TLC. 1 μL of reaction was spotted onto 20 x 20 cm glass backed TLC sheets (PEI Cellulose F, Millipore) 2 cm from the bottom. 13 samples were loaded onto each plate, allowed to dry and then put in the tank to develop fully, for approximately 1 h, in 50 mL 0.5 M ammonium formate pH 3.5. The plates were then dried and exposed to storage phosphor screen (Biorad) which was scanned via a Phosphoimager (Biorad), and the images analyzed via ImageLab software (Biorad).

3.4.4.2 CDA kinetic activity: assessment of linear range and extinction coefficients

CDA kinetic activity was assayed as described [Vincenzetti, 1996][Vincenzetti et al., 2004]. In order to start with our kinetic analysis we needed to determine:

- the linearity of the relationship between substrate concentration and measured absorbance;
- the extinction coefficients of product and substrate and their difference ($\Delta\epsilon$), to be

able to derive concentrations from the absorbance value of a determined reaction, according to the Lambert-Beer law;

- the linear range of product conversion.

Linear range of the substrate concentration versus absorbance To assess the linear range of the concentration of substrates versus absorbance, we measured their different absorbances at 260 nm at various concentrations. The range and the values are plotted in Figure 3.4. A good linear range was achieved between 5 μM and 200 μM, therefore we could use these substrates concentrations to run the assay.

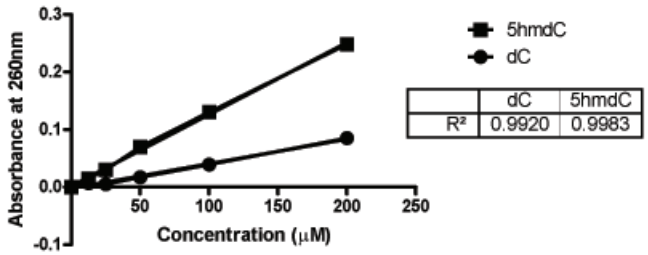


Figure 3.4: Linear range of two of the substrate used for CDA conversion. The square fit values of the linear range are indicated.

Extinction coefficient difference between product and substrate To be able to assess the concentration from absorbance readings in a reaction in which a product is formed, we needed to determine the difference between the extinction coefficient of the product and substrate ($\Delta\epsilon$), according to the Lambert-Beer law (mentioned before for HPLC). We assessed their readings at 260 nm. The path lenght (b) was the cuvette length, which is 1 cm. The values determined are indicated in Table 3.10

Linear range of the product formation We needed to assess the amount of enzyme that we could use to get a linear signal under initial velocity conditions. Initial velocity

Reaction	**Δ**ϵ[1/(M*cm)]
dC → dU	1274
5mdC → dT	4165.5
5hmdC → 5hmdU	5760.5
5fdC → 5fdU	-714
5azadC → 5azadU	-2071

Table 3.10: Extinction coefficient ($\Delta\epsilon$) differences between products and substrates of CDA

is satisfied when the substrate concentration is still unlimited and when there is no reaction inhibition by product, therefore the rate of product formation is proportional to the activity of the enzyme. To achieve this, we made different dilutions of the enzyme with a fixed concentration of the substrate and fixed assay conditions and plotted the absorbance versus time. Initial velocity conditions could be achieved in a time frame of 2 minutes with 45 ng of CDA for all the nucleosides, and 500 ng for 5hmdC. The results of the initial velocities are outlined in Figure 3.5.

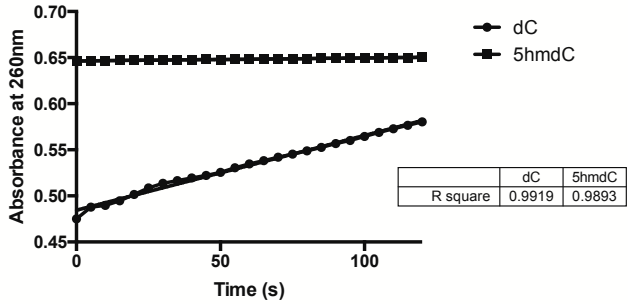


Figure 3.5: Linear range of the enzyme used for CDA conversion. Plotted are the linear relationships established between absorbance values and time. In the table square fit values of the linear range for 45 ng enzyme for dC and 500 ng for 5hmdC are indicated.

3.4.4.3 CDA kinetic activity: Michaelis-Menten

After having determined initial velocity and substrate concentrations from which we could get a proportional absorbance reading, according to the Lambert-Beer law, and the extinction coefficient difference at 260nm ($\Delta\epsilon$), we proceeded with our kinetic measure-

ments. Collected data were fitted to the following Michaelis-Menten equation [Michaelis and Menten, 1913] with Prism (GraphPad):

$$V = \frac{V_{max}[S]}{K_m + [S]} \tag{3.12}$$

where [S] is substrate concentration in M, V_{max} is expressed as M/min and K_m as M.

3.4.5 Molecular modeling

A tetramer was generated with CDA structure 1MQ0 [Chung et al., 2005] and subject to DockPrep in Chimera 1.8 (www.cgl.ucsf.edu/chimera). Substrates were dC (ZINC18286013) [Irwin et al., 2012], 5hmdC [Irwin et al., 2012] (ZINC77300654) and 5fdC (CSID:10291642) (www.chemspider.com) downloaded as .mol files and subjected to .mol2 files conversion in Chimera. Docking was subsequently performed with the online tool SwissDock (www.swissdock.ch/docking) [Grosdidier et al., 2011]. The file was then visualized with Chimera and a picture generated with the lowest ΔG docked ligand.

3.4.6 *In vitro* replication assay

The assay was carried out following [Planck and Mueller, 1977] for nuclear extract and cytoplasmic fraction preparation and [Hershey et al., 2005] for the replication assay. The assay mixture contained 0.3 mM of each canonical nucleotide, except in test case when dCTP was substituted by 5hmdCTP (Bioline). The reaction was stopped with the addiction of 0.1 M EDTA final. DNA was extracted with Phenol chloroform, RNase A/T1 (Thermo Fisher) treated, as described in the DNA extraction section and free nucleotides removed with a Mini Quick Spin DNA column (Roche) prior to HPLC assay.

3.5 Protein analysis

3.5.1 Protein extraction

10^6 cells were lysed with RIPA buffer (20 mM HEPES at pH 7.5, 300 mM NaCl, 5 mM EDTA, 10% glycerol, 1% Triton X-100), supplemented with protease inhibitors (Mini, Roche) and sonicated (Diagenode), 30 s on/off for 5 minutes at high settings. The lysate was incubated for 20 minutes on ice and then cleared by spinning the cells at 10,000 rcf for 10 min at 4 °C. Mouse tumors were lysed as cells, with the exception of the use of a Dounce homogenizer with approximately 10 strokes of pestle B and 10 strokes of pestle A, to homogenize the sample before sonication and preclearing. Protein was quantified by Bradford assay, as detailed below.

3.5.2 Protein quantification

Protein was quantified by Bradford assay (B6916, Sigma) in a 96-well plate format. In brief, 200μL of reagent were used per well. A standard curve was produced with BSA (Molecular Biology Grade, NEB, B9000S), by diluting it in water to 1 μg/μL, ranging from 1 to 6 μg/μL and with the addition of 1 μL of lysis buffer. 1 μL of sample were measured in triplicate. Absorbance was measured in a plate reader (Magellan, Tecan) at 595 nm with previous 5 s shaking.

3.5.3 SDS-PAGE electrophoresis

20-100 μg of protein were dissolved in sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) loading buffer (5X: 200 mM Tris-Cl (pH 6.8), 400 mM DTT, 8% SDS, 0.4% bromophenol blue, 40% glycerol) and boiled at 100°C for 5 minutes. SDS-electrophoresis was performed on 4-20% precast gradient gels (Invitrogen, Novex Tris-Glycine). Gels were assembled, loaded and run at 200 V for 1 h in 1 X SDS-running buffer

(25 mM of Tris base, 192 mM of glycine, and 0.1% of SDS) using an Invitrogen tank (Nupage). Pre-stained protein marker (Page Ruler Plus, Thermo Scientific) was used as a size standard.

3.5.4 Immunoblot analysis

Gels were transferred to a wet transfer cassette unit (Biorad) and the proteins blotted onto nitrocellulose membrane, previously activated with water, for 1.5 h at a constant voltage of 110 V, or 30 V at 4 °C when the transfer was run overnight. The transfer tank was filled with 1 X transfer buffer (25 mM Tris-HCl pH 7.6, 192 mM glycine, 20% methanol, 0.05% sodium dodecyl sulfate (SDS)). When the transfer process was complete, the membrane was stained with Ponceau S solution (P7170, Sigma) to determine the success of the protein transfer. The membranes were then washed in water and blocked in 5% fat-free powder milk (Marvel, UK) dissolved in 1 X TBST (137 mM NaCl, 2.7 mM KCl, 19 mM Tris base pH 7.4 and 0.05% Tween-20), at room temperature for 1 h. They were then incubated overnight at 4°C with the primary antibodies listed in Table 3.11 diluted in 5% milk-TBST solution. The following day, the membranes were washed 3 times in TBST buffer for 5 minutes, by shaking them on an orbital shaker (Stuart) at medium speed. Subsequently, they were incubated with peroxidase conjugated appropriate secondary antibody (α-mouse, Santa Cruz sc-2005 $1:20,000$; α-rabbit, Santa Cruz sc-2004, $1:20,000$) for half an hour at room temperature. Membranes were successively washed for 5 minutes in TBST buffer, by shaking them on an orbital shaker (Stuart) at medium speed, three times. Subsequently, chemiluminescent detection was performed with ECL Select (GE Healthcare) peroxidase substrate as in the manufacturer's manual, through a CCD camera using the ChemiDoc System (BioRad) with Image Lab software (BioRad, version 4.0). If it was necessary to probe the same membrane with another primary antibody, the membrane was immediately incubated with stripping solution (Restore Western Blot Stripping Buffer, Thermo Scientific) for 30 min at room temperature on an orbital

Antibody	Catalog number	WB Dilution	Species
α-CDA	Sigma, SAB1300717	$1:250$	rabbit
α-actin	Abcam, ab185058	$1:75,000$	HRP conjugated
α-vinculin	Sigma, V9264	$1:10,000$	mouse
α-β-tubulin	Sigma, T8328	$1:10,000$	mouse
α-TET2	Active Motif, 61389	$1:1,000$	mouse
α-Flag	Sigma, F3165	$1:1,000$	mouse
α-S	Thermo-Fisher, MA1-981	$1:10,000$	mouse
α-GFP	Abcam, ab290	$1:1,000$	rabbit
α-PAF1	Bethyl, A300-172A	$1:500$	rabbit
α-CTR9	Bethyl/ A301-395A	$1:10,000$	rabbit
α-LEO1	Bethyl/ A300-175A	$1:1,000$	rabbit
α-RTF1	Bethyl/ A300-179A	$1:10,000$	rabbit
α-CDC73	Bethyl/ A300-170A	$1:10,000$	rabbit
α-PolII-PS5	Abcam/ ab5131	$1:1,000$	rabbit

Table 3.11: Primary antibodies used in Western Blot analysis with relative dilutions and catalog numbers.

shaker. Following these steps, the membrane was re-blocked in 5% milk/TBST, before being incubated with a new primary antibody.

3.5.5 Immunoprecipitation

Cells growing on plates were washed three times with cold PBS, then scraped and lysed in IP buffer (20 mM Tris-HCl pH 8, 10% glycerol, 300 mM KCl, 1 mM EDTA, 0.1% NP-40) containing protease inhibitors (Mini, Roche) according to [Muntean et al., 2010]. The lysate was then placed on an eppendorf rotating wheel at 20 rpm speed (Stuart, SB3) for 20 minutes, prior to centrifugation at 20,000 rcf at 4°C for 20 minutes (Eppendorf). The resulting supernatant was transferred to a new eppendorf, and protein concentration determined by Bradford assay. 50-100 μg (1% of the total IP) of protein was removed to be used as input. For each immunoprecipitation, 1-2 mg of protein were pre-cleared with 30 μL of 50% protein G/A beads mix (Roche), previously equilibrated in IP buffer, for 1 h at 4°C on an eppendorf rotating wheel. The protein lysate was then centrifuged at 12,000

rcf and the supernatant moved to a new eppendorf, to which the blocked beads and 1-2 μg of purified antibody or 30 μL of 50% Flag affinity resin (Sigma M2, A2220) or anti S-tag coated agarose beads (69704 Millipore, Novagen) were added. The tubes were left rotating overnight at 4 °on a wheel. The next morning, the beads were washed four times with cold IP wash buffer (50 mM Tris-HCl pH 7.9, 500 mM KCl, 0.3% NP-40) [Muntean et al., 2010] and centrifuged at 5,000 rcf for 3 minutes at 4°C. The supernatant was aspirated and the bound fraction eluted either with SDS-PAGE sample buffer or, in case of FLAG tagged protein, competitively with FLAG peptide, following the manufacturer's instructions (F3290, Sigma Aldrich). The samples were then boiled at 100°C, centrifuged once again at 12,000 rcf for 1 minute, and the supernatant analyzed by SDS-PAGE gel electrophoresis and Western Blot.

3.5.6 Large scale affinity purification for MS analysis

Affinity purification for MS analysis was carried out as previously described (IP), but with 4 mg of protein extract. 50 μL of 50% anti S-tag coated agarose beads (69704 Millipore, Novagen) were used for each AP. The elution was carried out with 400 μl 100 mM glycine pH 2.5 for 30 minutes on a rotating wheel at 4 °C. 1/20 of the elute fraction was analyzed by SDS-PAGE followed by Ruby-Pro gel staining (Biorad) according to the manufacturer's instructions; 1/20 was analyzed by western blot and the rest was sent for MS analysis to Dr. Kessler's laboratory (Target Discovery Institute, Oxford University). The results obtained from the Kessler lab were analyzed with the STRING online tool (http://string-db.org) [Franceschini et al., 2012] to generate network maps and background subtraction for the most common contaminants in MS analysis was performed with the CRAPPOME database [Mellacheruvu et al., 2013] by setting a threshold of 100 of observed proteins over their total database entries.

3.5.7 Sucrose gradient

A 60% sucrose stock solution was prepared in water. Sucrose stock solution was diluted in IP buffer without glycerol (20 mM Tri-HCl pH 8.0, 300 mM KCl, 1 mM EDTA, 0.1% NP-40), supplemented with protease inhibitors to prepare a 5% and a 20% dilution. A 5-20% sucrose gradient was pre-assembled in polyallomer tubes (Beckman, 14 x 89 mm, 331372), with 6.3 mL of 20% sucrose at the bottom and 6.3 mL of 5% sucrose at the top. The gradient was made with Gradient master (BIOCOMP) by selecting rotor SW41 and program SHORT SUCROSE 5-20%. 1 mg of protein extract, prepared with IP buffer, without glycerol (20 mM Tri-HCL pH 8, 300 mM KCl, 1 mM EDTA, 0.1% NP-40), supplemented with protease inhibitors, was loaded on top of the assembled gradient. A 1% input was saved as a loading control. The tubes were centrifuged at 260,343 rcf for 15 h at 4°C with maxiumum acceleration and no brakes (Beckman, SW41 rotor). 24 fractions of 500 μL were subsequently collected with care from top to bottom, with numbering starting from the top. Proteins were extracted by the trichloroacetic acid (TCA) precipitation method. In brief, 89 μL of TCA (T6399, Sigma) was added to the fractions, followed by vigorous shaking and incubation at 4°C for 1 h. The fractions were then centrifuged at 20,000 rcf for 15 minutes at 4°C. Supernatant was aspirated and the protein-containing pellet washed with 500 μL acetone before being pelleted at 20,000 rcf for 15 minutes. Supernatant was discarded, the pellet air dried and resuspended in 50 μL SDS-PAGE loading buffer. 10 μL were loaded on 4-20% MIDI-Protean®Tris-glycine™Stain free gels (Biorad), electrophoresed and immunoblotted as before. A protein standard, composed of 6 different proteins of molecular weights comprised between 29 and 700 kDa (MWGF1000, Sigma Aldrich) was run under the same conditions as the samples and visualized by CCD camera on a ChemiDoc imager (Biorad), through Image Lab software (BioRad, version 4.0).

3.5.8 Chromatin immunoprecipitation

Chromatin immunoprecipitation was performed according to previous protocols [Frank et al., 2001]. Formaldehyde dissolved in phosphate buffered saline (PBS) to a final concentration of 1%, was added to cultured cells (approx. $10x10^6$ cells per immunoprecipitation) to fix them at room temperature for 10 minutes. Fixation was stopped by the addition of glycine to a final concentration of 0.125 M for 5 minutes. Plates were then rinsed twice with room temperature PBS. PBS was aspirated completely, and cells harvested in SDS buffer (100 mM NaCl, 50 mM Tris/HCL pH 8.1, 5 mM EDTA pH 8.0, 0.2% sodium azide, 0.5% SDS) containing protease inhibitors (10X, Roche). Cells were pelleted by spinning in a tabletop centrifuge for 10 min at 6,000 rcf at room temperature, and resuspended in 4 mL of ice-cold IP Buffer for sonication (IP buffer: 1 volume SDS Buffer + 0.5 volume Triton Dilution Buffer. Triton Dilution Buffer: 100 mM Tris/HCl pH 8.6, 100 mM NaCl, 5 mM EDTA, 0.2% sodium azide, 5.0% Triton X-100).

Samples were sonicated to an average length of chromatin of 500-1,000 bp according to the manufacturer's instructions (Covaris). To check the average length of the sonicated DNA, crosslinks were reverted by incubating a 50 μL aliquot of the sample for 30 min at 65°C, purifying the DNA with a QIAquick PCR Purification Kit (Qiagen) following the manufacturer's instructions, and loading it on a 2% agarose gel. Subsequently, the volume of the extract, was adjusted with IP buffer to 1 mL per IP. Lysates were pre-cleared with 30 μL of Protein-A beads (50% v/v in TE) for 1 hour at 4°C on a rotating wheel; then centrifuged for 10 minutes at 12,000 rcf on a tabletop centrifuge. After pre-clearing, 50 μL of each lysate was removed to be used as total input and stored at 4°C (5%).

The primary antibody was added (3 μg for each IP) and the mix incubated overnight at 4°C on a rotating wheel. After centrifugation for 20 minutes at 15,000 rcf at 4°C, the supernatant was added to 30 μL of 50% Protein-G beads, and incubated on a rotating wheel for three hours at 4°C. Beads were washed 3 times in 1 mL of Mixed Micelle Wash Buffer (150 mM NaCl, 20 mM Tris/HCl pH 8.1, 5 mM EDTA pH 8.0, 5.2% w/v sucrose,

0.02% sodium azide, 1% Triton X-100, 0.2% SDS), 2 times in 1 mL Buffer 500 (0.1% w/v deoxycholic acid, 1 mM EDTA, 50 mM HEPES pH 7.5, 500 mM NaCl, 1% v/v Triton X-100, 0.2% sodium azide), 2 times in 1 mL LiCl/detergent buffer (0.5% w/v deoxycholic acid (sodium salt), 1 mM EDTA, 250 mM LiCl, 0.5% v/v NP-40, 10 mM Tris/HCl pH 8, 0.2% sodium azide) and once in 1 mL Tris/EDTA pH 7.4.

120 μL of 2% SDS solution was added to both beads and input and samples were incubated overnight at 65°C to elute immune complexes and to reverse the crosslinks. Eluted DNA was purified with a QIAquick PCR Purification Kit (Qiagen) following the manufacturer's instruction. Samples were resuspended in 300 μL and the input in 600 μL of 10 mM Tris/HCl pH 8.0.

The antibody used was α-Flag (Sigma, F3165). For the negative control, a mouse anti-IgG antibody (sc-2004, Santa Cruz Biotechnologies) was used.

Bound regions were identified by running QPCR on 6 μL of the immunoprecipitated DNA with conditions as as outlined in the previous QPCR method section. Data analysis was performed using the ΔC_T method; in particular, fold enrichment in the IP was calculated as a percentage of the total DNA present in the input (see Fig. 3.6).

Initially the ΔC_T is calculated as follows:

$$\Delta C_T = C_T\,(input) - C_T\,(IP)\,. \tag{3.13}$$

Then, the quantity of DNA immunoprecipitated is derived, as a percentage of the total DNA in the input, as:

$$\%total = 2^{\Delta C_T} \times 2.5. \tag{3.14}$$

The list of primers used for ChIP-QPCR are listed in Table 3.12.

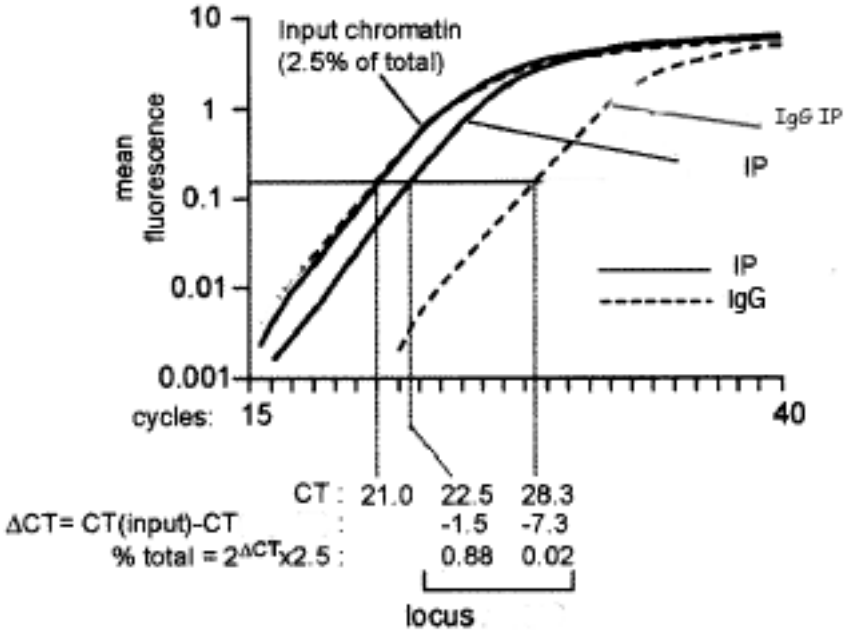


Figure 3.6: Quantification of chromatin immunoprecipitation by real-time PCR (adapted from [Frank et al., 2001]).

Primer Pair	Sequence	Target
SK615	TCTAACCTTTCCAAGTCCTCGTAAA	HOXA9
SK616	GGCGGGAAGTCGGAAACG	
SK619	ACGTTGGCCACAATTAAAACA	HOXA9
SK620	AAATCACTCCGCACGCTATT	
SK621	CCTGTGTGGCTTCTGAAACA	HOXA9
SK622	CAAATCGCATTGTCGCTCTA	
SK623	GGGAGACGGGAGAGTACAGA	HOXA9
SK624	GGCCAACGACGATTAAAAGA	
SK631	AGGCGTCTAACGCTATGCTC	MLL1
SK632	AAGAAGGAGGCGAGTTAGGG	
SK633	TTTGGAAAGAGCTCCAGCAT	MLL1
SK634	AAGCGCTGCCCAGTTTACTA	

Table 3.12: Primers used for ChIP-QPCR.

3.6 Microscopy techniques

3.6.1 Immunofluorescence

Cells were seeded on coverslips and stained by immunofluorescence with antibodies listed in Table 3.13, with their relative dilution. In brief, cells were first washed in PBS and fixed for 20 min at room temperature with 4% PFA. Cells were washed twice in PBS and permeabilized for 10 minutes in 0.2% Triton X-100. After 2 washes in PBS, cells were blocked for 1 h in 3% BSA solution prepared in PBS (Sigma Aldrich) and incubated with primary antibody overnight at 4°C in a humidified chamber. The following day, cells were washed 3 times in PBS and incubated with the appropriate Alexa conjugated secondary antibody (Table 3.13) and DAPI (Sigma Aldrich). Coverslips were then washed 3 times in PBS and mounted with mounting media (Vectashield). Tiled pictures were automatically taken with a Zeiss 710 microscope with a 20x lens.

Antibody	Catalog number	WB Dilution	Species raised
$\alpha-\gamma$-H2AX	Millipore, 05-636	1 : 500	mouse
α-TET2	Active Motif, 61389	1 : 200	mouse
α-Flag	Sigma, F3165	1 : 400	mouse
α-S	Thermo-Fisher, MA1-981	1 : 400	mouse
α-GFP	Abcam, ab290	1 : 400	rabbit
Alexa Fluor®488 α-rabbit	Life Technologies, A-11008	1 : 400	goat
Alexa Fluor®488α-mouse	Life Technologies, A-11001	1 : 400	goat
Alexa Fluor®546α-mouse	Life Technologies, A-11003	1 : 400	goat
Alexa Fluor®546α-rabbit	Life Technologies, A-11003	1 : 400	goat
Alexa Fluor®647α-rabbit	Life Technologies, A-21244	1 : 400	goat

Table 3.13: Antibodies used in immunofluorescence.

3.6.2 Proximity ligation assay

The proximity ligation assay (PLA) [Soderberg et al., 2006] is based on the direct recognition of a target protein by two primary antibodies raised in different species, and on their recognition by two respective secondary antibodies called PLA probes. PLA probes have a unique DNA strand attached to them which, when the secondary antibodies are in

close proximity, can be ligated and amplified through rolling circle amplification (RCA) by a DNA polymerase [Gusev et al., 2001], while incorporating a fluorochrome in the so-formed double stranded DNA. This results in the formation of bright spots that can be easily detected by confocal microscopy (Fig. 3.7).

To conduct the assay, we used α-S-tag (Thermo-Fisher - MA1-981, 1 : 400, mouse) and

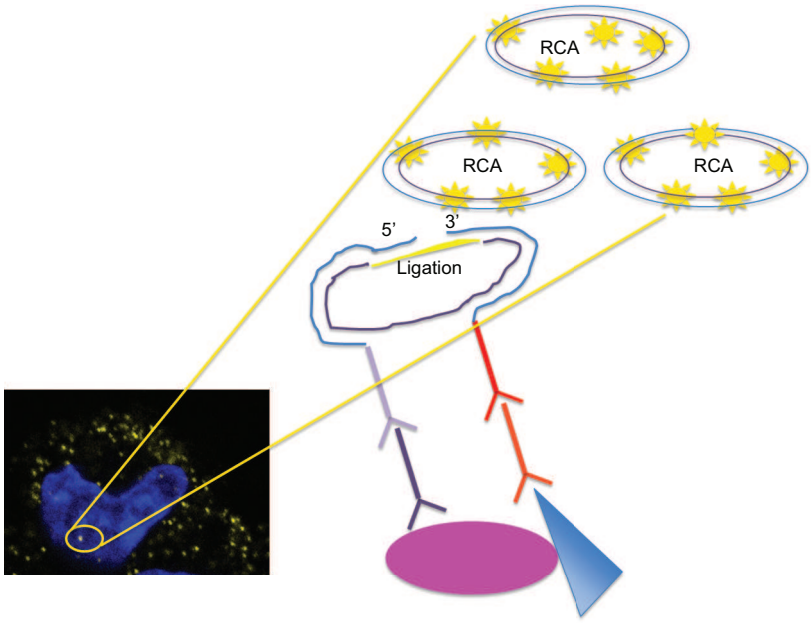


Figure 3.7: Proximity ligation assay: the method. A spot is generated when the two antibodies recognizing the target protein get in close proximity. They are conjugated to oligonucleotides that are ligated and amplified through rolling circle amplification (RCA). The amplified double stranded DNA can incorporate the fluorescent dye, resulting in the formation of localized dots in proximity of the detected proteins.

α-GFP (Abcam - ab290, 1 : 400, rabbit) antibodies against GFP-S-tagged WDR-61 as positive controls and α-GFP, α-Flag (Sigma - F3165, 1 : 400, mouse) to assess the interaction between Tet2 and WDR-61. Different antibody concentrations were assessed to determine which one would give the best signal/noise ratio; where noise was defined as the presence of a fluorescent signal in a negative control sample. The reaction progressed as for the immunofluorescence protocol outlined earlier up until the addition of the secondary antibodies. After this point, conditions specified in the datasheet of the Duolink®

In Situ Red Starter Kit Mouse/Rabbit (Duo92106, Sigma Aldrich) were used. Images were acquired with a 63x lens on a Zeiss 710 confocal microscope. For quantification, tiled images with 30 Z stacks were acquired and quantification was performed on their maximum intensity projection generated through ZEN software by overlaying all of the Z stacks. The number of dots was counted with ImageJ software [Schneider et al., 2012].

3.6.3 Histology

Organs and tumors were collected and immediately 10% formalin fixed for 48 h. They were then dehydrated in an ethanol series, cleared in xylene (Leica, TP1020) and embedded in paraffin wax (VWR, 361427G). 4 μm thick sections were cut with a microtome (Microme HM355S) with the help of I. Ratnayaka (Ludwig Institute). All sections were stained with hematoxylin and eosin. In brief, tissue sections were dewaxed with HistoClear (National Diagnostics) for 5 minutes, rehydrated through an ethanol series, immersed in Harris hematoxylin (Sigma, HHS128) for 3 minutes, differentiated with acid alcohol for 5 dips, blued in Scott Water for 30 seconds and then immersed in eosin for 5 minutes, with washes in water between each step. Sections were then dehydrated and permanently mounted (Vectamount – Vector Labs, CA, USA). To grade the level of fibrosis in the samples, a scale was created with the following criteria: 1 (fibrosis $\leq$ 5), 2 ($\leq$ 5 fibrosis $\geq$ 25), 3 ($\leq$ 26 fibrosis $\geq$ 50) and 4 (fibrosis $\geq$ 51) (Fig. 3.8). The tumor grading was done double blind in collaboration with Prof. R. Goldin (Imperial College, London). Tumors were additionally stained with Masson's Trichrome Stain Kit (Sigma Aldrich) according to the manufacturer's instructions.

3.6.4 Immunofluorescence on paraffin embedded samples

4 μm thick sections were re-hydrated through a series of alcohols and subjected to antigen retrieval, by boiling them for 2 minutes in a pressure cooker (Lacor, ES) in Tris buffer

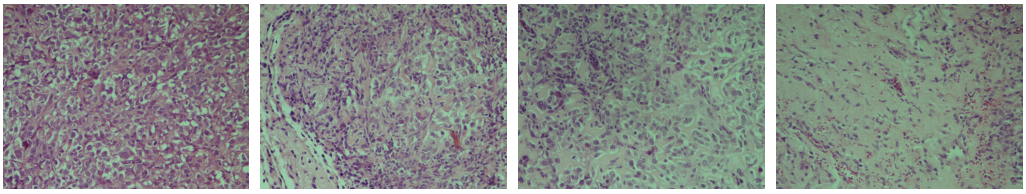


Figure 3.8: Histology fibrosis grading scale. Done in collaboration with Prof. R. Goldin (Imperial College).

pH 9 (10 mM Tris base, 0.05% Tween 20). They were then blocked in 3% BSA in PBS for 30 minutes and incubated overnight in a humidified chamber at 4°C with $\alpha-\gamma$H2A.X (1:200, Millipore, 05-636), $\alpha-$PH3 (1:200 Millipore, 06-570) or $\alpha-\beta-$catenin (1:250 BD Transduction Laboratories, 610153) and $\alpha-$CDA (1: 100 Sigma Aldrich, SAB1300717). The slides were then washed vigorously 3 times in PBS, and incubated for 1 h at room temperature with an appropriate Alexa546 and Alexa488 conjugated secondary antibody (1:400, Life Technologies) and DAPI (Sigma Aldrich). Coverslips were then washed 3 times in PBS and mounted with mounting media (Vectashield - Vector Labs, CA, USA). Images were acquired with a Zeiss 710 confocal microscope with a 20x lens. For quantification of the percentage of DNA damage and proliferation in tumors, tiled images with Z stacks were acquired to cover the entire central section of the tumor.

3.6.5 Microscopy quantification with Image J

All microscopy quantification was performed through Image J, with the help of M. Shipman (Ludwig Institute).

The amount of nuclear fluorescence for $\alpha-\gamma$-H2AX was quantified through ImageJ through the following steps:

1. Drag & drop .lsm file to ImageJ;
2. Image>Color>Split Channels;
3. Select the DAPI Channel;
4. Process>Filters>Gaussian Blur, 2px radius;
5. Image>Adjust>Threshold, Default, Check the Dark Background option;
6. Process>Binary>Fill Holes;
7. Process>Binary>Watershed;
8. Analyze>Analyze Particles>size=100-Infinity, circularity=0-1, show=nothing, exclude on edges, clear results, add to ROI;
9. Select the red channel image;
10. In the Region Manager window click measure;
11. Save the measurements from the menu in the results window.

An equal number of nuclei data were subsequently processed in Excel (Microsoft) by first generating an histogram of the frequency distribution of the values of signal intensities. Subsequently, the threshold was set after the most frequent intensities which appeared equally in all samples, independently of treatment. Percentages were then calculated by dividing the events counted after threshold subtraction for the number of events measured. The number of nuclei was quantified through the DAPI signal and the number of those positive for H3P signal (green) and $\alpha-\gamma$H2AX (red) was counted with ImageJ (Fig. 3.9). The following steps were used to count the DAPI signal:

1. Image>Properties (channels=1 slices=1 frames=1 unit=μm pixel_width =.415 pixel_height=.415 voxel_depth=.415);

2. Image>Color>Split Channels;
3. Select the DAPI Channel;
4. Duplicate the image;
5. Process>Filters>Gaussian Blur, 2px radius;
6. Image>Adjust>Threshold, Default, Check the Dark Background option;
7. Process>Binary>Convert to Mask;
8. Process>Binary>Fill Holes;
9. Process>Binary>Watershed;
10. Analyze>Analyze Particles>size=20-140, circularity=0-1, show=nothing, clear results, add to ROI;
11. Select the original DAPI Channel;
12. In the Region Manager window click measure;
13. Save the measurements from the menu in the results window.

For the green (H3P) and the red (γH2AX) channels, the same steps were run as for DAPI, except that the Threshold step was Yen dark and the particle size analyzed was 20-1400. The frequency of events was then calculated, by dividing the number of red and green measures for the total number of nuclei measured.

To count the number of spots in PLA stainings the following steps were used in Image J:

1. Drag & drop .lsm file to ImageJ,
2. Image>Color>Split Channels;
3. Select the Red Channel (PLA spots);

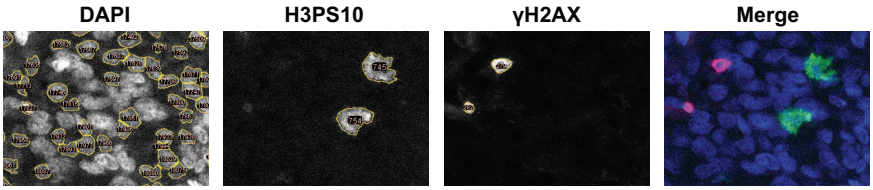


Figure 3.9: Quantification of $\alpha - \gamma$H2AX and $\alpha-$PH3: the method

4. Image>Adjust>Threshold, Default, Check the Dark Background option;
5. Save as ImageC3_threshold;
6. Select the Green Channel (GFP-S-WDR61);
7. Process>Filters>Gaussian Blur, 2px radius;
8. Image>Adjust>Threshold, Default, Check the Dark Background option;
9. Save as ImageC2_threshold;
10. Select the DAPI Channel;
11. Process>Filters>Gaussian Blur, 2px radius;
12. Image>Adjust>Threshold, Yen, Check the Dark Background option;
13. Save as ImageC3_threshold;

To measure the PLA spot numbers, we used the following steps:

1. Drag & drop .lsm file to ImageJ
2. Image>Color>Split Channels;
3. Select the RED Channel (PLA spots);
4. Open ImageC3_threshold;
5. Analyze>Analyze Particles>size=0.1-Infinity, clear results, add to ROI;

6. Select RED channel;

7. Analyze>Analyze Particles>size=0.1-Infinity, clear results, add to ROI;

8. In the Region Manager window uncheck and recheck Show All then click Measure.

3.7 Animal work

3.7.1 Toxicology and dose determination

Animal work was done after approval by the UK Home Office and University of Oxford Local Ethical review. The project license number was 30/3033 and personal license number was 30/10162. The compounds were dissolved in PBS and intraperitoneally (IP) injected in 3 animals (Athymic Nude: BALB/cOlaHsd-$Foxn1^{nu/nu}$, females, 5-7 week old from Harlan) per dosage in 200 μL volume. Animals were monitored for any deviation in normal behavior and sacrificed after 5 days for histopathology. At 30 minutes post injection, a few drops of blood were collected through tail vein bleeding with a Microvette CB300 (Sarstedt) collection device. The serum fraction was isolated by centrifugation according to the Microvette CB300 (Sarstedt) manufacturer's instructions.

3.7.2 MS analysis of serum samples

The molecules present in serum samples were extracted via methanol chloroform precipitation. In brief, samples were brought up to 200 μL with double distilled sterile water. 3 volumes of methanol and 150 μL of chloroform were added. The samples were vortexed and 450 μL of water added. After vortexing, they were centrifuged at maximum speed for 1 min. The aqueous phase containing the water soluble molecules was collected and dried in a Speedvac. In the same way, a standard curve was produced by dissolving known amounts of 5hmdC and 5fdC in fetal bovine serum (Biosera) and extracting them (Fig

3.10). The dried pellets were then sent for MS analysis to G. Berridge (TDI, Oxford).

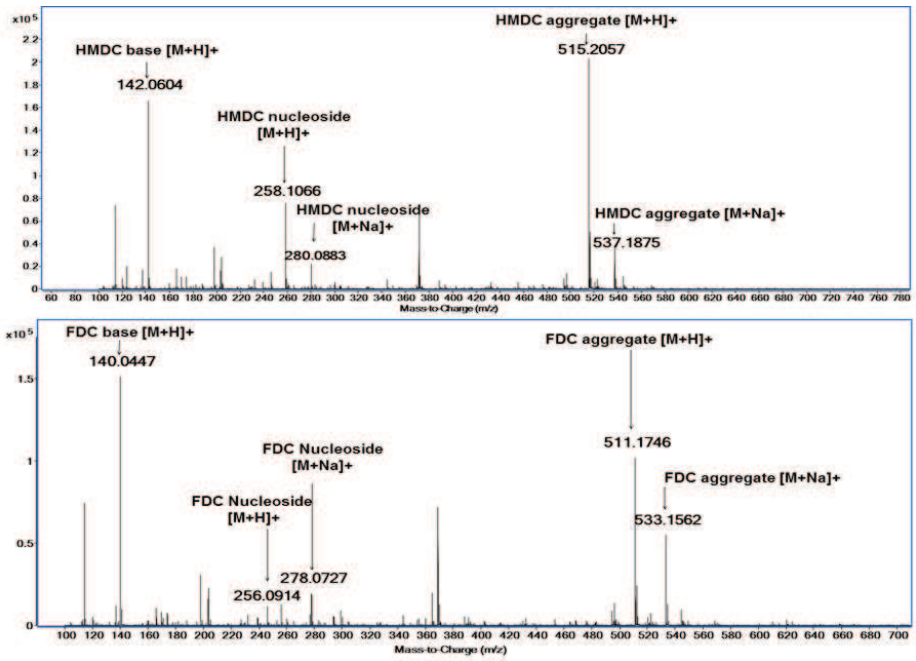


Figure 3.10: Mass spectrum of 5hmdC and 5fdC standards. This analysis was performed in collaboration with G. Berridge (TDI, Oxford).

3.7.3 Tumor growth assessment using xenograft assay

10^6 cells in a 50% suspension with MatriGel (200 μL) were subcutaneously injected into BALB/cOlaHsd-*Foxn*1$^{nu/nu}$ (Harlan, 5-7 week old) mice, 8 animals per group, in each flank following the scheme: SN12C/H1299 (left), SN12CshCDA8/H1299_ dsRedCDA (right). When the tumors reached palpable size, 8 mice were assigned to each treatment group: PBS, 100 mg/kg of 5hmdC and 100 mg/kg 5fdC. The drugs were dissolved in PBS and administered intraperitoneally in a volume of 0.2 mL per body weight every 72 h for a total of 4 injections. Tumor size was measured every 3 days by Vernier caliper (Sigma Aldrich) and the animal cohort sacrificed when the cumulative tumor diameter in the first animal reached 12 mm. Tumor volume was calculated assuming that the tumors were spheres with the following formula: $4/3\pi(D/2)^3$, in which D represents the diameter of

the tumor. Tumors were dissected and genotyped by PCR with primers amplifying the lentivirus genome used to transduce the originating cell lines (Table 3.14).

PP	Sequence	Target	T_{ann}
SK301 SK489	CTCAAATTCACGGTAAATGGCCCGCCTGG GAGTCAAACCGCTATCCACGC	dsRedCDA	56°C
SK251 SK847	AAACCCAGGGCTGCCTTGGAAAAG GAGGGCCTATTTCCCATGATTCC	shCDA_8	56°C

Table 3.14: Primers to genotype cancer cell line derived tumor xenografts.

3.8 Datasets used for the analysis of gene expression

To assess CDA expression levels in all the cell lines utilized in the study, the following datasets were mined: GSE36139 (GPL15308) [Barretina et al., 2012], GSE32474 (GPL570) [Pfister et al., 2009] and GSE1133 [Su et al., 2004].

GPL15308, GPL570 and GSE16515 [Pei et al., 2009] were analyzed directly on the NCBI portal with GEO2R to extract normalized values for CDA expression for all the cell lines included in the study. For cluster analysis, GSE32474 was analyzed directly on the GEO2R portal and the data analysis performed via t test, with Benjamini-Hochberg correction.

Expression values of *CDA* were obtained searching among human tissues datasets in GENT database (medical-genome.kribb.re.kr/GENT) and the plot was downloaded [Kim et al., 2011].

Chapter 4

Mammalian salvage of deoxycytidine derivatives

In mammalian DNA, cytosine can have different variables at the fifth position on its carbon ring. The first one to be discovered was 5-methylcytosine (5mC) in calf thymus DNA [Hotchkiss, 1948]. Following this discovery, much research was directed towards establishing the biological meaning of the modification. Consequently, several reports shed light on the repressive role of DNA methylation in gene expression, first *in vivo* -such as in the case of the promoter of the β-globin gene in multiple tissues [van der Ploeg and Flavell, 1980]- and later in cell lines transfected with methylated constructs [Busslinger et al., 1983]. Around the same time, DNA methylating enzymes were discovered: DNMT1 was purified from mammalian cells [Bestor and Ingram, 1983] and the gene later isolated [Bestor et al., 1988] and DNMT3 [Li et al., 1998] was subsequently identified through genomic studies. We now know significantly more about the role of 5mC and its involvement in signaling events that lie outside the coding DNA sequence, such as transcriptional repression, X chromosome inactivation, transposon and retroelement promoter silencing [Bestor, 2000]. More recently, it has been discovered through large genomic studies that DNA methylation

plays major roles in diseases, such as cancer, where its levels were observed to be globally decreased [Feinberg and Vogelstein, 1983][Gama-Sosa et al., 1983] or locally increased at promoters of oncosuppressors, such as p16 [Gonzalez-Zulueta et al., 1995]. Given the fundamental roles of 5mC mentioned above, strategies targeting DNA methylation pathways, have been investigated for their potential as anti-cancer therapies [Jones, 1984][Matsuda and Sasaki, 2004]. They are mainly pursued via the use of DNA methylation inhibitors, with the rationale of reactivating the expression of oncosuppressors, such as p16, that have become aberrantly silenced. While major research focus was on the role of DNA methylation in signaling events, another branch of research was directed towards identifying enzymes of the nucleotide salvage pathway that could have 5mC as a substrate, such as: cytidine deaminase (CDA), which would produce dT [Cacciamani et al., 1991]; and deoxycytidine kinase (DCK), which would produce 5mdCMP [Eriksson et al., 1991]. However, although 5mdCMP could be detected, a diphosphate form (5mdCDP) could not be produced [Vilpo and Vilpo, 1991].

In 2009, two seminal papers [Kriaucionis and Heintz, 2009][Tahiliani et al., 2009] were published that showed the presence of 5-hydroxymethylcytosine (5hmC) in mammalian DNA. Later, another two modifications were added: 5-formylcytosine (5fC) and 5-carboxylcytosine (5cadC) [Ito et al., 2011]. Since their discovery, a number of papers have related these new forms of cytosine to a possible DNA demethylation pathway [Tahiliani et al., 2009][Hajkova et al., 2010][Guo et al., 2011]. Other studies have focused their efforts on trying to understand the link between 5hmC and cancer. 5hmC is produced from 5mC by oxidation of its methyl group at the fifth position of the carbon ring, by the TET family of proteins [Iyer et al., 2009][Tahiliani et al., 2009], whose genes (*TET1* and *TET2*) are found to be mutated in cancer [Delhommeau et al., 2009][Lorsbach et al., 2003]. Moreover, levels of 5hmC in cancer have been found to be globally reduced [Yang et al., 2012][Kraus et al., 2012]. A few studies have investigated the effects that these modifications have on transcription elongation rate and DNA replication. It was shown, both with mammalian and

yeast RNA Polymerase II, that 5fC and 5caC are able to inhibit transcription elongation and promote pausing on templates containing these modified bases [Kellinger et al., 2012]. Moreover, in a tissue culture model it was demonstrated that templates containing 5fC and 5caC modestly inhibited replication, while 5hmC-containing plasmids showed no significant impairment [Ji et al., 2014].

Nucleotides are produced either *de novo* or recycled through the salvage pathway. Since *de novo* production of modified cytosines had never been demonstrated, it was decided to exclude the possibility that these modified cytosines could be derived from dC in the cell. Another possible source of nucleosides is the nucleotide salvage pathway. Therefore, there was a theoretical possibility of having external sources of 5hmC and 5fC, either unphosphorylated or in monophosphate forms, from intracellular sources such as DNA repair processes, with 5fC amount being negligible as compared to 5hmC, given its low abundance in the DNA [Ito et al., 2011]. Thus, it was first investigated whether the barrier to the incorporation of these newly discovered cytosines could lie in their incorporation into the DNA of replicating cells. Given the fact that nucleoside triphosphates cannot be imported by mammalian cells, whether or not 5hmdC, 5fdC and 5cadC could be produced by the nucleotide salvage pathway enzymes involved in the metabolism of dC, such as CDA, DCK and CMPK1, was also addressed.

4.1 Results

4.1.0.1 Evaluating the possibility of 5hmdCTP incorporation into the DNA during replication

Nucleotide production in cells is dependent on *de novo* biosynthesis or salvage pathways. Possible intracellular sources for nucleotide recycling lie on byproducts of DNA damage [Cooke et al., 2009], cellular nucleotide sanitation pathways [Mo et al., 1992][Tchigvintsev

et al., 2013], which both act as protective mechanisms against erroneous incorporation, or mitophagy, that is selective degradation of mitochondria upon starvation [Youle and Narendra, 2011]. External sources of nucleotides were excluded because mammalian cells have not evolved the ability to import nucleotide triphosphates. Nucleotide transporters have only been identified in Bacteria [Winkler, 1976], Plants [Kampfenkel et al., 1995] and Algae [Ast et al., 2009]. Conversely, mammalian cells possess a variety of nucleoside transporters that are divided into two families: concentrative nucleoside transporters (CNT) and equilibrative nucleoside transporters (ENT). The CNT family consists of 3 members that are responsible for nucleosides transport coupled to a sodium flux, while the ENT family has four members that mediate the diffusion of nucleosides and therefore display lower substrate specificity (reviewed in [Cano Soldado and Pastor-Anglada, 2012][Molina-Arcas et al., 2008]). It has been shown that both cellular proliferation [Soler et al., 2001] and fasting [Valdés et al., 2000] can upregulate nucleoside transport mechanisms, linking their regulation to increased nucleoside recycling and *de novo* biosynthesis. Given the sources of intracellular nucleotide recycling, it was speculated that 5hmdCTP could be found in actively replicating cells and thus, first it was investigated whether there is a barrier to 5hmdCTP incorporation in the replication machinery. To answer this question, given the lack of mammalian nucleotide triphosphates transporters, nucleotides had to be artificially delivered into the cells. Hence, three systems were established:

1. *In vitro* replication assay (IVRA);
2. Nucleoporation of 5hmdCTP in MDA-MB-231;
3. Lipofection of 5hmdCTP in H1299.

***In vitro* replication assay (IVRA)** The first system adopted [Planck and Mueller, 1977][Hershey et al., 2005] enabled the investigation of the research question in a cell free system, as isolated nuclei, that are permeable to nucleotides, from synchronized HeLa cells

that were actively replicating. A detectable amount of 5hmdC was measured by HPLC as compared to dCTP-treated nuclei (Figure 4.1).

In summary, IVRA showed that 5hmdCTP can be incorporated by the replication ma-

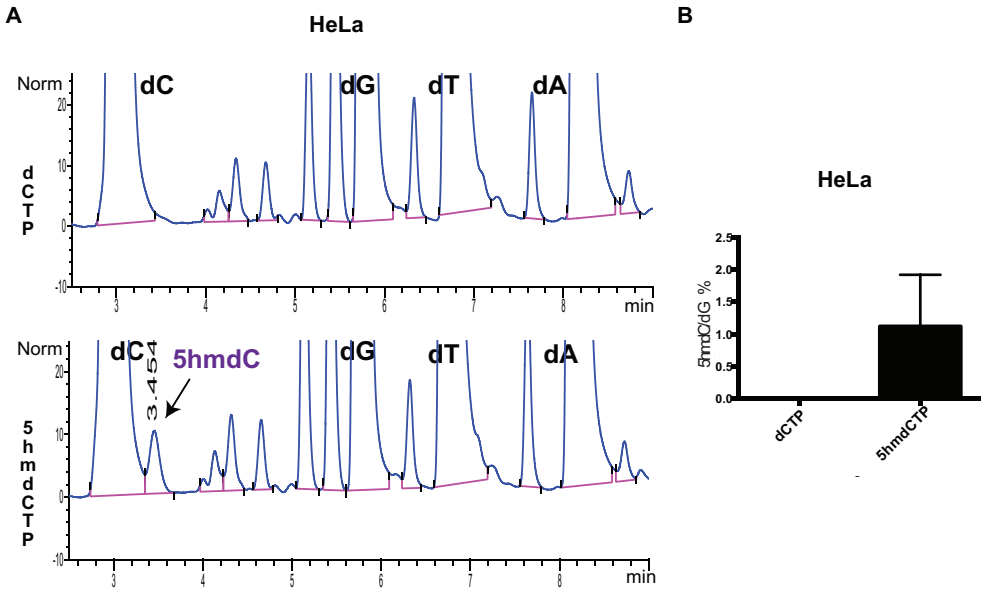


Figure 4.1: IVRA. A. Representative HPLC chromatograms from IVRA. 5hmdC is highlighted in the chromatograms with elution time. Norm. absorbance units; Min, minutes. B. Quantification of the amount of 5hmC incorporated in HeLa DNA. n.d. not detectable. (n =3. Whiskers indicate standard deviation).

chinery when it is available. No detectable amount of 5hmC was found in dCTP-treated samples.

Nucleoporation of 5hmdCTP in MDA-MB-231 The second system investigated was the nucleoporation of the triphosphate in MDA-MB-231 cells, a breast cancer cell line. In this case, small pores were created in the cell membrane via electric charges, enabling the triphosphates to enter the cells. With this method, 5hmdC could be detected in DNA 48 hours post nucleoporation, when at least one round of replication had occurred (Figure 4.2), and at a similar percentage to that shown in Figure 4.1 for IVRA.

Another difference that could be noticed on the HPLC chromatograms was a peak for

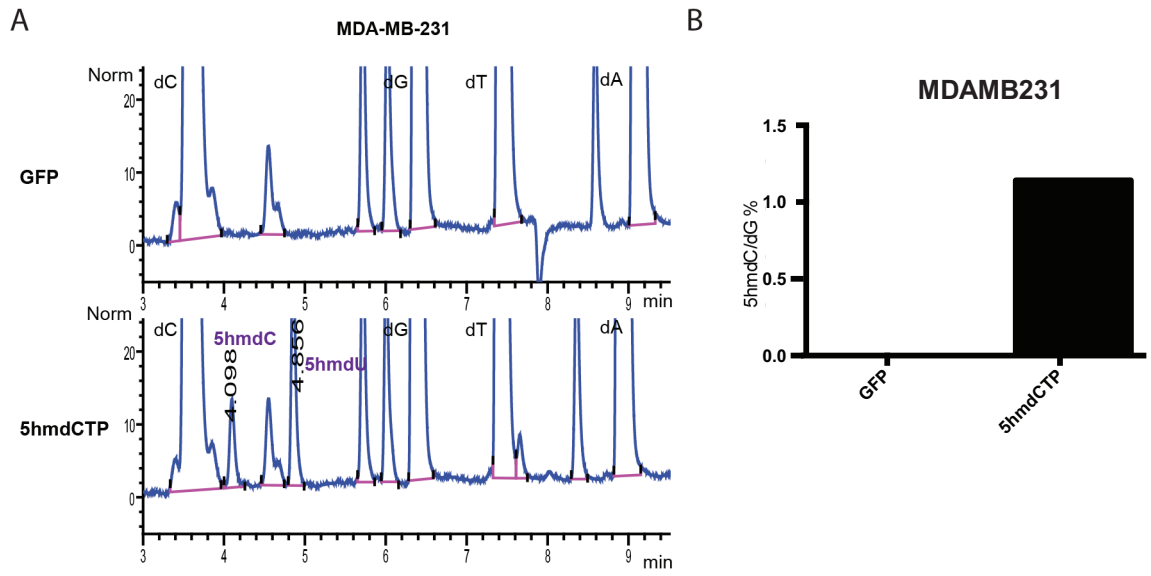


Figure 4.2: Nucleoporation of 5hmdCTP in MDA-MB-231. A. Representative HPLC chromatogram of GFP- and 5hmdCTP-transfected cells. Norm. absorbance units; Min. minutes. 5hmdC and 5hmdU are highlighted in the chromatograms with elution time. B. Quantification of the amount of 5hmdC incorporated into MDA-MB-231 DNA (n =3. Whiskers indicate standard deviation).

5hmdU. 5hmU has been shown to be present in mammalian DNA, originating from the oxidation of thymine both upon exposure of DNA to damaging agents, like ionizing radiation [Frenkel et al., 1985] or free oxygen radicals [Mouret et al., 1991], or from enzymatic conversion by the TET family enzymes [Pfaffeneder et al., 2014] or from 5hmC deamination by AID/APOBEC enzymes [Guo et al., 2011]. Furthermore, 5hmU has been shown to be incorporated in the DNA when cells are treated with 5hmdU [Kaufman, 1986]. Additionally, it has been demonstrated *ex vivo* that spontaneous deamination of 5hmdC is unlikely to occur in solution [Schiesser et al., 2013]. Among all of them, it could be speculated that 5hmU in the DNA of MDA-MB-231 cells may come from two routes: the enzymatic deamination of 5hmdC by CDA, which would generate 5hmdU that would then be incorporated; and the enzymatic deamination of 5hmdCMP to generate 5hmdUMP by DCTD. DNA damage could be excluded as a possible source of 5hmU, since such a peak was not observed on the chromatogram of untreated cells and cell death was not seen

following electroporation.

Lipofection of 5hmdCTP in H1299 The third system investigated was classical lipofection, which is known to give a high transfection efficiency, in H1299 cells, a lung cancer cell line. This method relies on the formation of positively charged liposomes around negatively charged particles, allowing them to cross the negatively charged plasma membrane [Chu et al., 2009]. The presence of 5hmC in DNA was confirmed at 48 hours post transfection, when at least one round of replication had occurred (see Fig. 4.3). This indicated that DNA polymerase was not a barrier for 5hmdCTP incorporation.

Lower incorporation of 5hmdC was observed in this cell line compared to electroporated

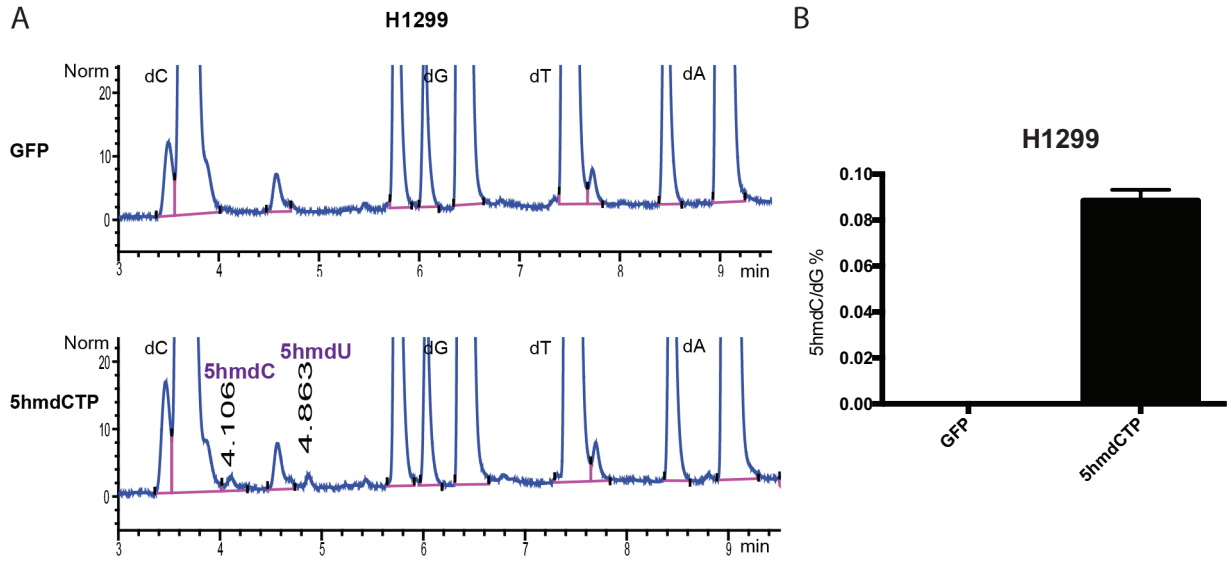


Figure 4.3: Lipofection of 5hmdCTP in H1299. A. Representative HPLC chromatograms of GFP- and 5hmdCTP-transfected cells. 5hmdC and 5hmdU are highlighted in the chromatograms with elution time. Norm. absorbance units; Min. minutes. B. Quantification of the amount of 5hmdC incorporated into H1299 DNA (n =3. Whiskers indicate standard deviation).

cells (Fig. 4.2). This reduced incorporation could be explained by the faster replication time of H1299 [Chan et al., 2012], which would dilute the incorporated nucleotide, and by the lower efficiency of lipofection, as compared to nucleofection [Maurisse et al., 2010]. 5hmdU could also be detected in the chromatograms of these 5hmdCTP-treated cell lines

(Fig. 4.3). Again, the possibility that it derived from DNA damage could be excluded, as it was not present in dCTP-treated cell lines.
From these sets of experiments, it could therefore be concluded that when 5hmdCTP is present, it can be incorporated by actively replicating cells.

4.1.0.2 Assessment of the possibility of monophosphate production

Having validated 5hmdCTP incorporation in the DNA of actively replicating cells, the decision was made to investigate whether 5hmdCTP could be produced in the cells. *De novo* pathways have never been shown to produce cytidine triphosphates modified at the fifth position. Therefore, attention was directed to the salvage pathway. An ideal candidate to understand whether the route of nucleoside triphosphate production could be undertaken, was deoxycytidine kinase (DCK), a protein of the nucleoside salvage pathway, which, in mammals, is responsible for the production of deoxycytidine monophosphate in the cytoplasm (dCMP) [Bohman and Eriksson, 1986][Hatzis et al., 1998] by catalyzing this reaction:

$$\mathrm{dC} + \mathrm{NTP} \rightleftharpoons \mathrm{dCMP} + \mathrm{NDP}$$

where NTP indicates any phosphate donor.
After purifying the enzyme, a radioactivity-based assay [Hazra et al., 2010] was adopted to enable the detection of product formation by 2D TLC, since the HPLC column available did not allow the separation of nucleoside phosphate forms. This method has the advantage of allowing the fast qualitative characterization of substrates of DCK, with the disadvantage of not being able to measure enzyme rates. Monophosphate production needs to be coupled to pyruvate kinase and lactate dehydrogenase reactions, resulting in the oxidation of NADH, to be able it to be monitored spectrophotometrically in real time [Agarwal et al., 1978]. Otherwise, mass spectrometry could be employed to quantify products as they were formed, thus allowing to additionally calculate live production rates.

Since we wanted to assess whether DCK could use 5hmdC, and the newly discovered forms of cytidine 5fdC and 5cadC as substrates, the radioactivity assay was suitable. 2D TLC allowed the separation of different phosphate forms (Fig. 4.4).

DCK was able to produce monophosphates from all of the cytidine variants assayed except for 5cadC. To try to qualitatively assess substrate preferences, DCK was incubated with all of the substrates in one reaction. Only dCMP and, in minor quantities, 5mdCMP could be obtained (Fig. 4.5). These data are in agreement with the current literature, indicating that DCK displays lower activity on cytosine when modified at the fifth carbon position [Eriksson et al., 1991].

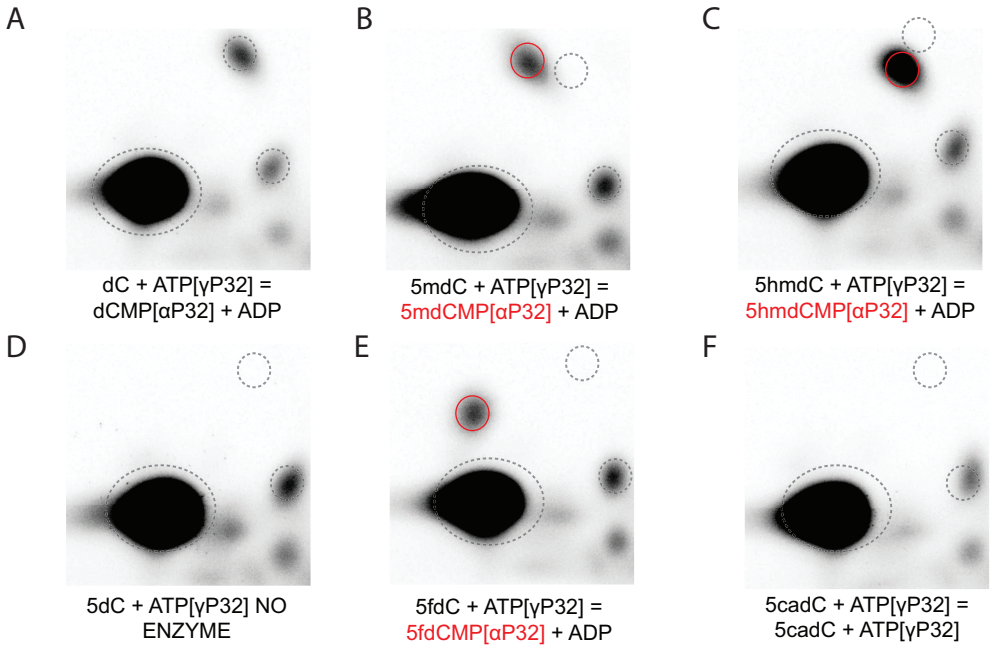


Figure 4.4: 2D TLC showing the DCK assay. A. dC conversion. B. 5mdC conversion. C.5hmdC conversion D. Background (no enzyme) E. 5fdC conversion F. 5cadC reaction. Black circles indicate reference points, including dCMP (top). Red circles indicate the product formed by each reaction (n =3).

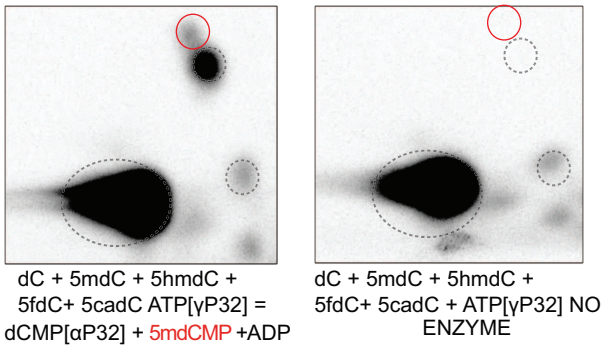


Figure 4.5: 2D TLC showing a DCK substrate competition assay. dC, 5mdC, 5hmdC, 5fdC and 5cadC are substrates. On the right is shown the relative no enzyme control. Black circles indicate reference points, including dCMP (top). Red circles show 5mdCMP (n =3).

4.1.0.3 Assessing the possibility of diphosphate production

In order to obtain triphosphates, a further step is necessary: diphosphate production. Subsequent to monophosphate production, diphosphates can be formed by CMPK enzymes [Maness and Orengo, 1975], which are able to catalyze the following reaction:

$$\text{(d)CMP} + \text{ATP} \rightleftharpoons \text{(d)CDP} + \text{ADP}$$

There are two variants of CMPK: one is cytosolic (CMPK1) [Van Rompay et al., 1999b] while the other resides in the mitochondria (CMPK2) [Xu et al., 2008]. CMPK1 was selected for further analysis, since we were interested in the cytosolic salvage of nucleotides, which is the main source for nucleotides for the DNA replication machinery [Pontarin et al., 2003]. After purification of the human enzyme from *E. coli*, its functionality was tested on cytosine derivatives (5hmdC, 5fdC and 5cadC) in a coupled assay with DCK, needed to obtain the eventual corresponding monophosphates. A radioactivity method was employed followed by 1D TLC, which allowed the visualization of multiple reactions on the same TLC plate, thus enabling the direct comparison of products (Fig. 4.6).

As indicated by the results, the enzyme CMPK1 was not able to produce diphosphates from cytosine modified at the fifth position, such as 5mdC, 5hmdC and 5fdC, while producing it from dCMP. Therefore, it can be concluded that 5hmdCTP and 5fdCTP can

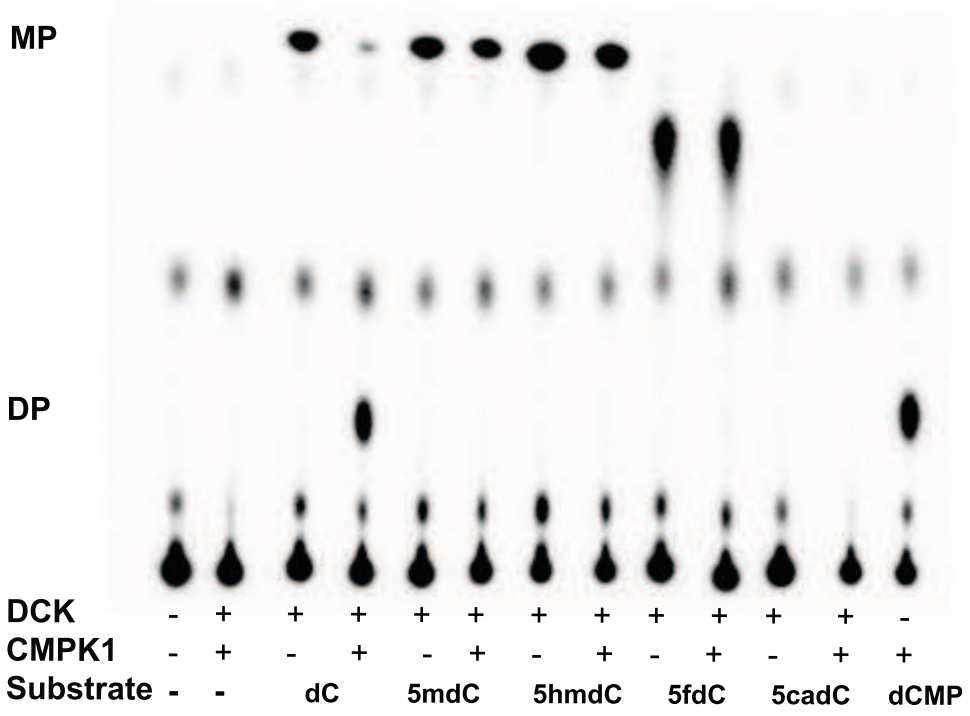


Figure 4.6: 1D TLC showing coupled DCK/CMPK1 assay. +/- indicate the presence/absence of the enzyme or substrate. MP indicates the migration level of monophosphates and DP the migration level of diphosphates (n =3).

not be obtained from 5hmdC.

4.1.0.4 Brief summary

The nucleotide salvage pathway is an important process by which cells recycle nucleosides and bases, to avoid the energy expenditure required for the *de novo* production of nucleotides. Studies suggest that the energy balance that governs protein and nucleotide biosynthesis can drive the evolution of gene expression [Wagner, 2005]. Moreover, the metabolic homeostasis between protein synthesis and nucleotide production can have an important role in the development and growth of tumors. For example, Myc driven lym-

phomas are dependent on PRPS2, a central player of the nucleotide synthesis pathway that is able to produce 5-phosphoribosyl-1-pyrophosphate (PRPP), the first step for *de novo* nucleotide biosynthesis and the purine salvage pathway [Cunningham et al., 2014]. In light of these hypotheses, any uncontrolled incorporation of crucial regulators of gene expression into the DNA, like modified forms of cytosines, could have devastating effects on the cell. To regulate this process, we and others [Vilpo and Vilpo, 1991] tested whether bases like 5mdC, 5hmdC, 5fdC and 5cadC could not be incorporated into the DNA of actively replicating cells, because of the inability of the salvage pathway to produce triphosphates from them.

The results presented here support this hypothesis. First, the capacity of the cellular polymerases to function as a barrier for the incorporation of these nucleotides was evaluated. Others have already shown that, when 5fdC and 5cadC are present in plasmid DNA transfected into mammalian cells, they can reduce the amount of its replication by 30%, while for 5hmdC there was no difference [Ji et al., 2014]. Additional lines of evidence have shown that 5fdC and 5cadC can reduce the rate of both mammalian and yeast RNA PolII elongation by 3-fold, when a single modification was present on the template DNA that was transcribed [Kellinger et al., 2012]. To provide additional insights into the regulation of these nucleotides in the cellular landscape, we decided to test whether 5hmdCTP could be incorporated into the replicating DNA. All of the systems that were employed to test this hypothesis - *in vitro* replication assay, nucleoporation and transfection of triphosphates - demonstrated the incorporation of 5hmdCTP in the DNA (Fig. 4.1, 4.2, 4.3). Consequently, it could be concluded that the polymerase in itself was not able to discriminate between dC and 5hmdCTP. The possibility that the rate of incorporation of 5hmdCTP is slower than that of dCTP cannot be excluded, as our experimental assay did not allow us to monitor the kinetics of incorporation. Further studies will be necessary to validate this finding, for example by employing a purified replication machinery or by repeating IVRA with a definite number of S-phase synchronized replicating cells

and labeled or radioactive 5hmdCTP, the levels of which can be quantified and discriminated from the endogenous form. Having established that the replication machinery is not able to exclude 5hmdCTP incorporation from the DNA, the possibility of whether nucleotides can be produced from 5mdC, 5hmdC, 5fdC and 5cadC by enzymes of the salvage pathway was assessed. DCK was the first enzyme examined, as it is known to generate monophosphate from dC. After purifying deoxycytidine kinase, it was verified that monophosphates were produced from all the nucleosides tested, except for 5cadC, for which a monophosphate was never observed (Fig. 4.4). The next step was to test for the production of the respective diphosphates (5mdCDP, 5hmdCDP and 5fdCDP) from 5mdCMP, 5hmdCMP and 5fdCMP. CMPK1 is the enzyme responsible of the production of cytidine diphosphates in the cytoplasm, where the majority of the nucleotides needed for replication are synthesized. Using this enzyme, diphosphates formation was not detected for 5hmdC and 5fdC, and the data obtained previously by Vilpo et al. [Vilpo and Vilpo, 1991], concerning the absence of 5mdCDP formation in competent HL-60 extracts, were confirmed. Therefore, triphosphate formation from cytosines modified at the fifth position via the cytosine salvage pathway was excluded. Consequently, their presence in DNA can only result from the enzymatic conversion of dC to 5mdC by DNMT enzymes and subsequent conversions of 5mdC to 5hmdC, 5fdC and 5cadC on polymerized chains, such as DNA, by the TET family of enzymes. It has been shown that DNMT accommodates and adapts to interact with the unmethylated DNA [Song et al., 2012], for which it possesses a DNA CXXC binding domain that mediates its recognition [Lee et al., 2001], making the conversion of single nucleosides unlikely. Moreover, 5mdC-treated cells incorporate dT [Jekunen et al., 1983], therefore cells are not able to produce triphosphates from 5mdC, limiting the presence of this base to the DNA chain only. For the TET family of enzymes, both their substrate dependency for 5mdC and the presence (for TET1 and TET3 [Iyer et al., 2009]) or the interaction (for TET2 [Ko et al., 2013]) with the CXXC domain, make the probability of these enzymes acting on 5mdC very low.

Chapter 5

5hmdC and 5fdC can impair cancer cell growth

A variety of nucleoside analogues and nucleobases are currently approved for the treatment of different kinds of cancer. The majority inhibit DNA replication, while others target epigenetic functions. For example, among the purine analogues there are: 2-Chlorodeoxyadenosine (cladribine) and 2-Fluororiboxyadenosine-monophosphate (fludarabine), used in low grade blood malignancies; 2-Chloro-arabinofluoro-2-deoxyadenosine (clofarabine), 2-amino-9-β-D arabinofuranosyl-6-methoxy-9H-purine (nelarabine) and 2-deoxycoformycin approved for different kinds of leukemia. Among the pyrimidine analogues, find treatment use: cytosine β-D-arabinofuranoside (cytarabine) in leukemia; 2,2-Difluorodeoxycytidine (gemcitabine) in different solid tumors; fluorouracil and 5-Deoxy-5-fluoro-N-[(pentyloxy)carbonyl]cytidine (capecitabine) in colorectal and breast cancer; 5-Fluoro-2-deoxyuridine in colon and kidney cancer; and 5-aza-2-cytidine (5azaC) and 5-aza-2-deoxycytidine (5azadC) in blood cancers. They exploit the same metabolic pathways as endogenous nucleosides or nucleotides, sometimes acting as antimetabolites and interfering with normal nucleotide metabolism, such as in the case of 5-Fluorouracil, which

is classified as a pyrimidine antagonist. Nucleoside and nucleotide analogs enter the cell via specific nucleoside transporters, through both passive and active mechanisms, in the latter case via processes coupled to sodium pumps. Inside the cells, the drugs depend upon nucleotide salvage pathway enzymes, such as nucleoside kinases like deoxycytidine kinase (DCK), for their phosphorylation. Furthermore, enzymes such as cytidine deaminase (CDA), can process pyrimidine based compounds rendering them inactive. The resistance of some cancer cells to the effects of nucleoside analogues is thought to be largely due to somatic changes in the tumor cells. In cancer cells, a deficiency in nucleoside transporters such as equilibrative nucleoside transporter 1 (ENT1) and intracellular nucleoside kinases such as DCK, or increased expression of CDA, are all correlated with reduced cytotoxicity of nucleoside analogues observed both *in vitro* and in the clinic [Jordheim and Dumontet, 2007]. Most of these mechanisms limit the time frame in which nucleoside analogs can be effectively used. Therefore, there is an unmet need for new drugs that have new mechanisms of action that can overcome, or exploit, resistance mechanisms as second line treatments (reviewed in[Jordheim et al., 2013]).

The earliest indications of a link between DNA methylation and cancer were derived from gene expression studies. In particular, it was found that some oncosuppressors, like p16, displayed abnormal promoter silencing caused by DNA methylation [Gonzalez-Zulueta et al., 1995]. Therapies inhibiting DNA methylation found a rationale for employment in cancer and are currently FDA approved. 5azaC and its deoxyribose analogue 5azadC, have been studied for their ability to inhibit DNA methylating enzymes (DNMT) [Taylor and Jones, 1982], treatment with them resulting in DNA demethylation [Jones and Taylor, 1980]. Resistance to these nucleoside analogs is seen in the clinic, with increasing evidences pointing towards the overexpression of CDA [Mercier et al., 2013], which would convert 5azaC in its inactive analogue 5-azauridine (5azaU), preventing it from mediating the inhibition of DNMTs. Another well known drawback of the treatment with 5azaC, is its chemical instability when in solution [Lin et al., 1981]. The increased rate of resistance

mechanisms and the poor bio-availability of the drug, make research for new nucleoside analogues that act on DNA methylation a high priority.

In 2009 5hmC had just been published as a base present in mammalian DNA [Kriaucionis and Heintz, 2009][Tahiliani et al., 2009] and there were numerous suggestions about its potential role as an intermediate of DNA demethylation [Tahiliani et al., 2009][Hajkova et al., 2010][Guo et al., 2011]. Evidences toward this direction have shown that, when TET enzymes are overexpressed, a global DNA demethylation occurs [Tahiliani et al., 2009]. As global DNA demethylation is also a consequence of 5azaC and 5azadC treatment [Jones and Taylor, 1980], it was speculated that 5hmdC could be incorporated into DNA and induce DNA demethylation. In order to investigate its suitability as a demethylating compound, it was first necessary to test the stability of 5hmdC compared to that of 5azadC, and later its biological effects on the proliferation of cancer cell lines. 5fdC and 5cadC were published at a later date, thus were included at a later point in the study.

5.1 Results

5.1.1 Assessment of the stability of the molecules

5-aza-deoxycytidine (5azadC), a cytosine analogue approved for the treatment of myelodysplastic syndrome, is very unstable at alkaline pH at 37°C, being almost fully degraded after 24 hours. The hydrolysis of 5-azacytidine in water and at alkaline pH progresses in few steps: first, with the opening of the triazine ring between C-6 and N-1, followed by loss of the formyl group that results in 1-β-D-ribofuranosyl-3-guanylurea [Lin et al., 1981]. Furthermore, the measured half-life of 5azadC in the blood is 20 minutes, due to clearance by liver and spleen cytidine deaminases [Karahoca and Momparler, 2013]. We therefore set out to compare the stability of dC, 5hmdC, 5fdC and 5cadC with that of 5azadC in water and in tissue culture medium (DMEM), through HPLC measurements, at 37°C

over a period of 10 days. This established the treatment conditions for the subsequent cell culture experiment.

The chromatograms for days 0, 1 and 10 are shown in the Fig. 5.1. The results obtained

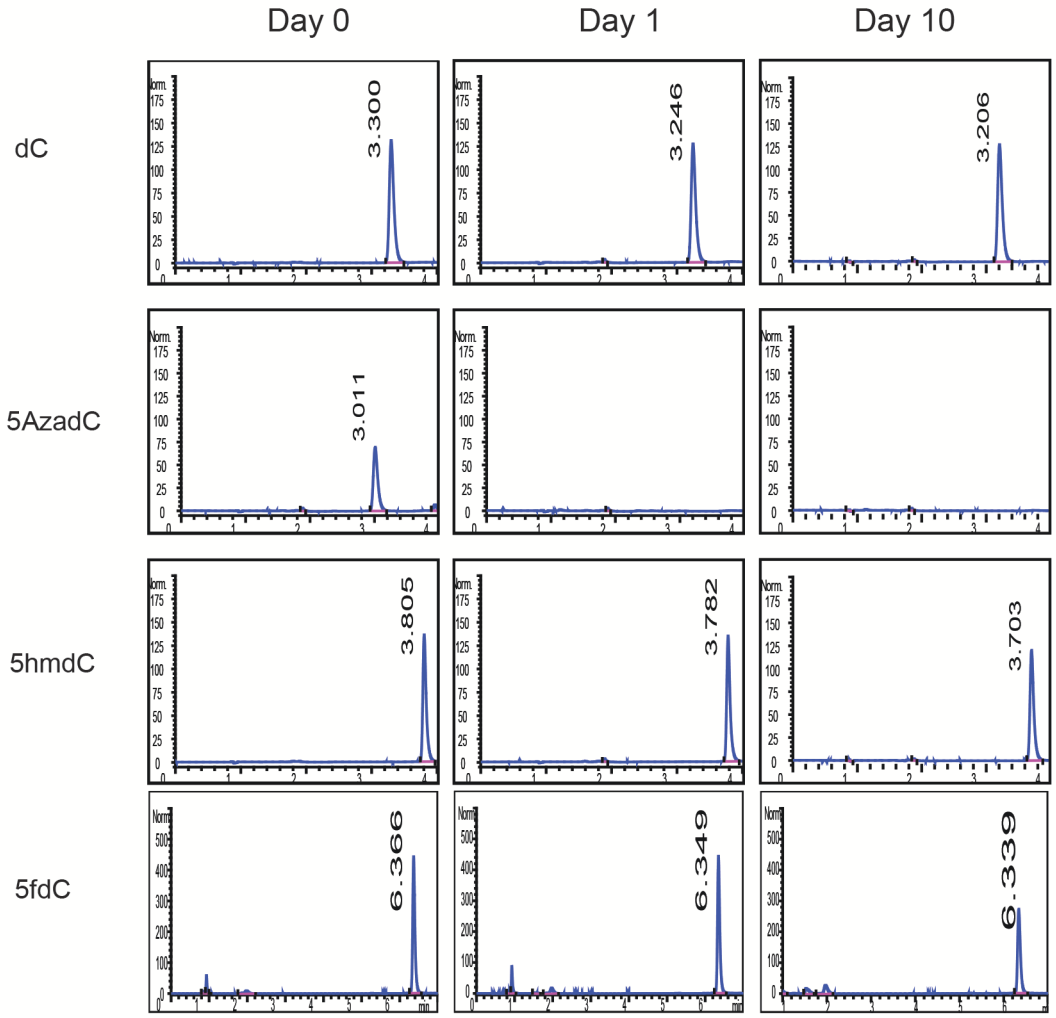


Figure 5.1: HPLC chromatograms highlighting the behavior of nucleosides in water at days 0, 1 and 10 at 37°C. Norm. absorbance units; Min. minutes. Representative picture (n =3).

showed that, except from 5fdC, the molecules were stable at the conditions assessed when compared to 5azadC, of which only 30% of the initial amount was left after 24 hours (Fig. 5.2). After 10 days, a 30% decrease of 5fdC was detected compared with day 0.

The results were reproduced for DMEM (Fig. 5.3), except for 5fdC which looked stable

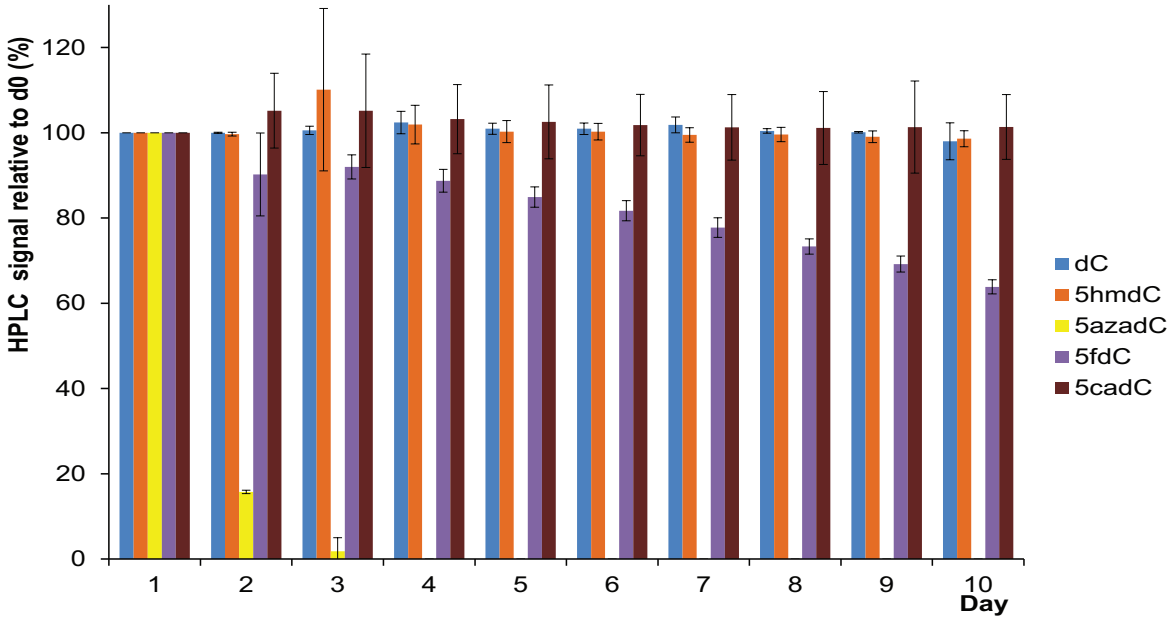


Figure 5.2: Stability of nucleosides in water at 37°C. Quantification of HPLC signal (Peak area) relative to day 0 (n=3. Whiskers indicate standard deviation).

over the period assessed. This might be due to a dependence of 5fdC on pH. DMEM is mildly alkaline (pH 7.4 which tends to increase upon exposure to carbon dioxide), compared to the HPLC grade water used in this analysis which is at pH 7.

With this study, it was established that the compounds could be tested in a tissue culture setting without them having to be replaced every day.

5.1.2 5hmdC inhibits cell growth in a subset of cancer cell lines tested

Initially, five cancer cell lines with different origins were tested that were readily available in the laboratory: LN18, Colo320, H1299, MDA-MB-231 and A375. Their characteristics are outlined in Table 5.1.

We sought to assess the biological effect of 5hmdC treatment over a different range of concentrations (0.1, 1 and 10 μM) that were determined based on the published literature on 5azadC treatment in the NCI60 cancer cell lines [Holbeck et al., 2010] where 50% growth

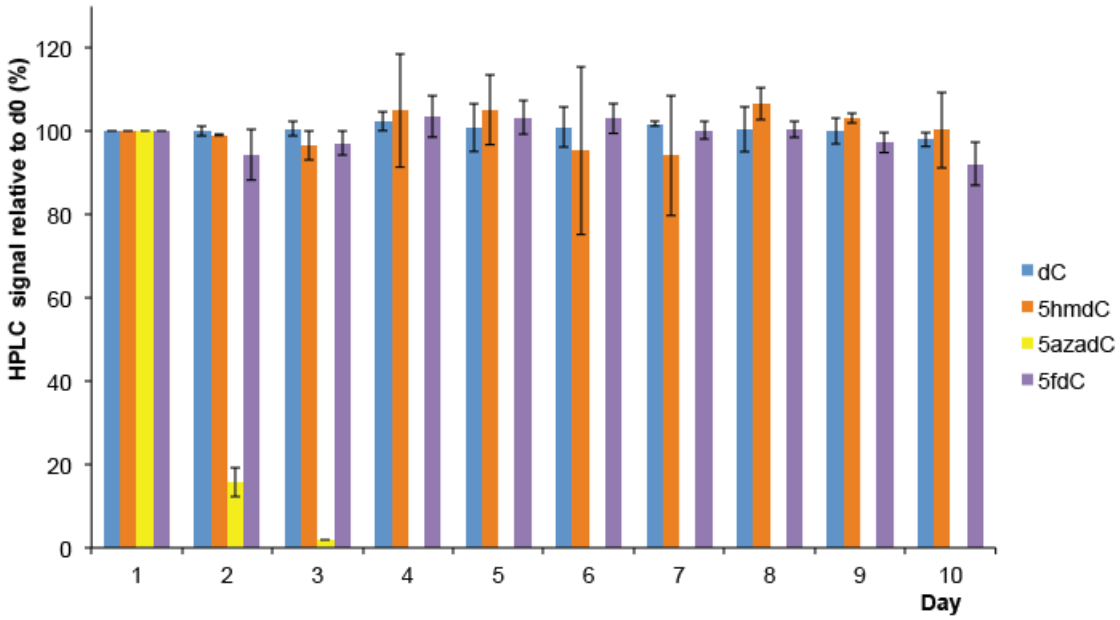


Figure 5.3: Stability of nucleosides in DMEM at 37°C. Quantification of HPLC signal (Peak area) relative to day 0 (n=3. Whiskers indicate standard deviation).

inhibition (GI50) over a 2 day period was observed with 100 μM. Growth was monitored over a 10 days period by measuring the growth rate of the 5hmdC-treated cells, compared to the ones treated with deoxycytidine (dC) to act as a negative control, as dC is the unmodified version of 5hmdC and the nucleotide that would be substituted if 5hmdC was incorporated into the DNA. A 10 days period was adopted to allow for replication and eventual DNA demethylation to occur. 5azadC was the positive control, as it is known to induce cell death via a DNA methylation targeting mechanism [Bender et al., 1998]. Compounds were replaced at day 1, 3, 5, 7, 9 and 10, on average every 48 h, as they had been assessed as being stable over this time frame.

Different behaviors could be observed (Fig. 5.4) in the cell line tested. Significant growth inhibition was observed in MDA-MB-231 and A375 cells. In particular, cells treated with 10 μM showed the most pronounced effects with death of MDA-MB-231 and an approximate 50% growth inhibition of A375 after 10 days. No significant growth inhibition

Cell line	Cancer origin	p53 status	p53 mutation	Doubling time (h)	Collection
Capan-2	Pancreatic adenocarcinoma	WT	n.a.	34	CCLE
BT-549	Breast ductal carcinoma	MUT	R249S	53.9	NCI60
MDA-MB-231	Breast adenocarcinoma	MUT	R280L	41.9	NCI60
MCF7	Breast adenocarcinoma	WT	n.a.	25.4	NCI60
HS578T	Breast carcinoma	MUT	V157F	53.8	NCI60
HCT-116	Colorectal carcinoma	WT	n.a.	17.4	NCI60
HCC-2998	Colon carcinoma	MUT	R213X	31.5	NCI60
SW-620	Colorectal carcinoma	MUT	R273H	20.4	NCI60
Colo-320	Colorectal carcinoma	MUT	R248W	24	CCLE
DU-145	Prostate adenocarcinoma	MUT	P223L + V274F	32.3	NCI60
PC-3	Prostate adenocarcinoma	MUT	A138X	27.1	NCI60
SN12C	Renal cell carcinoma	MUT	E336X	29.5	NCI60
MDA-MB-435	Melanoma	MUT	G266E	25.8	NCI60
A375	Melanoma	WT	n.a.	16	CCLE
HOP-92	Lung carcinoma	MUT	R175L	79.5	NCI60
NCI-H522	Lung adenocarcinoma	MUT	p.P191fs*57	38.2	NCI60
NCI-H1299	Lung carcinoma	NULL	DEL	25	CCLE
OVCAR-5	Ovarian carcinoma	WT	n.a.	48.8	NCI60
LN18	Glioblastoma	MUT	C238S	72	CCLE
BL-70	Burkitt lymphoma	MUT	R273C	42	CCLE
THP-1	Myeloid leukemia	DEL	DEL26A	42	CCLE

Table 5.1: Characteristics of the cell lines studied. p53 status is derived from the IARC p53 database [Petitjean et al., 2007]. Doubling time from the DTP-NIH screen for NCI-60 panel cell lines; Capan-2 [Sipos et al., 2003], LN-18 [Diserens et al., 1981], H1299 [Suzuki, 2004], BL-70 [Baran-Marszak et al., 2002], A375 [Benga, 2001], THP-1 [Tsuchiya et al., 1980], Colo-320 [Ahmed et al., 2013]. Collection: NCI60 [Shoemaker et al., 1988],CCLE [Barretina et al., 2012].

was observed for Colo320, H1299 and LN18 cells. A375 and MDA-MB-231 showed a dose dependent response, with MDA-MB-231 showing the most pronounced inhibition among the cells analyzed. 5azadC induced cell death at both 10 and 1 μM for all the cell lines tested, and approximately 50% growth inhibition after 10 days with 0.1 μM treatment. At this point, it was decided that a dosage of 10 μM would be used for the rest of the experiments, as this concentration had shown the greater efficacy in the first experiment for all the compounds tested.

5azadC is FDA approved for the treatment of acute myeloid leukemia and has shown greater success in blood tumors, compared to solid cancers. Therefore, 5hmdC was tested in a few cell lines of blood cancer origins, to understand whether it would show greater

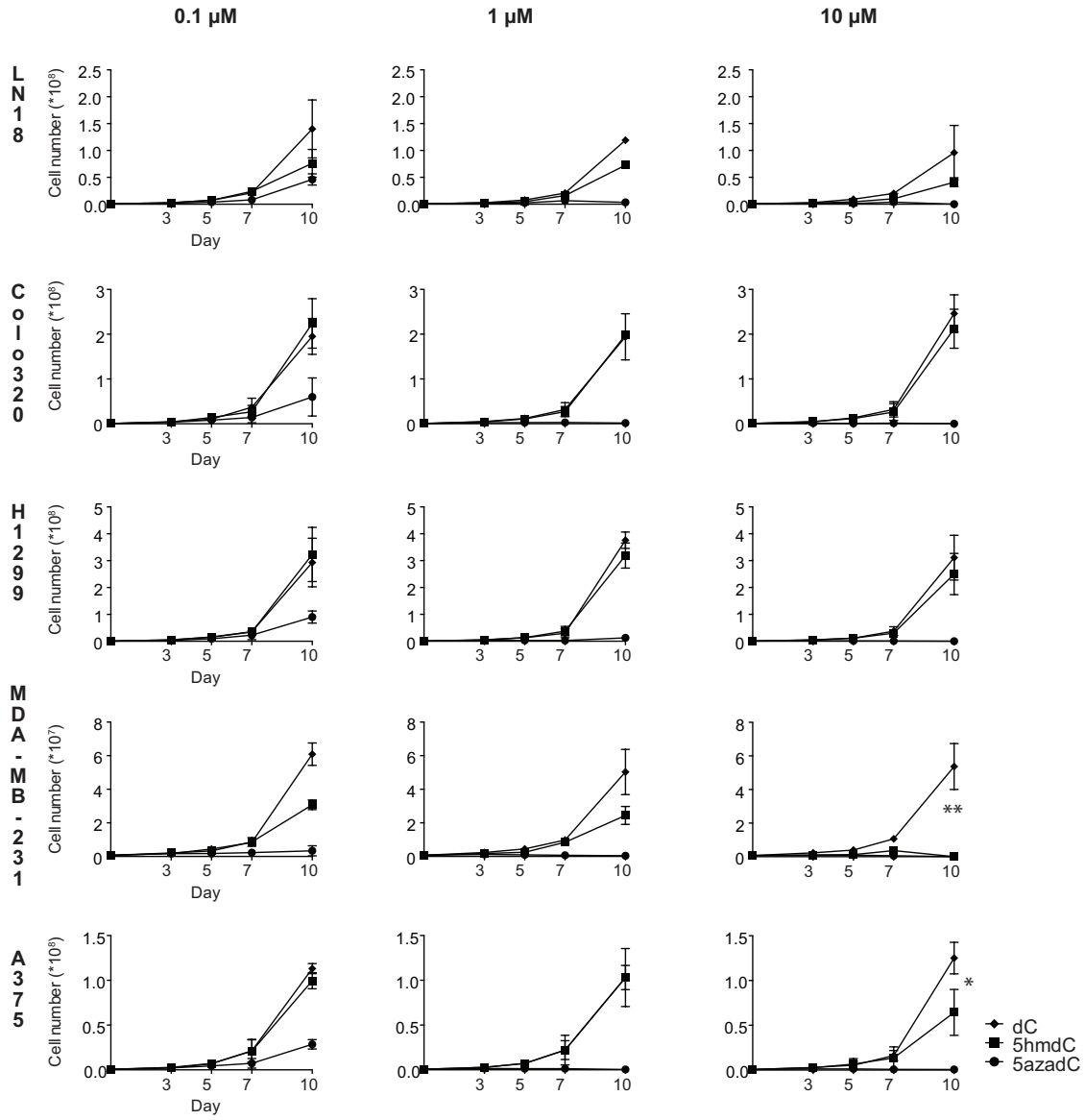


Figure 5.4: Growth rate of different cell lines at 0.1, 1 and 10 μM of dC, 5hmdC and 5azadC. 2-way ANOVA with Holm Sidak correction p (MDA-MB-231)=0.0016; p (A375) =0.0354. (n =3. Whiskers indicate standard deviation).

efficacy in this setting. BL-70 and THP-1 were selected due to their availability in the laboratory (Table 5.1).

Growth inhibition was not observed in these two blood cancer cell lines (Fig. 5.5). However, having found a cell line that had its growth inhibited by the treatment (MDA-

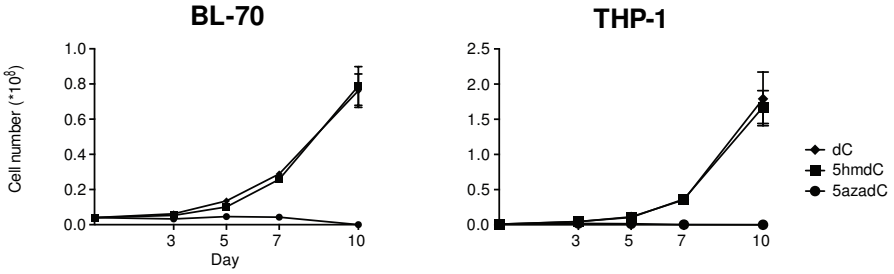


Figure 5.5: Growth rate at 10 μM of dC, 5hmdC and 5azadC of blood cancer cell lines (n =3. Whiskers indicate standard deviation).

MB-231), and even more importantly, belonging to the well characterized NCI60 panel of cancer cell lines [Scherf et al., 2000], the analysis was extended to more cancer cell lines belonging to the NCI60 panel. The NCI60 panel is a selection of 60 representative cancer cell lines that have been characterized in terms of their gene expression [Pfister et al., 2009], response to a wide variety of compounds, and methylation levels of promoters that are associated with CpG islands of genes that are relevant for cancer progression and development [Shen et al., 2007]. This would permit an eventual gene expression analysis with the available datasets, as a mean to identify the mechanism of response to 5hmdC treatment. Therefore, cell lines in the NCI60 panel that are related to MDA-MB-231 were selected for having common:

1. tumor subtype (breast cancer);
2. CpG islands methylation profile [Shen et al., 2007];
3. gene expression profile [Scherf et al., 2000];
4. drug response profile [Scherf et al., 2000].

Breast cancer cell lines

The following breast cancer cell lines were included: MCF7 and BT-549 (Table 5.1). Upon treatment, a significant reduction of approximately 40% was observed in the growth

of these cell lines after 10 days, indicating a slow sensitization of these cells to the action of the compound (Fig. 5.6). MCF7 cells did not show any proliferation impairment over the course of the treatment. Therefore, it was concluded that cancer type was not the main factor involved in the sensitization of MDA-MB-231 upon 5hmdC treatment.

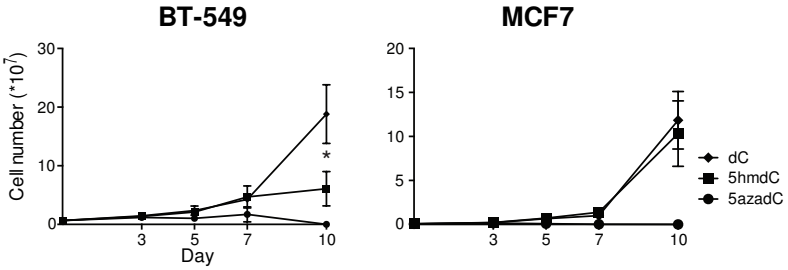


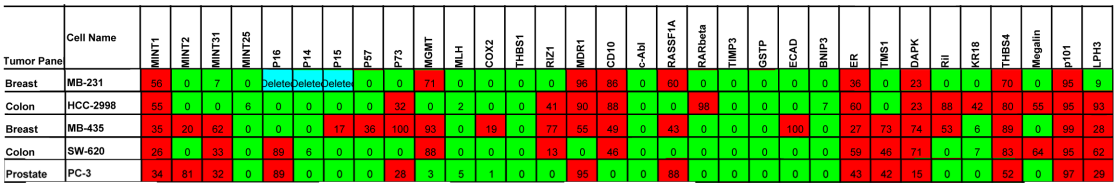
Figure 5.6: Growth rate of breast cancer cell lines at 10 μM of dC, 5hmdC and 5azadC . 2-way ANOVA with Holm Sidak correction, p= 0.0140 (n =3. Whiskers indicate standard deviation).

Cell lines with a common methylation profile on promoters-associated CpG islands

Next, given the suggested role of 5hmC in DNA demethylation [Tahiliani et al., 2009][Hajkova et al., 2010][Guo et al., 2011], it was hypothesized that the response was mediated by the reactivation of gene expression upon global DNA demethylation, as previously shown for 5azadC [Bender et al., 1998]. Therefore, for the next screen, cell lines were selected that displayed a similar promoter methylation status to that of MDA-MB-231 (Fig. 5.7): MDA-MB-435, HCC-2998, SW-620 and PC-3 (Table 5.1). In this reference study, the CpG islands of the promoters of 32 genes, which take part in different cellular processes, were bisulfite treated to understand their methylation status. In the group of cell lines selected, the cell cycle regulatory genes (*p16INK4a, p14ARF, p15INK4b, p57KIP*) and angiogenesis-related genes (*THBS1, THBS4, TIMP3, E-cadherin, DAPK*) were mostly unmethylated, while the genes mediating apoptosis (*TMS1, RIL, p73, BNIP3*) were mostly methylated. Other genes, including DNA repair genes (*MGMT, hMLH1*), drug

metabolism genes (*GSTP1, MDR1*), signal transduction genes (*ERa, RARb2, COX2, cABL, RASSF1A, p101, MINT31, MINT25*), transcription regulator genes (*RIZ1, KR18*), and others (*CD10, LPH3, Megalin, MINT1*) displayed more variable patterns [Shen et al., 2007] (Fig. 5.7).

Significant inhibition of growth after 10 days of treatment was detected in only in 2 can-



Tumor Pane	Cell Name	MINT1	MINT2	MINT31	MINT25	P16	P14	P15	P57	P73	MGMT	MLH	COX2	THBS1	RIZ1	MDR1	CD10	c-Abl	RASSF1A	RARbeta	TIMP3	GSTP	ECAD	BNIP3	ER	TMS1	DAPK	Ril	KR18	THBS4	Megalin	p101	LPH3
Breast	MB-231	56	0	7	0	Delete	Delete	Delete	0	0	71	0	0	0	0	96	86	0	60	0	0	0	0	0	36	0	23	0	0	70	0	95	9
Colon	HCC-2998	55	0	0	6	0	0	0	0	32	0	2	0	0	41	90	88	0	0	98	0	0	0	7	60	0	23	88	42	80	55	95	93
Breast	MB-435	35	20	62	0	0	0	17	36	100	93	0	19	0	77	55	49	0	43	0	0	0	100	0	27	73	74	53	6	89	0	99	28
Colon	SW-620	26	0	33	0	89	6	0	0	0	88	0	0	0	13	0	46	0	0	0	0	0	0	0	59	46	71	0	7	83	64	95	62
Prostate	PC-3	34	81	32	0	89	0	0	0	28	3	5	1	0	0	95	0	0	88	0	0	0	0	0	43	42	15	0	0	52	0	97	29

Figure 5.7: DNA methylation of 32 promoter associated CpG islands in selected cell lines of the NCI60 panel. Adapted from [Shen et al., 2007]. Red indicates methylation.

cer cell lines: PC-3 (60% inhibition) and HCC-2998 (50% inhibition) (Fig.5.8). These had common DNA methylation profiles on the CpG islands of promoters associated with genes involved in angiogenesis, drug metabolism and signal transduction. This might suggest that a gene expression mechanism is involved in the response to 5hmdC treatment, rather than a DNA demethylation-dependent process.

Cell lines with similar gene expression and drug response profiles

Once again, NCI60 panel linked resources like gene expression and drug sensitivity profiles, built upon common signatures of gene expression and response to classes of drugs, were used to examine cell lines with gene expression profiles similar to MDA-MB-231 (Fig. 5.9 A). Therefore, HOP-92 cells were included in the study (Table 5.1), which clustered together with MDA-MB-231. Among the cell lines with a similar drug response profile (Fig. 5.9 B), HS578T were analyzed (Table 5.1). However, since HOP-92 clustered close to MDA-MB-231 not only in the gene expression tree, but also the one related to drug

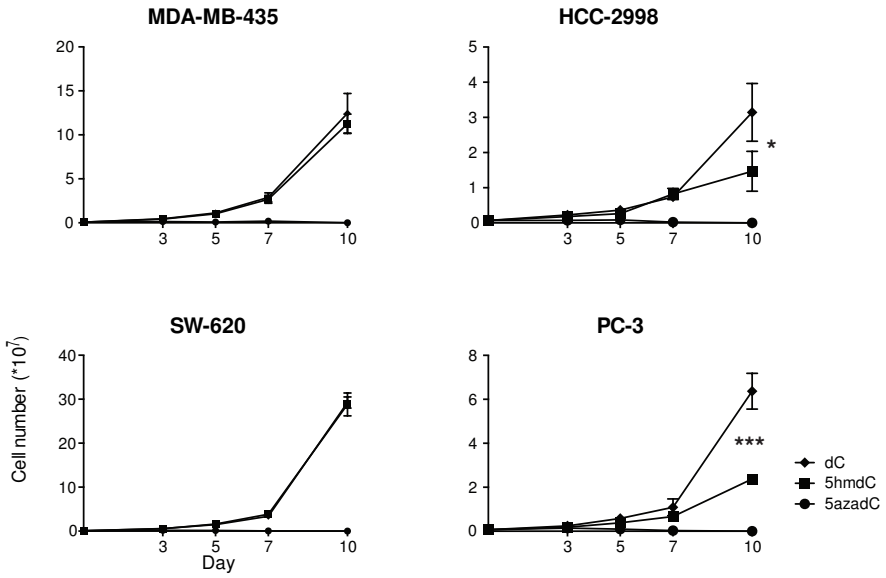


Figure 5.8: Growth rate of cell lines of the NCI60 with similar promoters CpG island methylation profile at 10 μM of dC, 5hmdC and 5azadC . 2-way ANOVA with Holm Sidak correction, p(PC-3) =0.0001; p (HCC-2998) =0.0324 (n =3. Whiskers indicate standard deviation).

response, the OVCAR-5 and DU-145 cell lines (Table 5.1) were included in the study, as they are closer on the gene expression tree than on the drug sensitivity one, to distinguish between the related contributions.

A significant decrease was detected in the number of cells treated with 5hmdC in HOP-92 (100%) and HS578T (50%), but only a minor decrease in DU-145 (20%) (Fig. 5.10). This suggests that the response is very similar among cells selected for common gene expression or for common drug response mechanisms, with a prevalence for genes involved in drug metabolism since the proximity among sensitive cell lines on this cluster tree was greater.

5.1.2.1 Brief summary

At this point the data were summarized by grouping the sensitive and non sensitive cell lines. To do so, the cell number ratio was plotted at day 10 between 5hmdC- and dC-

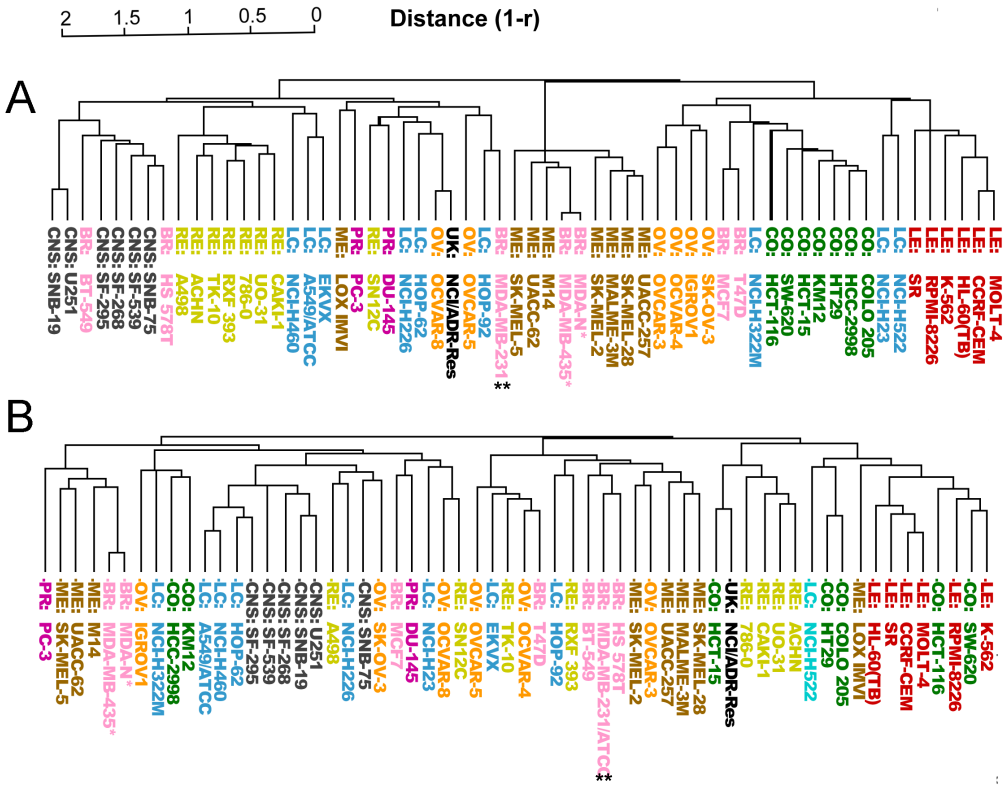


Figure 5.9: Cluster analysis of the NCI60 panel based on: gene expression (A) and drug sensitivity (B). Adapted from [Scherf et al., 2000].

treated cells. Cell lines with a ratio of less than 0.2 were classified as highly sensitive to the treatment, those between 0.2 and 0.8 as partially growth inhibited by the treatment, and those more than 0.8 as insensitive, where a value of 0.2 indicates 80% growth inhibition (Fig. 5.11). According to these criteria, HOP-92 and MDA-MB-231 were classified as sensitive to the growth inhibitory potential of 5hmdC, compared to dC; PC-3, HS578T, BT-549 and A375 were the cell lines that showed partial growth inhibition (e.g included between 80% and 20% growth inhibition), while OVCAR-5, H1299, THP-1, LN-18, Colo320, MDA-MB-435, SW-620, BL-70, DU-145, HCC-2998 and MCF7 were classified as insensitive (displaying less than 20% growth inhibition). p53 status (Table 5.1), which can be involved in the induction of apoptotic processes dependent on its activation, did not seem to be a prognostic factor for sensitivity to 5hmdC treatment.

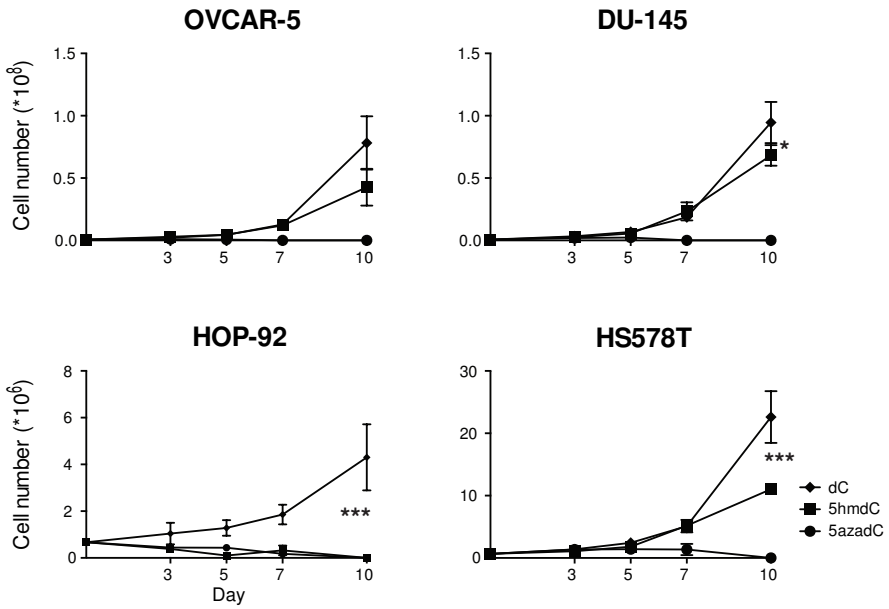


Figure 5.10: Growth rates of cell lines of the NCI60 panel with similar gene expression and drug response profiles. 2-way ANOVA with Holm Sidak correction, p (HOP-92) =0.0006; p (HS578T) =0.001 (n =3. Whiskers indicate standard deviation).

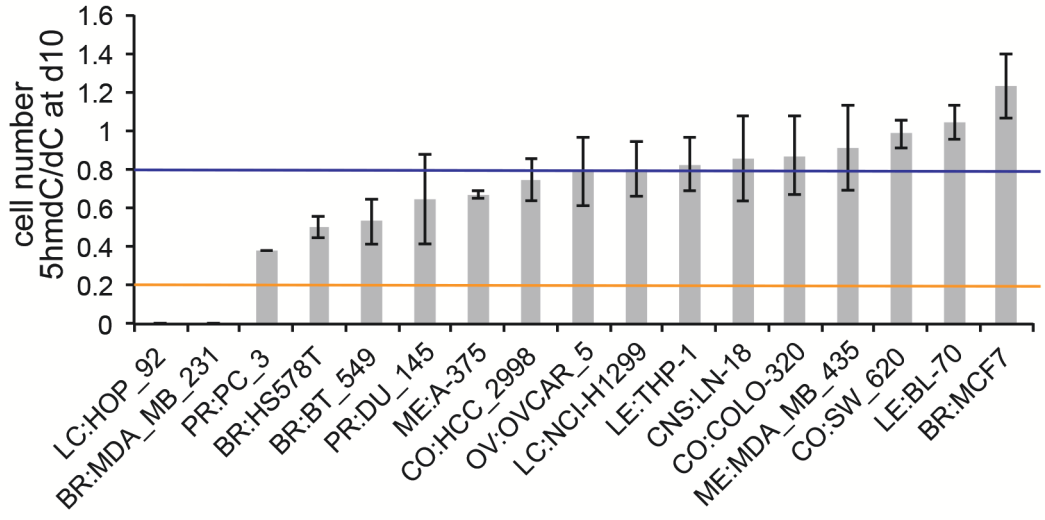


Figure 5.11: Cell number ratio between 5hmdC- and dC-treated cells at day 10. The threshold of 0.2 is shown in orange, the one of 0.8 in blue (n =3. Whiskers indicate standard deviation).

5.1.3 Gene expression data analysis identifies *CDA* as a potential gene involved in the response to 5hmdC treatment

Analysis of gene expression data-sets has had great impact on the drug discovery process [Liu, 2005]. Tools like the NCI60 panel initially, and the CCLE (Cancer Cell Line Encyclopedia) [Barretina et al., 2012] later on, were established to help scientists dissect pathways in a complex system like cancer that, in this way, could be standardized or modeled. Multiple data-sets are now available to the research community, enabling genetic analyses, such as gene expression analysis. It was, therefore, decided that gene expression analysis would be performed on data from the NCI60 panel [Pfister et al., 2009], utilizing the groups of cell previously defined. In particular, gene expression profiles of MDA-MB-231 and HOP-92 were compared to those of MCF7 and MDA-MB-435 to find any similarities between these responding cell lines, and to identify the differences between them and the non-responding group.

The results of this analysis pointed to cytidine deaminase (*CDA*) among the first hun-

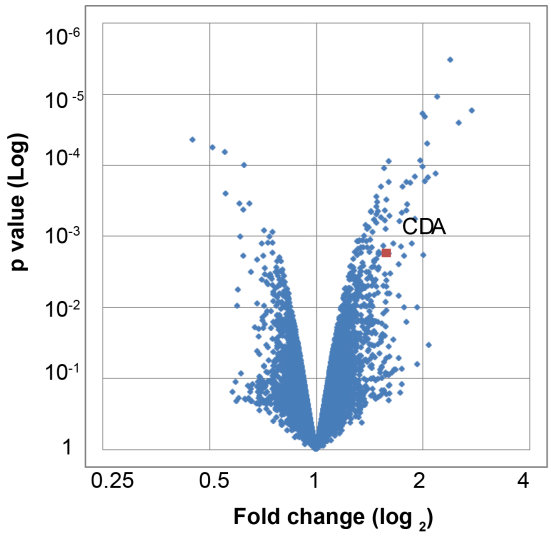


Figure 5.12: Gene expression analysis with NCI60 panel gene expression datasets [Pfister et al., 2009] (GEO2R). MDA-MB-231 and HOP-92 data were compared to those of MCF7 and MDA-MB-435. *CDA* is outlined in orange with a p value =0.001702. (t-test with Benjamini- Hochberg correction).

dred co-expressed genes (rank 79, p value =0.001702), which is involved in nucleotide metabolism and is differentially expressed among the two groups analyzed (Fig. 5.12 and Appendix 1 Table 1).

This suggested a working hypothesis (Fig. 5.13), according to which 5hmdC could be metabolized into the toxic nucleoside 5hmdU, and cause cell death following DNA damage induction once incorporated in the DNA [Kaufman, 1986]. It has been shown that cell death, via mechanisms independent of p53 status, follows 5hmdU incorporation as a result of the formation of double strand DNA breaks (DSB) [Mi, 2001]. Having previously shown that if 5hmdC entered the cell there would be no possibility of it forming the triphosphate 5hmdCTP (see Fig. 4.6), and in light of the cluster analysis results, it was assumed that:

- the route of deamination could be undertaken, which would produce 5hmdU;
- 5hmdU could then be triphosphorylated;
- such 5hmdUTP could be incorporated into the DNA causing DSB and cell death [Kaufman, 1986] [Vilpo and Vilpo, 1988].

5.1.3.1 Further selected cancer cell lines support the involvement of CDA in the response

To further strengthen this hypothesis, *CDA* expression data were extracted from the NCI60 and CCLE for the cell lines employed and for the highest *CDA*-expressing cell lines (Fig. 5.14). Thus, among the available highest expressing cell lines, the following were selected: SN12C, Capan-2, HCT-116 (Table 5.1). In addition, NCI-H522 (Table 5.1), which has comparatively low levels of *CDA*, was added as a negative control.

A significant reduction in growth (Fig. 5.15) was observed in Capan-2 (100%), SN12C (90%) and HCT-116 (20%) cells, but not in NCI-H522.

In order to confirm the expression of CDA in the aforementioned cell lines, western blotting

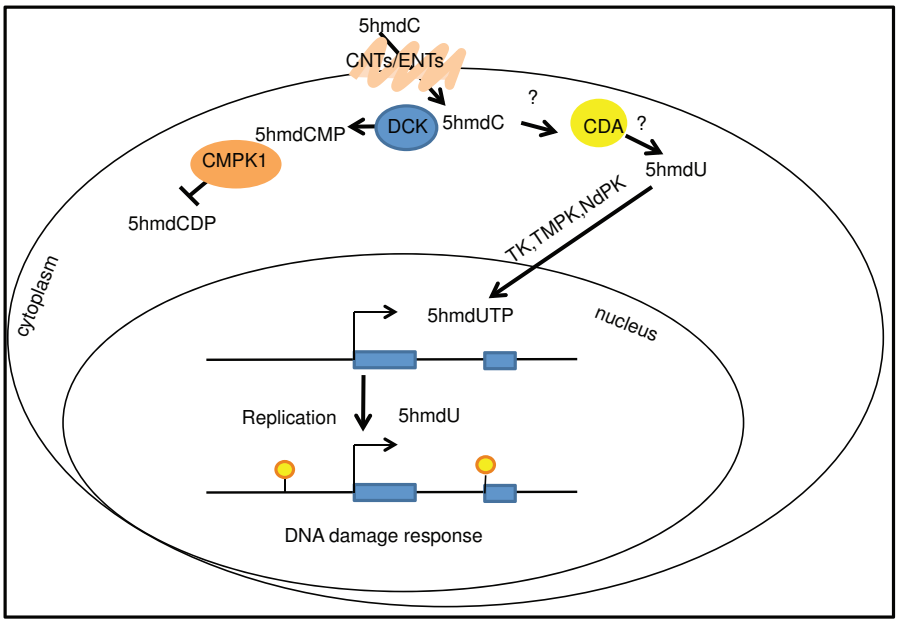


Figure 5.13: Working model.

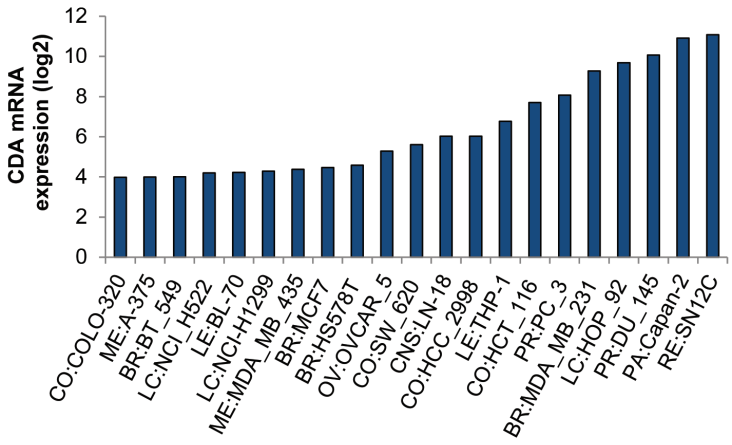


Figure 5.14: CDA expression in the cell line tested (Geo2R processed data from [Barretina et al., 2012] and [Pfister et al., 2009] datasets.)

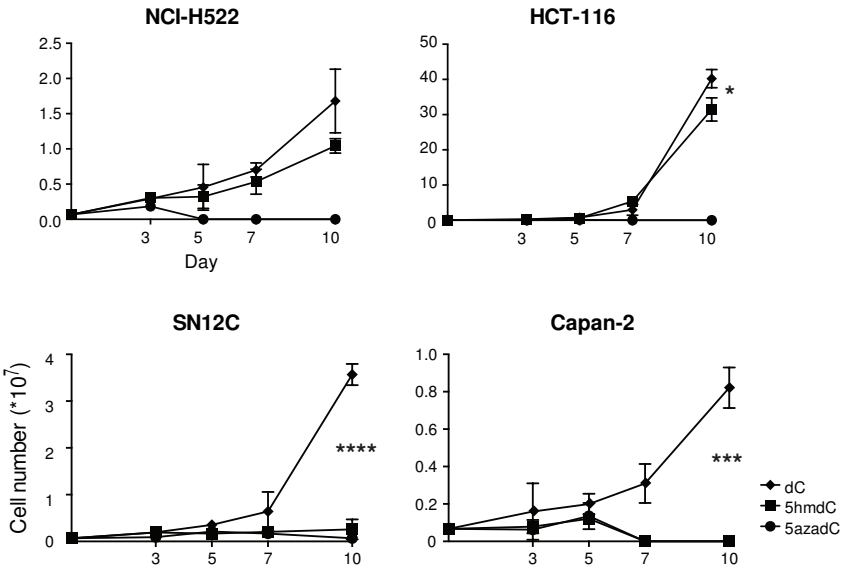


Figure 5.15: Growth rate of cell lines of the NCI60 with similar gene expression profiles. 2-way ANOVA with Holm Sidak correction p(HCT-116) =0.0358; p(SN12C) <0.0001; p(Capan-2) =0.0007 (n =3. Whiskers show standard deviation).

was performed to assess CDA levels. Expression of CDA was detected in MDA-MB-231 and SN12C, but not in H1299 and MCF7 (Fig. 5.16 A). Moreover, high expression of CDA was confirmed in Capan-2 and SN12C cells relative to MCF7 (Fig. 5.16 B). These results confirmed the levels of CDA detected by gene expression analysis.

Subsequently, the last group of cell lines tested was added to the plot in Figure 5.11,

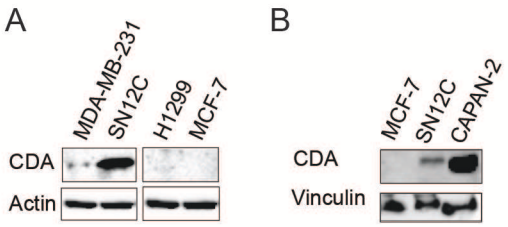


Figure 5.16: Western Blot of CDA levels in some of the analyzed cell lines. Vinculin and actin are the loading controls.

where the levels of *CDA* expression for all the cell lines tested were added to generate a plot that visually shows the correlation between CDA expression and growth inhibition

(Fig. 5.17). On the left, cell lines expressing the highest amount of *CDA* are the ones that show the strongest growth inhibition and *viceversa* for cell lines on the right.

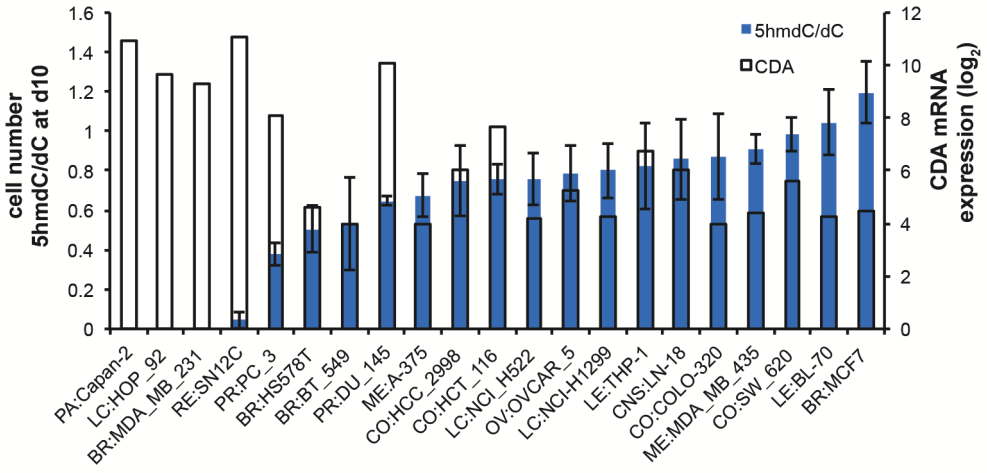


Figure 5.17: Cell number ratio dhmdC/dC relative to *CDA* expression.

5.1.4 Knock down of *CDA* in sensitive cell lines rescues the inhibition of proliferation

In order to confirm the involvement of *CDA* in the response to 5hmdC, MDA-MB-231 and SN12C cells were infected with lentiviruses carrying short interfering RNA (sh-RNA), to stably knockdown the levels of *CDA* and try to rescue the phenotype. The levels of CDA in both the cell lines, could be stably reduced, as validated by western blot and quantitative PCR (qPCR) (Fig. 5.18), with two different hairpins. Therefore, we obtained: MDA-MB-231 and SN12C sh-Luc, sh-CDA_0 and sh-CDA_8. The knockdown obtained with sh-0 was less efficient than sh-8, but all knockdowns reduced the cDNA and protein levels by at least 50%, relative to uninfected or sh-Luc infected cells.

Once the knockdowns were established, they were treated with 10μM 5hmdC. The phenotype was rescued in both cell lines. In particular, knockdown for MDA-MB-231 was

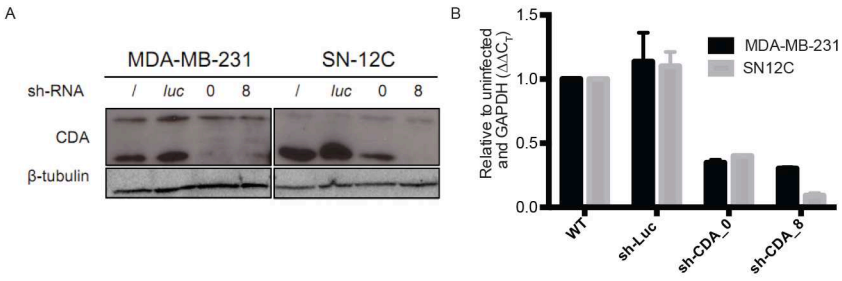


Figure 5.18: Western blot (A) and qPCR (B) of CDA levels in MDA-MB-231 and SN12C. β-tubulin is the loading control. *GAPDH* is the housekeeping gene (n =3. Whiskers show standard deviation).

rescued in the growth by 100% as compared to sh-Luc, when treated with 5hmdC. SN12C sh-0 cells were partially rescued, with only a two-fold growth advantage over the control infected cells (SN12C sh-Luc) and SN12C sh-8 of three-folds (Fig. 5.19). The greater rescue efficiency given by sh-8 can be attributed to it having the greatest knockdown efficiency.

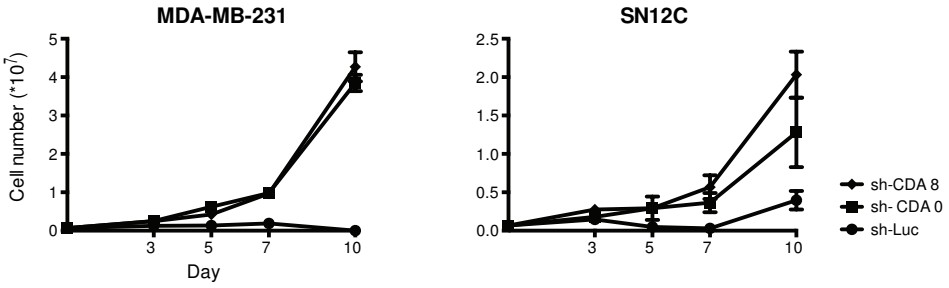


Figure 5.19: Treatment of MDA-MB-231 and SN12C knockdowns with 10μM 5hmdC (n =3. Whiskers show standard deviation).

5.1.5 Overexpression of *CDA* inhibits the growth of non-responding cell lines

To further support our hypothesis, regarding the involvement of *CDA* in the cellular response to 5hmdC, MCF7 and H1299 cells, which are not sensitive to 5hmdC treatment (Fig. 5.11), were infected with lentiviruses to stably overexpress *CDA* to sensitize them to

the treatment. CDA was stably overexpressed in both the cell lines, as shown in Figure 5.20 by western blot. Therefore MCF7dsRed_CDA and H1299dsRed_CDA were obtained. In the MCF7dsRed_CDA lane there was an additional band on the western blot, which was speculated to be due to a phosphorylated form of CDA that could be predicted at Y79 (Phosphosite extracted information) [Li et al., 2002].

Once the cell lines were established, they were treated with 10 μM 5hmdC. We were

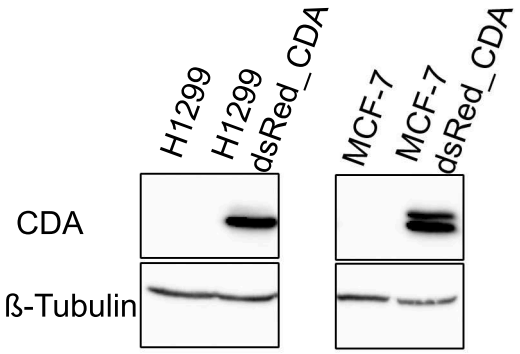


Figure 5.20: Western blot of *CDA* levels in H1299 and MCF7 cells. β-tubulin was used as a loading control.

able to fully inhibit the growth of MCF7-overexpressing cells and only partially the one of H1299-overexpressing cells (Figure 5.21). This might be due to the CDA status of these cell lines. The putative phosphorylation detected in CDA in MCF7 might influence its catalytic activity and thus the concentration of converted drug might be affected. Structural analysis suggests that Y79 is far away from the catalytic site, but might affect the assembly of the CDA tetramer necessary for catalysis (Appendix 1 Fig. 5).

5.1.6 Global and promoters associated DNA methylation levels are not affected

5azadC is an inhibitor of methyltransferases (*DNMTs*) and induces, upon treatment, a global decrease in DNA methylation levels and, as a consequence, the reactivation of the

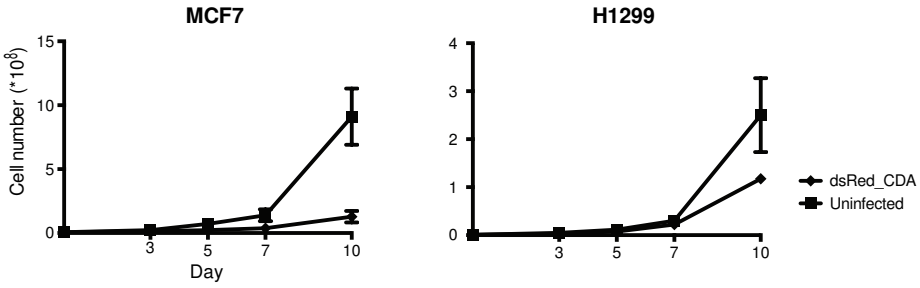


Figure 5.21: Treatment of overexpressing cell lines MCF7dsRed_CDA and H1299dsRed_CDA with 10 μM 5hmdC (n =3. Whiskers show standard deviation).

expression of some genes [Bender et al., 1998]. To test whether the treatment did not involve the DNA methylation pathway, the global levels of DNA methylation were checked by HPLC at day three after treatment with 10 μM 5hmdC in all the treated cell lines, and compared to the dC and 5azadC treated ones.

DNA methylation was globally decreased at day three after treatment with 5azadC by approximately 75% in all treated cell lines, while 5mC levels did not change upon 5hmdC treatment (Fig. 5.22), remaining similar to the control samples, i.e. between 0.5 - 1% of the total guanine amount.

Furthermore, the levels of expression of genes, whose promoters are known to be silenced by DNA methylation in some of the cell lines treated, were measured by qPCR. We selected: *PTEN* for LN18 [Rajendran et al., 2011] and MDA-MB-231 [Krawczyk et al., 2007]; *DERMO-1* for H1299 [Mao et al., 2011]; *TGF-3β* for A375 [Archey et al., 1999]; and *MLH1* for Colo320 [Sakamoto et al., 2001]. Significant re-expression of these genes at day three after 5hmdC treatment was not observed, as opposed to 5azadC treatment, which induced it of at least two folds (Fig. 5.23).

Given the results presented (Fig. 5.22, 5.23), it was excluded the possibility that reactivation of expression of genes normally silenced by DNA methylation or of a DNA demethylation pathway could be involved in the observed growth inhibitory effect of 5hmC.

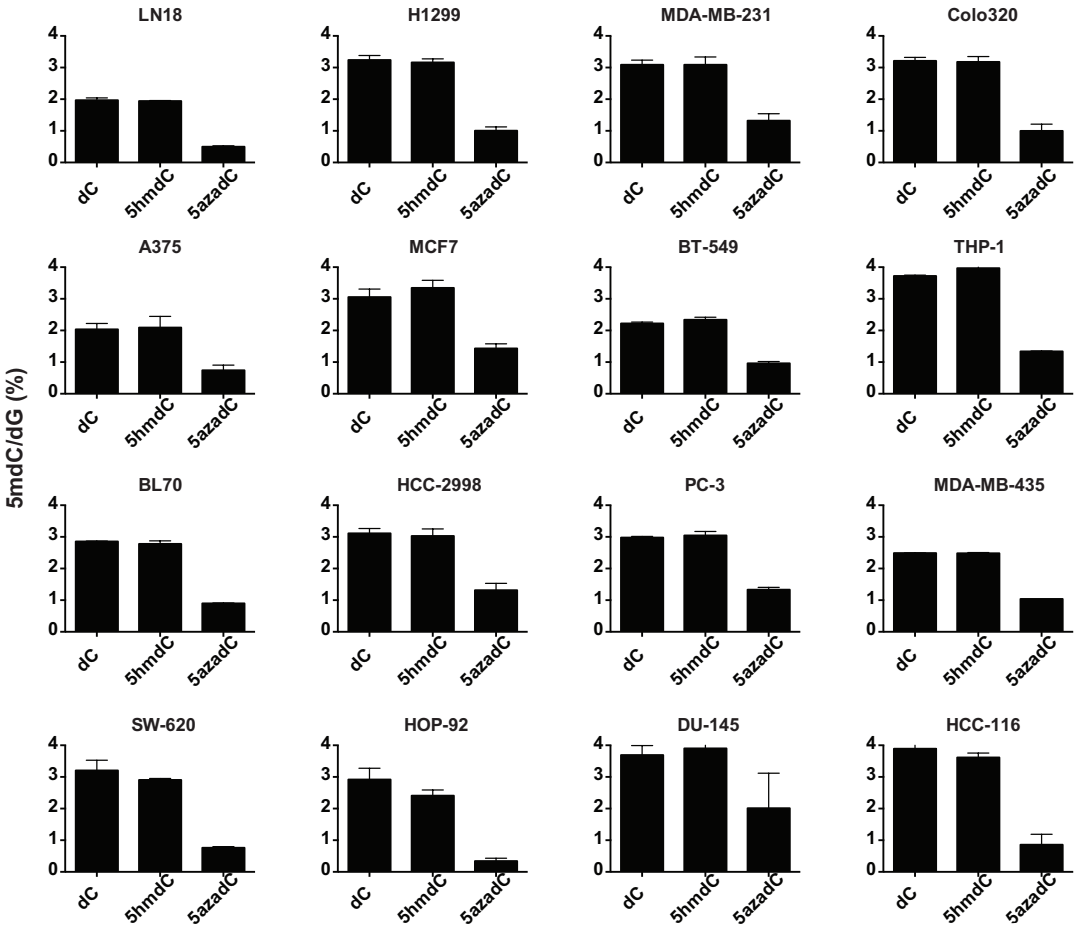


Figure 5.22: Global DNA methylation in treated cell lines with 10 μM dC, 5hmdC and 5azadC (5mdC/dG percentage shown) (n =3. Whiskers indicate standard deviation.

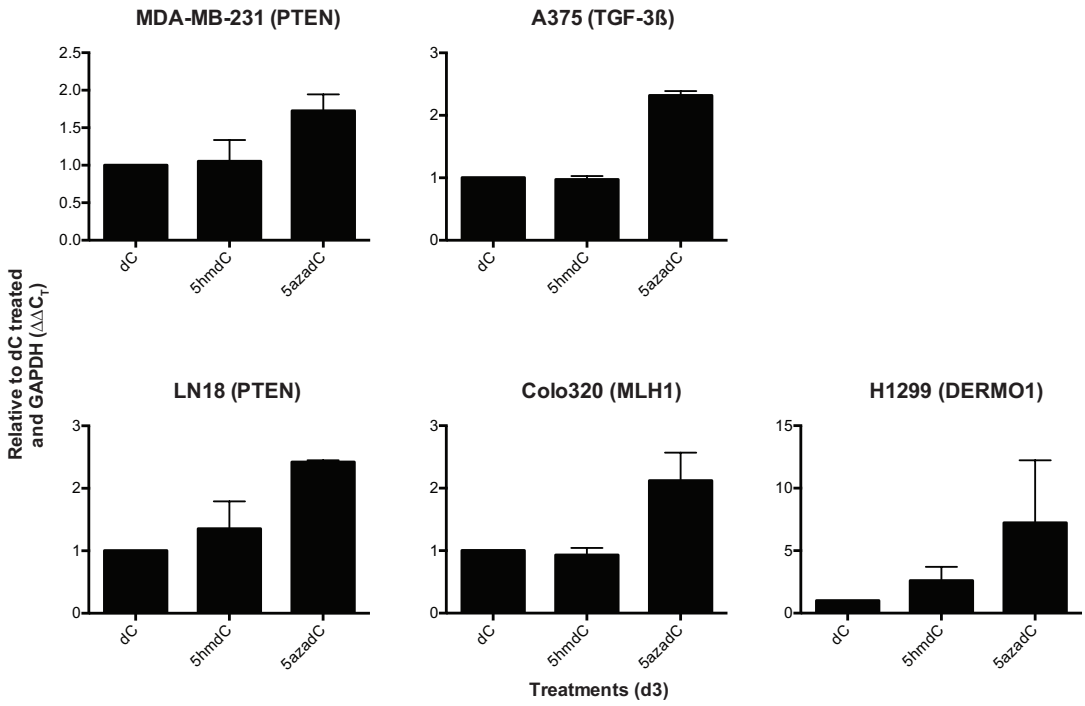


Figure 5.23: Expression of genes known to be silenced by promoter DNA methylation in cell lines treated with 10 μM dC, 5hmdC and 5azadC at day 3 (level relative to dC-treated cells and *GAPDH* shown) (n =3. Whiskers show standard deviation).

5.1.7 Cytidine deaminase can convert 5hmdC into 5hmdU

To further evaluate our hypothesis (Fig. 5.13), the possibility of 5hmdC being converted into 5hmdU was assessed. Cytidine deaminase [Cacciamani et al., 1991] [Vincenzetti, 1996] is an enzyme of the nucleotide salvage pathway that has been studied, primarily for medical reasons, with regards of its involvement in the deamination and inactivation of many nucleoside analogs used in cancer therapy [Lemaire et al., 2008][Bapiro et al., 2012]. It has been shown to be able to deaminate its primary substrate dC and additionally 5mdC, 5azadC and gemcitabine. Therefore, its deamination potential on the newly discovered forms of cytosine (5hmdC [Kriaucionis and Heintz, 2009][Tahiliani et al., 2009], 5fdC and 5cadC [Ito et al., 2011]) was assessed. Recombinant CDA was purified from *E. Coli* and its activity assessed spectrophotometrically, as previously shown [Vincenzetti et al., 2004],

with the end products confirmed by HPLC. The enzyme was able to successfully deaminate dC, 5hmdC and 5fdC, but not 5cadC (Fig. 5.24).

As shown by kinetic data (Fig. 5.25, Table 5.2), the best substrate was dC that had

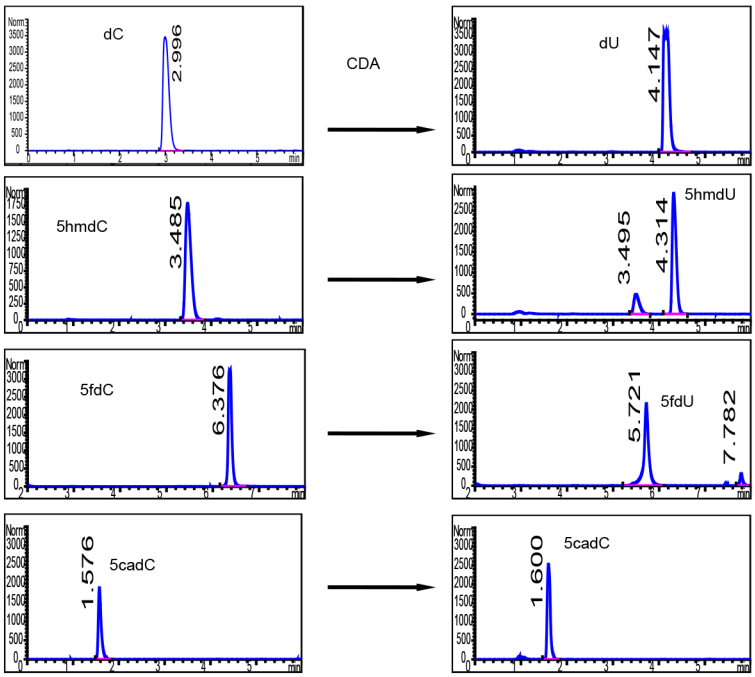


Figure 5.24: HPLC representative plots of CDA substrates and products (n =3).

Substrate	$\mathbf{K_m}(*\mathbf{10^{-5}M})$	$\mathbf{V_{max}}(\mathbf{M/min})$	$(\mathbf{K_{cat}})(*\mathbf{10^8})$
dC	2.15 ± 0.263	17.39 ± 0.660	6.42 ± 0.244
5mdC	4.88 ± 0.849	3.92 ± 0.095	1.45 ± 0.107
5hmdC	6.20 ± 1.05	1.229 ± 0.289	0.041 ± 0.0032
5fdC	6.85 ± 1.24	13.15 ± 1.134	4.85 ± 0.418
5azadC	5.15 ± 0.56	3.827 ± 0.180	1.417 ± 0.126

Table 5.2: Kinetic values of CDA deamination. K_m,V_{max} and K_{cat} are indicated with the respective unit of measure (n =3).

the lowest K_m and the highest V_{max}; 5mdC and 5azadC were recognized similarly and deaminated almost at the same rate (similar V_{max}); while 5fdC had the faster deamination rate compared to 5hmdC and the other nucleoside variants tested. 5cadC could not be

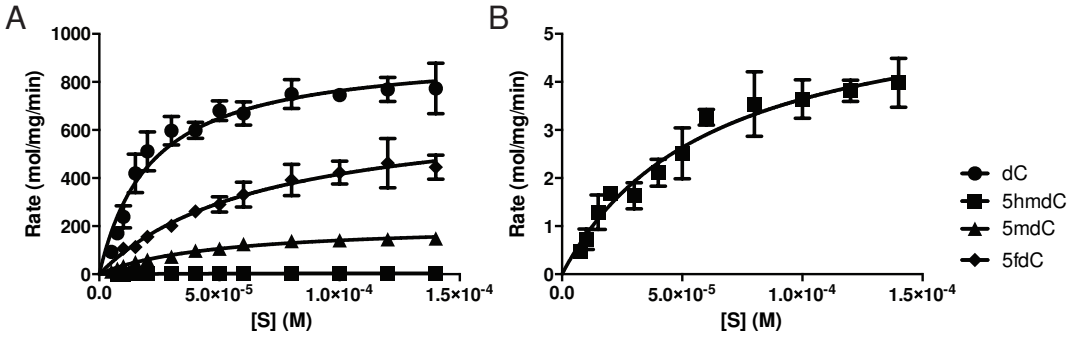


Figure 5.25: Kinetic plots (Michaelis Menten) of CDA activity.

deaminated, therefore, whether it could be used as an inhibitor of the conversion dC to dU at different concentrations was investigated. Inhibitors of *CDA*, like zebularine, have been developed [Kim et al., 1986] to overcome the resistance to nucleoside analogues caused by *CDA* overexpression. It was seen that from 0.1 M, 5cadC started to inhibit the reaction in a competitive manner, by decreasing the K_m and the V_{max}/K_m ratio (Table 5.3).

Moreover, 5hmdC and 5fdC were successfully modeled into the catalytic site of the crystal structure of CDA [Chung et al., 2005] (Fig. 5.26). CDA assembles as a tetrameter in which every subunit is able to deaminate one substrate [Vincenzetti, 1996]. Molecular docking suggested that 5fdC docks to the catalytic site nearly at 180° rotation, when compared to the unmodified cytidine, while positioning the amino group close to the active site with Zn^{2+} within it.

Inhibitor concentration [M]	$\mathbf{K_m}(*\mathbf{10^{-5}M})$	$\mathbf{V_{max}}(\mathbf{M/min/mg})$	$(\mathbf{V_{max}/K_m})(*\mathbf{10^5})$
0.4	3.79 ± 0.72	1120 ± 78.6	295.51
0.2	4.71 ± 1.16	998 ± 99.51	211.88
0.15	9.46 ± 0.321	732 ± 47	77.37
0.1	3.66 ± 0.75	801 ± 60.25	218.85
0.05	2.23 ± 0.75	905 ± 87.67	405.83

Table 5.3: Kinetic values of CDA deamination with 5cadC as an inhibitor. K_m,V_{max} and their ratio are indicated with the respective unit of measure (n =3).

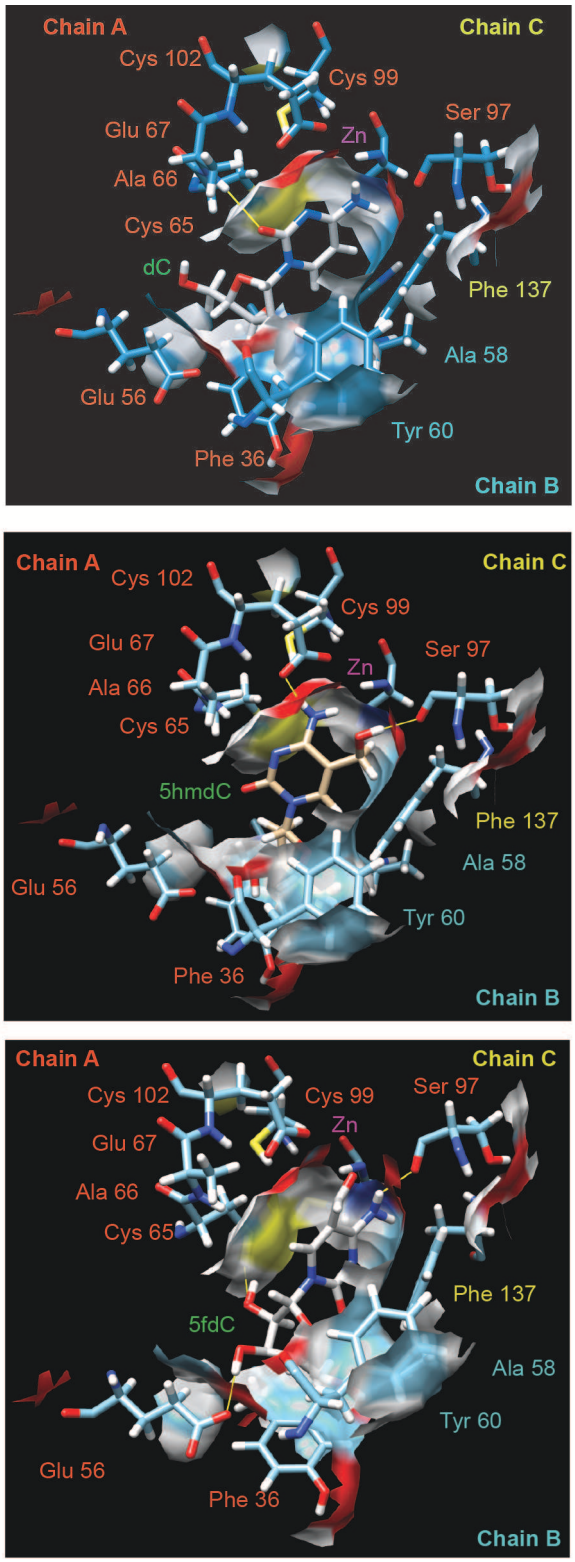


Figure 5.26: From top to bottom dC, 5hmdC and 5fdC modeling in the active site of CDA.

5.1.8 5hmdU can be detected in the DNA of treated cell lines

Subsequently, it was tested whether 5hmdU could be detected in the DNA of 5hmdC-treated cells. This was tested in the same HPLC chromatograms that were used to measure 5mdC levels. Variable levels of 5hmdU were found in the DNA of 5hmdC-treated cell lines (Fig. 5.27). It can be noted that the cell lines that responded to the treatment (e. g. MDA-MB-231, Capan-2, SN12C, HOP-92) had higher levels of 5hmdU in their DNA, at day 3 after the treatment. It was previously shown that 5hmdU/T substitutions of 1/150 could be toxic to the cell [Boorstein et al., 1987]. The percentages observed of 5hmdU were around 0.9% of T, making the hypothesis that this base can create a sufficient amount of DNA damage to kill the cell appear reasonable. As an example, HPLC chromatograms of

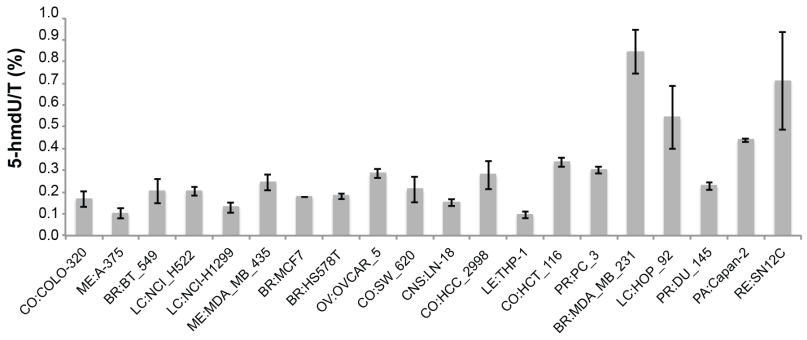


Figure 5.27: 5hmdU/dT (%) levels across 5hmdC-treated cell lines at day 3 (n =3. Whiskers indicate standard deviation).

dC- and 5hmdC-treated MDA-MB-231 cells are shown (Fig. 5.28). To further summarize the data for the treated cell lines, a relationship was established between the percentage of 5hmdU over T, versus the *CDA* expression levels from the NCI60 and CCLE datasets [Barretina et al., 2012][Pfister et al., 2009]). The linear correlation coefficient was 0.65, with cell lines with higher concentration of 5hmdU correlating well with higher *CDA* expression values (Fig. 5.29).

In summary, 5hmdU could be detected in the DNA of 5hmdC treated cell lines.

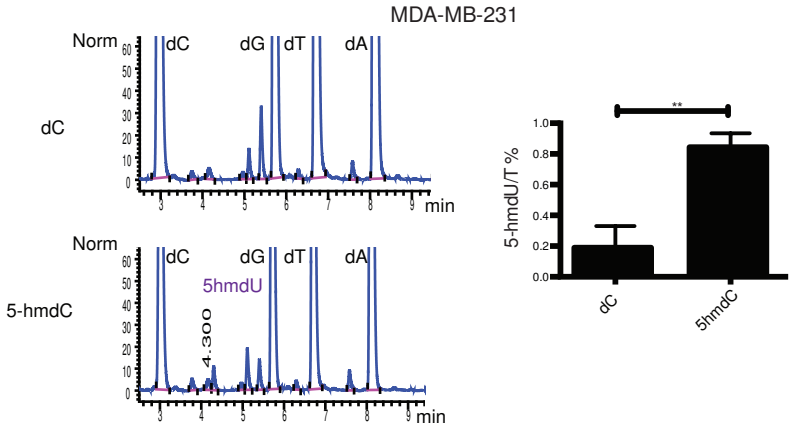


Figure 5.28: HPLC chromatograms of MDAMB231 dC- and 5hmdC-treated cells. 5hmdU is outlined in purple. Relative quantification is shown (n=3) *t*-test, p value =0.0057. Norm. absorbance units (n =3. Whiskers indicate standard deviation).

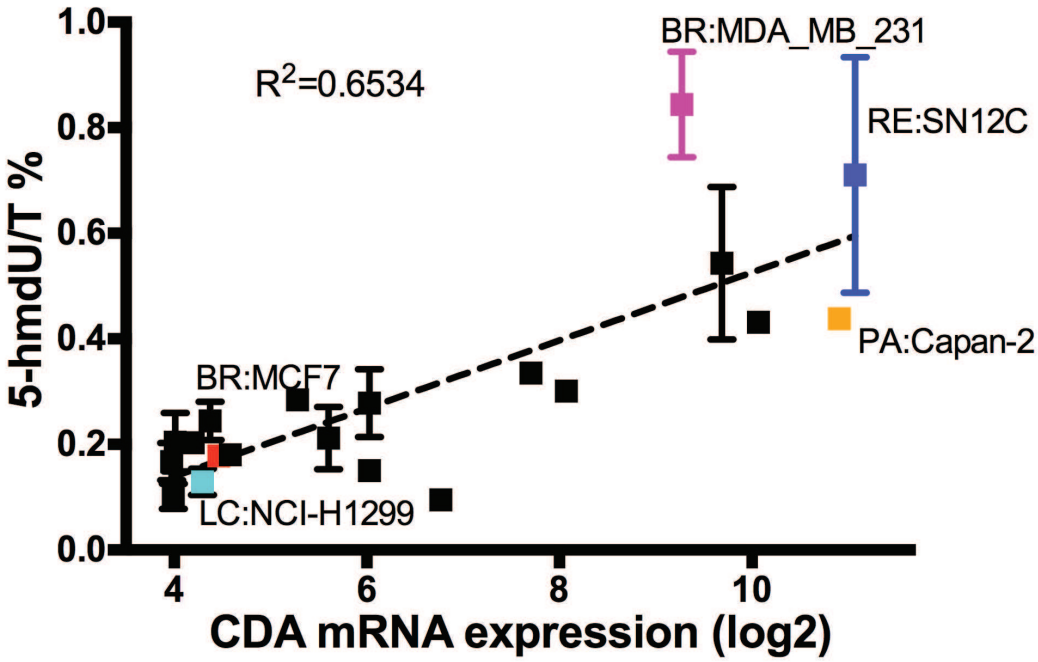


Figure 5.29: Correlation between *CDA* expression and 5hmdU levels in the DNA of 5hmdC-treated cells (R^2 is indicated. n =3. Whiskers indicate standard deviation).

5.1.9 5fdC can inhibit cell growth with the same specificity as 5hmdC

Following the finding that 5fdC is deaminated at a faster rate compared to 5hmdC (Table 5.2), whether or not the same could be observed in the cell lines of the extreme respondent and non respondent groups was tested. Experiments were conducted using the same concentrations of 5hmdC (0.1, 1 and 10 μM) as at the beginning, even though (according to kinetics data) 5fdC would show increased toxicity in *CDA* overexpressing cell lines at the lower dose tested, as compared to 5hmdC. MDA-MB-231 and SN12C were selected from the group of cell lines sensitive to 5hmdC treatment, and H1299 and MCF7 from the group of non-sensitive cell lines.

At day 3 of treatment 10 μM of 5fdC reduced proliferation by more than 50 % (Fig. 5.30), and by day 10 this was true also for 1 μM-treated cell lines, which was stronger than that of 5hmdC. To further validate the potency of the compound, the half effective concentration (EC_{50}) of 5hmdC and 5fdC in MDA-MB-231 and SN12C cells was calculated at day three. The EC_{50} of 5hmdC was higher compared to that of 5fdC, and the values correlated with *CDA* levels in these cell lines, being overall lower in SN12C (Fig. 5.31).

A further experimental system was used to assess the dependency of 5fdC and 5hmdC treatment on *CDA* expression. H1299 cells overexpressing CDA (H1299dsRed_CDA) were previously established. This cell line had a fluorescent marker (*dsRed*), cloned in the same open reading frame, upstream of *CDA* with an IRES sequence. Therefore, an experiment was set out in which, after mixing H1299 and H1299dsRed_CDA at an equal ratio, the cells could be treated with different concentrations of 5hmdC and 5fdC and the ratio of red and non-fluorescent cells could be monitored by FACS, over a period of ten days. At the end of the experiment (day 10), one population would have prevailed (Fig. 5.32). Furthermore, it was assessed whether 5hmdU was produced by H1299dsRed_CDA and secreted into the media where it could have exerted its inhibitory potential on the other cell line.

The *CDA* overexpressing population (dsRed) was almost abolished at the end of the

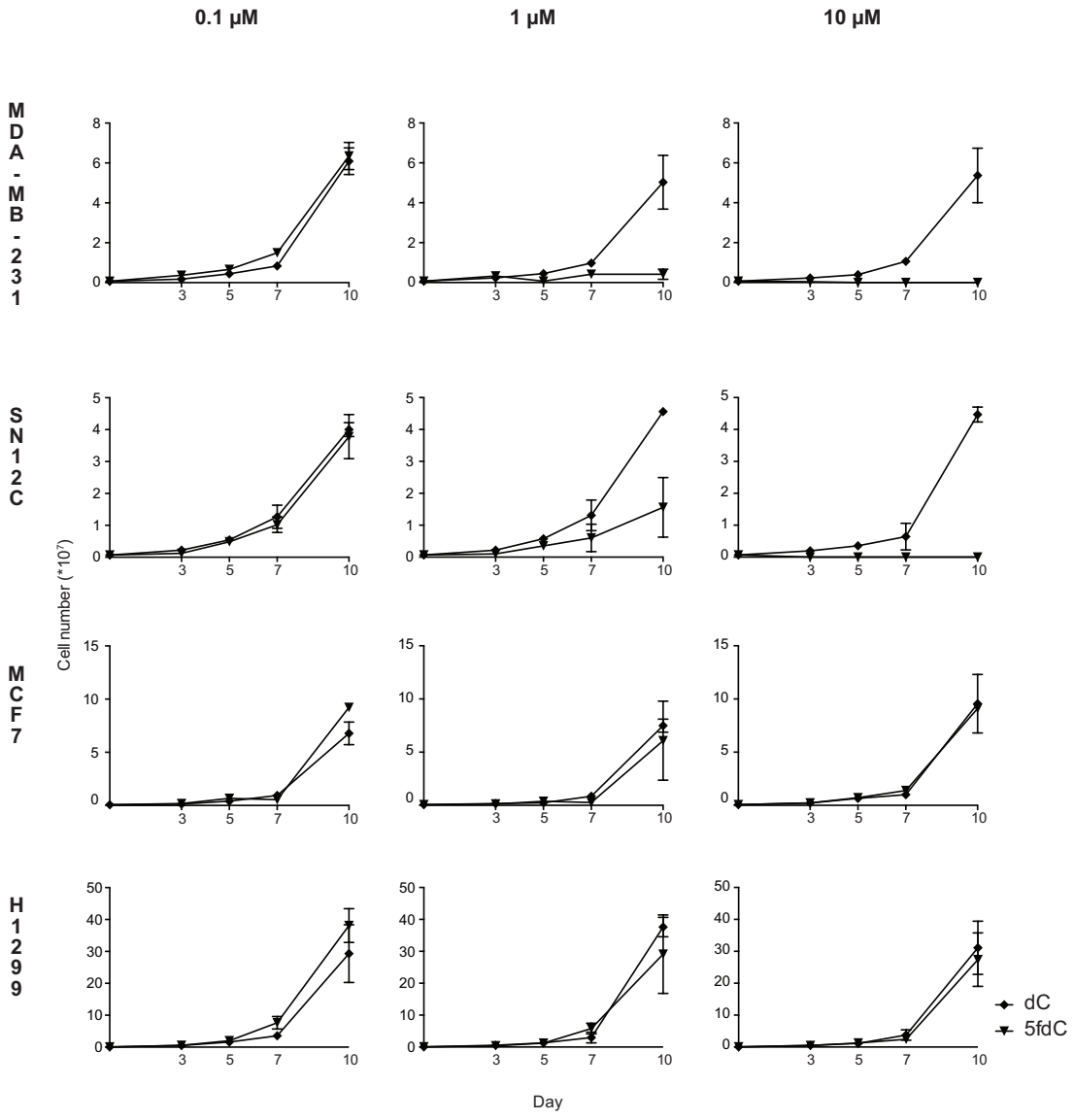
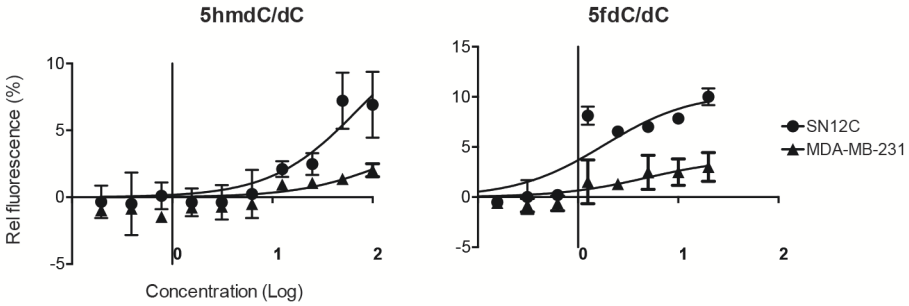


Figure 5.30: Growth rate of different cell lines at 0.1, 1 and 10 μM of dC and 5fdC (n =3. Whiskers indicate standard deviation).



Cell line / EC_{50}	5hmdC/dC (µM)	5fdC/dC (µM)
MDA-MB-231	131.5	4.8
SN12C	72.15	1.8

Figure 5.31: EC_{50} of 5hmdC and 5fdC in MDA-MB-231 and SN12C cells relative to dC at day 3 (n=2. Whiskers indicate standard deviation).

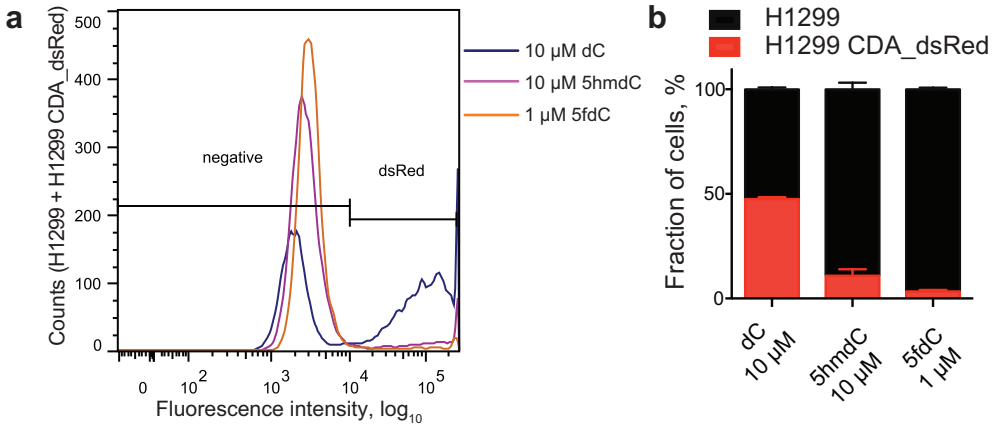


Figure 5.32: FACS competition experiment with H1299 and H1299dsRed_CDA. A. Example of FACS plot to show the two populations gated. B. Quantification of the two populations at day 10. Red indicates H1299dsRed_CDA (n =3. Whiskers show standard deviation. 10,000 events recorded).

treatment with 5hmdC and 5fdC, while in the control treatment the two populations were still present at about the same ratio. 5fdC was about 10 times more potent than 5hmdC, as suggested by the previous experiments.

5.1.10 5fdC treatment does not induce changes in global DNA methylation

As for 5hmdC treatment, global DNA methylation levels were checked in 5fdC-treated cell lines at day 3 at the lower dose of 1 μM, as at 10 μM most of the cells were dead and it was not possible to pursue the analysis.

No change in DNA methylation was detected compared to dC-treated cells (Fig. 5.33). In

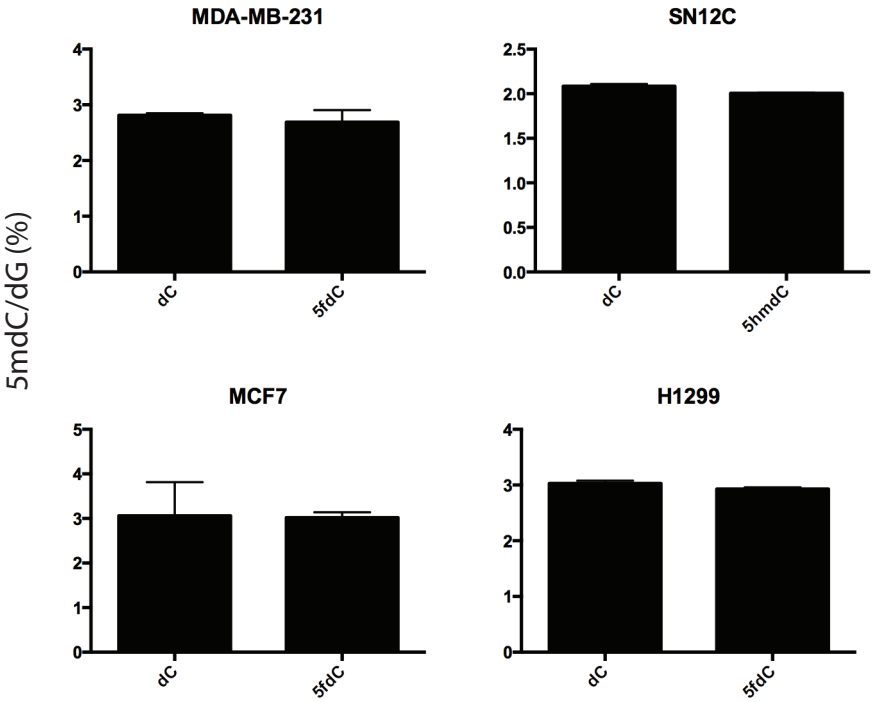


Figure 5.33: Global DNA methylation (5mdC/dG %) after 1 μM 5fdC treatment (n =3. Whiskers show standard deviation).

the same chromatograms, 5fdU was not be detected because it co-eluted with 5mdC on the HPLC column available.

5.1.11 5hmdC and 5fdC treatment induce DNA damage

The deamination of 5hmdC and 5fdC produces 5hmdU and 5fdU, respectively. These would be incorporated into the DNA after subsequent conversion to triphosphate mediated by thymidine kinases [Kaufman, 1986][Klungland et al., 2001] and induce a DNA damage response by the formation of double strand breaks [Boorstein et al., 1992] [Klungland et al., 2001][Mi, 2001]. Therefore, treated cells were stained at day 3 with the well-known marker of double strand breaks γH2AX [Rogakou et al., 1998][Rogakou et al., 1999]. γH2AX was induced in 5hmdC, 5fdC, 5azadC and UV-treated MDA-MB-231 cells (Fig. 5.35), while only in 5azadC and UV-treated H1299 (Fig. 5.35). Furthermore, the levels of DNA damage upon various treatments in different cell lines were quantified by counting the relative number of cells that stained positive for γH2AX (Fig. 5.36). A significant increase (at least 10%) in the total number of γH2AX positive cells was seen after both 5hmdC and dC treatment in MDA-MB-231 cells, after 5hmdC treatment in HOP-92 and after 5fdC treatment in SN12C, but not in H1299, MCF7 or SW-620 cells. The observed induction of γH2AX with high doses of 5azadC (10μM) was already documented in the literature [Palii et al., 2008].

5.1.12 5hmdC and 5fdC treatment and the cell cycle

When DNA damage occurs, the cell attempts to repair it. In order for it to have the possibility to do so, it activates cell cycle checkpoints, which stall the progression of the cell cycle before the completion of mitosis can occur [Hartwell and Weinert, 1989][Weinert and Hartwell, 1988]. Noticed the induction of γH2AX in responding cell lines, progression through the cell cycle was checked by propidium iodide staining. In the responsive cell lines (MDA-MB-231, CAPAN-2 and SN12C), 5hmdC and 5fdC induced an increase in the percentage of cells in S phase or an increase in G2/M by approximately 10%, which was common to 5fdC treated cells (Fig. 5.38 and 5.37). This could be due to a pause in the

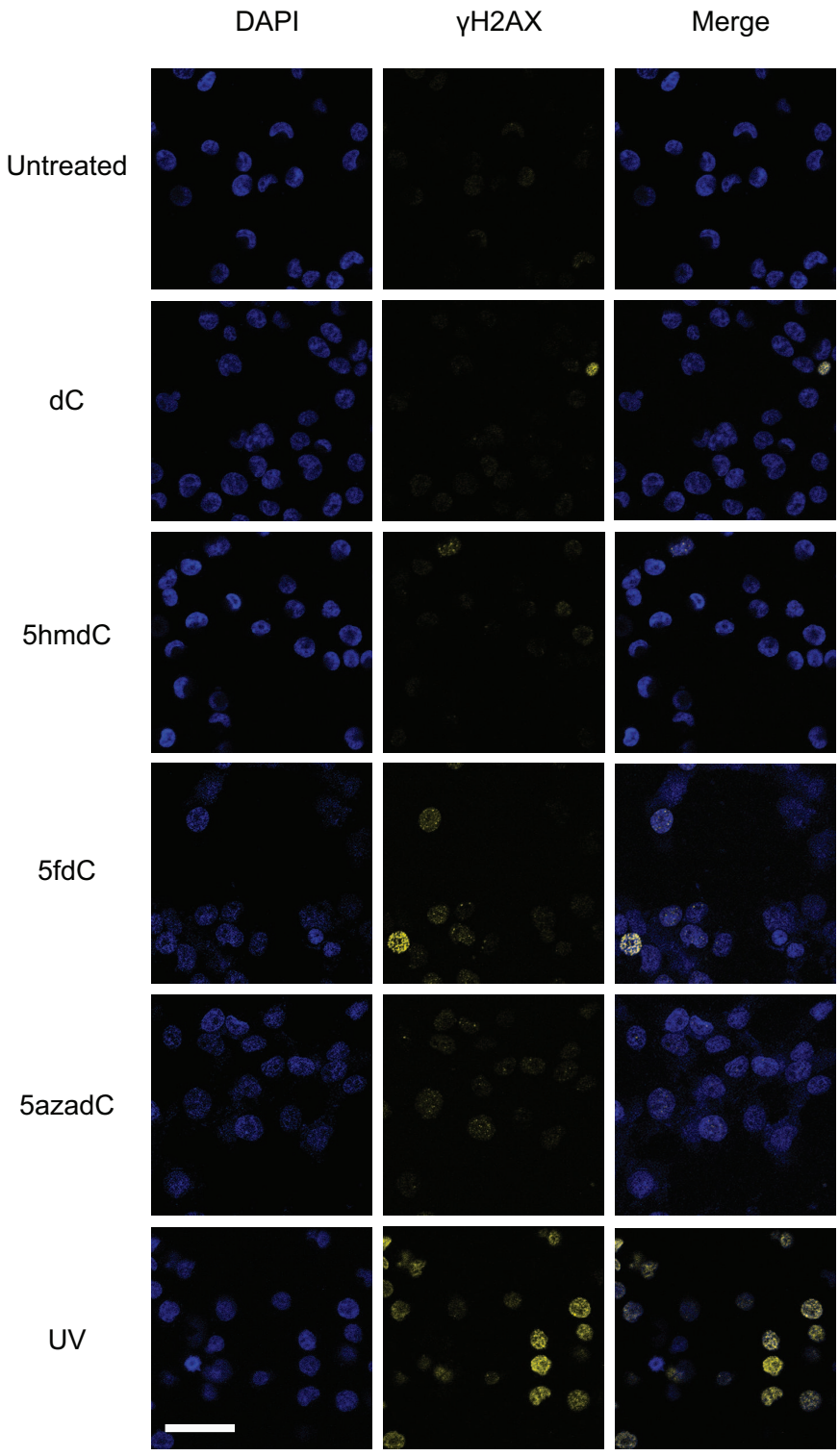


Figure 5.34: γH2AX levels in treated H1299 cells (Scale bar=50μm). Representative pictures (n =3).

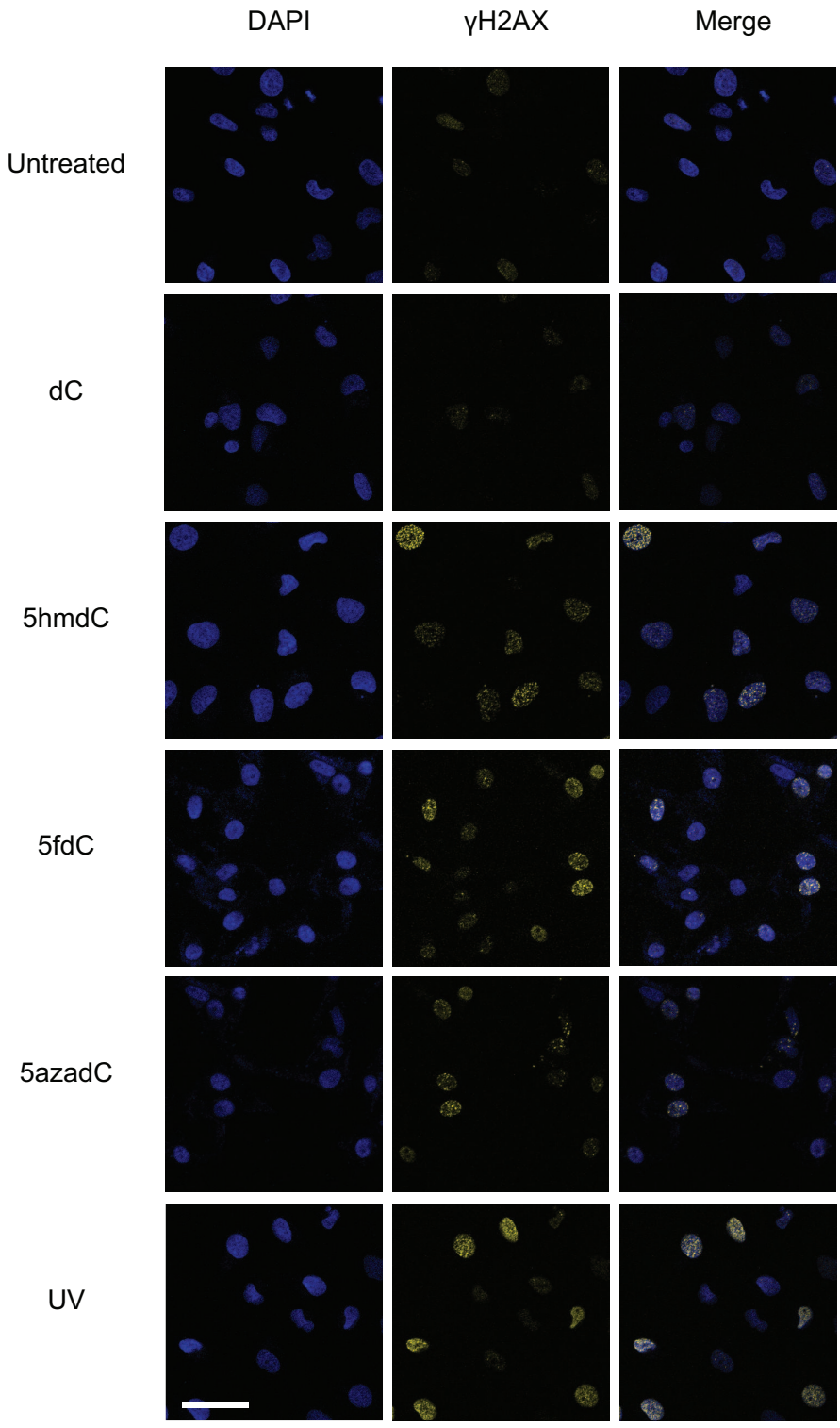


Figure 5.35: γH2AX levels in treated MDA-MB-231 cells (Scale bar =50 μm) Representative pictures (n =3).

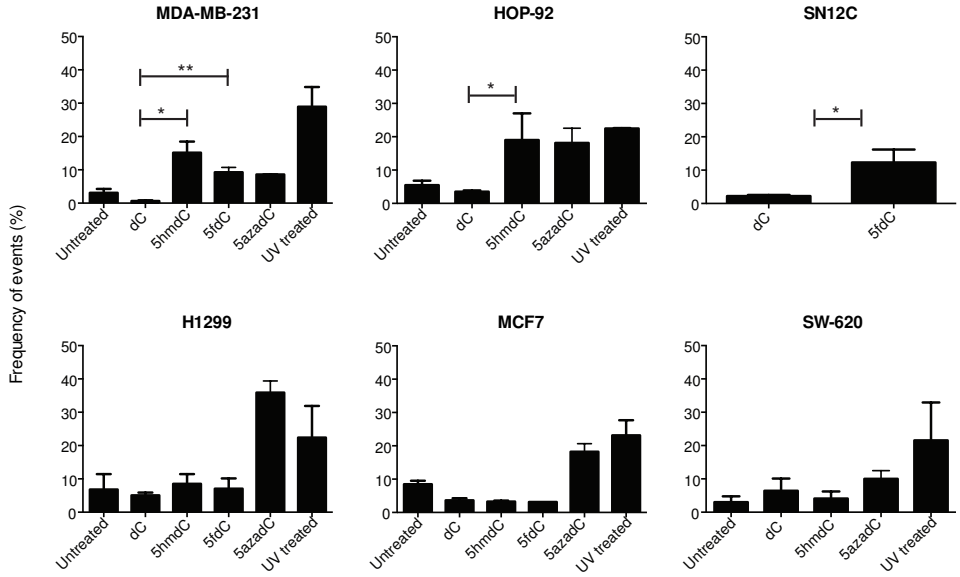


Figure 5.36: γH2AX quantification in treated cells with 10 μM dC, 5hmdC, 5azadC and 1 μM 5fdC. ANOVA with Sidak correction for multiple comparisons (MDA-MB-231, HOP-92, H1299, SW-620) and two tailed t test for SN12C. p (5hmdC vs dC MDA-MB-231) =0.0208; p (5fdC vs dC MDA-MB-231) =0.0022; p (5hmdC vs dC HOP-92) =0.0135; p (5fdC vs dC SN12C) =0.0392. (n =3. Whiskers indicate standard deviation.)

cell cycle taken to repair double strand breaks formed in the DNA by the incorporation of 5hmdU and 5fdU before progression into mitosis.

5.2 Brief summary

Many nucleoside analogs are currently used in cancer therapy. Their application in the clinic is becoming increasingly limited over time due to the insurgence of many resistance mechanisms, such as the repression of nucleoside transporter 1 (*ENT1*) and intracellular nucleoside kinases such as DCK, or the overexpression of *CDA* [Jordheim and Dumontet, 2007]. Moreover, DNA methylation nucleoside analogs (5azadC and 5azaC) are very unstable in solution, limiting their time frame of action once injected. Therefore, the possibility that the newly discovered forms of cytosine [Kriaucionis and Heintz, 2009][Tahiliani et al., 2009] 5hmdC, 5fdC and 5cadC [Ito et al., 2011] could be used as cytotoxic

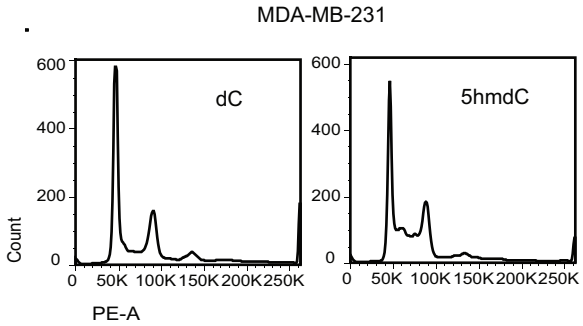


Figure 5.37: FACS cell cycle plot of 5hmdC- and dC-treated MDA-MB-231 cells (10,000 events recorded). Representative picture (n =3).

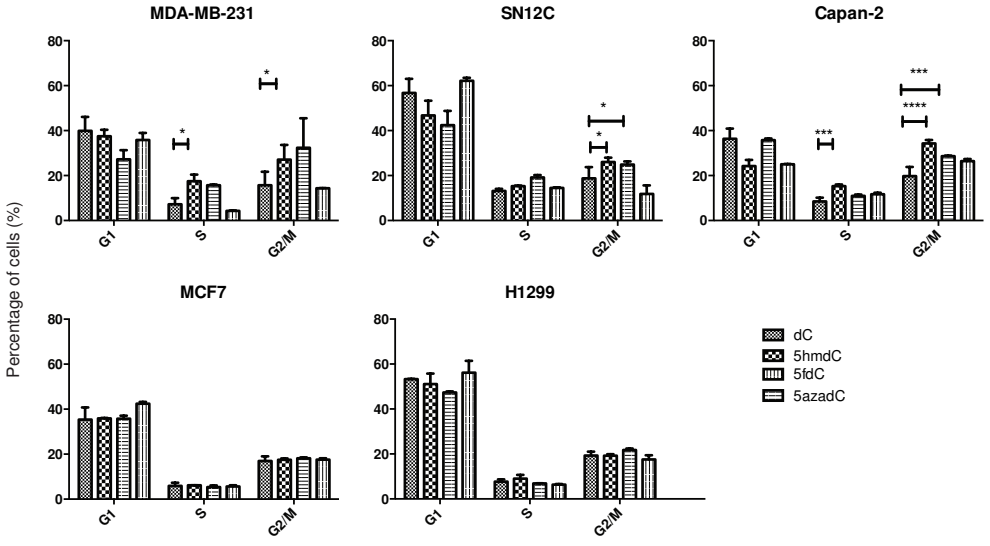


Figure 5.38: Quantification of the percentage of treated cells (10 μM dC, 5hmdC, 5azadC and 1 μM 5fdC) in each cell cycle stage by FACS. Two way ANOVA: p (S: 5hmdC vs dC MDA-MB-231) =0.0264; p (G2-M: 5hmdC vs dC MDA-MB-231) =0.0149; p (S: 5hmdC vs dC Capan-2) =0.0005; p (G2-M: 5hmdC vs dC Capan-2) <0.0001; p (G2-M: 5fdC vs dC Capan-2) =0.0007; p (G2-M: 5hmdC vs dC SN12C) =0.0276; p (G2-M: 5fdC vs dC SN12C) =0.0342. (n=3. Whiskers show standard deviation; 10000 events recorded).

compounds was investigated. Many papers have linked their respective bases to the role of intermediate in DNA demethylation pathways [Tahiliani et al., 2009][Hajkova et al., 2010][Guo et al., 2011]. Nucleoside analogs such as 5-azacytidine and its deoxyanalog 5-azadeoxycytidine, FDA approved drugs for the treatment of myelodysplastic syndromes, have been widely investigated for their role in inhibiting DNMTs, the enzymes responsible of DNA methylation [Jones, 1984]. This led to the hypothesis that 5hmdC, 5fdC and 5cadC could eventually lead to global DNA demethylation, by being incorporated into the DNA and then excised by the nucleotide excision repair pathway (SMUG1/TDG) per se, or following deamination by AID/APOBEC enzymes that would create mismatches (reviewed in [Gong and Zhu, 2011]).

Given the well known limitations regarding the stability of 5azadC, whether the compounds 5hmdC, 5fdC and 5cadC were more stable in aqueous environment at 37°C was assessed. Overall stability of these cytidine modified forms was observed over a period of 10 days except for 5fdC for which 30% was gradually lost between day 0 and day 10. On the other side, 5azadC was degraded in 24 h as previously described [Lin et al., 1981].

Having assessed the stability of the molecules over the time frame of the treatment window (10 days), their cytotoxic potential was determined in a subset of cancer cell lines (Table 5.1). It was found that 5hmdC can inhibit the growth of certain cancer cell lines, showing more potency in MDA-MB-231, HOP-92, Capan-2 and SN12C. Having utilized cancer cell lines belonging to the NCI60 panel or to the CCLE, it was possible to perform gene expression analysis of responding versus non responding cancer cell lines. In this way, *CDA* was identified as a gene differentially expressed between the two groups of cell lines. *CDA* is involved in the nucleotide salvage pathway where is responsible for the deamination of different forms of cytidine [Vincenzetti, 1996]. It was therefore hypothesized that 5hmdC could be deaminated to 5hmdU, and could then be incorporated into the DNA and induce DNA damage, as previously shown [Kaufman, 1986]. To investigate this, it was decided to first assess whether the gene was directly involved in the response that was ob-

served. Knockdowns of *CDA* in responsive cell lines were subsequently established and the gene was overexpressed in non-sensitive cell lines. The phenotype could be rescued in the knockdowns (Fig. 5.19) and overexpressing cell lines were partially sensitized (Fig. 5.21). Moreover, global DNA methylation levels did not differ among dC- and 5hmdC-treated cell lines, as compared to 5azadC-treated cells, which showed more than a 50% decrease in the levels of 5mdC. Furthermore, promoters known to be silenced in the treated cell lines, did not became re-expressed following treatment with 5hmdC, while a 2-fold induction of gene expression was observed in 5azadC-treated cells (Fig. 5.23).

It was considered that, if *CDA* was responsible for the observed phenotype, it should be able to deaminate 5hmdC. The human recombinant enzyme was purified and tested on 5hmdC and, since by that time 5fdC and 5cadC discovery had been reported [Ito et al., 2011], these variants were also assessed. Both 5hmdC and 5fdC could be deaminated (Fig. 5.24), but 5fdC was deaminated at faster rates compared to 5hmdC (Table 5.2). The deaminated forms would be 5hmdU and 5fdU, respectively. Previous studies have found that, if administered to cells, these nucleosides can be incorporated [Kaufman, 1986][Klungland et al., 2001] and confer toxicity to the cells by inducing DNA breaks and the base excision repair pathway [Mi, 2001]. As a consequence of deamination, it was hypothesized that 5hmdU could be found in the DNA of 5hmdC-treated cells. Subsequently, 5hmdU was detected in the DNA of 5hmdC-treated cell lines by HPLC, at levels proportional to CDA expression levels (Fig. 5.29).

Having found that 5fdC was deaminated at faster rates compared to 5hmdC, its growth inhibitory potential was tested in cell lines belonging to the same groups that had been defined with 5hmdC, predicting that the same dependency would be observed. The hypothesis was validated and the same groups of cells were sensitive or insensitive to 5fdC treatment, as they were to 5hmdC-treatment. Furthermore, 5fdC showed a 10 times greater growth inhibitory potential than 5hmdC, as predicted by kinetic data. Moreover, a documented increased mutagenic potential of 5fdC compared to 5hmC incorporation could

further explain this discrepancy [Miyashita et al., 2002][Klungland et al., 2001][Kamiya et al., 2002]. 5fdC was also unable to induce any global change in DNA methylation.
Lastly, whether or not a DNA damage response could be found in treated cell lines after they had undergone at least one replication was tested, at day three after treatment with 5hmdC and 5fdC. An increase in γH2AX positive cells in both treatment conditions could be detected in different sensitive cell lines, as assessed by confocal microscopy. Moreover, the treated and responsive cell lines were inhibited in their cell cycle progression, either in S or G2-M phase, in an attempt to repair the DNA damage that had accumulated.
In summary, in this chapter 5hmdC and 5fdC have been established as potential therapeutic agents *in vitro*, which can be added to the spectrum of nucleoside analogs currently used in the clinic.

Chapter 6

5hmdC and 5fdC are not toxic and can impair tumor growth *in vivo*

It was previously found that 5hmdC and 5fdC show anti-proliferative effects *in vitro* in cancer cell lines that overexpressed CDA. Furthermore, it was found that their growth inhibitory phenotype was closely related to their CDA mediated conversion into the nucleosides 5hmdC and 5fdU and their subsequent incorporation into the DNA. This further resulted in the induction of DNA damage and delay in named phases of the cell cycle. To assess the feasibility of using these named compounds as anticancer agents, the analysis needed to be extended to the *in vivo* setting. The assessment of the toxicology profile of the compounds was first needed. Mouse studies were the preferred choice since they are a good model to assess the physiological role of a named compound systematically. A further step was later necessary. *In vivo* drug preclinical efficacy studies can be conducted either in a genetic mouse model of cancer or with the use of mice with externally transplanted tumors (xenografts) from patient derived primary cells or from cancer cell lines. Cancer cell lines derived xenografts offer standardization and easy access, as compared to genetic mouse models or primary patient derived cell lines. Xenograft tumor mouse

models can be induced via the injection of cancer cell lines in two different sites: either different from the origin of the cancer cells (ectopic) or in the same cancer cell origin site (orthopic) [Jung, 2014]. Orthopic injection was considered, but since homogeneous conditions of drug delivery were desired between the cell lines tested, which had different cancer origins, and easy access to the tumors was needed for monitoring reasons, this option was discounted. Moreover, due to the facility of monitoring subcutaneously injected tumors, the number of animals employed in the study could be reduced by injecting two cell lines per animal, one in each flank. Since the cell lines utilized were of human origin, the use of immunocompromised mice (NOD/SCID) was necessary, with the obvious disadvantages of interspecies differences and lack of immune response that have already been documented in the literature [Jung, 2014]. Consequently, xenografts were ectopically established by subcutaneous injection of two of the cancer cell lines tested. Since 5hmdC and 5fdC had never undergone *in vivo* testing prior to this study to our knowledge, their toxicology profiles needed to be assessed before xenografts were established, to determine:

- that they did not induce adverse effects on their own at a given dosage;
- that they were able to enter the bloodstream;
- the maximum dosage that could be administered, without causing side effects.

After toxicology studies, it was checked that the cell lines selected would give tumors in this mouse model in the desired time-frame of one month, to allow a short treatment period. Thus, the same cell lines that responded to treatment *in vitro*, were injected into immunocompromised mice and, once tumors were formed, a treatment schedule was followed. Tumor parameters were measured (volume, weight) and, once dissected, tumors were subject to histopathological analysis and confocal microscopy analysis of proliferation and DNA damage markers.

6.1 Results

6.1.1 5hmdC and 5fdC do not show apparent toxicity *in vivo*

To assess the feasibility of the xenograft experiment, it was necessary to check whether the *in vivo* administration of 5hmdC and 5fdC would be toxic *per se*. Therefore, three dosages were tested for 5hmdC (25, 50 and 100 mg/kg) and four dosages for 5fdC (12.5, 25, 50 and 100 mg/kg), selected on the basis of other nucleoside analog injected compounds, such as gemcitabine in the pancreatic mouse model which is injected at 100 mg/kg every 72 h [Bapiro et al., 2012]. The drug was injected in four animals per group at day zero and the animals sacrificed and tissues dissected at day five to give a monitoring window for the evaluation of behavioral differences in the mice treated with 5hmdC and 5fdC, as compared to the vehicle treated ones (PBS). In this five days window, the mice were monitored for absence of swollen site of injection, loss of appetite or weight and behavioral differences as compared to the control injected mice. No changes in the monitored parameters were noted after injection. Therefore, at the end of this window, mice were sacrificed and tissue dissected to evaluate the presence of any abnormality in tissue architecture by histopathological analysis (with the help of the expertise of the pathologist Dr. R. Goldin), incorporation of 5hmdU in the DNA by HPLC analysis and DNA damage by confocal microscopy. To select the tissues to dissect, databases of mouse gene expression data were scanned [Su et al., 2004] and five organs were selected according to their levels of *Cda* expression (Fig. 6.1). They were:

- small intestine;
- kidney;
- pancreas;
- liver;

- lung.

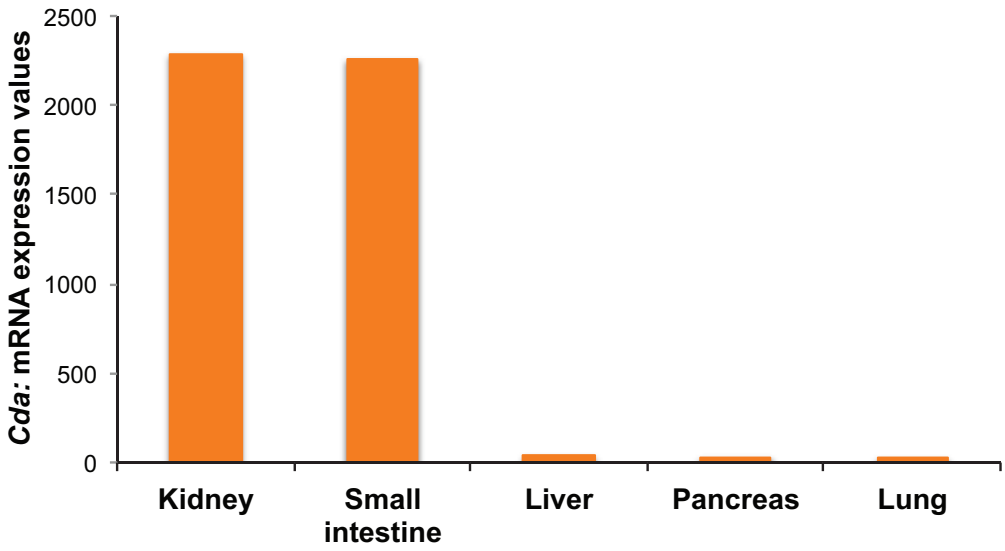


Figure 6.1: *Cda* expression in selected mouse tissues.

Due to its low Cda expression level, lung was selected as a negative control, where no significant changes were expected in the parameters due for analysis. Small intestine and kidney are the tissues where *Cda* expression is higher, therefore where most compound toxicity induced effects could be expected. In the following sections the data obtained for the highest injected dose of 5hmdC and 5fdC (100 mg/kg) will be summarized.

Small intestine

CDA expression is elevated in the intestine. It has been shown that the intestine, like bone marrow cells, is incapable of *de novo* nucleotides synthesis, relying on the salvage pathway for nucleotide provision [Mackinnon and Deller, 1973]. Histopathology was checked by hematoxylin and eosin staining and no gross abnormalities were detected upon 5hmdC and 5fdC administration, with major tissue architecture preserved (crypts, villi, lamina propria) (Fig. 6.2), despite CDA being expressed, as shown by confocal microscopy (Fig.

6.3).

In the DNA of the small intestine DNA methylation was measured by HPLC, but any

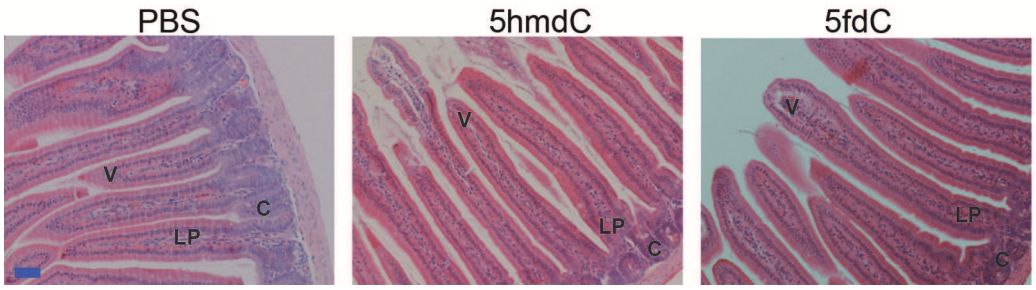


Figure 6.2: Hematoxylin and eosin staining in the small intestine upon PBS, 5hmdC or 5fdC administration (100 mg/kg). The *villi* V are lined by a simple columnar epithelium, which is continuous with that of the *crypts* C. The *lamina propria* LP extends between the crypt and into the core of each villus (vascular and lymphatic network necessary for absorption) (Scale bar =50 μm) (n =3 per group. Representative picture shown).

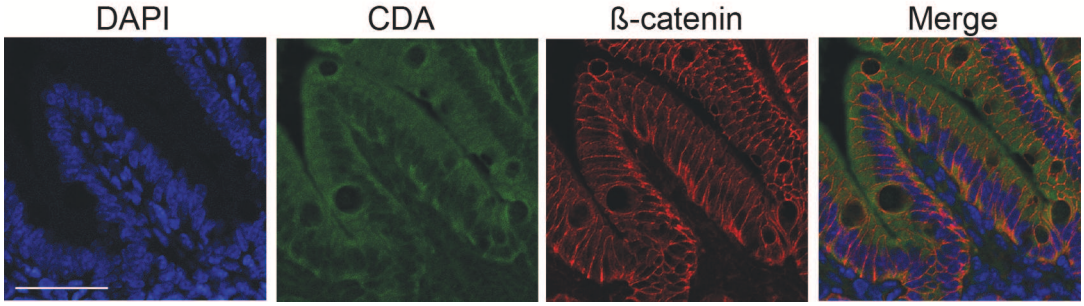


Figure 6.3: CDA staining in the small intestine (Scale bar =50 μm) (n =3 per group. Representative picture shown).

significant difference could not be detected between treated and untreated mice with regard to the content of 5mdC. 5hmdU was not detected. Few additional peaks differed between untreated and treated samples. They could be due to partial presence of RNA in the samples. No significant conclusions could be established due to variability for those peaks in the samples (Figure 6.4).

Subsequently, sections were stained for γH2AX and phosphorylated serine 10 of histone 3 (H3S10P), an established mitotic marker [Hendzel et al., 1997], to try to detect any DNA damage that might have occurred following 5hmdC and 5fdC administration, or any

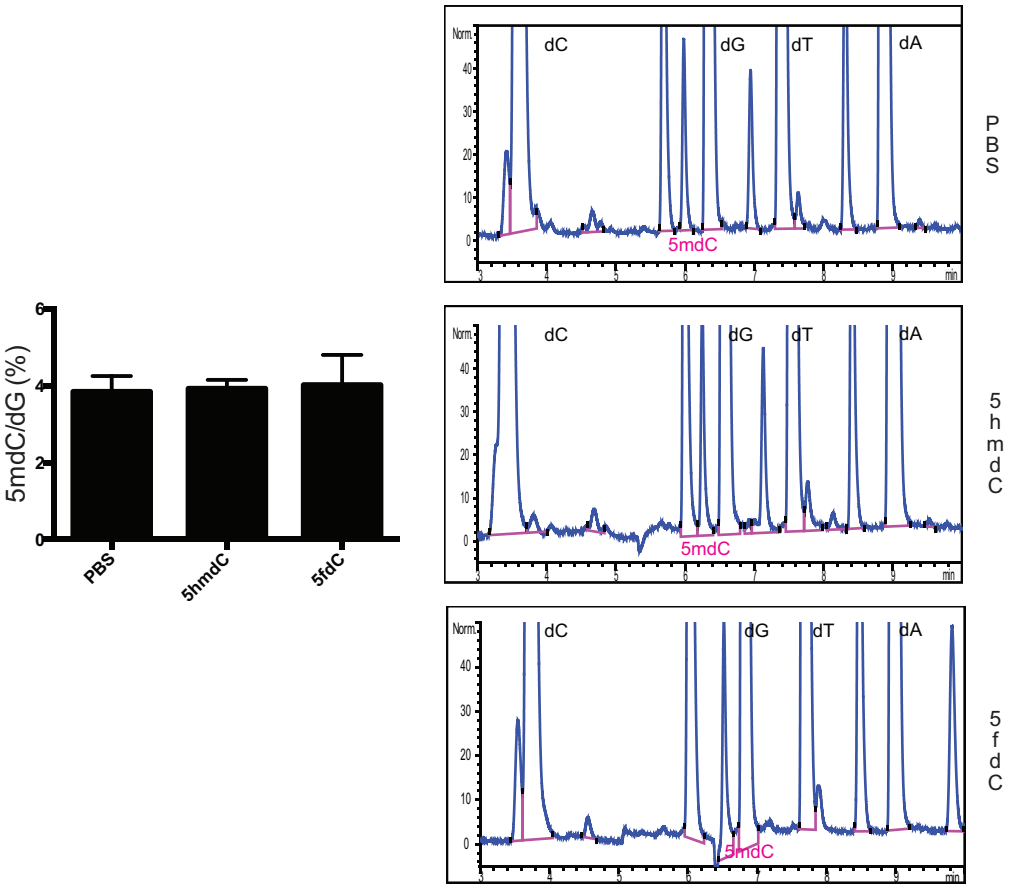


Figure 6.4: DNA methylation in the small intestine after treatment. Quantification and relative chromatograms are shown. Norm absorbance units (n =3 per group. Whiskers indicate standard deviation).

change in proliferation due to attempts by the cells to repair any DNA damage that might have been induced by the treatments. Positive staining was detected for PH3S10 at the bottom of the crypts, the site of most proliferation in the small intestine, but no changes were detected following 5hmdC or 5fdC treatment (Fig. 6.5).

Given the data presented, no overt toxicity was observed in the small intestine.

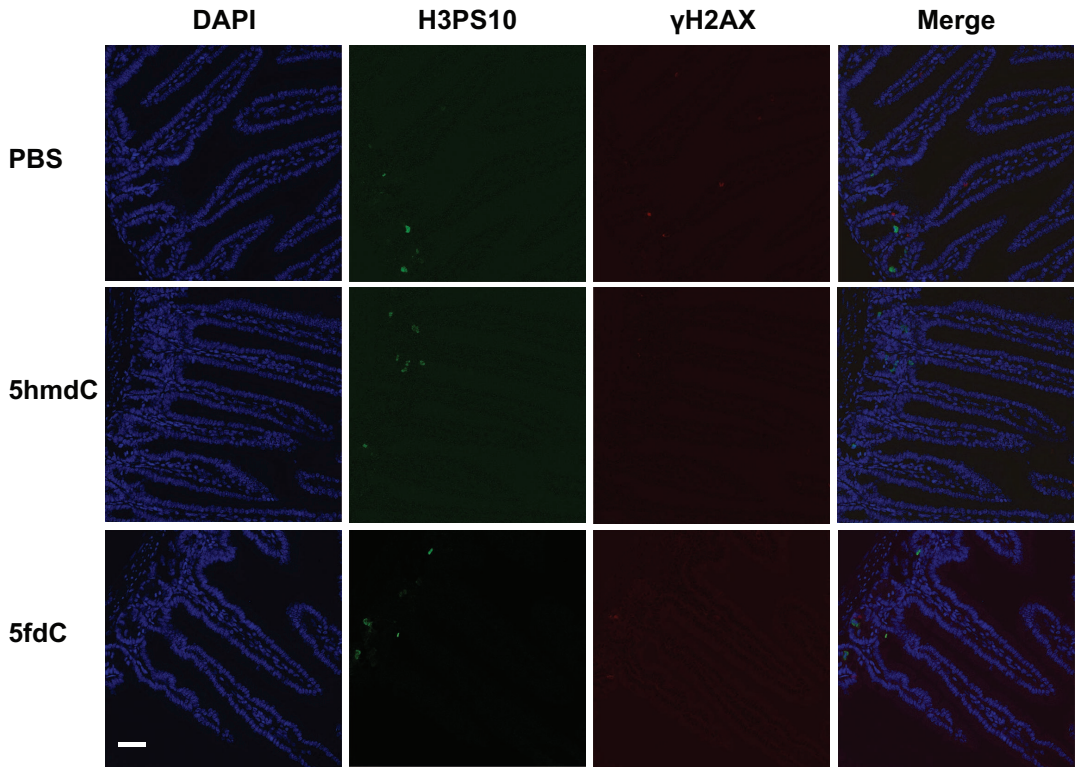


Figure 6.5: γH2AX in red and H3S10P in green in the small intestine after treatment.(Scale bar =50 μm) (n =3 per group. Representative picture shown).

Kidney

Another organ where the nucleotide salvage pathway plays a major role is the kidney as here the nucleoside transporters are highly expressed, which makes nucleoside reabsorption and metabolism likely [Rodríguez-Mulero et al., 2005]. Histopathology was assessed by hematoxylin and eosin staining and no gross abnormalities were detected upon 5hmdC and 5fdC adminstration compared to PBS treatment (Fig. 6.6), despite CDA being expressed (Fig. 6.7).

DNA methylation and hydroxymethylation was measured that was compatible with the high levels of 5hmdC previously observed in this tissue [Globisch et al., 2010], but no significant difference was detected between treated and untreated mice (Figure 6.8).

Furthermore, when sections were stained for γH2AX and H3S10P, no significant change

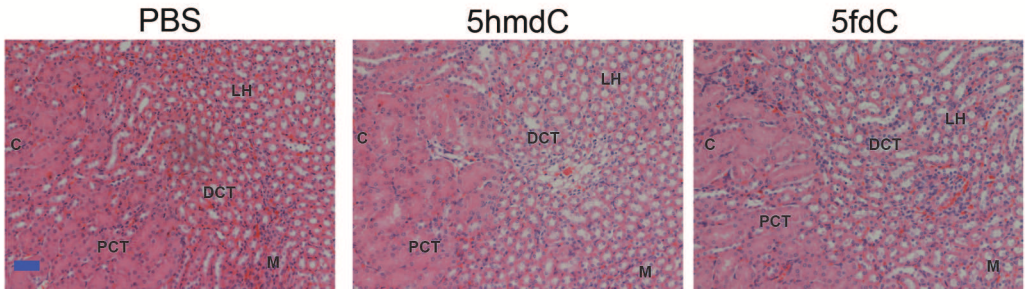


Figure 6.6: Hematoxylin and eosin staining in the kidneys upon PBS, 5hmdC or 5fdC administration (100 mg/kg) (Scale bar= 50 μm). Part of cortex (C) and medulla (M) can be seen. In the cortex/medulla are visible the proximal convoluted tubes (PCT), the loop of Henle (LH) and the proximal deconvoluted tubes (PDT) (n =3 per group. Representative picture shown).

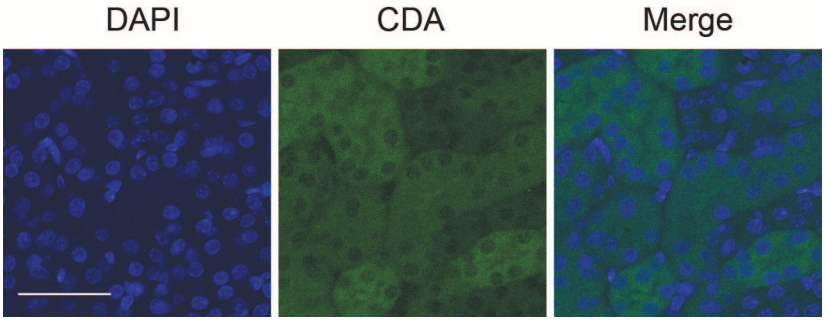


Figure 6.7: CDA staining in the kidneys (proximal convolute tuble shown) (Scale bar=50 μm) (n =3 per group. Representative picture shown).

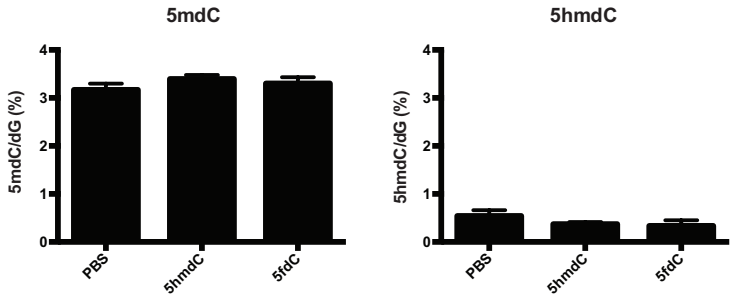


Figure 6.8: DNA methylation and hydroxymethylation in the kidneys after treatment. Quantification is shown (n =3 per group, whiskers indicate standard deviation).

between PBS-injected and 5hmdC/5fdC-injected animals was detected (figure 6.9).

Given the data presented, it can be concluded that no overt toxicity was observed in the kidneys.

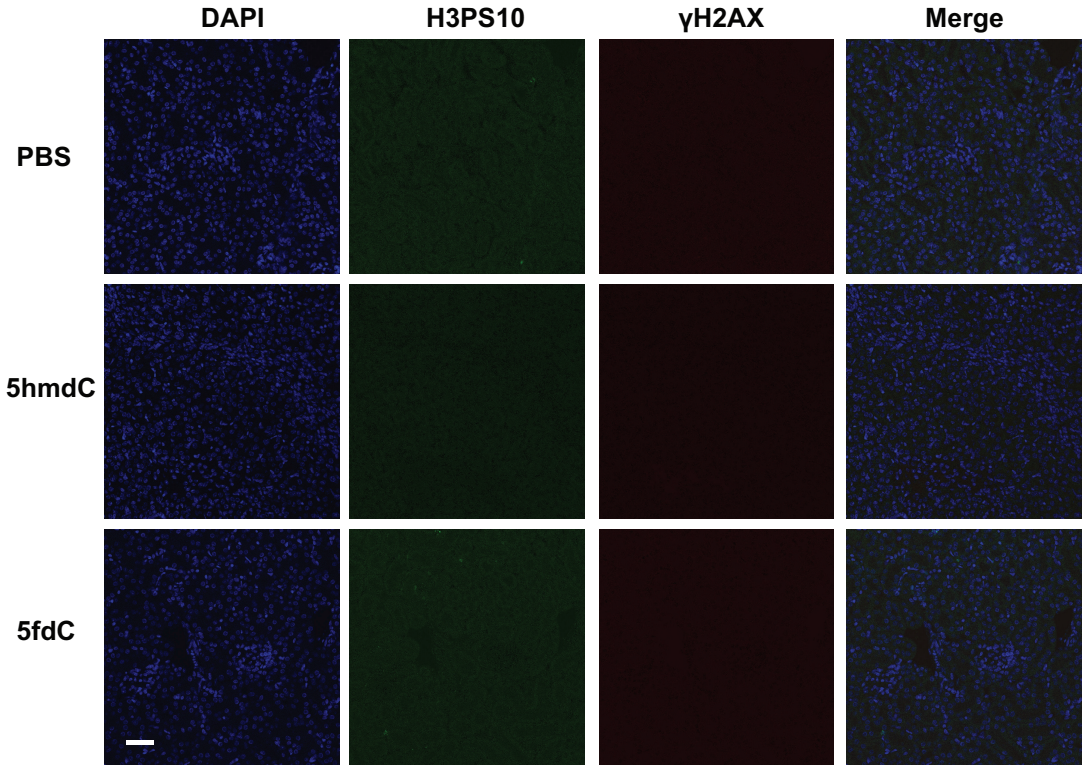


Figure 6.9: γH2AX in red and H3S10P in green in kidneys.(Scale bar=50 μm) (n =3 per group. Representative picture is shown).

Pancreas

In the pancreas, purine nucleotide biosynthesis and the salvage pathway have been studied [Meredith et al., 1995]. CDA expression levels are not elevated under normal conditions, but do became elevated in cancer [Bapiro et al., 2012]. Histopathology was assessed by hematoxylin and eosin staining and gross abnormalities were not detected upon 5hmdC and 5fdC administration. Intact islets were present in treated and untreated samples (Fig. 6.10). CDA was moderately expressed as shown by confocal microscopy (Fig. 6.11). DNA methylation and hydroxymethylation were measured in the DNA of the pancreas, but no significant difference was observed between treated and untreated mice (Figure 6.12).

Furthermore, when sections were stained for γH2AX and pH3S10P, any significant changes were not detected upon treatment. No substantial proliferation was detected

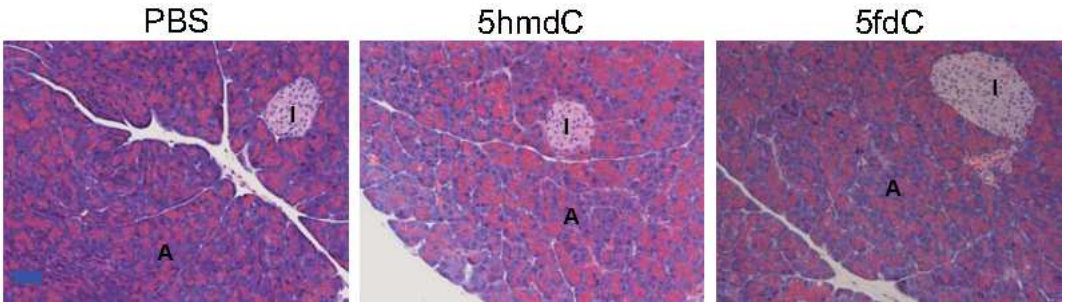


Figure 6.10: Hematoxylin and eosin staining of the pancreas upon PBS, 5hmdC or 5fdC administration (100 mg/kg). Exocrine pancreas constituted by glandular acini (A) and the endocrine pancreas constituted by islets of Langerhans (I) (Scale bar=50 μm) (n =3 per group. Representative picture is shown).

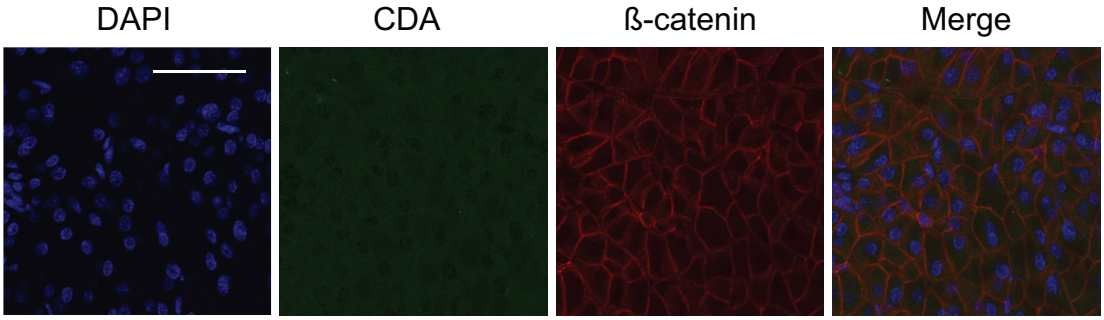


Figure 6.11: CDA staining in the pancreas.(Scale bar=50 μm) (n =3 per group. Representative picture is shown).

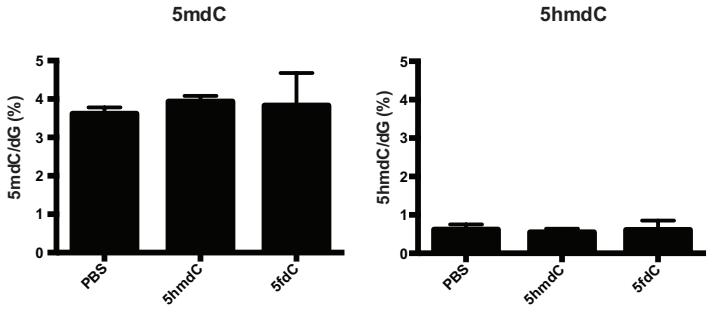


Figure 6.12: DNA methylation in the pancreas after treatment. Quantification is shown (n =3 per group, whiskers indicate standard deviation).

(Fig. 6.13).

Given the data presented, no overt toxicity was observed in the pancreas.

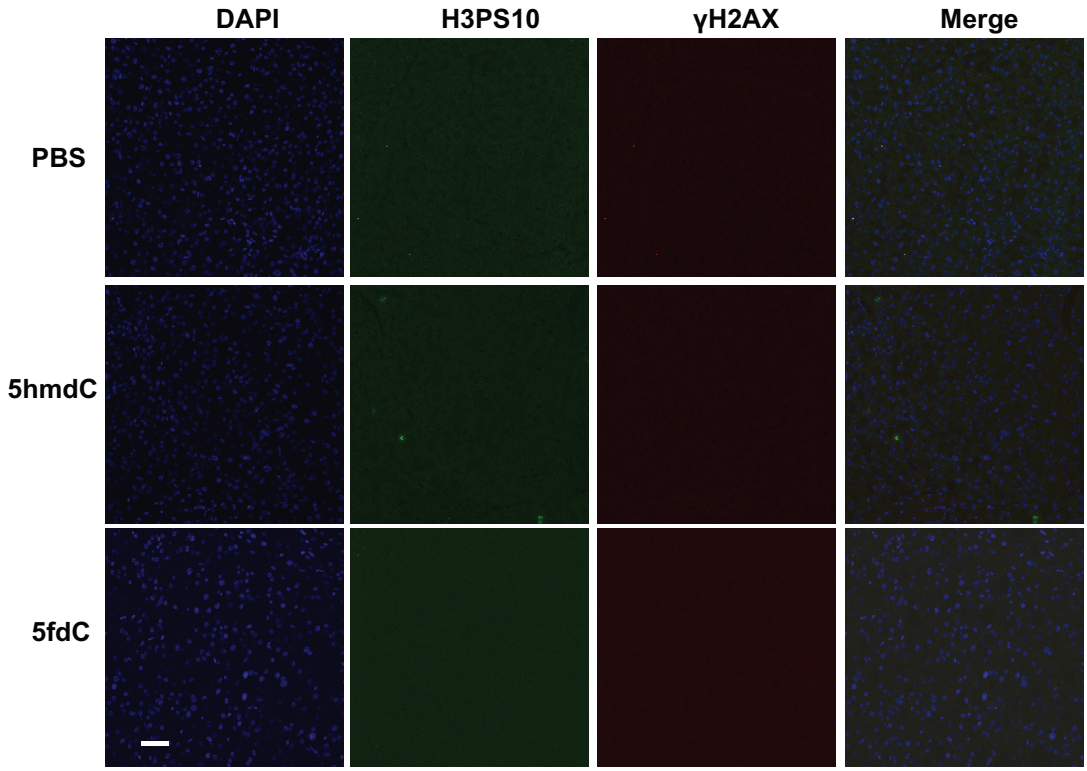


Figure 6.13: γH2AX in red and H3S10P in green in the pancreas (Scale bar =50 μm) (n =3 per group. Representative picture is shown).

Liver

It has been shown in primary cultures of rat hepatocytes that nucleotide salvage pathway enzymes are induced in the liver when it is stimulated to proliferate, even though they are expressed at basal levels in resting conditions [Mayer et al., 1990]. Histopathology was assessed by hematoxylin and eosin staining, and no significant difference was detected upon 5hmdC and 5fdC administration (Fig. 6.14), despite CDA being expressed as shown by confocal microscopy (Fig. 6.15).

In the DNA of the liver DNA methylation was measured, with no significant difference observed between treated and untreated mice (Fig. 6.16).

Furthermore, when sections were stained for γH2AX and H3S10P, a significant change was not detected upon treatment (Fig. 6.17). The low proliferation detected might coincide

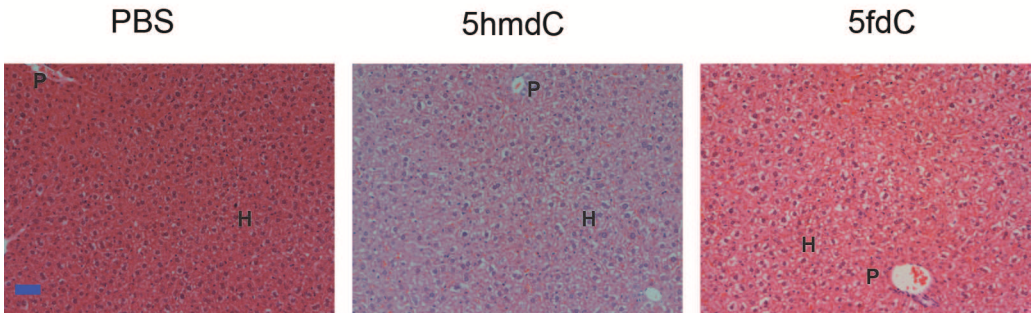


Figure 6.14: Hematoxylin and eosin staining in liver upon PBS, 5hmdC or 5fdC administration (100 mg/kg). Portal tract (P) and hepatocytes (H) are labeled (Scale bar =50 μm) (n =3 per group. Representative picture is shown).

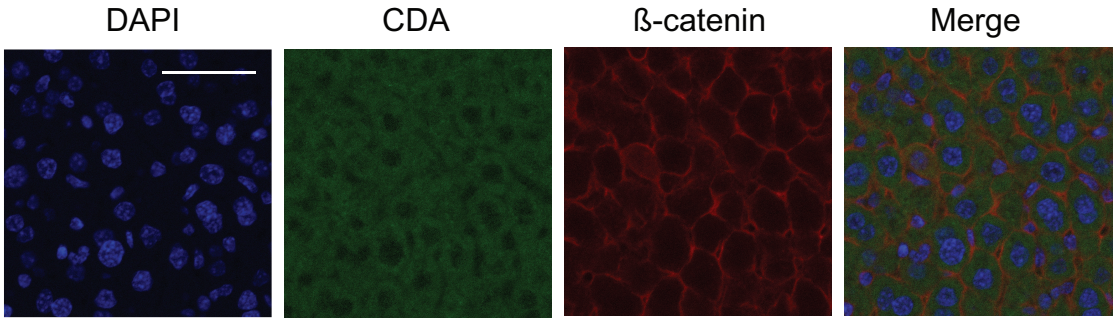


Figure 6.15: CDA staining in the liver (Scale bar=50 μm) (n =3 per group. Representative picture is shown).

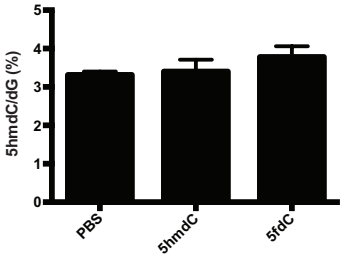


Figure 6.16: DNA methylation in liver after treatment. Quantification is shown (n =3 per group, whiskers indicate standard deviation).

with only basal activities of nucleotide salvage pathway enzymes, as previously shown [Mayer et al., 1990].

Given the data presented, no overt toxicity was observed in the liver.

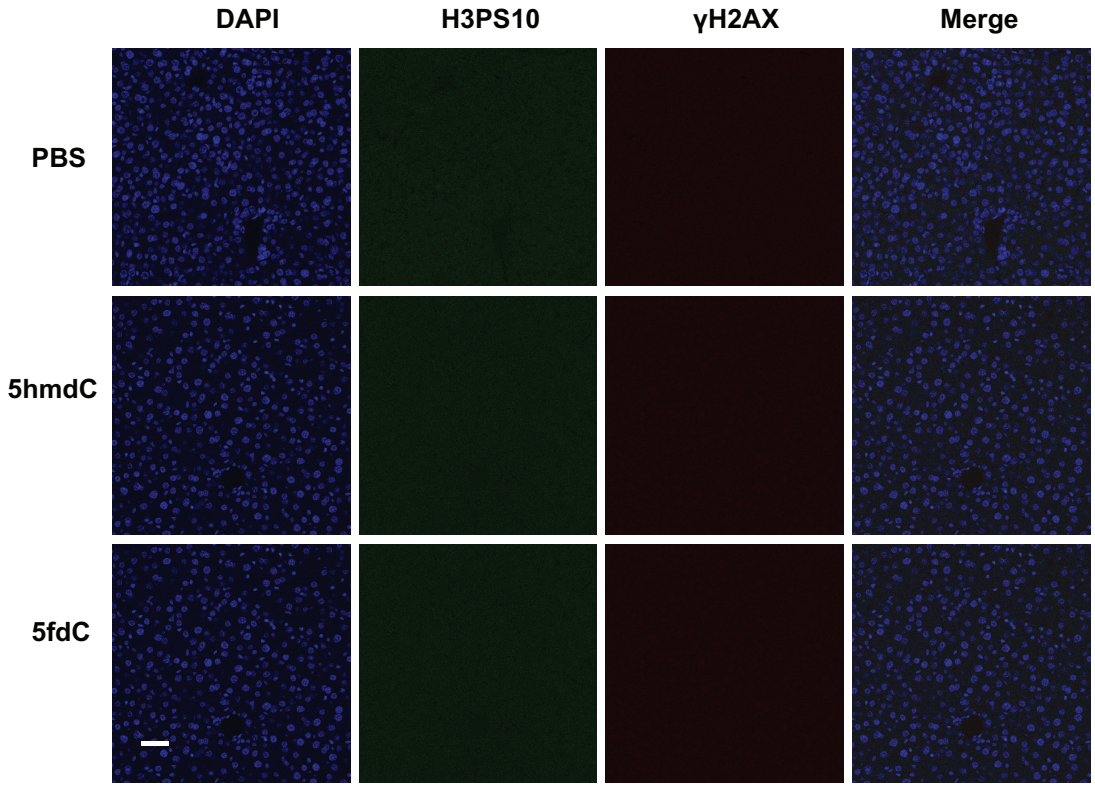


Figure 6.17: γH2AX in red and H3S10P in green in the liver.(Scale bar =50 μm). (n =3 per group. Representative picture is shown).

Lung

The lungs acted as a negative control due to low levels of Cda expression, but high vascularization, which could facilitate better delivery of the drug at the site of injection. Cytosolic pyrimidine salvage pathway enzymes such as TK and DCK are poorly active in this organ [Wang and Eriksson, 2010]. Histopathology was assessed by hematoxylin and eosin staining, and no gross abnormality was detected upon 5hmdC and 5fdC administration (Fig. 6.18). This was expected, since CDA expression is low, as shown by confocal microscopy (Fig. 6.19).

DNA methylation was measured, but no significant difference was detected between treated and untreated mice (Fig. 6.20).

Furthermore, when sections were stained for γH2AX and pH3S10P, there was no signifi-

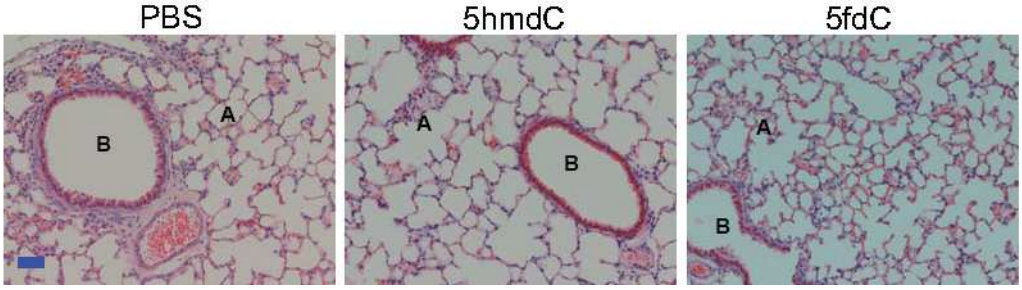


Figure 6.18: Hematoxylin and eosin staining in the lung upon PBS, 5hmdC or 5fdC administration (100 mg/kg). Bronchus (B) and alveoli (A) indicated (Scale bar =50 μm) (n =3 per group. Representative picture is shown).

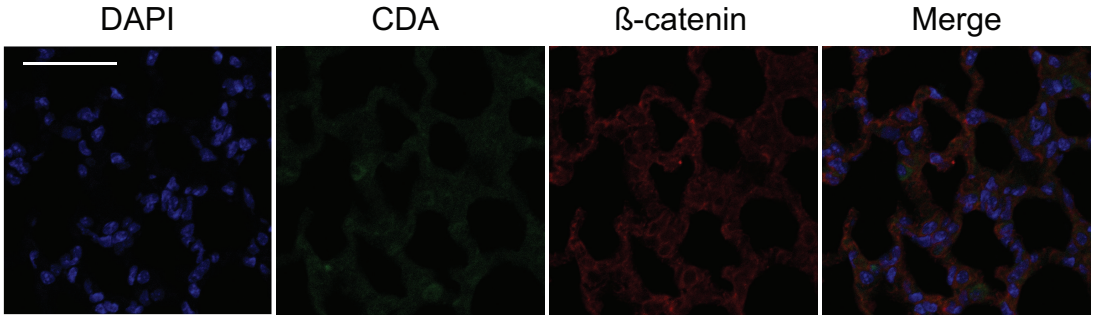


Figure 6.19: CDA staining in the lungs.(Scale bar =50 μm) (n =3 per group. Representative picture is shown).

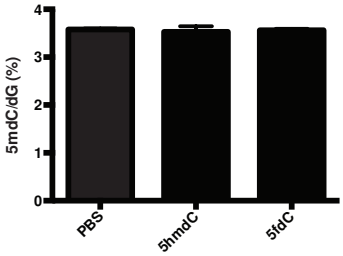


Figure 6.20: DNA methylation in the lungs after treatment. Quantification is shown (n =3 per group. Whiskers indicate standard deviation).

cant change upon treatment (Fig. 6.21).

Given the data presented, no overt toxicity was observed in the lungs.

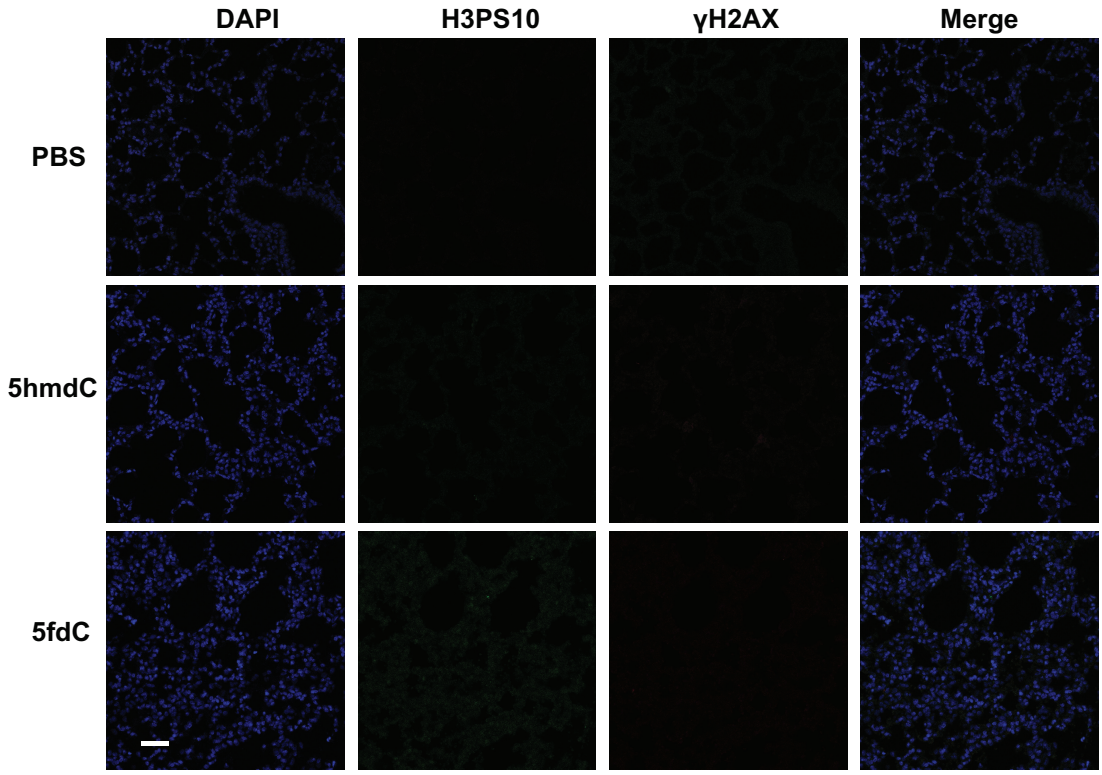


Figure 6.21: γH2AX in red and H3S10P in green in the lungs (Scale bar =50 μm) (n =3 per group. Representative picture is shown).

6.1.1.1 Brief summary of toxicology studies

In conclusion, no significant changes were detected, after the injection of 5hmdC and 5fdC, in mouse behavior and in the tissues analyzed for the parameters tested: proliferation, DNA damage or tissue abnormalities. Furthermore, no weight loss was observed over the time frame of the xenograft treatment window (Fig. 6.22).

6.1.2 Xenograft studies are feasible with the cell lines used for the *in vitro* experiments

Subsequently, it was necessary to assess whether the cell lines that were used for the earlier *in vitro* experiments would give tumors if implanted into nude mice. Moreover, tumors

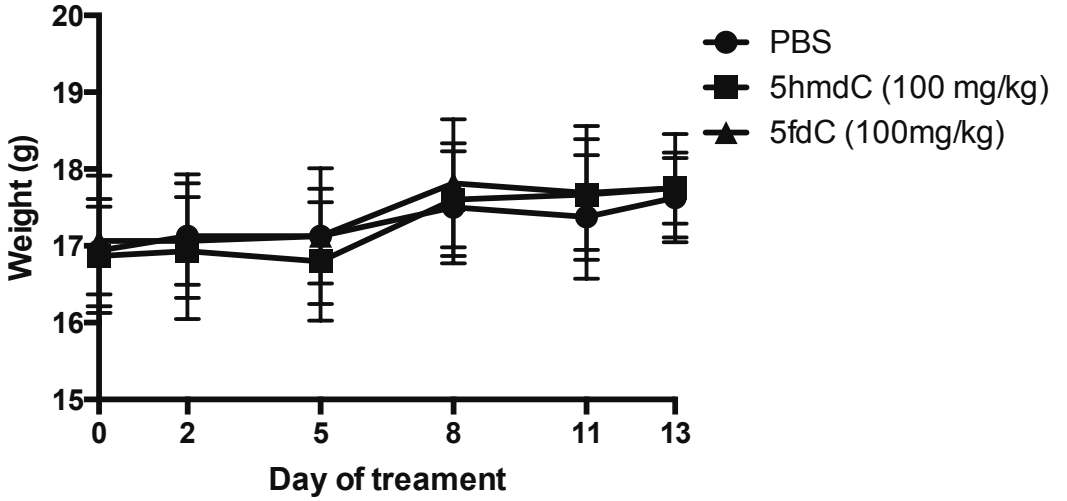


Figure 6.22: Mouse weight during the treatment with 100 mg/kg 5hmdC and 5fdC. (n= 16 per group. Whiskers indicates standard deviation).

needed to be obtainable in a time frame of twenty days, which would be reasonable for the compounds injections. Among the used cell lines, four were tested: MDA-MB-231, Capan-2, SN12C and H1299. They were selected according to published literature, which indicated that tumors could be obtained from them upon injection [Wang et al., 2004][Su et al., 2007][Zhong et al., 2012][Kadhim et al., 1997]. Subcutaneous tumors were obtained from the injected cell lines in the desired time frame of one month (Fig. 6.23).

SN12C and H1299 were chosen for further study, since knockdowns and *CDA*-overexpressing cell lines were already available for them, which had already been tested. This study allowed sample size for the next experiments to be calculated taking into account the variability observed in tumor growth and the difference among treated and untreated that would have been considered significant (30%). To observe a significant difference between treated and untreated mice, with this variability, we calculated that 8 mice were needed per treatment group (power analysis, p=0.05).

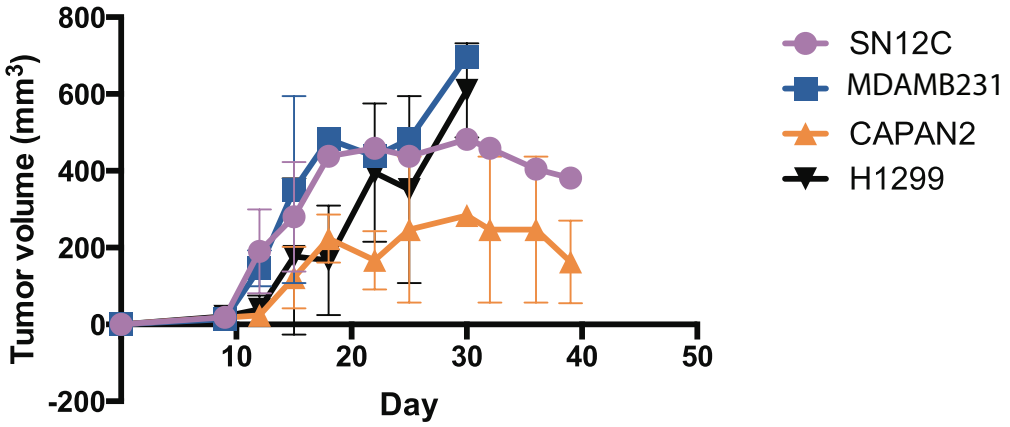


Figure 6.23: Cell lines tested for tumor formation (n =2 per group. Whiskers indicate standard deviation).

6.1.3 5hmdC and 5fdC can be detected in the bloodstream of injected animals

One of the parameters that pharmacokinetic studies evaluate is absorption, i. e. the entrance of the drug into the bloodstream. We decided to check this fundamental aspect of an *in vivo* drug study, to further understand whether the compound would reach the tumor via the bloodstream. Furthermore, the half-life of azacytidine is in the order of fifteen-thirty minutes in serum [Karahoca and Momparler, 2013]. Given the increased stability of our molecules at 37°C, we decided to check its presence in the blood thirty minutes after intraperitoneal injection. Therefore, aliquots of blood were taken from the tail veins of the mice and the presence of the drug measured by mass spectrometry for all the doses injected. 5hmdC and 5fdC were successfully detected in the blood at all of the dosages studied (Fig. 6.24).

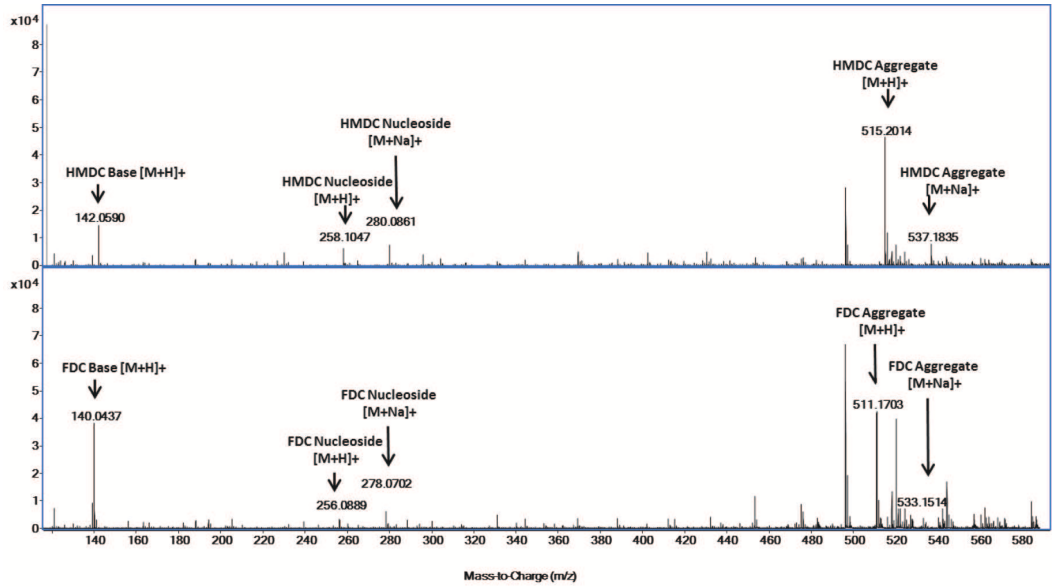


Figure 6.24: Mass spectrometry measurements of 5hmdC and 5fdC in the blood of injected animals (collaboration with B. Kessler laboratory). Figure prepared by G. Berridge.

6.1.4 Xenograft studies showed partial inhibition of tumor growth upon 5hmdC and 5fdC administration

Having set out the conditions for the xenograft study, the next step of the experiment was begun: the treatment of tumors injected into mice. This experiment would allow the efficacy of 5hmdC and 5fdC as anti-proliferative agents *in vivo*. It was started with the following cell lines:

- SN12C;
- SN12CshCDA_8, which showed the greatest *CDA* knock down efficiency;
- H1299;
- H1299dsRed_CDA.

Before injection, expression of CDA was validated by western blot (Fig. 6.25). Overexpression in H1299, and knockdown in SN12C, was confirmed.

For animal reduction reasons [Russell and Burch, 1959], it was decided that each flank

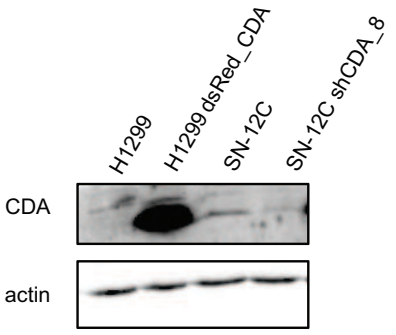


Figure 6.25: CDA expression in the cell lines chosen for the xenograft experiment.

of the animals used would be injected subcutaneously with different cell lines, in order to have the control of each cell line in the same animal. SN12C and H1299 were injected on the left side, while the knockdown and overexpressing cell lines were injected in the right. The intraperitoneal injections were done at 72 h intervals with 100 mg/kg of 5hmdC and 5fdC. Measurements were taken at the same interval, typically the day after each injection. Animals were sacrificed when the cumulative tumor size reached 1.2 cm in diameter, in accordance with the Home Office regulations (Fig. 6.26).

Treatment was started when tumors of palpable size were detected. The tumor volumes

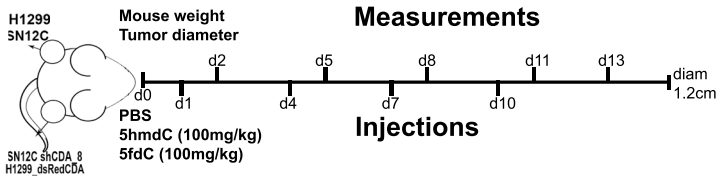


Figure 6.26: Injection scheme for the xenograft experiment.

of the cell lines injected showed a significant reduction in H1299dsRed_CDA and SN12C upon treatment with both 5hmdC and 5fdC (Fig. 6.27A-D). Tumor size was reduced when *CDA* was manipulated. In both H1299 and SN12C perturbation of *CDA* levels resulted

in tumors that were half the size of their wild type counterparts. This might indicate a strong dependency on the salvage pathway for growth and proliferation as previously published for the liver, where in hepatocytes it became highly active when they were induced to proliferate [Mayer et al., 1990]. When the tumors were weighted, no significant differences were detected in 5hmdC-treated tumors, but were detected for 5fdC in H1299dsRed_CDA, SN12C and SN12CshCDA_8 (Fig. 6.27 E-F). No significant weight difference was found between SN12C and SN12CshCDA_8 derived tumors treated with 5fdC. It might be speculated that, since the activity of *CDA* towards 5fdC is higher than that for 5hmdC, even the residual levels of this enzyme present in the knockdown might be sufficient to induce a response in SN12CshCDA_8 derived tumors. In this latter cell lines derived tumors, some of the 5fdC treated ones, were identified, post dissection as lymph nodes (Fig. 6.27 D) and thus excluded from further analysis.

After the experiment, western blot were run to detect CDA in some of the tumor samples; where expected CDA levels were detected. Additionally, tumors were genotyped to confirm the presence of the lentiviral construct and to rule out any possibility of tumor cross-contamination. PCR bands were only detected for tumors with the lentiviral vectors (Fig. 6.28). After dissection, tumors were subjected to the following analysis: histology (pathology and confocal microscopy studies) and DNA extraction (HPLC study). To be able to have enough material for the analysis mentioned, two groups were established with four animals each, since the size of some tumors (especially the cohort with 5fdC treatments) did not allow partitioning.

6.1.5 Histological features of the dissected tumors

Once dissected, the features of the tumors were assessed by hematoxylin and eosin staining. The first staining allowed the visualization of tumor characteristics such as mitosis, necrosis and apoptosis upon assessment from the pathologist (Fig.6.29 and 6.30). Masson's trichrome staining enabled characterization of another aspect of the tumors: fibrosis.

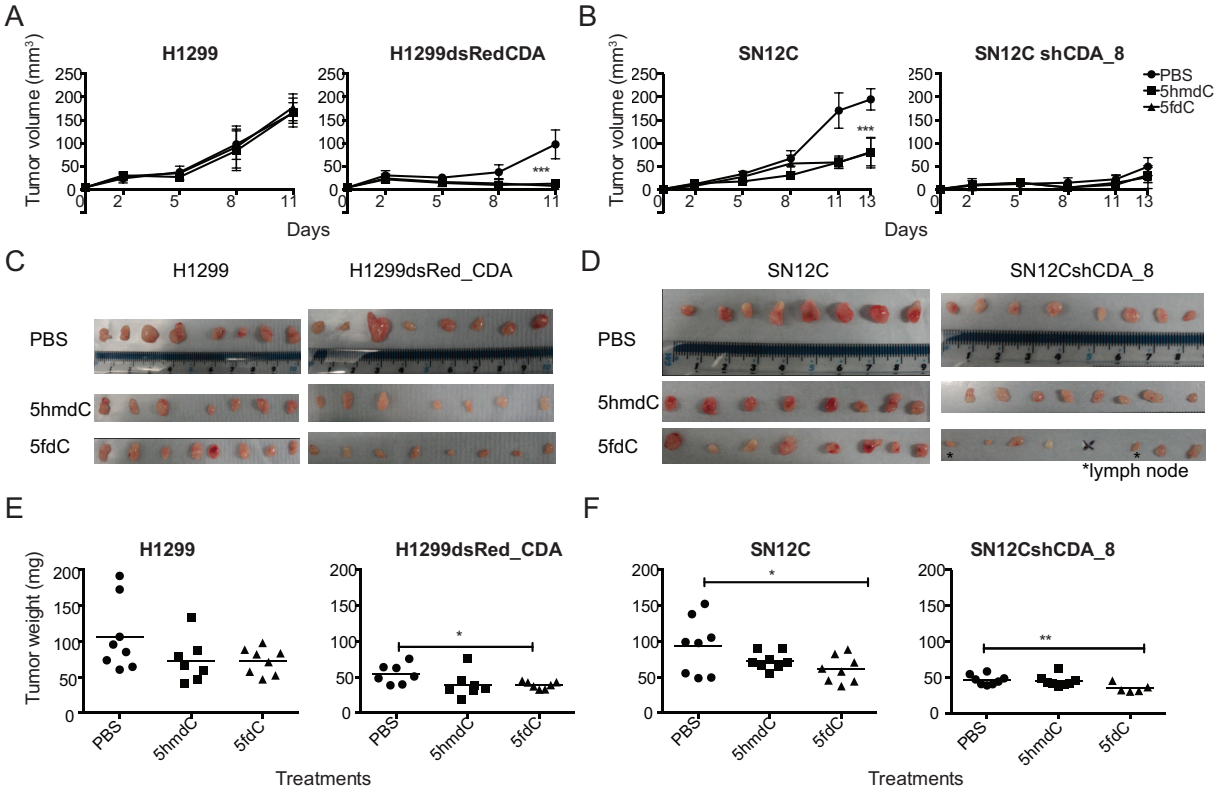


Figure 6.27: Tumor size measurements in the xenograft experiment. A. H1299 and H1299dsRed_CDA tumor volumes (mean with SD). Two-way ANOVA with repeated measures Holm-Sidak correction, p value < 0.0001 ($n = 8, 7, 8$ for PBS, 5hmdC and 5fdC groups). B.SN12C and SN12CshCDA_8 tumor volumes (mean with SD). Two way ANOVA with repeated measures Holm-Sidak correction, p value < 0.0001 ($n = 8$). C.H1299 and H1299dsRed_CDA dissected tumors. D.SN12C and SN12CshCDA_8 dissected tumors ($*$ indicates lymph nodes) E. H1299 and H1299dsRed_CDA tumor weights (Mean is shown; $n = 8, 7, 8$ for PBS, 5hmdC and 5fdC groups). ANOVA, p value $= 0.0431$ F. SN12C and SN12CshCDA_8 tumor weights.(Mean is shown;$n = 8$). ANOVA, SN12C p value $= 0.0230$, SN12CshCDA_8 p value $= 0.0099$.

The matrix is the scaffold that keeps the tumor in place, and fibrosis is what is left once the tumor regresses (Fig. 6.31 and 6.32). Although fibrosis was observed in the treated samples, the grading assessment from the pathologist (Dr. Goldin) did not result in a statistically significant increase, either due to the low sample number available for analysis or to absence of phenotype (Fig. 6.33).

Furthermore, H3PS10 and γH2AX were quantitatively assessed as markers for proliferation and DNA damage, in the central sections of the tumors that were scanned entirely

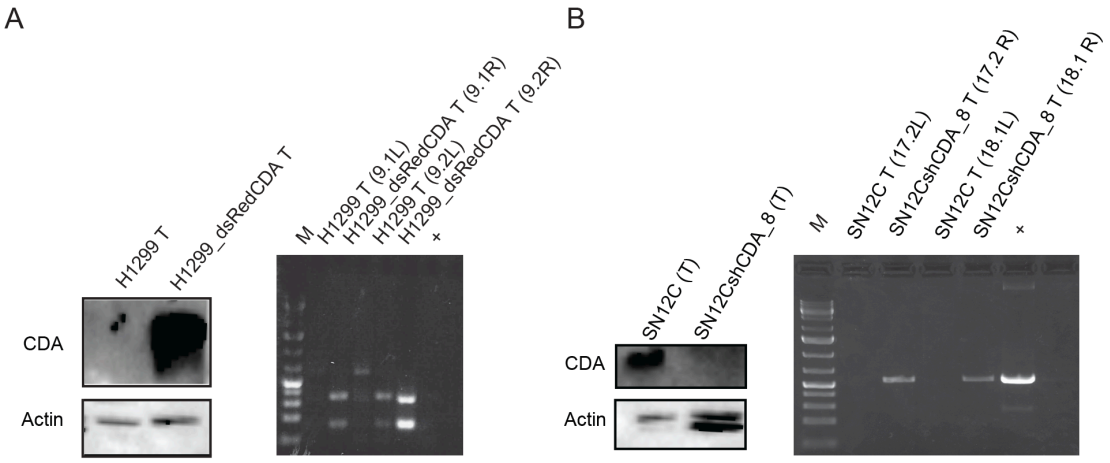


Figure 6.28: CDA expression in tumors from the xenograft experiment and genotyping. A. H1299 and H1299dsRed_CDA WB and genotyping; B. SN12C and SN12CshCDA_8 WB and genotyping. + indicates positive control fro genotyping.

by automated confocal microscopy. The number of stained cells was quantified (Fig. 6.34 and 6.35). Significant differences in proliferation were found in SN12C derived tumors. In particular, a decrease of 20% was noted in treated tumors, a two-fold increase in DNA damage in 5hmdC-treated tumors and of 50% in those 5fdC treated. The only significant difference detected in the *CDA* knockdown was a slight increase in proliferation in 5hmdC-treated tumors. This might be explained by cells trying to upregulate the salvage pathway during tumor growth, to efficiently recycle 5hmdC.

Significant differences in DNA damage were seen between the H1299dsRed_CDA treated and untreated tumors of about two-folds in 5hmdC treated tumors, and three-folds in 5fdC treated tumors. Furthermore, differences were detected in proliferation, but only in the 5fdC treated samples with a decrease of about two-fold (Fig. 6.35).

6.1.6 5hmdU is present in the DNA of 5hmdC-treated tumors

Following confocal microscopy analysis, where an increased percentage of DNA damage in 5hmdC and 5fdC-treated tumors was observed, it was decided to evaluate by HPLC

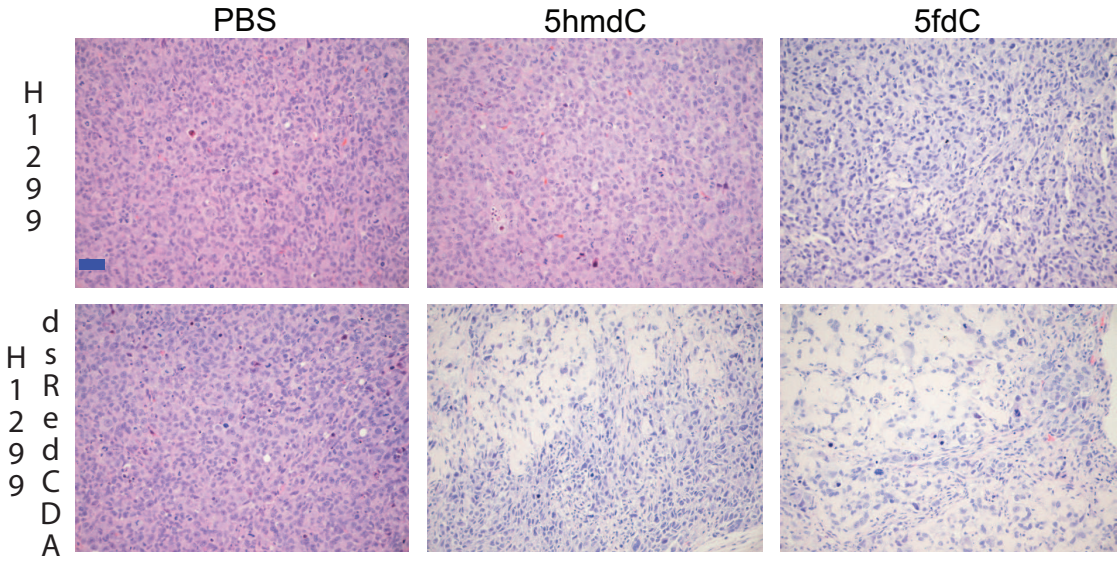


Figure 6.29: H&E staining of dissected tumors derived from H1299 and H1299dsRed_CDA. Nuclei in blue and, upon treatment, less dense nuclear regions and fibrotic areas (light pink) are visible. Representative pictures (n =4 per group).

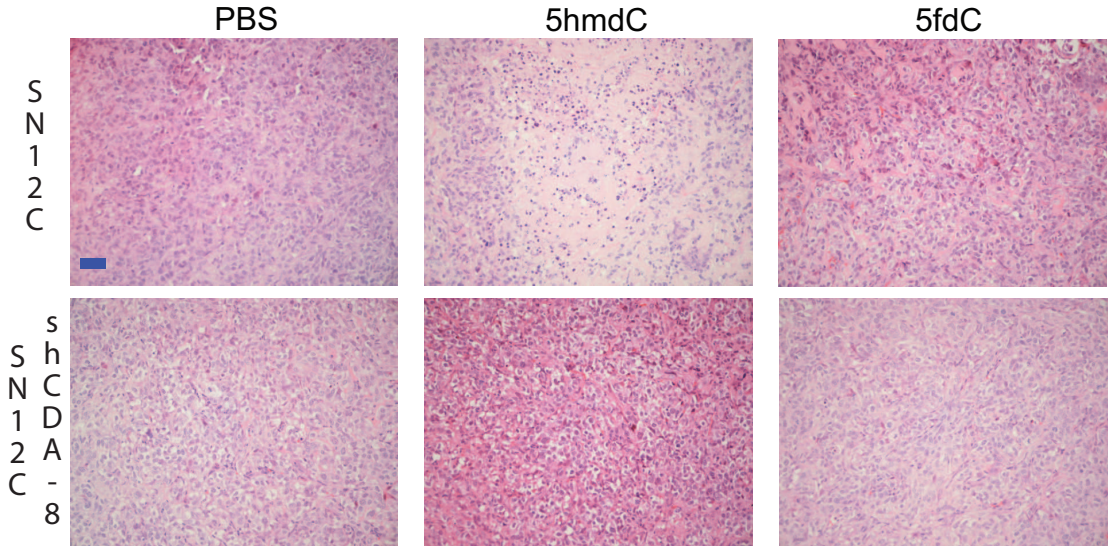


Figure 6.30: H&E staining of dissected tumors derived from SN12C and SN12CshCDA_8. Nuclei in blue and, upon treatment, less dense nuclear regions and fibrotic areas (light pink) are visible. Representative pictures (n =4 per group).

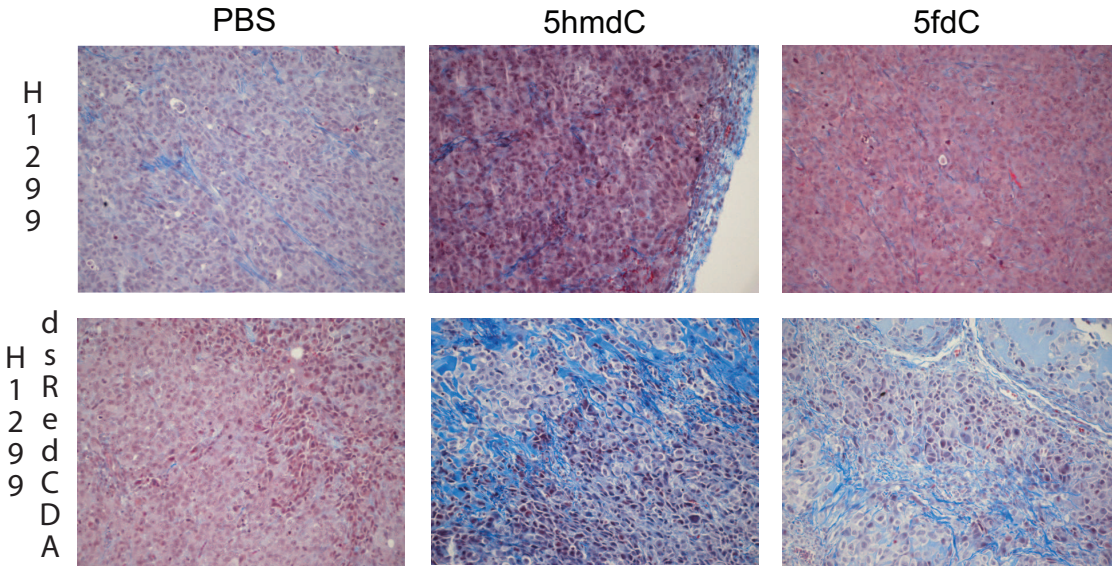


Figure 6.31: Masson's Trichrome staining of dissected tumors derived from H1299 and H1299dsRed_CDA. Nuclei in red and, upon treatment, less dense nuclear regions and fibrotic areas (blue) are visible. Representative pictures (n =4 per group).

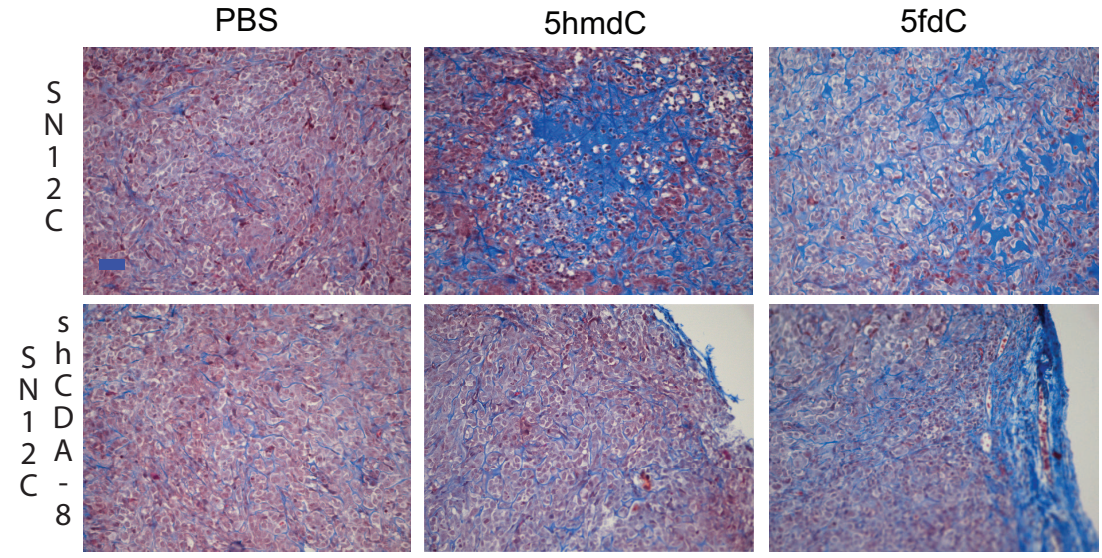


Figure 6.32: Masson's Trichrome staining of dissected tumors derived from SN12C and SN12CshCDA_8.Nuclei in red and, upon treatment, less dense nuclear regions and fibrotic areas (blue) are visible. Representative pictures (n =4 per group).

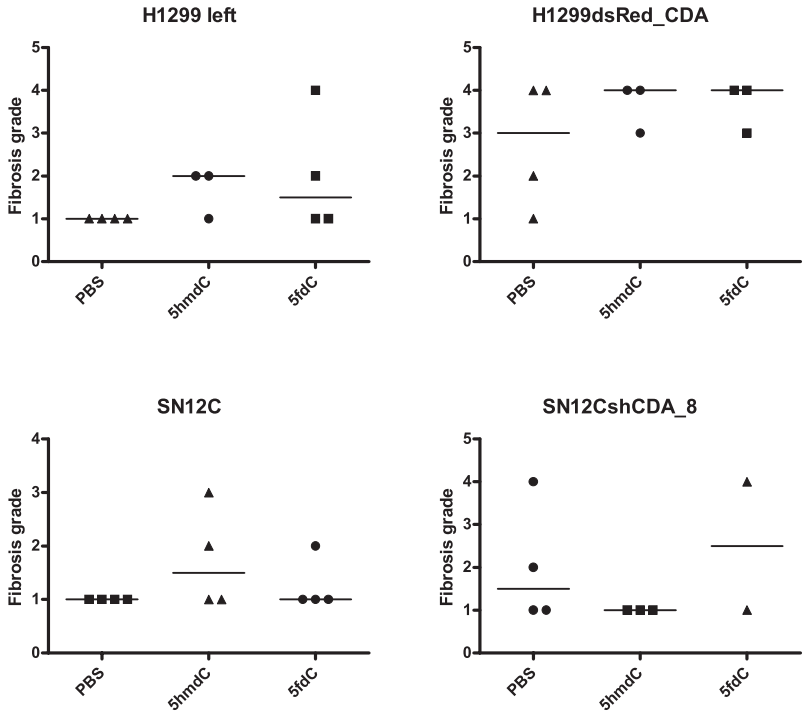


Figure 6.33: Fibrosis grading of dissected tumors (n =4 per group, except for the groups in which lymphnodes were excluded from the analysis). Bar indicates the mean.

the DNA extracted from the tumors for the presence of 5hmU. It was possible to detect 5hmdU in treated tumors (Fig. 6.36 and 6.37), but at smaller percentages compared to cell lines. It should be taken into account that the tumors were dissected at least three days after the last treatment and repair mechanism might have taken place, evident in the presence of increased γH2AX, induced to repair the accumulated 5hmdU which had not been enough to kill the cell.

6.2 Brief summary

After having assessed the feasibility of 5hmdC and 5fdC as possible therapeutic compounds *in vitro*, their feasibility was assessed as drugs *in vivo*. The first step was to evaluate whether they showed any sign of toxicity in single dose injected mice at the dosage window of 12.5 - 100 mg/kg. Having previously determined that their cytotoxic

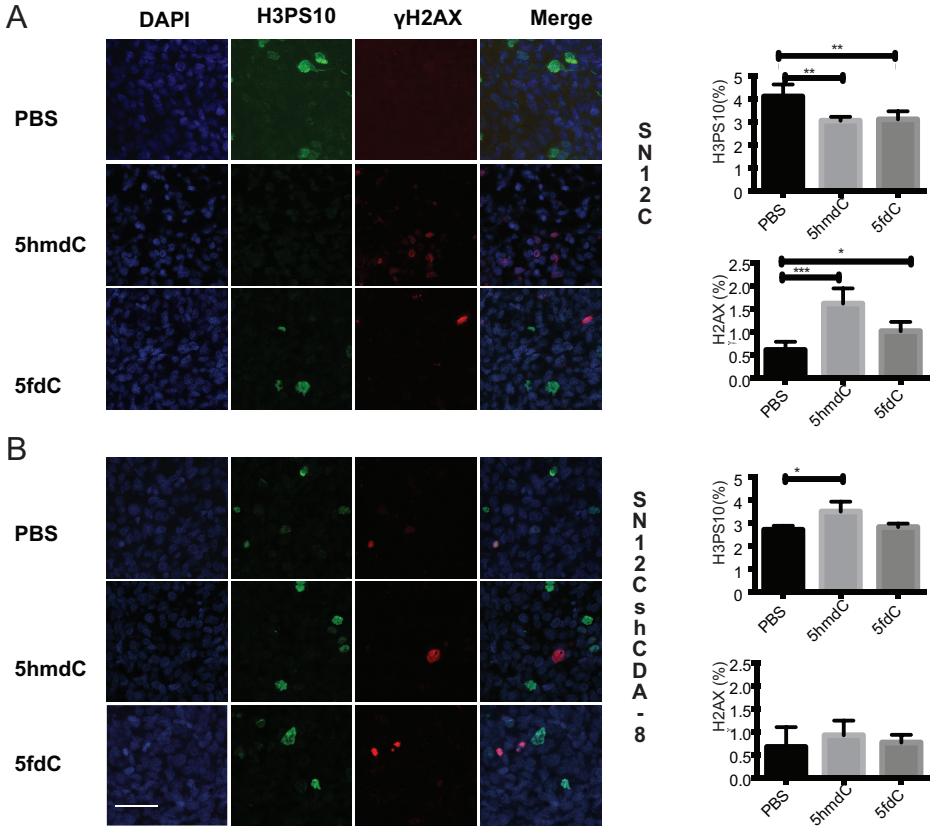


Figure 6.34: γH2AX and H3PS10 staining in dissected tumors: A. SN12C (A) and B. SN12CshCDA_8. (Scale bar = 50 μm). One way ANOVA, SN12C H3PS10 p value =0.0033 for PBS compared with 5hmdC and p =0.0046 with 5fdC, γH2AX p value =0.0003 for PBS compared with 5hmdC and p =0.0436 with 5fdC; SN12CshCDA_8 p =0.0130 for PBS compared with 5hmdC. Whisker bars indicate standard deviation (n =4 per group).

activities were dependent on CDA expression, we decided to examine, in addition to the general wealth of the animal, *post mortem* the tissues expressing considerable levels of CDA. Data mining of gene expression studies (Fig. 6.1) identified the mouse tissues with the highest CDA expression: small intestine, kidneys, liver and pancreas. Lung was used as a negative control due to its low CDA expression levels. None of the tissues analyzed showed histological features of toxicity, with the main structures preserved in all the organs. Furthermore, none of the DNA of the treated tissues showed any presence of 5hmdU in the DNA, or any increase in DNA damage and proliferation.

Since 5hmdC and 5fdC display good solubility in the aqueous environment, they were

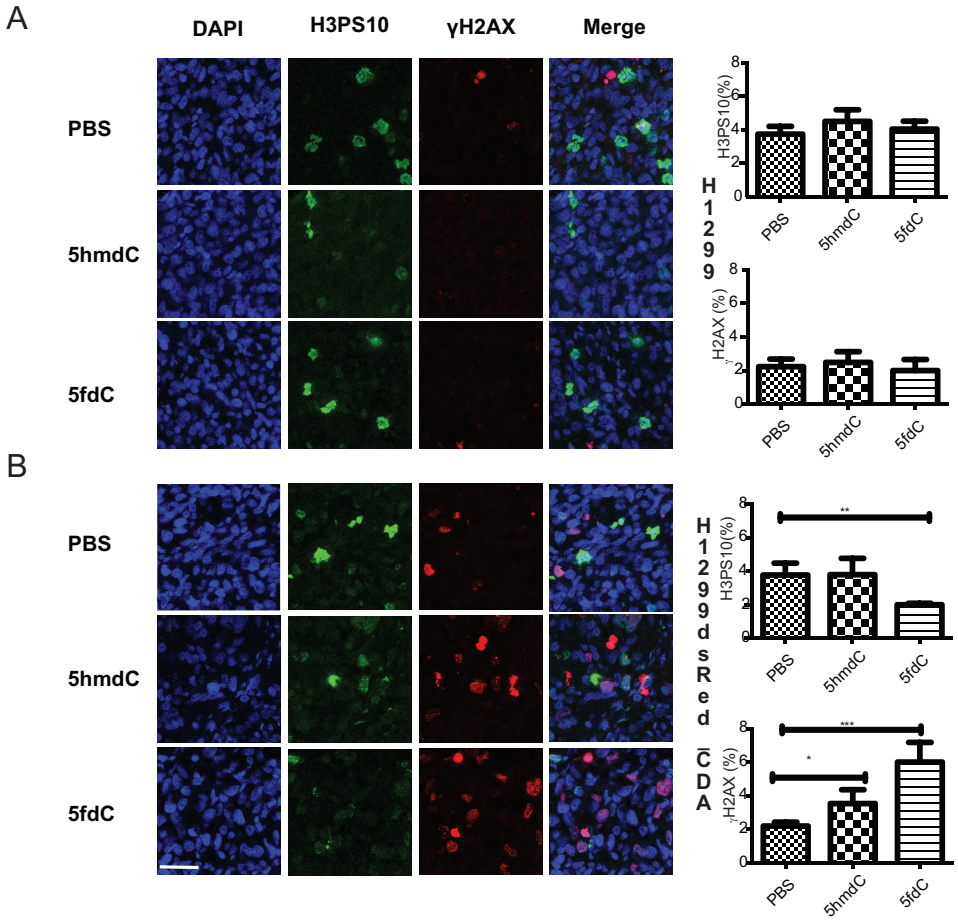


Figure 6.35: γH2AX and H3PS10 staining and relative quantification in dissected tumors: A. H1299 (A) and B. H1299dsRed_CDA(Scale bar = 50 μm). One way ANOVA, H3PS10 p value =0.0057, γH2AX p value $(*)$ =0.0491, p value$(***)$ =0.0001. Whisker bars indicate standard deviation (n =4 per group).

administered by intraperitoneal injection. Therefore, another parameter to define for a drug, according to pharmacological conventions, is the absorption, i.e. the presence of the drug in the bloodstream. Blood samples were collected thirty minutes after injection for all the dosages tested and the presence of the drug assessed by mass spectrometry. Both 5hmdC and 5fdC were detected in the blood of treated animals indicating that they were distributed in the body and that they could reach the site of the tumor.

Different options were available to test the tumor anti-proliferative potential of the compounds *in vivo*. Patient derived primary cells were excluded due to difficulties with their

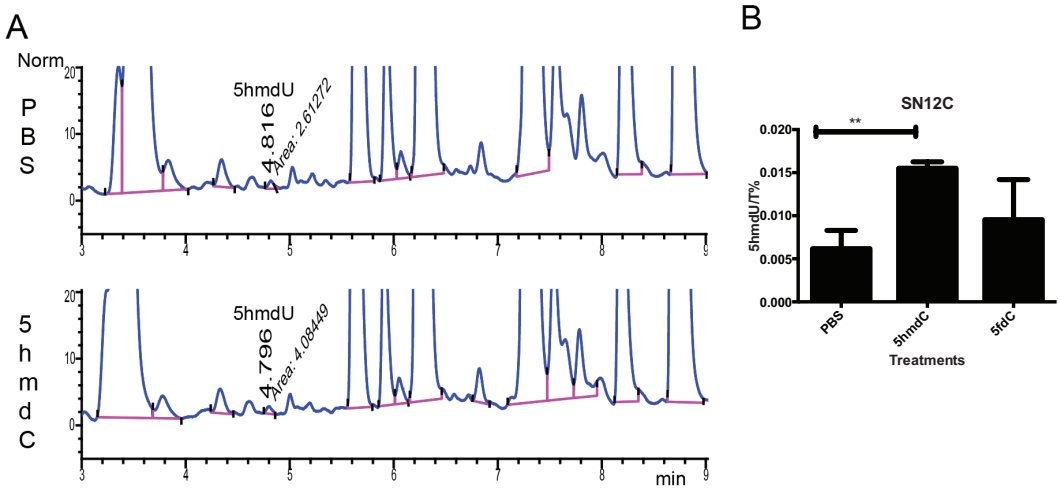


Figure 6.36: 5hmdU levels in DNA of SN12C tumors treated with 5hmdC or PBS: HPLC chromatograms (Norm = absorbance units). Highlighted is 5hmdU with relative quantification (n =4 per group, whiskers bars indicate standard deviation, one way ANOVA p =0.0041).

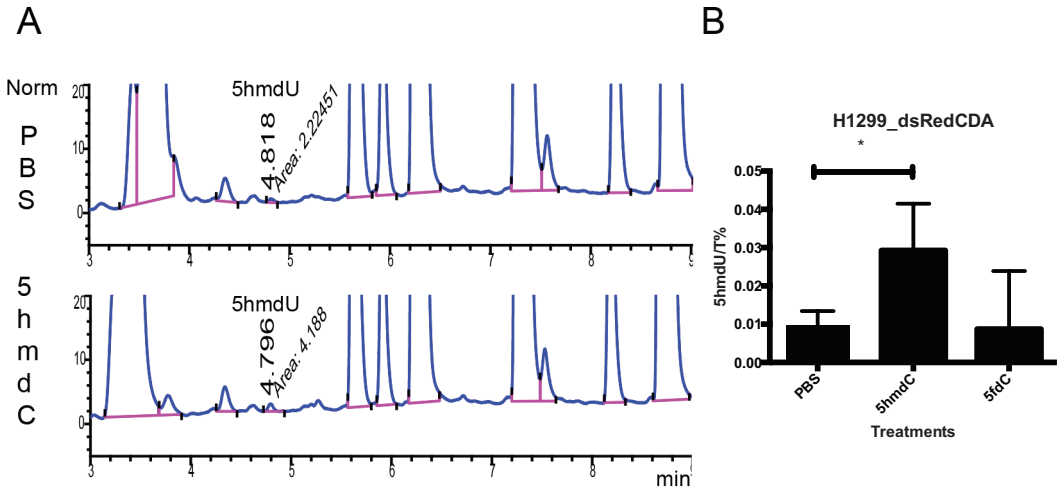


Figure 6.37: 5hmdU levels in DNA of H1299dsRed_CDA tumors treated with 5hmdC or PBS: HPLC chromatograms (Norm = absorbance units) and relative quantification. Highlighted is 5hmdU with relative quantification (n =4 per group, whiskers bars indicate standard deviation, one way ANOVA p =0.0395).

availability and necessity of characterization for CDA expression levels. Therefore, two possibilities remained such as: treatment of a suitable genetic mouse model of cancer, or the induction of xenografts by injection of the previously tested cancer cell lines. A suitable genetic tumor mouse model for use was the pancreatic tumor model KPC, gen-

erated by the Tuveson laboratory, which carries specific expression of endogenous mutant Kras and p53 alleles and overexpresses cytidine deaminase (CDA) [Bapiro et al., 2012]. It has been shown that nanoparticle albumin bound-Paclitaxel (nab-Paclitaxel) increases the availability of gemcitabine (dFdC), by suppressing the endogenous CDA overexpression in the KPC mouse model [Bapiro et al., 2012]. Gemcitabine is a substrate for CDA that converts it into the inactive metabolite deoxy-fluorouridine. This model was excluded for the first *in vivo* test due to limited availability. Therefore human cell lines derived xenografts in NOD/SCID mice were chosen as the preferred model for the next set of experiments with the drawback of lack of immune response that has already been documented in the literature as being disadvantageous for this set of studies [Jung, 2014].

The same cell lines were used that were sensitive and insensitive to 5hmdC and 5fdC earlier in *in vitro* studies. Having evaluated that they could produce tumors, the experiment was performed with SN12C and SN12CshCDA_8, and H1299 and H1299dsRed_CDA, which were injected into both flanks of nude mice and the treatment performed (injection scheme in Fig. 6.26). In total, 4 injections of 100 mg/kg of 5hmdC and 5fdC were administered to each animal. Significant differences were detected in the tumor volume of treated animals, both with 5hmdC and 5fdC, versus untreated in H1299dsRed_CDA and SN12C-derived tumors. No significant difference was detected in the weight of tumors in 5hmdC-treated animals in the same cell lines. This discrepancy might be explained by the fact that the measure of the volume might be less accurate, due to the assumption that the tumors were spherical. 5fdC treatment, which according to the *in vitro* data has greater potency, gave significance in H1299dsRed_CDA, SN12C and SN12CshCDA_8 derived tumors. Significance in the latter case might be explained by the high turnover rate of 5fdC by CDA and by the residual levels of the protein present in the tumor. An increase in γH2AX was detected in treated samples by quantitative confocal microscopy and a decrease in proliferation in H1299dsRed_CDA (only in 5fdC-treated tumors) and in SN12C, although this was less pronounced. Moreover, the presence of 5hmdU was checked in the DNA of

treated tumors and was detected at significant levels, relative to PBS and 5fdC-treated samples, in both H1299dsRed_CDA and SN12C-derived tumors.

In conclusion, the drugs showed potential for further development studies *in vivo*, both for their availability in the bloodstream and for their low toxicity.

Chapter 7

TET2 protein interaction network

Concurrent with the discovery of the presence of 5hmC in the DNA [Kriaucionis and Heintz, 2009][Tahiliani et al., 2009], the enzymes responsible of the conversion of 5mC to this new variant were identified [Tahiliani et al., 2009]. They belong to the TET family of iron and α-ketoglutarate (α-KG) dependent oxygenases, which in mammals has three members: *TET1*, *TET2* and *TET3*. One of them, Tet2, has been found extensively mutated in cancers, in multiple tissue types such as: bone, breast, central nervous system, endometrium, hematopoietic and lymphoid systems, kidney, large intestine, liver, lung, ovary, prostate, skin, alongside the upper aerodigestive and urinary tracts. However, the most prevalent mutations are in hematopoietic cancer types such as lymphomas and myeloproliferative neoplasms [Delhommeau et al., 2009]. Moreover, a causative link between myeloproliferation and Tet2 loss was established in mice [Moran-Crusio et al., 2011][Quivoron et al., 2011]. However, a detailed understanding of the function of TET2 is still needed to clarify the role of this gene in the formation of blood lineages. The large spectrum of mutations observed in myelodysplastic syndromes results in different molecular phenotypes, such that a strict correlation between *TET2* mutation and the global amount of 5hmC is not always seen. In fact, it has been shown that not only

patients with mutated *TET2*, but also some patients with wild type *TET2*, display low 5hmC levels in the DNA [Ko et al., 2010]. The few studies that tried to shed light on this apparent dilemma have reported mutations upstream of *TET2*. One example is represented in *IDH* mutant cancers, the enzyme producing the co-factor α-KG for the TET2 catalyzed reaction, which result in the same phenotype as *TET2* mutant ones [Figueroa et al., 2010]. Indeed, it has been shown that the production of 2-hydroxyglutarate by neomorphic IDH1/2 mutant proteins inhibits TET2's catalytic activity [Figueroa et al., 2010]. New avenues to try to explain this discrepancy and shed light on the function of TET2, might come from the study of its protein interaction partners.

At the start of the this research project no interactions had been published for TET2. Later, TET2 has been shown to interact with a number of proteins. The first was OGT [Chen et al., 2012b][Vella et al., 2013][Deplus et al., 2013], an O-linked N-acetylglucosamine transferase, which is able to glucosylate histone H2BS112 and to mediate the ubiquitination of the neighboring residue lysine 120 (H2BK120) by the BER1 enzymes [Fujiki et al., 2011], a modification linked to the recruitment of the SET/COMPASS complex and the subsequent production of H3K4me3 promoting the activation of gene expression [Lee et al., 2007]. Following the discovery of OGT, a number of other interacting partners have been found such as:

- IDAX, which is a CXXC domain-harboring protein, responsible for TET2 binding to DNA and, interestingly, it has been found to induce TET2 caspase-dependent degradation upon stem cell differentiation [Ko et al., 2013];
- EBF1 [Guilhamon et al., 2013], a transcription factor involved in B-cell differentiation [Treiber et al., 2010] that can contribute to DNA demethylation at certain gene promoters [Li et al., 2010];
- NANOG with whom it mediates the establishment of pluripotency in embryonic stem cells [Costa et al., 2013];

- PU.1 , which is a transcription factor involved in osteoclast [de la Rica et al., 2013], myeloid and lymphoid differentiation and alternative splicing [Guillouf, 2006], with whom promotes differentiation;
- PRDM14 with whom it participates in the establishment of DNA demethylation at pluripotency-related genes [Okashita et al., 2014];
- BER factors, like TDG, MBD4, SMUG1, NEIL1, NEIL2, NEIL3, PARP1, LIG3 and XRCC1 with whom it mediates demethylation [Müller et al., 2014].

The decision was made to dissect the protein interaction network of TET2 to gain insights on possible targeting mechanisms to chromatin or other potential regulatory mechanisms on 5hmC deposition. To do so, a yeast two-hybrid screen was initially adopted, subsequently followed by mammalian validation strategies of a selected candidate.

7.1 Results

In this section, the yeast two-hybrid data will be presented first, followed by the studies of protein interaction performed in mammalian cells.

7.1.1 A yeast two-hybrid screen to identify protein interaction partners of TET2

A fundamental consideration when studying protein interaction networks is tissue specificity. The presence of loss of function mutations in specific tumor sub-types suggested a functional link between the mutated gene and the tissue. Therefore, a technique was needed that would allow the study of the the TET2 interactome in bone marrow, where most of the mutations occur and where the expression of *TET2* is elevated [Su et al.,

2004][Wu et al., 2009] (Fig. 7.1).

To perform this experiment in a relevant tissue setting, the yeast two-hybrid technique

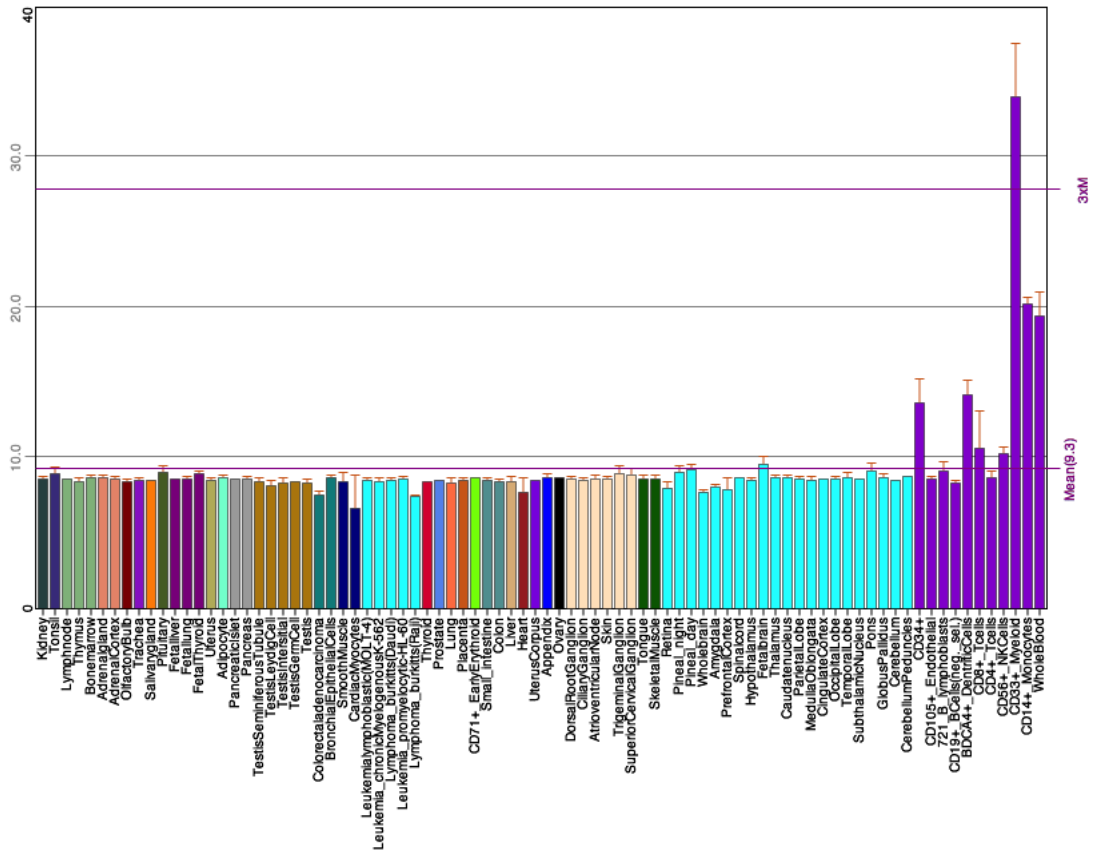


Figure 7.1: Tet2 expression in human tissues (downloaded from BioGPS database [Wu et al., 2009]).

was adopted, which relies on the modular nature of the GAL4 transcription factor that contains two domains: one for DNA binding and one for transcriptional activation, both necessary for the function of the protein [Fields and Song, 1989]. Fundamental for the technique is the generation of two hybrid proteins, between the candidates being studied and the modules of GAL4, to generate "bait" and "prey". If productive interaction occurs there will be a transcript, encoding one or more reporter proteins, which can be easily screened. Key advantages of the yeast two hybrid assay are its sensitivity and the flexibility. The sensitivity allows for the identification of weak and transient interactions,

because of the overexpression of the proteins, their compartmentalization in the nucleus where the interactions are monitored, and their amenability to high throughput, enabling the assessment of large amounts of candidates at the same time. The other advantage is the flexibility that gives the choice of the context in which to interrogate the interaction by allowing the use of libraries of prey candidates from certain tissues. A disadvantage of the yeast two-hybrid assay is the high number of false interacting partners that it can give. To minimize this, in addition to the selective media used to grow bait and prey strains (DDO: double drop out), in the two-hybrid system chosen there were four reporter genes, under the control of three distinct Gal4-responsive promoters. They were:

1. AUR1-C: encodes the enzyme inositol phosphoryl-ceramide synthase, which confers resistance (AbAr) to the otherwise highly toxic drug Aureobasidin A (AbA);

2. MEL-1: encodes α-galactosidase, which is secreted by the yeast cells, which turn blue in the presence of its chromagenic substrate X-α-Gal[1];

3. HIS3: when bait and prey proteins interact, Gal4-responsive His3 expression permits the cell to synthesize histidine and grow on single drop out for His (SD/His) medium;

4. ADE2: when two proteins interact, Ade2 expression is activated, allowing these cells to synthesize adenine and grow on single drop out for adenine (SD/Ade) medium.

The longest protein for which a successful yeast two-hybrid screen had previously been developed with the system used in this study (Matchmaker, Clontech) was 750 aminoacid (AA) long (datasheet). Therefore, to enhance the possibility of success with this screen, TET2 (which is 2002 AA long) was split into two fragments based on sequence conservation between species (ClustalW) and secondary structure prediction (Psipred) (Appendix 2 Fig. 6, 7). Consequently, TET2 was split at residue 733, which does not lie in a conserved region and does not participate in secondary structure formation according to prediction with

[1]X-α-Gal is not X-Gal, and is not a substrate for β-galactosidase.

Psipred. In 2013, the first protein structure of TET2's catalytic domain was released [Hu et al., 2013]. This was checked against our own predicted secondary structure and there was good agreement (Appendix 2 Fig. 8). The fragments generated tagged with GAL4 binding domain (bait) were: TET2A, covering the first 732 AA; and TET2B, covering the last 1269 AA. To exploit the flexibility of the yeast two-hybrid screen, a bone marrow derived prey library was used. The inaccessibility of human bone marrow and the samll amounts of mouse bone marrow make them unsuitable for large proteomic studies.
Several things needed to be checked before proceeding with the screen:

- expression of the constructs containing TET2A and TET2B upon transduction in yeast;
- toxicity of the expressed constructs;
- autoactivation of the expressed constructs.

Expression of the constructs: *TET2A* and *TET2B*

The expression of the two fragments in the yeast strain used to transform the bait (Y2HGOLD) was checked by PCR from the complementary DNA (cDNA). Bands were detected for both constructs, suggesting that they are expressed at the RNA level (Fig. 7.2).

Testing for toxicity of the constructs: *TET2A* and *TET2B*

To assess whether the expression of the constructs TET2A and TET2B induced growth toxicity, the transformants were grown on non selective complete (YPDA) and selective single drop out (SDO) media plates. In particular, TET2 containing vectors conferred auxotrophy to the transformants for the lack of tryptophan (Trp), thus they were able to grow on SD-Trp medium. By plating serial dilutions, the growth rates were compared of

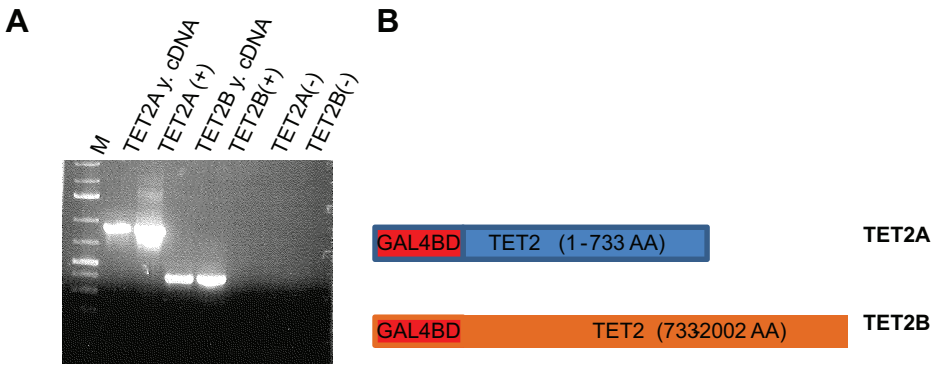


Figure 7.2: Expression of the constructs: *TET2A* and *TET2B* checked by PCR from cDNA. B. The constructs are schematized.

different yeast transformed for: empty vector, TET2A, TET2B, positive (pGBKT7p53, PGADT7-T) and negative control (pGBKT7-Lam, pGADT7-T) (Fig. 7.3). A growth

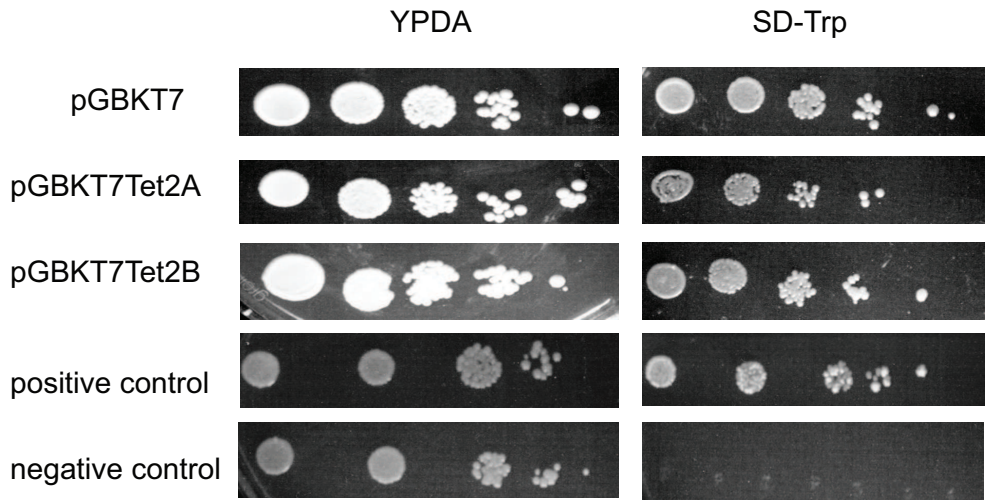


Figure 7.3: Testing the expressed constructs for growth toxicity. YPDA complete media; SDO/-Trp Single drop out media lacking tryptophan.

inhibition effect was not observed for the TET transformants on YPDA as compared to the controls, over a period of five days. On selective media, TET2A showed a mild phenotype compared to empty vector, as in the last serial dilution there were no colonies and in general there were less colonies across all the tested dilutions. This was not assessed to be a significant growth delay and the next phase of the screen was begun.

Testing for autoactivation of the constructs: *TET2A* and *TET2B*

Subsequently, it was necessary to assess whether the constructs were sufficient to activate the expression of the reporters, to determine whether they would be suitable for the next phases of the screen.

TET2A and TET2B were screened for auto-activation on the first two reporters, AUR1-C and MEL-1, that were the same two that would be screened first in a low stringency version of the screen. The colonies should have been able to grow on SD-Trp/X-α-Gal but not turn blue, because MEL-1, which is one of the reporter genes, should not be expressed without bait and prey interaction. The second reporter that the system provided to assess true positives was resistance to AbA. In this case, the colonies should not encode for AUR1-C, therefore would not be able to survive in the presence of the toxic compound (AbA). On DDO-Leu-Trp/X-α-Gal/AbA plates, they would not be able to survive due to a lack of leucine, which they cannot complement for. In addition, negative and positive control experiments were performed with the domains of proteins known to interact (p53/T antigen) or not (p53/lamin) on the same medium with which the constructs were observed (Fig. 7.4). No growth impairment or blue colonies were observed on SD-Trp/X-α-Gal for

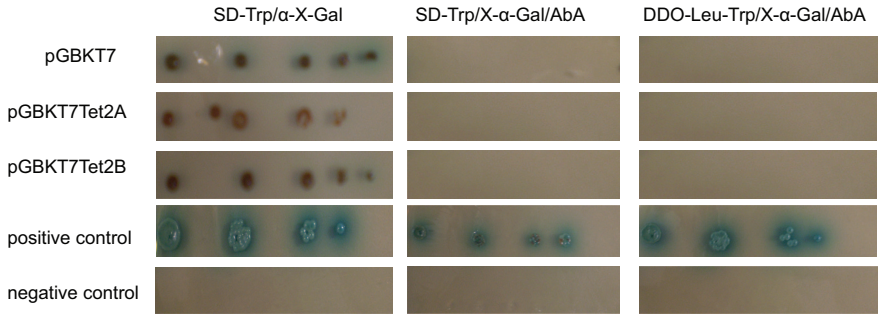


Figure 7.4: Controls for the expressed constructs TET2A and TET2B for autoactivation. (SD-Trp) single drop out for tryptophan; X-α-Galactose (X-α-Gal); aureobasadinA (AbA).

the transformed yeast. No colonies grew on the other selections, confirming that there was no autoactivation of TET2A and TET2B, thus they were used for the screen.

7.1.1.1 Mating identifies candidate proteins

Having established that the constructs were expressed, not toxic and did not autoactivate, the next phase of the yeast two-hybrid assay was begun which was the mating of prey and bait. The bait carrying strains TET2A and TET2B (a mating type) were mated with a library, derived from transcripts expressed in human bone marrow, of prey-transformed strains (α mating type).

The mating was judged successful by the formation of diploid colonies. The efficiency overcame the threshold of 2% and one million diploids to be screened for a screen to be judged successful (Table 7.1).

Following the mating, colonies were plated on both DDO/X/A media and quadruple drop

Bait	N of screened clones ($*10^6$)	Mating efficiency
TET2A(1-732AA)	12.4	4.3%
TET2B(733-2002AA)	21.96	10%

Table 7.1: Mating efficiency and number of screened clones for the Y2H screen.

out medium supplemented with X-α-Gal and Aureabasadin A (QDO/X/A). In this case, all the four reporters provided by the system were used, adding to AUR1-C and MEL-1 previously utilized, HIS-3 and ADE-2 for which the yeast with a true interaction would be additionally able to synthesize histidine and adenine. Colonies that survived in both conditions and turned blue were selected for further screening.

For TET2A 12 targets were identified: 3 were nuclear and the 9 were not. For TET2B, 21 targets were identified: 7 nuclear and 14 non-nuclear (Table 7.2 and 7.3).

Subsequently, positive prey clones were further assessed for autoactivation, by co-transforming them in yeast with empty pGBKT7 vector containing the GAL4 binding domain and screening them on DDO/X and QDO/X/A.

The ideal interaction partner would grow on DDO media and QDO/AbA media, and turn blue in the presence of X-α-Gal. Moreover, when transformed with an empty vector containing only the binding domain of GAL4, it should be able to grow on DDO,

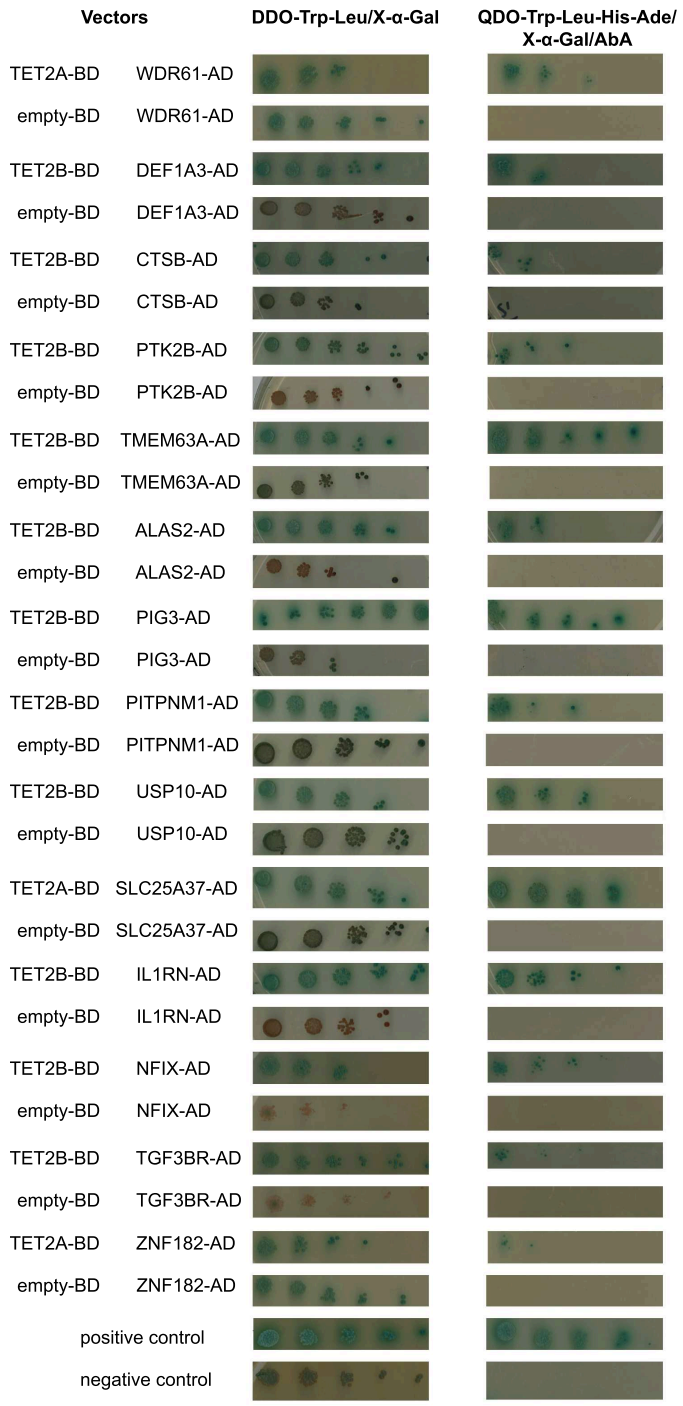


Figure 7.5: Interacting partners for TET2A and TET2B.

Gene	Total number of clones	N of identical clones
C1QA	1	1
FAM78A	5	3;2
hNRP-P2	2	1
HBA1	1	1
IGHC	1	1
IGHG2	3	3
PSAP	5	5
SLC25A37	1	1
SLC46A3	1	1
TMED2	1	1
WDR61	1	1
ZNF182	1	1

Table 7.2: List of interacting partners for TET2A. The ones localized in the nucleus are highlighted in yellow. Number of clones is indicated.

but not QDO/AbA media, and should not turn blue in the presence of X-α-Gal. The interaction partners that were validated showed these characteristics (Table 7.4 and Fig. 7.5). WDR61 and SLC25A37 showed autoactivation with X-α-Gal, forming blue colonies on DDO/X-α-Gal media. Since they were able to grow on QDO/AbA media only when co-transformed with the TET2A construct, they were considered as potential candidates for assessment via further mammalian screening.

One of the advantages of the Y2H assay is that it allows the detection of weak interactions; therefore, a system with multiple selection markers, can reveal the strength of an interaction from serial dilution plating. By comparing the colonies on QDO/AbA media in bait and prey, and in prey only, transformants, the strength of interaction, relative to one another can be inferred. In this case, this was well illustrated by TET2-WDR61 interaction versus TET2-NFIX, for example. Even though WDR61-AD with empty-BD alone was sufficient to activate X-α-Gal, it was not sufficient to grow on QDO/AbA. Therefore, the interaction was made stronger by TET2 either in terms of frequency of occurrence or stability. PIG3, on the other side, displayed strong interaction with TET2, as judged by the number of blue colonies obtained on QDO media.

Gene	Total number of clones	N of identical clones
ALAS2	3	3
B2M	1	1
DAPLE	1	1
CTSB	1	1
DEFA3	4	4
E4F1	1	1
HBB	2	1
HMGN2	1	1
IGHA1-2	3	3
IGHA1	5	5
IGHV3-30	1	1
IGLL5	1	1
IL1RN	1	1
NFIX	1	1
PIG3	1	1
PITPMN1	1	1
PSAP	8	7
PTK2B	1	1
TGFBR1	1	1
TMEM63A	1	1
USP10	1	1

Table 7.3: List of interacting partners for TET2B. The ones localized in the nucleus are highlighted in yellow. Number of clones is indicated.

Among all the candidates obtained, the first to be excluded were those that were non-nuclear. Next, being interested in the possible transcriptional effects of TET2, attention was focused on *WDR61*. WDR61 is part of the PAF complex, involved in co-transcriptional events such as mRNA elongation and the deposition of active histone marks across gene bodies [Kim et al., 2010], and of the SKI complex, which is involved in mRNA degradation [Zhu et al., 2005]. The PAF complex has been already linked to hematopoiesis due to its interaction with MLL1, a histone methyltransferase that is heavily mutated in blood cancers and catalyzes H3K4me3 formation, which augments the transcriptional activation of HoxA9 [Milne et al., 2010]. Conversely, knockdown of PAF complex members disrupts MLL and MLL fusion protein-recruitment to target loci, as well as MLL-mediated transcriptional activation. Deletions of MLL that abolish its interactions with PAFc also

Gene	Bait
ALAS2	TET2B
CTSB	TET2B
DEFA3	TET2B
IL1RN	TET2B
NFIX	TET2B
PIG3	TET2B
PITPMN1	TET2B
PTK2B	TET2B
TMEM63A	TET2B
USP10	TET2B
SLC25A37	TET2A
WDR61	TET2A
ZNF182	TET2A
TGFBR1	TET2B

Table 7.4: List of interacting partners for TET2A and TET2B. The ones localized in the nucleus are highlighted in yellow.

eliminate MLL-fusion construct-mediated cellular immortalization, indicating an essential function for this interaction in leukemogenesis [Muntean et al., 2010].

TET2, on the other hand, lacks an apparent DNA binding domain [Iyer et al., 2009] to target it to the DNA. Moreover, a high frequency of 5hmC deposition has been observed over gene bodies [Mellén et al., 2012], making the link between Paf complex and Tet2 a possible mechanism to explain this process.

7.1.2 WDR61 is a member of the PAF complex

WDR61 is a protein of the WDR family of proteins that is constituted by 7 WD repeat domains [Xu and Min, 2011]. These domains fold to form a three-dimensional protein structure called a beta propeller. Beta propellers consist of four stranded antiparallel β-sheet units (between four and eight) arranged in a circle, each one consisting of 40 aminoacids repeats that terminate with the tryptophan-aspartic acid (WD) dipeptide. Beta propellers allow WD proteins to act as a stable platform on which large protein complexes can assemble or disassemble. In a pull down assay, WDR61 was found to be

able to recognize H3K4me2 and H3K9me2 [Chan et al., 2009].

WDR61 is a protein of 34 kDa. The homolog of WDR61, SKI8, has been identified in yeast as part of the superkiller (SKI) complex, which is involved in mRNA degradation [Brown et al., 2000] and meiosis [Arora et al., 2004]. The mammalian WDR61 has been shown to be peculiarly different from that of yeast, as it associates primarily with the PAF complex and secondarily with a human SKI complex [Zhu et al., 2005] formed by TTC37 (homolog of Ski-3) and SKIV2L (homolog of Ski-2). It functions in the co-transcriptional regulation of mRNA synthesis, together with both the PAF and SKI complexes [Zhu et al., 2005].

To validate the earlier results, attempts were first made to isolate the PAF complex and determine how it related to TET2 in mammalian extracts. Since anti-WDR61 antibodies are not good enough for immunoprecipitation, a model system was set out to study the interaction. Two stable cell lines were obtained through lentiviral infection with a construct encoding for GFP-S-WDR61. Moreover, since the aim was to study the same portion of TET2 found to interact in the Y2H assay, we integrated a tetracycline inducible expression vector in the same cell line, that was either empty or encoded a 3XFl-TET2A.

By running a sucrose gradient, members of the complex and TET2 were co-fractionated. In particular PAF1 and CTR9 were shown in almost the same fractions, WDR61 (which was GFP tagged) and TET2 (both endogenous and TET2A) showed some degree of co-fractionation with members of the PAF complex (Fig. 7.6).

In order to further validate this finding, it was checked directly whether members of the complex could be coimmunoprecipitated (coIP) with WDR61. The same model system was used and LEO1 and CDC73 were validated as members of the same complex as WDR61. With the architecture of the complex previously published [Kim et al., 2010] (Fig. 7.7), where a direct interaction of WDR61 with CTR9 recruits other members of the complex, it can be inferred, under these conditions, that by coimmunoprecipitating CDC73 and LEO1 with WDR61, the whole complex is being immunoprecipitated.

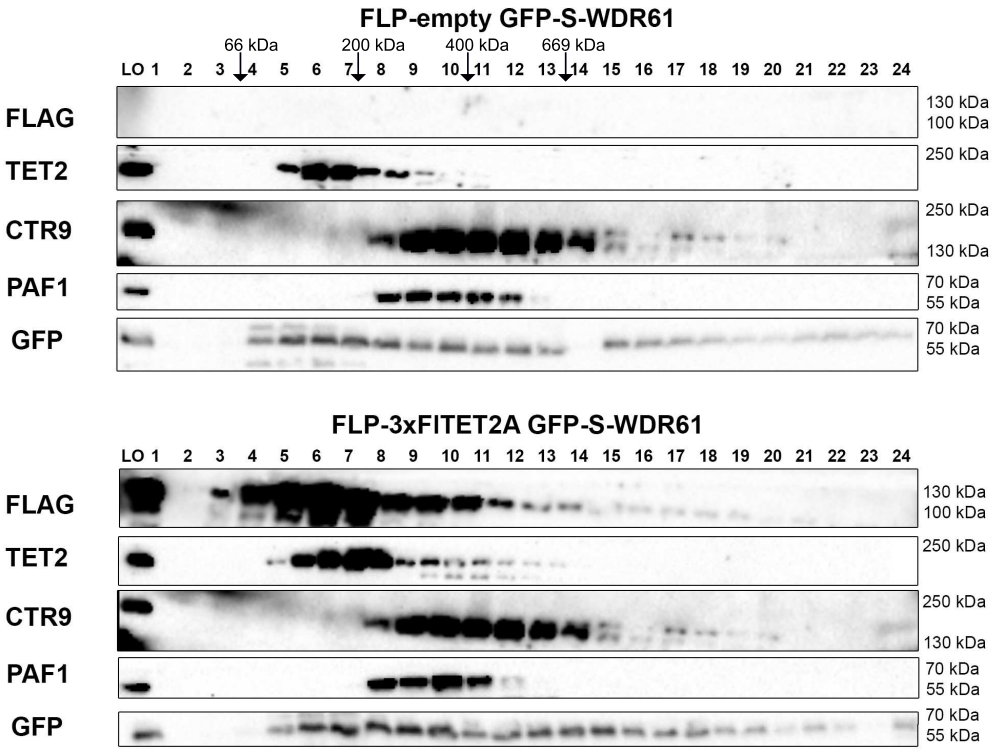


Figure 7.6: Sucrose gradient in FLP-empty vector GFP-S-WDR61 and FLP-TET2A GFP-S-WDR61. LO indicates the input.

7.1.3 Different approaches to validate Tet2-WDR61 interaction

Having previously established that the immunoprecipitation (IP) conditions were able to detect members of the PAF complex, we next asked whether TET2A interaction with WDR61 could be detected within the same system. Consequently, different approaches were set out:

- coimmunoprecipitation (Co-IP);
- affinity purification followed by mass spectrometry analysis (AP/MS);
- proximity ligation assay (PLA);
- colocalization assessed by confocal microscopy.

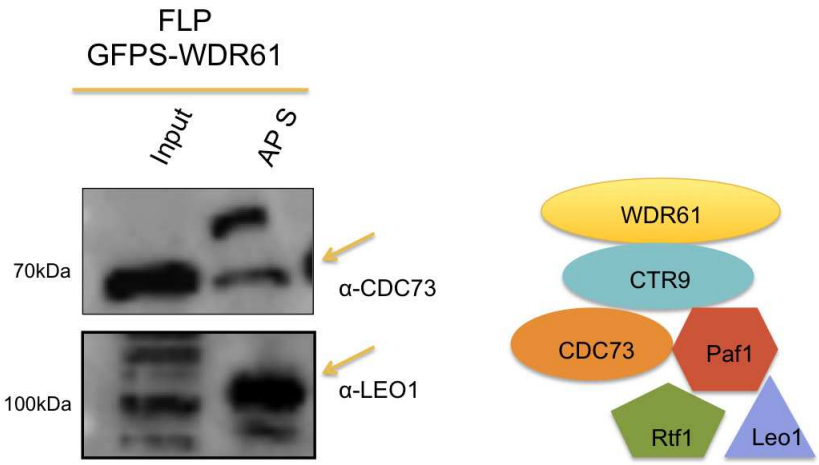


Figure 7.7: Co-IP of members of the PAF complex in FLP-empty vector GFP-S-WDR61 and FLP-TET2A GFP-S-WDR61 (At least 3 IP were repeated. Representative picture shown). Represented is the architecture of the PAF complex as described in [Kim et al., 2010].

The first two are common biochemical methods employed to detect interactions. The second two approaches are more suitable for the detection of weak or transient interactions, which are difficult to capture by biochemical approaches.

Co-IP does not show strong evidences of interaction

In the stable cell line expressing tagged TET2A and WDR61, we set out to coimmunoprecipitate TET2A and WDR61. The proteins were tagged with Flag and GFP-S tags; therefore, to IP, resins could be preloaded with antibodies or with peptides, respectively, against the Flag or the S tag. Additionally, all these antibodies were used to perform immunodetection, in addition to the anti-Tet2 and GFP antibodies that were available (Fig. 7.8). It was detected, upon affinity purification (AP) with the S tag antibody of S-WDR61, flagged TET2A (100 kDa) after long exposure with α-Flag and α-Tet2 antibodies (Fig. 7.8 A and B). Flag tagged Tet2 was immunoprecipitated with IP anti Flag (Fig. 7.8 C), and, some degree of co-IP was shown between FLP-TET2A, immunoprecipitated with α-Flag and α-Tet2 antibodies, and GFPS-WDR61 (70 kDa), detected by western blot with α-GFP antibody (Fig. 7.8 D).

Considering the weak evidence given by the immunoprecipitation results of TET2 po-

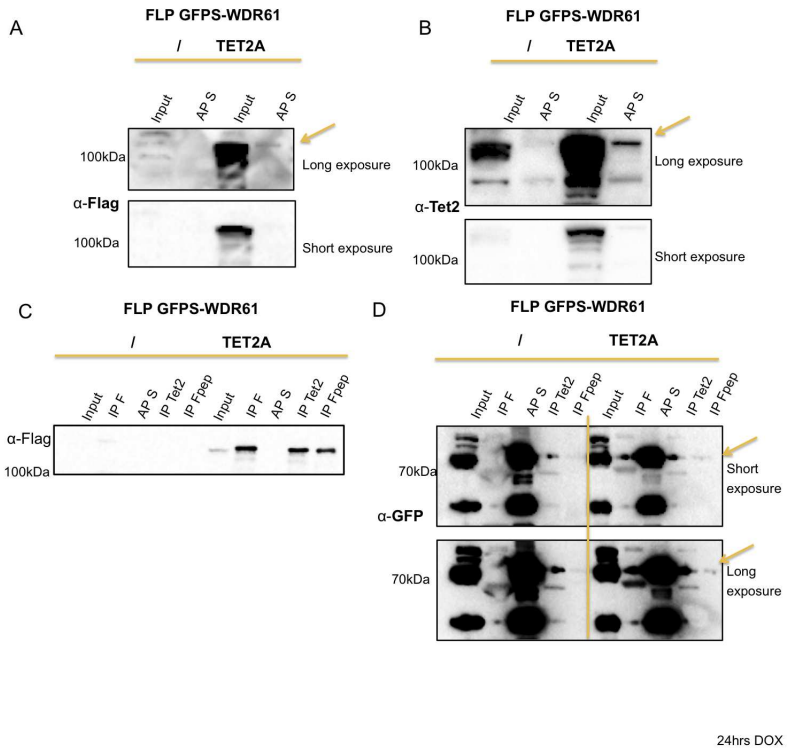


Figure 7.8: Co-IP of TET2A and WDR61 in FLP-empty vector GFPS-WDR61 and FLP-TET2A GFPS-WDR61. (At least 3 IP were repeated. Representative picture shown).

tentially interacting with WDR61, other strategies were used to assess whether this could occur. However, it could already be partially concluded, both from the Y2H results and from the co-IP, that the interaction is weak and not very frequent.

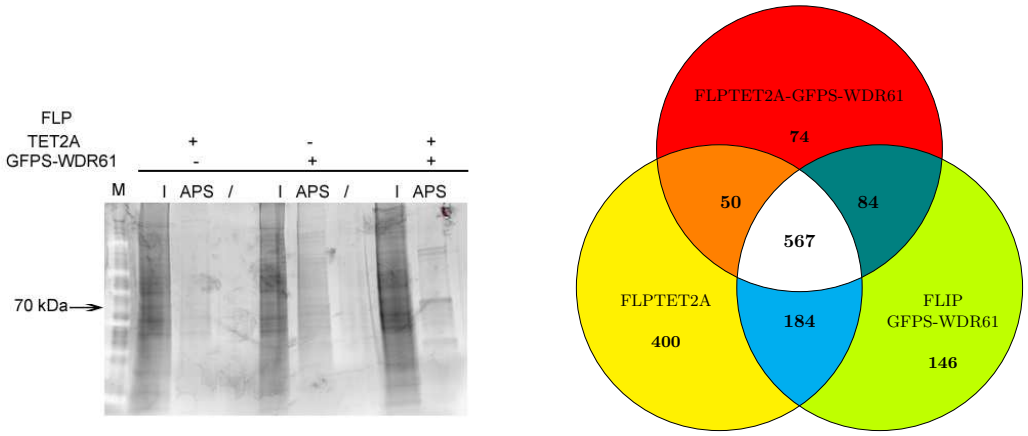
AP/MS for WDR61 does not detect TET2 despite detecting PAF complex members

The same cell lines were employed to perform affinity purification followed by mass spectrometry, which gives the potential to analyze thousands of potential interactions simultaneously and, at the same time, the sensitivity to detect weak interactions. It was decided that GFPS-WDR61 would be affinity purified with anti-S tag beads from three established

cell lines:

1. FLP-TET2A;
2. FLP GFPS-WDR61;
3. FLP-TET2A GFPS-WDR61.

The cell lines were chosen in a way that background subtraction could be performed for each AP. General background levels for the affinity purification anti-S would come from the first cell line. Additionally, subtraction of background interactions of WDR61 not related to TET2 were performed with the second cell line. GFPS-WDR61 was affinity



(a) (b)

Figure 7.9: AP/MS of GFP-S-WDR61 in the FLPN system: a)Ruby-pro staining of affinity purified samples. 0.5% of input and 8% of AP loaded. b) Venn diagram showing overlap of binders for FLPTET2A-GFPS-WDR61; FLPTET2A FLIP-GFPS-WDR61.

purified from both immunoprecipitations 2 and 3 alongside members of the Paf and Ski complex as: LEO-1, CDC73, CTR9, PAF1 and TTC37 (Ski-3 homolog) (Fig. 7.10 and Table 7.5). The results from the AP from FLIP-GFP-S-WDR61 confirmed that there was a significant enrichment of interaction in the network analysis performed after background and common contaminants in MS analysis subtraction (CRAPPOME database: [Mellacheruvu et al., 2013]). The AP from FLIPTET2A-GFP-S-WDR61 did not give the

expected results, and subsequent network analysis indicated that it was not enriched in interactions after background subtraction.

It was expected to increase the power of detection with the use of this approach in order

Target	Score	Mass	Num. of significant matches	Num. of significant sequences	emPAI	Rank
WDR61	7091	33731	309	25	403.64	1
TTC37	4457	177485	170	85	12.3	3
CTR9	903	134332	37	27	1.38	36
PAF1	823	60110	27	19	2.78	39
LEO1	461	75473	17	15	0.98	70
CDC73	75	60653	4	3	0.17	385

Table 7.5: AP/MS of GFP-S-WDR61 in the FLP system. PAF complex and SKI complex members with their relative scores are presented. emPAI is a measure of the relative abundance of a protein in the sample.

to be able to detect TET2A, which was only detected in small amounts with the coimmunoprecipitations. Despite being able to capture many interaction that so far have not been documented in complex with PAF or WDR61, TET2 was not detected. Therefore, it was concluded from this experiment that the strategy was suitable for the detection of interactions, even though the previous finding could not be validated using this approach. It might be relevant to repeat the experiment by IP/MS of TET2 in a suitable system such as the BM-HPC cell line, which fully recapitulates hematopoietic stem cells in their ability to repopulate blood lineages in a transplantation assay [Pinto do O, 1998][Pinto do O, 2002]. This could give insights into the protein network of TET2 in the same context that the question was interrogated in this study by Y2H.

Microscopy shows some degree of co-localization between WDR61 and TET2A

Since investigating the interaction biochemically was difficult to achieve due to the aforementioned reasons of weak interaction and TET2 possibly not being part of a stable complex, but instead possibly part of cotranscriptionally regulated event, attempts were made to try to solve this problem by adopting a microscopy approach. With the same cell lines that were previously utilized, immunoflorescence was performed to check for co-

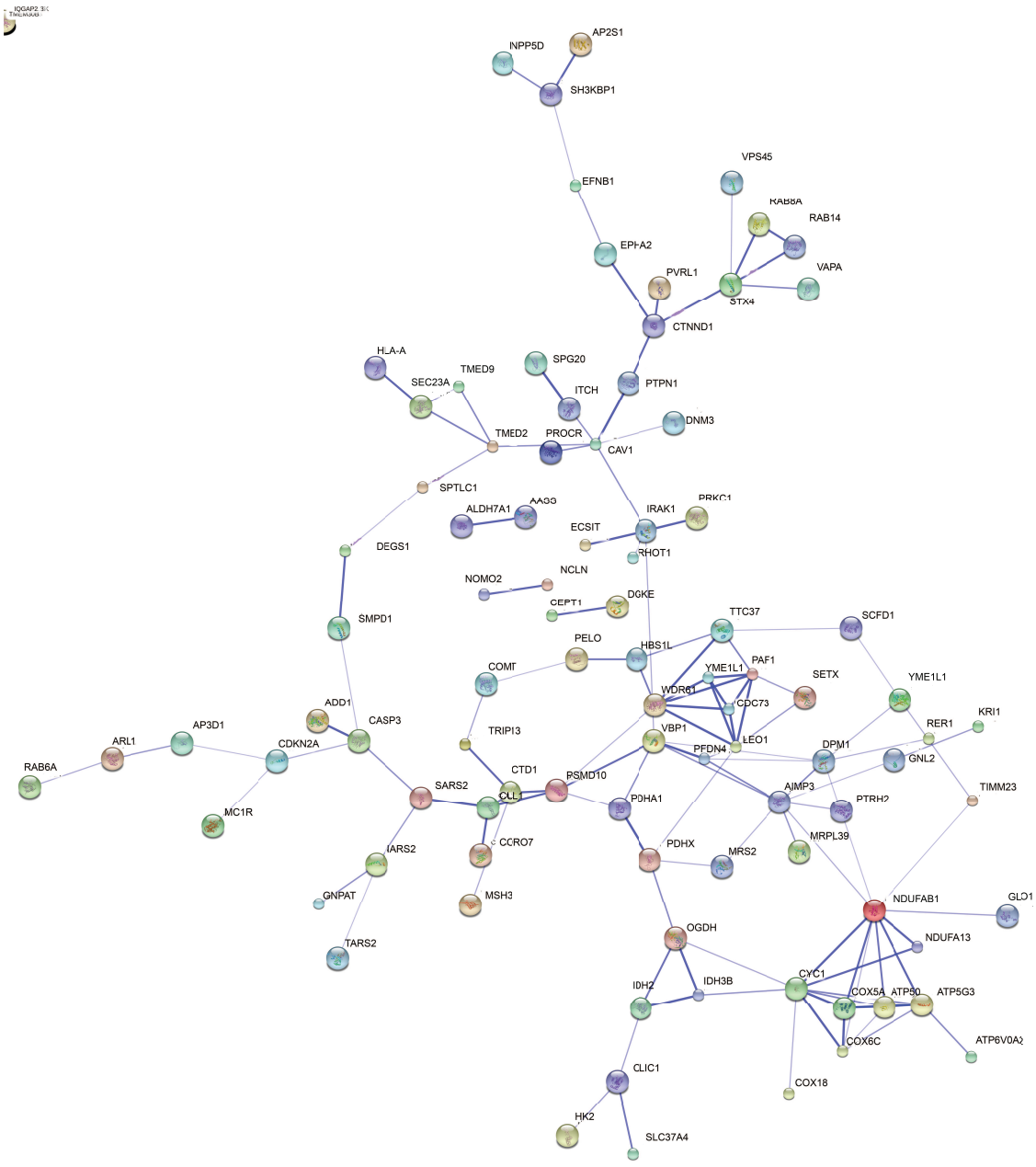


Figure 7.10: STRING network analysis of IP-MS data from FLPGFPS-WDR61. Confidence view, where darker lines indicate stronger associations. Interaction enrichment p value $= 9.79 * 10^{-8}$ calculated by STRING with the Random Graph with Given Degree Sequence (RGGDS) method [Pradines et al., 2005].

localization between the TET2 fragment and GFP-S-WDR61.

Some degree of co-localization between the TET2 fragment and WDR61 were observed.

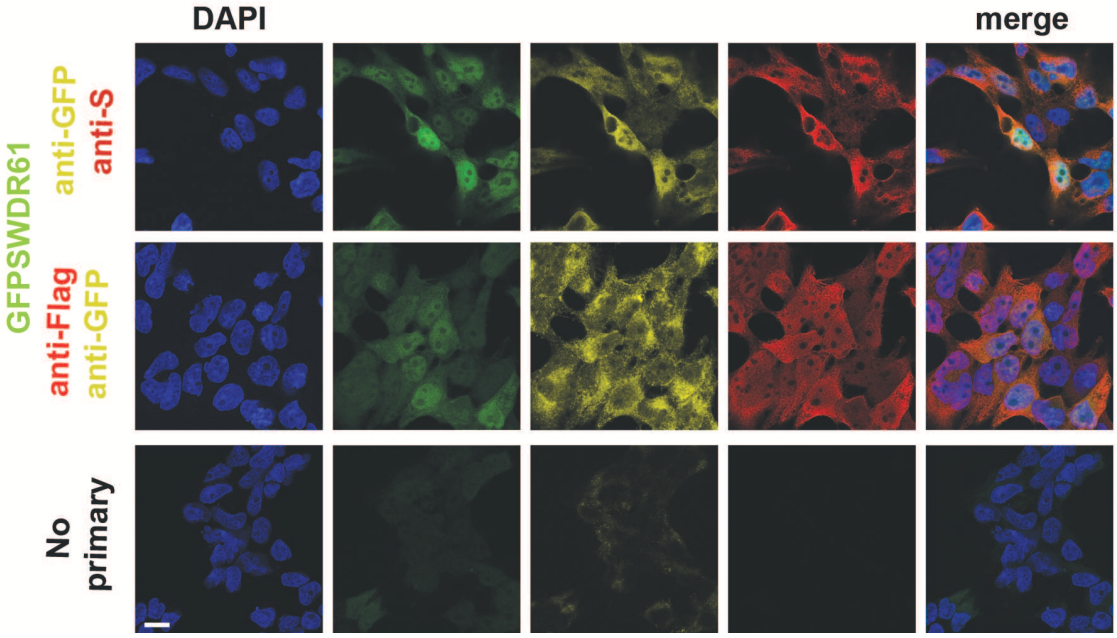


Figure 7.11: Assessment of co-localization between TET2A and GFPS-WDR61 in FLP-TET2A GFPS-WDR61. GFPS-WDR61 is displayed on the green channel (green label) and detected with anti-GFP (yellow label) and anti-S antibody (red label) in yellow and red respectively. TET2 is detected by staining with anti-Flag antibody (red label) on the second line. No primaries are shown on the third line. (n =3 at least. Representative picture shown.

It was notable that both proteins did not seem to localize only in the nucleus. WDR61 has already been documented in both the nucleus and the cytoplasm [Uhlén et al., 2005]. TET2 has been shown in the nucleus [Ko et al., 2010], however, the truncated protein might behave differently.

PLA shows interaction between WDR61 and TET2A

The proximity ligation (PLA) assay can be used to screen for weak interactions by microscopy [Soderberg et al., 2006]. It is based on the recognition of two proteins whose interaction is to be assessed, by two antibodies from different species. The secondary antibodies that will recognize them are conjugated to oligonucleotides that will circularize

by ligation if they are close enough (30-40 nm), to form a circular dsDNA. Rolling circle amplification will follow and will incorporate a florescent dye in the dsDNA. The assay was set out in FLIPNGFPS-WDR61. A positive control was established with antibodies against GFP and S tag that are on the same protein, and a negative control was set with the omission of primary antibodies (Fig. 7.12). An increase was detected in the number

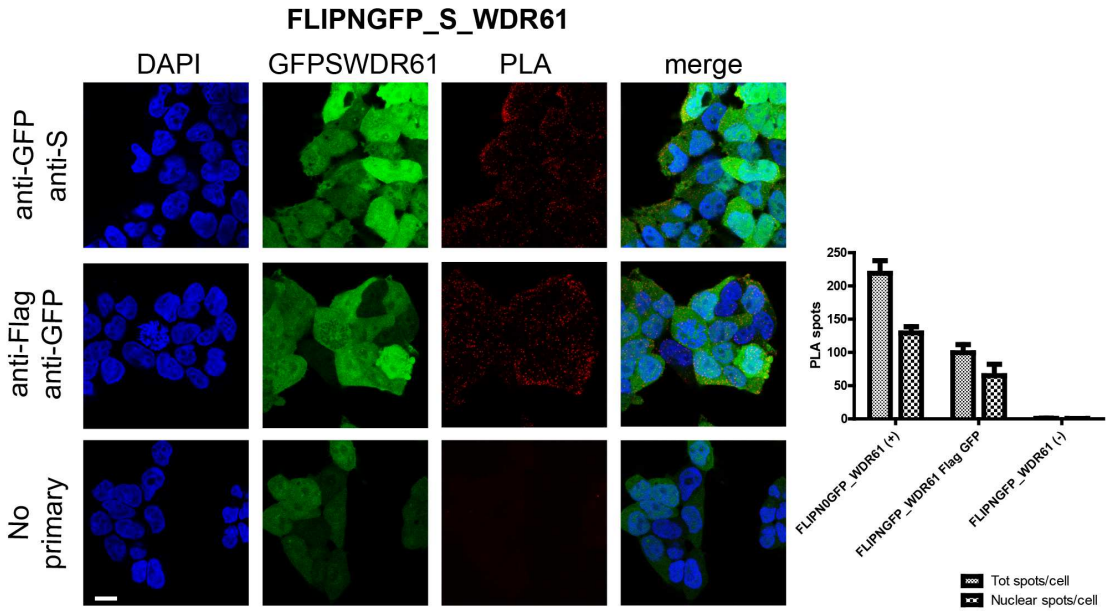


Figure 7.12: Proximity ligation assay between Tet2A and GFP-S-WDR61 in FLPTET2A GFPS-WDR61 cells. Antibody combinations are indicated on the left side. Signal detection on the top. Quantification of the average number of spots per cell is shown (approximately 200 cells counted for each experimental condition) (n =3. Whiskers indicate standard deviation).

of spots over background spots in the positive control under test conditions, indicating that the proteins displayed a certain degree of interaction in this *in vitro* non physiological setting. PLA is a useful tool as it allows, if antibodies are available, to perform interaction studies in a more physiological setting, such as in a tissue setting. Possible directions to further validate this finding could therefore be *in situ*.

7.1.4 A possible role of TET2 and PAF over the *HOXA9* locus

It has previously been shown that the PAF complex is present at the HOXA9 locus, where it orchestrates transcriptional activation together with MLL1 [Milne et al., 2010]. Moreover, by browsing a publicly available ChIP-seq dataset for TET2 in 293T cells, a TET2 peak was detected on the same locus [Deplus et al., 2013] and on MLL1 and HBB genes (Fig. 7.13). It was therefore decided that this finding needed to be validated, hence

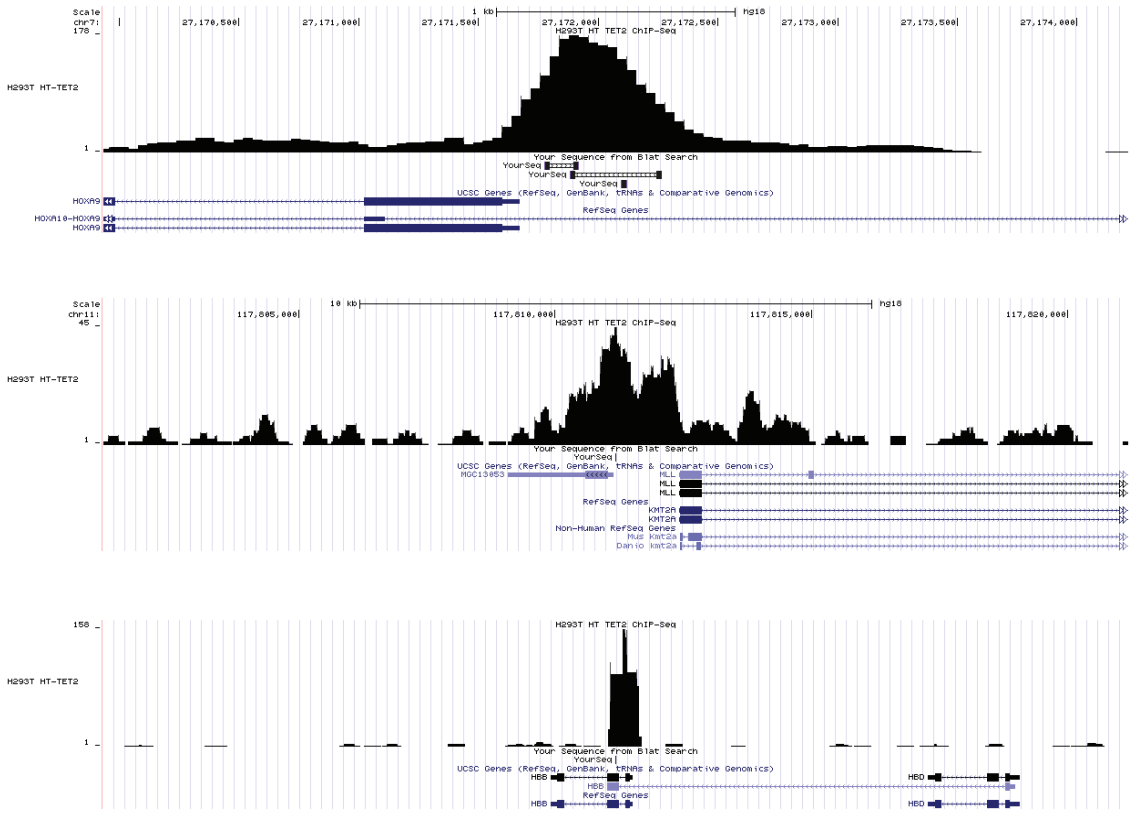


Figure 7.13: ChIP data obtained by browsing publicly available datasets over HOXA9, MLL1 and HBB promoters [Deplus et al., 2013].

a ChIP was performed in our 293T3xFlTet2 cell line, which carries a 3x Flag version of TET2 stably integrated into its genome. The Flag tag was used to perform a ChIP for TET2 in the system to validate the binding over HOXA9, MLL1 and HBB loci (Fig. 7.14). Moreover, HOXA9 was induced upon transient transfection with TET2 (FlTet2) or stable

TET2 expression (3xFlTet2) (Fig. 7.14).

Another study validated the role of Tet2 and 5hmC in transcriptional activation of the

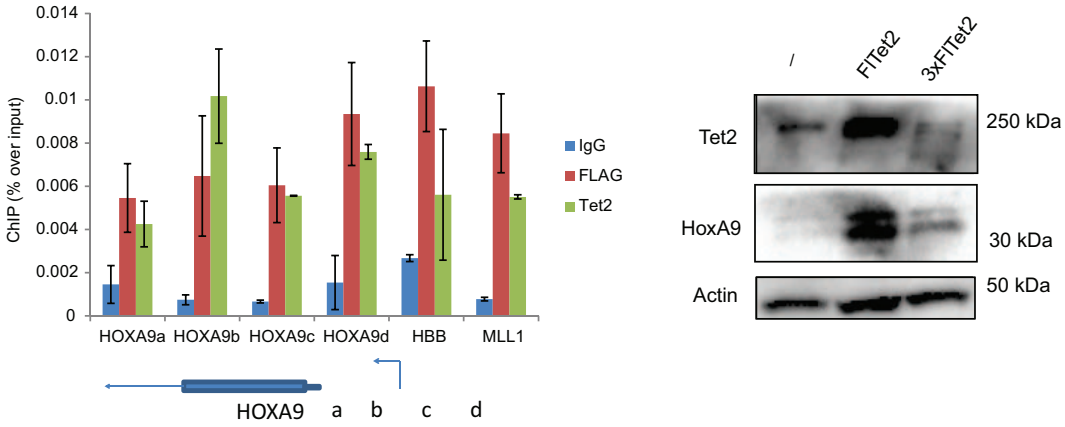


Figure 7.14: ChIP assay in 3xFlagTet2_293T and western blot of HOXA9. Primers span 150 bp each on the promoter region of HOXA9 (n = 3, whiskers bars indicate standard deviation.) Western blot of HOXA9 in transiently transfected TET2 cells (FlTet2) at 24 h post transfection or stable TET2 expression (3xFlTet2). Actin is used as loading control.

HOXA locus upon NT2 neuronal cell retinoic acid-induced differentiation [Bocker et al., 2012]. Moreover, these genes are all related to hematopoiesis, and HOXA9 has been shown to cause leukemia if overexpressed with Meis1 [Kroon et al., 1998]. Mice lacking HoxA9 have disrupted hematopoiesis, but retain normal levels of more primitive progenitors with long-term culture-initiating cell activity [Lawrence et al., 1997]. With this finding, a direct link was established between TET2 and myeloproliferation through Paf complex mediated-recruitment. Further studies will be needed to try to assess whether MEIS1 is regulated by TET2, and whether 5hmC is deposited over bound genes in the BM-HPC model system upon differentiation.

7.2 Brief summary

TET2 is a member of the TET oxygenase family of enzymes that catalyzes the α-KG and iron dependent conversion of 5mC to 5hmC [Tahiliani et al., 2009]. It has been found to

be mutated in a number of cancers, but the highest mutation incidence is in blood proliferative diseases such as myelodysplastic syndromes [Delhommeau et al., 2009], myeloproliferative neoplasms [Langemeijer et al., 2009], secondary neoplasms such as acute myeloid leukemia [Abdel-Wahab et al., 2010] and chronic myelomonocytic leukemia [Jankowska et al., 2009]. Much work has already been done to try to understand the connection of *TET2* with cancer. Loss of function studies in mice have modeled the human myeloproliferative phenotype with increased stemness, decreased differentiation potential, greater repopulating ability and a skew toward myeloid lineage [Moran-Crusio et al., 2011]. Other than genetic approaches, proteomic strategies have been pursued to try to identify protein functions. Here, the TET2 protein interaction network was investigated to try to get a better understanding of its role in disease establishment and contribution.

A number of protein partners for TET2 were published during the course of the study, mainly linking its function to transcription. One of the partner identified in more than one study was OGT [Chen et al., 2012b][Vella et al., 2013][Deplus et al., 2013], which is an O-linked N-acetylglucosamine transferase that is involved in transcriptional activation through recruitment of the SET/COMPASS histone methyltranferase [Lee et al., 2007]. Other partners further link TET2 with transcription and genomic targeting. IDAX, which is a CXXC domain-containing protein that binds to unmethylated CpG dinucleotides, localizes to promoters and CpG islands in genomic DNA and interacts directly with the catalytic domain of TET2 [Ko et al., 2013]. Other links with transcription and pluripotency maintenance, through the regulation of gene expression and promoter DNA methylation, include the transcription factors EBF1 [Guilhamon et al., 2013], NANOG [Costa et al., 2013] and PU.1 [de la Rica et al., 2013], the interaction with the transcriptional regulator PRDM14 [Okashita et al., 2014] and glycosylases of the BER pathway TDG and NEIL [Müller et al., 2014]. Among them the strongest evidences are for OGT and NANOG, which were observed multiple time with immunoprecipitations performed under both overexpression and endogenous settings and with stringent salt conditions. The oth-

ers might be peculiar to the system studied and less frequent or stable.

Here, the protein network of TET2 was investigated in an *in vitro* setting by adopting the yeast two-hybrid system. By exploiting one of the advantages of the system, which is its flexibility, the question could be investigated in a hematopoietic context by making use of a bone marrow-derived prey library. To overcome one of the major disadvantages of the method, a screening strategy was adopted that relied on the activation of four independent reporters upon positive interaction between prey and bait. Furthermore, to maximize the possibility of success, the expression of full length TET2 was avoided by dividing it into two parts following criteria that excluded regions of sequence conservation and secondary structure formation for the selection of the breakage point. The two fragments adopted as bait were TET2A (1-733 AA) and TET2B (734-2002 AA). After verifying that they did not result in cytotoxic effects or auto-activate the GAL4 responsive elements upon expression in yeast, the next step of the screening was begun, which was the mating of the bait transformants with the prey pre-transformed library. The mating screen resulted in the identification of possible candidates for the interaction, which were then also screened for auto-activation. Among the list of the ones obtained after filtering out for duplicates and auto-activating proteins, attention was focused on *WDR61*.

WDR61 is a member of the WDR family of proteins, constituted by 7 WD repeat domains [Xu and Min, 2011] which can recognize methylated histone lysine motifs [Chan et al., 2009]. WDR61 is conserved in yeast and it assembles into the PAF [Mueller and Jaehning, 2002] and SKI complexes [Brown et al., 2000][Zhu et al., 2005], which are dedicated to the co-transcriptional control of gene expression. The PAF complex has been found to participate in transcription elongation together with super elongator II (SII) [Kim et al., 2010], the process by which the RNA polymerase gets engaged in productive transcription after its release from the pre-initiation complex. Moreover, the PAF complex has been shown to be able to link elongation to histone modification by recruitment of the SET/COMPASS complex and deposition of H3K4me3, and of Dot1 with deposi-

tion of H3K79me3 [Krogan et al., 2003]. Additionally, the O-linked N-acetylglucosamine transferase OGT, which has previously been reported to interact with TET2 [Chen et al., 2012b][Vella et al., 2013][Deplus et al., 2013], is able to glucosylate H2BS112 and mediate the ubiquitination of H2BK120 by BER1 enzymes [Fujiki et al., 2011], a modification that is itself linked to the recruitment of the SET/COMPASS complex and to the production of H3K4me3, promoting the activation of gene expression [Lee et al., 2007]. It was intriguing to try to position TET2 in this landscape by adding a potential interaction with WDR61 and, therefore, with the PAF complex.

Another reason for interest in the PAF complex is the still unexplained role of 5hmC over gene bodies which has been found in highly expressed genes in the brain [Mellén et al., 2012], *in vitro* during neuronal differentiation [Hahn et al., 2013] and in both human and mouse ESCs [Xu et al., 2011][Stroud et al., 2011][Pastor et al., 2011][Wu et al., 2011a][Yu et al., 2012][Lister et al., 2013][Booth et al., 2012][Huang et al., 2014]. A possible hypothesis links TET2 to the deposition of 5hmC over gene bodies, as evidence suggests that levels of 5hmC are reduced primarily at this location upon Tet2 KD [Huang et al., 2014]. Speculation points to the fact that TET2 does not possess a CXXC binding domain and, thus, cannot be targeted efficiently to unmethylated CpG dinucleotides, which are known to be highly enriched in CpG islands present mainly at promoter regions [Bird, 1986].

To validate the yeast two-hybrid finding, cell lines were established carrying 3xFlTET2A and GFPS-WDR61. First it was validated that WDR61 could be co-purified with members of the PAF complex such as CDC73 and LEO-1 under the conditions used, and that some degree of co-fractionation could be detected in sucrose gradients between TET2, WDR61 and PAF complex members. Subsequently the analysis was continued using different approaches like affinity-purification associated with MS analysis and immunoblot detection, and microscopy techniques like proximity ligation assay and colocalization studies.

The biochemical approaches, together with the Y2H assay, revealed that the interaction is neither frequent or stable. Only weak bands were detected for the co-IP for both Tet2

and WDR61 when immunoprecipitating each other. The IP condition were found to be valid for the isolation of both PAF and SKI complex members, as they were detected in the MS results after background subtraction. Having established that biochemical techniques might not be sensitive enough for the detection of WDR61-TET2A interaction, efforts were directed towards microscopy techniques. First the degree of colocalization was assessed followed by PLA. In PLA, a spot can be detected only if proteins are in close proximity, therefore the technology is suitable to reveal weak and transient interactions that might happen in small fractions of cells. In the PLA assay, a number of spots was detected for TET2A GFPS-WDR61 over the background. In conclusion, some evidences has been obtained that this interaction occurs and future directions might test whether the reverse is true by IP/MS of TET2 in a hematopoietic setting, such as the BM-HPC cell line, which has been shown to be a valuable biochemical tool for the study of hematopoiesis as it overcomes limited hematopoietic stem cell availability. This cell line was established from stem cells through overexpression of Lhx2, and tested for engraftment in mice where it could fully recapitulate all the blood lineages [Pinto do O, 2002]. In this direction, attempts were made to link TET2 and the PAF complex to a possible function within the hematopoietic compartment. In the literature, the positive role of MLL1/PAF complex interaction over the transcription of the HoxA9 locus is well documented [Milne et al., 2010] and it has been known that HoxA9 deletion can induce myeloproliferation in a genetic mouse model [Lawrence et al., 1997]. Additional lines of evidences point to a direct role of MLL1/PAX complex interaction in the establishment of myeloid leukemia [Muntean et al., 2010]. The presence of TET2 over the *HOXA* locus has been documented in a neuronal cell line [Bocker et al., 2012]. This finding was validated here in HEK 293T (embyonic kidney fibrobalsts) cells, further confirming what had already been observed in publicly available ChIP datasets for TET2 [Deplus et al., 2013]. Furthermore, HOXA9 upregulation was detected following TET2 induction. The above finding further points towards a direct involvement of TET2 over the HOXA9 locus,

possibly mediated by its interaction with the PAF complex. Mutant TET2 could behave like MLL fusion protein and abnormally activate the transcription of the HoxA9 locus, necessary for normal myeloproliferation [Milne et al., 2010]. Future studies might look towards the mapping of the interaction to understand if any particular mutation is able to abolish/strengthen it. Finally, it could be established a direct link between TET2, cancer and 5hmC deposition over gene bodies. Further experiments to validate this, other than MS analysis, could assess the effects of the presence of 5hmC over the locus towards differentiation in the BM-HPC model-system.

Chapter 8

Discussion

In 2009, two groundbreaking papers discovered the presence of 5hmC in mammalian DNA [Kriaucionis and Heintz, 2009][Tahiliani et al., 2009]. 5hmC is derived from 5mC oxidation, which is catalyzed by the TET family of iron and α-ketoglutarate-dependent oxygenases [Tahiliani et al., 2009]. The same enzymes have subsequently been found to produce 5fC and 5caC [Ito et al., 2011]. TET1 was originally discovered fused to MLL1, forming the ten eleven translocation in leukemia that gave the Tet family its name [Lorsbach et al., 2003]. After the discovery of 5hmC, more papers came out documenting different TET2 mutations in many cancer types, with the highest mutation rate observed in blood cancers, particularly in acute myeloid leukemia [Delhommeau et al., 2009]. Moreover, 5hmC has been found to be globally depleted in the genome of multiple cancer types [Yang et al., 2012][Kraus et al., 2012], generally correlating with decreased TET expression (breast, pancreas, lung, liver and skin cancers) [Lian et al., 2012] or TET2 inactivating mutations (blood cancers) [Ko et al., 2010].

The aim of this research project was to characterize 5hmdC, 5fdC and 5cadC as possible cytotoxic agents for use in anti-cancer therapy (Chapter 5 and 6). Thus, the roles of 5hmdC and 5fdC were investigated in the context of the nucleotide salvage pathway (Chapter 4).

Additionally, given the already documented major roles of TET2 in cancer, attempts were made to dissect its protein interaction network to gain a better understanding of its physiological role, and that of 5hmC, in cancer development (Chapter 7).

8.1 5hmdC and 5fdC as possible cancer therapeutic agents

Since its discovery, one of the roles associated with 5hmC has been that of an intermediate for DNA demethylation, via both active and passive mechanisms [Tahiliani et al., 2009][Hajkova et al., 2010][Guo et al., 2011]. There has, therefore, been significant interest in it, given the prominent role of DNA methylation in biological processes such as imprinting, development, transposon silencing, X-chromosome inactivation and cancer. Cancer genomes generally display hypomethylation and site-specific CpG island promoter hypermethylation. DNA hypomethylation occurs at many genomic sequences, such as repetitive elements, retrotransposons and introns. However, the most recognized epigenetic disruption in human tumors is CpG island promoter hypermethylation-associated silencing of tumor suppressor genes such as p16, MLH1 (mutL homolog-1), BRCA1 (breast cancer–associated-1) and VHL (von Hippel-Lindau tumor suppressor) [Esteller, 2008]. The importance of the regulation of DNA methylation has been supported by the discovery of somatic mutations in enzymes involved in the pathway, such as the ones observed for DNMT3A [Ley et al., 2010], TET1 [Lorsbach et al., 2003] and TET2 [Delhommeau et al., 2009] in hematological malignancies. Moreover, aberrant DNA methylation has been proposed as one of the hallmarks of cancer [Esteller, 2007]. Numerous lines of investigations have, therefore, led to the development and FDA approval of therapies for acute myeloid leukemia that are based on nucleoside analogues (5-azacytidine and 5-azadeoxycytidine), able to mimic endogenous cytosine and to inhibit DNA methylation [Fenaux et al., 2009][Garcia-Manero et al., 2013]. In particular, azacytidine and its deoxynucleoside equivalent decitabine, when incorporated into the DNA, result in the

covalent trapping of DNMTs (DNA methyl transferase) enzymes on the base that they were attempting to methylate, which stops the progression of the replication fork until the covalent block is resolved by the degradation of DNMTs [Santi et al., 1984]. As a consequence of DNMT depletion, methylation marks are lost during DNA replication. One of the disadvantages in the use of these drugs in the clinic is their chemical instability and a consequent half-life of less than 24 hours in the aqueous environment [Lin et al., 1981]. Another drawback, is the expression of cytidine deaminase (CDA) in the kidney and liver [Camiener and Smith, 1965] which quickly converts them into the inactive DNA methylation inhibitors 5-azauridine (5azaU) and 5-azadeoxyuridine (5azadU) [Chabner et al., 1973]. Therefore, strategies to overcome the inherent instability of these analogs and their enzymatic conversion by CDA have been [Camiener, 1968][Driscoll et al., 1991], or are being attempted through the development of CDA inhibitors (tetrahydrouridine (THU) or zebularine derived) [Ferraris et al., 2014], or with different formulations like the conjugation of 5azadC to another nucleoside in the form of oligonucleotides [Stresemann and Lyko, 2008].

Since the result of TET enzymes overexpression, and that of treatment with 5-azacytidine and decitabine is global DNA demethylation [Tahiliani et al., 2009][Jones and Taylor, 1980], it was speculated that 5hmdC could be key in this process, even when administered to cells. It was hypothesized that it could eventually be incorporated into the DNA and induce DNA demethylation via mechanisms dependent on thymine DNA glycosylase (TDG) or activation-induced deaminase (AID)/APOBEC-mediated base excision repair (BER) pathway enzymes (reviewed in [Wu and Zhang, 2014]).

Once a nucleoside has entered a cell, through dedicated transporters, it has to overcome various obstacles before it can be incorporated into the DNA. Therefore, in order to verify the hypothesis whether 5hmdC treatment would induce DNA demethylation once administered to the cell, two strategies were adopted: the first one investigated the fate of 5hmdC and 5fdC in relation to nucleoside salvage pathway enzymes; while the second determined

the cytotoxic potential of the nucleosides.

The first strategy was investigated via a top-down approach, first testing whether the replication machinery would pose a barrier to 5hmdCTP incorporation. Previous reports have documented the role of 5hmC, 5fC and 5caC in DNA replication and DNA structure. In a tissue culture model it was demonstrated that templates containing 5fC and 5caC modestly inhibited replication, while 5hmC-containing plasmids caused no significant impairment [Ji et al., 2014]. Regarding Watson-Crick base pairing, no significant deviation was found in B-DNA structure, when 5hmC, 5fC and 5cadC were incorporated into oligonucleotides used to generate crystals [Renciuk et al., 2013]. Following these lines of investigation, this project assessed whether the cellular replication machinery imposed any barriers to 5hmdCTP incorporation. Given that mammalian cytoplasmic membranes are not permeable to nucleotide triphosphates, they were delivered either directly to nuclear extracts (IVRA) or through the usage of electroporation or vesicle mediated transfer. With these approaches, successful incorporation of 5hmdCTP into the DNA of replicating cells was achieved, with the drawback of not being able to assess the kinetics of incorporation (for which a labeled source of triphosphate would be necessary for accurate quantification).

Having assessed the successful incorporation of 5hmdCTP, the fate of 5hmdC once it entered the cell needed to be unraveled. Nucleosides are imported into the cell via nucleoside transporters which are broadly expressed across NCI60 panel cell lines, therefore it was assumed that 5hmdC would gain entry into the cells via any of these family members. Given the similarity of these cytidine variants with 5azadC, it was hypothesized that it would be preferentially imported via the same family of equilibrative nucleoside transporters (ENT) family members as documented previously [Damaraju et al., 2012]. Once it had entered the cell, the nucleoside could follow two main routes in the cytoplasm:

1. deamination by cytidine deaminase (CDA);

2. monophosphorylation by deoxycytidine kinase (DCK).

To tackle this hypothesis, two lines of investigation were followed: one that would see the treatment of a variety of cancer cell lines belonging to the NCI60 panel to potentially exploit the available gene expression resources for bioinformatic analysis; and the other that would start to biochemically dissect the pathway to nucleotide production, by purifying the enzymes believed to be involved.

In vitro cytotoxic potential of 5hmdC and 5fdC

After assessing the stability of 5hmdC, 5fdC and 5cadC under the conditions utilized for the assay, treatment of more than twenty cancer cell lines with 5hmdC over a 10 day time frame was begun. The discovery of 5fdC and 5cadC had not been reported at the start of the project, therefore they were included later in the study and tested on a smaller number of cell lines. The selection of cancer cell lines started by taking into account two criteria: availability and a broad range of tumor origins. The first cell lines tested included a breast cancer cell line (MDA-MB-231), which displayed a strong response to 5hmdC treatment, with 5hmdC treated cells dead at day 10 after treatment. Weaker growth inhibition was observed in the A375 cell line. No growth inhibition was observed for all the other cancer cell lines tested at the beginning, both derived from solid (LN18, H1299, Colo-320) and blood cancers (THP-1, BL70). With the initial idea to exploit genomic data to investigate the mechanism of cytotoxicity of the compound, it was necessary to find another cell line that behaved similarly to MDA-MB-231. In the search for cell lines displaying the same growth phenotype, few parameters were considered such as origin, common DNA methylation profile at promoters associated CpG islands of genes important for drug response [Shen et al., 2007], gene expression and drug response cluster analysis [Scherf et al., 2000]. By testing several cancer cell lines per group, cancer origin and common DNA methylation profiles at some promoters could be excluded as possible factors mediating the response. Finally, by making use of an available gene expression cluster analysis for cell lines belonging to the NCI60 panel [Scherf et al., 2000], a lung cancer cell line (HOP-92) was

found that behaved similarly to MDA-MB-231. Other than MDA-MB-231 and HOP-92, for which less than 20% of live cells were counted after treatment, a group of cell lines was found that displayed intermediate growth inhibition (defined by setting a percentage of cells remaining after the treatment comprised between 20 and 80%), and no growth inhibition (more than 80% of cells alive after the treatment).

To decode the mechanism of response, which in our hypothesis could be due to global induction of DNA demethylation, the global levels of 5mC and the levels of expression of genes known to be reactivated upon 5azadC administration were evaluated. No significant difference was observed in global 5mC content and gene expression levels between dC-treated and 5hmdC-treated cell lines. In comparison, 5azadC (which was the positive control for DNA demethylation) displayed both phenotypes.

Having excluded DNA demethylation as a possible cause of the observed cytotoxic response, available gene expression datasets [Pfister et al., 2009] were used for cluster analysis of two extreme responders (MDA-MB-231 and HOP-92) versus two cell lines insensitive to the treatment (MCF7 and MDA-MB-435). This analysis resulted in cytidine deaminase (*CDA*) being ranked 79th in the list of possible candidates commonly expressed in responsive cell lines, but differentially expressed in relation to the non-responding ones. No other gene known to be involved in nucleotide metabolism was found to be significantly differentially expressed. In cancer therapies development, large datasets (genomic, metabolomic) are currently exploited to model exhaustively the wide heterogeneity observed in cancer. In light of this, cancer cell panels represent a useful tool for better understanding how compounds act *in vitro*, before their *in vivo* development. A molecular understanding of drug action is necessary to limit its induced toxicity and to provide better and more personalized treatment option.

Nucleoside salvage of 5hmdC and 5fdC

Having assessed that 5hmdCTP, if present, could be incorporated by the replication machinery, it remained to be investigated whether the enzymes involved in the cytidine salvage pathway could use 5hmdC and 5fdC to produce their respective triphosphates. Hence, it was first investigated whether a monophosphate could be produced from 5hmdC, 5fdC and 5cadC. Nucleoside kinases capable of adding a monophosphate to nucleosides are present in both the cytoplasm (TK1, DCK, AK) and the mitochondria (TK2, DGK). Among them, the enzyme mainly responsible for the production of dCMP from dC is deoxycytidine kinase (DCK). Additionally, very low conversions of dC to dCMP are observed with TK1 [Johansson and Eriksson, 1996]. Given the highest affinity and activity of DCK for dC, the enzyme was purified and its activity assessed on 5hmdC, 5fdC and 5cadC. DCK has already been shown to be able to phosphorylate 5mdC, 5azadC and other nucleoside analogs, displaying lower V_{max} for cytidine modified at the fifth aromatic ring carbon position [Eriksson et al., 1991]. Monophosphates were obtained for 5hmdC and 5fdC, but not for 5cadC.

Following monophosphate production, the nucleotide can undertake one of two routes: either deamination by dCMP deaminase (DCTD) [Maley et al., 1993]) or diphosphorylation by cytosine monophosphate kinase (CMPK1). Since assessment of the possibility of triphosphate production from these modifications was desired, the focus of the study was restricted to CMPK1. CMPK1 is not the only kinase that can produce dCDP in the cytoplasm; AK5 [Van Rompay et al., 1999a], AK7, AK8 [Panayiotou et al., 2011] and the recently discovered AK9 [Amiri et al., 2013] all have minor activity towards dCMP compared with CMPK1. CMPK1 was purified and its phosphorylation potential assessed in a coupled assay with DCK. Diphosphates could be detected for dCMP, but not for 5hmdCMP and 5fdCMP. Therefore, it was concluded that a 5hmdCTP triphosphate could not be produced by the canonical route of cytosine salvage. This finding is in line with previous reports that were unable to obtain 5mdCTP from 5mdC [Vilpo and Vilpo, 1993].

CDA as an effector for 5hmdC and 5fdC salvage and cytotoxic activity

Having assessed that 5hmdC and 5fdC could not be converted into triposphates and the gene expression analysis having indicated that CDA could be mediating the observed cytotoxic effect of 5hmdC on cell lines, the hypothesis was made that 5hmdC could be converted into 5hmdU, which has already been shown to be incorporated into DNA and to induce toxicity if administered to cells and incorporated with a 5hmdU/T ratio of 1/150 [Boorstein et al., 1987][Kaufman, 1986].

To validate this hypothesis, the involvement of CDA was first confirmed in the growth inhibitory phenotype of 5hmdC displayed on other available cell lines with high *CDA* expression levels, such as SN12C and Capan-2 (renal and pancreatic cancer cell lines, respectively). Subsequently, stable knock-downs and overexpression of CDA were established in sensitive (MDA-MB-231 and SN12C) and non-sensitive cell lines (H1299 and MCF7) and treated with the same 5hmdC schedule used in previous experiments. The phenotype could be rescued in sensitive cell lines, and non-responding cell lines were sensitized to the treatment. Later, CDA was purified and the possibility of conversion of 5hmdC, 5fdC and 5cadC was assessed. Deamination of 5hmdC and 5fdC could be achieved, but not that of 5cadC, which instead displayed a mild inhibitory-potential over the deamination of the CDA-preferred substrate dC. 5fdC displayed the second highest V_{max} after dC, and 5hmdC the lowest. In between the two were 5azadC and 5mdC. The slow conversion rate observed for 5hmdC could be explained by an additional hydrogen bond formed between the enzyme and the hydroxyl group, shown in the modeling analysis, making the release of the product slower. In agreement with this finding, HPLC analysis detected 5hmdU in the DNA of 5hmdC-treated cells and its amount correlated well with CDA expression levels. In the literature, it had already been documented that cells treated with 5mdC did not acquire 5mC in their DNA, but instead gained dT after the deamination of 5mdC [Jekunen et al., 1983], and that the resulting cytotoxicity could be reversed by the addition of the CDA inhibitor tetrahydrouridine (THU) [Jekunen and Vilpo, 1984]. In conclusion,

the barrier to 5hmdC incorporation lies in the absence of diphosphate production in the cells. An interesting finding that further supports the evidence of low polymerase selectivity, recently revealed that, in *E. coli*, by overcoming the lack of importers for nucleotide triphosphates with the exogenous expression of genes from *Phaeodactylum tricornutum*, synthetic nucleotides could be imported, incorporated in the DNA upon replication and replicated [Malyshev et al., 2014] (Fig. 8.1).

Having seen that 5fdC could be deaminated at a faster rate compared to 5hmdC and that

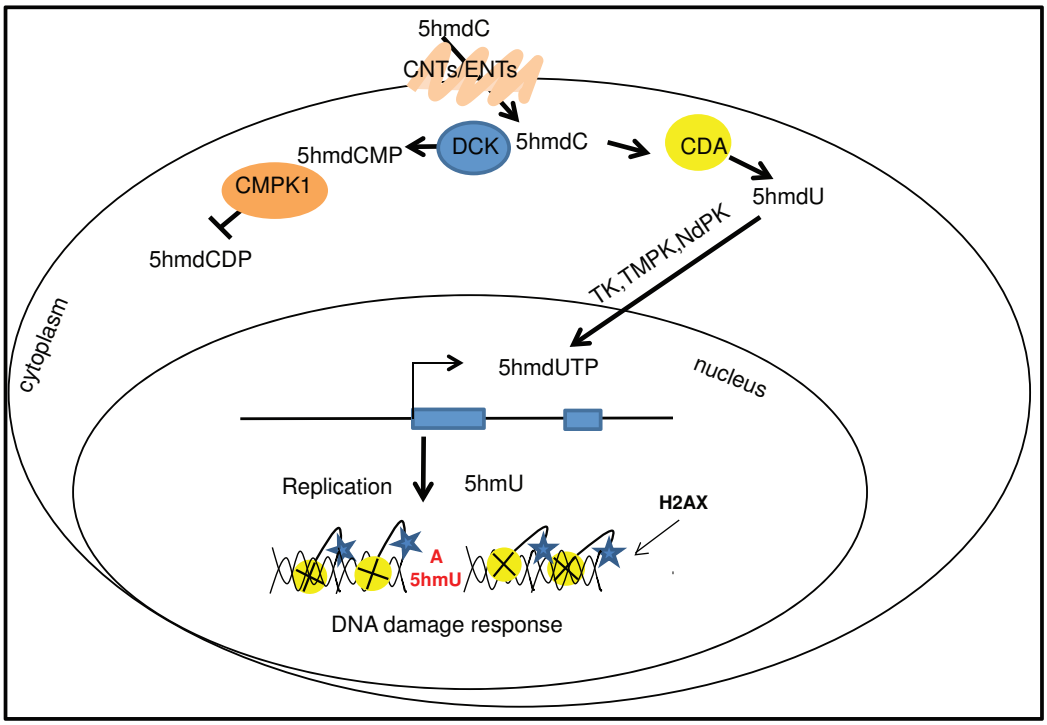


Figure 8.1: Proposed model for 5hmdC salvage and cytotoxicity. According to the data presented, 5hmdC enters into the cells and is either deaminated by CDA or monophosphorylated by DCK. The route to triphosphate is then blocked by CMPK1, which cannot form 5hmdCDP from 5hmdCMP. 5hmdU is then triphosphorylated and incorporated into DNA [Kaufman, 1986], where, following erroneous base pairing with A, it induces double strand breaks, γH2AX and replication arrest.

a diphosphate could not be produced from it, it was tested for cytotoxicity on the same cell lines in which a response to 5hmdC had been detected. A greater growth inhibitory effect for 5fdC, as compared to 5hmdC, could be detected in the same cell lines for which

less than 20% of cells remained at the end of the first experiment. The presence of 5fdU could not be detected in the DNA by the HPLC system in use. It might be useful to assess this again using another HPLC column or a different gradient, which would ensure the separation of 5mdC and 5fdU, which we found to overlap on the column and the conditions used in this study.

Further to the observed cytotoxicity of 5hmdC and 5fdC, an increase in phosphorylated H2AX could be detected, a known marker for double strand breaks and DNA damage [Rogakou et al., 1998][Rogakou et al., 1999]. Some delay in the cell cycle was also observed, such as an increase in the S or G2 phases, indicating damage-induced repair activity or replication arrest. No involvement of p53 could be inferred from the observed response, as the responding and non-responding cell lines had different p53 statuses. Another paper in the past verified the induction of double strand breaks in the DNA following 5hmdU treatment in a p53 knockout cell line, further supporting cell death being caused in treated cells by p53-independent routes [Mi, 2001].

In vivo evaluation of 5hmdC and 5fdC stability and cytotoxic potential

Further to the *in vitro* screen, the antitumoral activity of the compounds were evaluated *in vivo*. It was first established that 5hmdC and 5fdC did not show any sign of toxicity in the organs in which CDA is expressed. Furthermore, it was verified that the compounds were able to enter the bloodstream after intraperitoneal injection and persist there at least until 30 minutes after the injection. This was in comparison with the half life of 5azadC in animal blood, which ranges between 15 and 30 min [Covey and Zaharko, 1985]. However, a detailed pharmacokinetic study needs to be performed with more time points, to assess the persistence of the desired concentration of drug in the blood. Following the toxicological studies, xenograft tumor models were established with the cell lines utilized in the *in vitro* study (H1299, SN12C). Both "CDA elevated" and "CDA low expression" cell lines were injected into each flank of each animal tested, and the established tumors were treated

with 100 mg/kg of 5hmdC or 5fdC. A decrease in tumor volume in the CDA elevated cell lines could be detected. Tumor weight changes occurred mostly upon treatment with 5fdC, which could be explained by the increased potency displayed by this drug *in vitro*. Attempts were made to quantify fibrosis in the treated tumors, but significance could not be obtained due either to the low numbers of animals and the high variability observed or the absence of the phenotype. Improvements in the experimental setting could come from the accurate pre-selection of a homogeneous cohort of tumor induced xenografts. Following injection of the cell lines, slight variability in the timing of tumor induction was observed, which could have been normalized by increasing the number of injected animals and removing the ones that did not reach a homogeneous baseline at day 0 from the treatment schedule. However, this would have required a greater number of animals, which was not feasible for this first study. Small amount of 5hmdU could be detected in the DNA of high CDA-treated tumors, and this correlated with significant induction of γH2AX as measured by confocal microscopy. Certainly, the xenograft model used here is only a first step towards the *in vivo* development of clinically active compounds, therefore, more experiments are needed. Such future experiments could include, for example, a more physiological settings such as the pancreatic cancer mouse model $Kras^{G12D}$, $Trp53^{R172H}$ Pdx-1-Cre [Hingorani et al., 2005], which has been shown to faithfully recapitulate the clinical, histopathological, pharmacokinetic and molecular features of the human disease and which naturally overexpresses CDA [Bapiro et al., 2012].

CDA overexpression in cancer

Clinical interest around human CDA is due to its ability to deaminate both its natural substrate as well as several chemotherapeutic agents, such as the anti-cancer agents 5-azadeoxycytidine (5azadC) and difluorodeoxycytidine (gemcitabine). The deamination of these agents leads to their pharmacological inactivation. For this reason, many CDA inhibitors have been developed while others are under development. By data mining

databases and available genomic data, CDA expression levels were evaluated across a wide range of human cancers and CDA overexpression could be detected in pancreas, testis and stomach cancer (Fig. 8.2). In accordance with this, in mouse models of pancreatic cancer has been shown that treatment with nab-paclitaxel reduces the high CDA expression levels and it is able to further sensitize cells to the activity of gemcitabine [Bapiro et al., 2012]. Additionally, this finding has been validated in clinical trials in human pancreatic cancers underlying the fundamental importance of CDA expression in this setting in relation with response to treatment [Von Hoff et al., 2013].

In vitro, CDA overexpression has been clearly linked to a decreased sensitivity to 5azadC

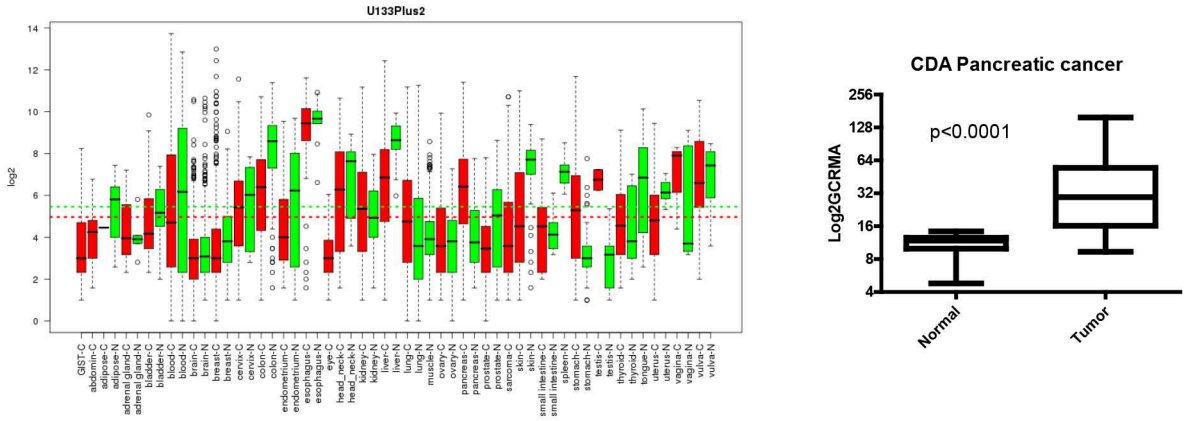


Figure 8.2: CDA expression across human cancers. Red indicates normal, green indicates cancer (data obtained from the GENT database and a pancreatic cancer dataset [Pei et al., 2009]).

[Eliopoulos et al., 1998][Kanno et al., 2007]. Indeed, when overexpressed it is able to rescue the toxicity induced by 5azadC, and only co-treatment with inhibitors of CDA like THU [Eliopoulos et al., 1998] or zebularine is able to reverse the growth phenotype [Lemaire et al., 2008]. Numerous reports link CDA overexpression, with poor clinical response to 5azadC treatment and poor prognosis [Qin et al., 2011][Mercier et al., 2013], with mention of high *CDA* expression observed in males [Mahfouz et al., 2013]. On the other side, *CDA* downregulation has been linked to increased toxicity of 5azadC treatment, with documented cases of death [Mercier et al., 2007]. Among the known genetic variants of CDA

there are K27Q and A70T, which are associated with the polymorphisms 79A>C and 208G>A [Kühn et al., 1993][Laliberté and Momparler, 1994][Yue et al., 2003]. They have been characterized enzymatically, and it has been shown that A70T displays generally low catalytic activity, while K27Q has enhanced catalytic activity towards dC and 5azaC, but not to 5azadC, compared to the K27 variant. On the other hand, enhanced inhibition by THU and zebularine was measured for K27Q [Micozzi et al., 2014]. In agreement with these findings in the clinic, CDA A70T has been associated with higher sensitivity to 5azadC and gemcitabine together with enhanced toxicity [Mercier et al., 2007][Giovannetti et al., 2008], while CDA K27Q-treated patients normally display lower efficacy of 5azaC treatment [Bhatla et al., 2009][Falk et al., 2013]. Moreover, further highlighting the importance of CDA in 5azadC treatment, Down's syndrome AML patients display a high cure rate compared to non-Down's syndrome patients. This is due to low CDA expression paralleled by high expression of DCK, due to inactivating mutations in the transcription factor GATA1 which is present at the *CDA* promoter, in all AML cases observed in Down's syndrome patients [Ge et al., 2004][Ge et al., 2005]. Lastly, another SNP on the promoter (-451C>T), which would lead to the loss of an ETS1 binding site and the gain of a LYF1 site, has been found to promote increased efficacy for cytarabine treatment, linked to decreased expression of CDA [Fitzgerald et al., 2006][Mahlknecht et al., 2009].

The results indicate that CDA overexpression would enable the conversion of both 5hmdC and 5fdC pro-drug into their toxic counterparts, mediating the induction of DNA damage and cell death. Further studies are therefore needed to assess whether sequential or combined treatment with 5azadC and 5fdC or 5hmdC could lead to better prognosis in these cases where CDA levels are elevated. According to the observed k_m values, combined treatment might not be useful, since 5azadC has the lower k_m of the three, which will always allow the compound to out-compete 5hmdC and 5fdC in the active site of the enzyme. On the other hand, it might be that once 5fdC "conquers" the active site of CDA, it would be able to be deaminated faster than any of the other two drugs (according to

the observed V_{max}), enabling the release of a high drug concentration. Therefore, it might prove useful to derivatize 5fdC to improve its affinity to CDA, compared with the one of 5azadC, potentially offering new treatment options. The relative stability of the compounds should also be taken into account and alternative injection schedules be proposed. CDA upregulation, following 5azadC treatment, has been widely proposed as a resistance mechanism [Steuart and Burke, 1971][Qin et al., 2011]. This, together with the increased resistance to targeted therapies that is observed in the clinic, make the efforts directed toward the development of secondary treatment regimes exploiting possible resistance mechanisms much needed. 5hmdC and 5fdC, despite not displaying nanomolar potencies *in vitro*, might offer possible treatment avenues in cases of multidrug resistance. Their activity is hugely dependent on DNA incorporation rate and DNA damage repair, therefore combinatorial treatment with other drugs inhibiting these processes might improve their potency. Following this line, it was previously shown that combination of 5hmdU with a PolyADP ribose inhibitor (PARP) showed an increased response due to inhibition of repair [Boorstein et al., 1987]. Other combinations possibilities might come from epigenetic therapies which are at the moment at the "tipping point" in cancer [Jones, 2014]. Promising results are being shown by the combination of DNA methylation and histone deacytylases inhibitors in solid tumors (small cell lung cancer)[Ahuja et al., 2014] and new avenues of therapies are being tested with the development of drugs targeting other epigenetic proteins (BRD4, DOT1L, EZH2). This approach might offer a priming step necessary to reprogram the cancer genome, making it more homogeneous to prime it to subsequent therapeutic steps, in which more selective inhibitors might be needed. Resistance to first line treatment is increasingly growing and as cancer is defined not only by a hyperproliferative state, but as well by an altered metabolism and by the interaction with the microenvironment and the with the immune system [Hanahan and Weinberg, 2011] targeted therapies, together with immuno-therapies or therapies yet to be developed replacing oncometabolites, offer new exciting treatment opportunities.

8.2 The TET2 interaction network

The ten eleven translocation (TET) family of proteins, despite being known for many years due to the role of *TET1* in translocations occurring in myeloid leukemia [Lorsbach et al., 2003][Ono et al., 2002], were only assigned a function in 2009, when it was discovered that they are responsible for the production of 5hmC in the DNA [Tahiliani et al., 2009]. Given the number of mutations observed for one of its family members, TET2, in myeloid malignancies, this investigation focused on trying to dissect its protein interaction network to shed light on its function in hematopoiesis. The flexibility of the yeast two-hybrid technology was exploited, which allowed a bone marrow-derived library of prey interaction candidates to be selected. Interrogating this question by biochemical techniques would have been complicated in a bone marrow hematopoietic stem cell setting, as cellular material is limited. The bait was constituted by TET2, which was split in two parts at aminoacid 733: TET2A (1-733AA) and TET2B (734-2002AA). Bait and prey were conjugated with the respective GAL4 DNA binding and activation domains and, whenever a productive interaction occurred, the proximity of the two domains activated the expression of reporter genes which were then screened for. A list of putative targets was obtained and validated for not inducing autoactivation when transformed with an empty binding domain. Even though the Y2H screen led to the conclusion that the strength of the interaction was weak, by observing autoactivation for one of the four reporters assessed, the TET2A putative interacting partner WDR61 was selected for further screening. WDR61 is a member of both the PAF and SKI complexes, and participates in the co-transcriptional regulation of RNA synthesis [Zhu et al., 2005][Kim et al., 2010]. Different reports have documented the presence of 5hmC over gene bodies in brain [Mellén et al., 2012], *in vitro* during neuronal differentiation [Hahn et al., 2013] and in both human and mouse ESCs [Xu et al., 2011][Stroud et al., 2011][Pastor et al., 2011][Wu et al., 2011a][Yu et al., 2012] [Lister et al., 2013][Booth et al., 2012][Huang et al., 2014]. Recent literature also points

to TET2 as a member of the family responsible for this deposition, since when it was knocked down, 5hmC levels over gene bodies were reduced [Huang et al., 2014]. Moreover, TET2 lacks a CXXC domain, which is responsible for DNA binding, therefore, alternative recruitment processes must occur to target it to the DNA. In this study it was speculated that WDR61 might be the protein responsible for TET2's targeting over gene bodies.

Screening was begun using a fibroblast-derived cell line (HEK 293T) which stably expressed tagged versions of WDR61 and, following doxycycline induction, TET2A. Coimmunoprecipitation strategies followed by immunoblotting or mass spectrometry (MS) were able to isolate members of the PAF and SKI complexes. Weak evidence of coimmunoprecipitation among the candidate proteins in the cell line studied were observed, leading to the conclusion that, if an interaction occurred, it must be infrequent and unstable. Therefore, microscopy approaches were employed to further validate this finding, which offer the advantage of observing single cell events. Colocalization was used, alongside the proximity ligation assay (PLA), a recently developed technology that allows the detection of fluorescent spots when close proximity among the candidate proteins occurs [Soderberg et al., 2006]. An increase in the number of spots detected per cell and some degree of colocalization between WDR61 and TET2A was seen, suggesting that the interaction might occur. Further validation steps in an endogenous setting are necessary to provide a final answer to the question. An ideal system with which to perform IP/MS analysis of Tet2 interacting partners is that of a recently established mouse cell line (BM-HPC), which fully recapitulates hematopoietic stem cells in its ability to reconstitute all blood lineages in irradiated mice following xenotransplantation [Pinto do O, 1998][Pinto do O, 2002]. This would allow the validation of the interaction in an endogenous bone marrow setting.

To provide a better understanding of the effect that the PAF complex might have on TET2, a 293T 3xFlTET2 cell line was established which could be used to detect TET2 binding over the promoter of *HOXA9*, a known target of the PAF complex [Milne et al.,

2010], further confirming data observed in a published dataset from the same cell line [Deplus et al., 2013]. Therefore, it was speculated that the targeting of the HOXA9 locus by TET2 might upregulate its expression levels. In both 293T cells transiently transfected with TET2 or stably overexpressing it, increased levels of HOXA9 were detected.

The data presented place TET2 in the context of transcriptional elongation and the co-transcriptional regulation of gene expression (Fig. 8.3). In particular, TET2 could promote the positive modulation of transcription, as it has been documented via the interaction with OGT and HCF of the SET/COMPASS complex, responsible for the production of H3K4me3 at gene promoters through its catalytic component SETD1 [Deplus et al., 2013]. Successively, transcription might be favored once the RNA polymerase was successfully engaged and the PAF complex was recruited by TET2 mediated 5hmC deposition over gene bodies (Fig. 8.3).

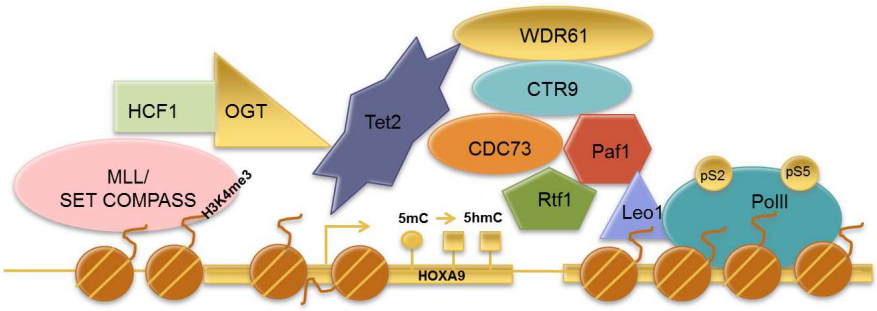


Figure 8.3: Proposed Tet2 working model. The interaction between TET2 and OGT has been widely documented [Chen et al., 2012b][Vella et al., 2013][Deplus et al., 2013], as well as the interaction with the SET/COMPASS complex [Deplus et al., 2013]. Functional interaction of Tet2 with WDR61, a member of the PAF complex, could occur on the HOXA9 locus and induce its transcription. It is hypothesized that 5hmC is deposited over this gene body following PAF-mediated TET2 recruitment. 5hmC has already been detected over this locus [Bocker et al., 2012].

Tet2 and cancer

TET2 has been shown to be highly mutated in hematopoietic cancers, probably linked to its high expression levels in this context. Tet2 knockout bone marrow cells display a

global reduction in H3K4me3 at Tet2-targeted promoters [Deplus et al., 2013], expansion of the hematopoietic stem cell compartment, biased differentiation towards the myelomonocytic lineage and myeloid cell transformation resembling MDS [Moran-Crusio et al., 2011]. Moreover, it has been shown that its presence is important for the induction of myeloid targets [Kallin et al., 2012] further supporting a potential role of TET2 in transcription. In embryonic stem cells, a reduction in the presence of 5hmC over gene bodies, mainly those of highly expressed genes, was observed upon Tet2 downregulation [Huang et al., 2014]. Moreover, *DNMT3A*, another gene important for the establishment of post replicative methylation is frequently mutated in blood cancer, sometimes together with TET2 [Wakita et al., 2012]. DNMT3A mutant AML and knockout mice hematopoietic stem cells have been shown to display global reduction in DNA methylation, localized at large genomic regions (canyons, or shores) of low CpG density, including gene bodies [Raddatz et al., 2012][Jeong et al., 2014][Russler-Germain et al., 2014]. HoxA9 and Meis1 have been found present in these regions in hematopoietic stem cells [Jeong et al., 2014]. Furthermore, Dnmt3a knockout mice display impaired hematopoietic differentiation [Challen et al., 2012], underlying the important role of DNA methylation regulation in this process, where it seems to be required for the establishment of defined transcriptional programs.
On the other hand, deletion of Paf complex members in yeast has been associated with growth defects ranging from mild in the case of RTF1 and LEO1 nulls, to strong for PAF1 and CTR9 [Squazzo et al., 2002], and sensitivity to the nucleotide depleting drug 6-azauracil [Costa and Arndt, 2000], establishing thus a link with transcriptional elongation [Exinger and Lacroute, 1992]. Given the prominent role of the PAF complex in transcriptional elongation [Zhu et al., 2005] [Kim et al., 2010] and its localization over gene bodies detected in both yeast [Krogan et al., 2002][Pokholok et al., 2002][Simic et al., 2003] and mammals [Rahl et al., 2010], this interaction might be the missing link between Tet2, 5hmC over gene bodies and the transcriptional effects of 5hmC on gene expression. Future efforts could be directed towards the precise mapping of the region of interac-

tion with its eventual correlation with a cancer-mutated residue. In TET2A (1-733AA) a number of mutations have been documented. Among these, the ones harboring missense mutations that might indicate a loss/gain of function and that are evolutionarily conserved are promising candidates for future screens (Fig. 8.4). The interaction could

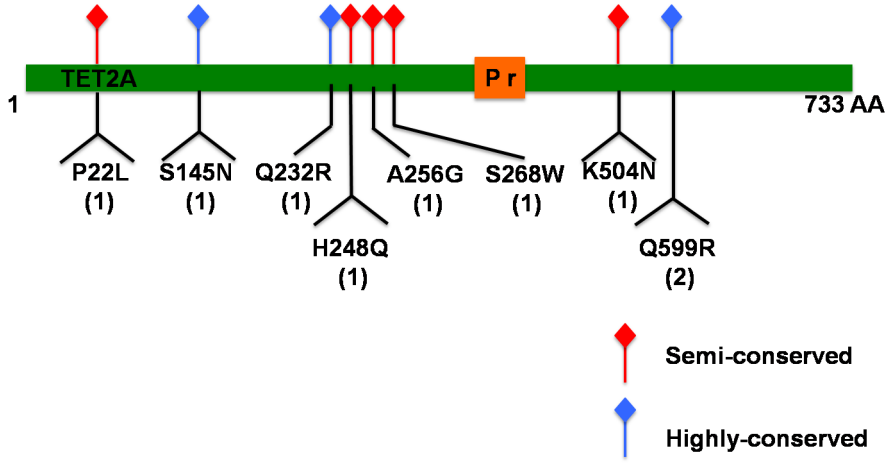


Figure 8.4: TET2A (1-733AA) missense mutations in conserved residues in hematopoietic malignancies. Residues selected for sequence conservation (Appendix 2). Semi-conserved indicates residues that are overall conserved except for *Danio rerio*. Pr indicates a proline rich domain of Tet2. The number of times the mutation has been observed is indicated in brackets (data obtained from COSMIC database [Forbes et al., 2001]).

either be lost in mutant case or strengthened. This could lead to a deregulated expression of *HOXA9*, and thus to myeloproliferation. The same has been documented in MLL rearranged leukemia that display an increased HOXA9 expression compared to wild type MLL [Milne et al., 2010]. Furthermore, MLL fusions are able to target Tet1 and increase the expression levels of one of its target, Hoxa9. In both Tet1 knockdown and knockout background the effects of MLLAF9 fusion on transformation are largely reduced and can be rescued by enforced expression of HoxA9 [Huang et al., 2013]. DNA methylation might regulate HoxA9 expression levels and its deregulation, both in DNMT3A mutants and TET2 mutants, might be fundamental for the correct maintenance of a balanced differentiation. If the interaction can be mapped to a single residue, a functional explanation could be provided for a direct role of TET2 mutations in leukemia. Furthermore, as for

the compound PRIMA-1 that could reactivate the transactivating function of mutated p53 and suppress tumor growth [Bykov et al., 2002], pharmacological targeting of particular mutations linked to a function, might prove useful for drug development.

Appendix 1

Rank	ID	P.Value	R	N R	logFC	Gene symbol
1	202628_s_at	3.32E-06	21.38705	2.0901	1.26	SERPINE1
2	211343_s_at	1.1E-05	16.8768	1.8351	1.14	COL13A1
3	202686_s_at	1.75E-05	15.3855	1.7073	1.46	AXL
4	212298_at	1.92E-05	15.10879	1.6802	9.99E-01	NRP1
5	204286_s_at	2.12E-05	14.80968	1.6494	1.02	PMAIP1
6	213524_s_at	2.6E-05	14.21676	1.5839	1.34	G0S2
7	217388_s_at	4.52E-05	-12.7312	1.388	-1.16	KYNU
8	202627_s_at	5.04E-05	12.45538	1.3457	1.04	SERPINE1
9	201417_at	5.75E-05	-12.1327	1.2935	-9.72E-01	SOX4
10	213712_at	6.65E-05	-11.7812	1.233	-8.55E-01	ELOVL2
11	209278_s_at	8.83E-05	11.12617	1.1092	9.77E-01	TFPI2
12	212771_at	9.1E-05	11.05784	1.0954	6.80E-01	FAM171A1
13	202454_s_at	0.000102	-10.8116	1.0441	-6.78E-01	ERBB3
14	219014_at	0.000109	10.66642	1.0127	1	PLAC8
15	205005_s_at	0.000113	10.57735	0.993	6.38E-01	NMT2
16	218718_at	0.000134	10.22245	0.911	1.12	PDGFC
17	203186_s_at	0.000148	10.02046	0.8617	9.28E-01	S100A4
18	204363_at	0.000151	9.982162	0.8522	1.05	F3
19	209369_at	0.000174	9.689985	0.7767	1.02	ANXA3
20	204285_s_at	0.000176	9.670323	0.7714	8.51E-01	PMAIP1
21	209270_at	0.000179	9.639462	0.7632	6.82E-01	LAMB3
22	201042_at	0.000183	9.592096	0.7504	8.84E-01	TGM2
23	211668_s_at	0.000204	9.380235	0.6917	8.11E-01	PLAU
24	205443_at	0.000205	9.374704	0.6901	6.08E-01	SNAPC1
25	203304_at	0.000256	-8.95716	0.5665	-8.45E-01	BAMBI
26	218322_s_at	0.000283	8.769381	0.5074	5.69E-01	ACSL5
27	205016_at	0.000313	8.592364	0.4495	6.59E-01	TGFA
28	209946_at	0.000334	8.473995	0.4095	5.75E-01	VEGFC
29	205286_at	0.000352	-8.38528	0.3789	-7.20E-01	TFAP2C
30	219648_at	0.000354	-8.37511	0.3754	-6.22E-01	MREG
31	212473_s_at	0.000361	8.341889	0.3637	8.53E-01	MICAL2
32	204222_s_at	0.000397	8.176546	0.3046	5.70E-01	GLIPR1
33	206373_at	0.000432	-8.0339	0.2519	-6.82E-01	ZIC1
34	205479_s_at	0.000445	7.986973	0.2342	8.51E-01	PLAU
35	212472_at	0.000445	7.984808	0.2334	6.49E-01	MICAL2
36	200878_at	0.000451	7.963447	0.2252	5.16E-01	EPAS1

Continued on next page

Table 1 – continued from previous page

Rank	ID	p Value	R	N R	logFC	Gene symbol
37	212509_s_at	0.000457	7.940827	0.2166	5.80E-01	MXRA7
38	207876_s_at	0.000471	7.892574	0.198	5.14E-01	FLNC
39	203939_at	0.000479	7.862631	0.1864	8.02E-01	NT5E
40	201425_at	0.000509	7.764303	0.1477	5.69E-01	ALDH2
41	201850_at	0.000524	7.716854	0.1288	6.92E-01	CAPG
42	204702_s_at	0.000555	7.625343	0.0916	6.08E-01	NFE2L3
43	220057_at	0.000579	7.555522	0.0628	9.24E-01	XAGE1
44	218711_s_at	0.000595	7.512537	0.0448	5.48E-01	SDPR
45	202558_s_at	0.000604	7.4888	0.0348	5.44E-01	HSPA13
46	201951_at	0.000617	7.455448	0.0207	6.48E-01	ALCAM
47	201983_s_at	0.000626	7.433107	0.0112	7.82E-01	EGFR
48	210845_s_at	0.000668	7.331663	-0.0327	5.64E-01	PLAUR
49	201466_s_at	0.000722	7.213481	-0.085	4.63E-01	JUN
50	203725_at	0.000769	7.117087	-0.1286	4.97E-01	GADD45A
51	219489_s_at	0.000771	7.113572	-0.1303	4.90E-01	NXN
52	204268_at	0.000801	7.056125	-0.1567	4.78E-01	S100A2
53	218451_at	0.000806	7.045857	-0.1615	6.53E-01	CDCP1
54	220615_s_at	0.000855	-6.95941	-0.2021	-4.88E-01	FAR2
55	210180_s_at	0.000884	-6.90878	-0.2263	-4.08E-01	TRA2B
56	219109_at	0.000889	6.901215	-0.2299	5.34E-01	SPAG16
57	203828_s_at	0.001007	6.719391	-0.319	4.48E-01	IL32
58	205097_at	0.001033	-6.68349	-0.337	-7.08E-01	SLC26A2
59	205287_s_at	0.001037	-6.678	-0.3398	-4.38E-01	TFAP2C
60	213603_s_at	0.001139	6.543999	-0.4085	5.10E-01	RAC2
61	205263_at	0.001167	6.509905	-0.4263	4.36E-01	BCL10
62	201795_at	0.001183	6.491913	-0.4357	4.10E-01	LBR
63	209723_at	0.001192	6.480356	-0.4418	3.87E-01	SERPINB9
64	201474_s_at	0.001209	6.461044	-0.452	4.74E-01	ITGA3
65	211962_s_at	0.001217	6.451736	-0.457	3.99E-01	ZFP36L1
66	210663_s_at	0.001266	-6.39681	-0.4863	-4.09E-01	KYNU
67	204653_at	0.001273	-6.38911	-0.4904	-4.49E-01	TFAP2A
68	215704_at	0.001289	6.372611	-0.4993	8.98E-01	FLG
69	218959_at	0.001303	-6.35727	-0.5076	-5.12E-01	HOXC10
70	203889_at	0.001307	6.353358	-0.5098	7.26E-01	SCG5
71	205006_s_at	0.001411	6.248839	-0.5671	6.31E-01	NMT2
72	203323_at	0.001421	6.239401	-0.5724	7.97E-01	CAV2
73	212256_at	0.001455	6.207026	-0.5905	4.82E-01	GALNT10
74	211732_x_at	0.00146	6.202649	-0.5929	5.22E-01	HNMT
75	213075_at	0.001467	6.196201	-0.5966	4.43E-01	OLFML2A
76	218880_at	0.001467	6.196158	-0.5966	4.23E-01	FOSL2
77	218980_at	0.001551	6.121294	-0.639	3.91E-01	FHOD3
78	203153_at	0.001699	6.000536	-0.709	5.88E-01	IFIT1
79	205627_at	0.001702	5.998686	-0.7101	6.58E-01	**CDA**
80	209817_at	0.001711	5.991519	-0.7143	3.62E-01	PPP3CB
81	201340_s_at	0.001715	5.988797	-0.7159	4.99E-01	ENC1
82	220235_s_at	0.001748	5.963318	-0.731	4.19E-01	LRIF1
83	220945_x_at	0.001751	-5.96112	-0.7323	-4.05E-01	MANSC1
84	204112_s_at	0.001798	5.926934	-0.7526	5.88E-01	HNMT
85	205596_s_at	0.001814	5.915408	-0.7595	6.84E-01	SMURF2

Continued on next page

Table 1 – continued from previous page

Rank	ID	p Value	R	N R	logFC	Gene symbol
86	215073_s_at	0.001838	5.898253	-0.7698	5.86E-01	NR2F2
87	209908_s_at	0.001888	5.863639	-0.7907	3.88E-01	TGFB2
88	201289_at	0.001895	5.858693	-0.7937	1.01	CYR61
89	210764_s_at	0.001912	5.847167	-0.8007	8.27E-01	CYR61
90	204654_s_at	0.001955	-5.81864	-0.8181	-4.94E-01	TFAP2A
91	207469_s_at	0.001959	-5.81638	-0.8195	-6.85E-01	PIR
92	201810_s_at	0.00204	-5.76414	-0.8516	-3.43E-01	SH3BP5
93	218854_at	0.00213	5.709487	-0.8857	5.67E-01	DSE
94	203741_s_at	0.002138	5.705166	-0.8884	3.51E-01	ADCY7
95	202291_s_at	0.002195	-5.67154	-0.9096	-5.51E-01	MGP
96	207419_s_at	0.00226	5.635078	-0.9328	5.45E-01	RAC2
97	210396_s_at	0.002313	-5.6065	-0.9511	-3.62E-01	LOC595101
98	218317_x_at	0.002327	-5.59848	-0.9562	-3.32E-01	SLX1A
99	209457_at	0.002376	5.573047	-0.9726	5.48E-01	DUSP5

Table 1: Gene expression analysis results comparing responding (MDA-MB-231 and HOP-92) and not responding (MCF7 and MDA-MB-435) cell lines data. First 99 hits ranked by p value (t-test with Benjamini-Hochberg correction).

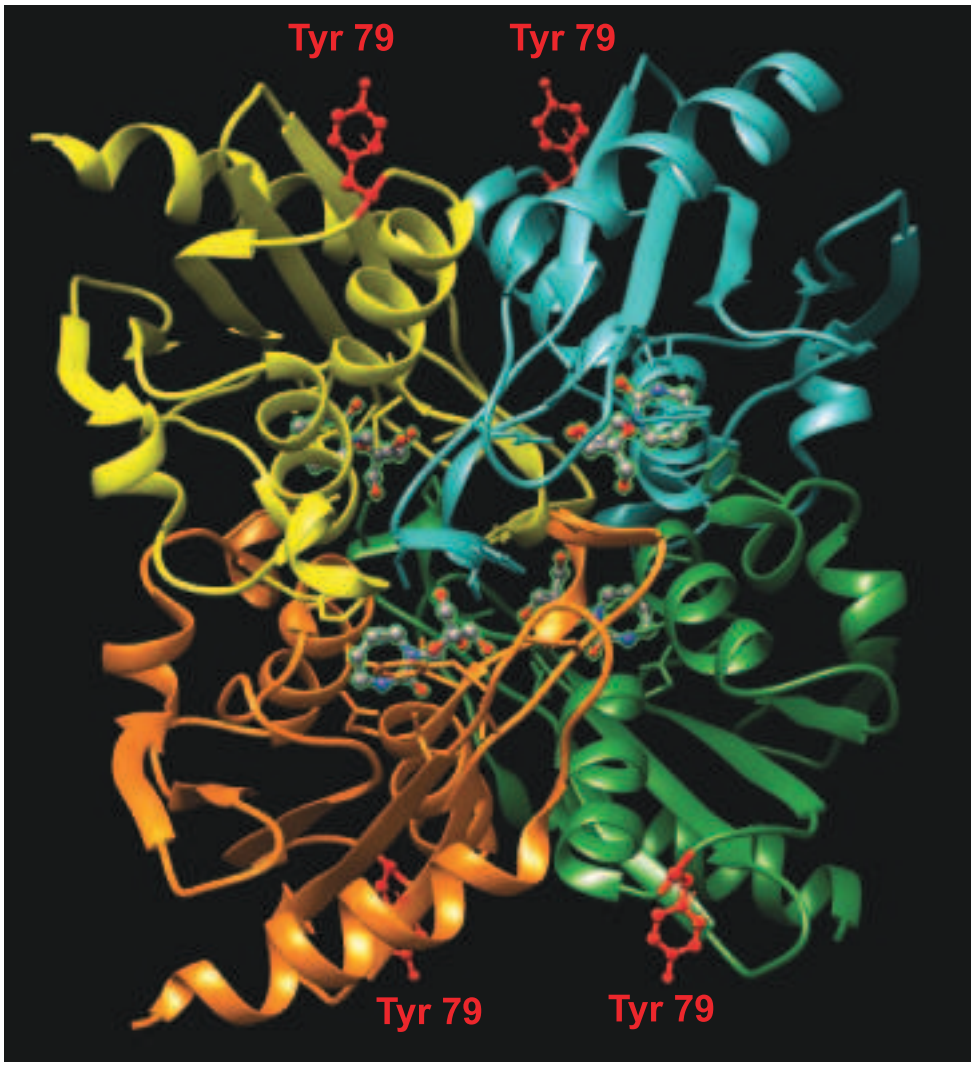
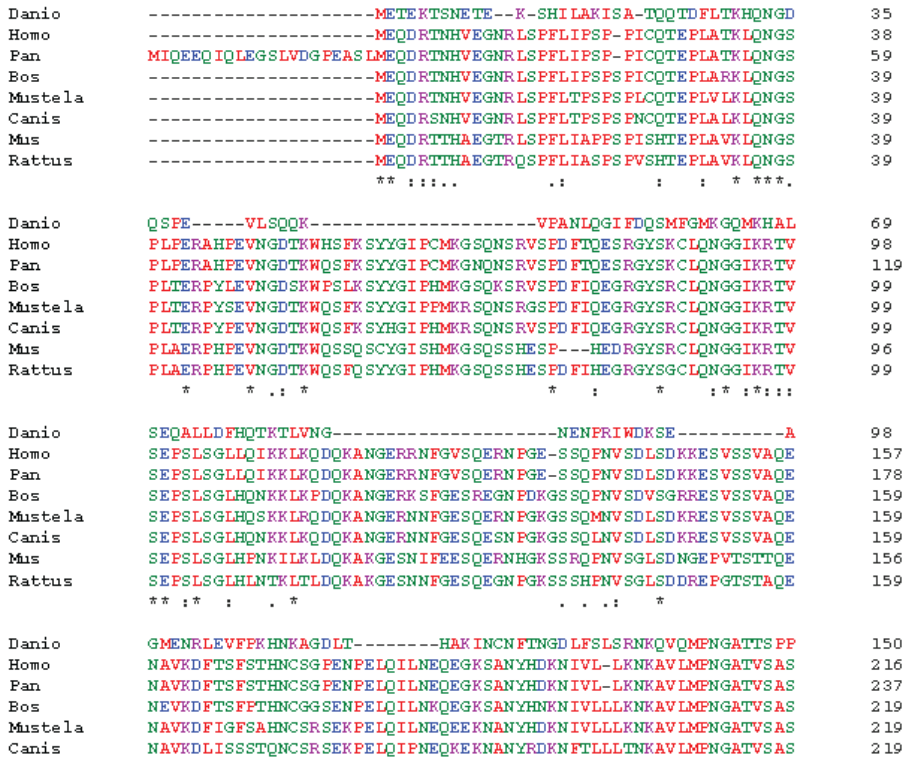


Figure 5: Localization of Y79, putative phosphorylation site of CDA. In different color the different monomers and in red/ balls sticks is highlighted the position of Y79 [Chung et al., 2005]. Zn is visible in grey and diazepinone riboside in its atom colors.

Appendix 2


Danio ----------METEKTSNETE--K-SHILAKISA-TQQTDFLTKHQNGD 35
Homo ----------MEQDRTNHVEGNRLSPFLIPSP-PICQTEPLATKLQNGS 38
Pan MIQEEQIQLEGSLVDGPEASLMEQDRTNHVEGNRLSPFLIPSP-PICQTEPLATKLQNGS 59
Bos ----------MEQDRTNHVEGNRLSPFLIPSPSPICQTEPLARKLQNGS 39
Mustela ----------MEQDRTNHVEGNRLSPFLTPSPSPLCQTEPLVLKLQNGS 39
Canis ----------MEQDRSNHVEGNRLSPFLTPSPSPNCQTEPLALKLQNGS 39
Mus ----------MEQDRTTHAEGTRLSPFLIAPPSPISHTEPLAVKLQNGS 39
Rattus ----------MEQDRTTHAEGTRQSPFLIASPSPVSHTEPLAVKLQNGS 39
** :::.. .: : : * ***.
Danio QSPE-----VLSQQK----------VPANLQGIFDQSMFGMKGQMKHAL 69
Homo PLPERAHPEVNGDTKWHSFKSYYGIPCMKGSQNSRVSPDFTQESRGYSKCLQNGGIKRTV 98
Pan PLPERAHPEVNGDTKWQSFKSYYGIPCMKGNQNSRVSPDFTQESRGYSKCLQNGGIKRTV 119
Bos PLTERPYLEVNGDSKWPSLKSYYGIPHMKGSQKSRVSPDFIQEGRGYSRCLQNGGIKRTV 99
Mustela PLTERPYSEVNGDTKWQSFKSYYGIPPMKRSQNSRGSPDFIQEGRGYSRCLQNGGIKRTV 99
Canis PLTERPYPEVNGDTKWQSFKSYHGIPHMKRSQNSRVSPDFIQEGRGYSRCLQNGGIKRTV 99
Mus PLAERPHPEVNGDTKWQSSQSCYGISHMKGSQSSHESP---HEDRGYSRCLQNGGIKRTV 96
Rattus PLAERPHPEVNGDTKWQSFQSYYGIPHMKGSQSSHESPDFIHEGRGYSGCLQNGGIKRTV 99
* * .: * * : * :* :*:::
Danio SEQALLDFHQTKTLVNG----------NENPRIWDKSE----------A 98
Homo SEPSLSGLLQIKKLKQDQKANGERRNFGVSQERNPGE-SSQPNVSDLSDKKESVSSVAQE 157
Pan SEPSLSGLLQIKKLKQDQKANGERRNFGVSQERNPGE-SSQPNVSDLSDKKESVSSVAQE 178
Bos SEPSLSGLHQNKKLKPDQKANGERKSFGESREGNPDKGSSQPNVSDVSGRRESVSSVAQE 159
Mustela SEPSLSGLHQSKKLRQDQKANGERNNFGESQERNPGKGSSQMNVSDLSDKRESVSSVAQE 159
Canis SEPSLSGLHQNKKLKQDQKANGERNNFGESQESNPGKGSSQLNVSDLSDKRESVSSVAQE 159
Mus SEPSLSGLHPNKILKLDQKAKGESNIFEESQERNHGKSSRQPNVSGLSDNGEPVTSTTQE 156
Rattus SEPSLSGLHLNTKLTLDQKAKGESNNFGESQEGNPGKSSSHPNVSGLSDDREPGTSTAQE 159
** :* : . * . . .: *
Danio GMENRLEVFPKHNKAGDLT--------HAKINCNFTNGDLFSLSRNKQVQMPNGATTSPP 150
Homo NAVKDFTSFSTHNCSGPENPELQILNEQEGKSANYHDKNIVL-LKNKAVLMPNGATVSAS 216
Pan NAVKDFTSFSTHNCSGPENPELQILNEQEGKSANYHDKNIVL-LKNKAVLMPNGATVSAS 237
Bos NEVKDFTSFPTHNCGGSENPELQILNKQEGKSANYHNKNIVLLLKNKAVLMPNGATVSAS 219
Mustela NAVKDFIGFSAHNCSRSEKPELQILNEQEEKNANYHDKNIVLLLKNKAVLMPNGATVSAS 219
Canis NAVKDLISSSTQNCSRSEKPELQIPNEQKEKNANYRDKNFTLLLTNKAVLMPNGATVSAS 219

```
Mus          SS--GADAFPTRNYNGVEIQV---LNEQEGEK-----GRSVTLLKNKIVLMPNGATVSAH      206
Rattus       ST--DIDAFPTRNYNGVENTE--IPEEQEGGN-----DRNALLLKNKAVLMPNGATVSAP      210
                       :*             :   .            ** * ******.*

Danio        SVKGTTGDLLEKTLSQYYPEHVSIAPQTNSS-QVDQVTSNLP---DQATSSPSFTSGFPI      206
Homo         SVEHTHGELLEKTLSQYYPDCVSIAVQK-TTSHINAINSQATNELSCEITHPSHTSGQIN      275
Pan          SMEHTHGELLEKTLSQYYPDCVSIAVQK-TTSHINAINSQATNELSCEITHPSHTSGQIN      296
Bos          SMENTHGELLEKTLSQYYPDCVSIAVQK-TTSHIHAINSQATNELSCEVTHPAHTSGQIN      278
Mustela      SMENTHGELLEKTLSQYYPDCVSVAVQK-TTSHIHAINSQATNELSCEITHPSHTSGQIN      278
Canis        SMENTHGELLEKTLSQYYPDCVSVAVQK-TTSHKHAINSQATNELSCEITHPSHTSGQIN      278
Mus          SEENTRGELLEKT--QCYPDCVSIAVQS-TASHVNTPSSQAAIELSHEIPQPSLTSAQIN      263
Rattus       SVEDTHGEFLEKT--QCYPDCVSMAVQNNTTPHVNALSGQAAIELPREITQPSLTSAQIN      268
             * : * *::****  * **: **:* *. :: : .  ..:           *: **.

Danio        SPQISASETLVKAPP------EVQNSNGYNPELAVNGYSSFAAEQ----KKPAYSLSDLP      256
Homo         SAQTSNSELPPKPAAVVSEACDADDADNASKLAAMLNTCSFQKPEQLQQQKSVFEICPSP      335
Pan          SAQTSNSELPPKPAAVVTEACDADDADNASKPAAMLNTCSFQKPEQLQQQKSVFEICPSP      356
Bos          FPQTSNSELPPEPAAAATESCEADG---ASKPAATLGACPFQKPEQ---QNSAFEIRPSL      332
Mustela      VPQTSNSELPPEPVAVVTEACDADK---ASKTGAVIGTCPFQKPEK---QKSVFDRCPSR      332
Canis        LPQTSNSELRPEPVAVVTEACDADK---ASKTAAMIGTCPFQKPEK---QKLLFDRCPSR      332
Mus          FSQTSSLQLPPEPAAMVTKACDADN---ASKPAIVPGTCPFQKAEH--QQKSALDIGPSR      318
Rattus       FSQTSSAQLPPEPAAMVTKACDLDS---ASKPAGVLGTCAFQKAEH--QQKSVLGIGPSS      323
               * *  :   :         : :     .        . *   :    ::

Danio        ELSQKDSESNVPSQTVVNGSSQIQEDREALPN-----------------DSECVNMFSKS      299
Homo         -----------AEN-NIQGTTKLASGEEFCSGSSSNLQAPGGSSERYLKQNEMNGAYFKQ      383
Pan          -----------AEN-NIQGTTKLASGEEFCSGSSSNLQAPGGSSERYLKQNEMNGAYFKQ      404
Bos          -----------AENSNIQATAKLGSGEEFCLGPSGNLQSPSGSSERYLKQNEMNGACFKQ      381
Mustela      -----------AENSTIQGTTELVSGEEFCSGSSSNLQAPGGSSERYLKQNETNGAYFKQ      381
Canis        -----------AENSNIQGTTKLVSGEEFCSGSSSNLQAPGGSSERYLKQNETNGAYFRQ      381
Mus          -----------AENKTIQGSMELF-AEEYYPSSDRNLQASHGSSEQYSKQKETNGAYFRQ      366
Rattus       --C---------AENINIQGNTKPF-AEEYYAGSDSDLQTSLGSSEQCIKQKETNGAYFRQ      372
                         .:  ::.. :    .*                    :.*      :.

Danio        NQNLSHGSYTLPPVTAGSPSQTSQTGQYNPQIKVPQK-----KTT-DENLHG------NR      347
Homo         SSVFTKDSFSAT--TTPPPPSQLLLSPPPPLPQVPQLPSEGKSTLNGGVLEEHHHYPNQS      441
Pan          SSVFTKDSFSAT--TTPPPPSQLLLSPPPPLPQVPQLPSEGKSTLNGGVLEEHHHYPNQS      462
Bos          SSVFTKDSFSAT--TTP-PPSQLLLSPPPPLPQVPQLPSQGNGTLNDGVLEELHHYPNQS      438
Mustela      SSVFSKDSFSAP--TTPPPPSQLRLSPPPPLPEVSQLPSEGKSTLNDGVLEEHHHYPNQS      439
Canis        SSVFTKDSFSAP--TTPP-PSQLLLSPPTPLPQVSQLPSEGKSTLNDGVLEEHHHYPNQS      438
Mus          SSKFPKDSISPT--TVTPPS-QSLLAPR----LVLQPPLEGKGALNDVALEEHHDYPNRS      419
Rattus       SSVFPEDSISAT--TMTPPS-PSPLAPR----PVLQLPSEGKGTLNDVALEEHHRYPNQS      425
             .. : . * :    *          .       * *       :     *.       .

Danio        DFSLLQQQT--HEPQAVE--------------SKLIPQ----ACT---------------      372
Homo         NTTLLREVKIEGKPEA-PPSQSPNPSTHVCSPSPMLSERPQNNCVNRNDIQTAGTMTVPL      500
Pan          NTTLLREVKIEGKPEA-PPSQSPNPSTHVCSPSPMLSERPQNNCVNRNDIQTAGTMTVPL      521
Bos          DPTLLREVKIEGQHEA-PPPQSPTPSAHVSNPSLMLPVRPQNNCVNKNDSRTPGTMTIPL      497
Mustela      NTALLREVKIEGQPEA-RPSPSPNPSTRASNHSLMLPERPQNNCVNQNDIQTPGTMTVPA      498
Canis        NTALLREVKIEGQCEA-QPSPSPNPSTHIPNHSLMLPERPQNNCVNKNDIQTPGMMTVPA      497
Mus          NRTLLREGKIDHQPKT-SSSQSLNPSVHTPNPPLMLPEQHQNDCG-------------SP      465
Rattus       DLTLSREGKTEHQPKKTTSSQSLNTSVHTPNPPLMLPEQHQNDCV-------------PP      472
             : :* :: .   : :                    ::       *

Danio        --SPS-------------QNRNN-DQDNKS--------------KLGLEKTLSKGTDPSS      402
Homo         CSEKTRPMSEHLKHNPPIFGSSGELQDNCQQLMRNKEQEILKGRDKEQTRDLVPPTQHYL      560
Pan          CSEKTRPMSEHLKHNPPIFGRSGELQDNCQQLMRNKEQEILKGRDKEQTRDLVPPTQHYL      581
Bos          GSEKTRQLSEHLKPNPPILRSGGDPQDHCQHLMGHKQQEILKSQDKEQTRDLMLPTQCYL      557
Mustela      CSEKARQISEHIMHNPPMLGSSGDPQDHCQRLMGHKEQDIPKGRDKEQTRGLVLPAQPSL      558
Canis        CSEKVRQISEHLKHNPTILGSSGDPQDHCQQLMGHREQDILKGRDKEQTRGLVLPAQPYL      557
Mus          SPEKSRKMSEYLMYYLPNHGHSGGLQEHSQYLMGHREQEIPKDANGKQTQGSVQ-----A      520
Rattus       CSEKTRKVSDYRMYYLPNHGHGGDLQENSQYLMGHGEHEILKDANKGQTQGSVL-----A      527
               .                    *:. .              .    :

Danio        QMEWIDLNSTSASTPSTEHSQMWDLFLPQTNIQQHTVPQSQSQCINSIPPNNFQSHNFSN      462
```

```
Homo        KPGWIELKAPRFHQAESH-LKRNE-------ASLPSILQYQPNLSNQMTSKQYTGNSNMP      612
Pan         KPGWIELKAPHFHQAESH-LKRNE-------ASLPSILQYQPNLSNQMTSKQYTGNSNMP      633
Bos         KPGWIELKAPHYRQAESH-LKCNE-------TTLRSILQYQSNPSNQMTSKQYTGNSNTP      609
Mustela     KPGWIELKPPHFHQAESH-PKCPE-------ASLRSVLQYQSNPANQMTSKQYTGNSSLP      610
Canis       KPGWIELKPPHFHQAESH-PKCHE-------ASLRSVLQYQSNPSNQMTSKQYTGNSSLP      609
Mus         APGWIELKAPNLHEALHQ-TKRKD-------ISLHSVLHSQTGPVNQMSSKQSTGNVNMP      572
Rattus      APGWIELKASHFHKALHQ-TECKA-------TSLHSGLRYQTSPTNQIGSIQSTGSVNMT      579
               **:*:         .  :             :  : *    *.:   :  .

Danio       SSFPSHEIEAQENKNMLPPTRFPHSAAPESQQRNSGMHNHINMQFNNPQH----------      512
Homo        GGLPRQAY---TQKTTQLEHKSQMYQVEMNQGQSQGT-VDQHLQFQKPSHQVHFSKTDHL      668
Pan         GGLPRQAY---TQKTTQLEHKSQMYQVEMNQGQSQGT-VDQHLQFQKPSHQVHFSKTDHL      689
Bos         GGLPGQAY---IQKIMQPEQRPPRYQVEMNQGQSQGT-VDQHLQFQNPSLQAHFSKVEPS      665
Mustela     GGLPGQAY---IQKIMQPEQRPQRYQGEMNREQSQGT-LDQQLQFQKPSLQVHFSKTDPS      666
Canis       GGLPGQAY---IQKIMQPEHRPQRYHGEMNQGQSQGT-VDQQLQFQKPSLQVHFSKTDPS      665
Mus         GGFQRLPY---LQKTAQPEQKAQMYQVQVNQGPSPGM-GDQHLQFQKALYQECIPRTDPS      628
Rattus      GEFQRPPY---IQKTAQPEHKSQMYQVEMNQGPSPGM-GDQHLQFQKTSYQECISRTDLP      635
            . :          :*      :         .:  . *   . .:**::

Danio       ---MSQKLCK----------NDQDQILSPS-------------------FLSQLPPTETH      540
Homo        PKAHVQSLCGTRFHFQQRADSQTEKLMSPVLKQHLNQQASETEPFSNSHLLQHKPHKQAA      728
Pan         PKAHVQSLCGTRFHFQQRADSQTEKLMSPVLKQHLNQQASETEPFSNSHLLQHKPHKQAA      749
Bos         PDTHRQPLCAPRWHFQQRPDPQTEKLMPPTLRQHLNQQTSETEPFSNSHLLQHKPHKQAA      725
Mustela     SKAHMQSLCTHRFYFEQRPDSQTEKLLPASLKQHLNQQASETEPFSNSHILQHKPHKQAA      726
Canis       SKAHMQSLCAHRFYFEQRPDSQTEKLIPPTLKQHLIQQASETEPFSNSHLLQHKPHKQAA      725
Mus         SEAHPQAPSVPQYHFQQRVNPSSDK--------HLSQQATETQRL--SGFLQHTPQTQAS      678
Rattus      PEA-----HGPQYHFHQRVDPSTDK--------Q---QATEAELF--SSFLQHTPHKQAA      677
                              . ::                          :*.: * .::

Danio       QQAT-STFSNQLPQDR-----------QNM---------QASPLSRDASRSQEVEPLLKK      579
Homo        QTQ--PSQSSHLP--------QNQQQQQKLQIKNKEEILQTFPHPQSNNDQQREGSFFGQ      778
Pan         QTQ--PSQNSHLP--------QNQQQQQKLQIKNKEEILQTFPHPQSNNDQQREGSFFGQ      799
Bos         QTQLPQ--NSHLS--------QNQQQQQKLQLKNKEQMPQTFSHLQGNSDEPREGTFFSQ      775
Mustela     QTQTQPSQNSLISQNQ--------QQPQKLQMKNKEQMPQTFSHPQGNNDQQREGSFFSQ      778
Canis       QTQTQPSQSSLISQNHHQQQQQQQQQQQKLQMKNKEQMPQTFSQSQGNNDQQREGSFFSQ      785
Mus         QTP--ASQNSNFPQI------CQQQQQQQLQRKNKEQMPQTFSHLQGSNDKQREGSCFGQ      730
Rattus      QTQ--ASQNSNFPQV------CQQQ--HQLQRKNKEQMPQTFSHLHGTNHEQSQGQCFGT      727
            *       .. :               :::         *:    :  . .      :

Danio       LKTEDYLHSEFQMLPAC-IKTPAQVQIS----SKAIRENFGFQDSSKSNTPSLFKE----      630
Homo        TKVEECFHGENQYSKSSEFETHN-VQMGLEEVQNINRRNSPYSQTMKSSACKIQVSCSNN      837
Pan         TKVEECFHGENQYSKSSEFETHN-VQMGLEEVQNINRRNSPYSQTMKSSACKIQVSCSNN      858
Bos         VKVEECFPGEDQYSKSSEFQTHN-PQMGLEQVQNINSINSPYGQILKSNANKVQISGSNH      834
Mustela     IKVEECFHGENQYLKSSEFQTPT-AQMGLEQVQNMNSRNPPYGQILKSNPSKLQIPCSNN      837
Canis       IKVEECFRGENQYLKSSEFQTHN-TQMGLEQGQNMNSRNPPYGQVLKSNAGKLQIACSNN      844
Mus         IKVEESFCVGNQYSKSSNFQTHNNTQGGLEQVQNINK-NFPYSKILTPNSSNLQILPSND      789
Rattus      IKAEESFCGENQYSKPSNFQTHS-AQGGLEQVQNVNR-SFPYAKIETPNSSKVQTLCSNY      785
             *.*: :    *    . ::*    * .    .:    .  : .  . .  .:

Danio       VEALTERKL---------------QQNTEC-SQNISTGQQIYQQPLQHRIYNHSEFSQSL      674
Homo        THLVSENKEQTTHPELFAGNKTQNLHHMQYFPNNV-----IPKQDLLHRCFQEQEQKSQQ      892
Pan         THLVSENKEQTTHPELFAGNKTQNLHHMQYFPNNV-----IPKQDLLHRCFQEQEQKSQQ      913
Bos         IQLVPDNKEQTVNSELFAGNKTPNVHHMQYFPNNV-----TPKQDILHRCFQQQEQKPQQ      889
Mustela     THLVPENKEQTINPELFAGNKTQNLHHRQYFPNNV-----TPKQDVLHRCFQEQEQKPQQ      892
Canis       THLVPENKEQIINSELFAGNKTQNLHHMQYFPNNV-----TPKQDVLHRCFQEQEQKPQQ      899
Mus         THPACEREQ-AL---HPVGSKTSNLQNMQYFPNNV-----TPNQDV-HRCFQEQAQKPQQ      839
Rattus      THPASESREQNLHLDLFIGNRTPNQQNTQCFPNNV-----TPSQDVAHRCFQEQEQKPHQ      840
             .   : .                  :. :   :*:       .* : ** ::..  .

Danio       IS--QSQHSQASDGLQIPVRTLTKAEF------HESFAQIQRESLSSNGAQMDVQRHAAL      726
Homo        ASVLQGYKNRNQDMSGQQAAQLAQQRYLIHNHANVFPVPDQGGSHTQTPPQKDTQKHAAL      952
Pan         ASVLQGYKNRNQDMSGQQAAQLAQQRYLIHNHANVFPVPDQGGSHTQTPPQKDIQKHAAL      973
Bos         ASVLQGYKSRNQDVSGQQAAQLAQQRYLMQNQANAFPVPDQTGTHVQTPPQKDIQKHAAL      949
Mustela     ASVLQGYRSRNQDMPGQQAAQLPQQRYLMQNQANAFPVSDQGGSHIQTPPQKDLQKHAAL      952
Canis       ASVLQGYRSRNQDMSGQQATQLPQQRYLMQNQPNTFPVSDQGGSHSQTPPQKDIQKHAAL      959
```

```
Mus         ASSLQGLKDRSQGESPAPPAEAAQQRYLVHNEAKALPVPEQGGSQTQTPPQKDTQKHAAL      899
Rattus      ASSLQGLKGRSRDEFSVPPAEATQARYLVHIQAKALPVPKQGGSHTQISPQKDTQKHAAL      900
             *  *. : :             : .:     .   .  *  :  .   * * *:****

Danio       RMHLLQRQERQSPNN------IRQMMPAIKSENCSGFEISAQPMEVQRQHEDHTKVLKAV      780
Homo        RWHLLQKQEQQQTQQPQTESCHSQMHRPIKVEPGCKPHACMHT--A----PPENKTWKKV      1006
Pan         RWHLLQKQEQQQTQQPQTESCHSQMHRPIKVEPGCKPHACMHT--A----PPENKTWKKV      1027
Bos         RWRLLQKQEQQQTQ-LQAESCHSQMQRPIKVEPGSKPHACMRPLSA----QPENKMWKKI      1004
Mustela     RWHLLQKQEQQQTQHPHTESCHSQMHRPIKVEPGSKPHTCMRSTSA----QPESKMWKKI      1008
Canis       RWHLLQKQEQQQTQHPQTESCHNQMHRPIKVEPGSKPHTCMRLMSA----QPENKTWKKT      1015
Mus         RWLLLQKQEQQQTQQ--SQPGHNQMLRPIKTEPVSKPSSYRYPLSP----PQENM--SSR      951
Rattus      RRLLLQKQEQQQIQQSQSESGHHQMPRPIKIEPGSKLSSCKHPTSG----PQENM--PKR      954
            *  ***:**:*. :          **   ** *  .                 ..

Danio       VKQEHPPSSCDENQQRSILATMEQQLKQYQLSPVFEKKSLVIKSPNKVKVEMAGAVTVIS      840
Homo        TKQENPPASCDNVQQKSIIETMEQHLKQFHAKSLFDHKALTLKSQKQVKVEMSGPVTVLT      1066
Pan         TKQENPPASCDNVQQKSIIETMEQHLKQFQAKSLFDHKALTLKSQKQVKVEMSGPVTVLT      1087
Bos         PKQEIPPPSCDSVQQRSILETMEQHLKQFQVKSLFDHKALNLKSQKQVKVEMSGPVTVLT      1064
Mustela     TKQEIPPPSCDNVQQRSILETMEQHLKQFQVKSLFDHRALTLKSQKQVKVEMSGPVTVLT      1068
Canis       TKQEIPPPSCDNVQQRSILETMEQHLKQFQVKSLFDHKALTLKSQKQVKVEMSGPVTVLT      1075
Mus         IKQEISSPSRDNGQPKSIIETMEQHLKQFQLKSLCDYKALTLKSQKHVKVP--TDI----      1005
Rattus      IKQESPLLSCDGGRTESIIETMERHLKQFQLKALCDYKILTLNSQRQGDVETATNA----      1010
             ***    * *  : .**: ***::***:: . : : : * ::* .: .*

Danio       TNAEGNAKEQ---CKQ------QHFTPPK-KIEPGLQSFLESPMKLLNTPIKNLMDTPLK      890
Homo        RQTTAAELDSHTPALEQQT-TSSEKTPTKRTAASVLNNFIESPSKLLDTPIKNLLDTPVK      1125
Pan         RQTTAAELDSHTPALEQQT-TSSEKTPTKRTAASVLNNFIESPSKLLDTPIKNLLDTPVK      1146
Bos         RQTTAAELESHTPVLEQQATPSSEKTPTKRTAGSVLNNFLESPSKLLDTPIKNLLDTPVK      1124
Mustela     RQTTAAELDSHTQALEQQATPSSEKTPTKRTAGSVLNNFLESPSKLLDTPIKNLLDTPVK      1128
Canis       RQTTAAELDSHTQALEQQATPSSEKTPTKRTAGSVLNNFLESPSKLLDTPIKNLLDTPVK      1135
Mus         ---QAAESENHARAAEPQA-TKS-------TDCSVLDDVSESDTPGE--------------      1041
Rattus      ---RSAGSENHTPAAEPQA-SLSEKTPTKRTADSVLTDVSESPSDSD--------------      1053
                .   :.     :      .       .    * .. **

Danio       TQYEIPSCHCVEQISERDEGPYYTHLGSASNVAGIRKIMEERSGMTGSAIRIEKVLYTGK      950
Homo        TQYDFPSCRCVEQIIEKDEGPFYTHLGAGPNVAAIREIMEERFGQKGKAIRIERVIYTGK      1185
Pan         TQYDFPSCRCVEQIIEKDEGPFYTHLGAGPNVAAIREIMEERFGQKGKAIRIERVIYTGK      1206
Bos         TQYDFPSCRCVEQIIEKDEGPFYTHLGAGPNVAAIREIMEERFGQKGKAIRIERVIYTGK      1184
Mustela     TQYDFPSCRCVEQIIEKDEGPFYTHLGAGPNVAAIREIMEERFGQKGKAIRIERVIYTGK      1188
Canis       TQYDFPSCRCVEQIIEKDEGPFYTHLGAGPNVAAIREIMEERFGQKGKAIRIERVIYTGK      1195
Mus         -QSQNGKCEGC-N-PDKDEAPYYTHLGAGPDVAAIRTLMEERYGEKGKAIRIEKVIYTGK      1098
Rattus      TPVEDISCEAC-KNAEKDEGPYYTHLGAGPNVAAIRTIMEERFGEKGKAIRIERVIYTGK      1112
               :  .*.   :  ::**.*:*****:. :**.** :**** * .*.*****:*:****

Danio       EGKSGLGCPIAKWVIRRANEDEKILVLVRERAGHSCETSCVVVVILIWEGIPTSLADRLY      1010
Homo        EGKSSQGCPIAKWVVRRSSSEEKLLCLVRERAGHTCEAAVIVILILVWEGIPLSLADKLY      1245
Pan         EGKSSQGCPIAKWVVRRSSSEEKLLCLVRERAGHTCEAAVIVILILVWEGIPLSLADKLY      1266
Bos         EGKSSQGCPIAKWVVRRSCSEEKLLCLVRERAGHTCEAAVIVILILVWEGIPVSLADKLY      1244
Mustela     EGKSSQGCPIAKWVVRRSCSEEKLLCLVRERAGHTCEAAVIVILILVWEGIPLSLADKLY      1248
Canis       EGKSSQGCPIAKWVVRRSCSEEKLLCLVRERAGHTCEAAVIVILILVWEGIPLSLADKLY      1255
Mus         EGKSSQGCPIAKWVYRRSSEEEKLLCLVRVRPNHTCETAVMVIAIMLWDGIPKLLASELY      1158
Rattus      EGKSSQGCPIAKWVYRRSSTEEKLLCLVRVRAKHTCDTAVIVIVILLWDGIPKPLASELY      1172
            ****. ******** **:  :**:* *** *  *:*::: ::: *::*:***  **..**

Danio       MELSDTLTKHGALTNRRCALNEERTCACQGLEADACGASFSFGCSWSMYYNGCKFARSKV      1070
Homo        SELTETLRKYGTLTNRRCALNEERTCACQGLDPETCGASFSFGCSWSMYYNGCKFARSKI      1305
Pan         SELTETLRKYGTLTSRRCALNEERTCACQGLDPETCGASFSFGCSWSMYYNGCKFARSKI      1326
Bos         SELTETLRKYGMLTNRRCALNEERTCACQGLDPDTCGASFSFGCSWSMYYNGCKFARSKI      1304
Mustela     SELTETLRKYGTLTNRRCALNEERTCACQGLDPETCGASFSFGCSWSMYYNGCKFARSKI      1308
Canis       SELTETLRKYGTLTNRRCALNEERTCACQGLDPETCGASFSFGCSWSMYYNGCKFARSKI      1315
Mus         SELTDILGKCGICTNRRCSQNETRNCCCQGENPETCGASFSFGCSWSMYYNGCKFARSKK      1218
Rattus      SELTEILSNRGICTNRRCAQNENRNCCCQGENPETCGASFSYGCSWSMYYNGCKFARSKN      1232
             **:: * : *  *.***: ** *.*.*** : ::******:*****************

Danio       PRKFKLLADDPKEEEKIEQNLQGLATYIAPVYKKMAPDAYSNQVEHENRAPDCRLGLKEG      1130
```

```
Homo        PRKFKLLGDDPKEEEKLESHLQNLSTLMAPTYKKLAPDAYNNQIEYEHRAPECRLGLKEG   1365
Pan         PRKFKLLGDDPKEEEKLESHLQNLSTLMAPTYKKLAPDAYNNQIEYEHRAPECRLGLKEG   1386
Bos         PRKFKLLGDDPKEEEKLESHLQNLSTLMAPTYKKLAPDAYNNQIEYEHRAPECRLGLKEG   1364
Mustela     PRKFKLLGDDPKEEEKLESHLQNLSTLMAPTYKKLAPDAYNNQIEYEHRAPECRLGLKEG   1368
Canis       PRKFKLLGDDPKEEEKLESHLQNLSTLMAPTYKKLAPDAYNNQIEYEHRAPECRLGLKEG   1375
Mus         PRKFRLHGAEPKEEERLGSHLQNLATVIAPIYKKLAPDAYNNQVEFEHQAPDCCLGLKEG   1278
Rattus      PRKFRLHGDEPKEEEKLGSHLQNLATVIAPIYKKLAPDAYRNQVEFEHRAIECRLGLKEG   1292
            ****:* . :*****:: ..** *:* :** ***:***** **:*.*.:* :* ******

Danio       RPFSGVTACLDFCAHAHRDLHNMQGGSTVVCTLTREDNREIGKIPEDEQLHVLPLYKPSS   1190
Homo        RPFSGVTACLDFCAHAHRDLHNMQNGSTLVCTLTREDNREFGGKPEDEQLHVLPLYKVSD   1425
Pan         RPFSGVTACLDFCAHAHRDLHNMQNGSTLVCTLTREDNREFGGKPEDEQLHVLPLYKVSD   1446
Bos         RPFSGVTACLDFCAHAHRDLQNMQNGSTLVCTLTREDNREIGGKPEDEQLHVLPLYKVSD   1424
Mustela     RPFSGVTACLDFCAHAHRDLHNMQNGSTLVCTLTREDNREIGGKPEDEQLHVLPLYKVSD   1428
Canis       RPFSGVTACLDFCAHAHRDLHNMQNGSTLVCTLTREDNREIGGKPEDEQLHVLPLYKVSD   1435
Mus         RPFSGVTACLDFSAHSHRDQQNMPNGSTVVVTLNREDNREVGAKPEDEQFHVLPMYIIAP   1338
Rattus      RPFSGVTACLDFSAHAHRDQQNMANGSTVVVTLTREDNREVGGQPEDEQLHVLPLYTIAT   1352
            ***********.**:*** :**  ***:* **.******.*  *****:****:*  :

Danio       TDEFGSAEAQLEKTKTGAIQVLSSFRR-QVRMLPEPAKSCRQRKLDAKRANKPNNNTPNS   1249
Homo        VDEFGSVEAQEEKKRSGAIQVLSSFRR-KVRMLAEPVKTCRQRKLEAK---KAAAEKLSS   1481
Pan         VDEFGSVEAQEEKKRSGAIQVLSSFRR-KVRMLAEPVKTCRQRKLEAK---KAAAEKLSS   1502
Bos         VDEFGSVEAQEEKKRNGAIQVLSSFRR-KVRMLAEPVKTGRQRKLEAK---KAAAEKLSS   1480
Mustela     VDEFGSVEAQEKKKQNGAIQVLSSFRR-KVRMLAEPVKTCRQRKLEAK---KAAAEKLSS   1484
Canis       VDEFGSVEAQEKKKQNGAIQVLSSFRR-KVRMLAEPVKTCRQRKLEAK---KAAAEKLSS   1491
Mus         EDEFGSTEGQEKKIRMGSIEVLQSFRRRRVIRIGELPKSCKKKAEPKKAKTKKAARKRSS   1398
Rattus      EDEFGSTEGQEEKILQGSIQVLHSFRRRRVRRLVDPPKNCRQKKL-----EAKKARKLSS   1407
             *****.*.* :*   *:*:** **** :*  : :  *. :::            .. .*

Danio       KTDNT----------------QQAKQKQTAYENPTVTGRGNMRTNLDSGHLPQAHAGHQP   1293
Homo        LENSSNKNEKEKSAPSRTKQTENASQAKQLAELLRLSGPVMQQSQQPQP---LQKQPPQP   1538
Pan         LENSSNKNEKEKSAPSRTKQTENASQAKQLAELLRLSGPVMQQSQQPQP---LQKQPPQP   1559
Bos         LETSTNKSEKEKSAASRLKQTENASQAKQLAELLQPPGPVMQLPQQPQSQPQLQKQLPQP   1540
Mustela     LENSANKNEKEKSASSRTKQTENASQAKQLAELLRLSGPVMQQPQQPPP-PPLQKQPPQP   1543
Canis       LENNTNKNEKEKSASSRTKQTENASQAKQLAELLRLSGPVMQQSQQPPP-PPLQKQPPQP   1550
Mus         LENCSSRTEKGKSS-SHTKLMENASHMKQMTAQPQLSGPVIRQPPTLQRHLQQGQRPQQP   1457
Rattus      LENCSNKNEKEKSS-SRTRPMENASPMIQMTAQLRLSGSVIQQPQSLQRHLQQVQRPPQP   1466
                :               ::*.           *                 :   **

Danio       QQQFPQ------------QTHPNP---------SY----ASPP--------FTRFPNASK   1320
Homo        QQQ--------QRPQQQQPHH-----PQTESVNSYSASGSTNPYMRRPNP-VSPYPNSSH   1584
Pan         QQQ--------QRPQQQQPHH-----PQTESVNSYSASGSTNPYMRRPNP-VSPYPNSSH   1605
Bos         QKQPPQLQQPQQQPPQQQPHHPLTNNSQSESVSSYSSSGSTNLYMRRPNA-MSPYPGSSH   1599
Mustela     QKQPPQH---------PPQHHPVTNNPQSEPVNSYSSSGSANLYIRRPNPVIQAYPSSSH   1594
Canis       QKQPPQQQQQQQQQ--QQQQHHPMTNNSQSESVNSYSSSGSANLYMRRPNPVIQAYPSSSH   1608
Mus         Q--PPQPQPQTTPQPQPQPQHIMPGNSQSV---GSHCSGSTSVYTRQPTP-HSPYPSSAH   1511
Rattus      Q--PP------------QPQHILPSNSQPV---GSHSSGSTNVYMRQPAP-LSPYPSSSH   1508
            *                  *            .     ::             :* :::

Danio       PSTPHPQTPSVDPYASPLHAPNSYINASNAPSPYSRSLAPSPLYNGYQCNGGIPMDNYHP   1380
Homo        TSDIYGSTSPMNFYSTSSQAAGSYLNSSNPMNPYPGLLNQNTQYPSYQCNGNLSVDNCSP   1644
Pan         TSDIYGSTSPMNFYSTSSQAAGSYLNSSNPMNPYSGLLNQNTQYPSYQCNGNLSVDNCSP   1665
Bos         TSDIYGGANPVNLYSTSSQAAGSYLNSSNPMSPYPGLLNQSNQYPSYQCNGNISVDNCSQ   1659
Mustela     TSDVYGSASPMNLYSTPSQATGSYLNSSNPMNPYTGLLNQNNQYPSYQCNGNVSVDNCSP   1654
Canis       TSDIYGGASPMNLYSTSSQATGSYLNSSNPMNPYTGLLNQNNQYPAYQCNGNVSVDNCSP   1668
Mus         TSDIYGDTNHVNFYPTSSHASGSYLNPSNYMNPYLGLLNQNNQYAPFPYNGSVPVDNGSP   1571
Rattus      AADMYGDANHVNFHPTSSHAAGLYLSPSNSMNPYPGLFNQNNQYAASQCNGSMSADSGPP   1568
             :  :  :  :: : :  :*   *:. **  .**   :  .  *     ** :  *.

Danio       YHSSNLKHPDMFHPQR----NP-----------LYSEQQYNAPQHYGVNYPPH---YGEA   1422
Homo        YLGSYSPQSQPMDLYRYPSQDPLSKLSLPPIHTLYQ-PRFGNSQSFTSKYLGYGNQNMQG   1703
Pan         YLGSYSPQSQPMDLYRYPSQDPLSKLSLPPIHTLYQ-PRFGNSQSFTSKYLGYGNQNMQG   1724
Bos         YLGSYSPQSQPMDLYRYPNQDPLSKLNLPPIHTLYQ-PRFGNSQSFTSKYLGYGNQNMQG   1718
Mustela     YLGSYSPQPQPMDLYRYPNQDPLSKLNLPPIHTLYQ-PRFGNSQNFSSKYLGYGNQTMQG   1713
Canis       YLGSYSPQSQPMDLYRYPNQDPLSKLNLPPIHTLYQ-PRFGNSQSFSSKYLGYGNQNIQG   1727
```

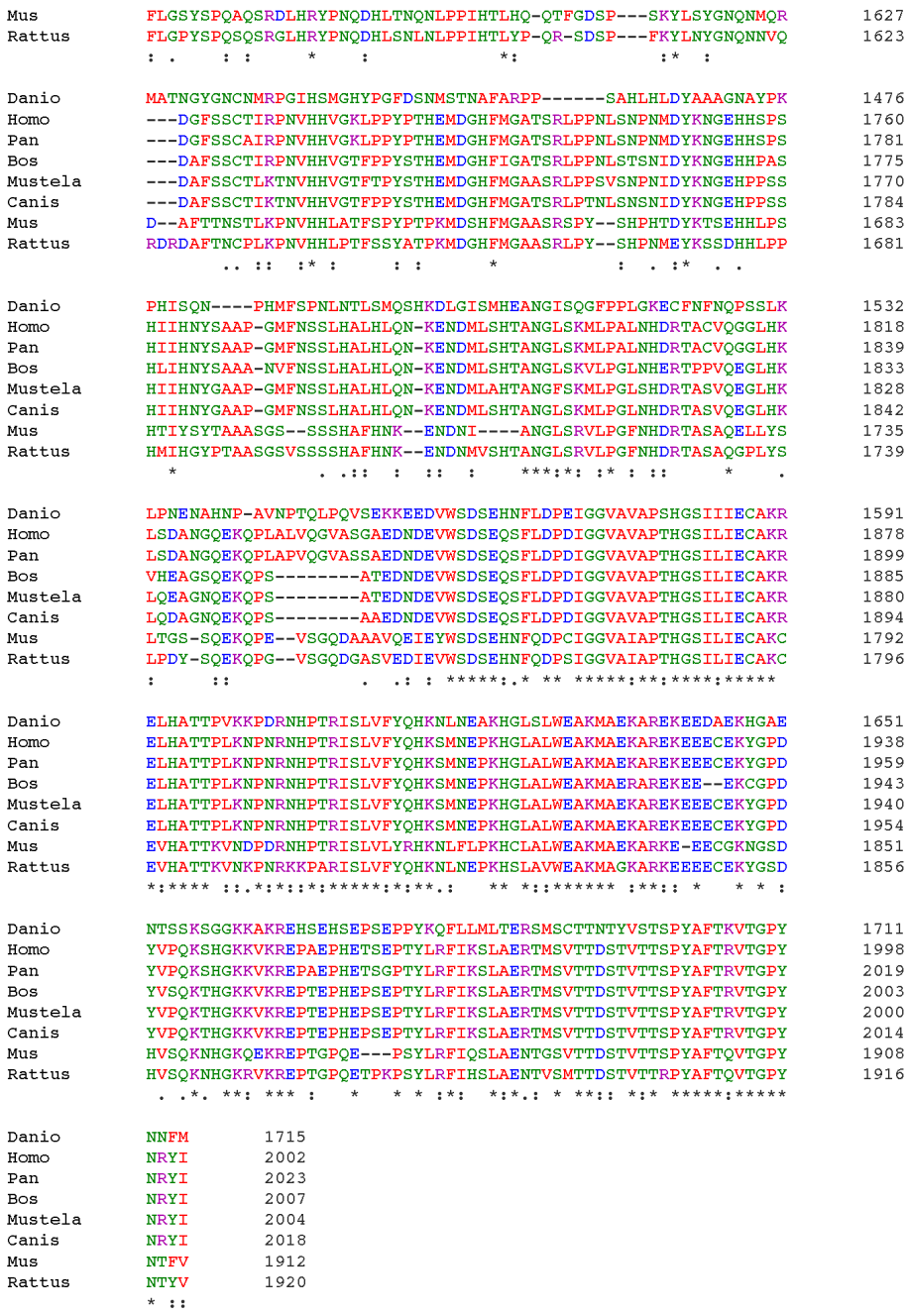
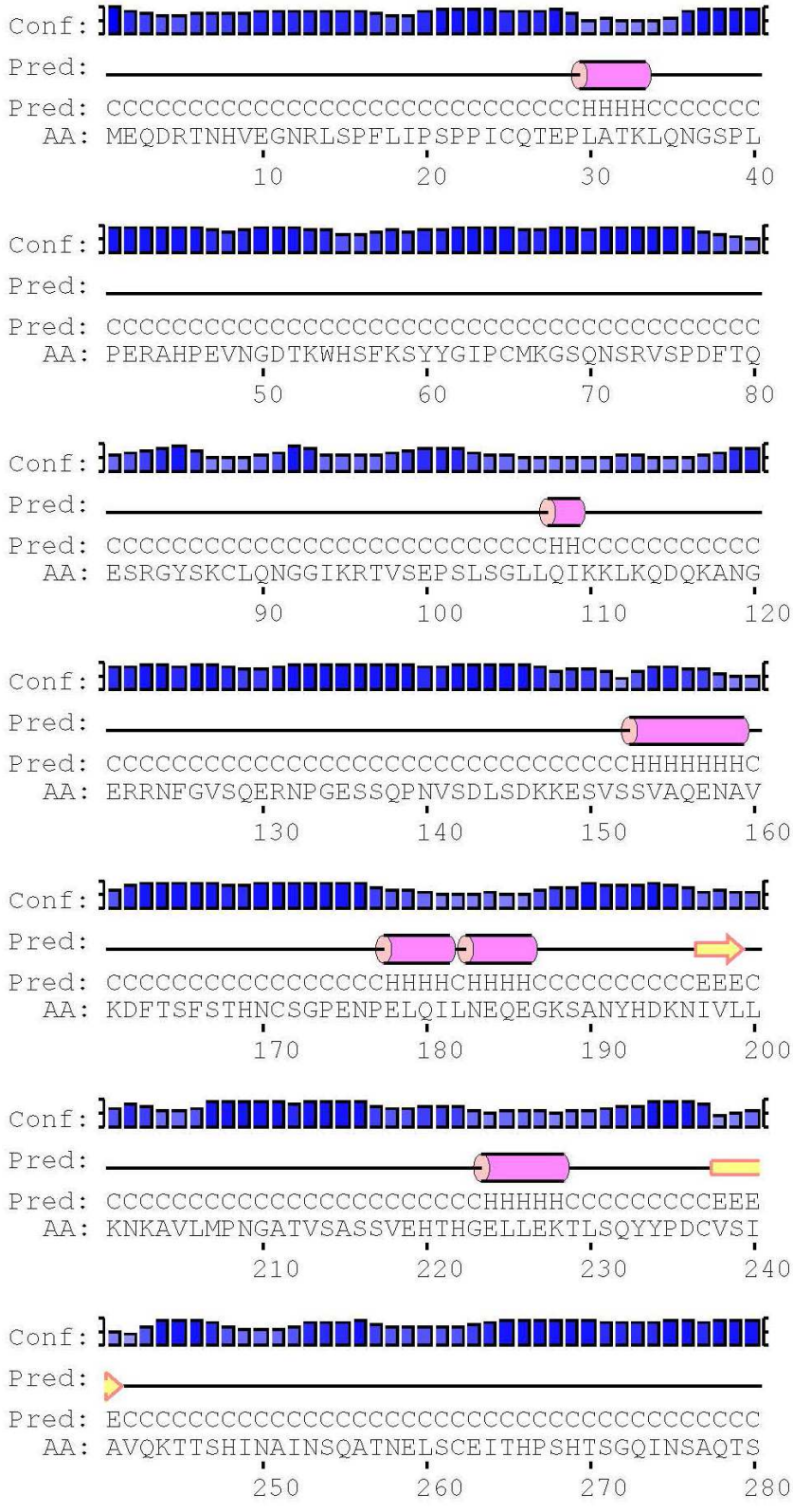
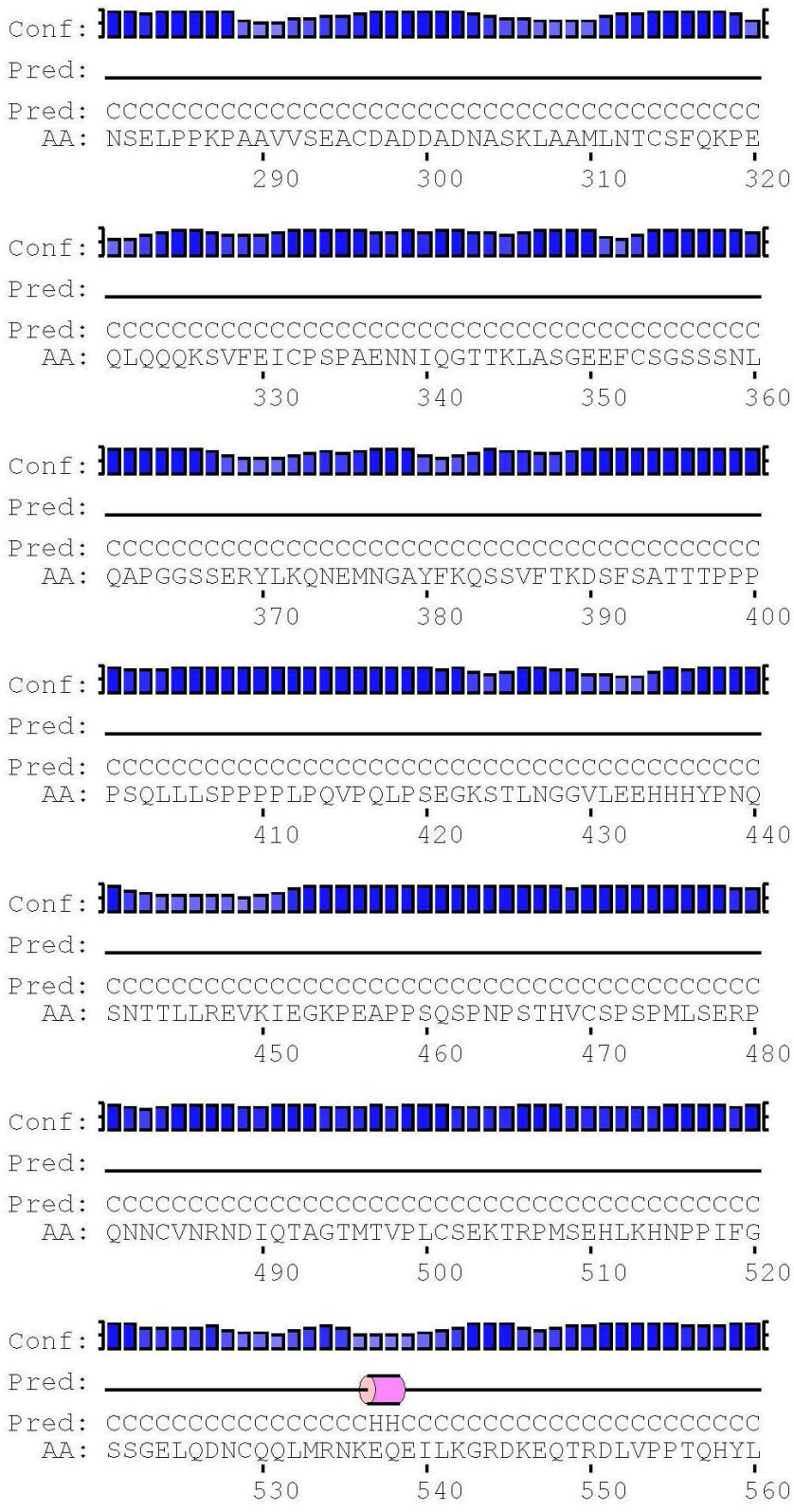
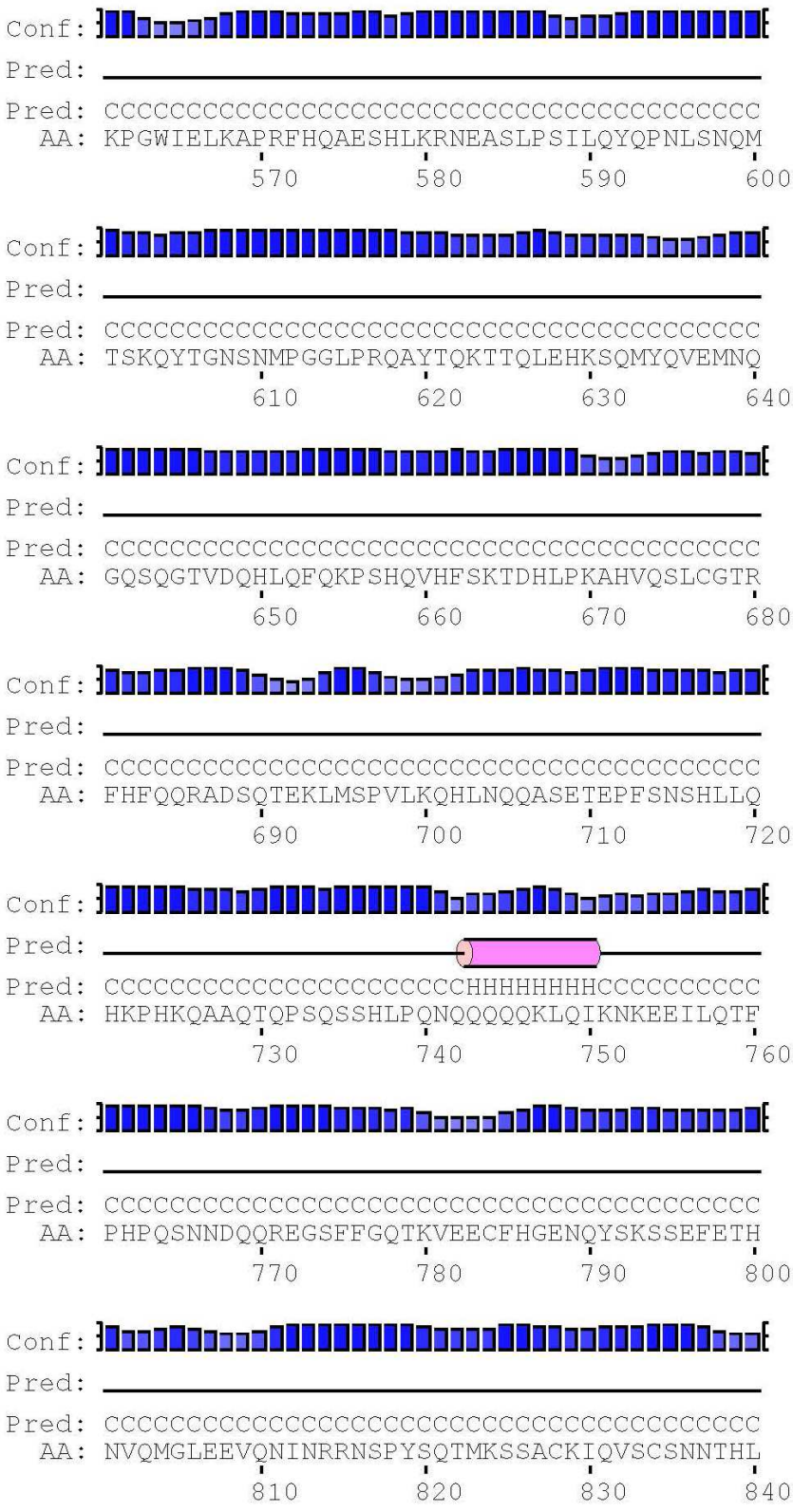
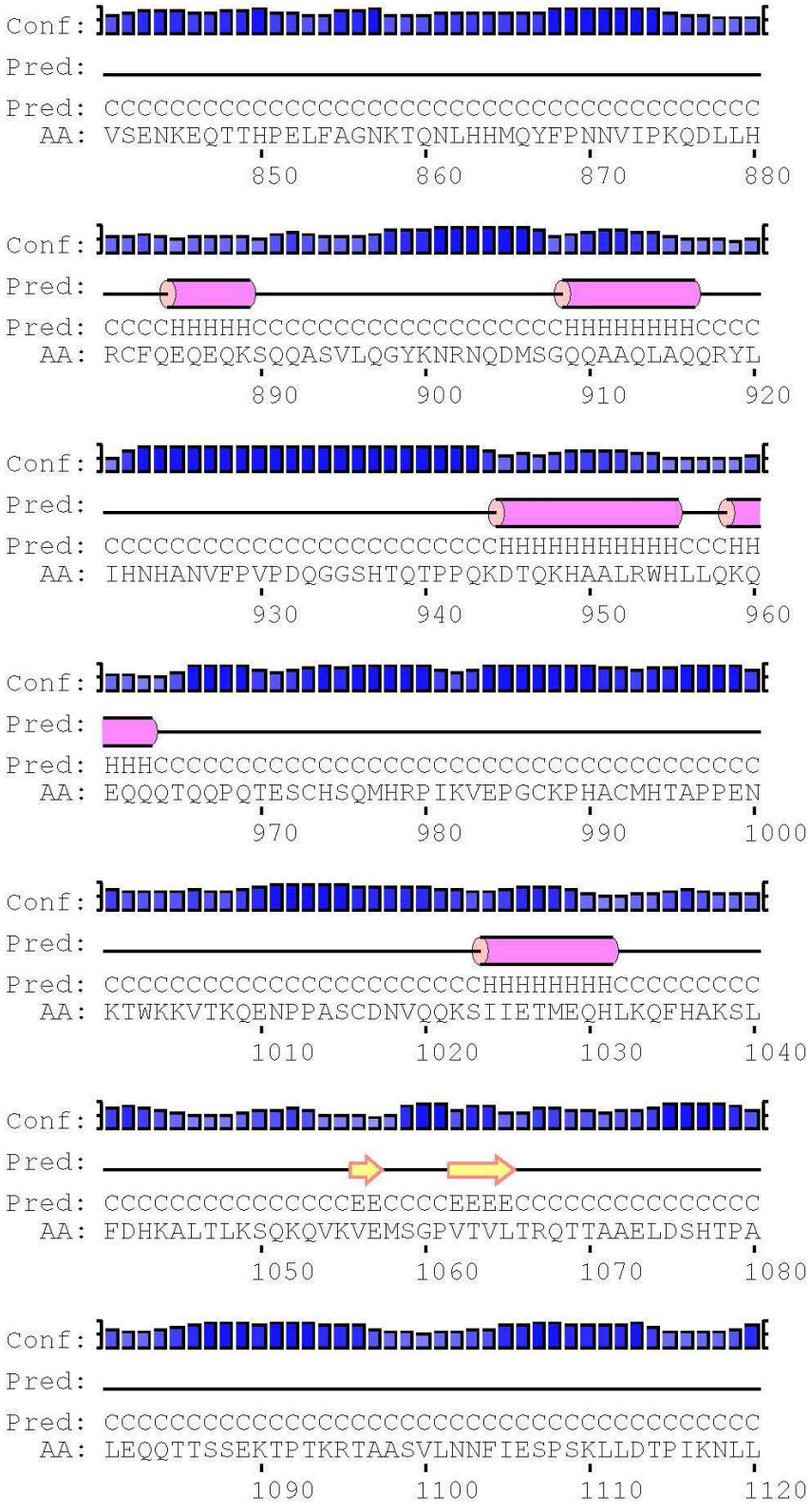


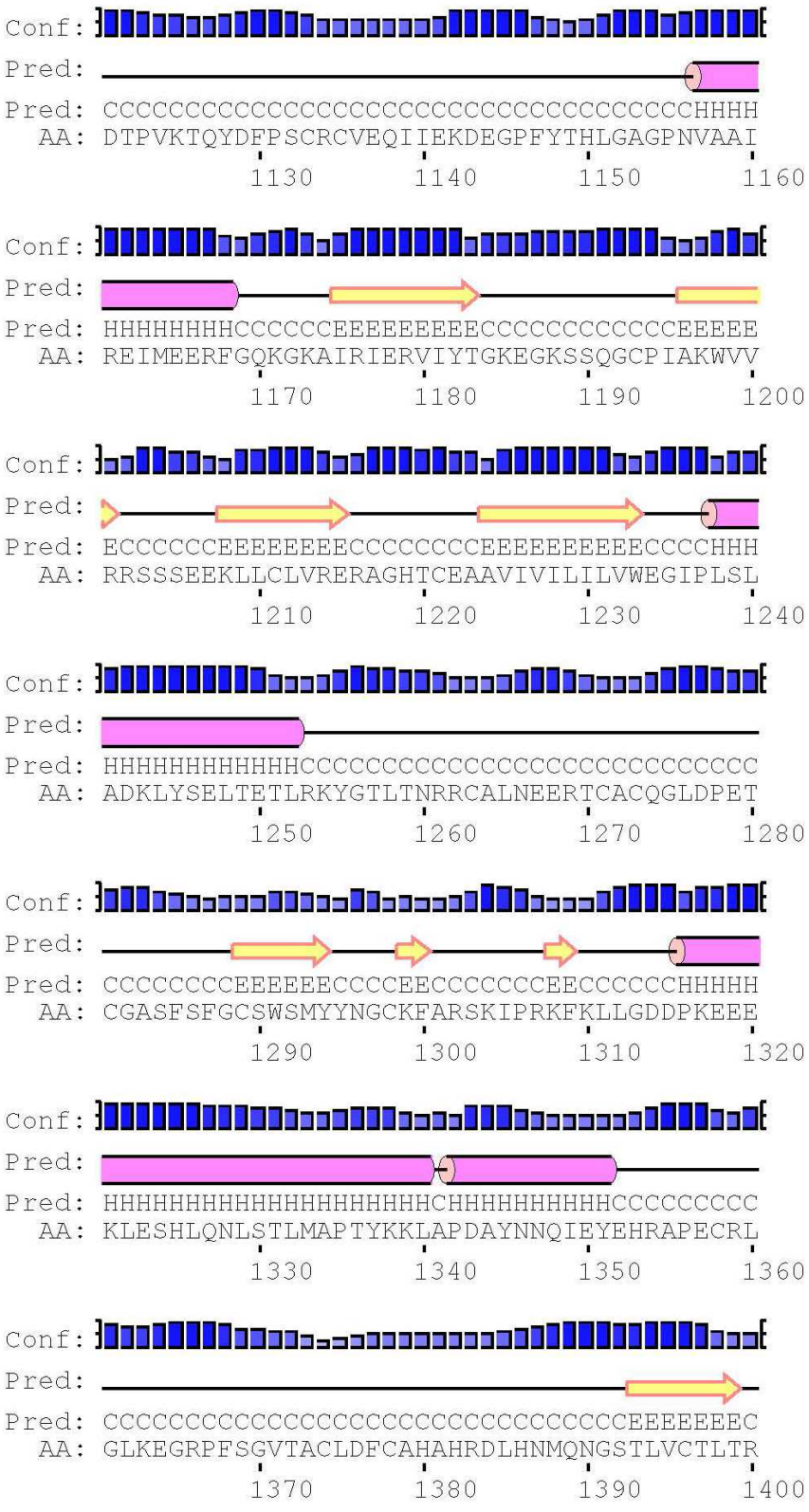
Figure 6: TET2 sequence alignment among: *Danio rerio, Pan troglodytes, Homo sapiens, Mus musculus, Bos taurus, Canis familiaris, Rattus norvegicus, Mustela putorius furo.*
∗ (asterisk) Indicates perfect conservation, **:** (colon) indicates conservation between groups of strongly similar properties; **.** (period) indicates conservation between groups of weakly similar properties.
Color coding: red indicates small hydrophobic aminoacids (AVFPMILW including Y), blue acidic aminoacids (DE), magenta basic aminoacids (RK, escluding H), green other aminoacids (STYHCNGQ including G) , grey is for unusual aminoacids.

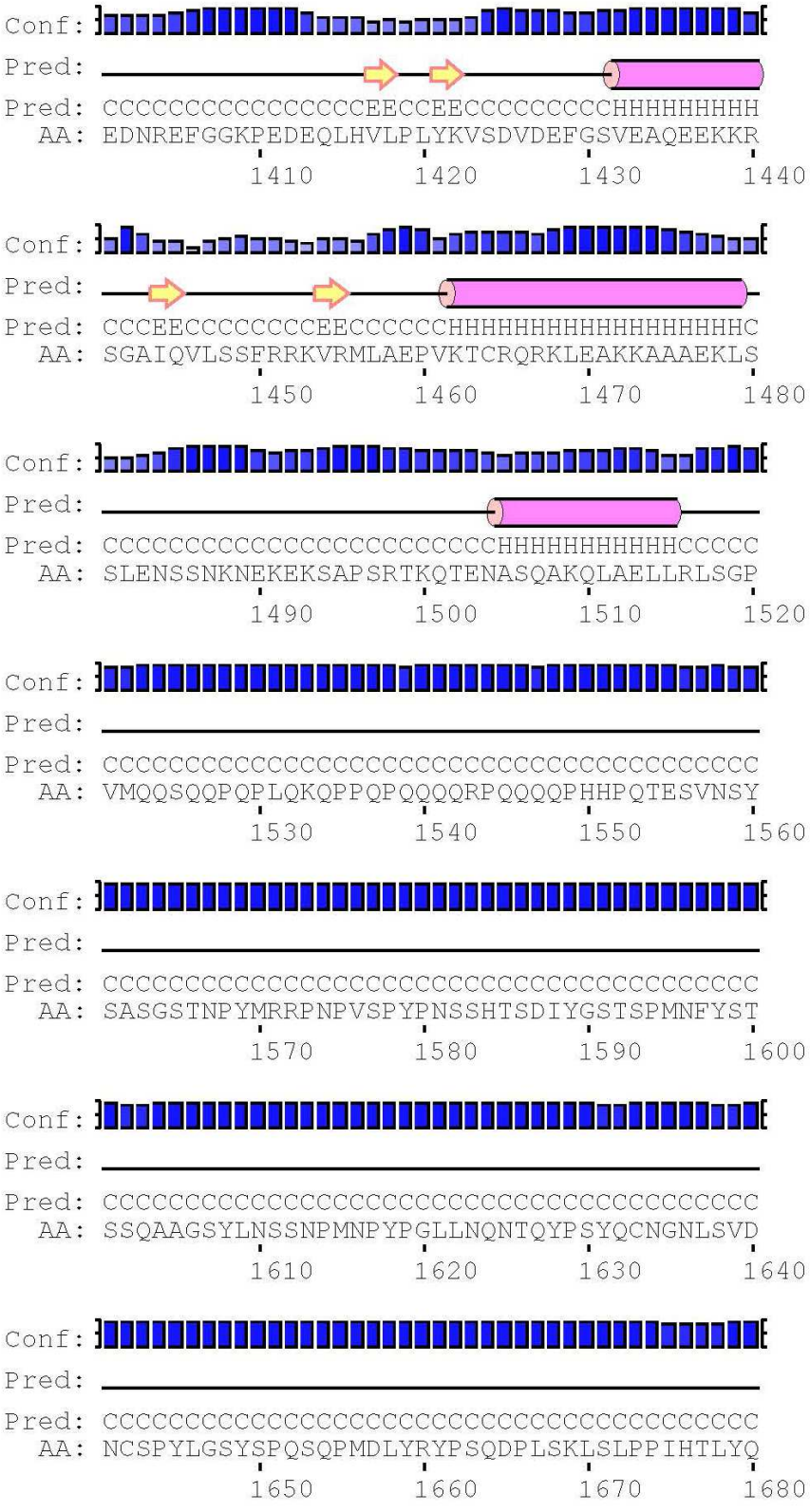

Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHCCCCCCC
AA: MEQDRTNHVEGNRLSPFLIPSPPICQTEPLATKLQNGSPL
10 20 30 40
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
AA: PERAHPEVNGDTKWHSFKSYYGIPCMKGSQNSRVSPDFTQ
50 60 70 80
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCHHCCCCCCCCCCC
AA: ESRGYSKCLQNGGIKRTVSEPSLSGLLQIKKLKQDQKANG
90 100 110 120
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHC
AA: ERRNFGVSQERNPGESSQPNVSDLSDKKESVSSVAQENAV
130 140 150 160
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCHHHHCHHHHCCCCCCCCCCEEEC
AA: KDFTSFSTHNCSGPENPELQILNEQEGKSANYHDKNIVLL
170 180 190 200
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCHHHHHCCCCCCCCCEEE
AA: KNKAVLMPNGATVSASSVEHTHGELLEKTLSQYYPDCVSI
210 220 230 240
Conf:
Pred:
Pred: ECCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
AA: AVQKTTSHINAINSQATNELSCEITHPSHTSGQINSAQTS
250 260 270 280

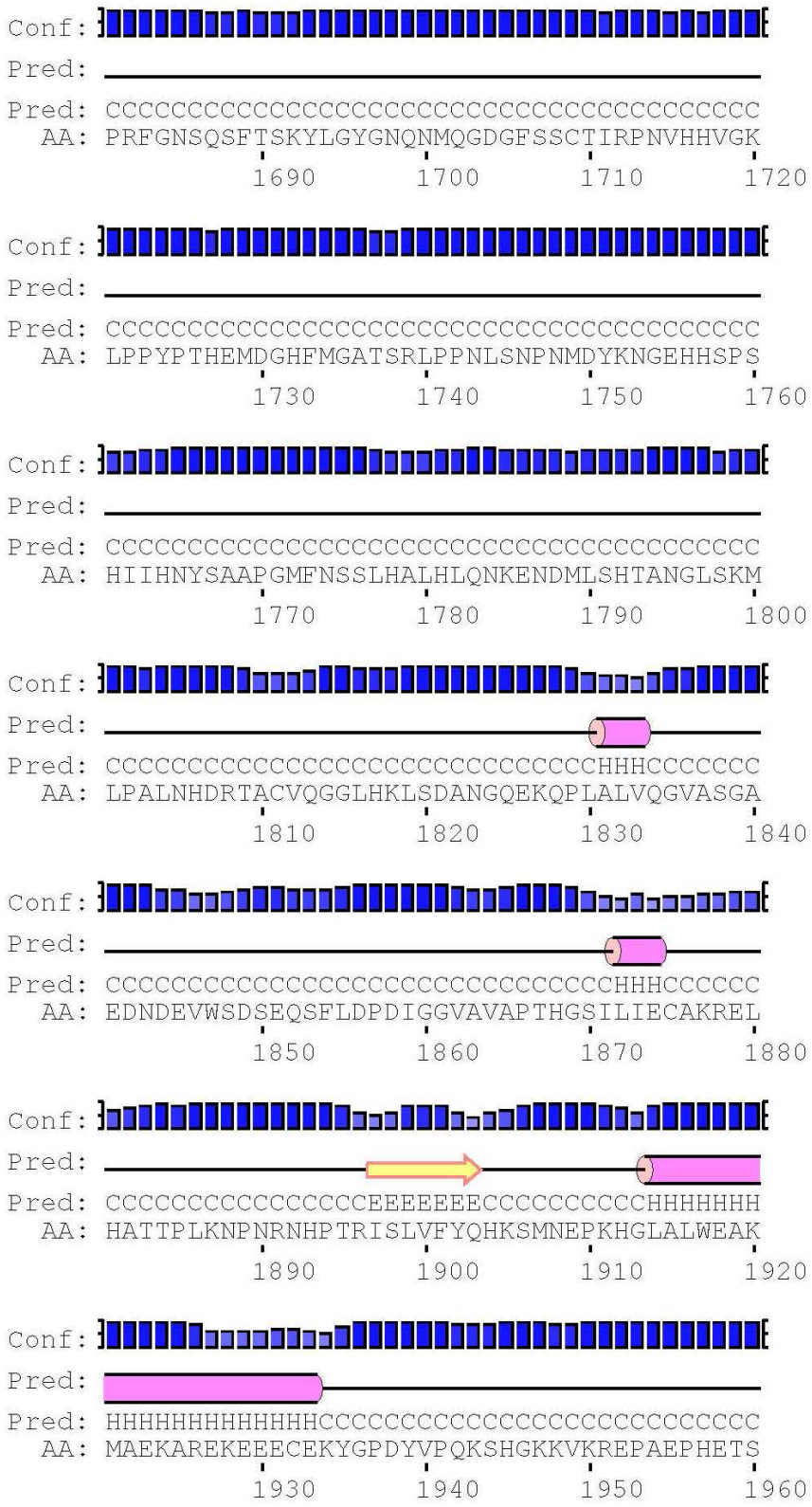

Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
AA: NSELPPKPAAVVSEACDADDADNASKLAAMLNTCSFQKPE
290 300 310 320
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
AA: QLQQQKSVFEICPSPAENNIQGTTKLASGEEFCSGSSSNL
330 340 350 360
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
AA: QAPGGSSERYLKQNEMNGAYFKQSSVFTKDSFSATTTPPP
370 380 390 400
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
AA: PSQLLLSPPPPLPQVPQLPSEGKSTLNGGVLEEHHHYPNQ
410 420 430 440
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
AA: SNTTLLREVKIEGKPEAPPSQSPNPSTHVCSPSPMLSERP
450 460 470 480
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
AA: QNNCVNRNDIQTAGTMTVPLCSEKTRPMSEHLKHNPPIFG
490 500 510 520
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCHHCCCCCCCCCCCCCCCCCCCCCC
AA: SSGELQDNCQQLMRNKEQEILKGRDKEQTRDLVPPTQHYL
530 540 550 560


Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
AA: KPGWIELKAPRFHQAESHLKRNEASLPSILQYQPNLSNQM
570 580 590 600
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
AA: TSKQYTGNSNMPGGLPRQAYTQKTTQLEHKSQMYQVEMNQ
610 620 630 640
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
AA: GQSQGTVDQHLQFQKPSHQVHFSKTDHLPKAHVQSLCGTR
650 660 670 680
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
AA: FHFQQRADSQTEKLMSPVLKQHLNQQASETEPFSNSHLLQ
690 700 710 720
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHCCCCCCCCCC
AA: HKPHKQAAQTQPSQSSHLPQNQQQQQKLQIKNKEEILQTF
730 740 750 760
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
AA: PHPQSNNDQQREGSFFGQTKVEECFHGENQYSKSSEFETH
770 780 790 800
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
AA: NVQMGLEEVQNINRRNSPYSQTMKSSACKIQVSCSNNTHL
810 820 830 840


Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
AA: VSENKEQTTHPELFAGNKTQNLHHMQYFPNNVIPKQDLLH
850 860 870 880
Conf:
Pred:
Pred: CCCCHHHHHCCCCCCCCCCCCCCCCCCHHHHHHHHCCCC
AA: RCFQEQEQKSQQASVLQGYKNRNQDMSGQQAAQLAQQRYL
890 900 910 920
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHCCCHH
AA: IHNHANVFPVPDQGGSHTQTPPQKDTQKHAALRWHLLQKQ
930 940 950 960
Conf:
Pred:
Pred: HHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
AA: EQQQTQQPQTESCHSQMHRPIKVEPGCKPHACMHTAPPEN
970 980 990 1000
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHCCCCCCCCC
AA: KTWKKVTKQENPPASCDNVQQKSIIETMEQHLKQFHAKSL
1010 1020 1030 1040
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCEECCCCEEEECCCCCCCCCCCCCCC
AA: FDHKALTLKSQKQVKVEMSGPVTVLTRQTTAAELDSHTPA
1050 1060 1070 1080
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
AA: LEQQTTSSEKTPTKRTAASVLNNFIESPSKLLDTPIKNLL
1090 1100 1110 1120


Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHH
AA: DTPVKTQYDFPSCRCVEQIIEKDEGPFYTHLGAGPNVAAI
1130 1140 1150 1160
Conf:
Pred:
Pred: HHHHHHHHCCCCCCEEEEEEEEECCCCCCCCCCCCCEEEEE
AA: REIMEERFGQKGKAIRIERVIYTGKEGKSSQGCPIAKWVV
1170 1180 1190 1200
Conf:
Pred:
Pred: ECCCCCCEEEEEEEECCCCCCCCEEEEEEEEEECCCCHHH
AA: RRSSSEEKLLCLVRERAGHTCEAAVIVILILVWEGIPLSL
1210 1220 1230 1240
Conf:
Pred:
Pred: HHHHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCC
AA: ADKLYSELTETLRKYGTLTNRRCALNEERTCACQGLDPET
1250 1260 1270 1280
Conf:
Pred:
Pred: CCCCCCCCEEEEEECCCCEECCCCCCCEECCCCCCCHHHHH
AA: CGASFSFGCSWSMYYNGCKFARSKIPRKFKLLGDDPKEEE
1290 1300 1310 1320
Conf:
Pred:
Pred: HHHHHHHHHHHHHHHHHHHHCHHHHHHHHHCCCCCCCCC
AA: KLESHLQNLSTLMAPTYKKLAPDAYNNQIEYEHRAPECRL
1330 1340 1350 1360
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCEEEEEEEC
AA: GLKEGRPFSGVTACLDFCAHAHRDLHNMQNGSTLVCTLTR
1370 1380 1390 1400


Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCEECCEECCCCCCCCCHHHHHHHHH
AA: EDNREFGGKPEDEQLHVLPLYKVSDVDEFGSVEAQEEKKR
1410 1420 1430 1440
Conf:
Pred:
Pred: CCCEECCCCCCCCEECCCCCCHHHHHHHHHHHHHHHHHC
AA: SGAIQVLSSFRRKVRMLAEPVKTCRQRKLEAKKAAAEKLS
1450 1460 1470 1480
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHCCCCC
AA: SLENSSNKNEKEKSAPSRTKQTENASQAKQLAELLRLSGP
1490 1500 1510 1520
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
AA: VMQQSQQPQPLQKQPPQPQQQQRPQQQQPHHPQTESVNSY
1530 1540 1550 1560
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
AA: SASGSTNPYMRRPNPVSPYPNSSHTSDIYGSTSPMNFYST
1570 1580 1590 1600
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
AA: SSQAAGSYLNSSNPMNPYPGLLNQNTQYPSYQCNGNLSVD
1610 1620 1630 1640
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
AA: NCSPYLGSYSPQSQPMDLYRYPSQDPLSKLSLPPIHTLYQ
1650 1660 1670 1680


Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
AA: PRFGNSQSFTSKYLGYGNQNMQGDGFSSCTIRPNVHHVGK
1690 1700 1710 1720
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
AA: LPPYPTHEMDGHFMGATSRLPPNLSNPNMDYKNGEHHSPS
1730 1740 1750 1760
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
AA: HIIHNYSAAPGMFNSSLHALHLQNKENDMLSHTANGLSKM
1770 1780 1790 1800
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHCCCCCCC
AA: LPALNHDRTACVQGGLHKLSDANGQEKQPLALVQGVASGA
1810 1820 1830 1840
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHCCCCCCC
AA: EDNDEVWSDSEQSFLDPDIGGVAVAPTHGSILIECAKREL
1850 1860 1870 1880
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCEEEEEEECCCCCCCCCCHHHHHHH
AA: HATTPLKNPNRNHPTRISLVFYQHKSMNEPKHGLALWEAK
1890 1900 1910 1920
Conf:
Pred:
Pred: HHHHHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCC
AA: MAEKAREKEEECEKYGPDYVPQKSHGKKVKREPAEPHETS
1930 1940 1950 1960

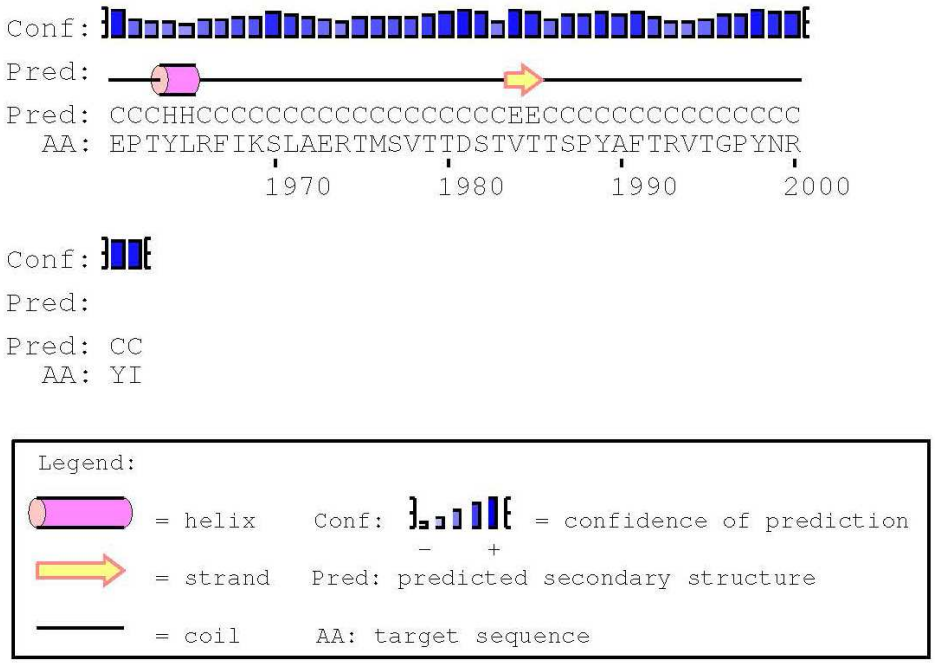


Figure 7: TET2 secondary structure prediction (Psipred). C indicates Coiled coil, H indicates Helix and E Strand.

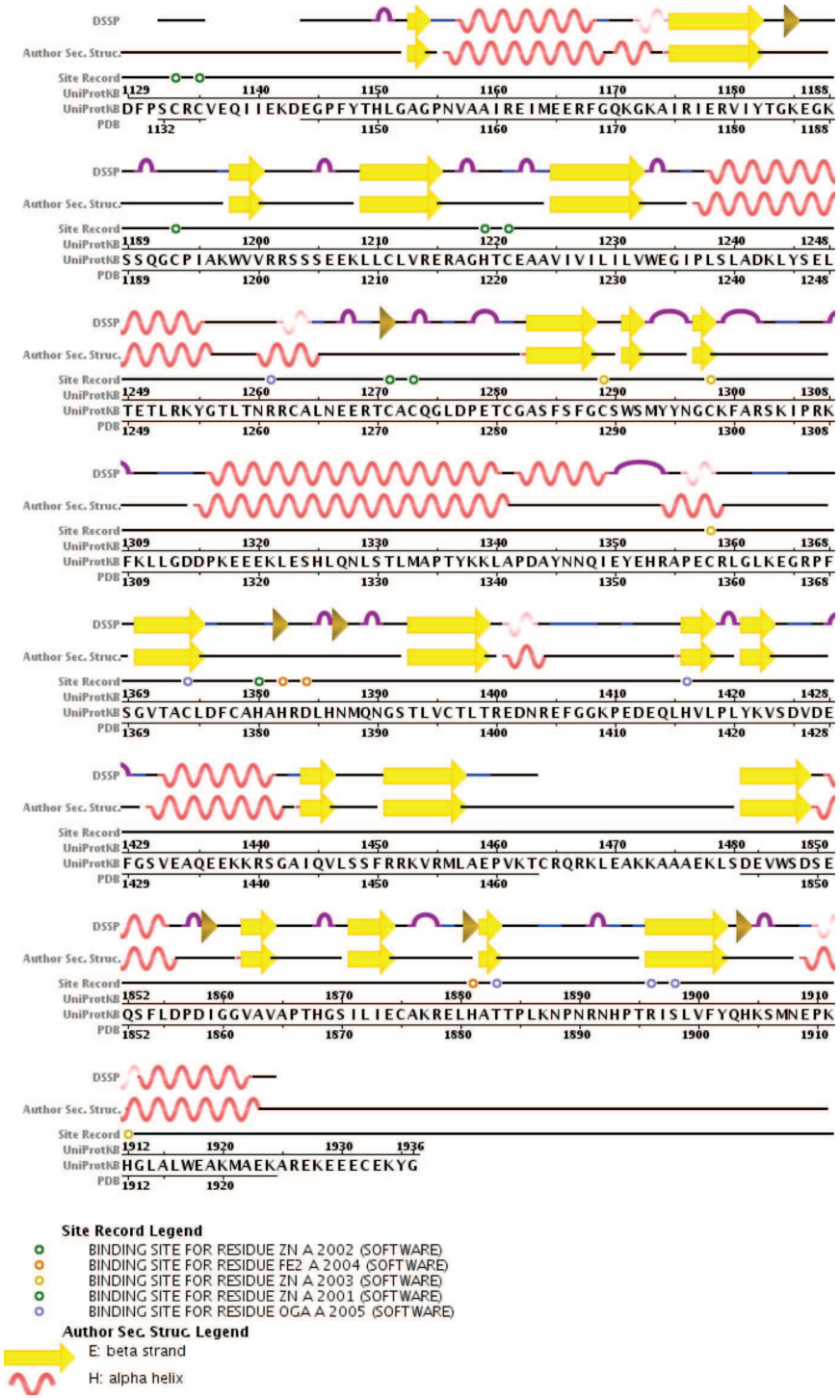


Figure 8: TET2 secondary structure according to the published crystal structure 4NM6 [Hu et al., 2013].

Bibliography

Ulla Aapola, Kazunori Shibuya, Hamish S Scott, Juha Ollila, Mauno Vihinen, Maarit Heino, Ai Shintani, Kazuhiko Kawasaki, Shinsei Minoshima, Kai Krohn, Stylianos E Antonarakis, Nobuyoshi Shimizu, Jun Kudoh, and Pärt Peterson. Isolation and Initial Characterization of a Novel Zinc Finger Gene, DNMT3L, on 21q22.3, Related to the Cytosine-5- Methyltransferase 3 Gene Family. *Genomics*, 65(3): 293–298, May 2000.

Omar Abdel-Wahab, Taghi Manshouri, Jay Patel, Kelly Harris, Jinjuan Yao, Cyrus Hedvat, Adriana Heguy, Carlos Bueso-Ramos, Hagop Kantarjian, Ross L Levine, and Srdan Verstovsek. Genetic analysis of transforming events that convert chronic myeloproliferative neoplasms to leukemias. *Cancer Research*, 70(2):447–452, January 2010.

Richard Abrams and Marian Bentley. Biosynthesis of nucleic acid purines. III. Guanosine 5-phosphate formation from xanthosine 5-phosphate and L-glutamine. *Archives of biochemistry and biophysics*, 79: 91–110, 1959.

K C Agarwal, R P Miech, and R E Parks, Jr. [64] Guanylate kinases from human erythrocytes, hog brain, and rat liver. *Methods in Enzymology*, 1978.

D Ahmed, P W Eide, I A Eilertsen, S A Danielsen, M Eknæs, M Hektoen, G E Lind, and R A Lothe. Epigenetic and genetic features of 24 colon cancer cell lines. *Oncogenesis*, 2(9):e71–, September 2013.

Nita Ahuja, Hariharan Easwaran, and Stephen B Baylin. Harnessing the potential of epigenetic therapy to target solid tumors. *Journal of Clinical Investigation*, 124(1):56–63, January 2014.

J Aimi, H Qiu, J Williams, H Zalkin, and J E Dixon. De novo purine nucleotide biosynthesis: cloning of human and avian cDNAs encoding the trifunctional glycinamide ribonucleotide synthetase-aminoimidazole ribonucleotide synthetase-glycinamide ribonucleotide transformylase by functional complementation in E. coli. *Nucleic Acids Research*, 18(22):6665–6672, November 1990.

S A Akman, D D Ross, H Rosen, C Salinger, P A Andrews, F E Chou, and N R Bachur. Growth inhibition by thymidine of leukemic HL-60 and normal human myeloid progenitor cells. *Cancer Research*, 41(6): 2141–2146, June 1981.

Ludmil B Alexandrov, Serena Nik-Zainal, David C Wedge, Samuel AJR Aparicio, Sam Behjati, Andrew V Biankin, Graham R Bignell, Niccolò Bolli, Åke Borg, and Anne-Lise Børresen-Dale. Signatures of mutational processes in human cancer. *Nature*, 2013.

Marjan Amiri, Francesca Conserva, Christakis Panayiotou, Anna Karlsson, and Nicola Solaroli. The human adenylate kinase 9 is a nucleoside mono- and diphosphate kinase. *The International Journal of Biochemistry & Cell Biology*, 45(5):925–931, May 2013.

F Antequera, M Tamame, J R Villanueva, and T Santos. DNA methylation in the fungi. *The Journal of biological chemistry*, 259(13):8033–8036, July 1984.

Francisco Antequera, Donald Macleod, and Adrian P Bird. Specific protection of methylated CpGs in mammalian nuclei. *Cell*, 58(3):509–517, January 1989.

W B Archey, M P Sweet, G C Alig, and B A Arrick. Methylation of CpGs as a determinant of transcriptional activation at alternative promoters for transforming growth factor-beta3. *Cancer Research*, 59(10):2292–2296, May 1999.

Elias S J Arnér and Staffan Eriksson. Mammalian deoxyribonucleoside kinases. *Pharmacology & Therapeutics*, 67(2):155–186, January 1995.

Charanjit Arora, Kehkooi Kee, Shohreh Maleki, and Scott Keeney. Antiviral protein Ski8 is a direct partner of Spo11 in meiotic DNA break formation, independent of its cytoplasmic role in RNA metabolism. *Molecular Cell*, 13(4):549–559, February 2004.

M Ast, A Gruber, S Schmitz-Esser, H E Neuhaus, P G Kroth, M Horn, and I Haferkamp. Diatom plastids depend on nucleotide import from the cytosol. *Proceedings of the National Academy of Sciences*, 106 (9):3621–3626, March 2009.

M De Vos H Judson D Hodge J Huntriss H M Picton E Sheridan DT Bonthron B E Hayward. Lack of involvement of known DNA methyltransferases in familial hydatidiform mole implies the involvement of other factors in establishment of imprinting in the human female germline. *BMC Genetics*, 4:2, 2003.

Ulrike Bacher, Claudia Haferlach, Susanne Schnittger, Alexander Kohlmann, Wolfgang Kern, and Torsten Haferlach. Mutations of the TET2 and CBL genes: novel molecular markers in myeloid malignancies. *Annals of Hematology*, 89(7):643–652, March 2010.

T E Bapiro, M P Lolkema, D I Jodrell, and D A Tuveson. nab-Paclitaxel Potentiates Gemcitabine Activity by Reducing Cytidine Deaminase Levels in a Mouse Model of Pancreatic Cancer. *Cancer discovery*, 2012.

Fanny Baran-Marszak, Remi Fagard, Bernard Girard, Sophie Camilleri-Bro euml t, Feng Zeng, Gilbert M Lenoir, Martine Rapha euml l, and Jean Feuillard. Gene Array Identification of Epstein Barr Virus-Regulated Cellular Genes in EBV-Converted Burkitt Lymphoma Cell Lines. *Laboratory Investigation*, 82(11):1463–1479, November 2002.

Janet M Barletta, Shirley Rainier, and Andrew P Feinberg. Reversal of Loss of Imprinting in Tumor Cells by 5-Aza-2-deoxycytidine. *Cancer Research*, 1997.

Tristan S Barnes, John H Bleskan, Iris M Hart, Katy A Walton, Jeffrey W Barton, and David Patterson. Purification of, Generation of Monoclonal Antibodies to, and Mapping of Phosphoribosyl N-Formylglycinamide Amidotransferase. *Biochemistry*, 33(7):1850–1860, February 1994.

Jordi Barretina, Giordano Caponigro, Nicolas Stransky, Kavitha Venkatesan, Adam A Margolin, Sungjoon Kim, Christopher J Wilson, Joseph Lehar, Gregory V Kryukov, Dmitriy Sonkin, Anupama Reddy, Manway Liu, Lauren Murray, Michael F Berger, John E Monahan, Paula Morais, Jodi Meltzer, Adam Korejwa, Judit Jane-Valbuena, Felipa A Mapa, Joseph Thibault, Eva Bric-Furlong, Pichai Raman, Aaron Shipway, Ingo H Engels, Jill Cheng, Guoying K Yu, Jianjun Yu, Peter Aspesi, Melanie de Silva, Kalpana Jagtap, Michael D Jones, Li Wang, Charles Hatton, Emanuele Palescandolo, Supriya Gupta, Scott Mahan, Carrie Sougnez, Robert C Onofrio, Ted Liefeld, Laura MacConaill, Wendy Winckler, Michael Reich, Nanxin Li, Jill P Mesirov, Stacey B Gabriel, Gad Getz, Kristin Ardlie, Vivien Chan, Vic E Myer, Barbara L Weber, Jeff Porter, Markus Warmuth, Peter Finan, Jennifer L Harris, Matthew Meyerson, Todd R Golub, Michael P Morrissey, William R Sellers, Robert Schlegel, and Levi A Garraway. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, 483(7391):603–607, March 2012.

Christina M Bender, Martha M Pao, and Peter A Jones. Inhibition of DNA Methylation by 5-Aza-2-deoxycytidine Suppresses the Growth of Human Tumor Cell Lines. *Cancer Research*, 1998.

Gh Benga. Basic studies on gene therapy of human malignant melanoma by use of the human interferon Beta gene entrapped in cationic multilamellar liposomes.: 1. Morphology and growth rate of six melanoma cell lines used in transfection experiments with the human interferon Beta gene. *Journal of Cellular and Molecular Medicine*, 5(4):402–408, October 2001.

M J Bessman, I R Lehman, J Adler, S B Zimmerman, E S Simms, and A Kornberg. ENZYMATIC SYNTHESIS OF DEOXYRIBONUCLEIC ACID. III. THE INCORPORATION OF PYRIMIDINE AND PURINE ANALOGUES INTO DEOXYRIBONUCLEIC ACID. *PNAS*, 44(7):633–640, July 1958.

T H Bestor. The DNA methyltransferases of mammals. *Human Molecular Genetics*, 9(16):2395–2402, October 2000.

T H Bestor and V M Ingram. Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA. *PNAS*, 80(18):5559–5563, September 1983.

T H Bestor and V M Ingram. Growth-dependent expression of multiple species of DNA methyltransferase in murine erythroleukemia cells. *PNAS*, 82(9):2674–2678, May 1985.

Timothy Bestor, Andrew Laudano, Robert Mattaliano, and Vernon Ingram. Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. *Journal of Molecular Biology*, 203(4):971–983, October 1988.

Deepika Bhatla, Robert B Gerbing, Todd A Alonzo, Heather Conner, Julie A Ross, Soheil Meshinchi, Xiaowen Zhai, Tiffany Zamzow, Parinda A Mehta, Hartmut Geiger, John Perentesis, and Stella M Davies. Cytidine deaminase genotype and toxicity of cytosine arabinoside therapy in children with acute myeloid leukemia. *British Journal of Haematology*, 144(3):388–394, February 2009.

V Bianchi and J Spychala. Mammalian 5'-Nucleotidases. *Journal of Biological Chemistry*, 278(47):46195–46198, November 2003.

A P Bird. Use of restriction enzymes to study eukaryotic DNA methylation: II. The symmetry of methylated sites supports semi-conservative copying of the methylation pattern. *Journal of Molecular Biology*, 118(1):49–60, January 1978.

A P Bird. CpG-rich islands and the function of DNA methylation. *Nature*, 321(6067):209–213, May 1986.

Adrian P Bird. CpG islands as gene markers in the vertebrate nucleus. *Trends in Genetics*, 3:342–347, January 1987.

Adrian P Bird and Edwin M Southern. Use of restriction enzymes to study eukaryotic DNA methylation. *Journal of Molecular Biology*, 118(1):27–47, January 1978.

Kathryn Blaschke, Kevin T Ebata, Mohammad M Karimi, Jorge A Zepeda-Martínez, Preeti Goyal, Sahasransu Mahapatra, Angela Tam, Diana J Laird, Martin Hirst, Anjana Rao, Matthew C Lorincz, and Miguel Ramalho-Santos. Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. *Nature*, 500(7461):222–226, August 2013.

Michael T Bocker, Francesca Tuorto, Günter Raddatz, Tanja Musch, Feng-Chun Yang, Mingjiang Xu, Frank Lyko, and Achim Breiling. Hydroxylation of 5-methylcytosine by TET2 maintains the active state of the mammalian HOXA cluster. *Nature Communications*, 3:818–, May 2012.

Christina Bohman and Staffan Eriksson. Purification and Properties of Human Deoxycytidine Kinase. In W L Nyhan, L F Thompson, and R W E Watts, editors, *Advances in Experimental Medicine and Biology*, pages 311–314–314. Springer US, Boston, MA, 1986.

Peter L Bonate, Larry Arthaud, William R Cantrell, Katherine Stephenson, John A Secrist, and Steve Weitman. Discovery and development of clofarabine: a nucleoside analogue for treating cancer. *Nature Reviews Drug Discovery*, 5(10):855–863, October 2006.

R J Boorstein, D D Levy, and G W Teebor. Toxicity of 3-aminobenzamide to Chinese hamster cells containing 5-hydroxymethyluracil in their DNA. *Cancer Research*, 47(16):4372–4377, August 1987.

R J Boorstein, L N Chiu, and G W Teebor. A mammalian cell line deficient in activity of the DNA repair enzyme 5-hydroxymethyluracil-DNA glycosylase is resistant to the toxic effects of the thymidine analog 5-hydroxymethyl-2-deoxyuridine. *Molecular and cellular ...*, 1992.

Michael J Booth, Miguel R Branco, Gabriella Ficz, David Oxley, Felix Krueger, Wolf Reik, and Shankar Balasubramanian. Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science*, 336(6083):934–937, May 2012.

D Bourc'his, G L Xu, C S Lin, B Bollman, and T H Bestor. Dnmt3L and the establishment of maternal genomic imprints. *Science*, 294(5551):2536–2539, December 2001.

Déborah Bourc'his and Timothy H Bestor. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature Cell Biology*, 431(7004):96–99, August 2004.

J Boyes and A Bird. Repression of genes by DNA methylation depends on CpG density and promoter strength: evidence for involvement of a methyl-CpG binding protein. *EMBO Journal*, 11(1):327–333, January 1992.

Dana Branzei and Marco Foiani. Regulation of DNA repair throughout the cell cycle. *Nature Reviews Molecular Cell Biology*, 9(4):297–308, February 2008.

Mandy M Brecqueville, Jérôme J Rey, François F Bertucci, Emilie E Coppin, Pascal P Finetti, Nadine N Carbuccia, Nathalie N Cervera, Véronique V Gelsi-Boyer, Christine C Arnoulet, Olivier O Gisserot, Denis D Verrot, Borhane B Slama, Norbert N Vey, Marie-Joelle MJ Mozziconacci, Daniel D Birnbaum, and Anne A Murati. Mutation analysis of ASXL1, CBL, DNMT3A, IDH1, IDH2, JAK2, MPL, NF1, SF3B1, SUZ12, and TET2 in myeloproliferative neoplasms. *Genes, Chromosomes and Cancer*, 51(8): 743–755, July 2012.

Ann-Marie Bröske, Lena Vockentanz, Shabnam Kharazi, Matthew R Huska, Elena Mancini, Marina Scheller, Christiane Kuhl, Andreas Enns, Marco Prinz, Rudolf Jaenisch, Claus Nerlov, Achim Leutz, Miguel A Andrade-Navarro, Sten Eirik W Jacobsen, and Frank Rosenbauer. DNA methylation protects hematopoietic stem cell multipotency from myeloerythroid restriction. *Nature Genetics*, 41(11): 1207–1215, October 2009.

G B Brown and P M Roll. The utilization of adenine for nucleic acid synthesis and as a precursor of guanine. *The Journal of biological chemistry*, 172(2):469–484, February 1948.

J T Brown, X Bai, and A W Johnson. The yeast antiviral proteins Ski2p, Ski3p, and Ski8p exist as a complex in vivo. *RNA*, 6(3):449–457, February 2000.

John H Bull and John C Wootton. Heavily methylated amplified DNA in transformants of Neurospora crassa. *Nature*, 310(5979):701–704, August 1984.

Lambert Busque, Jay P Patel, Maria E Figueroa, Aparna Vasanthakumar, Sylvie Provost, Zineb Hamilou, Luigina Mollica, Juan Li, Agnes Viale, Adriana Heguy, Maryam Hassimi, Nicholas Socci, Parva K Bhatt, Mithat Gonen, Christopher E Mason, Ari Melnick, Lucy A Godley, Cameron W Brennan, Omar Abdel-Wahab, and Ross L Levine. Recurrent somatic TET2 mutations in normal elderly individuals with clonal hematopoiesis. *Nature Genetics*, 44(11):1179–1181, November 2012.

M Busslinger, J Hurst, and R A Flavell. DNA methylation and the regulation of globin gene expression. *Cell*, 34(1):197–206, August 1983.

Vladimir J N Bykov, Natalia Issaeva, Alexandre Shilov, Monica Hultcrantz, Elena Pugacheva, Peter Chumakov, Jan Bergman, Klas G Wiman, and Galina Selivanova. Restoration of the tumor suppressor function to mutant p53 by a low-molecular-weight compound. *Nature Medicine*, 8(3):282–288, March 2002.

T Cacciamani, A Vita, G Cristalli, S Vincenzetti, P Natalini, S Ruggieri, A Amici, and G Magni. Purification of human cytidine deaminase: molecular and enzymatic characterization and inhibition by synthetic pyrimidine analogs. *Archives of biochemistry and biophysics*, 290(2):285–292, November 1991.

E E Cameron, K E Bachman, S Myöhänen, J G Herman, and S B Baylin. Synergy of demethylation and histone deacetylase inhibition in the re-expression of genes silenced in cancer. *Nature Genetics*, 21(1): 103–107, January 1999.

Gerald W Camiener. Studies of the enzymatic deamination of ara-cytidine—V. *Biochemical Pharmacology*, 17(9):1981–1991, September 1968.

Gerald W Camiener and Charles G Smith. Studies of the enzymatic deamination of cytosine arabinoside—I. *Biochemical Pharmacology*, 14(10):1405–1416, October 1965.

Pedro Cano Soldado and Marçal Pastor-Anglada. Transporters that translocate nucleosides and structural similar drugs: structural requirements for substrate recognition. *Medicinal research reviews*, 32(2): 428–457, 2012.

Floriana Capuano, Michael Mülleder, Robert Kok, Henk J Blom, and Markus Ralser. Cytosine DNA methylation is found in Drosophila melanogaster but absent in Saccharomyces cerevisiae, Schizosaccharomyces pombe, and other yeast species. *Analytical Chemistry*, 86(8):3697–3702, April 2014.

Jane D Carver and W Allan Walker. The role of nucleotides in human nutrition. *The Journal of Nutritional Biochemistry*, 6(2):58–72, 1995.

Bruce A Chabner, James C Drake, and David G Johns. Deamination of 5-azacytidine by a human leukemia cell cytidine deaminase. *Biochemical Pharmacology*, 22(21):2763–2765, 1973.

Grant A Challen, Deqiang Sun, Mira Jeong, Min Luo, Jaroslav Jelinek, Jonathan S Berg, Christoph Bock, Aparna Vasanthakumar, Hongcang Gu, Yuanxin Xi, Shoudan Liang, Yue Lu, Gretchen J Darlington, Alexander Meissner, Jean-Pierre J Issa, Lucy A Godley, Wei Li, and Margaret A Goodell. Dnmt3a is essential for hematopoietic stem cell differentiation. *Nature Genetics*, 44(1):23–31, January 2012.

Doug W Chan, Yi Wang, Meng Wu, Jiemin Wong, Jun Qin, and Yingming Zhao. Unbiased proteomic screen for binding proteins to modified lysines on histone H3. *PROTEOMICS*, 9(9):2343–2354, May 2009.

K-S Chan, C-G Koh, and H-Y Li. Mitosis-targeted anti-cancer therapies: where they stand. *Cell death & disease*, 3:e411, 2012.

Chun-Chang Chen, Keh-Yang Wang, and Che-Kun James Shen. The Mammalian de Novo DNA Methyltransferases DNMT3A and DNMT3B Are Also DNA 5-Hydroxymethylcytosine Dehydroxymethylases. *Journal of Biological Chemistry*, 2012a.

Qiang Chen, Yibin Chen, Chunjing Bian, Ryoji Fujiki, and Xiaochun Yu. TET2 promotes histone O-GlcNAcylation during gene transcription. *Nature*, 493(7433):561–564, December 2012b.

Y. Chu, M. Masoud, and G. Gebeyehu. Cationic lipids capable of facilitating transport of biologically active agents or substances into cells; lipid aggregate macromolecular complex interacts with cells making the polyanionic macromolecule available for absorption and uptake by the cell, January 20 2009. URL `http://www.google.com/patents/US7479573`. US Patent 7,479,573.

Sang J Chung, J Christopher Fromme, and Gregory L Verdine. Structure of Human Cytidine Deaminase Bound to a Potent Inhibitor. *Journal of Medicinal Chemistry*, 48(3):658–660, February 2005.

P F Coleman, D P Suttle, and G R Stark. Purification from hamster cells of the multifunctional protein that initiates de novo synthesis of pyrimidine nucleotides. *The Journal of biological chemistry*, 252(18): 6379–6385, September 1977.

Marcus S Cooke, Paul T Henderson, and Mark D Evans. Oxidative Stress-Induced Carcinogenesis and Its Prevention Guest Editor: Shinya Toyokuni Sources of Extracellular, Oxidatively-Modified DNA Lesions: Implications for Their Measurement in Urine. *Journal of Clinical Biochemistry and Nutrition*, 45(3): 255–270, 2009.

Daniel Cortázar, Christophe Kunz, Jim Selfridge, Teresa Lettieri, Yusuke Saito, Eilidh MacDougall, Annika Wirz, David Schuermann, Angelika L Jacobs, Fredy Siegrist, Roland Steinacher, Josef Jiricny, Adrian Bird, and Primo Schär. Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature*, 470(7334):419–423, April 2012.

Salvatore Cortellino, Jinfei Xu, Mara Sannai, Robert Moore, Elena Caretti, Antonio Cigliano, Madeleine Le Coz, Karthik Devarajan, Andy Wessels, Dianne Soprano, Lara K Abramowitz, Marisa S Bartolomei, Florian Rambow, Maria Rosaria Bassi, Tiziana Bruno, Maurizio Fanciulli, Catherine Renner, Andres J Klein-Szanto, Yoshihiro Matsumoto, Dominique Kobi, Irwin Davidson, Christophe Alberti, Lionel Larue, and Alfonso Bellacosa. Thymine DNA Glycosylase Is Essential for Active DNA Demethylation by Linked Deamination-Base Excision Repair. *Cell*, 146(1):67–79, July 2011.

P J Costa and K M Arndt. Synthetic lethal interactions suggest a role for the Saccharomyces cerevisiae Rtf1 protein in transcription elongation. *Genetics*, 156(2):535–547, October 2000.

Yael Costa, Junjun Ding, Thorold W Theunissen, Francesco Faiola, Timothy A Hore, Pavel V Shliaha, Miguel Fidalgo, Arven Saunders, Moyra Lawrence, Sabine Dietmann, Satyabrata Das, Dana N Levasseur, Zhe Li, Mingjiang Xu, Wolf Reik, José C R Silva, and Jianlong Wang. NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. *Nature*, 495(7441):370–374, March 2013.

Lucile Couronné, Christian Bastard, and Olivier A Bernard. TET2 and DNMT3A mutations in human T-cell lymphoma. *New England Journal of Medicine*, 366(1):95–96, January 2012.

Joseph M Covey and Daniel S Zaharko. Comparison of the in vitro cytotoxicity (L1210) of 5-Aza-2-deoxycytidine with its therapeutic and toxic effects in mice. *European Journal of Cancer and Clinical Oncology*, 21(1):109–117, January 1985.

J M Cunningham, E R Christensen, D J Tester, C Y Kim, P C Roche, L J Burgart, and S N Thibodeau. Hypermethylation of the hMLH1 promoter in colon cancer with microsatellite instability. *Cancer Research*, 58(15):3455–3460, August 1998.

John T Cunningham, Melissa V Moreno, Alessia Lodi, Sabrina M Ronen, and Davide Ruggero. Protein and Nucleotide Biosynthesis Are Coupled by a Single Rate-Limiting Enzyme, PRPS2, to Drive Cancer. *Cell*, 157(5):1088–1103, May 2014.

M H Taggart A P Bird D N Cooper. Unmethylated domains in vertebrate DNA. *Nucleic Acids Research*, 11(3):647, February 1983.

Vijaya L Damaraju, Delores Mowles, Sylvia Yao, Amy Ng, James D Young, Carol E Cass, and Zeen Tong. Role of Human Nucleoside Transporters in the Uptake and Cytotoxicity of Azacitidine and Decitabine. *Nucleosides, Nucleotides and Nucleic Acids*, 31(3):236–255, March 2012.

V J Davisson, W Sirawaraporn, and D V Santi. Expression of human thymidylate synthase in Escherichia coli. *The Journal of biological chemistry*, 264(16):9145–9148, June 1989.

Meelad M Dawlaty, Kibibi Ganz, Benjamin E Powell, Yueh-Chiang Hu, Styliani Markoulaki, Albert W Cheng, Qing Gao, Jongpil Kim, Sang-Woon Choi, David C Page, and Rudolf Jaenisch. Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. *Cell Stem Cell*, 9(2):166–175, August 2011.

Meelad M Dawlaty, Achim Breiling, Thuc Le, Günter Raddatz, M Inmaculada Barrasa, Albert W Cheng, Qing Gao, Benjamin E Powell, Zhe Li, Mingjiang Xu, Kym F Faull, Frank Lyko, and Rudolf Jaenisch. Combined Deficiency of Tet1 and Tet2 Causes Epigenetic Abnormalities but Is Compatible with Postnatal Development. *Developmental Cell*, 24(3):310–323, February 2013.

Meelad M Dawlaty, Achim Breiling, Thuc Le, M Inmaculada Barrasa, Günter Raddatz, Qing Gao, Benjamin E Powell, Albert W Cheng, Kym F Faull, Frank Lyko, and Rudolf Jaenisch. Loss of Tet enzymes compromises proper differentiation of embryonic stem cells. *Developmental Cell*, 29(1):102–111, April 2014.

L de la Rica, J Rodríguez-Ubreva, and M García. PU. 1 target genes undergo Tet2-coupled demethylation and DNMT3b-mediated methylation in monocyte-to-osteoclast differentiation. *Genome ...*, 2013.

B R de Saint Vincent, M Dechamps, and G Buttin. The modulation of the thymidine triphosphate pool of Chinese hamster cells by dCMP deaminase and UDP reductase. Thymidine auxotrophy induced by CTP in dCMP deaminase-deficient lines. *The Journal of biological chemistry*, 255(1):162–167, January 1980.

C H De Verdier and V R Potter. Alternative pathways of thymine and uracil metabolism in the liver and hepatoma. *Journal of the National Cancer Institute*, 24:13–29, January 1960.

Francois Delhommeau, Sabrina Dupont, Véronique Della Valle, Chloé James, Severine Trannoy, Aline Massé, Olivier Kosmider, Jean-Pierre Le Couedic, Fabienne Robert, Antonio Alberdi, Yann Lécluse, Isabelle Plo, François J Dreyfus, Christophe Marzac, Nicole Casadevall, Catherine Lacombe, Serge P Romana, Philippe Dessen, Jean Soulier, Franck Viguié, Michaëla Fontenay, William Vainchenker, and Olivier A Bernard. Mutation in TET2in Myeloid Cancers. *New England Journal of Medicine*, 360(22): 2289–2301, May 2009.

Rachel Deplus, Benjamin Delatte, Marie K Schwinn, Matthieu Defrance, Jacqui Méndez, Nancy Murphy, Mark A Dawson, Michael Volkmar, Pascale Putmans, Emilie Calonne, Alan H Shih, Ross L Levine, Olivier Bernard, Thomas Mercher, Eric Solary, Marjeta Urh, Danette L Daniels, and François Fuks. TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. *EMBO Journal*, 32(5):645–655, March 2013.

A C Diserens, N Tribolet, A Martin-Achard, A C Gaide, J F Schnegg, and S Carrel. Characterization of an established human malignant glioma cell line: LN-18. *Acta Neuropathologica*, 53(1):21–28, 1981.

J Doskočil and F Šorm. Distribution of 5-methylcytosine in pyrimidine sequences of deoxyribonucleic acids. *Biochimica et biophysica acta*, 55(6):953–959, June 1962.

John S Driscoll, Victor E Marquez, Jacqueline Plowman, Paul S Liu, James A Kelley, and Joseph J Barchi. Antitumor properties of 2(1H)-pyrimidinone riboside (zebularine) and its fluorinated analogs. *Journal of Medicinal Chemistry*, 34(11):3280–3284, November 1991.

John P Durham and David H Ives. Deoxycytidine Kinase. *Molecular pharmacology*, 1969.

John P Durham and David H Ives. Deoxycytidine Kinase. *Journal of Biological Chemistry*, 1970.

Daisy Dussoix and Werner Arber. Host specificity of DNA produced by Escherichia coli. *Journal of Molecular Biology*, 5(1):37–49, July 1962.

M Ehrlich, M A Gama-Sosa, L H Huang, R M Midgett, K C Kuo, R A McCune, and C Gehrke. Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. *Nucleic Acids Research*, 10(8):2709–2721, April 1982.

M Ehrlich, K F Norris, R Y Wang, K C Kuo, and C W Gehrke. DNA cytosine methylation and heat-induced deamination. *Bioscience reports*, 6(4):387–393, April 1986.

Nicoletta Eliopoulos, Denis Cournoyer, and Richard L Momparler. Drug resistance to 5-aza-2-deoxycytidine, 2, 2-difluorodeoxycytidine, and cytosine arabinoside conferred by retroviral-mediated transfer of human cytidine deaminase cDNA into murine cells. *Cancer Chemotherapy and Pharmacology*, 42(5):373–378, 1998.

Staffan Eriksson, Borys Kierdaszuk, Birgitte Munch-Petersen, Bo Oberg, and Nils Gunnar Johansson. Comparison of the substrate specificities of human thymidine kinase 1 and 2 and deoxycytidine kinase toward antiviral and cytostatic nucleoside analogs. *Biochemical and biophysical research communications*, 176(2):586–592, 1991.

Manel Esteller. Epigenetic gene silencing in cancer: the DNA hypermethylome. *Human Molecular Genetics*, 16 Spec No 1:R50–9, April 2007.

Manel Esteller. Epigenetics in Cancer. *New England Journal of Medicine*, 358(11):1148–1159, March 2008.

Fran oise Exinger and Fran ois Lacroute. 6-Azauracil inhibition of GTP biosynthesis in Saccharomyces cerevisiae. *Current Genetics*, 22(1):9–11, July 1992.

Ingrid Jakobsen Falk, Anna Fyrberg, Esbjörn Paul, Hareth Nahi, Monica Hermanson, Richard Rosenquist, Martin Höglund, Lars Palmqvist, Dick Stockelberg, Yuan Wei, Henrik Gréen, and Kourosh Lotfi. Decreased survival in normal karyotype AML with single-nucleotide polymorphisms in genes encoding the AraC metabolizing enzymes cytidine deaminase and 5-nucleotidase. *American Journal of Hematology*, 88(12):1001–1006, September 2013.

Andrew P Feinberg and Bert Vogelstein. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature*, 301(5895):89–92, January 1983.

Pierre Fenaux, Ghulam J Mufti, Eva Hellstrom-Lindberg, Valeria Santini, Carlo Finelli, Aristoteles Giagounidis, Robert Schoch, Norbert Gattermann, Guillermo Sanz, Alan List, Steven D Gore, John F Seymour, John M Bennett, John Byrd, Jay Backstrom, Linda Zimmerman, David McKenzie, C L Beach, and Lewis R Silverman. Efficacy of azacitidine compared with that of conventional care regimens in the treatment of higher-risk myelodysplastic syndromes: a randomised, open-label, phase III study. *The Lancet Oncology*, 10(3):223–232, March 2009.

Dana Ferraris, Bridget Duvall, Greg Delahanty, Bipin Mistry, Jesse Alt, Camilo Rojas, Christopher Rowbottom, Kristen Sanders, Edgar Schuck, Kuan-Chun Huang, Sanjeev Redkar, Barbara B Slusher, and Takashi Tsukamoto. Design, Synthesis, and Pharmacological Evaluation of Fluorinated Tetrahydrouridine Derivatives as Inhibitors of Cytidine Deaminase. *Journal of Medicinal Chemistry*, 57(6):2582–2588, March 2014.

R P Feynman and R Leighton. *"What Do You Care what Other People Think?": Further Adventures of a Curious Character*. Norton, 1988.

S Fields and O Song. A novel genetic system to detect protein-protein interactions. *Nature*, 340(6230): 245–246, July 1989.

Maria E Figueroa, Omar Abdel-Wahab, Chao Lu, Patrick S Ward, Jay Patel, Alan Shih, Yushan Li, Neha Bhagwat, Aparna Vasanthakumar, Hugo F Fernandez, Martin S Tallman, Zhuoxin Sun, Kristy Wolniak, Justine K Peeters, Wei Liu, Sung E Choe, Valeria R Fantin, Elisabeth Paietta, Bob Löwenberg, Jonathan D Licht, Lucy A Godley, Ruud Delwel, Peter J M Valk, Craig B Thompson, Ross L Levine, and

Ari Melnick. Leukemic IDH1 and IDH2 Mutations Result in a Hypermethylation Phenotype, Disrupt TET2 Function, and Impair Hematopoietic Differentiation. *Cancer cell*, 18(6):553–567, December 2010.

Sara M Fitzgerald, Rakesh K Goyal, William R A Osborne, Jennifer D Roy, John W Wilson, and R E Ferrell. Identification of functional single nucleotide polymorphism haplotypes in the cytidine deaminase promoter. *Human genetics*, 119(3):276–283, April 2006.

Jerry J Fong, Brenda L Nguyen, Robert Bridger, Estela E Medrano, Lance Wells, Shujuan Pan, and Richard N Sifers. β-N-Acetylglucosamine (O-GlcNAc) is a novel regulator of mitosis-specific phosphorylations on histone H3. *Journal of Biological Chemistry*, 287(15):12195–12203, April 2012.

S A Forbes, G Bhamra, S Bamford, E Dawson, C Kok, J Clements, A Menzies, J W Teague, P A Futreal, and M R Stratton. *The Catalogue of Somatic Mutations in Cancer (COSMIC)*. John Wiley & Sons, Inc., Hoboken, NJ, USA, May 2001.

Irving H Fox and William N Kelley. Human phosphoribosylpyrophosphate synthetase distribution, purification, and properties. *The Journal of biological chemistry*, 246(18):5739–5748, 1971.

A Franceschini, D Szklarczyk, S Frankild, M Kuhn, M Simonovic, A Roth, J Lin, P Minguez, P Bork, C von Mering, and L J Jensen. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Research*, 41(D1):D808–D815, December 2012.

Scott R Frank, Marianne Schroeder, Paula Fernandez, Stefan Taubert, and Bruno Amati. Binding of c-Myc to chromatin mediates mitogen-induced acetylation of histone H4 and gene activation. *Genes & ...*, 2001.

R E Franklin and R G Gosling. Molecular configuration in sodium thymonucleate. *Nature*, 171(4356): 740–741, April 1953.

K Frenkel, A Cummings, J Solomon, J Cadet, J J Steinberg, and G W Teebor. Quantitative determination of the 5-(hydroxymethyl)uracil moiety in the DNA of gamma-irradiated cells. *Biochemistry*, 24(17): 4527–4533, August 1985.

M Friedkin and D Roberts. The enzymatic synthesis of nucleosides. I. Thymidine phosphorylase in mammalian tissue. *The Journal of biological chemistry*, 207(1):245–256, March 1954a.

M Friedkin and D Roberts. The enzymatic synthesis of nucleosides. II. Thymidine and related pyrimidine nucleosides. *The Journal of biological chemistry*, 207(1):257–266, March 1954b.

Ryoji Fujiki, Waka Hashiba, Hiroki Sekine, Atsushi Yokoyama, Toshihiro Chikanishi, Saya Ito, Yuuki Imai, Jaehoon Kim, Housheng Hansen He, Katsuhide Igarashi, Jun Kanno, Fumiaki Ohtake, Hirochika Kitagawa, Robert G Roeder, Myles Brown, and Shigeaki Kato. GlcNAcylation of histone H2B facilitates its monoubiquitination. *Nature*, November 2011.

Helge Gad, Tobias Koolmeister, Ann-Sofie Jemth, Saeed Eshtad, Sylvain A Jacques, Cecilia E Ström, Linda M Svensson, Niklas Schultz, Thomas Lundbäck, Berglind Osk Einarsdottir, Aljona Saleh, Camilla Göktürk, Pawel Baranczewski, Richard Svensson, Ronnie P-A Berntsson, Robert Gustafsson, Kia Strömberg, Kumar Sanjiv, Marie-Caroline Jacques-Cordonnier, Matthieu Desroses, Anna-Lena Gustavsson, Roger Olofsson, Fredrik Johansson, Evert J Homan, Olga Loseva, Lars Bräutigam, Lars Johansson, Andreas Höglund, Anna Hagenkort, Therese Pham, Mikael Altun, Fabienne Z Gaugaz, Svante Vikingsson, Bastiaan Evers, Martin Henriksson, Karl S A Vallin, Olov A Wallner, Lars G J Hammarström, Elisee Wiita, Ingrid Almlöf, Christina Kalderén, Hanna Axelsson, Tatjana Djureinovic, Jordi Carreras Puigvert, Maria Häggblad, Fredrik Jeppsson, Ulf Martens, Cecilia Lundin, Bo Lundgren, Ingrid Granelli, Annika Jenmalm Jensen, Per Artursson, Jonas A Nilsson, Pål Stenmark, Martin Scobie, Ulrika Warpman Berglund, and Thomas Helleday. MTH1 inhibition eradicates cancer by preventing sanitation of the dNTP pool. *Nature*, 508(7495):215–221, April 2014.

Carlos M Galmarini, John R Mackey, and Charles Dumontet. Nucleoside analogues and nucleobases in cancer treatment. *The Lancet Oncology*, 3(7):415–424, July 2002.

Michael Y Galperin, Olga V Moroz, Keith S Wilson, and Alexey G Murzin. House cleaning, a part of good housekeeping. *Molecular Microbiology*, 59(1):5–19, January 2006.

Miguel A Gama-Sosa, Valerie A Slagel, Ronald W Trewyn, Ronald Oxenhandler, Kenneth C Kuo, Charles W Gehrke, and Melanie Ehrlich. The 5-methylcytosine content of DNA from human tumors. *Nucleic Acids Research*, 11(19):6883–6894, 1983.

Y Gao, J Chen, K Li, T Wu, B Huang, W Liu, and X Kou. Replacement of Oct4 by Tet1 during iPSC Induction Reveals an Important Role of DNA Methylation and Hydroxymethylation in Reprogramming. *Cell Stem Cell*, 2013.

G Garcia-Manero, E Jabbour, G Borthakur, S Faderl, Z Estrov, H Yang, S Maddipoti, L A Godley, N Gabrail, J G Berdeja, A Nadeem, L Kassalow, and H Kantarjian. Randomized Open-Label Phase II Study of Decitabine in Patients With Low- or Intermediate-Risk Myelodysplastic Syndromes. *Journal of Clinical Oncology*, 31(20):2548–2553, July 2013.

Yubin Ge, Tanya L Jensen, Mark L Stout, Robin M Flatley, Patrick J Grohar, Yaddanapudi Ravindranath, Larry H Matherly, and Jeffrey W Taub. The role of cytidine deaminase and GATA1 mutations in the increased cytosine arabinoside sensitivity of Down syndrome myeloblasts and leukemia cell lines. *Cancer Research*, 64(2):728–735, January 2004.

Yubin Ge, Mark L Stout, Dana A Tatman, Tanya L Jensen, Steven Buck, Ronald L Thomas, Yaddanapudi Ravindranath, Larry H Matherly, and Jeffrey W Taub. GATA1, Cytidine Deaminase, and the High Cure Rate of Down Syndrome Children With Acute Megakaryocytic Leukemia. *Journal of the ...*, 2005.

Kathrin K Geyer, Carlos M Rodríguez López, Iain W Chalmers, Sabrina E Munshi, Martha Truscott, James Heald, Mike J Wilkinson, and Karl F Hoffmann. Cytosine methylation regulates oviposition in the pathogenic blood fluke Schistosoma mansoni. *Nature Communications*, 2:424–424, January 2011.

A M Gholami, H Hahne, Z Wu, F J Auer, and C Meng. Global Proteome Analysis of the NCI-60 Cell Line Panel. *Cell reports*, 2013.

E Giovannetti, A C Laan, E Vasile, C Tibaldi, S Nannizzi, S Ricciardi, A Falcone, R Danesi, and G J Peters. Correlation Between Cytidine Deaminase Genotype and Gemcitabine Deamination in Blood Samples. *Nucleosides, Nucleotides and Nucleic Acids*, 27(6-7):720–725, July 2008.

Daniel Globisch, Martin Munzel, Markus Muller, Stylianos Michalakis, Mirko Wagner, Susanne Koch, Tobias Bruckl, Martin Biel, and Thomas Carell. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLOS ONE*, 5(12):e15367, 2010.

M Gold and J Hurwitz. The enzymatic methylation of ribonucleic acid and deoxyribonucleic acid. V. Purification and properties of the deoxyribonucleic acid-methylating activity of Escherichia coli. *The Journal of biological chemistry*, 239:3858–3865, November 1964.

David C Goldstone, Valerie Ennis-Adeniran, Joseph J Hedden, Harriet C T Groom, Gillian I Rice, Evangelos Christodoulou, Philip A Walker, Geoff Kelly, Lesley F Haire, Melvyn W Yap, Luiz Pedro S de Carvalho, Jonathan P Stoye, Yanick J Crow, Ian A Taylor, and Michelle Webb. HIV-1 restriction factor SAMHD1 is a deoxynucleoside triphosphate triphosphohydrolase. *Nature*, 480(7377):379–382, December 2011.

Zhizhong Gong and Jian-Kang Zhu. Active DNA demethylation by oxidation and repair. *Nature Publishing Group*, 21(12):1649–1651, August 2011.

M Gonzalez-Zulueta, C M Bender, A S Yang, T Nguyen, R W Beart, J M Van Tornout, and P A Jones. Methylation of the 5' CpG island of the p16/CDKN2 tumor suppressor gene in normal and transformed human tissues correlates with gene silencing. *Cancer Research*, 55(20):4531–4535, October 1995.

S D Gore. Combined DNA Methyltransferase and Histone Deacetylase Inhibition in the Treatment of Myeloid Neoplasms. *Cancer Research*, 66(12):6361–6369, June 2006.

H Gowher. DNA of Drosophila melanogaster contains 5-methylcytosine. *The EMBO Journal*, 19(24): 6918–6923, December 2000.

Aurelien Grosdidier, Vincent Zoete, and Olivier Michielin. SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Research*, 39(Web Server issue):W270–7, July 2011.

M Groudine and K Conkin. Chromatin structure and de novo methylation of sperm DNA: implications for activation of the paternal genome. *Science*, 228(4703):1061–1068, May 1985.

Tian-Peng Gu, Fan Guo, Hui Yang, Hai-Ping Wu, Gui-Fang Xu, Wei Liu, Zhi-Guo Xie, Linyu Shi, Xinyi He, Seung-gi Jin, Khursheed Iqbal, Yujiang Geno Shi, Zixin Deng, Piroska E Szabó, Gerd P Pfeifer, Jinsong Li, and Guo-Liang Xu. The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature*, 477(7366):606–610, September 2011.

Paul Guilhamon, Malihe Eskandarpour, Dina Halai, Gareth A Wilson, Andrew Feber, Andrew E Teschendorff, Valenti Gomez, Alexander Hergovich, Roberto Tirabosco, M Fernanda Amary, Daniel Baumhoer, Gernot Jundt, Mark T Ross, Adrienne M Flanagan, and Stephan Beck. Meta-analysis of IDH-mutant cancers identifies EBF1 as an interaction partner for TET2. *Nature Communications*, 4:2166–2166, January 2013.

C Guillouf. Spi-1/PU.1 Oncoprotein Affects Splicing Decisions in a Promoter Binding-dependent Manner. *Journal of Biological Chemistry*, 281(28):19145–19155, May 2006.

Fan Guo, Xianlong Li, Dan Liang, Tong Li, Ping Zhu, Hongshan Guo, Xinglong Wu, Lu Wen, Tian-Peng Gu, Boqiang Hu, Colum P Walsh, Jinsong Li, Fuchou Tang, and Guo-Liang Xu. Active and Passive Demethylation of Male and Female Pronuclear DNA in the Mammalian Zygote. *Cell Stem Cell*, September 2014.

Junjie U Guo, Yijing Su, Chun Zhong, Guo-li Ming, and Hongjun Song. Hydroxylation of 5-Methylcytosine by TET1 Promotes Active DNA Demethylation in the Adult Brain. *Cell*, 145(3):423–434, April 2011.

Junjie U Guo, Yijing Su, Joo Heon Shin, Jaehoon Shin, Hongda Li, Bin Xie, Chun Zhong, Shaohui Hu, Thuc Le, Guoping Fan, Heng Zhu, Qiang Chang, Yuan Gao, Guo-li Ming, and Hongjun Song. Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nature Neuroscience*, 17(2):215–222, December 2013.

Yuriy Gusev, Jason Sparkowski, Arumugham Raghunathan, Harley Ferguson, Jr., Jane Montano, Nancy Bogdan, Barry Schweitzer, Steven Wiltshire, Stephen F Kingsmore, and Warren Maltzman. Rolling circle amplification: a new approach to increase sensitivity for immunohistochemistry and flow cytometry. *The American journal of pathology*, 159(1):63–69, 2001.

Maria A Hahn, Runxiang Qiu, Xiwei Wu, Arthur X Li, Heying Zhang, Jun Wang, Jonathan Jui, Seung-gi Jin, Yong Jiang, Gerd P Pfeifer, and Qiang Lu. Dynamics of 5-hydroxymethylcytosine and chromatin marks in Mammalian neurogenesis. *CellReports*, 3(2):291–300, February 2013.

P Hajkova, S J Jeffries, C Lee, N Miller, and S P Jackson. Genome-Wide Reprogramming in the Mouse Germ Line Entails the Base Excision Repair Pathway. *Science*, 2010.

Douglas Hanahan and Robert A Weinberg. Hallmarks of Cancer: The Next Generation. *Cell*, 144(5): 646–674, March 2011.

Standish C Hartman. Phosphoribosyl Pyrophosphate Amidotransferase. *Journal of Biological Chemistry*, 1963.

L H Hartwell and T A Weinert. Checkpoints: controls that ensure the order of cell cycle events. *Science*, 1989.

H Hashimoto, Y Liu, A K Upadhyay, Y Chang, S B Howerton, P M Vertino, X Zhang, and X Cheng. Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Research*, 40(11):4841–4849, June 2012a.

Hideharu Hashimoto, Samuel Hong, Ashok S Bhagwat, Xing Zhang, and Xiaodong Cheng. Excision of 5-hydroxymethyluracil and 5-carboxylcytosine by the thymine DNA glycosylase domain: its structural basis and implications for active DNA demethylation. *Nucleic acids ...*, 2012b.

Hideharu Hashimoto, Xing Zhang, and Xiaodong Cheng. Excision of thymine and 5-hydroxymethyluracil by the MBD4 DNA glycosylase domain: structural basis and implications for active DNA demethylation. *Nucleic Acids Research*, 2012c.

S Hattman, C Kenny, L Berger, and K Pratt. Comparative study of DNA methylation in three unicellular eucaryotes. *Journal of bacteriology*, 135(3):1156–1157, September 1978.

P Hatzis, A S Al-Madhoon, M Jullig, T G Petrakis, S Eriksson, and I Talianidis. The Intracellular Localization of Deoxycytidine Kinase. *Journal of Biological Chemistry*, 273(46):30239–30243, November 1998.

Saugata Hazra, Stephan Ort, Manfred Konrad, and Arnon Lavie. Structural and Kinetic Characterization of Human Deoxycytidine Kinase Variants Able To Phosphorylate 5-Substituted Deoxycytidine and Thymidine Analogues,. *Biochemistry*, 49(31):6784–6790, August 2010.

Tapas K Hazra, Yoke W Kow, Zafar Hatahet, Barry Imhoff, Istvan Boldogh, Sanath K Mokkapati, Sankar Mitra, and Tadahide Izumi. Identification and characterization of a novel human DNA glycosylase for repair of cytosine-derived lesions. *The Journal of biological chemistry*, 277(34):30417–30420, August 2002.

Y F He, B Z Li, Z Li, P Liu, Y Wang, Q Tang, J Ding, Y Jia, Z Chen, L Li, Y Sun, X Li, Q Dai, C X Song, K Zhang, C He, and G L Xu. Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. *Science*, 333(6047):1303–1307, September 2011.

Asaf Hellman and Andrew Chess. Gene body-specific methylation on the active X chromosome. *Science*, 315(5815):1141–1143, February 2007.

Michael J Hendzel, Yi Wei, Michael A Mancini, Aaron Van Hooser, Tamara Ranalli, B R Brinkley, David P Bazett-Jones, and C David Allis. Mitosis-specific phosphorylation of histone H3 initiates primarily within pericentromeric heterochromatin during G2 and spreads in an ordered fashion coincident with mitotic chromosome condensation. *Chromosoma*, 106(6):348–360, 1997.

J G Herman, F Latif, Y Weng, M I Lerman, B Zbar, S Liu, D Samid, D S Duan, J R Gnarra, and W M Linehan. Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. *PNAS*, 91(21):9700–9704, October 1994.

Howard V Hershey, Joseph F Stieber, and Gerald C Mueller. DNA Synthesis in Isolated HeLa Nuclei. *European Journal of Biochemistry*, 34(2):383–394, March 2005.

Sunil R Hingorani, Lifu Wang, Asha S Multani, Chelsea Combs, Therese B Deramaudt, Ralph H Hruban, Anil K Rustgi, Sandy Chang, and David A Tuveson. Trp53 R172HandKras G12Dcooperate to promote chromosomal instability and widely metastatic pancreatic ductal adenocarcinoma in mice. *Cancer cell*, 7(5):469–483, January 2005.

S L Holbeck, J M Collins, and J H Doroshow. Analysis of Food and Drug Administration-Approved Anticancer Agents in the NCI60 Panel of Human Tumor Cell Lines. *Molecular Cancer Therapeutics*, 9 (5):1451–1460, May 2010.

J A Holden, G S Meredith, and W N Kelley. Human Adenine Phosphoribosyltransferase - Affinity Purification, Subunit Structure, Amino-Acid Composition, and Peptide-Mapping. *The Journal of biological chemistry*, 254(15):6951–6955, 1979.

N J Hoogenraad and D C Lee. Effect of uridine on de novo pyrimidine biosynthesis in rat hepatoma cells in culture. *The Journal of biological chemistry*, 249(9):2763–2768, May 1974.

R D Hotchkiss. The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. *The Journal of biological chemistry*, 175(1):315–332, August 1948.

C Y Howell, T H Bestor, F Ding, K E Latham, C Mertineit, J M Trasler, and J R Chaillet. Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene. *Cell*, 104(6):829–838, March 2001.

Chih-Hung Hsu, Kai-Lin Peng, Ming-Lun Kang, Yi-Ren Chen, Yu-Chih Yang, Chin-Hsien Tsai, Chi-Shuen Chu, Yung-Ming Jeng, Yen-Ting Chen, Feng-Mao Lin, Hsien-Da Huang, Yun-Yuh Lu, Yu-Ching Teng, Shinn-Tsuen Lin, Ruo-Kai Lin, Fan-Mei Tang, Sung-Bau Lee, Huan Ming Hsu, Jyh-Cherng Yu, Pei-Wen Hsiao, and Li-Jung Juan. TET1 Suppresses Cancer Invasion by Activating the Tissue Inhibitors of Metalloproteinases. *Cell reports*, 2(3):568–579, September 2012.

Chun-Mei Hu, Ming-Tyng Yeh, Ning Tsao, Chih-Wei Chen, Quan-Ze Gao, Chia-Yun Chang, Ming-Hsiang Lee, Jim-Min Fang, Sheh-Yi Sheu, Chow-Jaw Lin, Mei-Chun Tseng, Yu-Ju Chen, and Zee-Fen Chang. Tumor Cells Require Thymidylate Kinase to Prevent dUTP Incorporation during DNA Repair. *Cancer cell*, 22(1):36–50, July 2012.

Lulu Hu, Ze Li, Jingdong Cheng, Qinhui Rao, Wei Gong, Mengjie Liu, Yujiang Geno Shi, Jiayu Zhu, Ping Wang, and Yanhui Xu. Crystal Structure of TET2-DNA Complex: Insight into TET-Mediated 5mC Oxidation. *Cell*, 155(7):1545–1555, December 2013.

Hao Huang, Xi Jiang, Zejuan Li, Yuanyuan Li, Chun-Xiao Song, Chunjiang He, Miao Sun, Ping Chen, Sandeep Gurbuxani, Jiapeng Wang, Gia-Ming Hong, Abdel G Elkahloun, Stephen Arnovitz, Jinhua Wang, Keith Szulwach, Li Lin, Craig Street, Mark Wunderlich, Meelad Dawlaty, Mary Beth Neilly, Rudolf Jaenisch, Feng-Chun Yang, James C Mulloy, Peng Jin, Paul P Liu, Janet D Rowley, Mingjiang Xu, Chuan He, and Jianjun Chen. TET1 plays an essential oncogenic role in MLL-rearranged leukemia. *Proceedings of the National Academy of Sciences*, 110(29):11994–11999, July 2013.

Yun Huang and Anjana Rao. Connections between TET proteins and aberrant DNA modification in cancer. *Trends in Genetics*, August 2014.

Yun Huang, Lukas Chavez, Xing Chang, Xue Wang, William A Pastor, Jinsuk Kang, Jorge A Zepeda-Martínez, Utz J Pape, Steven E Jacobsen, Bjoern Peters, and Anjana Rao. Distinct roles of the methylcytosine oxidases Tet1 and Tet2 in mouse embryonic stem cells. *PNAS*, 111(4):1361–1366, January 2014.

Kilian V M Huber, Eidarus Salah, Branka Radic, Manuela Gridling, Jonathan M Elkins, Alexey Stukalov, Ann-Sofie Jemth, Camilla Göktürk, Kumar Sanjiv, Kia Strömberg, Therese Pham, Ulrika Warpman Berglund, Jacques Colinge, Keiryn L Bennett, Joanna I Loizou, Thomas Helleday, Stefan Knapp, and Giulio Superti-Furga. Stereospecific targeting of MTH1 by (S)-crizotinib as an anticancer strategy. *Nature*, 508(7495):222–227, April 2014.

Robert B Hurlbert and Harold O Kammen. Formation of cytidine nucleotides from uridine nucleotides by soluble mammalian enzymes: requirements for glutamine and guanosine nucleotides. *The Journal of biological chemistry*, 235(2):443–449, 1960.

John J Irwin, Teague Sterling, Michael M Mysinger, Erin S Bolstad, and Ryan G Coleman. ZINC: A Free Tool to Discover Chemistry for Biology. *Journal of Chemical Information and Modeling*, 52(7): 1757–1768, July 2012.

Ryo Ito, Shogo Katsura, Hiroki Shimada, Hikaru Tsuchiya, Masashi Hada, Tomoko Okumura, Akira Sugawara, and Atsushi Yokoyama. TET3-OGT interaction increases the stability and the presence of OGT in chromatin. *Genes to cells : devoted to molecular & cellular mechanisms*, 19(1):52–65, January 2014.

S Ito, L Shen, Q Dai, S C Wu, L B Collins, J A Swenberg, C He, and Y Zhang. Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. *Science*, 333(6047):1300–1303, September 2011.

Shinsuke Ito, Ana C D'Alessio, Olena V Taranova, Kwonho Hong, Lawrence C Sowers, and Yi Zhang. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature*, 466(7310):1129–1133, July 2010.

Lakshminarayan Iyer, Mamta Tahiliani, Anjana Rao, and L Aravind. Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle*, 8 (11):1698–1710, May 2009.

Rudolf Jaenisch, Frank Lyko, and Bernard H Ramsahoye. Development: DNA methylation in Drosophila melanogaster : Article : Nature. *Nature*, 408(6812):538–540, November 2000.

A M Jankowska, H Szpurka, R V Tiu, H Makishima, M Afable, J Huh, C L O'Keefe, R Ganetzky, M A McDevitt, and J P Maciejewski. Loss of heterozygosity 4q24 and TET2 mutations associated with myelodysplastic/myeloproliferative neoplasms. *Blood*, 113(25):6403–6410, June 2009.

A Jekunen and J A Vilpo. Reversal of deamination-related cytotoxicity of 5-methyl-2'-deoxycytidine by tetrahydrouridine in human leukemia cells. *Journal of the National Cancer Institute*, 73(5):1087–1091, November 1984.

Antti Jekunen, Matti Puukka, and Juhami Vilpo. Exclusion of exogenous 5-methyl-2'-deoxycytidine from DNA in human leukemic cells. *Biochemical Pharmacology*, 32(7):1165–1168, April 1983.

Mira Jeong, Deqiang Sun, Min Luo, Yun Huang, Grant A Challen, Benjamin Rodriguez, Xiaotian Zhang, Lukas Chavez, Hui Wang, Rebecca Hannah, Sang-Bae Kim, Liubin Yang, Myunggon Ko, Rui Chen, Berthold Göttgens, Ju-Seog Lee, Preethi Gunaratne, Lucy A Godley, Gretchen J Darlington, Anjana Rao, Wei Li, and Margaret A Goodell. Large conserved domains of low DNA methylation maintained by Dnmt3a. *Nature Genetics*, 46(1):17–23, January 2014.

Debin Ji, Changjun You, Pengcheng Wang, and Yinsheng Wang. Effects of Tet-Induced Oxidation Products of 5-Methylcytosine on DNA Replication in Mammalian Cells. *Chemical Research in Toxicology*, 27(7): 1304–1309, July 2014.

Lan Jiang, Jing Zhang, Jing-Jing Wang, Lu Wang, Li Zhang, Guoqiang Li, Xiaodan Yang, Xin Ma, Xin Sun, Jun Cai, Jun Zhang, Xingxu Huang, Miao Yu, Xuegeng Wang, Feng Liu, Chung-I Wu, Chuan He, Bo Zhang, Weimin Ci, and Jiang Liu. Sperm, but not oocyte, DNA methylome is inherited by zebrafish early embryos. *Cell*, 153(4):773–784, May 2013.

M Johansson and A Karlsson. Cloning of the cDNA and Chromosome Localization of the Gene for Human Thymidine Kinase 2. *Journal of Biological Chemistry*, 272(13):8454–8458, March 1997.

Magnus Johansson and Anna Karlsson. Differences in kinetic properties of pure recombinant human and mouse deoxycytidine kinase. *Biochemical Pharmacology*, 50(2):163–168, January 1995.

N G Johansson and S Eriksson. Structure-activity relationships for phosphorylation of nucleoside analogs to monophosphates by nucleoside kinases. *Acta biochimica Polonica*, 43(1):143–160, 1996.

Treat B Johnson and Robert D Coghill. RESEARCHES ON PYRIMIDINES. C111. THE DISCOVERY OF 5-METHYL-CYTOSINE IN TUBERCULINIC ACID, THE NUCLEIC ACID OF THE TUBERCLE BACILLUS 1. *Journal of the American Chemical Society*, 47(11):2838–2844, November 1925.

M E Jones, L Spector, and F Lipmann. CARBAMYL PHOSPHATE, THE CARBAMYL DONOR IN ENZYMATIC CITRULLINE SYNTHESIS 1. *Journal of the American Chemical Society*, 77(3):819–820, February 1955.

P A Jones. Effects of 5-azacytidine and its 2-deoxyderivative on cell differentiation and DNA methylation. *Pharmacology & Therapeutics*, 28(1):17–27, December 1984.

Peter A Jones. At the tipping point for epigenetic therapies in cancer. *Journal of Clinical Investigation*, 124(1):14–16, January 2014.

Peter A Jones and Shirley M Taylor. Cellular differentiation, cytidine analogs and DNA methylation. *Cell*, 20(1):85–93, May 1980.

L Jordheim and C Dumontet. Review of recent studies on resistance to cytotoxic deoxynucleoside analogues. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, 1776(2):138–159, December 2007.

Lars Petter Jordheim, David Durantel, Fabien Zoulim, and Charles Dumontet. Advances in the development of nucleoside and nucleotide analogues for cancer and viral diseases. *Nature Reviews Drug Discovery*, 12(6):447–464, June 2013.

Joohee Jung. Human Tumor Xenograft Models for Preclinical Assessment of Anticancer Drug Development. *Toxicological Research*, 30(1):1, March 2014.

R Jüttermann, E Li, and R Jaenisch. Toxicity of 5-aza-2-deoxycytidine to mammalian cells is mediated primarily by covalent trapping of DNA methyltransferase rather than DNA demethylation. *PNAS*, 91 (25):11797–11801, December 1994.

Garrett A Kaas, Chun Zhong, Dawn E Eason, Daniel L Ross, Raj V Vachhani, Guo-li Ming, Jennifer R King, Hongjun Song, and J David Sweatt. TET1 Controls CNS 5-Methylcytosine Hydroxylation, Active DNA Demethylation, Gene Transcription, and Memory Formation. *Neuron*, 79(6):1086–1093, September 2013.

S A Kadhim, T L Bowlin, W R Waud, E G Angers, L Bibeau, J M DeMuys, K Bednarski, A Cimpoia, and G Attardo. Potent antitumor activity of a novel nucleoside analogue, BCH-4556 (beta-L-dioxolane-cytidine), in human renal cell carcinoma xenograft tumor models. *Cancer Research*, 57(21):4803–4810, November 1997.

Eric M Kallin, Javier Rodríguez-Ubreva, Jesper Christensen, Luisa Cimmino, Iannis Aifantis, Kristian Helin, Esteban Ballestar, and Thomas Graf. Tet2 Facilitates the Derepression of Myeloid Target Genes during CEBPα-Induced Transdifferentiation of Pre-B Cells. *Molecular Cell*, 48(2):266–276, October 2012.

H Kamiya, N Murata-Kamiya, N Karino, and Y Ueno. Induction of T $\rightarrow$ G and T $\rightarrow$ A transversions by 5-formyluracil in mammalian cells. *... /Genetic Toxicology and ...*, 2002.

Karlheinz Kampfenkel, Torsten Mohlmann, Olaf Batz, Marc Van Montagu, Dirk Inze, and H Ekkehard Neuhaus. Molecular characterization of an Arabidopsis thaliana cDNA encoding a novel putative adenylate translocator of higher plants. *FEBS Letters*, 374(3):351–355, November 1995.

Yae Kanai, Saori Ushijima, Yukihiro Nakanishi, Michiie Sakamoto, and Setsuo Hirohashi. Mutation of the DNA methyltransferase (DNMT) 1 gene in human colorectal cancers. *Cancer letters*, 192(1):75–82, March 2003.

Syu-ichi Kanno, Takako Hiura, Takaharu Ohtake, Kimiko Koiwai, Hiroyoshi Suzuki, Mayuko Ujibe, and Masaaki Ishikawa. Characterization of resistance to cytosine arabinoside (Ara-C) in NALM-6 human B leukemia cells. *Clinica Chimica Acta*, 377(1-2):144–149, February 2007.

Metin Karahoca and Richard L Momparler. Pharmacokinetic and pharmacodynamic analysis of 5-aza-2-deoxycytidine (decitabine) in the design of its dose-schedule for cancer therapy. *Clinical Epigenetics*, 5 (1):3, 2013.

Mariko Katoh, Tomaž Curk, Qikai Xu, Blaž Zupan, Adam Kuspa, and Gad Shaulsky. Developmentally Regulated DNA Methylation in Dictyostelium discoideum. *Eukaryotic Cell*, 5(1):18–25, January 2006.

Elliot R Kaufman. Biochemical analysis of toxic effects of 5-hydroxymethyl-2-deoxyuridine in mammalian cells. *Somatic Cell and Molecular Genetics*, 12(5):501–512, September 1986.

Matthew W Kellinger, Chun-Xiao Song, Jenny Chong, Xing-Yu Lu, Chuan He, and Dong Wang. 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. *Nature Structural & Molecular Biology*, 19(8):831–833, July 2012.

Kristin Kemmerich, Felix A Dingler, Cristina Rada, and Michael S Neuberger. Germline ablation of SMUG1 DNA glycosylase causes loss of 5-hydroxymethyluracil- and UNG-backup uracil-excision activities and increases cancer predisposition of Ung-/-Msh2-/- mice. *Nucleic acids . . .*, 2012.

Chong Ho Kim, Victor E Marquez, David T Mao, David R Haines, and John J McCormack. Synthesis of pyrimidin-2-one nucleosides as acid-stable inhibitors of cytidine deaminase. *Journal of Medicinal Chemistry*, 29(8):1374–1380, August 1986.

Jaehoon Kim, Mohamed Guermah, and Robert G Roeder. The human PAF1 complex acts in chromatin transcription elongation both independently and cooperatively with SII/TFIIS. *Cell*, 140(4):491–503, February 2010.

Seon-Young Kim, Gwangsik Shin, Kang, Sungjin Yang, Su-Jin Baek, and Yong-Su Jeong. GENT: Gene Expression Database of Normal and Tumor Tissues. *Cancer Informatics*, 10:149, May 2011.

K Kincaid. Exploration of factors driving incorporation of unnatural dNTPS into DNA by Klenow fragment (DNA polymerase I) and DNA polymerase . *Nucleic Acids Research*, 33(8):2620–2628, April 2005.

A Klungland, R Paulsen, V Rolseth, Y Yamada, Y Ueno, P Wiik, A Matsuda, E Seeberg, and S Bjelland. 5-Formyluracil and its nucleoside derivatives confer toxicity and mutagenicity to mammalian cells by interfering with normal RNA and DNA metabolism. *Toxicology Letters*, 119(1):71–78, January 2001.

Myunggon Ko, Yun Huang, Anna M Jankowska, Utz J Pape, Mamta Tahiliani, Hozefa S Bandukwala, Jungeun An, Edward D Lamperti, Kian Peng Koh, Rebecca Ganetzky, X Shirley Liu, L Aravind, Suneet Agarwal, Jaroslaw P Maciejewski, and Anjana Rao. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature*, 468(7325):839–843, November 2010.

Myunggon Ko, Hozefa S Bandukwala, Jungeun An, Edward D Lamperti, Elizabeth C Thompson, Ryan Hastie, Angeliki Tsangaratou, Klaus Rajewsky, Sergei B Koralov, and Anjana Rao. Ten-Eleven-Translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice. *Proceedings of the National Academy of Sciences*, 108(35):14566–14571, August 2011.

Myunggon Ko, Jungeun An, Hozefa S Bandukwala, Lukas Chavez, Tarmo Äijö, William A Pastor, Matthew F Segal, Huiming Li, Kian Peng Koh, Harri Lähdesmäki, Patrick G Hogan, L Aravind, and Anjana Rao. Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. *Nature*, 497(7447):122–126, May 2013.

Kian Peng Koh, Akiko Yabuuchi, Sridhar Rao, Yun Huang, Kerrianne Cunniff, Julie Nardone, Asta Laiho, Mamta Tahiliani, Cesar A Sommer, Gustavo Mostoslavsky, Riitta Lahesmaa, Stuart H Orkin, Scott J Rodig, George Q Daley, and Anjana Rao. Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell Stem Cell*, 8(2):200–213, February 2011.

Arthur Kornberg, Steven B Zimmerman, S R Kornberg, and John Josse. Enzymatic synthesis of deoxyribonucleic acid. Influence of bacteriophage T2 on the synthetic pathway in host cells. *Proceedings of the National Academy of Sciences of the United States of America*, 45(6):772, 1959.

Theo F J Kraus, Daniel Globisch, Mirko Wagner, Sabina Eigenbrod, David Widmann, Martin Münzel, Markus Müller, Toni Pfaffeneder, Benjamin Hackner, Wolfgang Feiden, Ulrich Schüller, Thomas Carell, and Hans A Kretzschmar. Low values of 5-hydroxymethylcytosine (5hmC), the "sixth base," are associated with anaplasia in human brain tumors. *International Journal of Cancer*, 131(7):1577–1590, March 2012.

B Krawczyk, K Rudnicka, and K Fabianowska-Majewska. The effects of nucleoside analogues on promoter methylation of selected tumor suppressor genes in MCF-7 and MDA-MB-231 breast cancer cell lines. *Nucleosides, Nucleotides and Nucleic Acids*, 26(8-9):1043–1046, 2007.

S Kriaucionis and N Heintz. The Nuclear DNA Base 5-Hydroxymethylcytosine Is Present in Purkinje Neurons and the Brain. *Science*, 324(5929):929–930, May 2009.

Nevan J Krogan, Minkyu Kim, Seong Hoon Ahn, Guoqing Zhong, Michael S Kobor, Gerard Cagney, Andrew Emili, Ali Shilatifard, Stephen Buratowski, and Jack F Greenblatt. RNA polymerase II elongation factors of Saccharomyces cerevisiae: a targeted proteomics approach. *Molecular and Cellular Biology*, 22(20):6979–6992, October 2002.

Nevan J Krogan, Jim Dover, Adam Wood, Jessica Schneider, Jonathan Heidt, Marry Ann Boateng, Kimberly Dean, Owen W Ryan, Ashkan Golshani, Mark Johnston, Jack F Greenblatt, and Ali Shilatifard. The Paf1 Complex Is Required for Histone H3 Methylation by COMPASS and Dot1p: Linking Transcriptional Elongation to Histone Methylation. *Molecular Cell*, 11(3):721–729, January 2003.

Evert Kroon, Jana Krosl, Unnur Thorsteinsdottir, Soheyl Baban, Arthur M Buchberg, and Guy Sauvageau. Hoxa9 transforms primary bone marrow cells through specific collaboration with Meis1a but not Pbx1b. *The EMBO Journal*, 17(13):3714–3725, 1998.

Yotaro Kudo, Keisuke Tateishi, Keisuke Yamamoto, Shinzo Yamamoto, Yoshinari Asaoka, Hideaki Ijichi, Genta Nagae, Haruhiko Yoshida, Hiroyuki Aburatani, and Kazuhiko Koike. Loss of 5-hydroxymethylcytosine is accompanied with malignant cellular transformation. *Cancer Science*, 103 (4):670–676, April 2012.

K Kühn, W M Bertling, and F Emmrich. Cloning of a functional cDNA for human cytidine deaminase (CDD) and its use as a marker of monocyte/macrophage differentiation. *Biochemical and biophysical research communications*, 190(1):1–7, January 1993.

R D Ladner, D E McNulty, S A Carr, G D Roberts, and S J Caradonna. Characterization of distinct nuclear and mitochondrial forms of human deoxyuridine triphosphate nucleotidohydrolase. *The Journal of biological chemistry*, 271(13):7745–7751, March 1996.

U Lagerkvist, P Reichard, B Carlsson, and J Grabosz. The biogenesis of pentosenucleic acid pyrimidines in the Ehrlich ascites tumor. *Cancer Research*, 15(3):164–169, March 1955.

J Laliberté and R L Momparler. Human cytidine deaminase: purification of enzyme, cloning, and expression of its complementary DNA. *Cancer Research*, 54(20):5401–5407, October 1994.

Saskia M C Langemeijer, Roland P Kuiper, Marieke Berends, Ruth Knops, Mariam G Aslanyan, Marion Massop, Ellen Stevens-Linders, Patricia van Hoogen, Ad Geurts van Kessel, Reinier A P Raymakers, Eveline J Kamping, Gregor E Verhoef, Estelle Verburgh, Anne Hagemeijer, Peter Vandenberghe, Theo de Witte, Bert A van der Reijden, and Joop H Jansen. Acquired mutations in TET2 are common in myelodysplastic syndromes. *Nature Genetics*, 41(7):838–842, May 2009.

Louise Laurent, Eleanor Wong, Guoliang Li, Tien Huynh, Aristotelis Tsirigos, Chin Thing Ong, Hwee Meng Low, Ken Wing Kin Sung, Isidore Rigoutsos, Jeanne Loring, and Chia-Lin Wei. Dynamic changes in the human methylome during differentiation. *Genome Research*, 20(3):320–331, March 2010.

H J Lawrence, C D Helgason, G Sauvageau, S Fong, D J Izon, R K Humphries, and C Largman. Mice bearing a targeted interruption of the homeobox gene HOXA9 have defects in myeloid, erythroid, and lymphoid hematopoiesis. *Blood*, 89(6):1922–1930, March 1997.

J H Lee, K S Voo, and D G Skalnik. Identification and characterization of the DNA binding domain of CpG-binding protein. *The Journal of biological chemistry*, 276(48):44669–44676, November 2001.

Jung-Shin Lee, Abhijit Shukla, Jessica Schneider, Selene K Swanson, Michael P Washburn, Laurence Florens, Sukesh R Bhaumik, and Ali Shilatifard. Histone Crosstalk between H2B Monoubiquitination and H3 Methylation Mediated by COMPASS. *Cell*, 131(6):1084–1096, December 2007.

L S Lee and Y Cheng. Human thymidylate kinase. Purification, characterization, and kinetic behavior of the thymidylate kinase derived from chronic myelocytic leukemia. *The Journal of biological chemistry*, 252(16):5686–5691, August 1977.

Y P Lee. 5'-Adenylic acid deaminase. III. Properties and kinetic studies. *The Journal of biological chemistry*, 227(2):999–1007, August 1957.

Maryse Lemaire, Louise F Momparler, Noël J M Raynal, Mark L Bernstein, and Richard L Momparler. Inhibition of cytidine deaminase by zebularine enhances the antineoplastic action of 5-aza-2-deoxycytidine. *Cancer Chemotherapy and Pharmacology*, 63(3):411–416, April 2008.

Heinrich Leonhardt, Andrea W Page, Heinz-Ulrich Weier, and Timothy H Bestor. A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell*, 71(5):865–873, November 1992.

Eric Letouzé, Cosimo Martinelli, Céline Loriot, Nelly Burnichon, Nasséra Abermil, Chris Ottolenghi, Maxime Janin, Mélanie Menara, An Thach Nguyen, Paule Benit, Alexandre Buffet, Charles Marcaillou, Jérôme Bertherat, Laurence Amar, Pierre Rustin, Aurélien De Reyniès, Anne-Paule Gimenez-Roqueplo, and Judith Favier. SDH Mutations Establish a Hypermethylator Phenotype in Paraganglioma. *Cancer cell*, 23(6):739–752, June 2013.

P A Levene. On the biochemistry of nucleic acids. 2. *Journal of the American Chemical Society*, 32(2): 231–240, 1910.

P A Levene and W A Jacobs. Über die Hefe-Nucleinsäure. *Berichte der deutschen chemischen Gesellschaft*, 42(2):2474–2478, April 1909.

Timothy J Ley, Li Ding, Matthew J Walter, Michael D McLellan, Tamara Lamprecht, David E Larson, Cyriac Kandoth, Jacqueline E Payton, Jack Baty, John Welch, Christopher C Harris, Cheryl F Lichti, R Reid Townsend, Robert S Fulton, David J Dooling, Daniel C Koboldt, Heather Schmidt, Qunyuan Zhang, John R Osborne, Ling Lin, Michelle O'Laughlin, Joshua F McMichael, Kim D Delehaunty, Sean D McGrath, Lucinda A Fulton, Vincent J Magrini, Tammi L Vickery, Jasreet Hundal, Lisa L Cook, Joshua J Conyers, Gary W Swift, Jerry P Reed, Patricia A Alldredge, Todd Wylie, Jason Walker, Joelle Kalicki, Mark A Watson, Sharon Heath, William D Shannon, Nobish Varghese, Rakesh Nagarajan, Peter Westervelt, Michael H Tomasson, Daniel C Link, Timothy A Graubert, John F DiPersio, Elaine R Mardis, and Richard K Wilson. DNMT3AMutations in Acute Myeloid Leukemia. *New England Journal of Medicine*, 363(25):2424–2433, December 2010.

E Li, C Beard, and R Jaenisch. Role for DNA methylation in genomic imprinting. *Nature*, 366(6453): 362–365, November 1993.

En Li, Timothy H Bestor, and Rudolf Jaenisch. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*, 69(6):915–926, June 1992.

En Li, Masaki Okano, and Shaoping Xie. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nature Genetics*, 19(3):219–220, July 1998.

Huili Li, Katherine B Chiappinelli, Angela A Guzzetta, Hariharan Easwaran, Ray-Whay Chiu Yen, Rajita Vatapalli, Michael J Topper, Jianjun Luo, Roisin M Connolly, Nilofer S Azad, Vered Stearns, Drew M Pardoll, Nancy Davidson, Peter A Jones, Dennis J Slamon, Stephen B Baylin, Cynthia A Zahnow, and Nita Ahuja. Immune regulation by low doses of the DNA methyltransferase inhibitor 5-azacitidine in common human epithelial cancers. *Oncotarget*, 5(3):587–598, April 2015.

Joshua Li, Yuhong Ning, Warren Hedley, Brian Saunders, Yongsheng Chen, Nicole Tindill, Timo Hannay, and Shankar Subramaniam. The Molecule Pages database. *Nature*, 420(6916):716–717, December 2002.

Xiang Li, Wei Wei, Qiong-Yi Zhao, Jocelyn Widagdo, Danay Baker-Andresen, Charlotte R Flavell, Ana D'Alessio, Yi Zhang, and Timothy W Bredy. Neocortical Tet3-mediated accumulation of 5-hydroxymethylcytosine promotes rapid behavioral adaptation. *PNAS*, 111(19):7120–7125, May 2014.

Yingrui Li, Jingde Zhu, Geng Tian, Ning Li, Qibin Li, Mingzhi Ye, Hancheng Zheng, Jian Yu, Honglong Wu, Jihua Sun, Hongyu Zhang, Quan Chen, Ruibang Luo, Minfeng Chen, Yinghua He, Xin Jin, Qinghui Zhang, Chang Yu, Guangyu Zhou, Jinfeng Sun, Yebo Huang, Huisong Zheng, Hongzhi Cao, Xiaoyu Zhou, Shicheng Guo, Xueda Hu, Xin Li, Karsten Kristiansen, Lars Bolund, Jiujin Xu, Wen Wang, Huanming Yang, Jian Wang, Ruiqiang Li, Stephan Beck, Jun Wang, and Xiuqing Zhang. The DNA methylome of human peripheral blood mononuclear cells. *PLOS Biology*, 8(11):e1000533, 2010.

Christine Guo Lian, Yufei Xu, Craig Ceol, Feizhen Wu, Allison Larson, Karen Dresser, Wenqi Xu, Li Tan, Yeguang Hu, Qian Zhan, Chung-wei Lee, Di Hu, Bill Q Lian, Sonja Kleffel, Yijun Yang, James Neiswender, Abraham J Khorasani, Rui Fang, Cecilia Lezcano, Lyn M Duncan, Richard A Scolyer, John F Thompson, Hojabr Kakavand, Yariv Houvras, Leonard I Zon, Martin C Mihm, Jr., Ursula B Kaiser, Tobias Schatton, Bruce A Woda, George F Murphy, and Yujiang G Shi. Loss of 5-Hydroxymethylcytosine Is an Epigenetic Hallmark of Melanoma. *Cell*, 150(6):1135–1146, September 2012.

Christine Guo Lian, Shuyun Xu, Weimin Guo, Jian Yan, Maximilian Y M Frank, Robert Liu, Cynthia Liu, Ying Chen, George F Murphy, and Tao Chen. Decrease of 5-hydroxymethylcytosine in rat liver with subchronic exposure to genotoxic carcinogens riddelliine and aristolochic acid. *Molecular Carcinogenesis*, pages n/a–n/a, August 2014.

Kun-Tsan Lin, Richard L Momparlerm, and Georges E Rivard. High-performance liquid chromatographic analysis of chemical stability of 5-aza-2-deoxycytidine. *Journal of Pharmaceutical Sciences*, 70(11): 1228–1232, November 1981.

Jieh-Yuan Liou, Ginger E Dutschman, Wing Lam, Zaoli Jiang, and Yung-Chi Cheng. Characterization of human UMP/CMP kinase and its phosphorylation of D- and L-form deoxycytidine analogue monophosphates. *Cancer Research*, 62(6):1624–1631, March 2002.

Ryan Lister, Eran A Mukamel, Joseph R Nery, Mark Urich, Clare A Puddifoot, Nicholas D Johnson, Jacinta Lucero, Yun Huang, Andrew J Dwork, Matthew D Schultz, Miao Yu, Julian Tonti-Filippini, Holger Heyn, Shijun Hu, Joseph C Wu, Anjana Rao, Manel Esteller, Chuan He, Fatemeh G Haghighi, Terrence J Sejnowski, M Margarita Behrens, and Joseph R Ecker. Global epigenomic reconfiguration during mammalian brain development. *Science*, 341(6146):1237905, August 2013.

Chungang Liu, Limei Liu, Xuejiao Chen, Junjie Shen, Juanjuan Shan, Yanmin Xu, Zhi Yang, Lin Wu, Feng Xia, Ping Bie, Youhong Cui, Xiu-wu Bian, and Cheng Qian. Decrease of 5-Hydroxymethylcytosine Is Associated with Progression of Hepatocellular Carcinoma through Downregulation of TET1. *PLOS ONE*, 8(5):e62828, May 2013a.

Edison T Liu. Expression genomics and drug development: towards predictive pharmacology. *Briefings in functional genomics & proteomics*, 3(4):303–321, February 2005.

S Liu, J Wang, Y Su, C Guerrero, Y Zeng, D Mitra, P J Brooks, D E Fisher, H Song, and Y Wang. Quantitative assessment of Tet-induced oxidation products of 5-methylcytosine in cellular and tissue DNA. *Nucleic Acids Research*, 41(13):6421–6429, July 2013b.

Shuo Liu, Thomas L Dunwell, Gerd P Pfeifer, Jim M Dunwell, Ihsan Ullah, and Yinsheng Wang. Detection of oxidation products of 5-methyl-2'-deoxycytidine in Arabidopsis DNA. *PLOS ONE*, 8(12):e84620–e84620, January 2013c.

Daniel B Longley, D Paul Harkin, and Patrick G Johnston. 5-Fluorouracil: mechanisms of action and clinical strategies. *Nature Reviews Cancer*, 3(5):330–338, May 2003.

R B Lorsbach, J Moore, S Mathew, S C Raimondi, S T Mukatira, and J R Downing. TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23). *Leukemia*, 17(3):637–641, March 2003.

Chao Lu, Patrick S Ward, Gurpreet S Kapoor, Dan Rohle, Sevin Turcan, Omar Abdel-Wahab, Christopher R Edwards, Raya Khanin, Maria E Figueroa, Ari Melnick, Kathryn E Wellen, Donald M O'Rourke, Shelley L Berger, Timothy A Chan, Ross L Levine, Ingo K Mellinghoff, and Craig B Thompson. IDH mutation impairs histone demethylation and results in a block to cell differentiation. *Nature*, 483(7390): 474–478, February 2012.

Z H Lu, R Zhang, and R B Diasio. Purification and characterization of dihydropyrimidine dehydrogenase from human liver. *The Journal of biological chemistry*, 267(24):17102–17109, August 1992.

Eugene Lutsenko and Ashok S Bhagwat. Principal causes of hot spots for cytosine to thymine mutations at sites of cytosine methylation in growing cells. *Mutation Research/Reviews in Mutation Research*, 437 (1):11–20, July 1999.

M F Lyon. Gene action in the X-chromosome of the mouse (Mus musculus L.). *Nature*, 190:372–373, April 1961.

Crystal J MacKenzie and Toshi Shioda. COS-1 Cells as Packaging Host for Production of Lentiviruses. *Current Protocols in Cell Biology*, pages 26.7. 1–26.7. 15, 2011.

J J Mackenzie and L B Sorensen. Guanosine 5'-phosphate reductase of human erythrocytes. *Biochimica et biophysica acta*, 327(2):282–294, December 1973.

A M Mackinnon and D J Deller. Purine nucleotide biosynthesis in gastrointestinal mucosa. *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis*, 319(1):1–4, August 1973.

R Z Mahfouz, A Jankowska, Q Ebrahem, X Gu, V Visconte, A Tabarroki, P Terse, J Covey, K Chan, Y Ling, K J Engelke, M A Sekeres, R Tiu, J Maciejewski, T Radivoyevitch, and Y Saunthararajah. Increased CDA Expression/Activity in Males Contributes to Decreased Cytidine Analog Half-Life and Likely Contributes to Worse Outcomes with 5-Azacytidine or Decitabine Therapy. *Clinical Cancer Research*, 19(4):938–948, February 2013.

U Mahlknecht, C-L Dransfeld, N Bulut, M Kramer, C Thiede, G Ehninger, and M Schaich. SNP analyses in cytarabine metabolizing enzymes in AML patients and their impact on treatment response and patient survival: identification of CDA SNP C-451T as an independent prognostic parameter for survival. *Leukemia*, 23(10):1929–1932, May 2009.

Atanu Maiti and Alexander C Drohat. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *Journal of Biological Chemistry*, 286(41):35334–35338, October 2011.

F Maley and G F Maley. Nucleotide interconversions. II. Elevation of deoxycytidylate deaminase and thymidylate synthetase in regenerating rat liver. *The Journal of biological chemistry*, 235:2968–2970, October 1960.

F Maley and G F Maley. The regulatory influence of allosteric effectors on deoxycytidilate deaminase. *Curr Top Cell Regul*, 1972.

G F Maley and F Maley. Nucleotide interconversions in embryonic and neoplastic tissues. I. The conversion of deoxycytidylic acid to deoxyuridylic acid and thymidylic acid. *The Journal of biological chemistry*, 234:2975–2980, November 1959.

G F Maley, A P Lobo, and F Maley. Properties of an affinity-column-purified human deoxycytidylate deaminase. *Biochimica et biophysica acta*, 1162(1-2):161–170, March 1993.

Denis A Malyshev, Kirandeep Dhami, Thomas Lavergne, Tingjian Chen, Nan Dai, Jeremy M Foster, Ivan R Corrêa, and Floyd E Romesberg. A semi-synthetic organism with an expanded genetic alphabet. *Nature*, 509(7500):385–388, May 2014.

Patricia Maness and Antonio Orengo. Pyrimidine nucleoside monophosphate kinase from rat liver. *Biochemistry*, 14(7):1484–1489, April 1975.

Yubin Mao, Hween-Boon Toh, Zhijie Ding, Lifang Sun, Liang Hong, Chien-Shing Chen, and Xueji Wu. Differential methylation of CpG islands within the dermo1 gene promoter in several cancer cell lines. *Oncology reports*, 25(1):107–111, 2011.

C K Mathews. Deoxyribonucleotides as genetic and metabolic regulators. *The FASEB Journal*, June 2014.

Akira Matsuda and Takuma Sasaki. Antitumor activity of sugar-modified cytosine nucleosides. *Cancer Science*, 95(2):105–111, January 2004.

Rosalie Maurisse, David De Semir, Hamid Emamekhoo, Babak Bedayat, Alireza Abdolmohammadi, Hooman Parsi, and Dieter C Gruenert. Comparative transfection of DNA into primary and transformed mammalian cells from different lineages. *BMC Biotechnology*, 10(1):9, 2010.

Doris Mayer, Yutaka Natsumeda, Tadashi Ikegami, Mary Faderan, May Lui, Jahangir Emrani, Melissa Reardon, Edith Oláh, and George Weber. Expression of key enzymes of purine and pyrimidine metabolism in a hepatocyte-derived cell line at different phases of the growth cycle. *Journal of cancer research and clinical oncology*, 116(3):251–258, 1990.

R W McClard, M J Black, L R Livingstone, and M E Jones. Isolation and initial characterization of the single polypeptide that synthesizes uridine 5-monophosphate from orotate in Ehrlich ascites carcinoma. Purification by tandem affinity chromatography of uridine-5-monophosphate synthase. *Biochemistry*, 19(20):4699–4706, September 1980.

Elizabeth McFall and Boris Magasanik. The control of purine biosynthesis in cultured mammalian cells. *The Journal of biological chemistry*, 235(7):2103–2108, 1960.

N V McFerran, M Smyth, and B A Orsi. Regulation of cytidine aminohydrolase. *Biochemical Journal*, 114(1):8P–9P, August 1969.

Dattatreya Mellacheruvu, Zachary Wright, Amber L Couzens, Jean-Philippe Lambert, Nicole A St-Denis, Tuo Li, Yana V Miteva, Simon Hauri, Mihaela E Sardiu, Teck Yew Low, Vincentius A Halim, Richard D Bagshaw, Nina C Hubner, Abdallah al Hakim, Annie Bouchard, Denis Faubert, Damian Fermin, Wade H Dunham, Marilyn Goudreault, Zhen-Yuan Lin, Beatriz Gonzalez Badillo, Tony Pawson, Daniel Durocher, Benoit Coulombe, Ruedi Aebersold, Giulio Superti-Furga, Jacques Colinge, Albert J R Heck, Hyungwon Choi, Matthias Gstaiger, Shabaz Mohammed, Ileana M Cristea, Keiryn L Bennett, Mike P Washburn, Brian Raught, Rob M Ewing, Anne-Claude Gingras, and Alexey I Nesvizhskii. The CRAPome: a contaminant repository for affinity purification–mass spectrometry data. *Nature Methods*, 10(8):730–736, July 2013.

Marian Mellén, Pinar Ayata, Scott Dewell, Skirmantas Kriaucionis, and Nathaniel Heintz. MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell*, 151(7): 1417–1430, December 2012.

Cédric Mercier, Caroline Raynal, Laetitia Dahan, Adrien Ortiz, Alexandre Evrard, Charlotte Dupuis, Aurore Blesius, Muriel Duluc, Fleur Franceschini, Sarah Giacometti, Sébastien Salas, Gerard Milano, Roger Favre, Jean-François Seitz, and Joseph Ciccolini. Toxic death case in a patient undergoing gemcitabine-based chemotherapy in relation with cytidine deaminase downregulation. *Pharmacogenetics and genomics*, 17(10):841–844, October 2007.

Cedric Mercier, Raphaelle Fanciullino, Cindy Serdjebi, Joseph Ciccolini, Bertrand Pourroy, Gerard Milano, L'Houcine Ouafik, and Regis T Costello, Sr. Cytidine Deaminase Status As a Predictive Marker Of Response and Toxicity In Patients With Haematological Malignancies Treated With Azacytidine Or Cytarabine. *Blood*, 122(21):3862–3862, November 2013.

Melissa Meredith, Mary Rabaglia, and Stewart Metz. Cytosolic biosynthesis of GTP and ATP in normal rat pancreatic islets. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1266(1):16–22, April 1995.

L J Mi. Excessive base excision repair of 5-hydroxymethyluracil from DNA induces apoptosis in Chinese hamster V79 cells containing mutant p53. *Carcinogenesis*, 22(1):179–186, January 2001.

L Michaelis and M L Menten. Die kinetik der invertinwirkung. *Biochem. Z*, 49(333-369):352, 1913.

Daniela Micozzi, Francesco Martino Carpi, Stefania Pucciarelli, Valeria Polzonetti, Paolo Polidori, Santiago Vilar, Brian Williams, Stefano Costanzi, and Silvia Vincenzetti. Human cytidine deaminase: a biochemical characterization of its naturally occurring variants. *International journal of biological macromolecules*, 63:64–74, February 2014.

Catherine B Millar, Jacky Guy, Owen J Sansom, Jim Selfridge, Eilidh MacDougall, Brian Hendrich, Peter D Keightley, Stefan M Bishop, Alan R Clarke, and Adrian Bird. Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice. *Science*, 297(5580):403–405, July 2002.

R L Miller and N Ressler. The activity of serum deoxycytidylate deaminase in various diseases. *Clinica chimica acta; international journal of clinical chemistry*, 26(3):405–412, December 1969.

Thomas A Milne, Jaehoon Kim, Gang G Wang, Sonja C Stadler, Venkatesha Basrur, Sarah J Whitcomb, Zhanxin Wang, Alexander J Ruthenburg, Kojo SJ Elenitoba-Johnson, Robert G Roeder, and C David Allis. Multiple Interactions Recruit MLL1 and MLL1 Fusion Proteins to theHOXA9Locus in Leukemogenesis. *Molecular Cell*, 38(6):853–863, January 2010.

Tetsuro Miyashita, Akira Ono, and Shunji Izuta. Kinetic analysis of nucleotides incorporated opposite oxidized thymine bases on template DNA. *Nucleic acids research. Supplement (2001)*, (2):255–256, 2002.

Jin-Yao Mo, Hisaji Maki, and Mutsuo Sekiguchi. Hydrolytic elimination of a mutagenic nucleotide, 8-oxodGTP, by human 18-kilodalton protein: sanitization of nucleotide pool. *PNAS*, 89(22):11021–11025, 1992.

T Mohandas, R S Sparkes, and L J Shapiro. Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation. *Science*, 211(4480):393–396, January 1981.

S C Moldoveanu and V David. *Sample Preparation in Chromatography*. Journal of Chromatography Library. Elsevier Science, 2002.

Miriam Molina-Arcas, Laia Trigueros-Motos, F Javier Casado, and Marçal Pastor-Anglada. Physiological and pharmacological roles of nucleoside transporter proteins. *Nucleosides, Nucleotides and Nucleic Acids*, 27(6):769–778, June 2008.

E C Moore and P Reichard. ENZYMATIC SYNTHESIS OF DEOXYRIBONUCLEOTIDES. VI. THE CYTIDINE DIPHOSPHATE REDUCTASE SYSTEM FROM NOVIKOFF HEPATOMA. *The Journal of biological chemistry*, 239:3453–3456, October 1964.

Kelly Moran-Crusio, Linsey Reavie, Alan Shih, Omar Abdel-Wahab, Delphine Ndiaye-Lobry, Camille Lobry, Maria E Figueroa, Aparna Vasanthakumar, Jay Patel, Xinyang Zhao, Fabiana Perna, Suveg Pandey, Jozef Madzo, Chunxiao Song, Qing Dai, Chuan He, Sherif Ibrahim, Miloslav Beran, Jiri Zavadil, Stephen D Nimer, Ari Melnick, Lucy A Godley, Iannis Aifantis, and Ross L Levine. Tet2Loss Leads to Increased Hematopoietic Stem Cell Self-Renewal and Myeloid Transformation. *Cancer cell*, 20(1): 11–24, January 2011.

N Mourad and R E Parks. Erythrocytic nucleoside diphosphokinase. II. Isolation and kinetics. *The Journal of biological chemistry*, 241(2):271–278, January 1966.

J F Mouret, M Polverelli, F Sarrazini, and J Cadet. Ionic and radical oxidations of DNA by hydrogen peroxide. *Chemico-Biological Interactions*, 77(2):187–201, January 1991.

Cherie L Mueller and Judith A Jaehning. Ctr9, Rtf1, and Leo1 are components of the Paf1/RNA polymerase II complex. *Molecular and Cellular Biology*, 22(7):1971–1980, April 2002.

U Müller, C Bauer, M Siegl, and A Rottach. TET-mediated oxidation of methylcytosine causes TDG or NEIL glycosylase dependent gene reactivation. *Nucleic acids . . .*, 2014.

Andrew G Muntean, Jiaying Tan, Kajal Sitwala, Yongsheng Huang, Joel Bronstein, James A Connelly, Venkatesha Basrur, Kojo S J Elenitoba-Johnson, and Jay L Hess. The PAF Complex Synergizes with MLL Fusion Proteins at HOX Loci to Promote Leukemogenesis. *Cancer cell*, 17(6):609–621, June 2010.

Tomoya Muto, Goro Sashida, Motohiko Oshima, George R Wendt, Makiko Mochizuki-Kashio, Yasunobu Nagata, Masashi Sanada, Satoru Miyagi, Atsunori Saraya, Asuka Kamio, Genta Nagae, Chiaki Nakaseko, Koutaro Yokote, Kazuya Shimoda, Haruhiko Koseki, Yutaka Suzuki, Sumio Sugano, Hiroyuki Aburatani, Seishi Ogawa, and Atsushi Iwama. Concurrent loss of Ezh2 and Tet2 cooperates in the pathogenesis of myelodysplastic disorders. *Journal of Experimental Medicine*, 210(12):2627–2639, November 2013.

R Suwalski C O'Riley K Bojanowski R Yura N W Penn. The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid. *Biochemical Journal*, 126(4):781, February 1972.

Gergely N Nagy, Ibolya Leveles, and Beata G Vertessy. Preventive DNA repair by sanitizing the cellular (deoxy)nucleoside triphosphate pool. *FEBS Journal*, pages n/a–n/a, August 2014.

Yusaku Nakabeppu, Kunihiko Sakumi, Katsumi Sakamoto, Daisuke Tsuchimoto, Teruhisa Tsuzuki, and Yoshimichi Nakatsu. Mutagenesis and carcinogenesis caused by the oxidation of nucleic acids. *Biological chemistry*, 387(4):373–379, April 2006.

N Ohtani-Fujita, T Fujita, A Aoike, N E Osifchin, P D Robbins, and T Sakai. CpG methylation inactivates the promoter activity of the human retinoblastoma tumor-suppressor gene. *Oncogene*, 8(4):1063–1067, March 1993.

Masaki Okano, Daphne W Bell, Daniel A Haber, and En Li. DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. *Cell*, 99(3):247–257, October 1999.

Naoki Okashita, Yuichi Kumaki, Kuniaki Ebi, Miyuki Nishi, Yoshinori Okamoto, Megumi Nakayama, Shota Hashimoto, Tomohumi Nakamura, Kaoru Sugasawa, Nakao Kojima, Tatsuyuki Takada, Masaki Okano, and Yoshiyuki Seki. PRDM14 promotes active DNA demethylation through the ten-eleven translocation (TET)-mediated base excision repair pathway in embryonic stem cells. *Development*, 141 (2):269–280, January 2014.

Ryszard Olinski, Marek Jurgowiak, and Tomasz Zaremba. Uracil in DNA–its biological significance. *Mutation research*, 705(3):239–245, December 2010.

Ryoichi Ono, Tomohiko Taki, Takeshi Taketani, Masafumi Taniwaki, Hajime Kobayashi, and Yasuhide Hayashi. LCX, leukemia-associated protein with a CXXC domain, is fused to MLL in acute myeloid leukemia with trilineage dysplasia having t(10;11)(q22;q23). *Cancer Research*, 62(14):4075–4080, July 2002.

Steen K T Ooi, Chen Qiu, Emily Bernstein, Keqin Li, Da Jia, Zhe Yang, Hediye Erdjument-Bromage, Paul Tempst, Shau-Ping Lin, C David Allis, Xiaodong Cheng, and Timothy H Bestor. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature*, 448(7154):714–717, August 2007.

Stela S Palii, Beth O Van Emburgh, Umesh T Sankpal, Kevin D Brown, and Keith D Robertson. DNA methylation inhibitor 5-Aza-2-deoxycytidine induces reversible genome-wide DNA damage that is distinctly influenced by DNA methyltransferases 1 and 3B. *Molecular and Cellular Biology*, 28(2):752–771, January 2008.

C Panayiotou, N Solaroli, and A Karlsson. The many isoforms of human adenylate kinases. *... journal of biochemistry & cell biology*, 2014.

Christakis C Panayiotou, Nicola N Solaroli, Yunjian Y Xu, Magnus M Johansson, and Anna A Karlsson. The characterization of human adenylate kinases 7 and 8 demonstrates differences in kinetic parameters and structural organization among the family of adenylate kinase isoenzymes. *Audio, Transactions of the IRE Professional Group on*, 433(3):527–534, February 2011.

William A Pastor, Utz J Pape, Yun Huang, Hope R Henderson, Ryan Lister, Myunggon Ko, Erin M McLoughlin, Yevgeny Brudno, Sahasransu Mahapatra, Philipp Kapranov, Mamta Tahiliani, George Q Daley, X Shirley Liu, Joseph R Ecker, Patrice M Milos, Suneet Agarwal, and Anjana Rao. Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature*, 473(7347):394–397, May 2011.

M Pastor-Anglada, P Cano-soldado, E Errasti-murugarren, and F J Casado. SLC28 genes and concentrative nucleoside transporter (CNT) proteins. *Xenobiotica*, 38(7-8):972–994, August 2008.

K Patel, J Dickson, S Din, K Macleod, D Jodrell, and B Ramsahoye. Targeting of 5-aza-2-deoxycytidine residues by chromatin-associated DNMT1 induces proteasomal degradation of the free enzyme. *Nucleic Acids Research*, 38(13):4313–4324, July 2010.

Huadong Pei, Liang Li, Brooke L Fridley, Gregory D Jenkins, Krishna R Kalari, Wilma Lingle, Gloria Petersen, Zhenkun Lou, and Liewei Wang. FKBP51 Affects Cancer Cell Response to Chemotherapy by Negatively Regulating Akt. *Cancer cell*, 16(3):259–266, September 2009.

Audrey Petitjean, Ewy Mathe, Shunsuke Kato, Chikashi Ishioka, Sean V Tavtigian, Pierre Hainaut, and Magali Olivier. Impact of mutant p53 functional properties on TP53mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database. *Human Mutation*, 28(6): 622–629, June 2007.

Toni Pfaffeneder, Fabio Spada, Mirko Wagner, Caterina Brandmayr, Silvia K Laube, David Eisen, Matthias Truss, Jessica Steinbacher, Benjamin Hackner, Olga Kotljarova, David Schuermann, Stylianos Michalakis, Olesea Kosmatchev, Stefan Schiesser, Barbara Steigenberger, Nada Raddaoui, Gengo Kashiwazaki, Udo Müller, Cornelia G Spruijt, Michiel Vermeulen, Heinrich Leonhardt, Primo Schär, Markus Müller, and Thomas Carell. Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. *Nature Chemical Biology*, 10(7):574–581, May 2014.

T D Pfister, W C Reinhold, K Agama, S Gupta, S A Khin, R J Kinders, R E Parchment, J E Tomaszewski, J H Doroshow, and Y Pommier. Topoisomerase I levels in the NCI-60 cancer cell line panel determined by validated ELISA and microarray analysis and correlation with indenoisoquinoline sensitivity. *Molecular Cancer Therapeutics*, 8(7):1878–1884, July 2009.

Francesco M Piccolo, Hakan Bagci, Karen E Brown, David Landeira, Jorge Soza-Ried, Amelie Feytout, Dylan Mooijman, Petra Hajkova, Harry G Leitch, Takashi Tada, Skirmantas Kriaucionis, Meelad M Dawlaty, Rudolf Jaenisch, Matthias Merkenschlager, and Amanda G Fisher. Different roles for Tet1 and Tet2 proteins in reprogramming-mediated erasure of imprints induced by EGC fusion. *Molecular Cell*, 49(6):1023–1033, March 2013.

P Pinto do O. Expression of the LIM-homeobox gene LH2 generates immortalized Steel factor-dependent multipotent hematopoietic precursors. *The EMBO Journal*, 17(19):5744–5756, October 1998.

P Pinto do O. Hematopoietic progenitor/stem cells immortalized by Lhx2 generate functional hematopoietic cells in vivo. *Blood*, 99(11):3939–3946, May 2002.

S R Planck and G C Mueller. DNA chain growth in isolated HeLa nuclei. *Biochemistry*, 16(12):2778–2782, June 1977.

J Pliml and F Šorm. Synthesis of a 2-deoxy-D-ribofuranosyl-5-azacytosine. *Collection of Czechoslovak Chemical Communications*, 29(10):2576–2578, 1964.

Dmitry K Pokholok, Nancy M Hannett, and Richard A Young. Exchange of RNA polymerase II initiation and elongation factors during gene expression in vivo. *Molecular Cell*, 9(4):799–809, April 2002.

Giovanna Pontarin, Lisa Gallinaro, Paola Ferraro, Peter Reichard, and Vera Bianchi. Origins of mitochondrial thymidine triphosphate: dynamic relations to cytosolic pools. *PNAS*, 100(21):12159–12164, October 2003.

Horacio Pontis, Gertrud Degerstedt, and Peter Reichard. Uridine and deoxyuridine phosphorylases from Ehrlich ascites tumor. *Biochimica et biophysica acta*, 51(1):138–147, July 1961.

Joël R Pradines, Victor Farutin, Steve Rowley, and Vlado Dancík. Analyzing protein lists with large networks: edge-count probabilities in random graphs with given expected degrees. *Journal of computational biology : a journal of computational molecular cell biology*, 12(2):113–128, March 2005.

E Pronier, C Almire, H Mokrani, A Vasanthakumar, A Simon, B da Costa Reis Monte Mor, A Masse, J P Le Couedic, F Pendino, B Carbonne, J Larghero, J L Ravanat, N Casadevall, O A Bernard, N Droin, E Solary, L A Godley, W Vainchenker, I Plo, and F Delhommeau. Inhibition of TET2-mediated conversion of 5-methylcytosine to 5-hydroxymethylcytosine disturbs erythroid and granulomonocytic differentiation of human hematopoietic progenitors. *Blood*, 118(9):2551–2555, September 2011.

Emily L Putiri, Rochelle L Tiedemann, Joyce J Thompson, Chunsheng Liu, Thai Ho, Jeong-Hyeon Choi, and Keith D Robertson. Distinct and overlapping control of 5-methylcytosine and 5-hydroxymethylcytosine by the TET proteins in human cancer cells. *Genome Biology*, 15(6):R81, 2014.

Taichun Qin, Ryan Castoro, Samih El Ahdab, Jaroslav Jelinek, Xiaodan Wang, Jiali Si, Jingmin Shu, Rong He, Nianxiang Zhang, Woonbok Chung, Hagop M Kantarjian, and Jean-Pierre J Issa. Mechanisms of resistance to decitabine in the myelodysplastic syndrome. *PLOS ONE*, 6(8):e23372, 2011.

Chen Qiu, Ken Sawada, Xing Zhang, and Xiaodong Cheng. The PWWP domain of mammalian DNA methyltransferase Dnmt3b defines a new family of DNA-binding folds. *Nature Structural Biology*, February 2002.

Cyril Quivoron, Lucile Couronné, Véronique Della Valle, Cécile K Lopez, Isabelle Plo, Orianne Wagner-Ballon, Marcio Do Cruzeiro, Francois Delhommeau, Bertrand Arnulf, Marc-Henri Stern, Lucy Godley, Paule Opolon, Hervé Tilly, Eric Solary, Yannis Duffourd, Philippe Dessen, Hélène Merle-Beral, Florence Nguyen-Khac, Michaëla Fontenay, William Vainchenker, Christian Bastard, Thomas Mercher, and Olivier A Bernard. TET2 Inactivation Results in Pleiotropic Hematopoietic Abnormalities in Mouse and Is a Recurrent Event during Human Lymphomagenesis. *Cancer cell*, 20(1):276–276, January 2011.

Günter Raddatz, Qing Gao, Sebastian Bender, Rudolf Jaenisch, and Frank Lyko. Dnmt3a protects active chromosome domains against cancer-associated hypomethylation. *PLoS Genetics*, 8(12):e1003146, 2012.

Peter B Rahl, Charles Y Lin, Amy C Seila, Ryan A Flynn, Scott McCuine, Christopher B Burge, Phillip A Sharp, and Richard A Young. c-Myc Regulates Transcriptional Pause Release. *Cell*, 141(3):432–445, April 2010.

Ganeshkumar Rajendran, Karthik Shanmuganandam, Ameya Bendre, Dattatraya Muzumdar, Dattatreya Mujumdar, Abhay Goel, and Anjali Shiras. Epigenetic regulation of DNA methyltransferases: DNMT1 and DNMT3B in gliomas. *Journal of neuro-oncology*, 104(2):483–494, September 2011.

B H Ramsahoye, D Biniszkiewicz, F Lyko, V Clark, A P Bird, and R Jaenisch. Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *PNAS*, 97(10): 5237–5242, May 2000.

Peter Reichard. The enzymatic synthesis of ureidosuccinic acid in rat liver mitochondria. *Acta Chemica (Scand.)*, 8:795–805, 1954.

Peter Reichard. Interactions between deoxyribonucleotide and DNA synthesis. *Annual review of biochemistry*, 57(1):349–374, 1988.

Wolf Reik, Andrew Collick, Michael L Norris, Sheila C Barton, and M Azim Surani. Genomic imprinting determines methylation of parental alleles in transgenic mice. *Nature*, 328(6127):248–251, July 1987.

Daniel Renciuk, Olivier Blacque, Michaela Vorlickova, and Bernhard Spingler. Crystal structures of B-DNA dodecamer containing the epigenetic modifications 5-hydroxymethylcytosine or 5-methylcytosine. *Nucleic Acids Research*, page gkt738, 2013.

Cristina E Requena, Guiomar Pérez-Moreno, Luis M Ruiz-Pérez, Antonio E Vidal, and Dolores González-Pacanowska. The NTP pyrophosphatase DCTPP1 contributes to the homoeostasis and cleansing of the dNTP pool in human cells. *Biochemical Journal*, 459(1):171–180, March 2014.

W M Rideout, G A Coetzee, A F Olumi, and P A Jones. 5-Methylcytosine as an endogenous mutagen in the human LDL receptor and p53 genes. *Science*, 249(4974):1288–1290, September 1990.

Julien Rocquain, Nadine Carbuccia, Virginie Trouplin, Stéphane Raynaud, Anne Murati, Meyer Nezri, Zoulika Tadrist, Sylviane Olschwang, Norbert Vey, Daniel Birnbaum, Véronique Gelsi-Boyer, and Marie-Joelle Mozziconacci. Combined mutations of ASXL1, CBL, FLT3, IDH1, IDH2, JAK2, KRAS, NPM1, NRAS, RUNX1, TET2 and WT1 genes in myelodysplastic syndromes and acute myeloid leukemias. *BMC Cancer*, 10(1):401, 2010.

Sílvia Rodríguez-Mulero, Ekaitz Errasti-Murugarren, José Ballarín, Antonio Felipe, Alain Doucet, F Javier Casado, and Marçal Pastor-Anglada. Expression of concentrative nucleoside transporters SLC28 (CNT1, CNT2, and CNT3) along the rat nephron: effect of diabetes. *Kidney international*, 68(2):665–672, August 2005.

E P Rogakou, D R Pilch, A H Orr, V S Ivanova, and W M Bonner. DNA Double-stranded Breaks Induce Histone H2AX Phosphorylation on Serine 139. *Journal of Biological Chemistry*, 273(10):5858–5868, March 1998.

E P Rogakou, C Boon, C Redon, and W M Bonner. Megabase chromatin domains involved in DNA double-strand breaks in vivo. *The Journal of Cell Biology*, 146(5):905–916, September 1999.

Frederick M Rosenbloom. Inherited Disorder of Purine Metabolism. *JAMA*, 202(3):175–177, October 1967.

Andrii Rudenko, Meelad M Dawlaty, Jinsoo Seo, Albert W Cheng, Jia Meng, Thuc Le, Kym F Faull, Rudolf Jaenisch, and Li-Huei Tsai. Tet1 Is Critical for Neuronal Activity-Regulated Gene Expression and Memory Extinction. *Neuron*, 79(6):1109–1122, September 2013.

W M S Russell and R L Burch. *The principles of humane experimental technique*. Methuen, 1959.

David A Russler-Germain, David H Spencer, Margaret A Young, Tamara L Lamprecht, Christopher A Miller, Robert Fulton, Matthew R Meyer, Petra Erdmann-Gilmore, R Reid Townsend, Richard K Wilson, and Timothy J Ley. The R882H DNMT3A mutation associated with AML dominantly inhibits wild-type DNMT3A by blocking its ability to form active tetramers. *Cancer cell*, 25(4):442–454, April 2014.

Yoshio Sakamoto, Riko Kitazawa, Sakan Maeda, and Sohei Kitazawa. Methylation of CpG loci in 5-flanking region alters steady-state expression of adenomatous polyposis coli gene in colon cancer cell lines. *Journal of cellular biochemistry*, 80(3):415–423, 2001.

Raphael Salomon and Alvin M Kaye. Methylation of mouse DNA in vivo. *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis*, 204(2):340–351, April 1970.

H Sano and R Sager. Tissue specificity and clustering of methylated cystosines in bovine satellite I DNA. *PNAS*, 79(11):3584–3588, May 1982.

D V Santi, A Norment, and C E Garrett. Covalent bond formation between a DNA-cytosine methyltransferase and DNA containing 5-azacytosine. *PNAS*, 81(22):6993–6997, November 1984.

R G Sargent and C K Mathews. Imbalanced deoxyribonucleoside triphosphate pools and spontaneous mutation rates determined during dCMP deaminase-defective bacteriophage T4 infections. *The Journal of biological chemistry*, 262(12):5546–5553, April 1987.

Masato Sasaki, Christiane B Knobbe, Joshua C Munger, Evan F Lind, Dirk Brenner, Anne Brüstle, Isaac S Harris, Roxanne Holmes, Andrew Wakeham, Jillian Haight, Annick You-Ten, Wanda Y Li, Stefanie Schalm, Shinsan M Su, Carl Virtanen, Guido Reifenberger, Pamela S Ohashi, Dwayne L Barber, Maria E Figueroa, Ari Melnick, Juan-Carlos Zúñiga-Pflücker, and Tak W Mak. IDH1(R132H) mutation increases murine haematopoietic progenitors and alters epigenetics. *Nature*, 488(7413):656–659, August 2012.

Serge Saxonov, Paul Berg, and Douglas L Brutlag. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *PNAS*, 103(5):1412–1417, January 2006.

Uwe Scherf, Douglas T Ross, Mark Waltham, Lawrence H Smith, Jae K Lee, Lorraine Tanabe, Kurt W Kohn, William C Reinhold, Timothy G Myers, Darren T Andrews, Dominic A Scudiero, Michael B Eisen, Edward A Sausville, Yves Pommier, David Botstein, Patrick O Brown, and John N Weinstein. A gene expression database for the molecular pharmacology of cancer. *Nature Genetics*, 24(3):236–244, March 2000.

Lothar Schermelleh, Fabio Spada, Hariharan P Easwaran, Kourosh Zolghadr, Jean B Margot, M Cristina Cardoso, and Heinrich Leonhardt. Trapped in action: direct visualization of DNA methyltransferase activity in living cells. *Nature Methods*, 2(10):751–756, September 2005.

Stefan Schiesser, Benjamin Hackner, Toni Pfaffeneder, Markus Müller, Christian Hagemeier, Matthias Truss, and Thomas Carell. Mechanism and Stem-Cell Activity of 5-Carboxycytosine Decarboxylation Determined by Isotope Tracing. *Angewandte Chemie International Edition*, 51(26):6516–6520, May 2012.

Stefan Schiesser, Toni Pfaffeneder, Keyarash Sadeghian, Benjamin Hackner, Barbara Steigenberger, Arne S Schröder, Jessica Steinbacher, Gengo Kashiwazaki, Georg Höfner, Klaus T Wanner, Christian Ochsenfeld, and Thomas Carell. Deamination, Oxidation, and C–C Bond Cleavage Reactivity of 5-Hydroxymethylcytosine, 5-Formylcytosine, and 5-Carboxycytosine. *Journal of the American Chemical Society*, 135(39):14593–14599, October 2013.

D Schild, A J Brake, M C Kiefer, D Young, and P J Barr. Cloning of three human multifunctional de novo purine biosynthetic genes by functional complementation of yeast mutations. *PNAS*, 87(8):2916–2920, April 1990.

Caroline A Schneider, Wayne S Rasband, and Kevin W Eliceiri. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, 9(7):671–675, June 2012.

Somasekar Seshagiri, Eric W Stawiski, Steffen Durinck, Zora Modrusan, Elaine E Storm, Caitlin B Conboy, Subhra Chaudhuri, Yinghui Guan, Vasantharajan Janakiraman, Bijay S Jaiswal, Joseph Guillory, Connie Ha, Gerrit J P Dijkgraaf, Jeremy Stinson, Florian Gnad, Melanie A Huntley, Jeremiah D Degenhardt, Peter M Haverty, Richard Bourgon, Weiru Wang, Hartmut Koeppen, Robert Gentleman, Timothy K Starr, Zemin Zhang, David A Largaespada, Thomas D Wu, and Frederic J de Sauvage. Recurrent R-spondin fusions in colon cancer. *Nature*, 488(7413):660–664, August 2012.

Lanlan Shen, Yutaka Kondo, Saira Ahmed, Yanis Boumber, Kazuo Konishi, Yi Guo, Xinli Chen, Jill N Vilaythong, and Jean-Pierre J Issa. Drug Sensitivity Prediction by CpG Island Methylation Profile in the NCI-60 Cancer Cell Line Panel. *Cancer Research*, 2007.

Li Shen, Hao Wu, Dinh Diep, Shinpei Yamaguchi, Ana C D'Alessio, Ho-Lim Fung, Kun Zhang, and Yi Zhang. Genome-wide Analysis Reveals TET- and TDG-Dependent 5-Methylcytosine Oxidation Dynamics. *Cell*, 153(3):692–706, April 2013.

Bertrum Shied, Parithychery R Srinivasan, and Ernest Borek. Deoxyribonucleic acid methylase of mammalian tissues. *Biochemistry*, 7(1):280–285, January 1968.

Robert Shoemaker, Monks A, Alley MC, Scudiero DA, Fine DL, McLemore TL, Abbott BJ, Paull KD, Mayo JG, and Boyd MR. Development of human tumor cell line panels for use in disease-oriented drug screening. *Progress in clinical and biological research*, 276:265–286, January 1988.

Rajna Simic, Derek L Lindstrom, Hien G Tran, Kelli L Roinick, Patrick J Costa, Alexander D Johnson, Grant A Hartzog, and Karen M Arndt. Chromatin remodeling protein Chd1 interacts with transcription elongation factors and localizes to transcribed genes. *EMBO Journal*, 22(8):1846–1856, April 2003.

M W MW Simmen, S S Leitgeb, J J Charlton, S J SJ Jones, B R BR Harris, V H VH Clark, and A A Bird. Nonmethylated transposable elements and methylated genes in a chordate genome. *Science*, 283 (5405):1164–1167, February 1999.

R L Sinsheimer. The action of pancreatic deoxyribonuclease. II. Isomeric dinucleotides. *The Journal of biological chemistry*, 215(2):579–583, July 1955.

Bence Sipos, Simone Moser, Holger Kalthoff, Virag Torok, Matthias Lohr, and Gunter Kloppel. A comprehensive characterization of pancreatic ductal carcinoma cell lines: towards the establishment of an in vitro research platform. *Virchows Archiv : an international journal of pathology*, 442(5):444–452, May 2003.

Ola Soderberg, Mats Gullberg, Malin Jarvius, Karin Ridderstraale, Karl-Johan Leuchowius, Jonas Jarvius, Kenneth Wester, Per Hydbring, Fuad Bahram, Lars-Gunnar Larsson, and Ulf Landegren. Direct observation of individual endogenous protein complexes in situ by proximity ligation. *Nature Methods*, 3 (12):995–1000, December 2006.

C Soler, J García-Manteiga, R Valdés, J Xaus, M Comalada, F J Casado, M Pastor-Anglada, A Celada, and A Felipe. Macrophages require different nucleoside transport systems for proliferation and activation. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, 15(11):1979–1988, September 2001.

Chun-Xiao Song, Keith E Szulwach, Qing Dai, Ye Fu, Shi-Qing Mao, Li Lin, Craig Street, Yujing Li, Mickael Poidevin, Hao Wu, Juan Gao, Peng Liu, Lin Li, Guo-Liang Xu, Peng Jin, and Chuan He. Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell*, 153(3):678–691, April 2013a.

J Song, M Teplova, S Ishibe-Murakami, and D J Patel. Structure-Based Mechanistic Insights into DNMT1-Mediated Maintenance DNA Methylation. *Science*, 335(6069):709–712, February 2012.

Su Jung Song, Keisuke Ito, Ugo Ala, Lev Kats, Kaitlyn Webster, Su Ming Sun, Mojca Jongen-Lavrencic, Katia Manova-Todorova, Julie Teruya-Feldstein, David E Avigan, Ruud Delwel, and Pier Paolo Pandolfi. The oncogenic microRNA miR-22 targets the TET2 tumor suppressor to promote hematopoietic stem cell self-renewal and transformation. *Cell Stem Cell*, 13(1):87–101, July 2013b.

Su Jung Song, Laura Poliseno, Min Sup Song, Ugo Ala, Kaitlyn Webster, Christopher Ng, Gary Beringer, Nicolai J Brikbak, Xin Yuan, Lewis C Cantley, Andrea L Richardson, and Pier Paolo Pandolfi. MicroRNA-antagonism regulates breast cancer stemness and metastasis via TET-family-dependent chromatin remodeling. *Cell*, 154(2):311–324, July 2013c.

Cornelia G Spruijt, Felix Gnerlich, Arne H Smits, Toni Pfaffeneder, Pascal W T C Jansen, Christina Bauer, Martin Münzel, Mirko Wagner, Markus Müller, Fariha Khan, H Christian Eberl, Anneloes Mensinga, Arie B Brinkman, Konstantin Lephikov, Udo Müller, Jörn Walter, Rolf Boelens, Hugo van Ingen, Heinrich Leonhardt, Thomas Carell, and Michiel Vermeulen. Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell*, 152(5):1146–1159, February 2013.

Sharon L Squazzo, Patrick J Costa, Derek L Lindstrom, Kathryn E Kumer, Rajna Simic, Jennifer L Jennings, Andrew J Link, Karen M Arndt, and Grant A Hartzog. The Paf1 complex physically and functionally associates with transcription elongation factors in vivo. *EMBO Journal*, 21(7):1764–1774, April 2002.

Michael B Stadler, Rabih Murr, Lukas Burger, Robert Ivanek, Florian Lienert, Anne Schöler, Erik van Nimwegen, Christiane Wirbelauer, Edward J Oakeley, Dimos Gaidatzis, Vijay K Tiwari, and Dirk Schübeler. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature*, 480 (7378):490–495, December 2011.

M J Steenman, S Rainier, C J Dobry, P Grundy, I L Horon, and A P Feinberg. Loss of imprinting of IGF2 is linked to reduced expression and abnormal methylation of H19 in Wilms' tumour. *Nature Genetics*, 7(3):433–439, June 1994.

C D Steuart and P J Burke. Cytidine deaminase and the development of resistance to arabinosyl cytosine. *Nature: New biology*, 233(38):109–110, September 1971.

R L Stone, H Zalkin, and J E Dixon. Expression, purification, and kinetic characterization of recombinant human adenylosuccinate lyase. *The Journal of biological chemistry*, 268(26):19710–19716, September 1993.

Carlo Stresemann and Frank Lyko. Modes of action of the DNA methyltransferase inhibitors azacytidine and decitabine. *International Journal of Cancer*, 123(1):8–13, 2008.

H Stroud, S Feng, S Morey Kinney, and S Pradhan. 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. *Genome ...*, 2011.

A I Su, T Wiltshire, S Batalov, H Lapp, K A Ching, D Block, J Zhang, R Soden, M Hayakawa, G Kreiman, M P Cooke, J R Walker, and J B Hogenesch. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proceedings of the National Academy of Sciences*, 101(16):6062–6067, April 2004.

Yunpeng Su, Janelle Ortiz, Shihui Liu, Thomas H Bugge, Ravibhushan Singh, Stephen H Leppla, and Arthur E Frankel. Systematic urokinase-activated anthrax toxin therapy produces regressions of subcutaneous human non-small cell lung tumor in athymic nude mice. *Cancer Research*, 67(7):3329–3336, April 2007.

Miao Sun, Chun-Xiao Song, Hao Huang, Casey A Frankenberger, Devipriya Sankarasharma, Suzana Gomes, Ping Chen, Jianjun Chen, Kiran K Chada, Chuan He, and Marsha R Rosner. HMGA2/TET1/HOXA9 signaling pathway regulates breast cancer growth and metastasis. *Proceedings of the National Academy of Sciences*, 110(24):9920–9925, June 2013.

M Suzuki. RNA Interference-Mediated Knockdown of DNA Methyltransferase 1 Leads to Promoter Demethylation and Gene Re-Expression in Human Lung and Breast Cancer Cells. *Cancer Research*, 64 (9):3137–3143, May 2004.

Keith E Szulwach, Xuekun Li, Yujing Li, Chun-Xiao Song, Hao Wu, Qing Dai, Hasan Irier, Anup K Upadhyay, Marla Gearing, Allan I Levey, Aparna Vasanthakumar, Lucy A Godley, Qiang Chang, Xiaodong Cheng, Chuan He, and Peng Jin. 5-hmC–mediated epigenetic dynamics during postnatal neurodevelopment and aging. *Nature Neuroscience*, 14(12):1607–1616, October 2011.

M Tahiliani, K P Koh, Y Shen, W A Pastor, H Bandukwala, Y Brudno, S Agarwal, L M Iyer, D R Liu, L Aravind, and A Rao. Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science*, 324(5929):930–935, May 2009.

M Tatibana and K Ito. Control of pyrimidine biosynthesis in mammalian tissues. I. Partial purification and characterization of glutamine-utilizing carbamyl phosphate synthetase of mouse spleen and its tissue distribution. *The Journal of biological chemistry*, 244(19):5403–5413, October 1969.

S M Taylor and P A Jones. Mechanism of action of eukaryotic DNA methyltransferase. Use of 5-azacytosine-containing DNA. *Journal of Molecular Biology*, 162(3):679–692, December 1982.

Shirley M Taylor and Peter A Jones. Multiple new phenotypes induced in 10T1/2 and 3T3 cells treated with 5-azacytidine. *Cell*, 17(4):771–779, August 1979.

Anatoli Tchigvintsev, Dmitri Tchigvintsev, Robert Flick, Ana Popovic, Aiping Dong, Xiaohui Xu, Greg Brown, Wenyun Lu, Hong Wu, Hong Cui, Ludmila Dombrowski, Jeong Chan Joo, Natalia Beloglazova, Jinrong Min, Alexei Savchenko, Amy A Caudy, Joshua D Rabinowitz, Alexey G Murzin, and Alexander F Yakunin. Biochemical and Structural Studies of Conserved Maf Proteins Revealed Nucleotide Pyrophosphatases with a Preference for Modified Nucleotides. *Chemistry & biology*, 20(11):1386–1398, November 2013.

A Tefferi. Novel mutations and their functional and clinical relevance in myeloproliferative neoplasms JAK2, MPL, TET2, ASXL1, CBL, IDH and IKZF1 2010. *Leukemia*, 24(6):1128–1138, April 2010.

M Toyota, N Ahuja, M Ohe-Toyota, J G Herman, S B Baylin, and J P Issa. CpG island methylator phenotype in colorectal cancer. *PNAS*, 96(15):8681–8686, July 1999.

Thomas Treiber, Elizabeth M Mandel, Sebastian Pott, Ildiko György, Sonja Firner, Edison T Liu, and Rudolf Grosschedl. Early B Cell Factor 1 Regulates B Cell Gene Networks by Activation, Repression, and Transcription- Independent Poising of Chromatin. *Immunity*, 32(5):714–725, May 2010.

Shigeru Tsuchiya, Michiko Yamabe, Yoshiko Yamaguchi, Yasuko Kobayashi, Tasuke Konno, and Keiya Tada. Establishment and characterization of a human acute monocytic leukemia cell line (THP-1). *International Journal of Cancer*, 26(2):171–176, August 1980.

Teruhisa Tsuzuki, Akinori Egashira, Hisato Igarashi, Tomoo Iwakuma, Yoko Nakatsuru, Yohei Tominaga, Hisaya Kawate, Kazuki Nakao, Kenji Nakamura, Fumio Ide, Shinobu Kura, Yusaku Nakabeppu, Motoya Katsuki, Takatoshi Ishikawa, and Mutsuo Sekiguchi. Spontaneous tumorigenesis in mice defective in the MTH1 gene encoding 8-oxo-dGTPase. *Proceedings of the National Academy of Sciences*, 98(20): 11456–11461, September 2001.

Sevin Turcan, Daniel Rohle, Anuj Goenka, Logan A Walsh, Fang Fang, Emrullah Yilmaz, Carl Campos, Armida W M Fabius, Chao Lu, Patrick S Ward, Craig B Thompson, Andrew Kaufman, Olga Guryanova, Ross Levine, Adriana Heguy, Agnes Viale, Luc G T Morris, Jason T Huse, Ingo K Mellinghoff, and Timothy A Chan. IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. *Nature*, 483(7390):479–483, February 2012.

M K Turner, R Abrams, and I Lieberman. Levels of ribonucleotide reductase activity during the division cycle of the L cell. *The Journal of biological chemistry*, 243(13):3725–3728, July 1968.

Mathias Uhlén, Erik Björling, Charlotta Agaton, Cristina Al-Khalili Szigyarto, Bahram Amini, Elisabet Andersen, Ann-Catrin Andersson, Pia Angelidou, Anna Asplund, Caroline Asplund, Lisa Berglund, Kristina Bergström, Harry Brumer, Dijana Cerjan, Marica Ekström, Adila Elobeid, Cecilia Eriksson, Linn Fagerberg, Ronny Falk, Jenny Fall, Mattias Forsberg, Marcus Gry Björklund, Kristoffer Gumbel, Asif Halimi, Inga Hallin, Carl Hamsten, Marianne Hansson, My Hedhammar, Görel Hercules, Caroline Kampf, Karin Larsson, Mats Lindskog, Wald Lodewyckx, Jan Lund, Joakim Lundeberg, Kristina Magnusson, Erik Malm, Peter Nilsson, Jenny Ödling, Per Oksvold, Ingmarie Olsson, Emma Öster, Jenny Ottosson, Linda Paavilainen, Anja Persson, Rebecca Rimini, Johan Rockberg, Marcus Runeson, Åsa Sivertsson, Anna Sköllermo, Johanna Steen, Maria Stenvall, Fredrik Sterky, Sara Strömberg, Mårten Sundberg, Hanna Tegel, Samuel Tourle, Eva Wahlund, Annelie Waldén, Jinghong Wan, Henrik Wernérus, Joakim Westberg, Kenneth Wester, Ulla Wrethagen, Lan Lan Xu, Sophia Hober, and Fredrik Pontén. A human protein atlas for normal and cancer tissues based on antibody proteomics. *Molecular & cellular proteomics : MCP*, 4(12):1920–1932, December 2005.

Andreas Untergasser, Ioana Cutcutache, Triinu Koressaar, Jian Ye, Brant C Faircloth, Maido Remm, and Steven G Rozen. Primer3-new capabilities and interfaces. *Nucleic acids ...*, 2012.

Stuart Linn Werner Arber Urs Kühnlein. HOST SPECIFICITY OF DNA PRODUCED BY Escherichia coli, XI. In vitro MODIFICATION OF PHAGE FD REPLICATIVE FORM. *Proceedings of the National Academy of Sciences of the United States of America*, 63(2):556, June 1969.

R Valdés, M A Ortega, F J Casado, A Felipe, A Gil, A Sánchez-Pozo, and M Pastor-Anglada. Nutritional regulation of nucleoside transporter expression in rat small intestine. *Gastroenterology*, 119(6):1623–1630, December 2000.

L H T van der Ploeg and R A Flavell. DNA methylation in the human $\gamma\delta\beta$-globin locus in erythroid and nonerythroid tissues. *Cell*, 19(4):947–958, April 1980.

M B Van der Weyden and W N Kelly. Human adenylosuccinate synthetase. Partial purification, kinetic and regulatory properties of the enzyme from placenta. *The Journal of biological chemistry*, 249(22): 7282–7289, November 1974.

A R Van Rompay, M Johansson, and A Karlsson. Identification of a novel human adenylate kinase. cDNA cloning, expression analysis, chromosome localization and characterization of the recombinant protein. *European journal of biochemistry / FEBS*, 261(2):509–517, April 1999a.

A R Van Rompay, M Johansson, and A Karlsson. Phosphorylation of nucleosides and nucleoside analogs by mammalian nucleoside monophosphate kinases. *Pharmacology & Therapeutics*, 2000.

A R Van Rompay, A Norda, K Lindén, M Johansson, and A Karlsson. Phosphorylation of uridine and cytidine nucleoside analogs by two human uridine-cytidine kinases. *Molecular pharmacology*, 59(5): 1181–1186, May 2001.

An R Van Rompay, Magnus Johansson, and Anna Karlsson. Phosphorylation of Deoxycytidine Analog Monophosphates by UMP-CMP Kinase: Molecular Characterization of the Human Enzyme. *Molecular pharmacology*, 1999b.

Pietro Vella, Andrea Scelfo, Sriganesh Jammula, Fulvio Chiacchiera, Kristine Williams, Alessandro Cuomo, Alessandra Roberto, Jesper Christensen, Tiziana Bonaldi, Kristian Helin, and Diego Pasini. Tet Proteins Connect the O-Linked N-acetylglucosamine Transferase Ogt to Chromatin in Embryonic Stem Cells. *Molecular Cell*, 49(4):645–656, February 2013.

Juhani A Vilpo and Leena A Vilpo. Biochemical mechanisms by which reutilization of DNA 5-methylcytosine is prevented in human cells. *Mutation Research/DNAging*, 256(1):29–35, January 1991.

Juhani A Vilpo and Leena M Vilpo. Metabolism, Incorporation into DNA, and Interactions with 1-β-d-Arabinofuranosylcytosine of 5-Hydroxymethyl-2-deoxyuridine in Human Promyelocytic Leukemia Cells (HL-60). *Cancer Research*, 1988.

Juhani A Vilpo and Leena M Vilpo. Nucleoside monophosphate kinase may be the key enzyme preventing salvage of DNA 5-methylcytosine. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 286(2):217–220, April 1993.

S Vincenzetti. Recombinant Human Cytidine Deaminase: Expression, Purification, and Characterization. *Protein Expression and Purification*, 8(2):247–253, September 1996.

S Vincenzetti, P L Mariani, N Cammertoni, V Polzonetti, P Natalini, B Quadrini, R Volpini, and A Vita. Isoenzymatic forms of human cytidine deaminase. *Protein Engineering ...*, 2004.

Torkild Visnes, Berit Doseth, Henrik Sahlin Pettersen, Lars Hagen, Mirta M L Sousa, Mansour Akbari, Marit Otterlei, Bodil Kavli, Geir Slupphaug, and Hans E Krokan. Uracil in DNA and its processing by different DNA glycosylases. *... of the Royal ...*, 2009.

Daniel D Von Hoff, Thomas Ervin, Francis P Arena, E Gabriela Chiorean, Jeffrey Infante, Malcolm Moore, Thomas Seay, Sergei A Tjulandin, Wen Wee Ma, Mansoor N Saleh, Marion Harris, Michele Reni, Scot Dowden, Daniel Laheru, Nathan Bahary, Ramesh K Ramanathan, Josep Tabernero, Manuel Hidalgo, David Goldstein, Eric Van Cutsem, Xinyu Wei, Jose Iglesias, and Markus F Renschler. Increased survival in pancreatic cancer with nab-paclitaxel plus gemcitabine. *New England Journal of Medicine*, 369(18):1691–1703, October 2013.

A Wagner. Energy Constraints on the Evolution of Gene Expression. *Molecular Biology and Evolution*, 22 (6):1365–1374, March 2005.

S Wakita, H Yamaguchi, I Omori, K Terada, T Ueda, E Manabe, S Kurosawa, S Iida, T Ibaraki, Y Sato, T Todoroki, T Hirakawa, T Ryotokuji, K Arai, T Kitano, Y Mitamura, F Kosaka, K Dan, and K Inokuchi. Mutations of the epigenetics-modifying gene (DNMT3a, TET2, IDH1/2) at diagnosis may induce FLT3-ITD at relapse in de novo acute myeloid leukemia. *Leukemia*, 27(5):1044–1052, November 2012.

C P Walsh, J R Chaillet, and T H Bestor. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nature Genetics*, 20(2):116–117, October 1998.

L Wang and S Eriksson. Tissue Specific Distribution of Pyrimidine Deoxynucleoside Salvage Enzymes Shed Light on the Mechanism of Mitochondrial DNA Depletion. *Nucleosides, Nucleotides and Nucleic Acids*, 29(4-6):400–403, June 2010.

Lei Wang, Yong Pan, and Jia Le Dai. Evidence of MKK4 pro-oncogenic activity in breast and pancreatic tumors. *Oncogene*, 23(35):5978–5985, June 2004.

Yu Wang and Yi Zhang. Regulation of TET protein stability by calpains. *CellReports*, 6(2):278–284, January 2014.

J D Watson and F H Crick. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, 171(4356):737–738, April 1953a.

J D Watson and F H Crick. Genetical implications of the structure of deoxyribonucleic acid. *Nature*, 171 (4361):964–967, May 1953b.

T A Weinert and L H Hartwell. The RAD9 gene controls the cell cycle response to DNA damage in Saccharomyces cerevisiae. *Science*, 241(4863):317–322, July 1988.

M Wigler, D Levy, and M Perucho. The somatic replication of DNA methylation. *Cell*, 24(1):33–40, April 1981.

Kristine Williams, Jesper Christensen, Marianne Terndrup Pedersen, Jens V Johansen, Paul A C Cloos, Juri Rappsilber, and Kristian Helin. TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature*, 473(7347):343–348, April 2011.

H H Winkler. Rickettsial permeability. An ADP-ATP transport system. *Journal of Biological Chemistry*, 1976.

S F Wolf, D J Jolly, K D Lunnen, T Friedmann, and B R Migeon. Methylation of the hypoxanthine phosphoribosyltransferase locus on the human X chromosome: implications for X-chromosome inactivation. *PNAS*, 81(9):2806–2810, May 1984.

Edmund Wong, Kan Yang, Mari Kuraguchi, Uwe Werling, Elena Avdievich, Kunhua Fan, Melissa Fazzari, Bo Jin, Anthony M C Brown, Martin Lipkin, and Winfried Edelmann. Mbd4 inactivation increases C to T transition mutations and promotes gastrointestinal tumor formation. *PNAS*, 99(23):14937–14942, November 2002.

D M Woodcock, P J Crowther, and W P Diver. The majority of methylated deoxycytidines in human DNA are not in the CpG dinucleotide. *Biochemical and biophysical research communications*, 145(2): 888–894, June 1987.

Chunlei Wu, Camilo Orozco, Jason Boyer, Marc Leglise, James Goodale, Serge Batalov, Christopher L Hodge, James Haase, Jeff Janes, Jon W Huss, and Andrew I Su. BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biology*, 10(11):R130, 2009.

H Wu, A C D'Alessio, S Ito, Z Wang, K Cui, K Zhao, Y E Sun, and Y Zhang. Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. *Genes & Development*, 25(7):679–684, April 2011a.

Hao Wu and Yi Zhang. Reversing DNA methylation: mechanisms, genomics, and biological functions. *Cell*, 156(1-2):45–68, January 2014.

Hao Wu, Ana C D'Alessio, Shinsuke Ito, Kai Xia, Zhibin Wang, Kairong Cui, Keji Zhao, Yi Eve Sun, and Yi Zhang. Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature*, 473(7347):389–393, May 2011b.

Yongyan Wu, Zekun Guo, Ye Liu, Bo Tang, Yi Wang, Liping Yang, Juan Du, and Yong Zhang. Oct4 and the small molecule inhibitor, SC1, regulates Tet2 expression in mouse embryonic stem cells. *Molecular biology reports*, 40(4):2897–2906, April 2013.

Anton Wutz, Oskar W Smrzka, Norbert Schweifer, Karl Schellander, Erwin F Wagner, and Denise P Barlow. Imprinted expression of the Igf2r gene depends on an intronic CpG island. *Nature*, 389(6652): 745–749, October 1997.

G R Wyatt. The purine and pyrimidine composition of deoxypentose nucleic acids. *Biochemical Journal*, 48(5):584–590, May 1951.

Mengtao Xiao, Hui Yang, Wei Xu, Shenghong Ma, Huaipeng Lin, Honguang Zhu, Lixia Liu, Ying Liu, Chen Yang, Yanhui Xu, Shimin Zhao, Dan Ye, Yue Xiong, and Kun-Liang Guan. Inhibition of α-KG-dependent histone and DNA demethylases by fumarate and succinate that are accumulated in mutations of FH and SDH tumor suppressors. *Genes & Development*, 26(12):1326–1338, June 2012.

Chao Xu and Jinrong Min. Structure and function of WD40 domain proteins. *Protein & cell*, 2(3):202–214, March 2011.

Jie Xu, Yue-Ying Wang, Yu-Jun Dai, Wu Zhang, Wei-Na Zhang, Shu-Min Xiong, Zhao-Hui Gu, Kan-Kan Wang, Rong Zeng, Zhu Chen, and Sai-Juan Chen. DNMT3A Arg882 mutation drives chronic myelomonocytic leukemia through disturbing gene expression/DNA methylation in hematopoietic cells. *Proceedings of the ...*, 2014.

Y Xu, M Johansson, and A Karlsson. Human UMP-CMP Kinase 2, a Novel Nucleoside Monophosphate Kinase Localized in Mitochondria. *Journal of Biological Chemistry*, 283(3):1563–1571, January 2008.

Yufei Xu, Feizhen Wu, Li Tan, Lingchun Kong, Lijun Xiong, Jie Deng, Andrew J Barbera, Lijuan Zheng, Haikuo Zhang, Stephen Huang, Jinrong Min, Thomas Nicholson, Taiping Chen, Guoliang Xu, Yang Shi, Kun Zhang, and Yujiang Geno Shi. Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. *Molecular Cell*, 42(4):451–464, May 2011.

Yufei Xu, Chao Xu, Akiko Kato, Wolfram Tempel, Jose Garcia Abreu, Chuanbing Bian, Yeguang Hu, Di Hu, Bin Zhao, Tanja Cerovina, Jianbo Diao, Feizhen Wu, Housheng Hansen He, Qingyan Cui, Erin Clark, Chun Ma, Andrew Barbara, Gert Jan C Veenstra, Guoliang Xu, Ursula B Kaiser, X Shirley Liu, Stephen P Sugrue, Xi He, Jinrong Min, Yoichi Kato, and Yujiang Geno Shi. Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for Xenopus eye and neural development. *Cell*, 151(6):1200–1213, December 2012.

Shinpei Yamaguchi, Kwonho Hong, Rui Liu, Li Shen, Azusa Inoue, Dinh Diep, Kun Zhang, and Yi Zhang. Tet1 controls meiosis by regulating meiotic gene expression. *Nature*, 492(7429):443–447, December 2012.

Kimi Yamakoshi and Nobuyoshi Shimoda. De novo DNA methylation at the CpG island of the zebrafishno tail gene. *genesis*, 37(4):195–202, December 2003.

H Yang, Y Liu, Y Yang, P Wang, S H Kim, S Ito, and C Yang. Oncometabolite 2-Hydroxyglutarate Is a Competitive Inhibitor of α-Ketoglutarate-Dependent Dioxygenases. *Cancer cell*, 2011.

H Yang, Y Liu, F Bai, J-Y Zhang, S-H Ma, J Liu, Z-D Xu, H-G Zhu, Z-Q Ling, D Ye, K-L Guan, and Y Xiong. Tumor development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation. *Oncogene*, 32(5):663–669, March 2012.

Hui Yang, Huaipeng Lin, Haiyan Xu, Lei Zhang, Lu Cheng, Bo Wen, Jianyong Shou, Kunliang Guan, Yue Xiong, and Dan Ye. TET-catalyzed 5-methylcytosine hydroxylation is dynamically regulated by metabolites. *Nature Publishing Group*, 24(8):1017–1020, June 2014.

Jeffrey A Yoder, Neilesh S Soman, Gregory L Verdine, and Timothy H Bestor. DNA (cytosine-5)-methyltransferases in mouse cells and tissues. studies with a mechanism-based probe. *Journal of Molecular Biology*, 270(3):385–395, July 1997a.

Jeffrey A Yoder, Colum P Walsh, and Timothy H Bestor. Cytosine methylation and the ecology of intragenomic parasites. *Trends in Genetics*, 13(8):335–340, August 1997b.

Richard J Youle and Derek P Narendra. Mechanisms of mitophagy. *Nature Reviews Molecular Cell Biology*, 12(1):9–14, January 2011.

J D Young, S Y M Yao, L Sun, C E Cass, and S A Baldwin. Human equilibrative nucleoside transporter (ENT) family of nucleoside and nucleobase transporter proteins. *Xenobiotica*, 38(7-8):995–1021, August 2008.

Miao Yu, Gary C Hon, Keith E Szulwach, Chun-Xiao Song, Liang Zhang, Audrey Kim, Xuekun Li, Qing Dai, Yin Shen, Beomseok Park, Jung-Hyun Min, Peng Jin, Bing Ren, and Chuan He. Base-Resolution Analysis of 5-Hydroxymethylcytosine in the Mammalian Genome. *Cell*, 149(6):1368–1380, June 2012.

Lijie Yue, Yutaka Saikawa, Kazuhisa Ota, Motohiro Tanaka, Ryosei Nishimura, Takahiro Uehara, Hideaki Maeba, Takashi Ito, Takuma Sasaki, and Shoichi Koizumi. A functional single-nucleotide polymorphism in the human cytidine deaminase gene contributing to ara-C sensitivity. *Pharmacogenetics*, 13(1):29–38, January 2003.

Haikuo Zhang, Xin Zhang, Erin Clark, Michelle Mulcahey, Stephen Huang, and Yujiang Geno Shi. TET1 is a DNA-binding protein that modulates DNAmethylation and gene transcription via hydroxylation of5-methylcytosine. *Nature Publishing Group*, 20(12):1390–1393, November 2010a.

Run-Rui Zhang, Qing-Yan Cui, Kiyohito Murai, Yen Ching Lim, Zachary D Smith, Shengnan Jin, Peng Ye, Luis Rosa, Yew Kok Lee, Hai-Ping Wu, Wei Liu, Zhi-Mei Xu, Lu Yang, Yu-Qiang Ding, Fuchou Tang, Alexander Meissner, Chunming Ding, Yanhong Shi, and Guo-Liang Xu. Tet1 regulates adult hippocampal neurogenesis and cognition. *Cell Stem Cell*, 13(2):237–245, August 2013.

Y Zhang, R Jurkowska, S Soeroes, A Rajavelu, A Dhayalan, I Bock, P Rathert, O Brandt, R Reinhardt, W Fischle, and A Jeltsch. Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. *Nucleic Acids Research*, 38(13):4246–4253, July 2010b.

Yingqiang Zhong, Zhongsheng Xia, Juan Liu, Ying Lin, and Hui Zan. The effects of cyclooxygenase-2 gene silencing by siRNA on cell proliferation, cell apoptosis, cell cycle and tumorigenicityof Capan-2 human pancreatic cancer cells. *Oncology reports*, 27:1003–1010–, January 2012.

Bing Zhu, Subhrangsu S Mandal, Anh-Dung Pham, Yong Zheng, Hediye Erdjument-Bromage, Surinder K Batra, Paul Tempst, and Danny Reinberg. The human PAF complex coordinates transcription with events downstream of RNA synthesis. *Genes & Development*, 19(14):1668–1673, July 2005.

Daniel Zilberman, Devin Coleman-Derr, Tracy Ballinger, and Steven Henikoff. Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. *Nature*, 456(7218):125–129, September 2008.