Role of histone variant macroH2A and other chromatin regulators in genome regulation and response to drugs

Marguerite-Marie Le Pannérer

TESI DOCTORAL UPF / 2021

THESIS SUPERVISOR
Marcus Buschbeck
Department of experimental and health sciences
UNIVERSITAT POMPEU FABRA

À ceux et celles qui m’ont accompagnés

Acknowledgment

The doctoral thesis represents a long-term work, and for this reason, constitutes the thread of a slice of its author's life. Many people thus find themselves, fortuitously or not, for better or worse, between the doctoral student and her doctorate work. It is some of these people that I would like to briefly highlight in these acknowledgements. I would like to warmly thank all the people who helped me during the elaboration of my thesis and especially my supervisor, Marcus Buschbeck, for his interest and support, his advices. This work would not have been possible without the support of the Spanish Ministry of Science and innovation and the Josep Carreras leukaemia institute , which allowed me, thanks to a research allowance and various financial aids, to serenely devote myself to the elaboration of my thesis. I would like to thanks the members of the lab, and the persons whom gave me support during those intense years.

Abstract

The relevance of epigenetics is increasingly recognized in haematopoietic diseases such as myelodysplastic syndromes (MDS) and acute myeloid leukaemia (AML). Variants of histones, such as macroH2As, and chromatin remodellers, such as the CHRAC complex, are key players in these epigenetic regulations. On the other hand, non-coding DNA regions, including long interspersed nuclear elements (LINEs), also show increasing relevance, notably in cancer.

Here, we studied the relationship between LINEs and macroH2As in a hepatocarcinoma cell line. Computational analysis and experimental work led to conflicting results regarding a potential repressive role of macroH2As on LINEs. In addition, we studied the role of macroH2As as well as other chromatin regulators in the response to the epigenetic treatments Azacitidine and Decitabine in MDS/AML cell lines and a cohort of patient's samples. We were able to show that reduced expression of macroH2A1 and CHRAC complex component, BAZ1A, sensitized MDS/AML cell lines to epigenetic treatments. Furthermore, in patient sample, CHRAC complex components were less expressed in ten-eleven translocase 2 (TET2) mutants and in non-responders to Azacitidine.

Epigenetics plays an important, but very complex role in malignant diseases and more studies are necessary to increase our understanding in its regulation and function depending on the context.

Resumen

Cada vez se reconoce más la importancia de la epigenética en enfermedades hematopoyéticas como los síndromes mielodisplásicos (SMD) y la leucemia mieloide aguda (LMA). Las variantes de histonas, como las macroH2A, y los remodeladores de la cromatina, como el complejo CHRAC, son actores claves en estas regulaciones epigenéticas. Por otra parte, las regiones de ADN no codificante, incluidos los elementos nucleares intercalados largos (LINEs), también muestran una relevancia creciente, especialmente en el cáncer.

Aquí estudiamos la relación entre los LINEs y las macroH2As en una línea celular de hepatocarcinoma. El análisis computacional y el trabajo experimental condujeron a resultados contradictorios respecto a un potencial papel represivo de las macroH2As sobre los LINEs. Además, estudiamos el papel de las macroH2As, así como de otros reguladores de la cromatina, en la respuesta a los tratamientos epigenéticos Azacitidina y Decitabina en líneas celulares de SMD/LMA y en una cohorte de muestras de pacientes. Pudimos demostrar que la reducción de la expresión de macroH2A1 y del componente del complejo CHRAC, BAZ1A, sensibilizó a las líneas celulares de SMD/LMA a los tratamientos epigenéticos. Además, en la muestra de pacientes, los componentes del complejo CHRAC se expresaban menos en los mutantes de la *ten-eleven* translocasa 2 (TET2) y en los que no respondían a la Azacitidina.

La epigenética desempeña un papel importante, pero muy complejo, en las enfermedades malignas y son necesarios más estudios para aumentar nuestra comprensión de su regulación y función dependiendo del contexto.

Resumé

L'importance de l'épigénétique dans les maladies hématopoïétiques telles que les syndromes myélodysplasiques (SMD) et la leucémie myéloïde aiguë (LMA), est de plus en plus reconnue. Les variants d'histones, tels que les protéines macroH2As, et les complexes de remodelage de la chromatine, tels que le complexe CHRAC, sont des acteurs clés de ces régulations épigénétiques. En outre, les régions d'ADN non codantes, y compris les longs éléments nucléaires intercalés (LINEs), présentent également un intérêt croissant, notamment dans les cancers.

Nous étudions ici la relation entre les LINEs et les protéines macroH2As dans une lignée cellulaire d'hépatocarcinome. L'analyse computationnelle et le travail expérimental ont conduit à des résultats contradictoires concernant un rôle potentiellement répressif des protéines macroH2As sur les LINEs. De plus, nous avons étudié le rôle des protéines macroH2As, ainsi que d'autres régulateurs de la chromatine, dans la réponse aux traitements épigénétiques, Azacitidine et Décitabine, dans des lignées cellulaires SDM/LMA et dans une cohorte d'échantillons de patients. Nous avons pu démontrer que l'expression réduite du gène *MACROH2A1* et de la protéine du complexe CHRAC, *BAZ1A*, sensibilisait les lignées cellulaires MDS/AML aux traitements épigénétiques. En outre, dans l'échantillon de patients, les protéines du complexe CHRAC étaient moins exprimés chez les mutants de la translocase 2 (TET2) et les non-répondants à l'Azacitidine.

L'épigénétique joue un rôle important, mais très complexe, dans les maladies malignes et des études supplémentaires sont nécessaires pour mieux comprendre sa régulation et sa fonction en fonction du contexte.

Abbreviation

aza	azanucitidine
AML	Acute Myeloid Leukaemia
ASXL1	additional sex comb like 1
AZN	Azanucleoside
BAF	BRG1/BRM-associated factor
BAZ1A/B	Bromodomain Adjacent To Zinc Finger Domain 1A/B
BCL2L10	Bcl-2-like 10
CBP	CREB Binding Protein
CDA	Cytidine Deaminase
CDR	Common Deleted Region
CENPA	Centromere protein A
CHD	chromodomain helicase DNA-binding protein
ChIP	Chromatin Immunoprecipitation
CHRAC1	Chromatin Accessibility Complex Subunit 1
CREBBP	CREB Binding Protein, CBP
CTCF	CCCTC binding factor
ctrl	control
DNA	deoxyribonucleic acid
DNMT1	DNA Methyltransferase 1
DNMT3A	DNA Methyltransferase 3 Alpha
EDTA	Ethylenediaminetetraacetic acid
ENCODE	ENCyclopedia Of DNA Elements Project
ERCC2	ERCC excision repair 2
EZH2	Enhancer Of Zeste 2 Polycomb Repressive Complex 2 Subunit
FBOX11	F-Box Protein 11

FLYWCH1	FLYWCH-Type Zinc Finger 1
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GFP	green fluorescent protein
HDAC	Histone deacetylases
HUSH	Human Silencing Hub
HELLS	Helicase, Lymphoid Specific
HP1	Heterochromatin protein 1
HMM	hidden Markov model
HSC	haematopoietic stem cells
IDH	Isocitrate dehydrogenase
IP	immunoprecipitation
INO80	NOsitol-requiring mutant 80
ISWI	imitation SWI
L1HS	L1 Homo sapiens
LCS	leukaemic stem cell
LINEs	Long interspersed nuclear elements
LTR	long terminal repeat
MDS	Myelodysplastic syndromes
mH2A	MACROH2A
MRPL4	Mitochondrial Ribosomal Protein L4)
MRPL52	Mitochondrial Ribosomal Protein L52
MRPS26	Mitochondrial Ribosomal Protein S26
MRTO4	MRT4 Homolog, Ribosome Maturation Factor
NAA10	N-Alpha-Acetyltransferase 10
NAA15	N-alpha-acetyltransferases 15
NAD+	Nicotinamide adenine dinucleotide
NSUN3	NOP2/Sun RNA Methyltransferase 3
ORF	open reading frame
P300	E1A binding protein p300, Ep300
PBMC	peripheral blood mononuclear cell
PCR	polymerase chain reaction
PGK	phosphoglycerate kinase 1 promoter
PFA	Paraformaldehyde
PMSF	phenylmethylsulfonyl fluoride
POLE3	DNA Polymerase Epsilon subunit 3
POLR1A	RNA Polymerase I Subunit A

POLR1B	RNA Polymerase I Subunit B
POLR1C	RNA Polymerase I Subunit C
POLR1E	RNA Polymerase I Subunit E
POLR3D	RNA Polymerase III Subunit D
POLR3H	RNA Polymerase III Subunit H
PTM	post-transcriptional modification
PWWP2B	PWWP domain containing 2B
RT-qPCR	quantitative polymerase chain reaction
rDNA	ribosomal DNA
RING1	Ring Finger Protein 1
RIOX2	Ribosomal Oxygenase 2
RNA	Ribonucleic acid
RPLP0	60S acidic ribosomal protein P0
RRM1/2	Ribonucleotide Reductase Regulatory Subunit M1/2
RRP9	Ribosomal RNA Processing 9
RUNX1	RUNX family transcription factor 1
SAT	pericentromeric satellite repeats
shRNA	short hairpin RNA
snRNA	small nuclear RNA
sAML	secondary AML
SAT	satellite DNA
SINE	short interspersed nuclear elements
SLC28A3	Solute Carrier Family 28 Member 3
SLC29A1	Solute Carrier Family 29 Member 1
SMARCA5	SWI/SNF Related, Matrix Associated, Actin Dependent Regulator Of Chromatin, Subfamily A, Member 5
srpRNA	signal recognition particle RNA
SRSF2	Serine And Arginine Rich Splicing Factor 2
SSFV	spleen focus-forming virus
SUMO	Small Ubiquitin-like Modifier
SWI/SNF	SWItch/sucrose non-fermentable
TET2	ten-eleven translocase 2
TP53	Tumour protein 53
TPRT	target-primed reverse transcription
tRNA	transfer RNA

Tris	tris(hydroxymethyl)aminomethane
UCOE	ubiquitous chromatin opening element
UCK	Uridine-Cytidine Kinase
UTR	untranslated region
WHO	World Health Organization

Contents

Acknowledgments **5**

1 Introduction **19**
1.1 Epigenetics and chromatin 21
1.1.1 Concept of epigenetics and chromatin 21
1.1.2 Epigenetic regulation 23
1.2 MacroH2A histone variants 34
1.2.1 The MacroH2A histone variant family 34
1.2.2 Known molecular and physiological roles of macroH2As 35
1.3 Long interspersed nuclear elements 40
1.3.1 Quick overview of transposable elements . . 40
1.3.2 LINE-1 41
1.3.3 LINEs regulations and their relevance 43
1.4 Myelodysplastic syndromes 46
1.4.1 Normal haematopoiesis 46
1.4.2 Myelodysplastic syndrome 47
1.4.3 Acute myeloid leukaemia 52
1.4.4 An overview of molecular alterations in MDS and AML 52
1.4.5 Treatments 57

2 Rational and objectives **63**

3 Results **67**
3.1 Results I: macroH2As association with LINEs *in silico* 69
3.1.1 MacroH2As are associated with LINEs and simple repeats 69
3.1.2 The association is less evident *in vitro* 72
3.1.3 MacroH2As are overlapping with LINEs' youngest element 76
3.1.4 Previous macroH2A specific enrichment on 3'UTR of younger LINEs unconfirmed 82
3.1.5 MacroH2A deficiency has no effect on LINE-1 expression 85
3.1.6 Result I: conclusions 89
3.2 Results II: chromatin regulators as potential therapeutic targets in clonal hematopoietic stem cell diseases 91
3.2.1 An unbiased approach to finding new therapeutic targets 91
3.2.2 Validation of selected targets - Response to treatment 98
3.2.3 In a mixed population only BAZ1A depletion shows higher sensitivity to Azacitidine treatment 102
3.2.4 Confirming results in two additional cell lines 104
3.2.5 The sensitizing effect of BAZ1A and macroH2A1 depletion in response to azanucleoside treatment is cell line dependant 108
3.2.6 Both macroH2A1 and BAZ1A depletion partially affect apoptosis in response to Decitabine in MOLM-13 112
3.2.7 MacroH2A1 depletion increases cell death by Bcl2 inhibitor, ABT-199 in MOLM-13 cells . . 114
3.2.8 Results II: conclusions 115
3.3 Results III: gene expression in MDS patient samples 117
3.3.1 Cohort characteristics 117

3.3.2 Selection of genes for a panel expression analysis . . . 118
3.3.3 High variability of expression profile in patient samples . . . 126
3.3.4 Differential expression depending on key gene mutational status . . . 128
3.3.5 Gene expression as an indicator for response to treatment . . . 131
3.3.6 Overall survival in high-expression versus low-expression . . . 136
3.3.7 Results III: conclusions . . . 140

4 Discussion 141
4.1 Discussion I: the challenge of studying repetitive sequences . . . 143
4.1.1 The relevance of studying LINEs in relation to macroH2As . . . 143
4.1.2 Dichotomy of the results . . . 144
4.2 Discussion II: cell lines-based approach to clinical conditions . . . 147
4.3 Discussion III: of the importance of epigenetics . . . 150
4.3.1 The relation between mutational status and gene expression . . . 150
4.3.2 The relation between response to Azacitidine and gene expression . . . 154
4.3.3 Patient samples: benefit and limitations . . . 157

5 Conclusions 159

6 Methods 165
6.1 Molecular biology . . . 167
6.1.1 RNA expression . . . 167
6.1.2 Protein expression analysis . . . 173
6.1.3 Cloning . . . 175
6.1.4 Chromatin immunoprecipitation . . . 178

6.2 Cell culture 180
6.2.1 Culture conditions 180
6.2.2 Gene transduction 181
6.2.3 Screen 182
6.2.4 Dose-response 183
6.2.5 Competitive growth assay 183
6.2.6 Flow Cytometry 185
6.3 Bioinformatic analysis 187
6.3.1 External data sources 187
6.3.2 Association analysis with ChIP-seq data . . . 188
6.3.3 Mapping reads to a LINEs consensus genome 189
6.4 Antibodies 190
6.5 Cell lines 191
6.6 Oligonucleotides 191
6.7 Plasmids with short hairpin 196
6.8 Plasmides 196
6.9 Reagents 196
6.10 Kits 198
6.11 Disposable labware 199

Chapter 1

Introduction

Introduction

In this introduction I provide information on various biological concepts. It is by no means exhaustive, but it aims to provide key information and context for the understanding of my doctorate work. As all this work was completed using human cell lines and samples, all statements refer to human biology unless otherwise indicated.

1.1 Epigenetics and chromatin

1.1.1 Concept of epigenetics and chromatin

Epigenetics was defined in 1942 by Conrad H. Waddington as "*the branch of biology which studies the causal interactions between genes and their products, which bring the phenotype into being*" (C H Waddington, 1942). He described epigenetics as a mountain landscape, with a marble at the top being an undifferentiated cell and epigenetics shaping different valleys below guiding the marble while it rolls down toward one or another differentiated state (**Figure 1.1**).

Since then, scientific knowledge has grown and now the study of epigenetics encompasses a group of molecular mechanisms and marks, affecting DNA compaction and transcription and allowing a cell to differentiate. Epigenetics are the missing link between the genotype and the phenotype. To further explain epi-

genetic mechanisms, we need to define chromatin, as one influences the other.

Chromatin is contained inside the nucleic compartment and formed by RNA, DNA and proteins woven into its structure. When talking about chromatin there are two important things to take into account: it is three-dimensional, and it is a dynamic structure varying not only between cells but also during the lifespan of a cell itself (Hauer and Gasser, 2017).

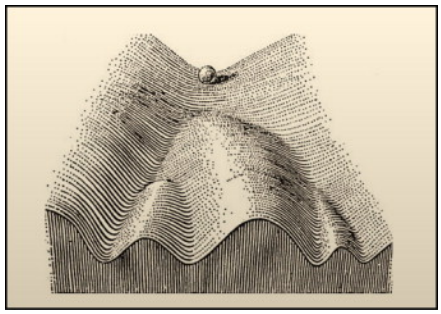


Figure 1.1: Waddindton's epigenetic landscape. The ball represents a non differentiated cell that could roll down different valleys formed by epigenetic regulations to become different mature cells.(from C. Waddington, 1957)

On a large scale, the unit of the chromatin is the chromosome which occupies a relatively separate space within the nucleus called chromosome territory (Misteli, 2020). Within these chromosome territories the chromatin is organized in compaction and in all three dimensions, with a tendency of high compaction toward the periphery of the nucleus and low compaction toward the interior creating a gradient of transcriptional activity. These states of compaction created the first distinction within the chromatin due to their staining density. As a result, it is categorized in two main types: the euchromatin and the heterochromatin (Hildebrand and Dekker, 2020). Euchromatin is transcribed by the transcription machinery lightly packed. Heterochromatin mainly includes tandem repeats, such as microsatellites, minisatellites and transposons repeats tightly packed, as well as genes whose state of compaction and transcription vary in function of cell type and degree of differen-

tiation. The DNA forms loops and nods, which affect transcription, replication, and recombination, modulating the level of compaction locally. Key players in the looping system feature the cohesin complex that interacts with the CCCTC binding factor (CTCF)(D. C. Wang et al., 2018).

For a better understanding of the different levels of compaction we need to move to a smaller scale. The nucleosome serves as the base for the compaction of the chromatin. It is a chromatin sub-unit composed of 146-147 pairs of nucleotides wrapped in two loops around a protein core of eight proteins called histones (Luger et al., 1997). These protein cores are made of two H2A histones, two H2B histones, two H3 histones and two H4 histones. Histone H1 is not part of the nucleosome but binds to it and to the extra nucleosomal DNA stabilising the higher-level structure of the chromatin. Locally, nucleosome positioning and density modulate the compaction of chromosomal regions and the looping system affecting DNA processing. In more recent years attention has been given to chromatin motion as a separate phenomenon from compaction status (Shaban et al., 2020).

1.1.2 Epigenetic regulation

As previously stated, epigenetics are the missing link between genotype and phenotype. Therefore, it is tightly regulated by various mechanisms.

DNA methylation

One of the most well known epigenetic regulations is independent of nucleosomes because it directly affects the DNA molecule. DNA methylation is the addition of a methyl group to a cytosine by DNA methyltransferases (DNMTs) (Gujar, Weisenberger and Liang, 2019). In human cells, there are two main DNMTs:

DNMT1, which maintains the DNA methylation during mitotic division allowing a stable transmission of methylation profile between somatic cells; DNMT3 is responsible for the *de novo* cytosine methylation as well as the maintenance of methylation in both undifferentiated somatic cells and embryonic cells. The DNA methylation is known to be a repressive mark, which means that methylated genes are not transcribed. It is also a reversible modification, wherein the demethylation is done by the ten-eleven translocases (TET) (Wu and Zhang, 2017). TET enzymes with co factors oxydize the methylated cytosine leading to its replacement by a non-methylated cytosine.

Histone post-transcriptional modifications

Another mechanism of regulation involves post transcriptional modification of histones (PTM) also known as histone marks. It is a core mechanism of both cellular and hereditary epigenetic transmission and regulation. Histone PTMs include methylation, acetylation, ubiquitination, SUMOylation, citrullination, glycosylation, ADP-ribosylation and phosphorylation (Zhao and Garcia, 2015). The two most understood types of histone PTMs are methylation and acetylation, which are mutually exclusive. These modifications are done on the tail of some histones, mainly H3 and H4.

The transfer of a methyl group from an S-adenosyl methionine to a lysine or arginine residue is done by specific members of the histone methyltransferase family. These are lysine methyltransferases and protein arginine methyltransferases respectively. Likewise, removal of a methyl group is done by lysine demethylases and protein arginine demethylases. Some lysines can be methylated, dimethylated or trimethylated and some arginines dimethylated (Audia and Campbell, 2016).

An acetyl group can be added by histone acetyltransferases or removed by histone deacetylases. Histone acetyltrans-

ferases can be divided in four families: the General Control Non-Derepressible 5 related N-Acetyltransferases family; the MYST family including MOZ, Ybf2/Sas3, Sas2 and Tat Interacting Protein 60; the P300/CBP family, including Adenoviral E1A-associated protein of 300kDa (p300) and the CREB-binding protein (CBP); and the last family regroups transcriptional co-activators, and steroid receptor co-activators with acetyltransferase activities (Wapenaar and Dekker, 2016). Histone deacetylases are separated in four classes, I to IV: classes I, II and IV are classical, but Class III - the sirtuins - have the specificity of being nicotinamide adenine dinucleotide (NAD+)- dependent (Park and Kim, 2020).

Beyond the biochemical modification, histone PTMs affect nuclear mechanisms. In 1964 Allfrey *et al.* discovered histone acetylation (Allfrey, Faulkner and Mirsky, 1964). Over twenty years later a general correlation between histone acetylation and gene activity was demonstrated (Hebbes, Thorne and Crane-Robinson, 1988). Since then, this type of correlation has been refined and detailed. For example, it is now established that histone H3 (H3) lysine 9 (K9) trimethylation (me3) is a hallmark of constitutive heterochromatin and under acetylation of pericentromeric regions necessary for a proper mitosis (Taddei et al., 2001). H3S10 phosphorylation is essential for proper condensation and segregation of the chromosome (Wei et al., 1999). H3K27me3 is associated promoters, poised enhancers, gene silencing (Rada-Iglesias et al., 2011) whereas H3K27ac is associated with active enhancers (Creyghton et al., 2010). The complexity of these PTMs led to the controversial hypothesis from Brian D. Strahl and C. David Allis of the existence of a “histone code” on top of the genetic code for other proteins to read and to trigger response (Strahl and Allis, 2000). The hypothesis is attractive but falls short in a few situations. Thanks to technological innovation a more global approach has been developed, integrating histone marks ChIP-seq data and now including the transcription factor and insulator, CTCF. In addition, various modellings of chromatin states emerged when using

a multivariate hidden Markov model (HMM) . Chromatin states are defined based on different combinations of histone modifications and correspond to different functional regions such as promoter-associated, transcription-associated, active intergenic, large-scale repressed and repeat-associated states (Ernst and Kellis, 2010).

Histone variants

Another way to affect and regulate chromatin is through histone variants. Variants of histones are histones, but they are substituting some of the replication-coupled histones, formally called "canonical histones". It should be noted that there is no known variant of H4 histone and H2A has the most variants with 8 different identified variants in somatic cells: H2A.X, H2A.Z.1, H2A.Z.2.1, H2A.Z.2.2, H2A.Bbd, macroH2A1.1, macroH2A1.2 and macroH2A2 (Buschbeck and Hake, 2017) (**Figure 1.2)**. Histone H3 has six variants, namely: H3.3, CENP-A, H3.1T, H3.5, H3.X (H3.Y.2) and H3.Y (H3.Y.1) . TSH2B, H2BFWT, H3.1t and H3.5 are known as testis specific histone variants. The first difference to consider when talking about histone variants is their deposition. Replication-coupled histones are globally evenly distributed across the genome, whereas histone variants are specifically deposited in specific points in space and time to complete their function. The histone deposition is mediated by histone chaperones that intervene at different steps. Indeed, histone chaperones can contribute to the pre-formation of the nucleosome in the cytoplasm (Cook et al., 2011), the entry in the the nucleus (Campos et al., 2010), the nucleosome assembly (Nakatani et al., 2004) and the histone exchange in an already formed nucleosome (Lewis et al., 2010). The second point is that each variant of histone has specific properties and they take part in a large range of physiological roles. For example: H2A.X has an additional C-terminal motif phosphorylated by DNA damage checkpoint kinases on Ser139, to generate γH2A.X at the start of the double strand DNA damage

response mechanism (Rogakou et al., 1998). H2A.Z incorporation increases accessibility for the repair machinery to the DNA for both homologous and non-homologous combinations (Giaimo et al., 2019). CENP-A is incorporated in the pericentromeric region and is essential for chromosome segregation (Black et al., 2007), (Lacoste et al., 2014). For instance, the sub-family called macroH2A histones has been shown to maintain the integrity of the 3D nuclear architecture (Douet et al., 2017). This family is a focus of the work in this thesis; I will expand on it later in this introduction. Histone variants are also subject to PTMs, some similar to their replication-coupled counterpart, others specific unto themselves (Corujo and Buschbeck, 2018). Lastly, the timing of expression of each variant is specific and the disruption of that timing can lead to disease such as cancer (Buschbeck and Hake, 2017),(Corujo and Buschbeck, 2018).

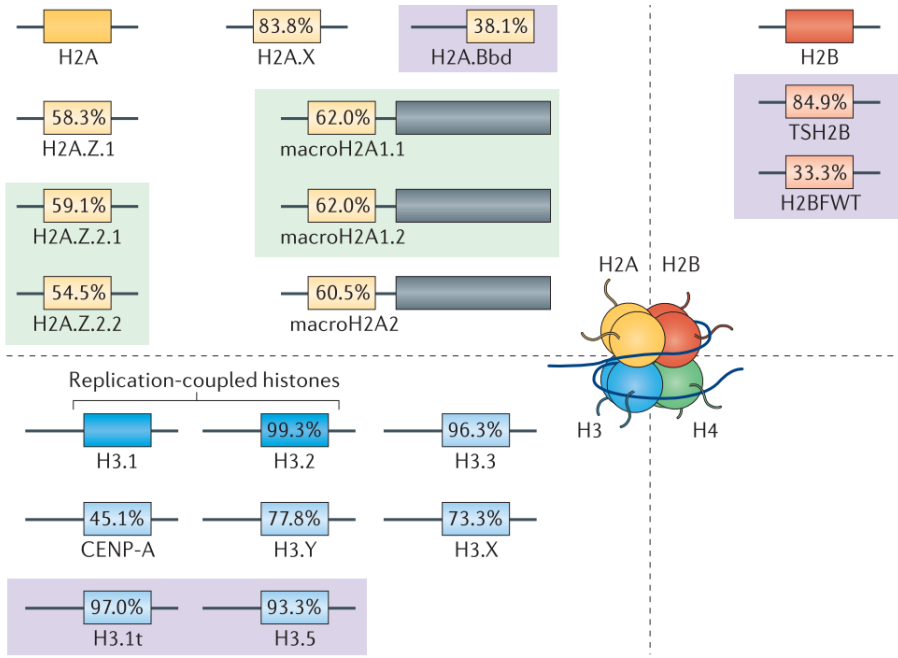


Figure 1.2: A depiction of human variants of histone H2A (yellow), H2B (orange) and H3 (blue), with variants shown in pale yellow, pale orange and pale blue, respectively. Alternative splice isoforms are in light green boxes. Percentages indicate total amino acid sequence conservation (% sequence identity) of the variants relative to their replication-coupled counterparts (for H3, two replication-coupled isoforms are present (H3.1 and H3.2); in the figure, sequence conservation was calculated for H3.1). CENP-A, histone H3-like centromeric protein A; H2BFWT, histone H2B type WT; TSH2B, testis-specific histone H2B.(image taken from Buschbeck and Hake, 2017)

Chromatin remodellers

Lastly, the position of the nucleosomes themselves adds a layer of complexity in epigenetic regulation. Chromatin remodellers regulate the chromatin by re-positioning, ejecting, or modifying nucleosomes. Overall, they regulate the accessibility to the DNA template and by extension all processes relying on it, such as transcription, DNA replication and DNA repair. They were first discovered in yeast, thus most of the names and classifications are based on the yeast homologues. When we are talking about

chromatin remodellers, we are referring to protein complexes which house a catalytic subunit containing a SWI2/SNF2-family ATPase domain that performs DNA translocation along the histone core of the nucleosome, as well as accessory subunits for target recognition, specificity, and modulation of the ATPase activity. There are four categories of remodelling complexes which are based on the sequence homology of the catalytic ATPase and the accessory subunits: SWItch/sucrose non-fermentable (SWI/SNF), imitation SWI (ISWI), chromodomain-helicase DNA-binding protein (CHD), and INOsitol-requiring mutant 80 (INO80). ISWI and CHD are generally formed by four proteins or less and they are implicated in nucleosome assembly, spacing and maturation (Figure 1.3). SWI/SNF and INO80 are made of up to 15 proteins. SWI/SNF cover many activities such as histone dimer or nucleosome ejection, and nucleosome sliding while INO80 are mainly active in histone exchanges (Figure 1.3). Each category of chromatin remodellers contains many complexes based on the subunit compositions which vary in function of the cell-type, the tissue, or the stage of development(Figure 1.4).

SWI/SNF complexes are well conserved from yeast to human, they are quite large , 2MDa, chromatin remodelling complexes (Kadoch and Crabtree, 2015). Three mammalian SWI/SNF complexes have been identified: BRG1/BRM-associated factor (BAF), polybromo-associated BAF, and non-canonical BAF. They all contain an ATPase subunit, SWI/SNF Related, Matrix Associated, actin dependent Regulator of Chromatin 2 (SMARCA2) or SMARCA4, that bind both the DNA and histones and co-factors with various enzymatic activities (bromodomain, chromodomain, helicase, zinc-finger...). Globally, SWI/SNF complexes are responsible for removing a nucleosome, sliding it along the DNA or switching a histone dimer.

INO80 replace replication-coupled histones H2A by H2A.Z influencing both transcription and replication (Brahma et al., 2017). They exhibit DNA helicase activity and binds to specialized

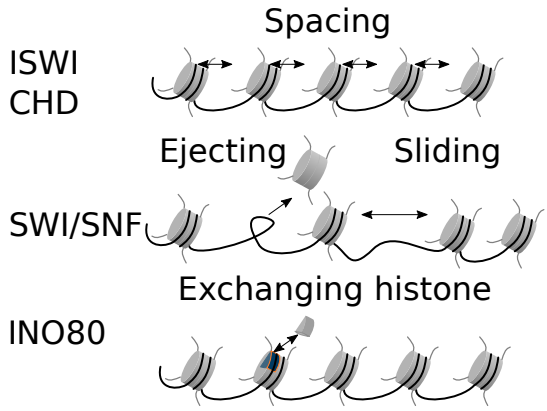


Figure 1.3: Overview of the functions ofATP-dependent chromatin remodeling complexes Chromatin remodelers have different functions depending on their composition. Part of ISWI and CHD complexes are involved in nucleosome spacing. SWI/SNF complexes are implicted in ejecting and sliding nuclesome along the genome. INO80 complexes are mainly part of the histone exchange activity.

DNA structures. Recently human INO80 have been classed in two groups: canonical INO80 and Non-Canonical INO80 (Runge, Raab and Magnuson, 2018).

Chromodomain-helicase DNA-binding proteins (CHDs) help assemble or disassemble nucleosomes and space them in cooperation with ISWI complexes (Marfella and Imbalzano, 2007). CHDs all contain a chromodomain and are sub-categorised into different subgroups according to their secondary domain: Chd1-Chd2 contain DNA-binding domain located in the C-terminal region; Chd3-Chd4 all have a paired N-terminal PHD (plant homeodomain) Zn-finger-like domain; Chd5-Chd9 contain a paired Brahma and Kismet domain, a switching-defective protein-3 adaptor-2 nuclear receptor co-repressor, a transcription factor IIIB domain,a CR domain, and a DNA-binding domain. CHDs are very heterogeneous, they recognise different marks - for example, hCHD1 binds to H3K4me2/3 via tandem chromodomains (Sims et al., 2007).

ISWI complexes were first discovered in drosophila and later many complexes were identified in humans: NURF (ATPase

motor of the nucleosome remodelling factor), ACF (ATP-utilizing chromatin assembly and remodelling factor), CHRAC (chromatin assembly complex), RSF, NoRC, WICH, CERF, and finally, SNF2H-cohesin (Hasan and Ahuja, 2019). They pair one of two ATPase subunits, SMARCA5 (SNF2H) or SMARCA1 (SNF2L), with one or two of the regulatory subunits: bromodomain adjacent to zinc-finger domain protein 1A (BAZ1A or ACF1), BAZ1B (WSTF), BAZ2A (TIP5), BAZ2A, bromodomain and PHD finger-containing transcription factor (BPTF), cat eye syndrome critical region protein 2 (CECR2), retinoblastoma-binding protein 4 (RBBP4), RBBP7 or remodelling, chromatin accessibility complex protein 1 (CHRAC1 or CHRAC15, DNA polymerase epsilon subunit 3 (POLE3 or CHRAC17) and spacing factor 1 (RSF) (Figure 1.4). I will focus more on the CHRAC complex as it is relevant for the rest of this thesis.

CHRAC complex is formed by SMARCA5, POLE3, CHRAC1 and BAZ1A. It facilitates DNA replication by sliding nucleosomes without causing major trans displacement of histones. As mention previously SMARCA5 is the ATPase subunit of the CHRAC complex but also of most of the ISWI complexes. It is highly conserved across mammals and has been shown to be essential to haematopoiesis (Kokavec et al., 2017). POLE3 contains a histone fold domain and a unique acidic C terminus. It has been shown in mice to be involved in T and B cell development (Siamishi et al., 2020). CHRAC1 also contains a histone fold domain and forms a stable complex with POLE3 to bind to the DNA (Poot et al., 2000). Finally BAZ1A is a non-catalytic subunit but it is essential to the sliding activity of the complex (Racki et al., 2009).

Epigenetic regulations a collaborative game

Another way of classifying enzymes and complexes involved in epigenetic regulations is by dividing them between writers, readers and erasers. It is the interplay of “writers” and “erasers”

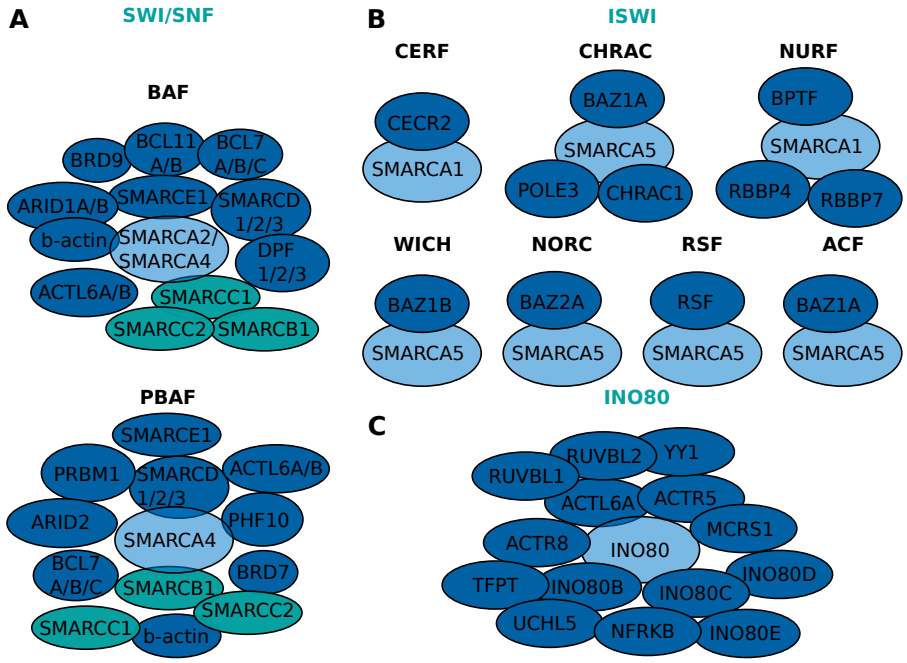


Figure 1.4: Overview of the subunit compositions of the ATP-dependent chromatin remodelling complexes. Subunits that comprise the mammalian (**A**) SWI/SNF complexes, (**B**) ISWI complexes, and (**C**) INO80 complex (for SWI-SNF complexes: orange color corresponds to catalytic ATPase subunits, green color corresponds to core subunits, and blue color corresponds to accessory subunits; for ISWI and INO80 complexes: orange color corresponds to catalytic ATPase subunits). For subunits that are separated by dashes, only one of the subunits is present in a given complex. Subunit composition might be different based on tissue/cell types. SWI/SNF noncanonical complex ncBAF and INO80 subfamily complexes p400 and SRCAP are not included in the schematic. Modified from Hasan and Ahuja, 2019

that allows the fine tuning of histone methylations for the “readers” to interpret. For example, PRC2(Polycomb Repressive Complex 2) is part of the writers (Cao and Zhang, 2004) while PRC1 reads methylation of H3K27 and triggers chromatin compaction (Cao et al., 2002). H3K9me is recognised by the chromodomain of heterochromatin protein 1 (HP1) leading to the establishment and maintenance of heterochromatin.

It is important to understand that all these regulations do

not happen in parallel and independently of each other, but it is an extremely complex interplay. For example, lysine acetylation was thought to render nucleosomes less stable by neutralizing the lysine positive charge (Hongs et al., 1993), thus facilitating transcription; however, now studies show that it plays a role in the recruitment of chromatin remodellers containing a chromodomain that open the chromatin for transcription (Lachner et al., 2001). The NuRD complex contains Chd3/4 ATPase as well as two histone acetylases, HDAC1 and HDAC2, which allow it to recognize histone acetyl-lysine marks and remove the acetyl group. This complex also interacts with methylated DNA via MBD2 (Y. Zhang et al., 1999). Nucleosomes containing the histone variants macroH2A1 are more stable and have been shown *in vitro* to be impeding chromatin remodelling by ACF (ISWI) and SWI/SNF complexes as well as lysine acetylation by p300-dependent histone acetylation (Angelov et al., 2003 and Doyen et al., 2006).

1.2 MacroH2A histone variants

1.2.1 The MacroH2A histone variant family

The histone variant family, macroH2A, was discovered almost thirty years ago (Fried and Pehrson, 1992). This family consists of three isoforms: macroH2A1.1, macroH2A1.2 and macroH2A2 (Buschbeck and Hake, 2017). The first two are splicing isoforms from the gene, *MACROH2A1* whereas the third is the product of a separate gene, *MACROH2A2*, previously named *H2AFY* AND *H2AFY2* (Fried and Pehrson, 1992) respectively.

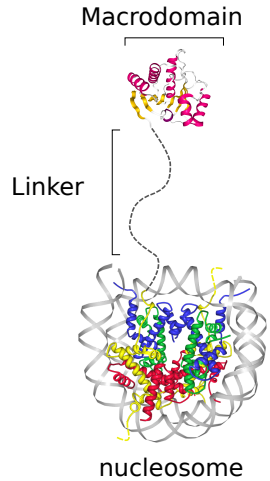


Figure 1.5: Structure of a nucleosome containing a macroH2A histone variant based on crystal structure. The nucleosome represented is made of one H2A and one macroH2A histones (both in yellow), H2B (red), H3 (green) and H4 (blue) and DNA (grey). The macrodomain of macroH2A is colored by secondary structure (α-helices in magenta and β-sheets in orange). The linker is represented with dashes as no crystal structure of it exists.The picture of the nucleosome is based on protein data bank ID 3REH (H. Wu et al., 2014) and generated with ProteinWorkshop (Moreland et al., 2005), the picture of the macroH2A macrodomain is based on PDB ID 2FXK (Kustatscher et al., 2005) and generated with NGL viewer (Rose and Hildebrand, 2015).

In my thesis I focus on human macroH2As. When not making the distinction between the splicing isoforms, I'll use macroH2A1 and macroH2As when referring to all three proteins. All three proteins are composed of three parts: a core histone similar to their replication-coupled counterpart H2A (65% se-

quence identity), a large globular macro domain, and an unstructured linker joining the both parts (Changolkar, Singh and Pehrson, 2008) (**Figure 1.5**). The linker plus the macrodomain are about three times the molecular weight of the core histone part, thus the incorporation of a macroH2A histone variant is a considerable modification of the nucleosome. The three proteins did not appear simultaneously but most likely evolved from one another by gene and exon duplication (Guberovic et al., 2021). The two splicing isoforms differ only in one exon (exon V) which affects the binding pocket situated in the macrodomain (Kustatscher et al., 2005). All three macrodomains show different binding properties. MacroH2A1.1 binds ADP-ribose and by consequence auto-ADP-ribosylated PARP1 whereas macroH2A1.2 and macroH2A2 do not (Timinszky and Ladurner, 2014). Binding partners for the two later have yet to be find. They also show different pattern of expression. For example, macroH2A1.2 is much more expressed in mouse ES cells than the two others (Creppe et al., 2012). MacroH2A1.1 and 1.2 often have opposite patterns of expression: the former is more expressed in differentiated cells whereas the latter is mainly expressed in proliferative cells (Buschbeck and Hake, 2017). MacroH2A1.2 and macroH2A2 are both concentrated in the inactive X chromosome but have a different pattern of expression in the liver and the kidney (Costanzi and Pehrson, 2001).

1.2.2 Known molecular and physiological roles of macroH2As

As previously stated, macroH2As were discovered almost thirty years ago. Since then, they have been shown to be implicated in numerous physiological and pathological processes; however, there are still a lot of unknowns at the molecular level.

There are two H2A histones in each nucleosome. Nucleosomes containing macroH2As tend to have one variant and one replication-coupled histone H2A, and increased stability *in vitro*

(Chakravarthy et al., 2005). It was suggested that macroH2As nucleosomes are refractory to p300-dependent histone acetylation and chromatin remodelling by SWI/SNF and ACF (ISWI complex) (Doyen et al., 2006), but a later study actually indicates that macroH2As nucleosomes are an excellent substrate for ISWI (Chang et al., 2008).

In 2005 it was shown that macroH2A1.1 macrodomain can bind to ADP-ribose and ADP-ribosylated proteins (Kustatscher et al., 2005). Later, it was shown that by binding ADP-ribose, macroH2A1.1 inhibits Poly [ADP-ribose] polymerase 1 (PARP1) activity (Timinszky and Ladurner, 2014). Inhibition of PARP1 reduces the nuclear consumption of NAD+ and impacts cellular NAD+ homeostasis (Posavec Marjanović et al., 2017). An important point to take from this mechanism is that this regulation does not involve a direct effect of macroH2A1.1 on gene expression. MacroH2A1 is also a crucial partner of polycomb repressive complex 1 for the stable X chromosome inactivation (Hernández-Muñoz et al., 2005). Large macroH2As-bound regions overlap with the repressive

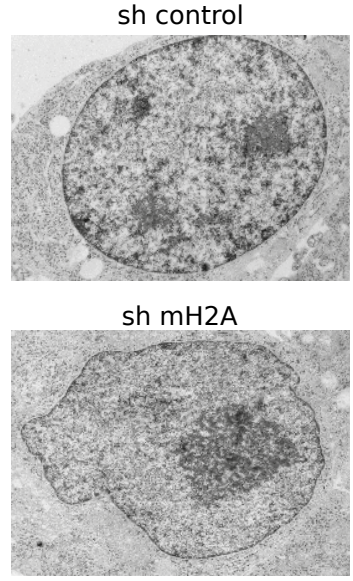


Figure 1.6: Transmission electron microscopy images of nuclei from HepG2 cells control and knock down for macroH2As. Illustration of the disruption of the nuclei structure and the heterochromatin by the loss of macroH2As. (from Douet et al., 2017)

mark, histone H3 lysine 27 trimethylation (H3K27me3), (Gamble et al., 2010). And at the nuclear scale, macroH2As maintain chromatin organisation in HepG2, as their knockdowns disrupt the heterochromatin architecture (Douet et al., 2017) (**Figure 1.6**). MacroH2As appear to contribute to the robustness of overall gene expression patterns and the reduction of transcriptional noise (Lavigne et al., 2015). MacroH2As are also involved in the silencing of specific genes such as B-cells, *CXCL8*, coding for interleukin-8 (Agelopoulos and Thanos, 2006), and during neuronal development, the *HOXA* cluster (Buschbeck et al., 2009). No sequence motif has been identified for macroH2As deposition nor has a chaperone but there is growing evidence of the involvement of facilitates chromatin transcription complex (FACT) (Z. Sun et al., 2018; Ni and Muegge, 2021).

Over the years it has been shown that macroH2As are implicated in many major cell processes (Figure 1.7). They are a positive factor in senescence and differentiation. MacroH2A levels generally increase during pluripotent stem cell differentiation and promote the differentiation of embryonic stem cells (Barrero et al., 2013; Creppe et al., 2012). They are involved in the development and differentiation of neurons (Buschbeck et al., 2009). MacroH2A1 is necessary for senescence-associated secretory phenotype (H. Chen et al., 2015). On the other hand, macroH2As restrict cell reprogramming, notably the re-activation of inactivated X chromosome (Pasque et al., 2011). There is some controversy on the role of macroH2A1 in stemness. The knock-down of macroH2A1 has been shown to increase stem-like phenotype (Lo Re et al., 2018), but the same group in a macroH2A1.1 knock-down mouse also led to a reduction of haematopoietic stem cells after high-dose irradiation (Bereshchenko et al., 2019). MacroH2A1.1 ectopic expression inhibits epithelial-mesenchymal transition induction (Hodge et al., 2018). MacroH2As seem to have several functions by interfering in different cellular processes.

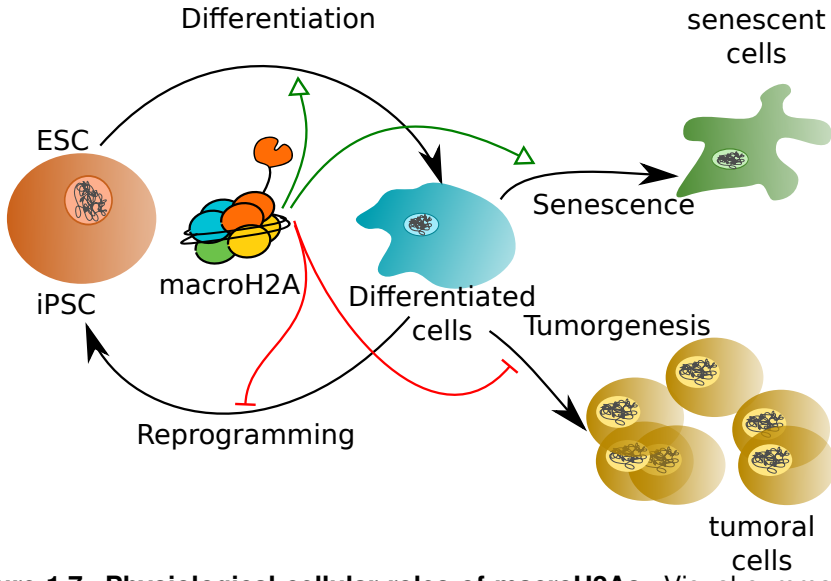


Figure 1.7: Physiological cellular roles of macroH2As. Visual summary of different function of macroH2As. Embryonic Stem Cell (ESC) / Induced Pluripotent Stem Cell (iPSC). Green arrows active function, red arrows inhibitory activity

MacroH2As in cancers

MacroH2As also have been shown to be relevant in different types of cancers. It is mainly referred to as a tumor repressor. MacroH2As depletion or isoform switch are associated with more aggressive tumors in teratoma, breast, bladder, prostate and colon cancers (Dardenne et al., 2012; Sporn et al., 2009; Vieira-Silva et al., 2019). MacroH2A1.2 also contributes to the regulation of the expression of the tumor suppressor, TP53, reducing transcriptional noise (Pliatska et al., 2018). In prostate cancer, lower expression of macroH2A1.1 is associated with less-differentiated tumours (Vieira-Silva et al., 2019). MacroH2A1.2 modulates osteoclastogenic potential of cancer cells (J.-M. Kim et al., 2018). In melanoma, macroH2A1.2 is shown to influence melanoma malignancy (Kapoor et al., 2010). Unexpectedly, the presence of macroH2A1 prevents senescence by a DNA-demethylation agent, Azacitidine

in HepG2 cells (Borghesan et al., 2016). MacroH2A's mutation is not known as a driver for any type of cancer but there is growing interest in its roles in many types of cancers.

1.3 Long interspersed nuclear elements

1.3.1 Quick overview of transposable elements

In the 1940s, Barbara McClintock first described and theorised on the role of ”controlling elements” in maize (B. McClintock, 1951) (Barbara McClintock, 1956). Since then, the scientific community has come a long way in characterising the mobile elements in plants as well as in animals. 45% of the human genome is composed of transposable elements which are divided in class I - the most abundant being the retrotransposons - and class II - DNA transposons - ”only ” 2% of the human genome (Kazazian and Moran, 1998). Class II transposable elements move by ”cut and paste” as the DNA sequence itself is moved around the genome, whereas class I elements move by ”copy and paste”, making an RNA copy that will embed at a different site of the genome. Thus there are two major differences between these two classes: one leaves a copy behind and thus multiplies itself, whereas the other has a finite number of elements that moves around the genome. Class II is further divided between long-terminal repeat (LTR)-containing elements and non-LTR elements, which are composed of short interspersed nuclear elements (SINEs), long interspersed nuclear elements (LINEs) and SINE-VNTR-Alu elements (SVAs). As the name states, the first one is flanked by LTR sequences which are reminiscent from retroviruses. LINEs can further be classified into five major super-families: R2, LINE-1, RTE-1, I, and Jockey. The LINE-1 retrotransposons are the only active, the most abundant, and consequently the most important, human mobile element. They are also the most relevant to our understanding of this work. Thus, I will focus on them and not delve into the details of other transposable elements.

1.3.2 LINE-1

LINE-1 cover around 17% of human genomics sequences. There are 6,000 full-length LINE-1 elements in the human genome, of which 100 are classified as active elements or retrotransposition competent. To be classified as active, a LINE-1 element has to be 6 kb long and contain a 5'-untranslated region (5'-UTR) - including an internal promoter -, two open reading frames (ORF1 and ORF2), and the 3'-UTR, which includes a poly-A tail. There is variation to this structure among subsets of LINE-1 but it is the most frequent and the minimal accepted structure.

LINE-1 retrotransposition

The LINE-1 retrotransposition cycle is a multi-step process. First a full length mRNA copy is made and transported to the cytoplasm, where it is translated to two proteins: ORF1p and ORF2p. ORF1p is a p40 protein with RNA-binding and chaperone activities, while ORF2p is a protein of 150 kDa with endonuclease and reverse transcriptase activities (Long, Prescott and Wysocka, 2016). These two proteins plus a copy of the mRNA re-enter the nucleus where ORF2p cuts the genomic DNA in a new site and synthesize the complementary DNA sequence in continuity from the 3' end to the 5'end (**Figure 1.8**). This mechanism of insertion explains why over 90% of L1 elements are 5' truncated during the insertion step. (Smit et al., 1995). Thus it is the combined functions of ORF1p and ORF2p that enables LINE-1 autonomous activity.

LINE-1 families and subfamilies

New insertions of LINE-1 repeatedly happened across time and species-evolution. In parallel, numerous mutations of the sequences occurred upon insertion and by nucleotide substitution (Konkel, Walker and Batzer, 2010). This led to sequence deriv-

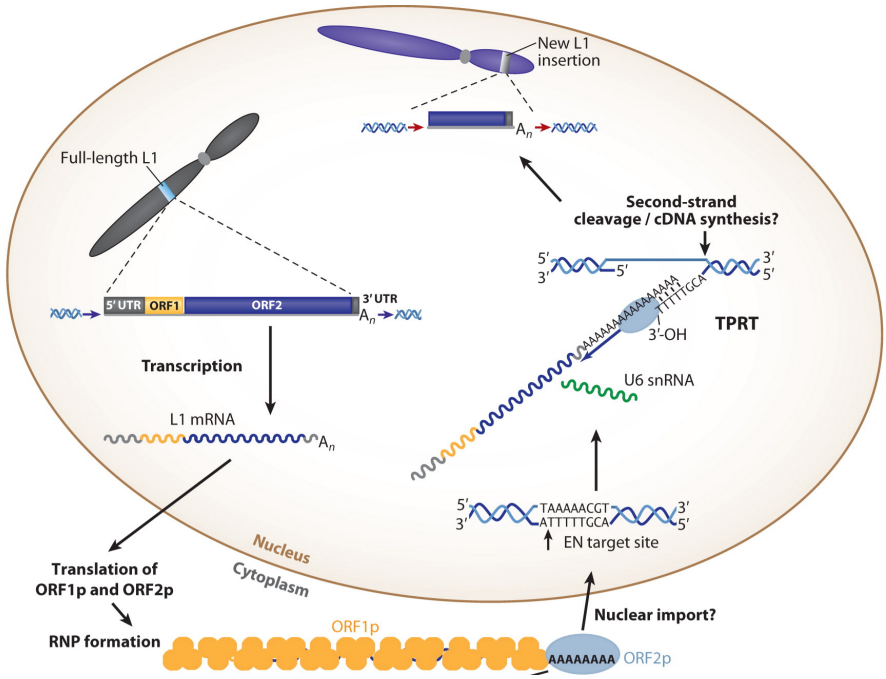


Figure 1.8: Retrotransposition mechanism of a LINE-1. First a full length LINE-1 (L1) (blue stripe on grey chromosome) is transcribed to a L1 mRNA which is then exported to the cytoplasm. Then the mRNA is translated in two proteins: ORF1p (yellow circles) and ORF2p (blue ovals), with the mRNA they form a ribonucleoprotein formation. This formation enters in the nucleus where retrotransposition occurs by target-site primed reverse transcription (TPRT). At the insertion site the ORF2p endonuclease (EN) activity cuts the a single-strand of DNA which then pairs with the polyadenosine tail of the mRNA. The reverse transcriptase part of ORF2p then recognises the 3'hydroxyl group (3'OH) exposed to synthesised the cDNA in continuity of the genomic DNA cut strand. lastly the second strand of DNA is cleaved and re-ligated to complete the integration (blue stripe on purple chromosome). *Modified from* Beck et al., 2011

ation, where the oldest members are more deviant and become retrotransposition incompetent. LINE-1 are divided into families according to the common phylogenetic class. For example, L1M are found in all mammalians, L1PB and L1PA in primates. Each family is composed of different subfamilies based on ORF1 and ORF2 homology (Figure 1.9) (Khan, Smit and Boissinot, 2006).

L1PA, 3'UTR evolution

In primates, three distinct lineages can be found: L1MA1-4, L1PB1-3 and L1PA1-17. L1PA1 is the youngest subfamily of the youngest family L1PA; L1PA1 elements are specific to Homo Sapiens and thus also called L1HS (**Figure 1.9**). The three distinct lineages evolved in parallel until at some point L1MAs and L1PBs became inactive. L1PAs evolved over time, with each new subfamily successively replacing the previous one. In humans, mainly L1HS remains active.

1.3.3 LINEs regulations and their relevance

If transposable element activity was not regulated, there would be many, possibly harmful, consequences. For example, numerous insertions would elongate the genome, potentially impacting the overall chromatin structure. The insertion site is also relevant: if the element inserts itself in the middle of a gene, the protein product would not be properly produced anymore; and if it inserts itself close to a promoter or an enhancer, it could modify gene expression. Thus, most transposable elements are being silenced through redundant epigenetic mechanisms. Most of methylated cytosines are on transposable elements sequences (Jansz, 2019). It has been shown in knock down experiments of DNMTs in murine cells that transposable elements expression increases whether DNMT1 or 3 is targeted (Walsh, Chaillet and Bestor, 1998; Arand et al., 2012). LINE-1 elements were shown to be silenced via DNA methylation in normal tissues as well as in human ESCs (Florl et al., 1999; Castro-Diaz et al., 2014). The Human Silencing Hub (HUSH) complex with TRIM28/KAP1 has been shown to contribute to the silencing of L1 elements via both H3K9me3 and DNA methylation (Robbez-Masson et al., 2018). Both polycomb complexes (PRC1 and PRC2) have been shown to suppress transposable elements via H3K27 trimethylation in murine cells (Leeb et al.,

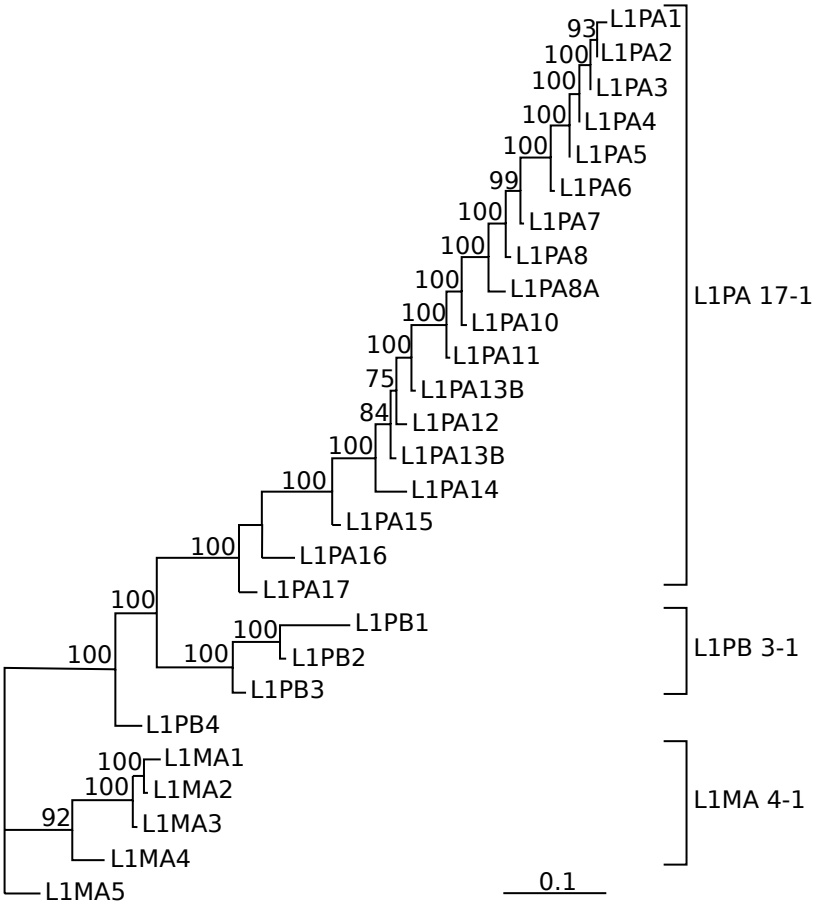


Figure 1.9: Phylogeny of L1 consensus sequences. This maximum likelihood tree is based on the family forms a separate lineage that consensus sequences of the ORF1 and ORF2 of 27 L1 families. The numbers above the nodes indicate the percentages of time the labeled node present in 1000 bootstrap replicates of the data. Modified from Khan, Smit and Boissinot, 2006

2010). Deregulation of LINE-1 has been shown to be implicated in oncogenesis. In fact, their expression increases genome instability which is a hallmark of cancer (Hanahan and Weinberg, 2011).

1.4 Myelodysplastic syndromes

1.4.1 Normal haematopoiesis

Before studying any pathology, it is important to understand the affected physiological process in healthy conditions. I present here an overview of the normal haematopoiesis. First, a definition: haematopoiesis is the process by which blood cells are formed, from haematopoietic stem cells (HSC) to fully differentiated circulating cells and each of their intermediaries. There is a lot of variability between animal species and here I focus on the human haematopoiesis. The haematopoiesis produces from a small number of HSC and progenitors an impressive 500 billion cells each day (Fliedner et al., 2002). In general, progenitors are found in the bone marrow whereas differentiated cells are found in the blood or in the tissues. Haematological cells are classically divided between the myeloid and the lymphoid lineages. The myeloid lineage matures in erythrocytes, thrombocytes, monocytes, macrophages, myeloid dendritic cells and granulocytes. The lymphoid lineage matures in lymphocytes B and T, Natural Killer cells and lymphoid dendritic cells, Figure 1.10. Recent advances in genomics showed that the frontier between both lineage is less stringent than we first thought, and progenitors are primed for one or the other, but it is not definitive (Liggett and Sankaran, 2020). The haematopoietic system is responsible for many physiological functions: the myeloid lineage is responsible for oxygen transport, haemostasis, and innate immunity while the lymphoid lineage is responsible for the body's adaptive immunity. At the functional level, HSC are defined by their ability to self-renew and to differentiate in all haematopoietic lineage. As a matter of note, both these capacities are at the essence of bone marrow transplant as treatment for certain haematopoietic diseases. In addition, the stemness capacity of HSC decreases with age (Van Zant and Liang, 2012). At the molecular level, HSC are defined by the ab-

sence of lineage specific cell surface marker (Lin) and the high expression of both stem cell antigen 1 (SCA1) and the tyrosine-protein kinase (KIT or CD117). This phenotype is known as Lin-SCA1+KIT+ (LSK) (Okada et al., 1992). Haematopoiesis is a multi-step process (**Figure 1.10**). First, HSC divide asymmetrically in HSC and multipotent progenitors, which then can differentiate into either common myeloid progenitor or lymphoid-primed multipotent progenitor. Both can differentiate in megakaryocyte/erythroid progenitor, granulocyte/macrophage progenitor or common lymphoid progenitor (Orkin and Zon, 2008). Lymphoid-primed multipotent progenitors differentiate preferentially in lymphoid-primed multipotent progenitor. Common myeloid progenitor differentiate preferentially in the two others but, as mentioned earlier, lineage commitment is now thought to not be as systematic as once thought (Liggett and Sankaran, 2020). Depending on the branch, the number of intermediary progenitors varies be two and five before a fully differentiated cell.

1.4.2 Myelodysplastic syndrome

Myelodysplastic syndrome (MDS) is a consequence of dysregulations of this finely regulated process. It affects myeloid lineage due to an abnormal differentiation and uncontrolled proliferation of malignant leukaemic stem cell (LSC) clones, hindering the normal haematopoiesis (Sperling, Gibson and Ebert, 2017). It is characterized by clonal haematopoiesis, one or more cytopenias (decreased cell number of a cell type), and abnormal cellular maturation. As this type of abnormality increases with age, MDS is the most frequent haematological disease in the elderly. Consequently, MDS classifications have evolved quite a bit in the past 40 years, from the French-American-British cooperative study group ((Bennett et al., 1982) to the World Health Organization (WHO) classification established in 2001 and revised in 2008 and 2016 (Arber et al., 2016). This latest classification is based on the dysplastic

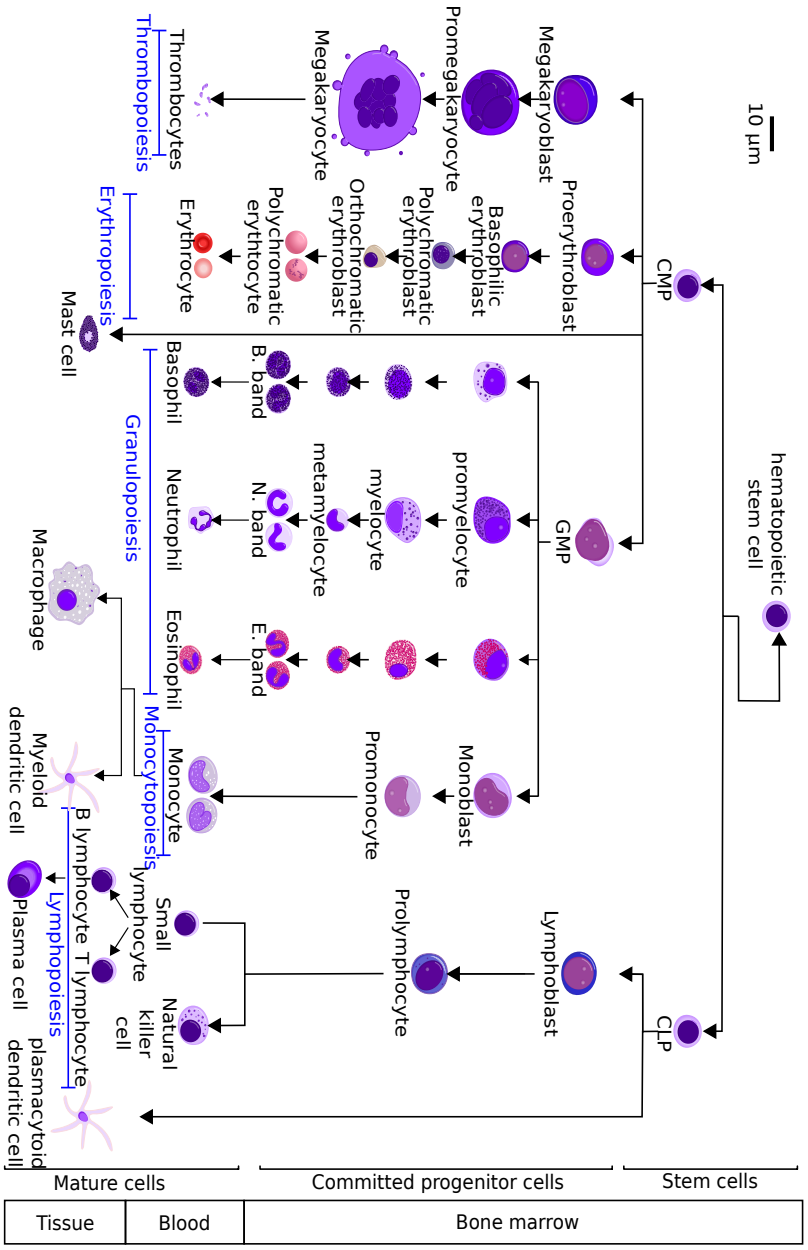


Figure 1.10: Haematopoiesis This diagramm shows the standard differentiation intermediaries of haematopoiesis. Cell size is done at scale. The morphological characteristics of the cells are based on Wright's stain, May-Giemsa or May-Grünwald Giemsa stain. Common myeloid progenitor (CMP), granulocyte/macrophage progenitor (GMP), common lymphoid progenitor (CLP) Figure adapted from Wikipedia.

lineages, the cytopenias, the presence of erythroblasts with iron-loaded mitochondria, the percentage of blast in the bone marrow and the peripheral blood as well as the cytogenetic alterations. **Table 1.1** details the 2016 WHO classification, but it can be summarised in three major categories: MDS with ring sideroblast, MDS with excess of blasts and the unclassifiable gathering of all other types. Among karyotypic alterations, special attention is given to the deletion of the long arm of chromosome 5 (del(5q)) because of its frequency but also because it has a specific symptomatology. Patients with the 5q- syndrome MDS often present erythroid hypoplasia, macrocytic anaemia, normal to elevated platelets count, and a preponderance of monolobulated megakaryocytes. MDS has a very broad cytogenetic and mutational landscape which is being studied more and more in depth in an effort to refine the classification and adapt treatment (Bersanelli et al., 2021; Adema et al., 2020; Papaemmanuil et al., 2013; Haferlach et al., 2014).

Name	Dysp lin-eages	Cytopenias	Ring sider-oblasts %	BM and PB blasts	Cytogenetics by conventional karyotype analysis
MDS with single lineage dysplasia (MDS-SLD)	1	1 or 2	$<$ 15%/$<$ 5%†	BM $<$5%, PB $<$1%, no A.r.	Any, unless fulfills all criteria for MDS with isolated del(5q)
MDS with multilineage dysplasia (MDS-MLD)	2 or 3	1-3	$<$15%/$<$5%†	BM $<$5%, PB $<$1%, no A.r.	Any, unless fulfills all criteria for MDS with isolated del(5q)
MDS with ring sideroblasts (MDS-RS)					
MDS-RS with single lineage dysplasia (MDS-RS-SLD)	1	1 or 2	$\geq$15%/$\geq$5%†	BM$<$5%, PB $<$1%, no 5% rods	Any, unless fulfills all criteria for MDS with isolated del(5q)
MDS-RS with multilineage dysplasia (MDS-RS-MLD)	2 or 3	1-3	¿15%/$\geq$5%†	BM $<$5%, PB $<$1%, no 5% rods	Any, unless fulfills all criteria for MDS with isolated del(5q)
MDS with isolated del(5q)	1-3	1-2	None or any	BM $<$5%, PB $<$1%, no A.r.	del(5q) alone or with 1 additional abnormality except -7 or del(7q)
MDS with excess blasts (MDS-EB)					
MDS-EB-1	0-3	1-3	None or any	BM 5%-9% or PB 2%-4%, no A.r.	Any

MDS-EB-2	0-3	1-3	None or any	BM 10%-19% or PB 5%-19% or A.r.	Any
MDS, unclassifiable (MDS-U)					
with 1% blood blasts	1-3	1-3	None or any	BM $<$5%, PB = 1%,‡ no A.r.	Any
with single lineage dysplasia and pancytopenia	1	3	None or any	BM $<$5%, PB $<$1%, no A.r.	Any
based on defining cytogenetic abnormality	0	1-3	$<$15%§	BM $<$5%, PB $<$1%, no A.r.	MDS-defining abnormality
Refractory cytopenia of childhood	1-3	1-3	None	BM $<$5%, PB $<$2%	Any

Table 1.1: WHO 2016 classification for MDS based on PB and BM findings and cytogenetics. Cytopenias defined as: hemoglobin $<$10 g/dL; platelet count $<$100 x10_9 /La and absolute neutrophil count $<$1.8x10_9 /L. Rarely, MDS may present with mild anemia or thrombocytopenia above these levels. PB monocytes must be $<$1x10_9 /L. Ring sideroblasts as % of marrow erythroid elements. A.r. =A.r. † If SF3B1 mutation is present. ‡ One percent PB blasts must be recorded on at least 2 separate occasions. § Cases with $\geq$15% ring sideroblasts by definition have significant erythroid dysplasia, and are classified as MDSTRSTSLD. Table adapted from Arber et al., 2016

1.4.3 Acute myeloid leukaemia

Leukaemia are the results of the disruption of the finely tuned process that is haematopoiesis. They are divided in four major groups: acute myeloid leukaemia (AML), chronic myeloid leukaemia, acute lymphoblastic leukaemia, and chronic lymphocytic leukaemia. Leukaemias are all the results of the anormal development of undifferentiated stem or progenitor cells in the bone marrow. Each group of leukaemia is divided in subgroups depending on the symptomatology and cytogenetics. AML is the most common type of acute leukaemia in adults (Germing et al., 2008). Like MDS, it is characterized by clonal haematopoiesis, one or more cytopenias, and abnormal cellular maturation but with a percentage of blasts superior to 20 percent in peripheral blood and bone marrow. 30 to 40 percent of MDS cases transform into secondary AML (sAML). The transformation to sAML is seen either as the consequence of an accumulation of mutational events (Lindsley et al., 2015) or of a single 'tipping point' mutation (Shukron et al., 2012). There is a correlation between this transformation and activating mutations in signaling pathways (T. Kim et al., 2017).

1.4.4 An overview of molecular alterations in MDS and AML

Molecular alterations studied for MDS and AML can be divided in different categories: loss or gain of the totality or part of a chromosome, and both germline and somatic mutations.

Chromosomal abnormalities More than half of MDS patients show abnormal karyotypes. I will only address the most frequent alterations. The deletion of the long arm of chromosome 5 (del(5q)), complete or partial, is seen in 10-15% of MDS patients (Komrokji et al., 2013). Several research groups have been trying to identify

the minimal deletion called common deleted regions (CDR) leading to the symptomatology associated to the del(5q). It has been narrowed down to the 5q31.1 and 5q32-33.3 regions (Ebert, 2011) and (Douet-Guilbert et al., 2012). No single mutated gene has been identified as the trigger of the 5q syndrome. In terms of prognosis, MDS del(5q) is only associated with a favourable prognosis. In contrast, MDS del(5q) plus additional chromosomal or mutational alterations is associated with a poor prognosis and increased probability of sAML. AML del(5q) is also associated with a poor prognosis.
10% of MDS patients show a deletion of the long arm of chromosome 7 (del(7q)), complete or partial. Several CDRs have been identified, namely 7q22, 7q32-33 and 7q35 (Hirai et al., 1987). In both MDS and AML, del(7q) is associated with a poor prognosis.
The third most frequent karyotype abnormalities in MDS is the trisomy 8, 5-10% (Schanz et al., 2011). It tends to appear late in the disease development and lead to sAML. Three or more cytogenetic abnormalities are defined as complex karyotypes. Complex karyotypes are frequent in MDS and are associated with poor prognoses.

Gene mutations The increased possibility for NGS led to numerous studies on patient samples, giving more information on the recurrence of mutated genes. They can be separated into six major pathways: histone modification (ASXL1, EZH2), DNA methylation (TET2, DNMT3A, IDH1, IDH2), signal transduction (RTK, FLT3, KIT, NRAS, KRAS, PTPN11), transcriptional regulation (RUNX1, TP53), cohesion complex (SMC1, SMC3, RAD21, STAG1/2) and RNA splicing (SF3B1, SRSF2, U2AF1, ZRSR2), (**Figure 1.11**) from (Awada, Thapa and Visconte, 2020) shows the main mutated genes in each of these pathways. The literature is quite extensive on many of these but I'd like to draw your attention to genes involved in epigenetic regulation. DNA methylation is associated with gene expression silencing. Thus, its dysregulation up or down

can lead to abnormal gene expression and abnormal cellular proliferation. As mentioned earlier, DNMT3A establishes *de novo* methylation. Its mutation is seen in 11-13% of MDS patient and in higher proportion in AML patients (Ley et al., 2010). MDS patients with mutations of *DNMT3A* mutations correlate with a worse overall survival and a quicker progression to AML (Walter et al., 2011). The methylcytosine dioxygenase 2 (TET2) has an opposite function to DNMT3A by inducing oxidation of methylated cytosine and its excision. 30-50% of MDS patients and 30% of sAML patients present a deleterious mutation of TET2 (Jankowska et al., 2009). The correlation with the outcome is unclear as various studies came to different conclusions. While some studies showed a correlation with a favourable prognosis in MDS (Kosmider et al., 2009), it is associated with decreased overall survival in AML (Abdel-Wahab et al., 2009). Isocitrate dehydrogenases 1 and 2 act as co-factors of TET2 enzymes (Xu et al., 2011). Their mutation can modify their enzymatic capacity, leading to an increase of 2-hydroxyglutarate. 2-hydroxyglutarate is a competitive inhibitor of 2-ketoglutarte-dependant enzymes such as TET2; consequently, *IDH* 1 and 2 mutations result in a general increase of histone and DNA methylation (Xu et al., 2011; Figueroa et al., 2010). *IDH*s mutations are found in different proportion in myeloid malignancies: 5% of MDS (Medeiros et al., 2017), 9.7% of sAML and 20% of AML (Di Nardo et al., 2016). The impact of *IDH* mutational status is not clear yet as the nature of the mutation has an effect but it is associated to disease progression in both myeloproliferative neoplasm and high-risk MDS (Calvert et al., 2017). Histone modification as we explained earlier is a major mechanism of epigenetic regulation, and as such its alteration can affect cellular physiology. Enhancer Of Zeste 2 Polycomb Repressive Complex 2 Subunit (EZH2) is as its name indicates part of the polycomb repressive complex 2 (PRC2). Polycomb repressive complex 2 plays an important role on the deposition of the H3K27me2 and H3k27me3 repressive marks. The alteration of the EZH2 function through its gene mutation can

trigger aberrant expression of an oncogene or the silencing of a tumour suppressor gene (Lund, Adams and Copland, 2014). Loss of function mutation occurs in 5% of MDS and correlates with poor outcomes (Nikoloski et al., 2010). Polycomb repressive complex 2 interacts with additional sex comb-like 1 (ASXL1) to mediate deubiquitination of histone H2A monoubiquitinated, preventing gene silencing (Kweon et al., 2019). Nonsense and frameshift mutation of *ASXL1* cause a reduced expression of ASXL1, thus reducing the repressive activity of polycomb repressive complex 2 (Abdel-Wahab et al., 2012). *ASXL1* is mutated in 20% of MDS patients and 6-30% of AML patients. In both diseases it is associated with worse prognosis and outcome (Thol et al., 2011). Lastly, tumour suppressor 53 (TP53) is not an epigenetic regulator, but it is extremely relevant in both MDS and AML. TP53 is well known for its role in multiple cellular processes, such as cell cycle arrest, cell senescence, apoptosis and differentiation. Loss of function mutation of *TP53* has been shown to correlate with worse prognoses in MDS and in AML, no matter the co-alterations (Kulasekararaj et al., 2013).

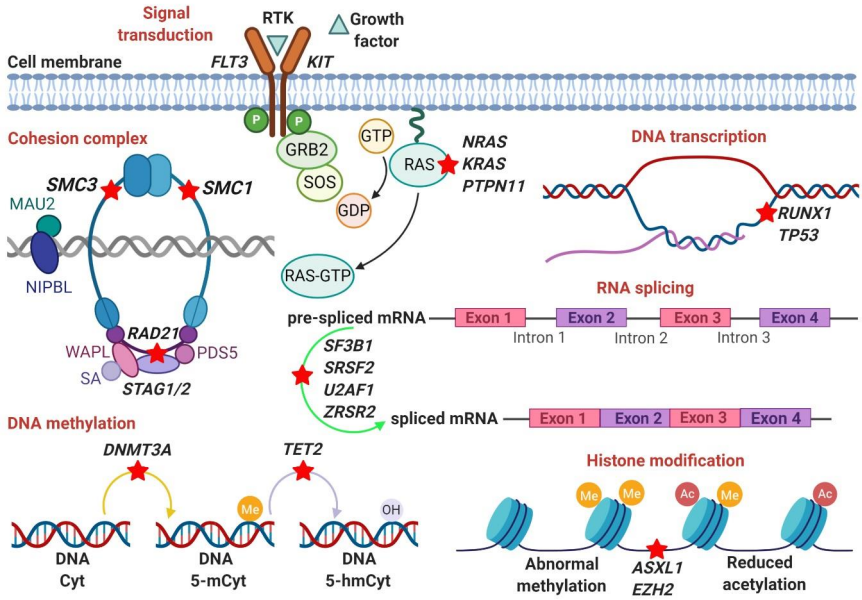


Figure 1.11: The altered genetic pathways in MDS. The figures illustrate the main genetic pathways involved in the pathogenesis of MDS including: Signal transduction (KIT, FLT3, RTK), Cohesion complex (RAD21, SMC1/3, STAG1/2), DNA transcription (TP53, RUNX1), RNA-splicing (SF3B1, SRSF2, U2AF1, ZRSR2), DNA methylation (DNMT3A, TET2), Histone modification (ASXL1, EZH2). From (Awada, Thapa and Visconte, 2020)

1.4.5 Treatments

The only curative treatment for MDS is allogeneic haematopoietic stem cell transplantation, but it comes with a significant morbidity (Saber and Horowitz, 2016) and is often unsuitable for elderly patients, especially when associated with comorbidities. Thus, alternative treatments are often used. When it comes to treatment strategy, MDS patients are divided based on the risk of their MDS transforming into AML (Chamseddine et al., 2016). TP53 mutation test or p53 immunostaining is also recommended for prognostic information. For "lower risk" patients, the aim is to manage the symptomatology, especially anaemia which affects 50% of patients (Bowen and Hellstrom-Lindberg, 2001). Red blood cell transfusions or erythropoietin are used to palliate anaemia (Cazzola, Della Porta and Malcovati, 2008; Gascón et al., 2019). Higher infection rates due to a weakened immune system are counter-balanced by an anti-microbial agent. Among the "lower risk" category, MDS patients with 5q deletion syndrome are treated by Lenalidomide as it increases apoptosis in malignant cells with del(5q). The European Medicines Agency only approved it for isolated del(5q) whereas the FDA approved its use across all MDS del(5q). Only over half of patients respond to Lenalidomide treatment (Giagounidis et al., 2014).

For "higher risk" situations, patients often undergo treatment by hypomethylating agent (HMA), before allogeneic haematopoietic stem cell transplantation when possible. The HMAs 5-azacitidine (Azacitidine), and 5-aza-2'-deoxycytidine (Decitabine) are called azanucleosides (AZN), nucleoside analogues. Since 1968, Decitabine, (**Figure 1.12 C**) has been identified as a suitable treatment for leukaemia in mice (Veselý, Čihák and Šorm, 1968). Since then, multiple studies have refined the use of Decitabine in treatment, uncovering its specific inhibitory effect on DNA methylation and therefore as a hypomethylating agent (HMA) (Vos and Overveld, 2005). In 2006 it was approved by the FDA as treat-

ment for MDS (Jabbour et al., 2008). It is often used in higher-risk MDS cases at a low-dose schedule of 15 mg/m2 every 8 h for 3 days repeated every 6 weeks (Lübbert et al., 2011), or with higher-dose intensity of 20 mg/m2 over 5 days repeated every 28 days (Kantarjian et al., 2007). Azacitidine (5-azacytidine) is also a azanucleoside-based therapy (**Figure 1.12 B**). Unlike Decitabine, it also is integrated in RNA, which interferes with protein synthesis (Gnyszka, Jastrzebski and Flis, 2013; Kaminskas et al., 2005). They are currently the best treatment option for "higher risk" cases in general, but half of the patients do not respond to AZN treatment and the other half eventually acquires a resistance to AZN. Global methylation in MDS patients has been shown to be a poor predictor of the response to AZN (Blum et al., 2007) and (Soriano et al., 2007). Thus, numerous studies have explored mechanisms of resistance to AZN or biomarkers to predict response to treatment. Expression level of nucleoside transporter proteins and enzymes involved in the metabolism of AZN have shown to correlate with the response to treatment. The cytidine deaminase/ deoxycytidine kinase ratio has been shown to be higher in non-responders to Decitabine (Qin et al., 2011). They are both proteins involved in the metabolism of decitabine but only cytidine deaminase degrades Azacitidine. Uridine-cytidine kinase 1 has been shown to have lower expression in non-responders to Azacitidine (Valencia et al., 2014), as well as Uridine-cytidine kinase 2, and its overexpression can restore sensitivity to treatment. (Gu et al., 2021). Compared to others, such as the previously mentioned AZN, Azacitidine has the particular ability to insert not only in DNA but also in RNA. Ribonucleotide reductase is made of the two subunit RRM1 and RRM2. The expression of BCL2L10, a member of the Bcl2 family preventing cell apoptosis, positively correlates with AZN resistance (Cluzeau et al., 2012).

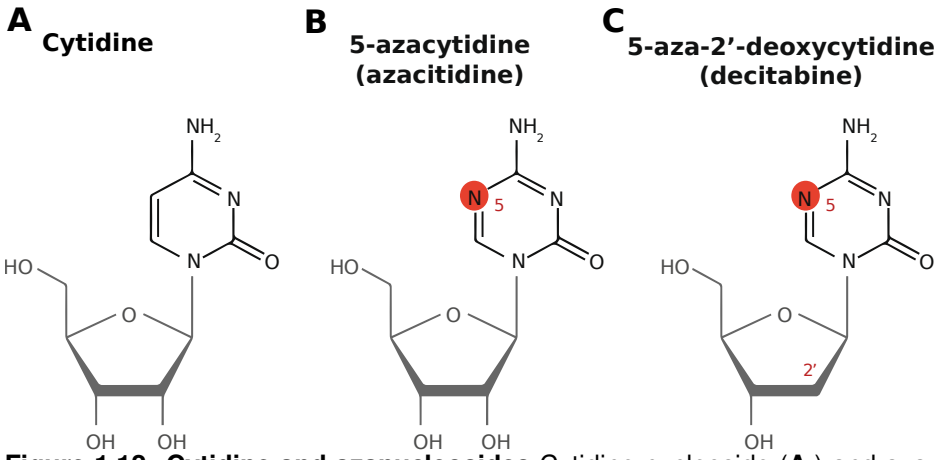


Figure 1.12: Cytidine and azanucleosides Cytidine nucleoside (**A.**) and azanucleoside (**B-C.**) chemical structures. Sugar moieties are indicated in grey and chemical changes between cytidine nucleoside and azanucleosides are highlighted in *red*. Taken from Diesch et al., 2016

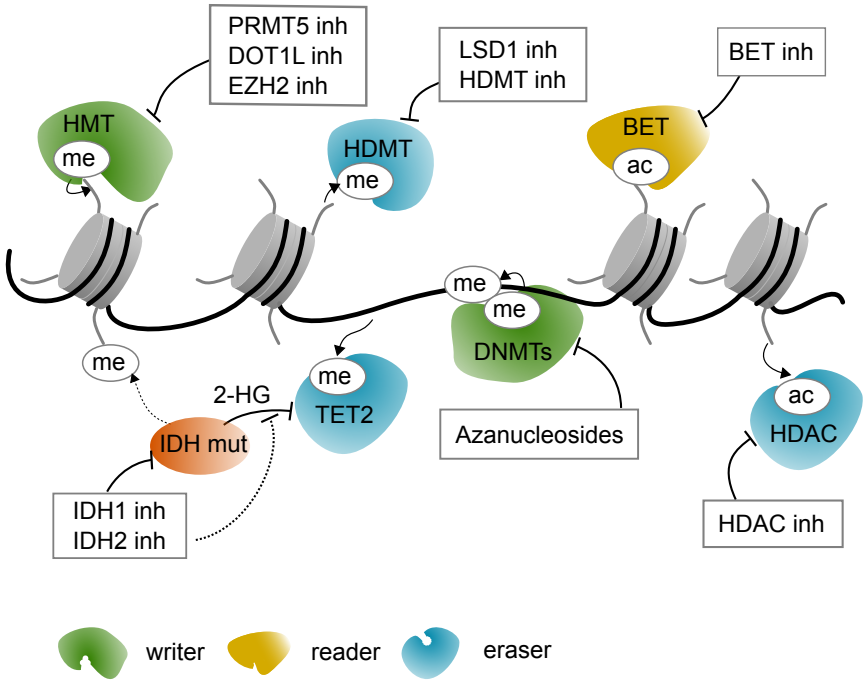


Figure 1.13: Overview of epigenetic drugs for MDS and sAML therapy. Many epigenetic enzymes are involved in the regulation of gene function. These can be broadly classified into "writers", which add specific marks to core histones, namely methyl (me) and acetyl (ac) groups; "readers", which identify the marks; and "erasers", which remove these marks. DNA methyltransferases (DNMTs) methylate DNA, thereby silencing certain tumour suppressor gene expressions. Hypomethylating agents, such as azanucleosides, are thought to reduce DNMT activity, thus reactivating silenced genes. Histone methyltransferase (HMT) inhibitors (inh) that target mutated PRMT5, DOT1L and EZH2 seek to re-stabilize perturbed histone methylation states. Histone deacetylase (HDAC) inhibitors restore histone acetylation, thus activating gene expression to promote differentiation and apoptosis. Ten-eleven translocation (TET) enzymes catalyze the demethylation of 5-methylcytosine to 5-hydroxymethylcytosine to induce active DNA demethylation. Isocitrate dehydrogenase (IDH) inhibitors reduce total serum levels of the oncometabolite, 2-hydroxyglutarate, restoring normal TET2 activity and DNA and histone methylation levels. Bromodomain and extra-terminal domain (BET) inhibitors mainly target BRD4, which normally promotes transcription of oncogenes, such as MYC, by binding acetylated histones. Taken from Maher et al., 2021

In parallel, more and more molecules are developed for more stratified treatment to answer to the heterogeneity seen in MDS. Among these, an increasing number of epigenetics drugs are interfering with enzymes affecting chromatin (Berdasco and Esteller, 2019). The **Figure 1.13** provides an overview of epigenetic drugs for MDS and sAML therapy. Some of these include inhibitors of histone methylation or demethylation, acetylation, but also of TET2 co-factors, IDH1 and 2 (Maher et al., 2021). Most of those therapies are currently used after failure of the transplantation or HMA treatments. There is still a need for better prediction of response and therapy in MDS.

Chapter 2

Rational and objectives

Rational and Objectives

The full spectrum of epigenetic implications in cellular processes is still not fully understood. The histone variant macroH2A has for example been shown to be involved in many biological processes but we lack an understanding of its function at the molecular level.

In myelodysplastic syndrome (MDS) there is increasing evidence of deregulation of genes coding for chromatin regulators. For MDS patients, the only curative option is allogeneic haematopoietic stem cell transplantation, but not all patients are eligible for it. All other high-risk patients are treated with nucleoside analogues such as Azacitidine. However, only the half of treated patients responds. Given that azacitidine acts in the context of chromatin, we hypothesized that chromatin regulators could affect the response to the treatment. Following this rationale, we assumed that a better understanding of chromatin regulation in the context of azacitidine treatment could potentially provide novel tools such as biomarkers and combinatorial drug targets that would allow to improve and personalize the for the treatment of MDS patients.

The aim of my PhD thesis has been to investigate the role of histone variant macroH2A and other chromatin regulators in genome regulation and response to drugs with the following specific objectives:

- To characterize the relation of the genomic distribution of macroH2A histone variants with repeat sequences and their regu-

lation by applying computational and experimental approaches.

- To functionally assess chromatin regulators including macroH2A histone variants as modulators of the response to treatments acting in the context of chromatin using MDS-derived secondary AML cell lines.
- To functionally assess chromatin regulators including macroH2A histone variants as modulators of the response to treatments acting in the context of chromatin using MDS-derived secondary AML cell lines.
- To analyse the correlation between mutation status, expression of genes encoding chromatin regulators and the response to Azacitidine in a cohort of MDS patients.

Following these three specific objectives, I have structured my PhD research in three parts and summarize the results in three dedicated chapters.

Chapter 3

Results

Results

3.1 Results I: macroH2As association with LINEs *in silico*

Based on previous results obtained by the group, we knew that a knock-down of all macroH2As in the hepatocarcinoma cell line, HepG2, leads to a global disruption of chromatin organization (**Figure 1.5**). A substantial portion of the genome is constituted by various categories of repeats, we hypothesized that such a global phenotype might be linked to interactions between macroH2As and repeat sequences. We have chosen HepG2 cells for our analysis as they are particularly well characterized as one of the main cell lines of the ENCyclopedia Of DNA Elements Project (ENCODE). It includes epigenetic marks data (Sloan et al., 2016).

3.1.1 MacroH2As are associated with LINEs and simple repeats

The first step of our study was to confirm in an unbiased way if macroH2As were interacting with any type of repeat sequences. Chromatin immunoprecipitation-sequencing (ChIP-seq) profiles give information on the local interaction of proteins with DNA. Skimming over macroH2As ChIP-seq profiles in HepG2 cells, we observed what seemed like a recurrent presence of long inter-

spersed nuclear elements (LINEs) in areas covered by macroH2As enrichment. An example is shown in **Figure 3.1 A**. To confirm this observation, we performed an association analysis of macroH2As peaks with annotated repeats. An association study calculates the overlap between two region sets and compares it to the overlap obtained between the first element and a set of randomized genomic positions of the same length and distribution than that of the second set. Here the first set is made up of regions of macroH2A enrichment defined as ChIP-seq peaks (from Douet et al., 2017). The second set was extracted from the annotation data from RepeatMasker (Smit, Hubley and Green, 0), sorted between small nuclear RNA, transfer RNA, signal recognition particle RNA, simple repeat, satellite, long terminal repeat, short interspersed nuclear element, long interspersed nuclear element (LINE). Then we used regionR (Gel et al., 2015) to do an association analysis, z-scores were normalized by $\sqrt{n}$, n being the number of elements in each category (**Table 3.1** second column). The z-score evaluates the strength of the association while the adjusted p-value evaluate the statistical significance of the result. Thus, if the adjusted p value is not significant the z-score is considered null (**Table 3.1**). Based on this association analysis, both macroH2A1 and macroH2A2 are associated with simple repeats, LINEs, and long terminal repeats. MacroH2A1 is also associated with short interspersed nuclear elements (**Figure 3.1 B**).

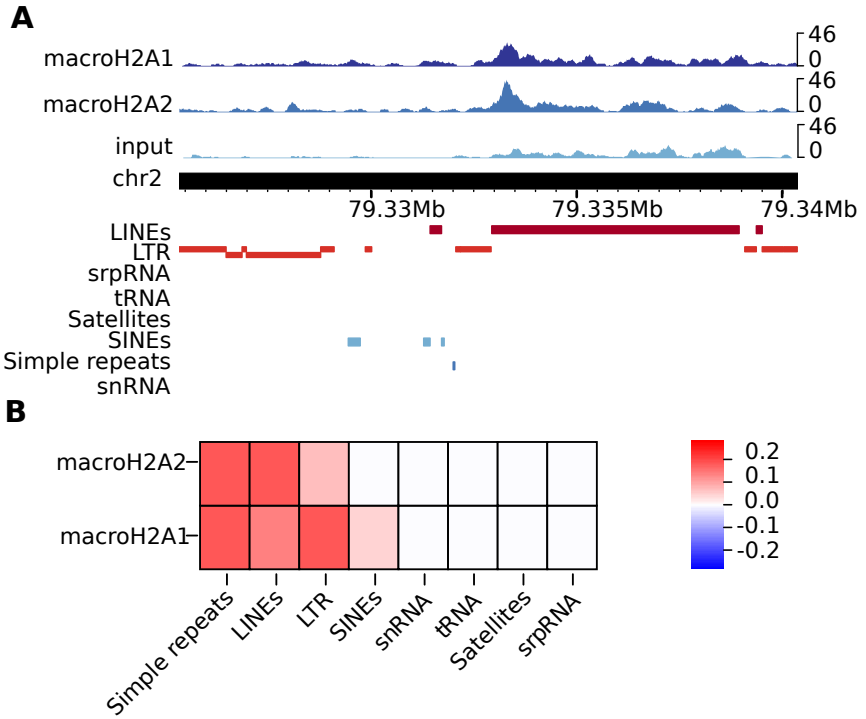


Figure 3.1: MacroH2A1/2 ChIP-seq shows enrichment on LINEs. A. ChIP-seq profiles of macroH2A1 and macroH2A2 and input, as well as genome repeats annotation according to RepeatMasker (Smit, Hubley and Green, 0) **B.** Association of macroH2A1 enriched regions in HepG2 with repeat classes indexed in Repeatmasker. Association was calculated with the permutation-test based R-package regioneR (Gel et al., 2015) performing 10000 permutations using circularRandomizeRegions. The z-score used was normalized by dividing by $\sqrt{n}$ in which n refers to the number of regions in the region sets tested. Colour grading correlates with normalized z-score and blue and red colours denote negative and positive association, respectively.

		macroH2A1		macroH2A2	
name	**n element**	**adjusted p-value**	**norm. z-score**	**adjusted p-value**	**norm. z-score**
snRNA	4386	0.8671	0*	0.2324	0*
Simple repeat	417913	0.0008	0.2775	0.0008	0.2282
SINEs	1793723	0.0008	0.1827	0.0008	0.071
Satellite	9566	0.1592	0*	0.5609	0*
tRNA	2002	0.5012	0*	0.5609	0*
srpRNA	1481	0.8671	0*	0.561	0*
LTR	717656	0.0008	0.0457	0.1685	0*
LINEs	1498690	0.0008	0.1403	0.0008	0.2788

Table 3.1: Computational results of overlap analysis between macroH2A peaks and repeats. Normalisation of the z-score is obtained by $\sqrt{n}$, n being the number of elements in each category. Normalised (norm.) z-score* values automatically set to 0 as adjusted p-value < 0.01.

3.1.2 The association is less evident *in vitro*

To confirm the local enrichment of macroH2A1 and macroH2A on sequences annotated as LINEs and long terminal repeats, we used ChIP coupled to semiquantitative PCR (ChIP-qPCR) on the same cell line, HepG2. ChIP-seq is a whole genome sequencing technique whereas ChIP-qPCR uses primer pairs to target specific sequences and results in the local relative enrichment. As we are studying sequences in many copies and with variations, we had to first properly validate the primers found in the bibliography. In order to assess the specificity of the results we used primers amplifying various type of repeats: 5S ribosomal DNA (Shen et al., 2013), DXZ4, a repeat present in the X chromosome and marked by Histone 3 lysine 9 trimethylation (H3K9me3) in male cells (Chadwick, 2008), long terminal repeats 1 (Montoya Durango et al., 2009), pericentromeric satellite repeats (SAT2, αSAT) (Zeng et al., 2009), and LINEs (Douet et al., 2017). We used a serial dilution of genomic DNA and set the qPCR reaction to assess the conditions and the quality of the primer pairs.

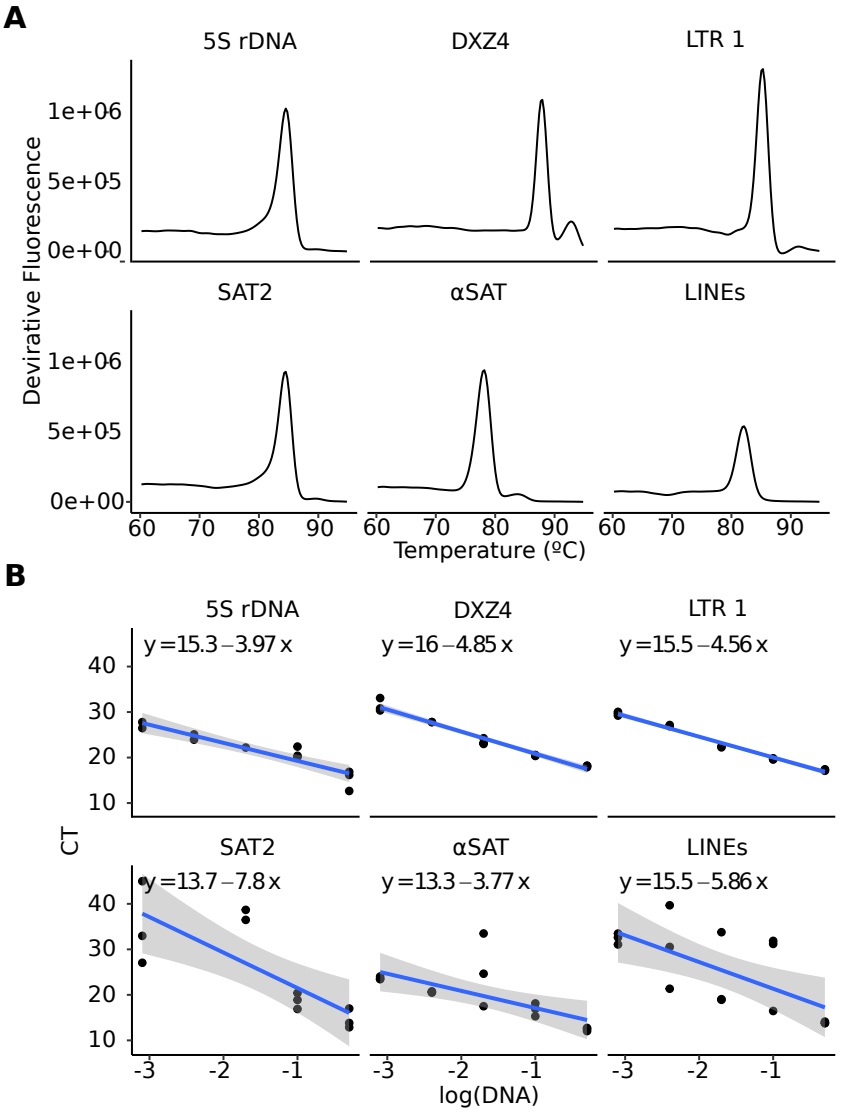


Figure 3.2: Validation of primers for RT-PCR quantification. A. Melting curve of 6 different primer sets, each specific to repeats: 5S ribosomal DNA (rDNA), DXZ4, long terminal repeats (LTR) retrotransposons, pericentromeric satellite repeats (SAT2, αSAT), and long interspersed nuclear elements (LINEs) **B.** Standard curves for each primer set were obtained using gDNA serial dilutions.

A single peak in the melting curve confirmed the amplification of a sole product by each primer pairs (**Figure 3.2 A**). The standard curve allowed to confirm that amplifications were efficient and dose dependent. The best fit line equations were calculated for each standard curve, the slope designates the geometric efficiency allowing to calculate the real efficiency used for ChIP-qPCR analysis (**Figure 3.2 B**).

To ensure the specificity of the results, we used both HepG2 control cells (control shRNA) and HepG2 deficient of all macroH2A proteins (*MACROH2A1/2* shRNA) that were previously validated ((Kozlowski et al., 2018)). We used ChIP-qPCR targeting histone H3 to measure nucleosomal occupancy and did not detect any difference between both cell lines (**Figure 3.3 C**). Histone H3K9me3 was used as an additional control as it is known to be enriched on repetitive sequences. We can see that normalised enrichment for histone H3K9me3 was independent from the expression status of macroH2As except for the negative control for which the signal drops in macroH2As deficient HepG2 cells. All sequences, except SAT2 and LINEs, showed a certain enrichment of macroH2As (**Figure 3.3 D**). To evaluate the level of enrichment we used two single copy references. The first locus is known to be enriched (positive control) and the second locus (negative control) for macroH2As enrichment ((Douet et al., 2017)). The normalised enrichment ratios were significantly lower in the knock-down cell line when using antibody against either macroH2As (**Figure 3.3 A-B**), confirming the specificity of macroH2As antibodies. Since we were using enrichment normalised for 100% coverage, results should be the same for both single copy and multi-copy targets, but none of the targets showed levels of enrichment close to the positive control. Also, to be noted, due to high variability of the measurements between experiments, the normalised enrichment ratios were not statistically different between both cell lines. These results did not allow to confirm nor to disprove the potential association of macroH2A with repeats suggested by the bioinformatic

association study.

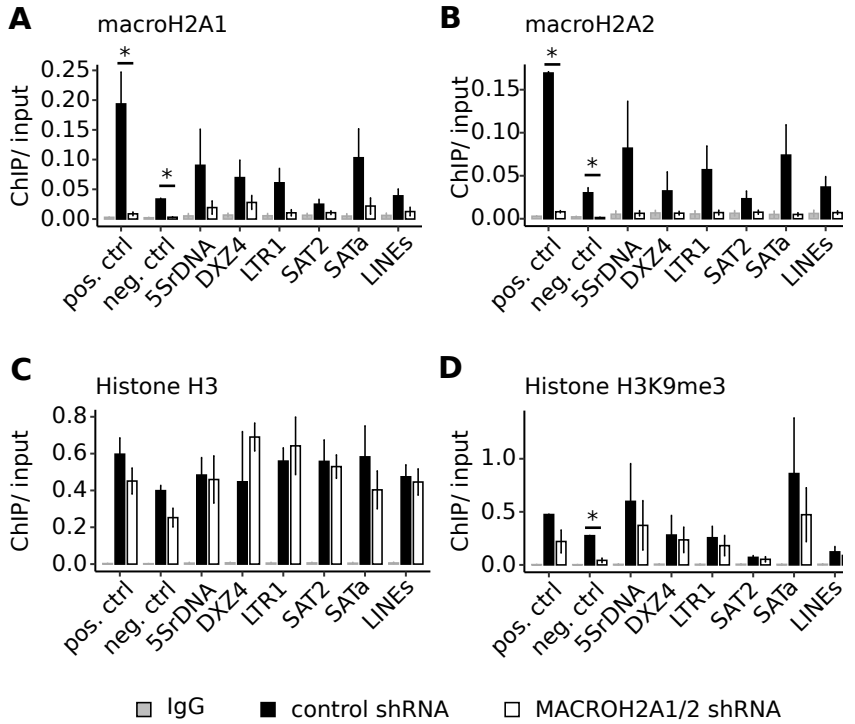


Figure 3.3: Protein enrichment on repeat sequences. Histone variant macroH2A1(**A**), histone variant macroH2A2 (**B**), histone H3 (**C**), and histone H3K9me3 (**D**) occupancy on various repeats was analysed by ChIP-qPCR in HepG2 control cells (control shRNA) and HepG2 cells deficient of all macroH2A proteins (*MACROH2A1/2* shRNA). The upstream promoter regions of *LAMA5* and *GRXCR1* served as positive and negative controls, respectively, for macroH2As-enrichment. IgG was used as background control. Data is shown as mean + s.e.m. (n=3). p-values by T-test; * $p<0.05$.

3.1.3 MacroH2As are overlapping with LINEs' youngest element

The high variation between experiments pointed toward a technical issue regarding the selected targets due to the elevated number of copies in the genome. Thus, we decided to look at the association of macroH2As with repeats in greater details to identify with which subfamilies macroH2As overlap more precisely. LINEs duplicated and their sequences naturally drifted over time (Wagstaff, Barnerβoi and Roy-Engel, 2011).

We focused on the youngest LINEs (L1PAs) as older elements have larger sequence variability. We first started with a computational approach and later used the results to guide the *in vitro* experiments. L1PAs were sorted between the different subfamilies, each subfamily still retaining a high number of copies (**Table 3.2**). As mentioned in the introduction L1PA1 is also called L1HS and designates the most recent elements, whereas L1PA17 encompasses the oldest L1PA elements (**Figure 3.4 A**). We used regionR to perform a local permutation-based association test of each subfamily with macroH2A1 (Gel et al., 2015). As stated above this

L1PA subfamily	number of elements
L1HS	1544
L1PA2	4917
L1PA3	10645
L1PA4	11921
L1PA5	11338
L1PA6	5977
L1PA7	13043
L1PA8	8128
L1PA8A	2490
L1PA10	7176
L1PA11	4184
L1PA12	1781
L1PA13	9093
L1PA14	3070
L1PA15	8445
L1PA15-16	1410
L1PA16	14051
L1PA17	4866

Table 3.2: Counts of L1PAs subfamilies based on RepeatMasker data

test calculates local overlap of
macroH2As on the set of regions compared to a random set of regions with similar characteristics, here we also shifted the analysis on both sides recalculating in bins of 1000bp over 10 000bp on each side. A positive association has a z-score above 2 and a negative association has a z-score under -2, anything in between is considered as no association. We found that macroH2A1 showed a rather narrow local association with most L1PAs (**Figure 3.4 B**). The trend decreased with increasing age of subfamilies, from L1PA8 to L1PA17 there was less and less overlap.

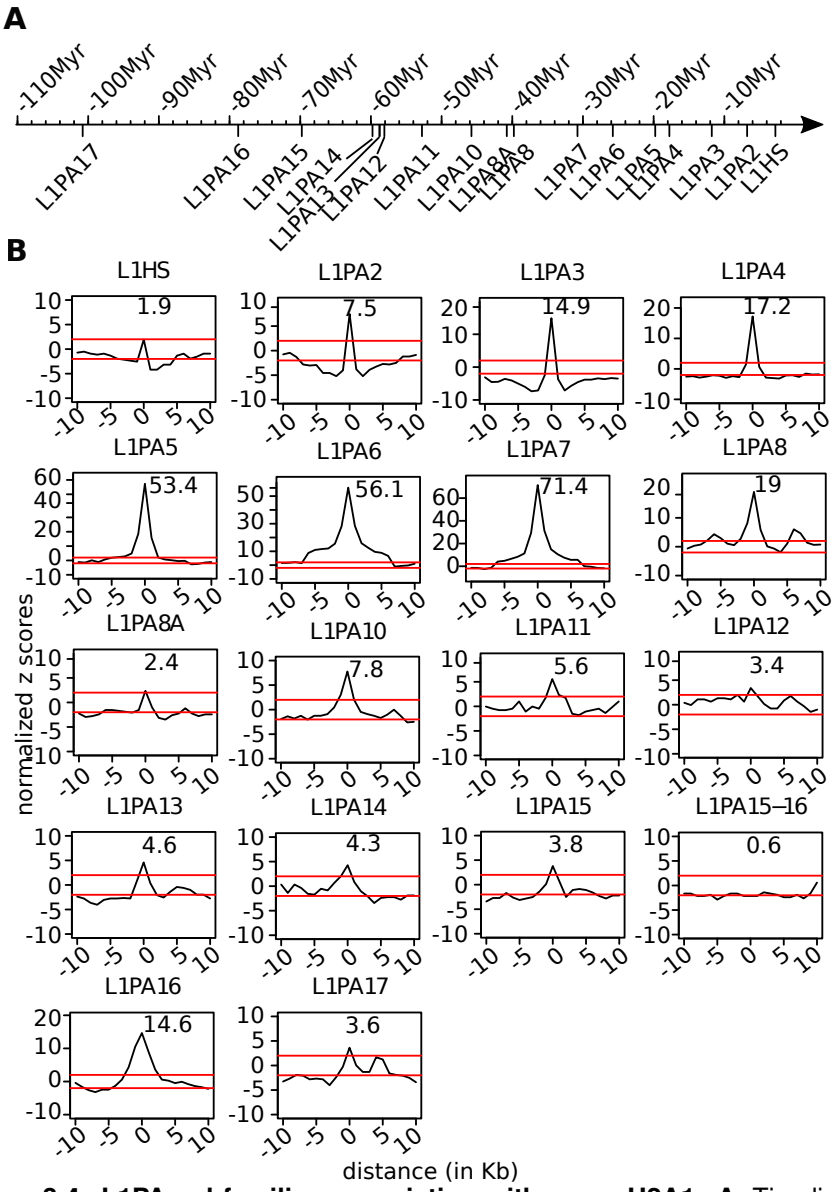


Figure 3.4: L1PA subfamilies association with macroH2A1. A. Timeline of the apparition of L1PA subfamilies **B.** Associations of macroH2A1 with each L1PA subfamilies in a window of 10kb and a permutation of 1000 times represented by the shifted z-score, red lines set at 2 and -2 indicating threshold of association and anti-association.

All these results suggested that the overlap was not on a broader area but instead very localised. We decided to look closer at the position of macroH2As' enrichment on the L1PA sequences and to assess if we could narrow down to a section of the L1PA sequences. We based our analysis on a published method (Z. Sun et al., 2018), and used reported consensus sequences for L1PAs and L1PBs (Khan, Smit and Boissinot, 2006). The analysis can be divided into five main steps. First, we performed the standard ChIP-seq alignment on the unmasked human genome. In this process only unique reads were aligned and reads with more than one possible alignment were randomly distributed between the different possible loci. Second, we extracted all reads aligning on sequences annotated as LINEs by RepeatMasker to obtain a library of reads selected by their positional assignment. Third we proceeded to a stringent filtering of reads excluding any read with more than 3 mismatches, only partially aligned (soft clip) or aligned on two positions for over 48bp. This step aimed at eliminating reads that could be from similar but non identical sequences. Fourth we re-aligned the reads on a virtual genome made up only of the consensus sequences of L1PAs and L1PBs to generate a coverage profile. The last step was to use input to normalize to remove background signal (**Figure 3.5 A**). In our analysis, we also included CTCF ChIP-seq data as a control data set which is known not to be overlapping with L1PAs, thus could be used to assess the specificity of enrichments. Newly analysed ChIP-seq data showed an enrichment of both macroH2As,higher for macroH2A2, on the 3'UTR of the youngest LINEs and no such enrichment was found on the oldest ones (**Figure 3.5 B**). The absence of enrichment of CTCF on LINE-1 was confirmed. These results confirmed the results of the overlap study macroH2As overlap with younger LINE-1 and it also positioned more precisely this overlap to the 3'UTR end of these LINE-1.

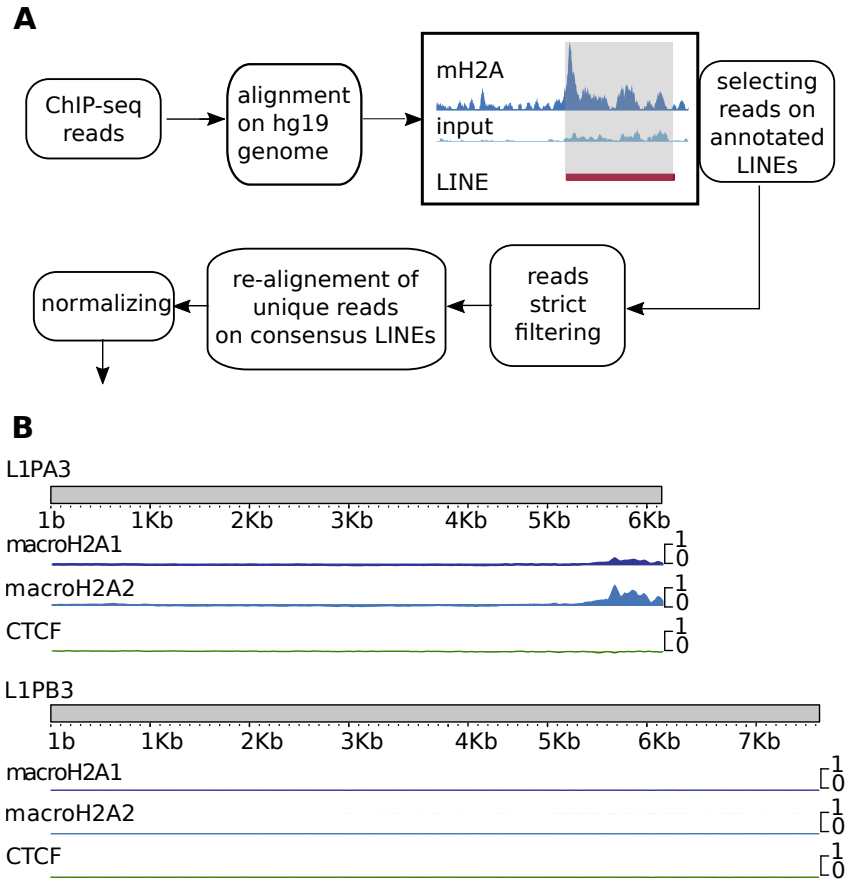


Figure 3.5: ChIP-seq reanalysis on LINE-1 consensus sequences. A. Schematics of the pipeline of based on Z. Sun et al., 2018. ChIP-seq data are first aligned to the human genome (hg19). Reads aligned on LINE annotated region (grey box) are then extracted. Reads with partial alignment (soft clipping), more than three mismatches, or a secondary alignment of over 48 bases are removed. Filtered reads are subsequently re-aligned to a virtual genome consisting of LINEs consensus sequences to obtain compiled reads. Input is used for normalization. **B.** Coverage profiles for macroH2A1, 2 and CTCF ChIP-seq on L1PA3 and L1PB3 consensus sequences as example of young and older LINE-1.

3.1.4 Previous macroH2A specific enrichment on 3'UTR of younger LINEs unconfirmed

Based on this novel analysis, we chose to perform a ChIP-qPCR targeting each end of L1PA3 and a L1PB3 as a negative control to confirm the results *in vitro* (**Figure 3.6 A**). As previously done in Figure 3.2, we first validated amplicon melting curve for unique peak (**Figure 3.6 B**). We calculated the real primer efficiency based on the standard curve using genomic DNA as template (**Figure 3.6 C and Table 3.3**). We also confirmed that with a smaller number of potential targets the amplification gave more reliable quantification with less variability.

	start	**end**	**amplicon length (bp)**	**calculated efficiency**
L1PA3 3' UTR	5737	5973	236	1.91
L1PA3 5' UTR	722	987	265	2.14
L1PB3	7027	7208	181	2.02

Table 3.3: L1PA primer pairs position and efficiency

We then performed ChIP-qPCR using the validated primer pairs to validate the *in silico* results showing an enrichment of macroH2As on the 3'UTR of younger LINE-1. H3 immunoprecipitation did not any show difference in nucleosomal occupancy between aither HepG2-derived cell lines (**Figure 3.7 C**). H3K9me3 immunoprecipitation showed much higher normalized enrichment for L1PA3 5'UTR position than single copy controls but with also a high standard deviation thus statistically non-significant compared to signal in the macroH2As deficient HepG2 cells (**Figure 3.7 D**). MacroH2A1 immunoprecipitation showed a small but specific enrichment in L1PA 3'UTR. For the other two LINE targets, the variability was too high to conclude on local enrichment as there was no statistical difference between control HepG2 and macroH2As deficient HepG2 cells (**Figure 3.7 A**). For macroH2A2, immunoprecipitations gave the same profiles of local enrichment as for

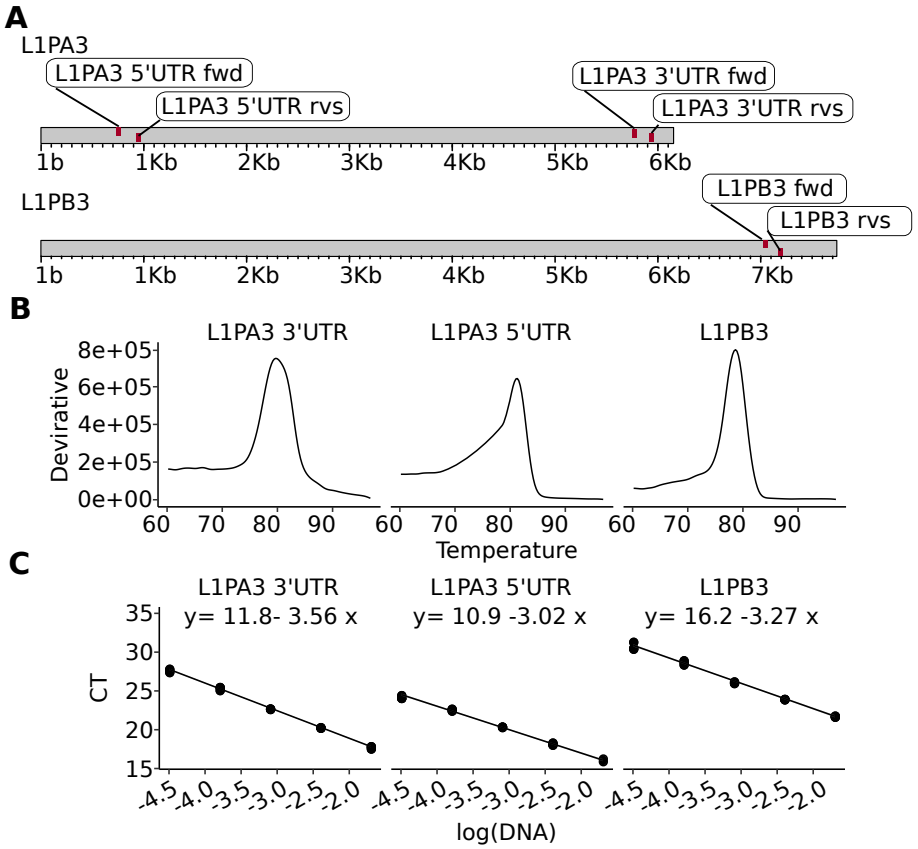


Figure 3.6: Validation of primers for qPCR quantification based on LINE-1 consensus sequence. A. Positioning of three primers set, each specific to one subfamily of LINEs and one position based on consensus sequences. **B.** Melting curve of the three different amplicons. **C.** Standard curves for each primer set was obtained using gDNA serial dilutions.

macroH2A1, but standard deviation was high and resulted not statistically significant (**Figure 3.7 B**). The absence of statistical difference of enrichment between both HepG2 cell lines when using macroH2As antibodies for repetitive sequences may indicate a technical limitation. In conclusion these experiments did not allow to conclude on the specificity of macroH2As enrichment on the 3'UTR of younger LINEs.

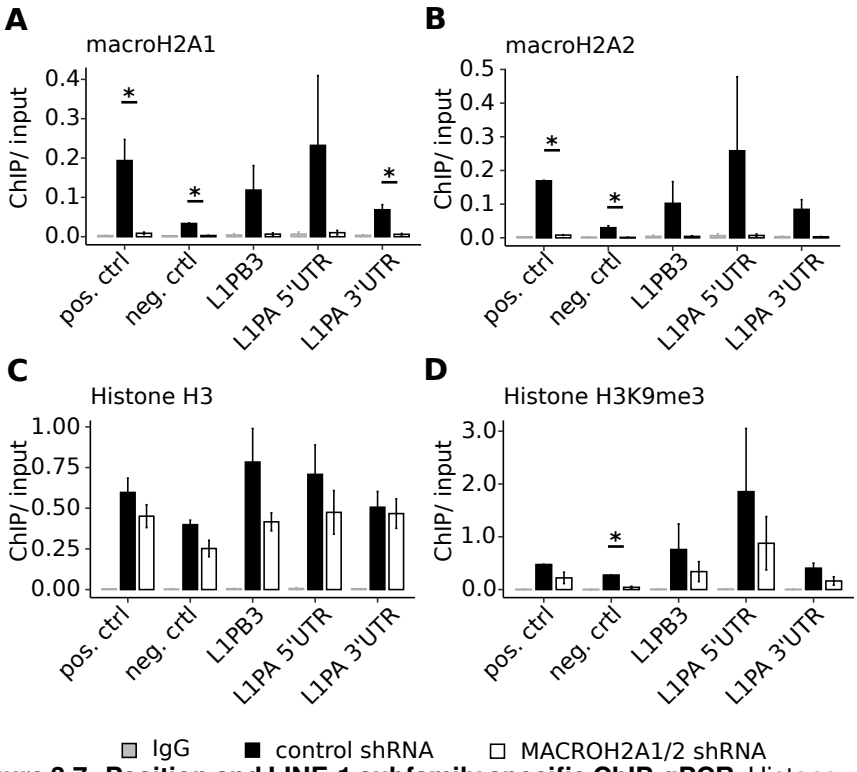


Figure 3.7: Position and LINE-1 subfamily specific ChIP-qPCR. Histone variant macroH2A1(**A**), histone variant macroH2A2 (**B**), histone H3 (**C**), and histone H3K9me3 (**D**) on various position of LINE-1 repeats were analysed by ChIP-qPCR in HepG2 control cells (control shRNA) and HepG2 cells deficient of all macroH2A proteins (*MACROH2A1/2* shRNA). The upstream promoter regions of *LAMA5* and *GRXCR1* served as positive and negative controls, respectively, for macroH2As-enrichment. IgG was used as background control. Data is shown as mean + s.e.m. (n=3). p-values by T-test; * $p<0.05$.

3.1.5 MacroH2A deficiency has no effect on LINE-1 expression

The next question was if the change of chromatin architecture upon the knock-down of macroH2As in HepG2 correlates with an effect on the expression levels of LINE-1. We were particularly interested in younger LINEs since the ChIP-seq data suggested a preferential enrichment of macroH2A. To address this question, we decided to measure the mRNA level in control cells and macroH2As deficient cells. LINEs are multi-copies and there is no exon/intron structure thus primers amplifying LINEs' mRNA also detect the considerable number of genomic DNA copies. Therefore detection of mRNA levels requires additional technical controls. The pair of primers we used targeted mostly the 5'/ORF1 half of the mRNA as it has the least number of DNA copy due to the transposition mechanism (**Figure 1.8**). We selected the primers from the literature and confirmed amplification of single amplicon by plotting the melting curve (**Figure 3.8 A**). When performing the cDNA synthesis, we added an additional control sample without retrotranscriptase, the difference of detection between samples with or without retrotranscriptase was then compared. Both housekeeping single copy genes showed a major difference between both samples, whereas the gap was, as expected, much smaller for LINEs primers but still statistically significant (**Figure 3.8 B**). Although the experimental conditions were not ideal, they still provided an adequate tool to compare the expression of LINEs in both cell line.

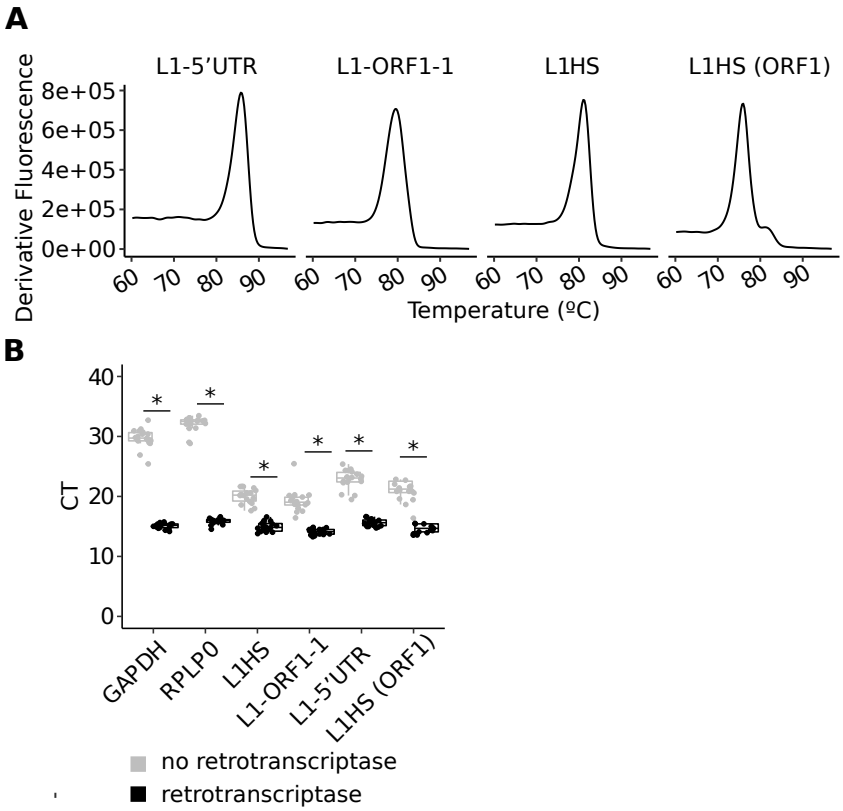


Figure 3.8: Validation of primers for mRNA quantification of LINE-1 expression. A. Melting curve of the four different amplicons of LINEs' mRNA. **B.** Cycle threshold (CT) comparison from samples prepared with or without retrotranscriptase, to determine DNA contamination in samples. Data is shown as mean + s.e.m. (n=3). p-values by T-test; * $p<0.05$.

As explained in the introduction only L1HS are considered to be active LINE-1 and, when activated, are transcribed in a unique mRNA. Thus, all four validated primer pairs allowed the quantification of the same mRNA. When we compared the mRNA level using the four primer pairs, we could see that three out of four did not show a statistical difference of transcript abundance between both HepG2 derived cell line (**Figure 3.9**). Unexpectedly the fourth one showed a slight, but statistically significant, decrease of transcript abundance in macroH2A-deficient cells.

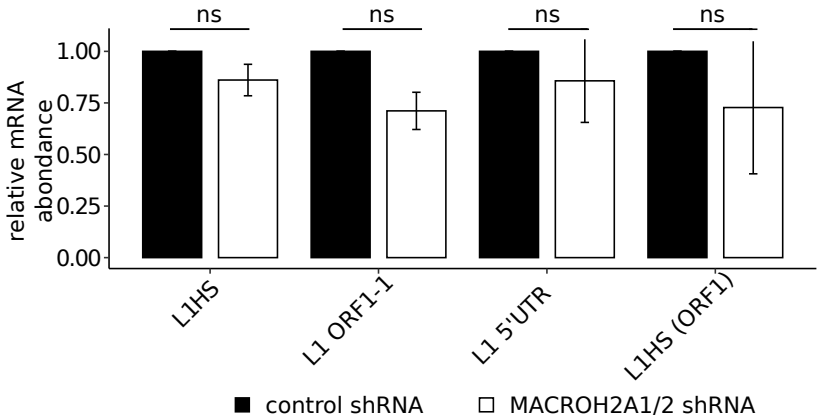


Figure 3.9: Relative LINE-1 mRNA abundance in HepG2 control cells and HepG2 cells deficient of all macroH2A proteins. Relative expression of RNA measured with four different primer pairs and determined by RT-qPCR. Values are normalized to the endogenous control genes GAPDH and RPLP0. Data is shown as mean + s.e.m. (n=3), statistics based on T-test

3.1.6 Result I: conclusions

Overall analysing previously generated ChIP-seq data, we found evidence suggesting an association between macroH2As and both LINEs and simple repeats. More specifically, we found a preferential association with the 3' end of younger members of the L1PA sub family. However, we were unable to reproduce these results *in vitro*. We consider two possible explanations for this discrepancy. Either the nature of these repeats' sequences or the incompleteness of the reference genome sequence introduced a bias generating an artefact at the level of the bioinformatic analysis. The applied experimental tools were neither specific nor sensitive enough to detect the multi-copy loci of interest.

Contributions This work was conceived by Marcus Buschbeck, Roberto Malinverni, and me, it was supervised by Marcus Buschbeck, Roberto Malinverni and David Corujo. All wet-lab experiments were conceived and performed by me with some support from David Corujo. Apart from the ChIP-seq reanalysis performed by Roberto Malinverni, all analysis were done by me under the guidance of Roberto Malinverni for the computational ones.

3.2 Results II: chromatin regulators as potential therapeutic targets in clonal hematopoietic stem cell diseases

As mentioned in the introduction, to this day the only curative treatment for MDS patients is allogeneic haematopoietic stem cell transplantation, but not all MDS patients are eligible for it. For high-risk MDS patients, the current best alternative is Azacitidine treatment to which only half of the patients responds. In this section I will explain our *in vitro* approach to find a better treatment strategy for these patients. Azacitidine is an ”epidrug” thus we decided to explore in a less biased approach the therapeutic potential of epigenetic regulators.

3.2.1 An unbiased approach to finding new therapeutic targets

We performed a loss of function short hairpin (shRNA) screen using the hEpi9 library of shRNAs (Diesch et al., 2021). The library targets 912 epigenetic regulatory and chromatin remodeler genes with 8 shRNA for each target. The mir-E shRNA expression vector cSGEP was used for genome integration. This vector includes a ubiquitous chromatin opening element (UCOE), a spleen focus-forming virus (SSFV) promoter followed by the reporter GFP gene and the short hairpin sequence, a human phosphoglycerate kinase 1 promoter followed by a puromycin resistance gene (**Figure 3.10 A**). The UCOE element combined with the SSFV promoter has an anti-silencing effect allowing a higher probability of expression of both the GFP and the shRNA independently of the site of integration into genomic DNA of the target cell (Müller-Kuller et al., 2015). SKK-1 cells were infected with the vector at extremely low viral titre to ensure that the rate of infection was of maximum

one shRNA per cell to avoid double knock-downs. After puromycin selection cells were split in two, one set being treated with Azacitidine every two days, the other set left untreated (**Figure 3.10 B**) (3 replicates for each condition). In both conditions (treated and untreated), cells were harvested after 21 days, genomic DNA extracted followed by library preparation and high-throughput sequencing of shRNA guide strands. We treated cells with a low dose of Azacitidine (0.075µM) to deplete cells with higher sensitivity to the treatment while limiting the effect on other cells. Bioinformatic analysis of the sequencing data were done by comparing the abundance of each shRNA sequence in treated versus untreated samples. The genes could fall in four distinct categories: knock-down deleterious for cell survival (absent from both conditions), knock-down inducing resistance to Azacitidine treatment (increased in treated cells), knock-down sensitizing to Azacitidine treatment (decreased in treated cells), and knock-down not affecting response to treatment (equal proportion in both conditions).

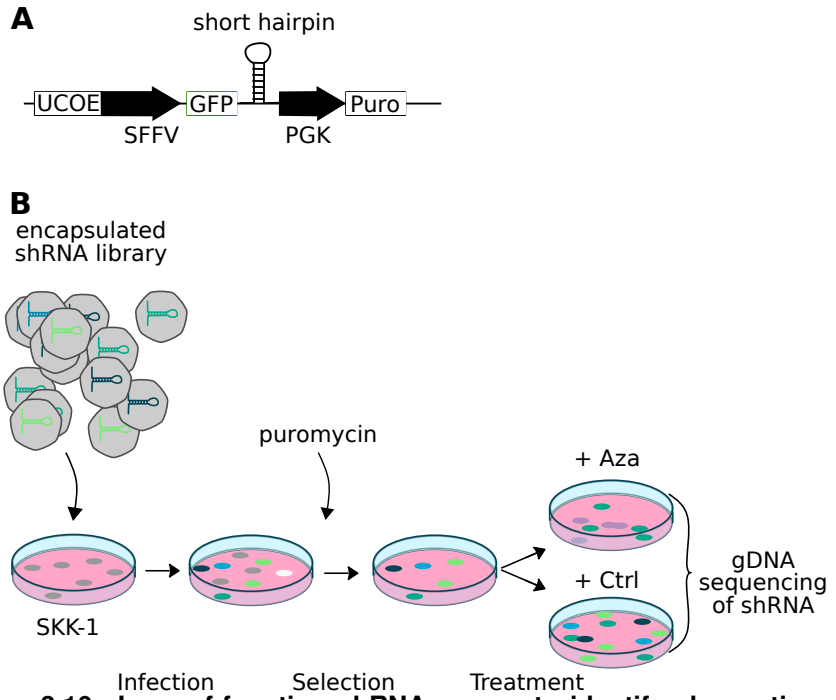


Figure 3.10: Loss-of-function shRNA screen to identify chromatin regulators sensitizing to Azacitidine treatment. **A.** Visual representation of the key element of the lentiviral cSGEP backbone vector used for the screen. It includes the ubiquitous chromatin opening element (UCOE) and schematic of the cSGEP backbone vector containing the UCOE, spleen focus-forming virus promoter (SFFV), GFP fluorescent marker gene, shRNA integration site, human phosphoglycerate kinase 1 promoter (PGK) and puromycin (puro) resistance gene. **B.** Workflow of the loss-of-function screen in SKK-1 cells using shRNA library targeting genes encoding chromatin regulators.

Hits were determined based on the fold change of abundance, using all short hairpins targeting the same gene and following the same trend. Genes of interest for this thesis are highlighted in **Figure 3.11**. N-alpha-acetyltransferases, *NAA15*, a component of the NatA complex binding to ribosomes and co-translationally acetylating proteins was found among the top hits of increased abundance in the Azacitidine treated cells compared to non-treated. And thus, its depletion led to resistance to Azacitidine in SKK-1 cells. Conversely, within the top hits of decreased abundance, we found several genes encoding for proteins of the CHRAC complex, BAZ1A, BAZ1B, CHRAC1, SMARCA5 and POLE3, indicating that the depletion of these genes increases sensitivity to the treatment (**Table 3.4**).

Based on these preliminary results, we hypothesized that targeting the CHRAC complex could be a strategy to sensitize cells to Azacitidine treatment.

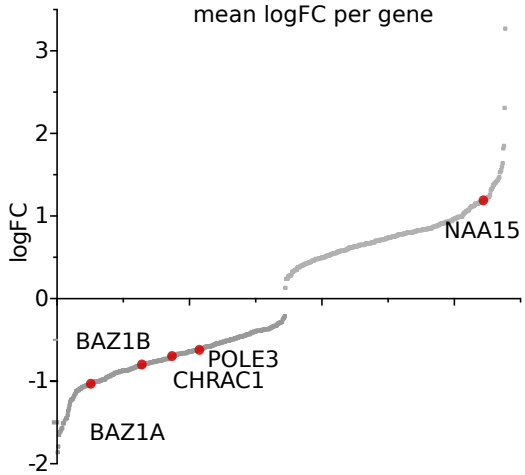


Figure 3.11: Hits from the loss-of-function screen in SKK-1 cells. Plot of genes corresponding to enriched or depleted shRNAs. Hits which have been further characterized, are named.

Gene	Mean logFC
BAZ1A	-1.031
SMARCA5	-0.981
BAZ1B	-0.797
CHRAC1	-0.699
POLE3	-0.6181
NAA15	1.021

Table 3.4: Table of the mean of log of fold change of the different shRNAs following the same trend for each hit.

To confirm hits from the loss-of-function screen, we used the same approach as for the screen to generate SKK-1 cell lines with stable knock-downs for two selected hits of interest *NAA15, BAZ1A*. cSGEP *renilla* shRNA (ctrl) served as control for the integration. We decided to also include in our study shRNAs targeting *MACROH2A1* (not included in the hEpi19 library), as it is located on the long arm of chromosome 5 (at 5q31.1), which is depleted in 20 percent of MDS patients (Haase et al., 2007). Three days after infection, cells with integrated vector sequence were selected using puromycin, and success of the selection was assessed by measuring the percentage of cells expressing GFP. A similar vector but with a neomycin resistance gene instead of puromycin (SGEN vector) was used as control for the selection, and therefore cells with integrated SGEN vector were killed by puromycin treatment. Knock-down efficiency of two separate shRNA (1 and 2) against *MACROH2A1* was assessed by mRNA level measurement (**Figure 3.12 A**) using three separate sets of primers to measure separate splicing-isoform-specific mRNA as well as overall mRNA level. The mRNA level of control cells was set to 1 for comparison. SKK-1*MACROH2A1* shRNA 1 and 2 cells showed a strong reduction of *MACROH2A1* mRNA level compared to control (around 12%) which came from a reduction of both isoforms *MACROH2A1.1* (reduction of over 70%) and *MACROH2A1.2* mRNA (reduction of over 90%) in three separate measurements (**Figure 3.12**

A). At the protein level, this reduction of expression was confirmed for both SKK-1 *MACROH2A1* shRNA cell lines for overall macroH2A1 as well as macroH2A1.2. Western blot showed no detectable level of macroH2A1.1 in control and knock-down cell lines (**Figure 3.12 E**). Similarly, knock-down efficiencies for both *BAZ1A* shRNAs were assessed at the mRNA level (**Figure 3.12 B**), and protein level (**Figure 3.12 D**). A statistically significant decrease of 30% can be seen in SKK-1 shRNA 1 cells at the RNA level, and the protein level of BAZ1A is decreased in both knock-down cell lines. *NAA15* mRNA level was decreased below 40% by integration of corresponding shRNA 1 (**Figure 3.12 C**).

Figure 3.12: A-C. Knock-down efficiency of shRNAs measured by RT-qPCR, ctrl cells with cSGEP renilla shRNA were set to 1. Data represent the mean ±SEM of three independent measurements. **A.** Relative expression to control cell line of *MACROH2A1* overall, *MACROH2A1.1* and *MACROH2A1.2* splicing isoforms with two separate shRNAs. n=3. **B.** Relative expression of *BAZ1A* with two separate shRNAs. n=3. **C.** Relative expression of *NAA15* with a shRNA. n=3. **D-E.** Western blot using SKK-1 cells with stable integration of BAZ1A shRNAs (**D.**) MACROH2A1 shRNAs (**E.**). Antibodies against macroH2A1 overall, macroH2A1.1 and macroH2A1.2, were used. H3 detection serves as loading control. p-value by T-test, *: $p<0.05$.

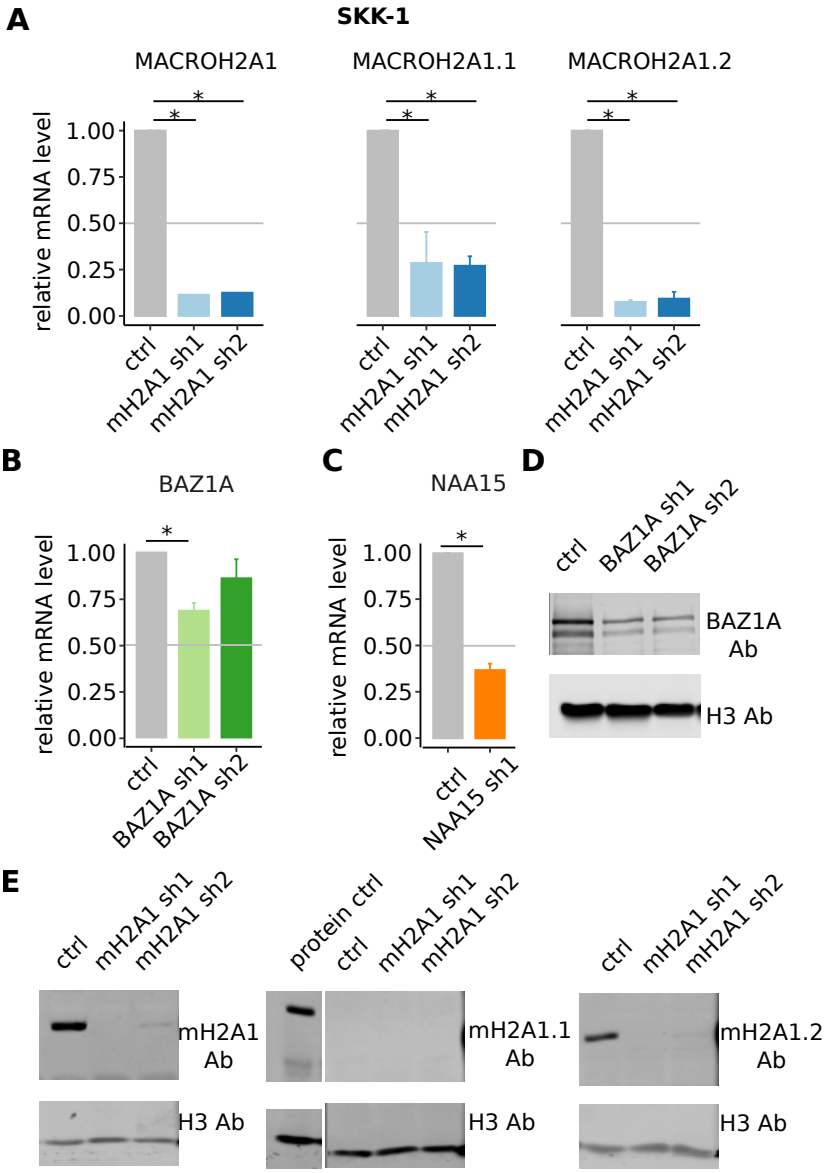


Figure 3.12: Validation of target genes' knock-down in SKK-1 cells.

3.2.2 Validation of selected targets - Response to treatment

To verify the results from the shRNA screen we determined the dose response of the newly generated knock-down SKK-1 cell lines to Azacitidine by co-staining with Dapi and MitoTracker after four days of treatment. In line with the results from the screen, we saw that the knock-down of *NAA15* in SKK-1 cells significantly decreases cell sensitivity to Azacitidine compared to control cells (**Figure 3.13 A**). We could also show the increased sensitivity to Azacitidine in cells with knock-down*BAZ1A* (**Figure 3.13 B**) and *MACROH2A1* for mid-range concentration (**Figure 3.13 C**), with two separate shRNAs for each target. To summarize, we could confirm that NAA15 depletion increased SKK-1 resistance to Azacitidine, and BAZ1A and macroH2A1 depletions had the opposite effect.

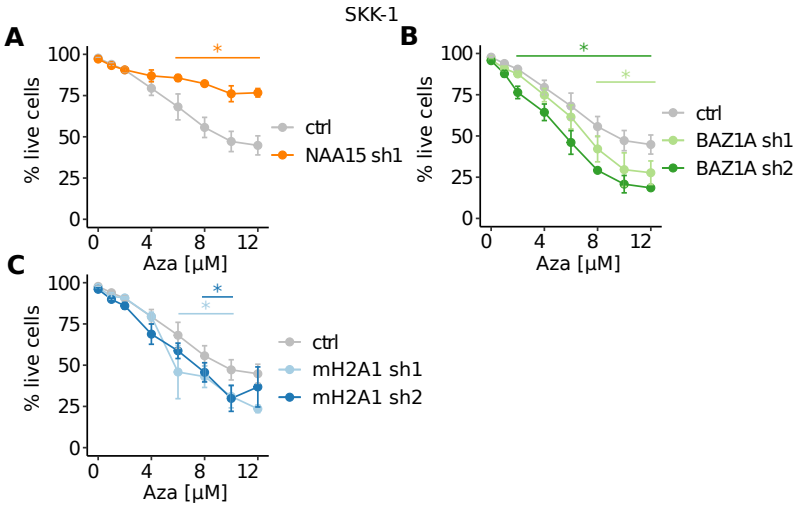


Figure 3.13: Effect of shRNA integration on Azacitidine treatment in SKK-1 cells. Cells are stably expressing cSGEP *renilla* shRNA (ctrl), *NAA15* shRNA 1 (**A.**), *BAZ1A* shRNA 1 or shRNA 2 (**B.**), *MACROH2A1* shRNA 1 or shRNA 2 (**C.**). **A-C.** Percentage of live SKK-1 cells determined by DAPI/MitoTracker were measured after 4 days of treatment with indicated concentrations of Azacitidine. n=3, p-value determined by ANOVA test, *: $p<0.05$.

We decided to test the specificity of these effects to Azacitidine treatment by using the Azacitidine analogue, Decitabine. Like for the treatment with Azacitidine, the knock-down of *NAA15* also led to resistance to treatment with low Decitabine concentration (**Figure 3.14 A**). Knock-down of *BAZ1A* did not show any difference in response to treatment compared to control cells (**Figure 3.14 B**). Both *MACROH2A1* knock-down cell lines showed a significant increase of sensitivity to Decitabine in the lower concentration range (**Figure 3.14 C**). In conclusion, the effect of NAA15 and macroH2A1 depletion on the response to treatment was not specific to Azacitidine, but could also be observed in response to treatment with Decitabine, whereas the sensitizing effect of BAZ1A depletion was specific to Azacitidine in SKK-1 cells.

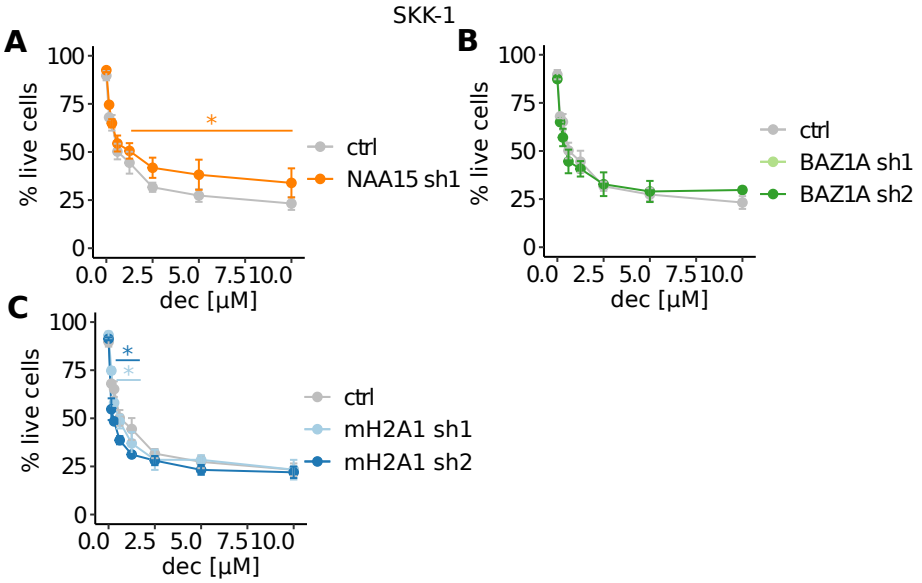


Figure 3.14: Effect of shRNA integration on Decitabine treatment in SKK-1 cells. Cells are stably expressing cSGEP renilla shRNA (ctrl), NAA15 shRNA 1 (**A.**), BAZ1A shRNA 1 or shRNA 2 (**B.**), MACROH2A1 shRNA 1 or shRNA 2 (**C.**). **A-C.** Percentage of live SKK-1 cells determined by DAPI/MitoTracker were measured after 4 days of treatment with indicated concentrations of Decitabine. n=3, p-value determined by ANOVA test, *: $p<0.05$.

3.2.3 In a mixed population only BAZ1A depletion shows higher sensitivity to Azacitidine treatment

To be one step closer to actual clinical conditions, we studied the effect of Azacitidine treatment on different knock-downs in a non-uniform pool of cells. A competitive cell growth assay allows to assess the growth capacity of the knock-downs compared to the parental cell line. As explained graphically in **Figure 3.15 A**, each daughter cell line was mixed with parental SKK-1 cells at a ratio of either 1 to 1, or 1 to 9, and the evolution of GFP positive cells was measured every other day during eight days. **Figure 3.15 B** shows the percentage of GFP positive cells in each population, confirming that parental SKK-1 cells do not have intrinsic green fluorescence and that all newly generated cell lines have close to 100% GFP positive cells. Thus, cells from the parental or daughter cell lines can easily be distinguished by flow cytometry. Results are plotted in **Figure 3.15 C**, which shows that using both ratios and all five daughter cell lines in normal cell culture conditions, the cell percentages are stable overtime. This indicates that the knock-downs do not give any growth advantage or disadvantage compared to the parental cells. We then treated the cells seeded in 1 to 1 ratio with 4µM Azacitidine, a concentration that affects cell survival without killing the whole cell population (70% cell survival based on **Figure 3.13 A**). **Figure 3.15 D** shows that in a mixed population, daughter/parental cell ratios were not affected by the treatment except for the knock-down of *BAZ1A*. Indeed in this case the daughter cells were depleted faster than the parental cell line in response to Azacitidine treatment. Taken together, in the presence of both parental and daughter cell lines, we did not observed a sensitizing effect of macroH2A1 to Azacitidine treatment. Neither did NAA15 depletion increase cell resistance to Azacitidine treatment in SKK-1 cells (**Figure 3.13**). BAZ1A depletion remained a sensitizing factor to the treatment in these conditions.

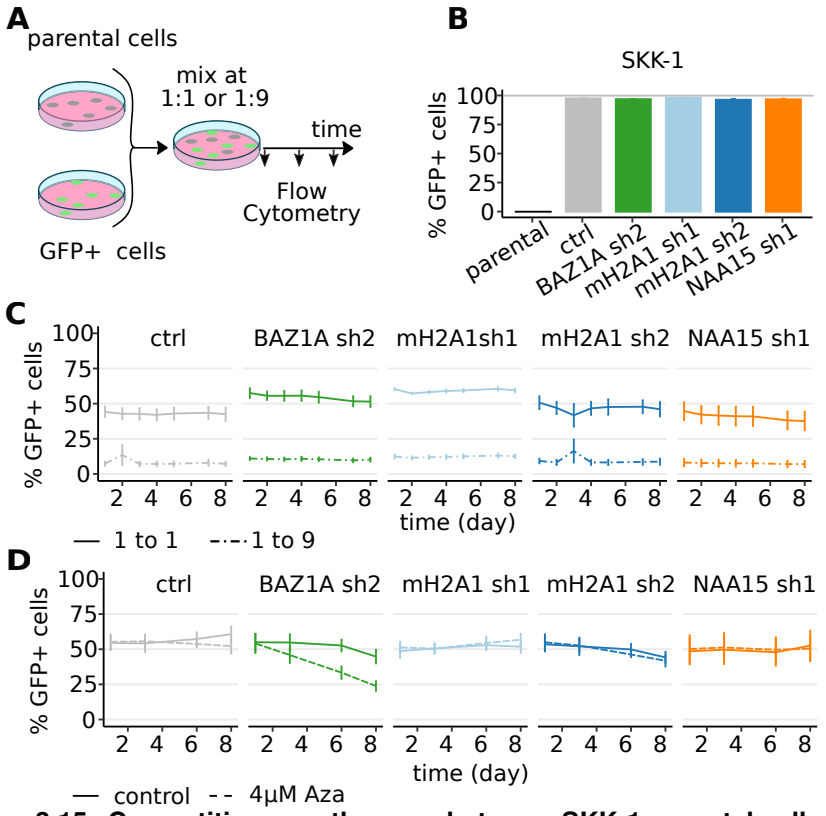


Figure 3.15: Competitive growth assay between SKK-1 parental cells and SKK-1 knock-down cells. A. Workflow of competitive growth assay. Daughters knock-down cells (GFP positive) were mixed at different ratios (either 1 to 1 or 1 to 9) with parental cells (GFP negative) and GFP positive cell percentage was measured at different time points. **B.** Flow cytometry measurement of the percentage of cells expressing GFP in the different cell lines: parental SKK-1 and daughter SKK-1 stably expressing cSGEP *renilla* shRNA (ctrl), *BAZ1A* shRNA 2, *MACROH2A1* shRNA 2 and *NAA15* shRNA 1. n=2 C. Percentage of GFP positive cells over the course of 8 days in the different cell lines, starting ratios 1:1 and 9:1. n=3. **D.** Percentage of GFP positive cells over the course of 8 days in the different cell lines treated with Azacitidine (Aza) at 4µM or same volume media (control) in corresponding wells at days 0 and 6. n=3, statistical significance evaluated by ANOVA test, *: $p<0.05$.

3.2.4 Confirming results in two additional cell lines

To further explore the sensitizing effect of *BAZ1A* and *MACROH2A1* knock-down in response to drug treatment, we expanded the assays to a second MDS derived cell line, MOLM-13. MOLM-13 cells have been isolated from a MDS patient already progressed to sAML. We also included another leukaemia cell-line, HL-60. HL-60 are derived from an acute promyelocytic leukaemia and are a highly proliferative cells.

The shRNA integration was done exactly as previously described for SKK-1 cells (see **Table 3.2.1**), and the cell selection was assessed by measuring the GFP positive cell percentage using flow cytometry after puromycin treatment. **Figure 3.16 A** shows a drop of over 50% of mRNA level of *MACROH2A1.1* and *MACROH2A1.2* by both short hairpins. This result was confirmed at the protein level by western blotting using an antibody against macroH2A1 (**Figure 3.16 C**). As for the *BAZ1A* knock-down by the shRNA 2, the mRNA level showed a non-significant mRNA level decrease (**Figure 3.16 B**). Western blot is still pending to confirm the decrease of BAZ1A protein level.

Figure 3.17 A shows a significant drop of mRNA levels of *MACROH2A1.1* and *MACROH2A1.2* in both shRNAs cell lines. This result was confirmed at the protein level by western blotting using a macroH2A1 antibody (**Figure 3.17 C**). The relative reduction in *BAZ1A* gene expression could not be confirmed due to variation (**Figure 3.17 B**) but at the protein level there was a significant decrease of expression (**Figure 3.17 D**).

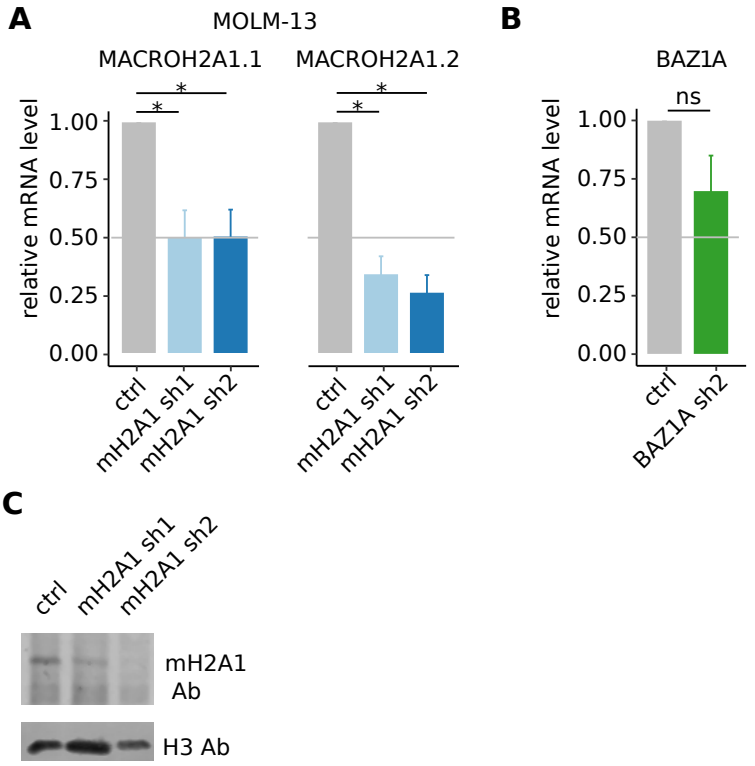


Figure 3.16: Validation of knock-down of expression of target genes in MOLM-13 cells. A-B. Knock-down efficiency of shRNAs measured by RT-qPCR, ctrl cells with cSGEP renilla shRNA (ctrl) were set to 1. Data represent the mean ±SEM of n independent measurements. **A.** Relative expression of *MACROH2A1.1* and *MACROH2A1.2* isoforms with two separate shRNAs. N=5, **B.** Relative expression of *BAZ1A* with *BAZ1A* shRNA 2. n= 4 **C.** Western blot showing reduction of protein level of macroH2A1s in MOLM-13 cells compared to the ctrl achieved by stable integration of shRNA targeting *MACROH2A1*. p-value by T-test, *: $p<0.05$

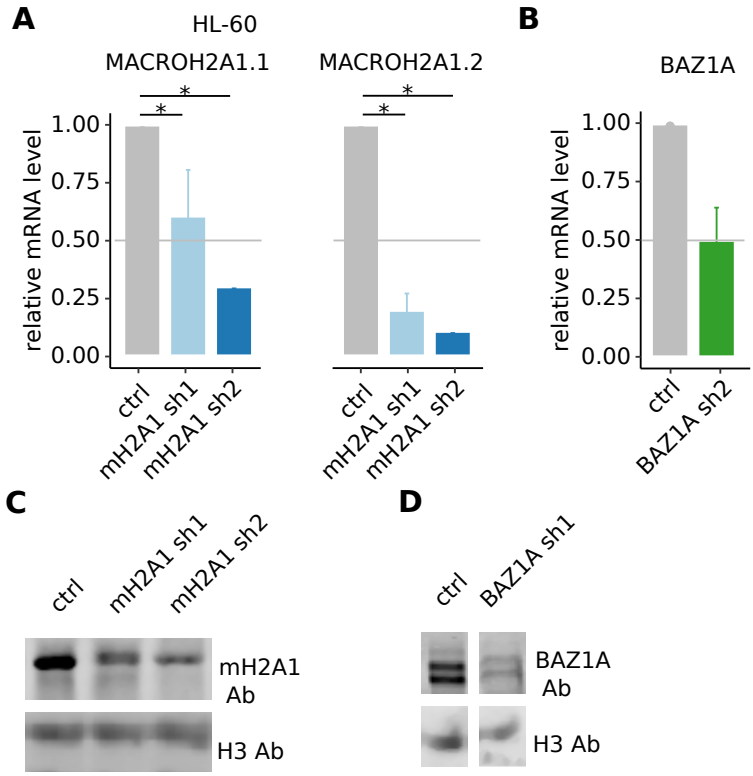


Figure 3.17: Validation of knock-down of expression of target genes in HL-60 cells. A-B. Knock-down efficiency of shRNAs measured by RT-qPCR, control cells with cSGEP *renilla* shRNA (ctrl) were set to 1. Data represent the mean ±SEM of n independent measurements. **A.** Relative expression of *MACROH2A1.1* and *MACROH2A1.2* isoforms with two separate shRNAs. n=2, p-value by T-test, *: $p<0.05$ **B.** Relative expression of *BAZ1A* with *BAZ1A* shRNA 2. n= 3 **C.** Western blot showing reduction of protein level of macroH2A1 in HL-60 cells compared to the control achieved by stable integration of shRNAs targeting *MACROH2A1*. **D.** Western blot showing reduction of protein level of BAZ1A in HL-60 cells compared to the control achieved by stable integration of a shRNA targeting *BAZ1A*.

3.2.5 The sensitizing effect of BAZ1A and macroH2A1 depletion in response to azanucleoside treatment is cell line dependant

MOLM-13 and HL-60 cell lines were treated with increasing concentrations of Azacitidine and the cell survival was measured on the fourth day by flow cytometry after co-staining with Dapi and MitoTracker. The integration of *BAZ1A* shRNA 2 did not affect the response to Azacitidine or to Decitabine in MOLM-13 cells (**Figure 3.18 A and C**). HL-60 sensitivity to Azacitidine treatment was not affected (**Figure 3.18 B**). *BAZ1A* knock-down in the same cell line increased slightly but statistically live cell percentages for the mid-range concentrations of Decitabine (1.25-5µM), we observed opposite results compared to previous cell lines (**Figure 3.18 D**). Lower Decitabine concentrations were used to treat MOLM-13 as they are much more sensitive to Decitabine than SKK-1 (**Figure 3.14**) and HL-60 cells (**Figure 3.18 D**). However, when MOLM-13 cells were treated with with the chemotherapy drug cytosine arabinoside, the integration of *BAZ1A* shRNA 2 showed a significant decrease of cell survival (**Figure 3.18 E**). Taken together, the results suggest that BAZ1A depletion had different effects in response to treatment by azanucleosides, effects that are dependant of both the cell line and the treatment. This suggests that BAZ1A depletion might not directly affect drug sensitivity but rather indirectly, which would explain the variability of the observed response.

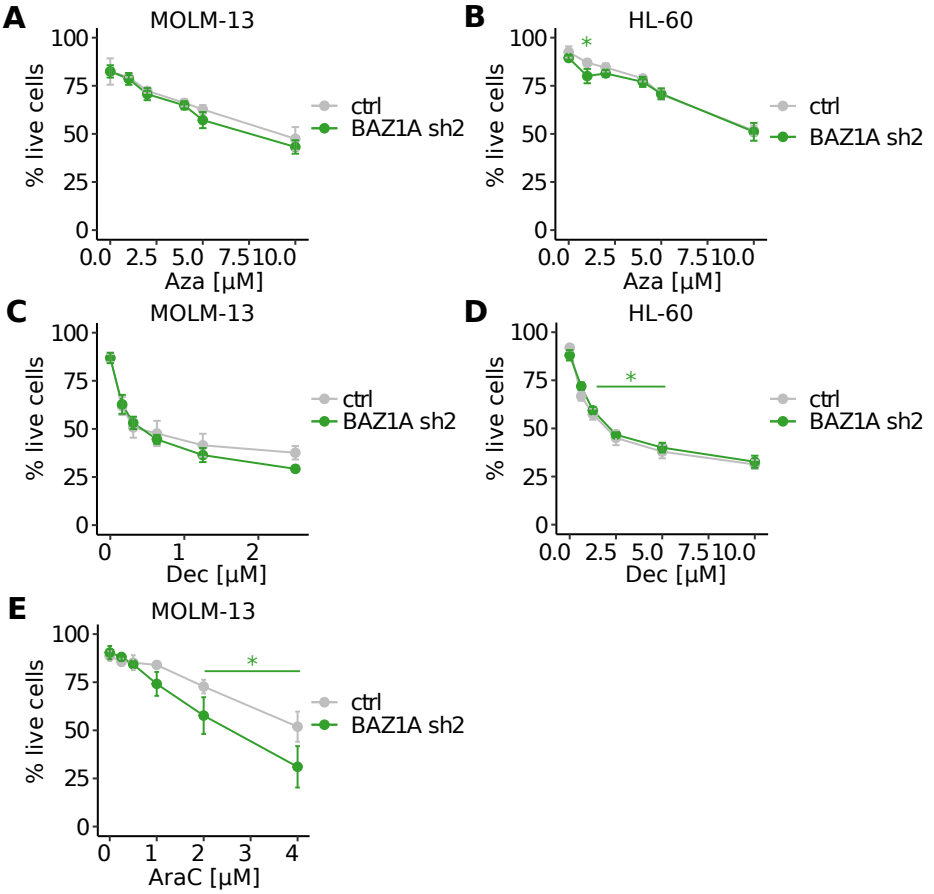


Figure 3.18: Effect of BAZ1A knock-down on different treatments in MOLM-13 and HL-60 cells. A-F. Percentage of live cells determined by DAPI/MitoTracker after 4 days of treatment. Cells are stably expressing cSGEP *renilla* shRNA (ctrl), *BAZ1A* shRNA 2, and treated with corresponding concentration of azanucleoside analogues. **A** MOLM-13 (n=3), and **B** HL-60 (n=4) were treated with Azacitidine. **C** MOLM-13 (n=4), and **D** HL-60 (n=4) were treated with Decitabine. **E.** MOLM-13 (n=3) were treated with cytosine arabinoside (AraC) . p-value determined by ANOVA test, *: $p<0.05$

The integration of both *MACROH2A1* shRNAs showed a slight but significant decrease of cell survival when treated with 5μM of Azacitidine, shRNA 2 also showed the same effect when treated with lower doses of Azacitidine (**Figure 3.19A**) however there was no effect in the response of HL-60 (**Figure 3.19B**). Both shRNAs integration increased the sensitivity to Decitabine treatment in MOLM-13 (**Figure 3.19C**) but only the less effective knock down of macroH2A1 sensitized HL-60 (**Figure 3.19D**). Cytosine arabinoside treatment was also more lethal in MOLM-13 expressing less macroH2A1 (**Figure 3.19E**). Overall MOLM-13 were sensitized to azanucleoside treatments when expressing lower level of macroH2A1 whereas HL-60 response was mostly not affected.

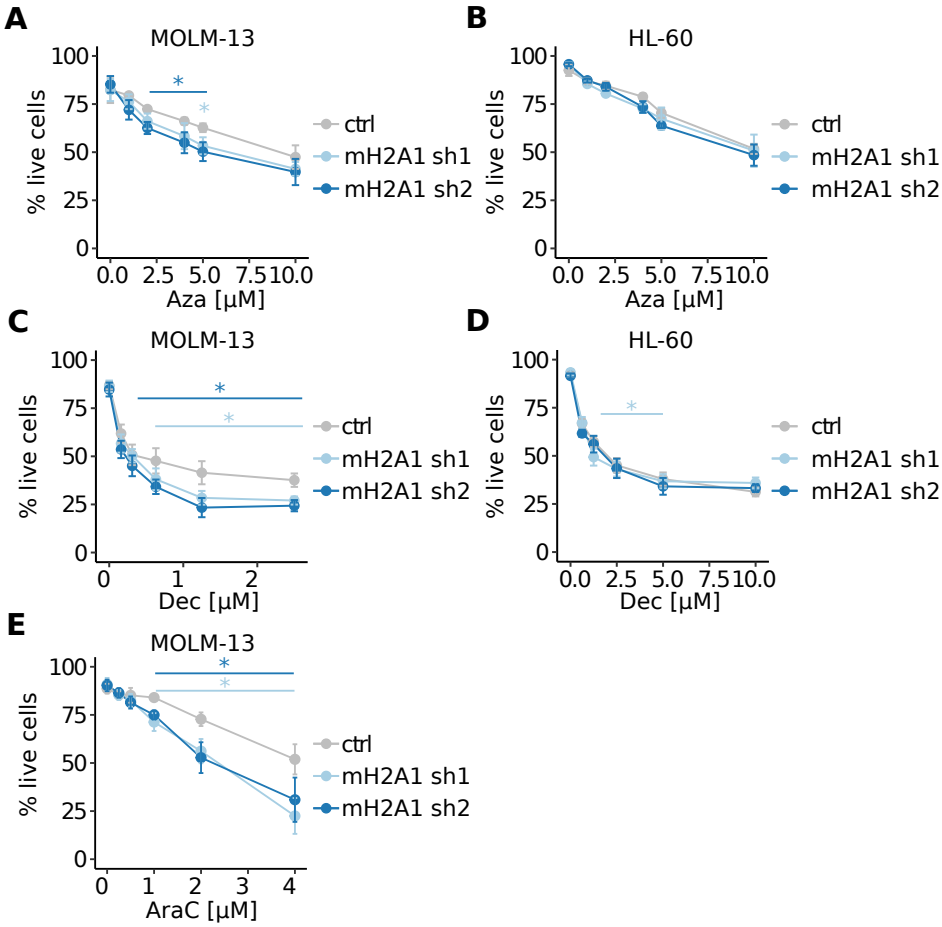


Figure 3.19: Effect of macroH2A1 knock-down on different treatments in MOLM-13 and HL-60 cells. A-F. Percentage of live cells determined by DAPI/MitoTracker after 4 days of treatment. Cells stably expressed cSGEP *renilla* shRNA (ctrl), *MACROH2A1* shRNA 1 or shRNA 2 and were treated with corresponding concentration of azanucleoside analogues. **A** MOLM-13 (n=3), and **B** HL-60 (n=4) were treated with Azacitidine. **C** MOLM-13 (n=4), and **D** HL-60 (n=4) were treated with Azacitidine. **E.** MOLM-13 (n=3) were treated with cytosine arabinoside (AraC). p-value determined by ANOVA test, *: $p<0.05$

3.2.6 Both macroH2A1 and BAZ1A depletion partially affect apoptosis in response to Decitabine in MOLM-13

To confirm the viability results in MOLM-13 cells, apoptosis was measured by flow cytometry after co-staining with Annexin V and Dapi on the fourth day of treatment with Decitabine. **Figure 3.20 A and B** show over 75% total apoptotic cells in all cell lines even at the lowest dose used (1µM). Co-staining by Annexin V and Dapi allows to distinguish between early and late apoptotic cell states, **Figure 3.20 C and E** show that 50% of cells were in early apoptosis. **Figure 3.20 D and F** show that there was a significantly lower percentage of late apoptotic cells in all knocks down cell lines when compared to ctrl cells. It can be the consequence of two things: either the overall apoptosis was reduced in the knockdown cells, however, this is not seen in the total apoptosis measured; or cell death was triggered earlier, thus these cells are not detected in the assay.

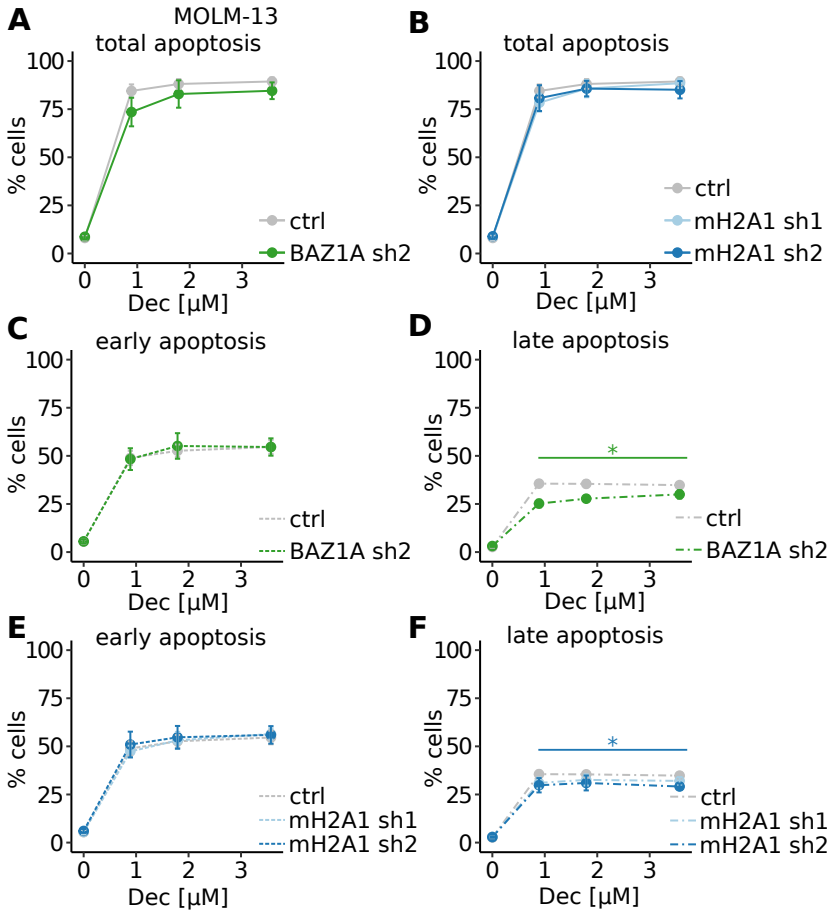


Figure 3.20: Effect of shRNA integration on apoptosis triggered by Decitabine treatment in MOLM-13 cells. Results of MOLM-13 cells stably expressing cSGEP *renilla* shRNA (ctrl), *BAZ1A* shRNA 2, *MACROH2A1* shRNA 1 or shRNA 2 determined after 4 days of treatment with indicated concentrations of Decitabine **A-B.** Percentage of total apoptotic cells, sum of early and late apoptotic cells, determined by co-staining with Annexin V and DAPI. n= 3. **C-D.** Percentage of early apoptotic cells (Annexin V positive and Dapi negative). n=3. **E-F.** Percentage of late apoptotic cells (Annexin V and Dapi positive). n=3. p-value determined by ANOVA test, *: $p<0.05$.

3.2.7 MacroH2A1 depletion increases cell death by Bcl2 inhibitor, ABT-199 in MOLM-13 cells

Due to these inconclusive results we decided to further explore the effect on apoptosis of both knock-downs in MOLM-13 cells. ABT-199, also called Venetoclax, is a Bcl-2-selective inhibitor, thus it modifies the anti/pro-apoptotic balance and triggers apoptosis. It has been shown to have selective effect in lymphoma (Souers et al., 2013). In recent years, several clinical studies investigated the use of ABT-199 alone or in combination with HMAs in MDS patients, with promising results (Ball et al., 2020). We treated the cells with low doses of ABT-199. There was no difference in response between *BAZ1A* knock-down and the control (**Figure 3.21 A**), whereas *MACROH2A1* knock-downs showed increase sensitivity to ABT-199 treatment at higher doses (**Figure 3.21 B**). In conclusion, BAZ1A depletion did not affect response to ABT-199 treatment. In contrast, the effect of ABT-199 on cell death was increased by macroH2A1 depletion suggesting a sensitizing effect to ABT-199 in response to macroH2A1 depletion in MOLM-13.

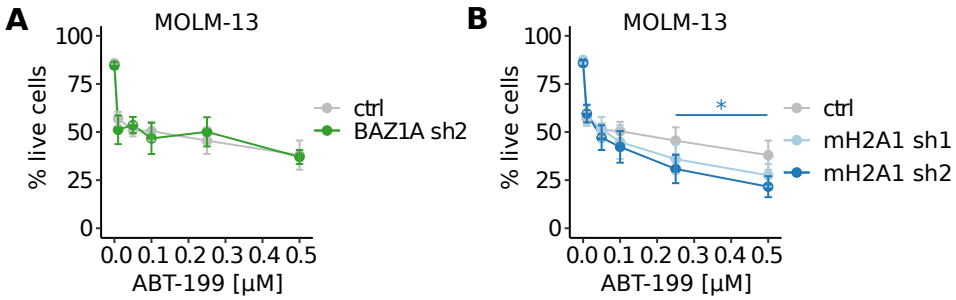


Figure 3.21: Effect of shRNA integration on ABT-199 treatment in MOLM-13 cells. Percentage of live MOLM-13 cells determined by DAPI/MitoTracker after 4 days of treatment with indicated concentrations of the Bcl-2 selective antagonist, ABT-199. Cells are stably expressing cSGEP *renilla* shRNA (ctrl), *BAZ1A* shRNA 2 (**A**), *MACROH2A1* shRNA 1 or shRNA 2 (**B**). n=4, p-value determined by ANOVA test, *: $p<0.05$.

3.2.8 Results II: conclusions

Using an unbiased approach and current knowledge on MDS, we identified a CHRAC complex component, BAZ1A, as well as macroH2A1s as potential targets for co-treatment or response indicators for treatment by azanucleosides. We could confirm that a partial loss of expression of NAA15 increased SKK-1 resistance to Azacitidine and Decitabine but did not affect cell proliferation. Partial loss of expression of *BAZ1A* had different effect on the response to azanucleoside treatments depending on which treatments and the cell lines. While our results are preliminary, we could demonstrate in two MDS derived cell lines that a partial loss of expression of macroH2A1s increased cell death by azanucleoside treatments and pro-apoptotic ABT-199. Provided that there is a frequent loss of the *MACROH2A1* gene in MDS, this effect should be further characterized.

Contributions This work was conceived by Marcus Buschbeck, Jeannine Diesch, and me, it was supervised by Marcus Buschbeck and Jeannine Diesch. The screens were done and analysed by Jeannine Diesch, Michael Maher, and Raquel Casquero. The rest of the experimental work and analysis were done by me with some support from Jeannine Diesch and Raquel Casquero.

3.3 Results III: gene expression in MDS patient samples

Cell lines are extremely useful tools in basic research, but they have numerous limitations when it comes to more translational research. Thus, we decided to assess the relevance of section 3.2 results and expand our study in a cohort of patient samples.

3.3.1 Cohort characteristics

We used a cohort of 35 patient samples, collected between 2004 and 2014. According to WHO 2008 (Vardiman et al., 2009), 20 (57%) had MDS, 2 (6%) had chronic myelomonocytic leukaemia type 2, one had primary AML and 12 (34%) had AML/MDS either therapy-related or secondary. Samples were part of a larger cohort used for a mutational analysis of the most mutated genes in MDS/AML in Kuendgen et al., 2018. In the present cohort, the most frequent mutations were SRSF2 (49%), ASXL1 and RUNX1 (29%) **Table 3.5**. One patient had no detectable mutations, and 27 patients (77%) had more than one mutation. Regarding cytogenetics, 15 patients (43%) had normal karyotypes, and 1 had only one cytogenetic alteration, all others presented more than one alteration. Among these, 6 of them (17%) had a loss of chromosome 7, 7 (20%) had a deletion 5q and 4 (11%) a trisomy of chromosome 8. Samples were taken at diagnostic and patients underwent between 3 and 25 Azacitidine treatment cycles with an average of 6.8 cycles.

Gene	Wild type	Mutated	NA
ASXL1	57.1% (20)	42.9% (15)	
DNMT3A	88.6% (31)	11.4% (4)	
EZH2	91.4% (32)	8.6% (3)	
FLT3.LM	91.4% (32)	8.6% (3)	
IDH1	80.0% (28)	17.1% (6)	2.9% (1)
IDH2	88.6% (31)	11.4% (4)	
KRAS	94.3% (33)	5.7% (2)	
MLL.PTD	94.3% (33)	5.7% (2)	
NRAS	82.9% (29)	17.1% (6)	
RUNX1	68.6% (24)	28.6% (10)	2.9% (1)
SF3B1	91.4% (32)	8.6% (3)	
SRSF2	42.9% (15)	48.6% (17)	8.6% (3)
TET2	68.6% (24)	31.4% (11)	
TP53	82.9% (29)	17.2% (6)	

Table 3.5: Summary of mutations present in the cohort

3.3.2 Selection of genes for a panel expression analysis

We decided to study the expression of a panel of genes using Nanostring technology and to assess any potential correlation between the response to Azacitidine treatment and the expression of genes of interest. We extracted RNA from the bone marrow aspirates and ,using Nanostring technology, we measured the RNA level for 50 probes (**Figure 3.22 A**). Nanostring chip allows to quantify mRNA for a selected panel of genes using specific probes coupled with fluorochrome barcodes (**Figure 3.22 B**). It prevents any amplification artefacts as there is no PCR step. We designed the panel based on literature and on our results from the loss-of-function consolidated by publicly available expression data.

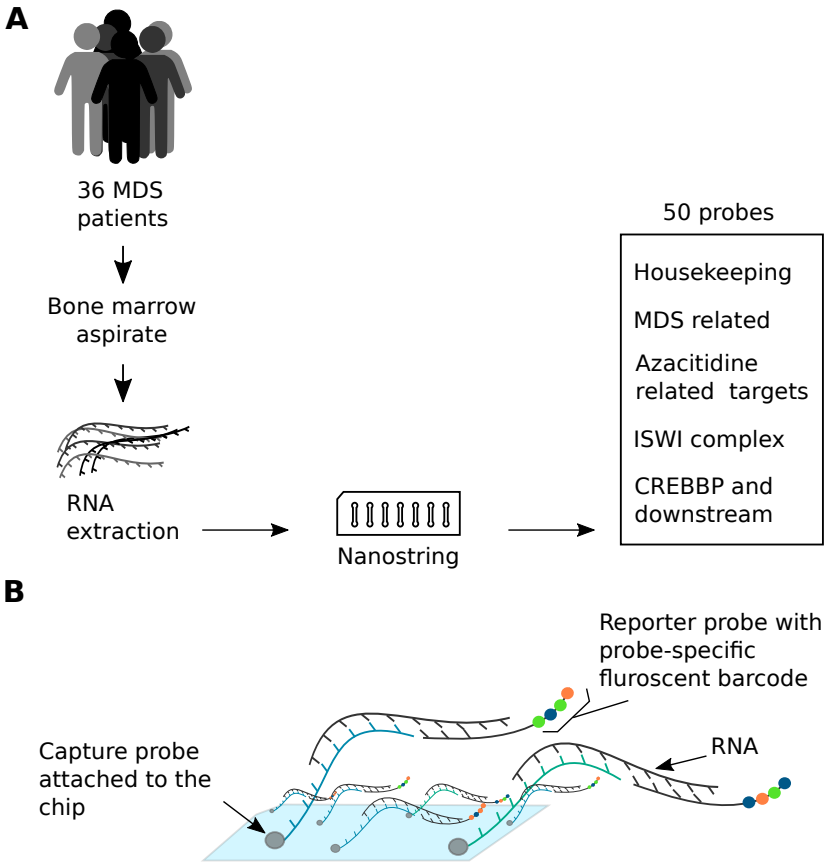


Figure 3.22: Gene expression patient samples flow. A. Flow of experiment: 35 bone marrow aspirations from patients were frozen, later RNA was extracted and analysed by Nanostring RNA chip using 50 gene probes. **B.** Representation of the Nanostring technology. RNA is captured using two-part probes: they hybridize with each end of the target RNA, the first part fixes the complex to the chip by biotin-streptavidin interaction, the second bear the unique fluorochrome sequence for quantifiable, probe specific readout.

The panel can be broken-down the following way, more details in Table 3.6 :

- 3 housekeeping genes,
- 6 genes known to be related to MDS,
- 10 genes involved in the response to Azacitidine treatment,
- 6 hits from the screen,
- 6 genes coding for proteins related to the ISWI complex,
- and 15 genes downstream of CREBBP, which was also in the top hits from the screen and led to the development of a project and an article (Diesch et al., 2021). It was led by another researcher thus I will not expand much on the topic.
- We also included both *MACROH2A1* and *MACROH2A2*. *MACROH2A1* is located on the long arm of chromosome 5 and frequently present in single copy in MDS, and *MACROH2A2* was used to evaluate any potential compensation effect between the isoforms.

Housekeeping targets

TUBB	encodes a beta tubulin protein
GUSB	encodes a hydrolase that degrades glycosaminoglycans, Glucuronidase Beta
GAPDH	encodes a member of the glyceraldehyde-3-phosphate dehydrogenase protein family, Glyceraldehyde-3-Phosphate Dehydrogenase

known relevant targets in MDS

TP53	transcription factor and known tumor suppressor (Welch et al., 2016)
ASXL1	Histone-binding protein (Prats-Martín et al., 2020)
IDH1	Catalyze the oxidative decarboxylation of isocitrate to 2-oxoglutarate (Di Nardo et al., 2016)
IDH2	Catalyze the oxidative decarboxylation of isocitrate to 2-oxoglutarate (Di Nardo et al., 2016)

DNMT3A	DNA methylteransferase citepYang2019AberrantImplications
TET2	Methylcytosine dioxygenase (Jankowska et al., 2009)
Azacitidine metabolism	
CDA	Cytidine deaminase, levels correlate with response to treatment.(Qin et al., 2011)
UCK1	Uridine-cytidine kinase, lower expression in nonresponders. (Valencia et al., 2014)
UCK2	Uridine-cytidine kinase, overexpression can restore sensitivity to treatment.(Gu et al., 2021) (Sripayap et al., 2014)
SLC29A1	Responsible for cellular uptake of AZN, when inhibited effects of AZN strongly reduced, it is a possible biomarker for response. (Ueda, Hosokawa and Iwakawa, 2015)
SLC28A3	AZN transporter. (Damaraju et al., 2012)
SLC29A1	Responsible for cellular uptake of AZN, when inhibited effects of AZN strongly reduced, it is a possible biomarker for response. (Ueda, Hosokawa and Iwakawa, 2015)
SLC28A3	AZN transporter. (Damaraju et al., 2012)
Azacitidine resistance	
RRM1	Overexpression has been associated with resistance (Aimiuwu et al., 2012)
RRM2	Overexpression has been associated with resistance (Aimiuwu et al., 2012)
BCL2L10	High expression correlates positively with AZN resistance (Cluzeau et al., 2012)
Interference RNA screen hits	
NAA10	knock-down confers apoptotic resistance to DNA-damaging agents (Arnesen et al., 2006)
NAA15	knock-down confers apoptotic resistance to DNA-damaging agents (Arnesen et al., 2006)
HELLS	SNF2 family member lymphoid-specific helicase, can be involved in DNA strand separation, including replication, repair, recombination, and transcription; required for de novo or maintenance DNA methylation

CREBBP	contribute to cell differentiation and stem cell maintenance in haematopoesisgene fusion with MOZ and MLL in AML (Borrow et al., 1996; Satake et al., 1997)
RING1	ring finger protein 1 PCR1
DNMT1	DNA (cytosine-5-)-methyltransferase 1
NSUN3	NOP2/Sun domain family, member 3. Putative RNA methyltransferase in mitochondria; initiates 5-formylcytidine biogenesis in mitochondrial tRNA(Met); knock-down reduces mito protein synthesis and oxygen consumption.
PWWP2B	PWWP domain containing 2B

ISWI components

CHRAC1	component of ISWI complex, regularly spaces nucleosomes and thus generates higher order chromatin and chromatin organization required for BER, NER
BAZ1A	also called ACF1, forms complex with CHRAC1 and SMARCA5 to mediate nucleosome positioning and relaxation of DNA promotes nucleosome sliding important during replication and transcription promote recovery from DNA damage
BAZ1B	also called WSTF, phosphorylates H2Ax in response to DNA damage, part of WICH complex together with SMARCA5
POLE3	also called CHRAC17, DNA Polymerase Epsilon 3. Involved in chromatin remodeling and DNA replication
SMARCA5	also called SNF2h, forms complex with BAZ1A and CHRAC

CREBBP and its downstream targets

CREBBP	Main hit from the screen. It contribute to cell differentiation and stem cell maintenance in haematopoesis,gene fusion with MOZ and MLL in AML (Borrow et al., 1996; Satake et al., 1997)
POLR1B	Common component of RNA polymerases I
POLR1C	Common component of RNA polymerases I and III
POLR1E	Common component of RNA polymerases I
POLR3D	Common component of RNA polymerases III

POLR3H	Common component of RNA polymerases III
MRPS26	Small subunit of ribosome
RRP9	U3 Small Nucleolar Ribonucleoprotein-Associated 55 KDa Protein
RRP1	ribosomal RNA processing protein 1
MALSU1	Mitochondrial Assembly Of Ribosomal Large Subunit 1
MRPL4	Mitochondrial Large Ribosomal Subunit Protein UL4m
MRPL52	Mitochondrial Large Ribosomal Subunit Protein ML52
RIOX2	Ribosomal Oxygenase 2
MRTO4	MRT4 Homolog, Ribosome Maturation Factor
P300	contribute to cell differentiation and stem cell maintenance in haematopoesis,gene fusion with MOZ in AML (Borrow et al., 1996)

Table 3.6: Gene panel for RNA expression in patient samples

BloodSpot is a database of haematopoietic cells in health and disease, regrouping RNA-sequencing data sets, statistics, and visualization tools (Bagger, Kinalis and Rapin, 2019). Using data from the BloodPool AML samples vs. normal cells, we looked at the difference of expression between normal haematopoietic cells and MDS patient samples. Results are represented in **Figure 3.23**, showing that *BAZ1A, BAZ1B* and *CHRAC1* of the CHRAC complex were all less expressed in MDS patient samples. This was also the case with *MACROH2A1*, which was not surprising provided that at least one copy of the gene is lost in 20% of MDS patients. There was no difference of expression of *POLE3, SMARCA5*, or *MACROH2A2*. Among the other hits from the screen, *CREBBP, HELLS* and *NAA15* are over-expressed in MDS patient samples compared to healthy cells and *ERCC2, FLYWCH1, NAA10* and *RING1* showed a decrease of expression (**Figure 3.24**). In this analysis we confirmed that the expression of most of the selected hits was different in MDS patient samples and thus may be relevant in MDS and worth analysing considering the response to Azacitidine treatment.

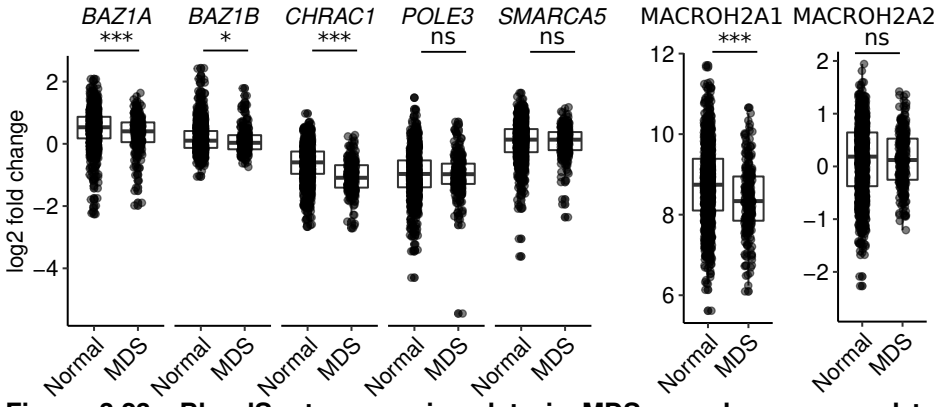


Figure 3.23: BloodSpot expression data in MDS samples compared to healthy samples of CHRAC complex components and macroH2As. Boxplots of gene expression profiles using oligonucleotide microarray chips in hematopoietic cell samples. Samples are categorized between normal (n= 989) and MDS (n= 228). Boxplots represent the first and third quartiles and the median of expression. p-values by T-test, *** $p<0.001$, ns = no significative.

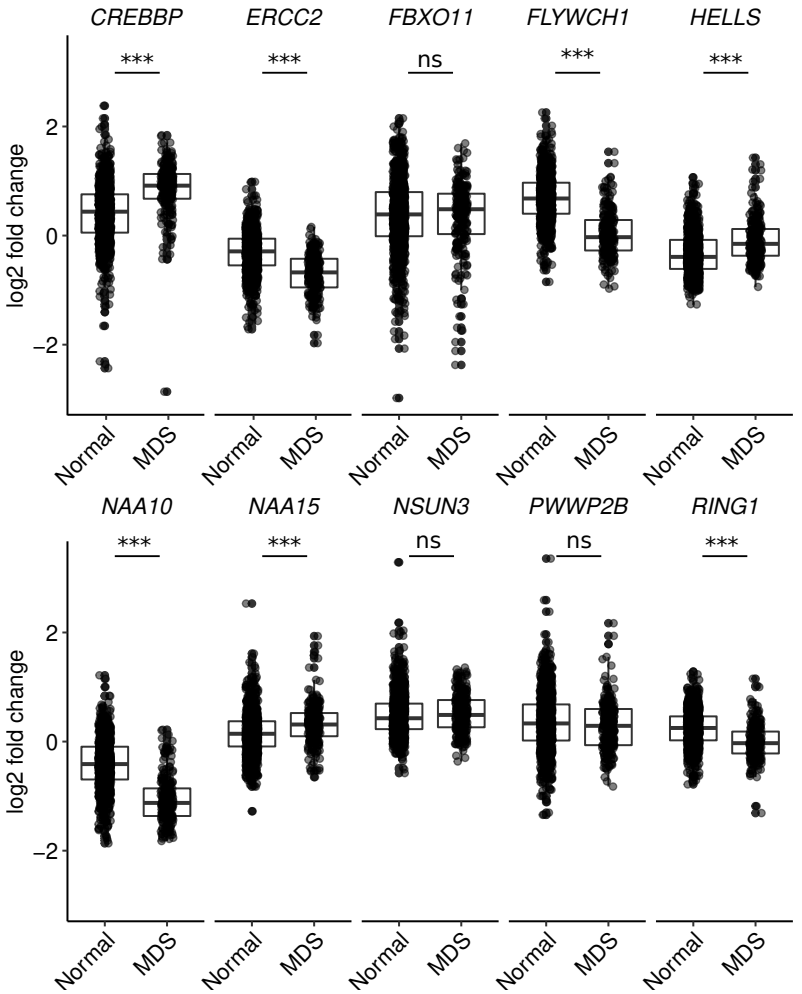


Figure 3.24: BloodSpot expression data in MDS samples compared to healthy samples of selected hits. Boxplots of gene expression profiles using oligonucleotide microarray chips in hematopoietic cell samples. Samples are categorized between normal (n= 989), MDS (n= 228). Boxplots represent the first and third quartiles and the median of expression. p-values by T-test, *: $p<0.05$; ** $p<0.01$ *** $p<0.001$, ns = no significative.

3.3.3 High variability of expression profile in patient samples

After the standard technical quality check, explained in Methods, we looked first at the raw data to assess the quality of our read out. One of the housekeeping genes, GAPDH, showed high variation between the patient samples, and thus was excluded for the normalisation (**Figure 3.25 A**). Overall, in all groups of probes, we detected a mix of low and higher counts. Also, there was a lot of variation among the samples, which is expected in patient samples (**Figure 3.25 B-G**).

Figure 3.25: Raw Nanostring counts of the different RNA targets. Each point represents the count for one sample, boxplots represent the first and third quartiles and the median of counts for each probe. Probes were grouped based on their selection for the panel. **A.** Housekeeping genes, **B.** proteins related to Azacitidine metabolism, **C.** MDS related genes, **D.** CHRAC complex components, **E.** screen hits, **F.** *MACROH2As*, **G.** CREBBP related genes.

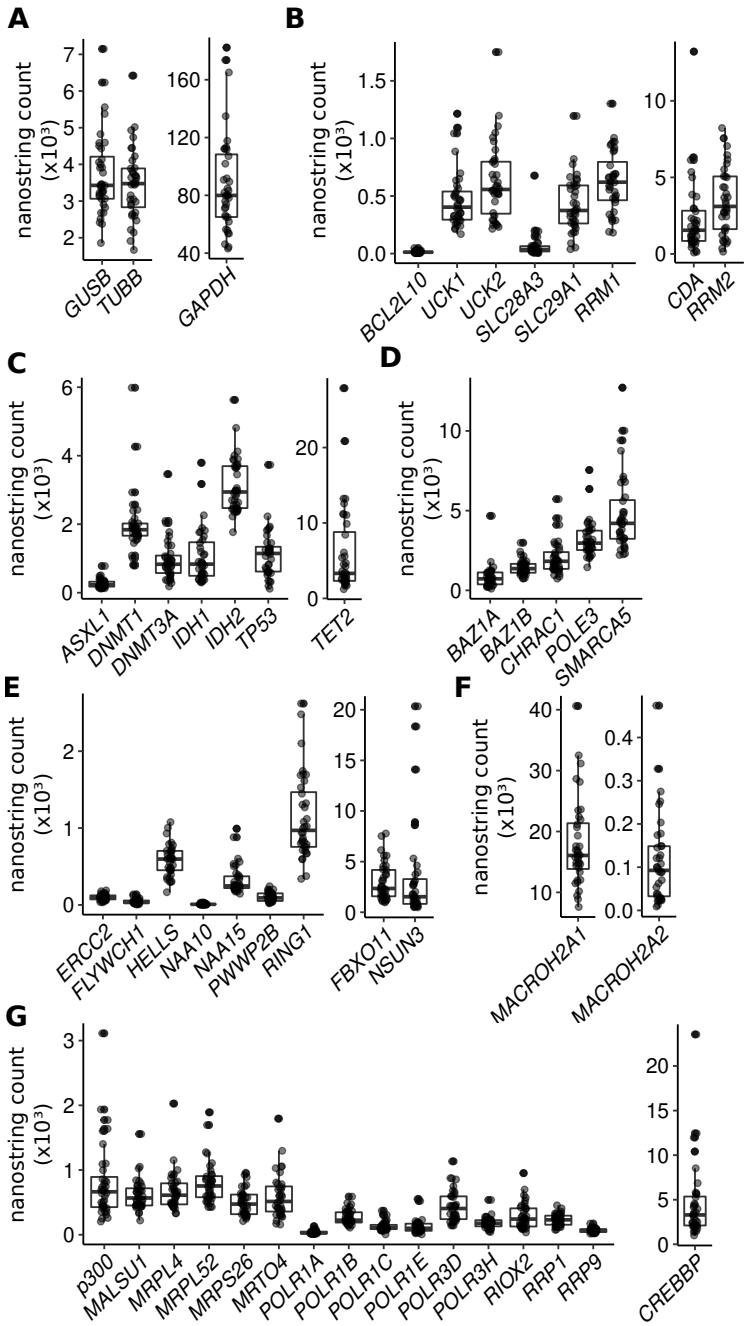


Figure 3.25: Raw Nanostring counts of the different RNA targets.

3.3.4 Differential expression depending on key gene mutational status

The first question we had was whether the normalized expression of any of our genes of interest was affected by the mutational status of the patient. Here, I will focus on genes that show a statistical difference of expression between mutated and wild-type samples in our cohort (**Figure 3.26**). Interestingly, *NAA15, RING1, ASXL1* as well as downstream targets of CREBBP *MALSU1, MRPL4, MRPS26, RIOX2, POLR1B, POLR3B, POLR3H* and *RRP1* were all less expressed in samples with Tet methylcytosine dioxygenase 2 (*TET2*) mutation (**Figure 3.26 A**). Another surprising result was that *TET2* mutation did not affect *TET2* expression in our cohort. *TET2* mutation was also associated with a decrease of expression of several epigenetic factors: *BAZ1A*,*CHRAC1, SMARCA5* and *DNMT1* (**Figure 3.26 A**). mutation of the transcription factor *RUNX1*, is associated with poor prognosis in MDS patients. We found that anti-apoptotic member of the Bcl2 family, *BCL2L10* and azanucleoside transporter, *SLC28A3* were both down regulated in *RUNX1* mutated samples (**Figure 3.26 B**). Serine/arginine-rich splicing factor 2 (*SRSF2*) mutation is also associated with an adverse prognostic impact on survival and MDS progression. In our cohort, the gene encoding for F-Box Protein 11, *FBXO11*, showed an increase of expression in *SRSF2* mutated samples while RING-type E3 ubiquitin transferase (*RING1*), ribonucleotide reductase regulatory subunit M2 (*RRM2*) and *BCL2L10* were down-regulated (**Figure 3.26 C**). Finally, the gene coding for tumour protein P53, *TP53* leads to adverse prognosis. Its mutation correlated with its own lower expression as well as *TET2*, NOP2/Sun RNA methyltransferase 3 (*NSUN3*) and macroH2A1 (**Figure 3.26 D**). *ASXL1*, *IDH1*, and *NRAS* mutational status did not show different expression for any of the genes in our panel.

The small number of patients with mutation of *DNMT3A, EZH2, FTL3* length, *IDH2, KRAS, MLL* partial tandem duplication,

and *SF3B1* in our cohort did not permit a similar analysis. Overall, most notably, the expression of the CHRAC complex components was affected by *TET2* mutation and *BCL2L10* expression was affected by both *RUNX1* and *SRSF2* mutations.

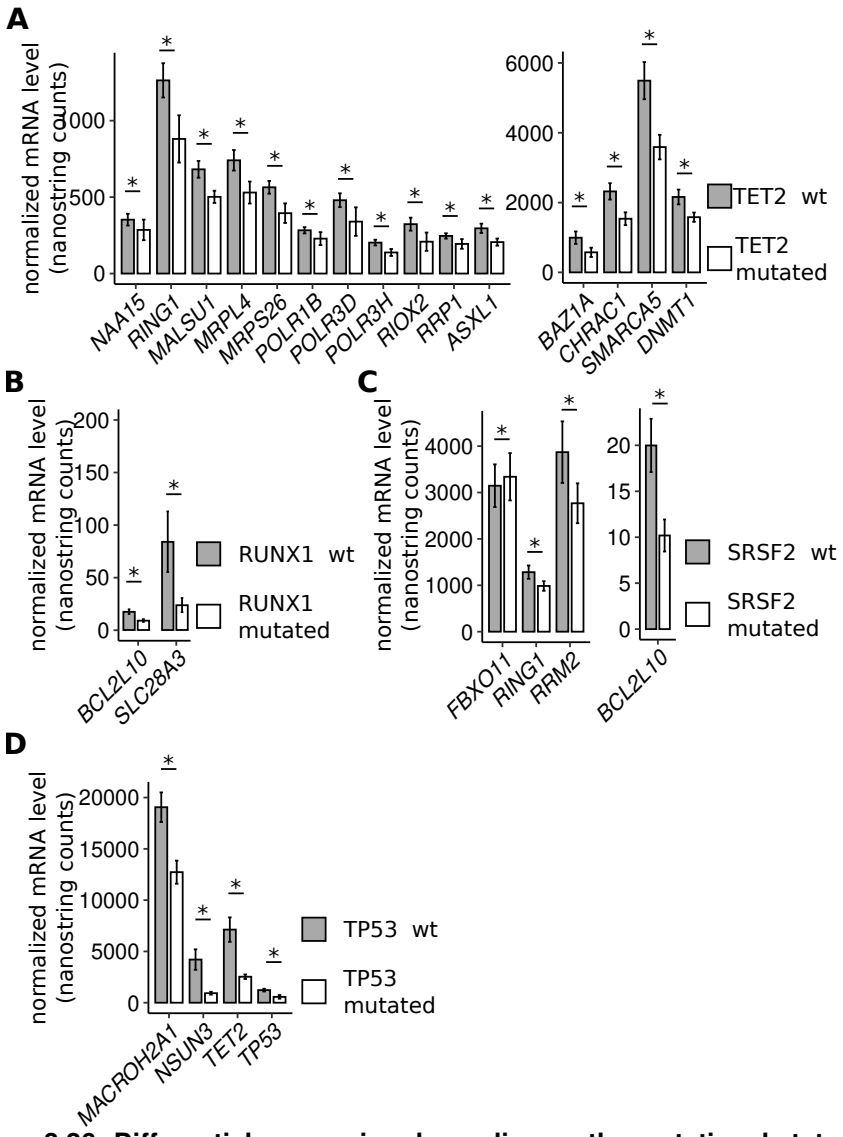
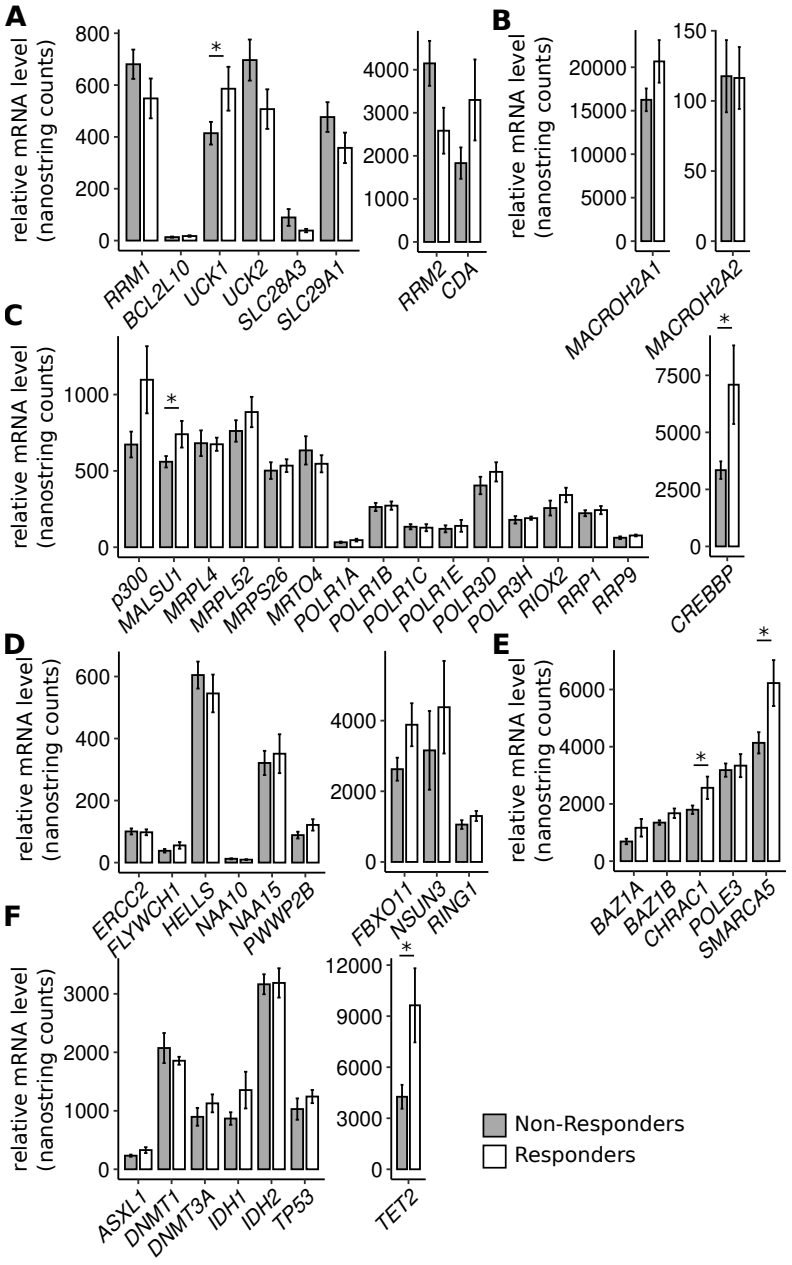


Figure 3.26: Differential expression depending on the mutational status of key genes. Relative mRNA level measured by Nanostring and normalised to the mean of two housekeeping genes from bone marrow aspirates from 35 patients. Patients are divided according to the mutational status of a key gene. **A.** 24 samples had no mutation of *TET2* versus 11 samples with mutated *TET2*. **B.** 24 samples had no mutation of *RUNX1* versus 10 samples with mutated *RUNX1*. **C.** 15 samples had no mutation of *SRSF2* versus 17 samples with mutated *SRSF2*. **D.** 29 samples had no mutation of *TP53* versus 6 samples with mutated *TP53*. p value from either Wilcoxon or Welsh test (more detail in Methods) *: $p<0.05$.

3.3.5 Gene expression as an indicator for response to treatment

To determine differential expression between patients that responded or not responded to Azacitidine treatment, we split the patient samples into responders (n=16) and non-responders (n=19) and compared the expression for each gene of interest in both groups. Among genes related to Azacitidine metabolism and response only *UCK1* showed a higher expression in responders (**Figure 3.27 A**). *RRM2* showed an opposite trend. Those results are aligned with the literature (Aimiuwu et al., 2012; Valencia et al., 2014). *MACROH2A1* showed the same trend, namely an increase of expression in responders whereas *MACROH2A2* showed no change of expression between both groups (**Figure 3.27 B**). *CREBBP*, *P300* and *MALSU1* had higher expression in responder samples (**Figure 3.27 C**). The CHRAC complex components *CHRAC1* and *SMARC5* were up-regulated in responder samples versus non-responders (**Figure 3.27 E**). Similarly, *BAZ1A* and *BAZ1B* showed the same trend, namely an increase of expression in responders. None of the hits from the screen showed a statistical difference of expression but *FLYWCH1, PWWP2B* and *NSUN3* showed a slight increase of expression in responders in our cohort (**Figure 3.27 D**). Among the genes whose mutational status is relevant in MDS, *TET2* is the only one that had a higher expression in responders (**Figure 3.27 F**). The statistical analysis is summarized in **Table 3.7**. Overall, the results showed a lower expression of CHRAC complex components as well as *MACROH2A1* in samples from patient that did not respond to Azacitidine treatment.

Figure 3.27: Expression levels of key genes depending on the patient response to Azacitidine treatment. Relative mRNA level measured by Nanostring and normalised to the mean of two housekeeping genes from bone marrow aspirates from 35 patients. Samples were categorized between non-responders (n=22), and responders (n=13). Probes have been grouped as explained in the text. **A** Housekeeping targets, **B** Azacitidine metabolism and response targets, **C** MDS relevant target, **D** ISWI components, **E** hits from shRNA screen, **F** macroH2As, and **G** CREBBP and its downstream targets. *: $p<0.05$.


A
B
C
D
E
F
relative mRNA level (nanostring counts)
RRM1
BCL2L10
UCK1
UCK2
SLC28A3
SLC29A1
RRM2
CDA
MACROH2A1
MACROH2A2
p300
MALSU1
MRPL4
MRPL52
MRPS26
MRTO4
POLR1A
POLR1B
POLR1C
POLR1E
POLR3D
POLR3H
RIOX2
RRP1
RRP9
CREBBP
ERCC2
FLYWCH1
HELLS
NAA10
NAA15
PWWP2B
FBXO11
NSUN3
RING1
BAZ1A
BAZ1B
CHRAC1
POLE3
SMARCA5
ASXL1
DNMT1
DNMT3A
IDH1
IDH2
TP53
TET2
Non-Responders
Responders

Probe name	Mean non responders	SEM	Mean responders	SEM	p value
significant (p-val<0.05)					
TET2	4184.93	674.10	9068.55	2094.08	0.01
CREBBP	3286.42	372.33	6688.41	1645.29	0.02
SMARCA5	4069.56	360.21	5959.65	788.27	0.02
ASXL1	228.25	18.10	321.44	47.44	0.04
MALSU1	554.70	35.89	718.63	83.62	0.05
CHRAC1	1775.27	141.93	2475.28	372.90	0.05
UCK1	406.40	42.38	571.03	79.71	0.05
trend (p-val<0.09)					
RRM2	4070.90	503.36	2595.07	494.04	0.06
p300	661.67	81.36	1043.82	210.31	0.06
IDH1	867.93	102.53	1383.46	291.70	0.06
PWWP2B	86.14	10.72	123.42	17.13	0.06
BAZ1A	668.34	93.47	1146.46	285.84	0.07
MACROH2A1	16245.37	1246.80	20671.16	2267.93	0.07
FBXO11	2577.23	314.02	3692.09	594.75	0.08
FLYWCH1	37.29	5.49	55.79	9.91	0.09
not significant (p-val>0.09)					
CDA	1787.18	350.60	3115.30	890.10	0.12
RRP9	61.00	8.27	78.63	5.72	0.13
UCK2	688.47	75.92	520.06	71.82	0.14
POLR1A	31.81	4.24	44.23	8.45	0.16
RIOX2	250.18	47.04	346.41	44.18	0.17
RRM1	672.47	54.49	549.00	71.11	0.17
SLC29A1	469.53	55.17	357.23	54.36	0.18
RING1	1037.13	120.13	1280.54	133.50	0.20
POLR3D	395.53	55.07	500.59	58.71	0.22
NAA10	12.07	1.55	9.25	1.62	0.24
DNMT3A	872.19	147.37	1128.28	141.96	0.25
SLC28A3	86.53	31.06	40.28	6.28	0.25
TP53	1016.32	174.83	1285.18	112.00	0.27
MRPL52	752.32	67.78	876.85	92.34	0.28

RRP1	218.15	19.27	251.34	26.15	0.31
BCL2L10	13.27	2.02	16.66	2.78	0.32
DNMT1	2058.73	245.47	1846.09	63.27	0.50
POLR1E	118.54	21.97	144.86	36.26	0.51
NSUN3	3043.95	1068.91	4131.97	1238.48	0.52
POLR3H	176.25	23.54	196.01	11.32	0.53
MRTO4	625.73	88.82	550.95	52.05	0.53
MRPS26	493.03	52.85	537.41	38.80	0.55
HELLS	590.75	43.77	549.05	56.42	0.56
NAA15	315.62	37.57	351.96	58.03	0.59
POLR1B	260.02	25.79	281.01	26.59	0.59
POLE3	3152.37	222.63	3317.32	373.72	0.69
POLR1C	133.66	16.12	128.93	21.40	0.86
ERCC2	99.63	9.27	97.88	8.71	0.90
IDH2	3142.55	164.05	3127.53	239.58	0.96
MRPL4	672.49	80.73	667.64	40.57	0.96
MACROH2A2	113.35	24.90	114.84	20.50	0.97

Table 3.7: Nanostring results statistical analysis. Summarized data from 36 patient samples expression level measuredby nanostring. t-test between 22 non-responders and 14 responders.

3.3.6 Overall survival in high-expression versus low-expression

Another way to analyse the expression data from patient samples is to correlate gene expression with patient survival. For this, the median expression of each gene was calculated, and the samples split between above or below median expression. The two subgroups were compared based on the median survival (**Table 3.8**). In **Figure 3.28** the overall survival depending on the expression of selected target genes can be seen. Interestingly, the two different *MACROH2A* isoforms show opposite trends. Patients expressing lower levels of *MACROH2A2* have a median survival of nearly 25 months compared to 16 months for the higher level (**Figure 3.28**). While patients expressing lower levels of *MACROH2A1* have a median survival of close to 14 months compared to nearly 25 months for the higher level (**Figure 3.28**). None of the CHRAC complex components show a significant correlation between expression and survival (**Table 3.8**).

Probe name	lower expression median survival (months)	higher expression median survival (months)	Square Chi	p value
significant				
RING1	13.667	26.233	6.785	0.009
NSUN3	11.767	24.500	5.262	0.022
SLC28A3	24.833	14.500	4.123	0.042
trend				
MRTO4	20.433	15.767	3.581	0.058
MACROH2A2	24.833	15.667	3.470	0.062
MACROH2A1	13.667	24.500	2.903	0.088
UCK2	20.433	15.767	2.851	0.091
not significant				
FBXO11	14.500	24.500	2.526	0.112
PWWP2B	14.500	23.700	2.244	0.134

CREBBP	14.500	23.700	2.135	0.144
BCL2L10	15.667	24.833	2.084	0.149
SMARCA5	15.167	24.500	2.016	0.156
ERCC2	15.167	24.500	1.988	0.159
TET2	14.500	23.700	1.843	0.175
MRPL52	15.167	23.700	1.594	0.207
UCK1	15.167	23.700	1.580	0.209
ASXL1	15.167	23.700	1.569	0.210
RRP9	17.633	18.900	1.566	0.211
BAZ1B	15.667	23.700	1.353	0.245
NAA10	18.900	17.633	1.324	0.250
POLE3	16.100	18.900	1.311	0.252
TP53	15.167	20.567	1.058	0.304
MALSU1	15.167	24.500	1.053	0.305
POLR1E	15.667	20.567	1.006	0.316
P300	16.100	20.433	0.868	0.351
RRM1	23.700	15.667	0.753	0.386
RRM2	23.700	15.667	0.737	0.391
CDA	17.633	18.900	0.671	0.413
POLR1B	17.633	20.567	0.604	0.437
RRP1	17.633	20.567	0.557	0.456
DNMT3A	15.167	23.700	0.552	0.458
POLR1C	17.633	20.567	0.502	0.479
IDH1	17.633	18.900	0.413	0.521
MRPS26	15.667	23.700	0.406	0.524
FLYWCH1	14.500	23.700	0.376	0.540
BAZ1A	15.167	20.567	0.290	0.590
POLR1A	14.500	20.433	0.270	0.603
MRPL4	15.167	23.700	0.229	0.633
POLR3H	18.900	16.100	0.145	0.704
SLC29A1	17.633	20.433	0.140	0.708
IDH2	17.633	18.900	0.137	0.711
RIOX2	16.100	23.700	0.084	0.771
NAA15	18.900	17.633	0.076	0.783
CHRAC1	16.100	20.433	0.042	0.838
DNMT1	17.633	20.567	0.039	0.843

HELLS	17.633	18.900	0.017	0.897
POLR3D	15.667	20.567	0.003	0.960

Table 3.8: *Analysis of patient survival based on gene expression*

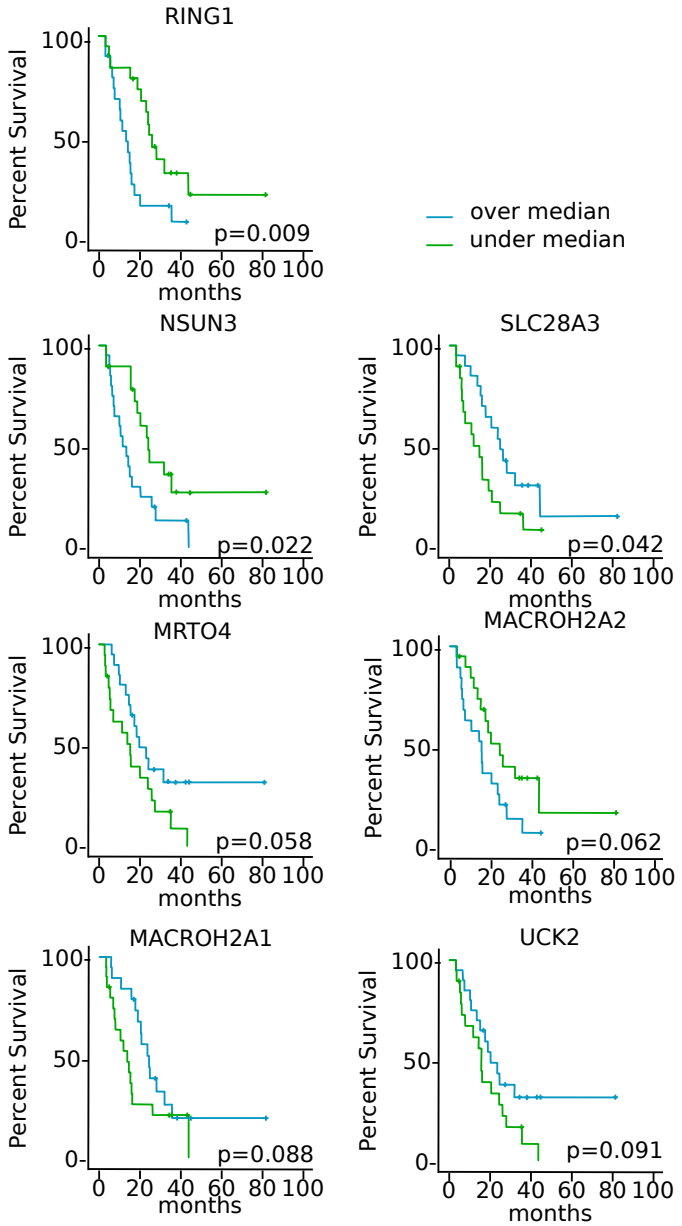


Figure 3.28: Overall survival by subgroups. Subgroups of patients were formed based on relative expression of genes measured by Nanostring, patients were categorized by a given gene expression as above or under the global median of expression. Only genes with lowest p value are plotted, *RING1, NSUN3, SLC28A3, MRTO4, MACROH2A2, MACROH2A1* and *UCK2*. n= 35.

3.3.7 Results III: conclusions

Using a cohort of patient samples, we analysed the mRNA expression for several genes of interest using the Nanostring technology. We could show that the CREBBP pathway was affected by *TET2* mutation. Higher levels of CREBBP were measured in patient who responded to Azacitidine treatment, but the level of expression did not affect the overall survival of patients. The levels of expression of various component of the CHRAC complex were also affected by *TET2* mutation and two of them (CHRAC1 and SMARCA5) had higher level of expression in samples from responders to azacitidine treatment. However here too it did not reflect on the overall survival of patients. *MACROH2A1* expression level was lower in TP53 mutant samples, it showed a slightly higher expression in patient responding to treatment which translated in the clinic to a longer overall survival.

Contributions This work was conceived by Marcus Buschbeck, Jeannine Diesch, and me, it was supervised by Marcus Buschbeck and Jeannine Diesch. The PBMC extraction at sample reception was done by Raquel Casquero. The sample preparation was done by me. The survival analysis was made by Olga Garcia. The rest of the analysis were done by Jeannine Diesch and me.

Chapter 4

Discussion

Discussion

The work presented in this thesis covers a wide range of subjects that could be discussed, I decided to focus on three main points. First, I will talk about the challenge of studying repetitive sequences and how to overcome it in the future. Then, I will discuss the relevance of macroH2A1 in MDS/AML in the light of our results and the current literature. And finally, I will discuss the role of gene expression as an informative read out in MDS/AML.

4.1 Discussion I: the challenge of studying repetitive sequences

4.1.1 The relevance of studying LINEs in relation to macroH2As

The ”coding DNA” is limited to 1-2% of the genome, historically the scientific community used to consider the remaining and main portion as unused and referred to it as ”junk DNA”. Indeed, the Human Genome Project completed in 2003 led to the identification of only around 20 000 genes instead of the expected 100 000. We have come a long way since then notably with the ENCODE project, which, in 2012, showed that 80% of the genome had a ”biochemical activity”. The definition was broad and any form of RNA transcription, whether it led to protein translation or not,

was labeled as functional. Now we are adding an additional complexity layer by studying the genome in 3D and part of this initial ”junk DNA” is now considered relevant for the spatial organization of the chromatin. This very brief summary of the evolution of the scientific knowledge regarding the relevance of all DNA sequences explains the interest we saw in studying the link between repeats and the histone variants macroH2As.

The project started on the basis of a ChIP-seq analysis showing that macroH2As were found in nucleosomes in some regions containing repetitive sequences. In our computational approach we could show a specific association of macroH2As with evolutionary younger LINEs and more specifically with the 3’ end (subsection 3.1.3). On one hand, younger LINEs are the only functional, thus potentially active, LINEs as with time the sequences of older ones have drifted. There is growing evidence of consequences of reactivation of LINEs in cancers (Tang et al., 2017; Guler et al., 2017). The inactivity of functional LINEs is attributed to epigenetic repressive marks such as histone H3 lysines 9 and 27 methylation, H2A.Z and HP1α (Rangasamy, 2013), all marks of the heterochromatin keeping LINEs in a silenced state. At the level of genome architecture, recent evidence showed that LINE-1 are preferentially more present in heterochromatic B compartments (Lu et al., 2021). On the other hand, macroH2As cooperate with other epigenetic factors such as core polycomb repressive complex protein, EZH2 (Buschbeck et al., 2009), both deacetylases HDAC1 and HDAC2 (Chakravarthy et al., 2005) in repressed regions. Thus, macroH2As interaction with LINEs could contribute to establish or maintain repressive mechanism on LINEs.

4.1.2 Dichotomy of the results

In our bench-based experiment we could not confirm a local enrichment of macroH2As on younger LINEs, nor the 3’end of LINEs. To explain this discrepancy there are two main possibilities:

technical issues did not allow to reproduce the ChIP-seq data even though there is a specific enrichment of macroH2As on LINEs, or there is a bias of analysis of the ChIP-seq data leading to artifactual enrichment. I'll briefly discuss a few points for each hypothesis.

Considering a technical bias, deposition of the insulator protein CTCF on the genome is well-studied and has not been associated with LINEs of any age or on any end. Thus, we used the CTCF ChIP-seq data as a control for the specificity of our computational approach and neither in the first analysis nor in the realignment on LINEs only, CTCF showed any association with LINEs. On the experimental side, repeated and repetitive sequences represent a challenge to design specific primers for PCR, independent if it is for expression or for ChIP. Additionally, primer design is based on the genome assembly, which is not complete and of variable quality especially in such regions.

Considering an analysis bias, the genome assembly covers 99% of gene-containing parts of human sequences but there are still some poor quality or misassembled sequences especially among repeats. This means that it is possible that reads were misaligned on known repeats when they actually came from a different part of the genome, which creates a false enrichment. Younger LINEs have less sequence variations, and thus, during a strict alignment on consensus sequences, reads from older elements are more likely to be excluded than reads from younger elements. Moreover, due to the mode of duplication of LINEs, the 5' end is the most truncated part and thus less present, which overall introduces an additional bias in the alignment process towards the 3'end. As for the control, since CTCF deposition is very localized on a set of specific sequence motifs, while macroH2As are much more widely spread with whole area covered, the patterns of deposition of these proteins on the genome are too different to be subjected to the same bias. Therefore, CTCF might not be the optimal control for the re-alignment experiment. On the bench side, we carefully validated tools to assess their specificity both in silico using websites

such as primer blast and primer3, or *in vitro* by standard curve and melting curve. We also used single-loci control to assess the quality of the ChIP experiment and knock-down of macroH2As to confirm the quality and the specificity of the immuno-precipitation.

With the current data, we could not discard either hypothesis, more experiments are needed to do so. For example, the use of long-read sequencing technologies would reduce the risk of misalignment, and therefore the bias in the analysis. Developing non-PCR based assays, for example in microscopy coupling LINE-1 FISH with macroH2As immunostaining could confirm the co-localisation of the two elements.

4.2 Discussion II: cell lines-based approach to clinical conditions

According to both The International Prognostic Scoring System-Revised (IPSS-R) (Greenberg et al., 2012) and the WHO Prognostic Scoring System (WPSS)(Malcovati et al., 2007), isolated 5q deletion are classed as low-risk/lower-risk of progression in AML. *MACROH2A1* is on the long arm of chromosome 5 and both macroH2A1 isoforms of the histone variant were shown to be downregulated at the RNA level in MDS patients with a 5q deletion compared to healthy donors (Bereshchenko et al., 2019). The histone variant macroH2A1 is implicated in many processes and has been described as having tumor suppressive and promoting functions (Hsu et al., 2021). Taking this into account, we decided to test the effect of the knock-down of macroH2A1s in three different cell lines: the two MDS/sAML derived cell lines SKK-1 and MOLM-13, and the cell line HL-60 derived from an acute promyelocytic leukemia. We could demonstrate that partial loss of macroH2A1 sensitized to some level both SKK-1 and MOLM-13 to treatments, while it mostly did not affect HL-60 cells. From these results we can discuss different aspects: How does the partial loss of macroH2A affect the response to Azacitidine and Decitabine treatment? And what could be the function of macroH2A1 in this context?

The hypomethylating agents Decitabine and Azacitidine are the recommended therapeutic treatment options for MDS patients with high IPSS score. Both azanucleosides integrate in DNA during DNA synthesis leading to an inhibition of DNA methyltransferase proteins, proper replication as well as DNA repair processes in cancerous cells (Hollenbach et al., 2010). In contrast to Decitabine though, Azacitidine gets mainly integrated into RNA, which results in an inhibition of protein synthesis (Diesch et al., 2016). Both treatments show different effects depending on the doses: at low doses they reactivate silenced genes and promotes cellular

differentiation due to DNA hypomethylation, while at higher doses they have mainly cytotoxic effects. In our experiments, the sensitizing effect of macroH2A1s knock-downs treated with Decitabine was mainly seen at low doses in all three cell lines. The response to Azacitidine treatment was not affected by the level of macroH2A in HL-60 and was mostly increased at low doses in MOLM-13 and SKK-1 (subsection 3.2.5 and subsection 3.2.2). The greater effect on Decitabine treatment compared to Azacitidine treatment implies that it is due to the integration of the azanucleoside into DNA, rather than the RNA-dependent effects that only Azacitidine induces. Furthermore, the effect at low doses indicate it is predominantly due to the DNA hypomethylation rather than the cytotoxic effects caused by both treatments.

It has been shown that there is a global hypomethylation in MOLM-13 and HL-60 treated with Azacitidine (Gawlitza et al., 2019). Furthermore, there is a change in methylation of genes' promoters in the apoptotic pathways such as *TP53*, *GADD45A* (Growth arrest and DNA damage-inducible, alpha factor) and *BID* (BCL2-associated transcription factor BH3-interacting domain death agonist) in MOLM-13 treated with Decitabine or Azacitidine (Janotka et al., 2021). Surprisingly, the increase in methylation is not consistent, and it does not correlate with a change in expression of those genes, which suggests it is not enough to alter those apoptotic steps. Hollenbach and colleagues also demonstrated that Azacitidine-mediated cell death is only partially due to an increase in apoptosis(Hollenbach et al., 2010). We did not detect a change in global apoptosis in MOLM-13 cells with macroH2A1 knock-down treated with Decitabine, whereas treatment with the pro-apoptotic ABT-199 was more cytotoxic in those cells (subsection 3.2.7). These results suggest that macroH2A1 loss probably affect the apoptotic pathway, but it might not be its main contribution to the increased effect in response to low doses of Azacitidine and Decitabine.

It also has been shown in a melanoma cell line that mac-

roH2A1 overexpression or knock-down affect cyclins, by increasing or decreasing cell proliferation, respectively (Lei, Long and Li, 2014). We did not observe any striking difference of proliferation in our cell lines in normal untreated conditions, but as the incorporation of Azacitidine into DNA depends on active proliferation further experiments should be conducted to measure the impact of the knock-down on DNA synthesis to confirm or discard this possibility.

Since there was a difference in the response observed in the different cell lines, looking at their mutational profile might give another possible explanation for this effect. SKK-1 cells harbor mutations in genes coding for the transcriptional co-repressor BCOR, and the RNA splicing factor U2AF1 and MOLM-13 in E3 ubiquitin ligase *CBL*, cytokine receptor *FLT3*, lysine methyltransferase *KMT2A* and neurofibromin *NF1*. Both cell lines are p53 wild-type (Palau et al., 2017). HL-60, derived from an acute promyelocytic leukemia, has many cytogenetic alterations and, importantly, is TP53 null (Wolf and Rotter, 1985). While there are certainly a lot of differences between the three cell lines used here, it would be interesting to test the implication of TP53 mutation. Even though MDS with a deletion 5q are considered low risk of progression to sAML, an additional *TP53* mutation increases the risk of progression (Jädersten et al., 2011).

Although our results do not establish a clear role of macroH2A1 in MDS/AML. Published data sets show that *MACROH2A1* is underexpressed in MDS and overexpressed in AML compared to normal haematopoietic stem cells (Hsu et al., 2019; Bagger, Kinalis and Rapin, 2019). This variation of expression reinforces the need to further explore the role of macroH2A1s in MDS/AML and their potential role in therapy as a biomarker or a target.

4.3 Discussion III: of the importance of epigenetics

Considering our results and the current literature, I will first discuss the link between known mutations and gene expression, then the value of gene expression as a predictive factor during diagnosis and finally the growing interest in harmonization of the experimental techniques to increase the reproducibility and comparison of clinical results.

4.3.1 The relation between mutational status and gene expression

MDS is an extremely heterogeneous disease. No driver mutations have been identified and only a few gene mutations are used routinely for diagnostic and treatment strategies. Among the 14 mutations frequently observed in MDS over half of them are epigenetic factors (Kuendgen et al., 2018). The relevance of studying epigenetic factors in MDS and AML is not new. There are several articles trying to combine mutational status and gene expression data in MDS and AML by different approaches, for example clustering patients based on their expression profile and then overlapping with the mutational status (Gerstung et al., 2015). Here, we approached the question the other way around, dividing the patient cohort dependent on their mutational status and then analyzing differential expression for a set of genes. We focused on a subset of mutations, *RUNX1, SRSF2, TP53* and *TET2*, for which we had at least 6 mutant samples to allow statistical analysis (subsection 3.3.4). I will comment here on the hits from the analysis in the context of the current knowledge.

Runt-related transcription factor 1 is a protein implicated in haematopoiesis especially in the differentiation between the lymph-

oid and myeloid lineage. The gene coding for this protein, *RUNX1*, is among the most frequently mutated genes in MDS. In our cohort, nearly 30% of the samples were annotated as mutant for *RUNX1* (subsection 3.3.1). At the level of expression, both anti -apoptotic *BCL2L10* and the transporter *SLC28A3* showed low expression levels and even lower in *RUNX1* mutant samples. *RUNX1* expression has been linked to mir18a which impacts *BL2L10* expression in HepG2 cells (X. Wang et al., 2018), the mutation could thus indirectly impact the gene expression. Chen and colleagues also recently showed the downregulation of *SLC28A3* in *RUNX1* mutants in a cohort of AML samples (Y. Chen et al., 2021). As SLC28A3 is involved in the response to treatment with Azacitidine and Decitabine and RUNX1 mutation is frequently detected, further exploration of the relation between both proteins might be interesting for treatment strategy.

SRSF2 mutations are normaly seen in 11%–15% of patients with MDS (Bejar, Levine and Ebert, 2011), and in half of our cohort. Serine and arginine rich splicing factor 2, SRSF2, contributes to both constitutive and alternative mRNA splicing by recognition of the 3'end of pre-mRNA (Masaki et al., 2019). *FBXO11* showed higher expression and *RING1* and *RRM2* showed lower expression in mutant samples. F-box only protein 11 is part of a ubiquitination complex that indirectly impacts the differentiation of B-cells and plasma cells, it is also a potential tumor suppressor in AML and the same study also showed an enrichment of *SRSF2* mutations in samples with a high expression of *FBXO11* (Schieber et al., 2020). RING1 is also part of the ubiquitination machinery. Mutation of *SRSF2* has already been shown to have a direct impact through alternative splicing of MDM2 (Comiskey et al., 2020) and an indirect one on protein ubiquitination (Pellagatti et al., 2018). This suggests that SRSF2 may impact not only on the splicing but also the expression of ubiquitination machinery components.

TP53 mutation is common in many cancers, it codes for p53 protein which is a pivotal point in many cellular processes,

DNA repair, cell cylcle arrest, apoptosis ect. As such it is identified as a gatekeeper for tumor suppression, though when mutated it is often overexpressed and contributes to tumour development (L. Zhang et al., 2017). In our cohort, 17% of the samples are TP53 mutant. Lower RNA level of *TP53* were found in the mutant samples, which could be either due to the mutation leading to a lower transcription or that the mutant mRNA was not detected by the assay. *TP53* and *TET2* mutations are often correlated, though in our cohort only one patient had both mutations. Interestingly, at the protein level *TP53* loss-of-function mutation has been recently linked to a decrease of the degradation of TET2 in a colon cell line (J. Zhang et al., 2019). We can hypothesize that the decrease of the protein's degradation triggers a decrease of its mRNA level in a retro-feedback loop. *NSUN3* and *MACROH2A1* expressions were also both strongly decreased in the mutant samples. MacroH2A1.2 has been known to directly contribute to the regulation of the expression of TP53 (Pliatska et al., 2018), and in alignment with our results, the mutation of *TP53* was shown to indirectly contribute to the repression of macroH2A1 (De Barrios et al., 2017). In our *in vitro* experiment, we used MOLM-13 (p53wt) and HL60(p53*null*), but the knock-down of *MACROH2A1*'s expression mainly affected the response to Azacitidine in MOLM-13. Considering the interdependent relation of macroH2A1 and p53, we suggest that the difference in response is, at least partially, due to the level of p53.

TET2 is a key player in DNA demethylation, it oxidizes methylated cytosine triggering their excision and replacement by a non-methylated cytosine (Kunimoto and Nakajima, 2021). Thereby TET2 activity impacts gene transcription. Thus it is not surprising that its mutation significantly impacted the mRNA level of various targets in our study. *TET2* mRNA levels were not significantly different based on its mutational status, which means the mutation did not affect its transcription or detection by the assay, while it does not give any indication on the activity of mutant protein. For most of the differentially expressed genes, based on the current

knowledge, it is difficult to say if it is a direct effect of the mutation or an indirect consequence. For example, there are no articles describing the relationship between *TET2* expression or mutation and *RING1* expression. Neither did we find any study linking TET2 mutation or expression to the expression of the gene coding for the mammalian mitochondrial ribosomal subunits *MALSU1*, *MRPL39* and *MRPS26*. But all three genes showed a lower expression in *TET2* mutant samples, which suggests a potential influence on the mito-ribosomal complex. It is known that TET2 activity has an impact on the mitochondrial DNA methylation and ATP level (Ji et al., 2018), and thus it could indirectly impact the expression of mitochondrial ribosomal subunits. It has been observed in other cohorts that in MDS and AML *TET2* mutation is associated with ASXL1 mutation (Rocquain et al., 2010; Bejar et al., 2012). In our case, 4 out of 11 TET2 mutant samples were also ASXL1 mutated. It can be also noted that TET2 mutant but ASXL1 wild-type patients show a better response to Azacitidine (Bejar et al., 2014; Kuendgen et al., 2018). We did not reproduce these results here as out of 11 patients with *TET2* mutation only two responded to treatment and one was mutant for *ASXL1.*

Three of the four components of the CHRAC1 complex were downregulated in samples from patients with *TET2* mutations. Some prior articles loosely established a relation between both. In breast cancer microRNAs, TET2, BAZ1A and SMARCA5 were shown to be down-regulated in tumorous versus healthy tissues (Lee et al., 2013). In a melanoma cell line, the expression of BAZ1A, among 585 other genes, was shown to be TET2/glucose dependent (D. Wu et al., 2018). Various microRNAs (200c, 100, 19c) have been studied in relation to SMARCA5 and TET2, both downstream and upstream. So far, no direct or indirect mechanism of interaction has been proposed, but considering our results this could be the starting point of a mechanistic study of the relation between TET2 and the CHRAC complex.

Overall, several interesting findings came from this ana-

lysis, notably the relation between the CHRAC complex and TET2 but also the mito-ribosomal machinery, and TP53 and macroH2A1s. This could be the starting point of different more mechanistic studies for a better understanding of the development and progression of MDS.

4.3.2 The relation between response to Azacitidine and gene expression

The second part of our analysis combining gene expression with the response to Azacitidine, could be more relevant in the clinic as predictive or monitoring factor of response. Our analysis provided two separate read-outs relating gene expression and clinical outcome: the response to treatment and the overall survival. Here, I will put both in the perspective of the current knowledge. Before going into details, the expression measurement was done on samples at diagnosis, thus the impact of the expression of several of the studied genes, especially genes implicated in Azacitidine metabolism, can only be used to conclude the predictive value of the levels of expression.

UCK1 is an enzyme essential for the metabolism of Azacitidine and its integration into RNA and DNA (Diesch et al., 2016). Levels of the kinase UCK1 were higher in responders compared to non-responders to Azacitidine treatment, in accordance with what has been previously reported (Valencia et al., 2014). This reinforces the idea of using *UCK1* expression as a predictive biomarker of the response to Azacitidine treatment. Among the rest of the genes encoding nucleoside transporter proteins and enzymes involved in the metabolism of Azacitidine, there were no statistical difference between responders and non-responders, but we can see some tendencies aligned with current knowledge. For example, the level of the degradation enzyme, CDA, was shown to correlate with the response (Qin et al., 2011), and in our cohort we see the same tendency. More recent studies showed that CDA ex-

pression affects response to Azacitidine treatment and is also dependent on gender (Mahfouz et al., 2013), and that CDA inhibitors increased Decitabine efficiency (Gu et al., 2021). Those studies are not directly comparable as some are done in mice and others *in vitro* with cell lines or a mix of AML and MDS samples, but these conflicting results suggest that CDA alone is not a suitable candidate as a predictive biomarker. The expression of *BCL2L10* has been reported as a marker of the response to Azacitidine treatment (Cluzeau et al., 2012) but in our cohort as well as in others we did not see this effect (K. Kim et al., 2020). These discrepancies might be explained by the difference of composition of the cohort and/or the definition of the responders/non-responders, as there is no distinction made between the partial and complete responders. This explanation is backed-up by the data of overall survival showing that *SLC28A3* and *UCK2* levels inversely correlate with the overall survival. In MDS samples, it has been shown that the kinase *UCK2* level did not change between responders and non-responders (Valencia et al., 2014) but that increasing its expression restores sensibility to treatment (Gu et al., 2021). These can appear as conflicting results, but we can suppose that independently of the starting level of UCK2, the expression may be affected differently in response to azanucleoside treatments.

We included in our panel numerous genes that were implicated in protein synthesis considering some of our previous results (Diesch et al., 2021). CREBBP was a hit in the loss-of-function screen and in our study we could show that CREBBP inhibitors had a significant impact on protein synthesis and could be used in combination with Azacitidine. In this cohort both *CREBBP* and *MALSU1* were unexpectedly downregulated in non-responders, but as their levels were quite high it does not exclude that the inhibition of CREBBP combined with Azacitidine can positively impact the overall survival.

Based on our loss-of-function screening we also included the components of the CHRAC complex, the ATPase sub-

unit SMARCA5, and its co-factors BAZ1A, CHRAC1 and POLE3. Their level of expression did not impact the overall survival but *CHRAC1* and *SMARCA5* showed a decrease of expression in non-responders. In CD34+ AML progenitors *SMARCA5* expression was shown to be higher than after complete haematologic remission, and demethylation by Azacitidine treatment restored positioning of SMARCA5 on the genome (Z. Wang et al., 2021). *SMARCA5* and *BAZ1A* expressions were both shown to be upregulated in mES cells upon increase of DNA methylation(Saksouk et al., 2014). *POLE3* mutants showed an impact in T and B cell development (Siamishi et al., 2020). This suggests that the CHRAC complex or some of its components might play a role in AML/MDS due to aberrant DNA methylation. We showed *in vitro* that the knock-down of BAZ1A had a variable sensitizing effect to Azacitidine in some cell lines, the difference of impact may be linked to global or local methylation levels. Although our study is not conclusive on the effect of BAZ1A expression in the response to treatments, taken together with the results of the expression analysis in diagnostic patient samples and the hints in literature of a relation with other components of the CHRAC complex, it seems that it would be worth to further explore the role of BAZ1A in MDS/AML, especially with the newly found BAZ1A inhibitors (Yang, Zhou and Zhong, 2021).

TET2 is mainly studied in terms of mutation in the context of MDS/AML. We found some data of expression but not related to the response to Azacitidine treatment. The expression did not affect the overall survival, but responders had higher levels than non-responders. *TET2* mutations are consistently associated with a decrease in activity, which could mimic a lower expression level. There is a controversy on the prognostic value of *TET2* mutation as it has been associated with both favorable and worse prognostics (Bejar et al., 2014; Kuendgen et al., 2018).

4.3.3 Patient samples: benefit and limitations

In this work we used cell lines for fundamental research as well as for more translational research and we used frozen samples from patients and patient data. Cell lines confer great advantages in terms of reproducibility and logistics (easy access, unlimited amount of cells etc). They are a great tool for both fundamental and translational research but they also come with limitations as *in vivo* there is much more heterogeneity. Access to patient samples is always limited, though working on liquid tumours such as MDS and AML, it is a bit easier as it is for solid tumors. In the future, we could further expand our study to combine our findings with larger data sets. As it is a disease with intrinsic and extrinsic heterogeneity and the validation of predictive markers requires a large and maybe more defined cohort of patients. Many other studies are using MDS/AML samples and as there is an international effort to standardize diagnostic criteria some comparison can be done to further validate findings and increasing credibility. When doing such comparison, we often are faced with differences of cohort composition, size, differences in read-outs (e.g., mutation, expression, proteomics) and differences in techniques (e.g. RNAseq, Nanostring) which limit comparability. The size of the cohorts and subgroups (e.g. responders/non-responders, TET2 mutants, etc.) defines the statistical power of the analysis. There are more and more articles trying to compile data from different studies, which also face the same technical issues. For better understanding of the disease, there has been global effort to harmonize the diagnostic criteria with the WHO nomenclature, but also to share clinical information with initiatives like the European MDS Registry (EUMDS) that regroup information on treatment and disease from patients who consent to participate from 18 countries. The number of biobanks is increasing, to collect, process and store voluntarily donated samples, like the IJC Leukemia and other blood disease Sample Collection. These initiatives and the efforts of har-

monization of the experimental techniques have already started to increase the benefit of using clinical samples with more robust statistics. This harmonization is one of the key aims of the HARMONY alliance. Researchers still need to improve on communication to a larger audience to explain the importance of increasing samples collection and standardized read-outs in order to improve knowledge and *in fine* patient care and quality of life.

Chapter 5

Conclusions

In the first part of this study, we analysed the relation between macroH2As and repetitive sequences and came to the following conclusions:

1. The computational analysis massive parallel sequencing data of macroH2A-enriched chromatin fractions suggests a specific e enrichment on repeat elements of the class of LINEs, specifically the 3'end of younger LINEs.

2. Experimental approaches including PCR analysis after chromatin immunoprecipitation of macroH2A1 do not confirm these previous results

3. Perturbing the levels of macroH2As does not affect the transcription of LINEs.

In the second part of this study, we functionally tested chromatin regulators in cell culture models with the long-term aim to identify novel potential therapeutic targets for MDS and sAML therapy. Specifically, we applied a loss-of-function screen that allowed us to derive the following conclusions:

1. *NAA15* contributes to the sensitivity of SKK-1 cells to Azacitidine and Decitabine and its downregulation can lead to resistance.

2. Genetic inhibition of components of the chromatin remodelling complex CHRAC such as *BAZ1A* sensitizes some MDS-derived cells to Azacitidine treatment.

3. The knockdown of *MACROH2A1* has variable effects on the response of AML cells to drugs. These effects include increases in sensitivity of SKK-1, MOLM-13 and HL-60 to azanucleoside and the Bcl-2 inhibitor ABT-199.

In the last part of this work, we studied the expression of a set of chromatin regulators and other disease-related genes in cohorts of MDS/AML patients at diagnosis. The conclusions from this study are the following:

1. Several genes encoding chromatin regulators are deferentially expressed comparing bone marrow samples from MDS patients and healthy donors. This includes various component of the CHRAC complex, MacroH2A1, CREBBP.

2. Mutation status correlates with differences in expression of some chromatin regulator-encoding genes. For example, the expression of various components of the CHRAC complex is decreased in patients with *TET2* mutations.

3. The expression of several chromatin regulators genes differs between responders and non-responders. CHRAC1 and SMARCA5 have higher expression in responders compared to non-responders.

4. The expression of RING1, NSUN3 and SLC28A is associated with overall survival. For MACROH2A1 and genes encoding members of the CHRAC complex we do not observe significant associations. The statistical power is limited by the small size of our cohort.

Chapter 6

Methods

Methods

6.1 Molecular biology

6.1.1 RNA expression

RNA isolation

RNA from cell lines The RNA samples were prepared according to recommendations from the Maxwell® RSC simplyRNA Cells Kit on ice till loading on the instrument.

Cells, from a dry frozen pellet or a fresh pellet, were homogenized in 200µL of chilled 1-Thioglycerol/Homogenization Solution, then 200µL of Lysis buffer were added for lysing cells, (vortexing till homogenization). The mix was transferred to the cartridge. 5µL of DNAse I solution to the wells 4 and 40µL of nuclease free water to the 0.5mL elution tubes. Finally, the cartridges are loaded in the Maxwell® RSC Instrument and starting the RSC simplyRNA Cell method. At the end of the run samples were kept on ice for the next step or transferred to safe-lock Eppendorf tubes and stored at -80°C.

RNA from patient samples RNA extraction of patient samples was prepared using Qiagen RNeasy kit according to recommendations.

Cells were frozen in 300µL RLT buffer + b-Mercapto and stored in -80°C. Samples were thawed out on ice, transferred to the column and spun 15s at 8000g. 350µL of RW1 buffer was added to the column then spun 15s at 8000g. DNase digestion was done on-column by adding 10µL of DNase I mixed with 70µL of RDD buffer and incubate it 15min at room temperature. Then successive washes were performed by adding 350µL of RW1 buffer, then 700µL, then twice 500µL of RPE buffer and centrifuging 15s at 8000g between each step. The last centrifugation is prolonged to 2min. Membrane was dried by a 1min centrifugation at full speed. Then RNA was eluted using 30µL of water 1min at 8000g. RNA concentration and purity were measured using a nanodrop. 100ng in 5µL final volume (20ng/µL) were prepared in a safe-lock tube and stored at -80°C (as well as the stock) before being sent for the Nanostring assay.

In each tube are added in order 8µl of mastermix (Reporter barcoded probes + hybridisation buffer containing known concentrations of positive and negative controls), 5µl of RNA sample and 2 µl of capture probes and then incubated 20h in a thermocycler (65ºC, lead at 70ºC) for hybridisation. Then hybridised material is passed to the prep-station where the hybridized molecules are captured on the chip and the rest is washed out. On the digital analyser 555 frames are taken for each samples and each barcoded probes is counted. The analysis was done using nSolver 4.0. Samples need to pass four parameters of quality control: imaging QC, binding density, positive control and limit of detection. Image QC evaluate the percentage of usable frames (¡75%), binding density is a range of molecule per µm2 for a reliable count (0.5-2.25), positive control evaluate the deviation of the measurement compared to the theoretical curve, and limit of detection is set based on the measured standard curve . Next Negative Background Threshold Parameters were confirmed. Next both Positive Control Normalization and mRNA CodeSet Content Normalization

were run to obtain normalized data by two housekeeping genes (GUSB and TUBB).

cDNA synthesis by reverse transcription

The cDNA synthesis was done according to recommendation from the First Strand cDNA Synthesis kit K1612.

1µg of RNA as template, and each sample is processed with and without the reverse transcriptase (control for DNA contamination). All reagents were mixed according to Table 6.1, then load on thermocycler for 5 minutes at 25°C, 60 minutes at 37°C, and 5 minutes at 70°C. Then stored at -20°C.

Reagents	Quantity per well
RNA	1µg
Random hexamer primer	1µL
Reaction Buffer	4µL
Ribolock	1µL
dNTP	2µL
Reverse Transcriptase	2µL
Water, nuclease-free	top up to 20µL

Table 6.1: cDNA synthesis mix composition

Quantitative real-time PCR (RT-qPCR)

Standard curve For each primer pairs, the validity was first evaluated using a standard curve to confirm that the measurements were quantitative followed by both an analysis of the melting curve and a migration on an agarose gel to confirm only one product was amplified.

To set the standard curve genomic DNA or a pool of cDNA (for ChIP or for expression analysis respectively) were diluted in series in water at 1/2, 1/10, 1/50, 1/250, 1/1250.

A mix containing 5µL SYBRGreen (Roche, Switzerland), 0.5µL of forward and 0.5µL of reverse primer (stock at 10 µM) and 2µL of water was mixed with 2µL of DNA template in a well of 384-well plate while on ice, three technical replicates were included for each point. RT-qPCR conditions are listed in **Table 6.2**, reactions were run on QuantStudio™ 12K Flex Real-time PCR System from Thermo Fisher Scientific. Using R, results were plotted CT over log (DNA dilution), and the equation of the best fit line was used for calculation of the real efficiency, $e = 10^{-1/slope}$.

Step	Temperature (°C)	Duration	Cycles
Preincubation	95	8 min	1
Amplification	95 62 72	10 sec 20 sec 15 sec	45
Melting curve	95 95 to 62	15sec 1min	1
Cooling	40	30 sec	1

Table 6.2: qPCR conditions

Quantitative PCR For each combination sample/primer pair, two technical replicates were prepared (three for targets related to repeats). A mix containing 5µL SYBRGreen (Roche, Switzerland), 0.5µL of forward and 0.5µL of reverse primer (stock at 10 µM)

and 2µL of water was mixed with 2µL of DNA template in a well of 384-well plate while on ice. RT-qPCR conditions are listed in **Table 6.2**, reactions were run on QuantStudio™ 12K Flex Real-time PCR System from ThermoFisher Scientific. cDNA was top upped to 70µL before loading the 2µL of template. GAPDH and RPLP0 primers are included and used as housekeeping genes. Samples without reverse transcriptase are loaded in parallel as a control of DNA contamination of the RNA extract for LINEs expression experiment.
Results were analyzed using R. For expression, data analysis is done in R by normalizing with the mean of both housekeeping gene. All quantifications of RNA expression are relative, so that the chosen control sample will have an expression of 1 for all target genes and the rest of the samples will have an expression that represents a multiple or fraction of the expression in the control. RT-qPCR raw data were normalized by the mean of the housekeeping genes. Each experiment was averaged between technical replicates then variance was verified by *var.test* before applying a *t.test*.

Nanostring processing

In each tub were added in order 8µl of mastermix (Reporter barcoded probes and hybridisation buffer), 5µl of RNA sample and 2 µl of capture probes and then incubated 20h in a thermocycler (65ºC, lid at 70ºC) for hybridisation.

Then samples were transferred to the prep-station where the hybridized probes are captured on a cartridge covered by straptavidin and the excess was removed. Cartridge was then scanned on the digital analyser , by an automated fluorescence microscope. 555 frames are taken for each samples and each barcoded probes were counted. The analysis was done using nSolver 4.0. Samples need to pass four parameters of quality control: imaging QC, binding density, positive control and limit of detection. Im-

age QC evaluate the percentage of usable frames ($<$75%), binding density is a range of molecule per µm2 for a reliable count (0.5-2.25), positive control evaluate the deviation of the measurement compared to the theoretical curve, and limit of detection is set based on the measured standard curve . Next Negative Background Threshold Parameters were confirmed. Next both Positive Control Normalization and mRNA CodeSet Content Normalization were run to obtain normalized data by two housekeeping genes (GUSB and TUBB). For both RT-qPCR and Nanostring data, statistical analysis was done using R. Visualization and statistical analysis were done using R. Each time two data sets of expression were compared (either responder vs non responders, or non-mutated vs mutated for gene X), Statistical analysis was run only if each set had at least 6 measurements. The normal distribution of both sets was evaluated using *shapiro.test*, if both were normally distributed a Welsh test was used using *t.test(Var.equal=FALSE)* otherwise a two-sided Wilcoxon test was applied (*wilcox.test*).

6.1.2 Protein expression analysis

Protein extraction and quantification

Cells were collected and washed with PBS twice, then they were re-suspended in 200-300µL (depending to the pellet's size) of RIPA buffer (**Table 6.3**) for lysis on ice during 10min. The lysate were transferred to a sonication 1.5mL tube and sonicated 5-7 cycles (30sec ON/30sec OFF), using Bioruptor® Pico sonication device. The sonicate was then passed through a 1mL syringe 4-8 times to shred long DNA molecule, then it was centrifuge at 10 000rpm during 5min to pellet insoluble debris and recover soluble proteins in the supernatant.

Reagent	Concentration
NaCl	150mM
Tris-HCl, pH 8.0	50mM
NP-40	1%
Sodium Deoxycholate	0.5%
SDS	0.1%
EDTA	5mM
PMSF	1mM

Table 6.3: RIPA buffer composition

Protein quantification was done using BCA Protein Assay Kit (ThermoFisher) using recommended protocol, (working range = 20–2000 µg/mL). BCA standards were prepared beforehand by dilution in series and kept at -20°C. Concentration ranged from 25 to 2000 µg/mL. 25 µL of each standard or sample (diluted at 1/2 or 1/10) were pipetted to a 96-well microplate in triplicate. Then 200µL of the mixed reagents (50:1, Reagent A: B) were added. The plate was protected from light with foil and left to incubate 30min at 37°C. The absorbance was then measured at 562nm on a plate reader (Synergy H1 Hybrid Multi-Mode Reader). To calculate the protein concentration the average measurement of the blank standard replicates was first subtracted to all measurement. Then

the standard curve was plotted using each BSA standard corrected measurement vs. its concentration in µg/mL, then the equation of the curve $y = A \times x + B$ is determined. Sample concentration is calculated according to $x = \frac{(y-B) \times dilution\,factor}{A}$ where x is the sample concentration.

Western blot (SDS-PAGE)

Samples' concentrations were homogenized with the appropriate volume of Laemli (5X) and buffer, then samples were heat up to 95°C during 5min. Then corresponding volumes were loaded on the migration gel. We used different percentage of polyacrylamide depending on the protein according to their molecular weights : 14% for histone H3 (15 kDa) only, 10% for both histone H3 and macroH2As (40 kDa) and 6% for BAZ1A(178 kDa) (see **Table 6.4**. Migration was run at 36 mA for 45-90 min to achieve optimal separation and then transferred at 220 mA (90min for H3 and macroH2A, overnight for BAZ1A) onto a nitrocellulose membrane (GE healthcare). After transfer, protein load was checked by Ponceau staining and the membranes were blocked using 5% non-fat milk (Nestlé) in TBST for 30 min. Membranes were incubated with primary antibodies o/n at 4°C on a orbital shaker. The next day, membranes were washed thrice with TBST for 10 min and incubated with secondary Fluorophore-conjugated secondary antibodies (LI-COR Biosciences IRDye 680RD and IRDye 800CW) (1:20000) for 1h at RT. Membranes were then washed again thrice with TBST for 10min. The dried membranes were scanned with an Odyssey® CLx Imager. All information regarding antibodies can be found in the Material chapter. Image Studio software was used for signal quantification and H3 signal served for normalization.

Reagent	6%	10%	14%
Acrylamide 30%	1.95mL	3.3mL	4.6mL
Tris 1.5M pH 8.8 + SDS	2.5mL	2.5mL	2.5mL
H2O	5.75mL	4.4mL	56.2mL
APS 10%	80µL	80µL	80µL
Temed x2	8µL	8µL	8µL

(a) Polyacrylamide resolving gel

Reagent	Volume for 1 gel
Acrylamide 30%	500µL
Tris 0.5M pH 6.8 + SDS	380µL
H2O	2.1mL
APS 10%	30µL
Temed x2	3µL

(b) Polyacrylamide stacking gel

Table 6.4: Polyacrylamide gel

6.1.3 Cloning

shRNA The *BAZ1A* and *NAA15* ShRNA shRNA fragment were PCR-amplified using primers 5'-miRE-Xho1 and 3'-miRE-EcoR1 and cloned into the cSGEP plasmid encoded with a lentiviral backbone as before (Fellmann et al., 2013). BAZ1A and NAA15 ShRNA sequences were also acquired from the Fellmann et al., 2013 study. The Ultramer oligos were dissolved in 120 µL of ddH2O to give a stock concentration of 1 µg/µL. For each oligo, a PCR was set up with the forward and reverse primers flanking the gene-specific guide sequences at EcoRI and XhoI restriction digestion sites (composition see **Table 6.5**). Amplification of the Ultramer oligos was confirmed by running 2µL of the PCR product on a 2% (weight/volume) agarose gel, which showed a single band 131 bp in size. 25 µL each of four amplified PCR reactions were pooled for subsequent

column purification. The QIAquick PCR Purification Kit (QIAGEN, Germany) was used according to the manufacturer's instructions with the sole modification of eluting the oligos with 30 µL of EB after incubating 10 minutes on column.

The *MACROH2A1* shRNAs were already cloned in another backbone. All sequences can be found in Material chapter

Reagent	Volume (µL)
Phusion Hot Start II HF PCR master mix	25
Forward primer (5'miRE-XhoI)	2.5
Reverse primer (3'miRE-EcoRI)	2.5
Template (Ultramer)	1
Water	19

Table 6.5: PCR master mix for ultramer amplification

98°C	30 sec	
98°C 66°C 72°C	10 sec 30 sec 30sec	35 cycles
72 °C 4°C	10 min	

Table 6.6: PCR conditions (shRNA Ultramer amplification)

shRNA insertion in cSGEP The purified PCR products and the plasmids were then digested using EcoR1 (10 U/µL) and Xho1 (10 U/µL) restriction enzymes in 2x Tango Buffer (ThermoFisher Scientific, USA). In parallel, digestion of the cSGEP plasmid vector was also carried out. Digestion reaction lasted 60 minutes at

37°C. The digested cSGEP plasmid vector was treated with alkaline phosphatase to avoid self-ligation. The digested samples were then run on a 2% (weight/volume) agarose gel to confirm the presence of the 110 bp band of the inserts and approximately 8 kb of the digested plasmid. The band was then excised using the QIAquick Gel Extraction Kit (QIAGEN, Germany), according to the manufacturer's instructions, and eluted in 50 µL. DNA concentration was measured by NanoDrop. The purified digested oligos were ligated using T4 DNA Ligase (ThermoFisher Scientific, USA) with the insert (4 ng), cSGEP vector (100 ng) in a molar ratio of approximately 3:1. A vector only control was also included to confirm that there was no re-ligation. The ligation mix was incubated overnight at 16°C.

The next day, 5 µL of the ligated reaction was transformed into chemically competent Stbl3 E. coli (ThermoFisher Scientific) and incubated on ice for 30 minutes. Meanwhile 1 mL/sample of SOC medium (super optimal broth with catabolites repression) was pre-warmed at 37°C. The mix was placed at 42°C for 45 seconds and returning to ice for two minutes, for heat shock. 300 µL pre-warmed SOC was added and incubated at 37°C with 300 rpm shaking to allow the bacteria to recover. The bacteria mix was centrifuged for 2 minutes at 2000 rpm. Most of the supernatant was removed, leaving around 50 µL with the pellet. The pellet of the bacteria mix was then plated on LB-ampicillin agar plates and incubated overnight at 37°C. The next day, successful transformations were visualized with the growth of ampicillin-resistant colonies. 10 colonies per shRNA pool were picked using a pipette tip and incubated overnight at 37 °C in 3mL of LB with Ampicillin. The next day, plasmids were extracted using PureLink Quick Plasmid Miniprep Kit (ThermoFisher Scientific), as per manufacturer's instructions. DNA of the isolated plasmids was then quantified using NanoDrop. Samples containing 500 ng DNA mixed with the ZUB-SEQ-SH primer were then sent for sequencing to confirm shRNA sequence's integration.

6.1.4 Chromatin immunoprecipitation

Nuclei extraction and chromatin preparation

Between 20 and 30 million HepG2 cells are trypsinized and fixed in a PBS solution containing 1% PFA and 10% FBS for 10 min at RT in rotation. Glycine (0.125M) is used to stop the reaction in rotation for 5 min at RT, then cells are washed with cold PBS. The cell pellet is either lysed directly or stored at -80°C.

For the nuclei extraction, the resuspended cells are incubated for 30min rotating at 4°C in 500 µL lysis buffer I (5 mM PIPES pH 8, 85 mM KCl, 0.5% NP-40, 1 mM PMSF, 50 µg/mL leupeptin). Lysed cells are centrifuged at 5000 rpm for 5 min at 4°C, the supernatant is discarded and the pellet is resuspended in 500 µL lysis buffer II (1% SDS, 10 mM EDTA pH 8, 50 mM TRIS pH 8, 1 mM PMSF, 50 µg/mL leupeptin) and incubated for 30min at 4°C in rotation. Chromatin is sheared by sonication in a Bioruptor® Pico (Diagenode, 10 cycles 30sec ON/30 sec OFF at high intensity). Insoluble debris are removed by centrifugation for 10 min at 10000rpm.

Samples are kept on ice at 4°C while an aliquot is used to assess the quality of the chromatin. 10 µL of lysate is mixed with 90 µL of water, de-crosslinked for 1 h at 65°C, purified using a PureLink™ Quick Gel Extraction and PCR Purification Combo Kit (Invitrogen), quantified using a Nanodrop (Thermo Scientific) and 500ng is run in a 1% agarose gel electrophoresis. Chromatin fragments should be between 300-500bp.

Immunoprecipitation

Based on aliquot quantification, 15 µg of sheared chromatin is used for each immunoprecipitation (IP) reaction diluted

with 9 volumes of IP dilution buffer (1% Triton X-100, 150 mM NaCl, 2 mM EDTA pH 8, 20 mM TRIS pH 8, 1mM PMSF).

Samples are precleared with 20 µL of Magna ChIP™ protein A/G magnetic beads (Merck Millipore) for 2h at 4°C in rotation. 10% of the precleared lysate is taken as input and stored on ice. Between 1 and 5 µg of antibody is added and the IP samples are incubated o/n at 4°C in rotation. The following day 20 µL of Magna ChIP™ protein A/G magnetic beads (Merck Millipore) are added to capture the antibody-bound chromatin fraction for 2h at 4°C in rotation. Beads are spun separated with a magnet (DynaMag™-2 Magnet) and washed twice at RT with each of the following buffers in this order: mixed micelle was buffer (140 mM NaCl, 20 mM Tris-HCl pH 8, 5 mM EDTA pH 8, 5% sucrose w/v, 1% Triton X-100 v/v, 0.2% SDS, 0.02% NaN3), LiCl/detergent wash (0.5% deoxycholic acid sodium salt w/v, 1 mM EDTA, 250 mM LiCl, 0.5% NP-40 v/v, 10 mM Tris-HCl pH 8) and buffer 500 (0.1% deoxycholic acid sodium salt w/v, 1mM EDTA, 50mM HEPES pH 7.5, 500 mM NaCl, 1% Triton X-100 v/v, 0.02% NaN3). First wash is done quickly resuspending the beads and the second in a rotating wheel for 3 min. Finally, a quick wash with TE buffer (10mM Tris-HCl pH8, 1 mM EDTA pH 8) was performed. The enriched chromatin is recovered by resuspending the beads in 200 µL of elution buffer (1%SDS, 100 mM NaHCO3), 30 sec of vortexing and incubation for 30 min at RT in rotation twice.

The input samples were included again in the experimental procedure at this point. Crosslink is reversed by adding NaCl to a final concentration of 200 mM and incubating the samples o/n at 65°C shaking. The following day the proteins in the samples, are digested by adding 2 µL of Proteinase K 10 mg/mL, 16 µL 1M TRIS pH 6.5 and 8 µL 500 mM EDTA and incubating for 2 h at 45°C.

After digestion, the DNA is purified using ChIP DNA Clean

and Concentrator columns, eluding in 90ul of water.

Quantification

DNA is measured as in **section 6.1.1**, with the input samples being diluted 10-fold prior loading into the plate. Efficiency of each primer pair was determined and used to calculate the relative ChIP enrichment. The input of each sample was used as the "control" to obtain a relative enrichment over input, using the same formula than for RNA expression. The relative enrichment was then converted to a percentage. The percentage of input value is used in all the ChIP plots.

$$input\% = \frac{100}{2^{\Delta Ct(Ct^{ChIP} - Ct^{input})}}$$

For statistical analysis, each experiment was averaged between technical replicates then variance was verified by *var.test* before applying a *t.test*.

6.2 Cell culture

6.2.1 Culture conditions

adherent cells HepG2 and 293T cells were cultured in DMEM (Dulbecco's Modified Eagle Medium) supplemented with 10% FBS, L-Glutamine and penicillin-streptomycin on treated culture ware at 37°C in 5% CO2. For passage, media was removed then cells were washed with sterile PBS once before adding Trypsin-EDTA and incubating 4-7min at 37°C. Cells were re-suspended in media for counting (Neubauer chamber) and diluting for seeding.

suspension cells MOLM-13, SKK-1 and HL-60 were cultured in RPMI (Roswell Park Memorial Institute) supplemented with 10%

FBS, L-Glutamine and penicillin-streptomycin on non-treated culture ware at 37°C in 5% CO2. For passage, cell's concentration was measured using cell counter Countess™ II for dilution. During cell maintenance then percentage of live cells was lower than 75%, cells were spun for 10 min at 200g, and supernatant removed, and fresh media added to remove a maximum of dead cells.

concentration ($\times 10^6/mL$)	SKK-1	MOLM-13	HL-60
Maintenance	0.2-1.5	0.4-1	0.2-1.5
Infection	3.1		
Survival assay	0.11	0.087	0.87
Competitive growth assay	0.11		

Table 6.7: Suspension cells working concentration

6.2.2 Gene transduction

4×10^6 HEK293t cells were seeded in 8mL DMEM media in 100 mm culture plates and returned to the incubator overnight so as to allow them to adhere to the culture plate. The next morning a DNA master mix containing the packaging plasmid psPax2 and the envelope-producing plasmid pCMV-VSV-G and short hairpin plasmid then CaCl2 was added (see Table 6.8. The mix is added to HEPES-buffered saline (HBS) drop by drop while vortexing and incubated for 15 minutes at room temperature. Lastly the mix was added to the cell plate and cells were returned to the 37°C incubator until the afternoon when the media is changed for fresh media.

The supernatant, containing viruses, was collected the next day in the morning and replaced by fresh growth media and again in the afternoon. The viral supernatant was either used straight away or frozen at -80°C till used. When recovering viral supernatant, it was passed through a 0.2µm filter using a syringe prior to storing in safe-lock tubes. During all these steps and the following ones, extra measures were taken to avoid any spillage of live viruses, any disposable plastic was decontaminated with

bleach and exposed to UV light for 30min before being carefully disposed.

Reagent	per 10cm plate
pasPAX2 pCMV-VSVG shRNA plasmid dH2O	10µg 3µg 10µg top up to 350µl
CaCl2 2mM	50µl
HEPES-buffered saline (HBS) 2X	400µl

Table 6.8: Transduction mix

Viral infection Polybrene was added to viral supernatant at 8 µg/mL before adding to cell in suspension (for concentration see Table 6.7) in 6-well treated plate. Cell plates were spun 1400rpm, 30min, 37° to favor cell infection by viruses and then place back in the incubator. The process is done twice (morning and afternoon). The next day cells were passage and the following day they were treated with puromycin to select infected cells (small aliquot was passed through flow cytometry to measure the percentage of infected/GFP positive cells). 2-3 days later the GFP positive cells percentage was measured to confirm selection. As an additional control a non-puromycin-resistant plasmid was used and after selection all cells must be dead.

The knock-down was confirmed by RT-qPCR and western blot when possible.

6.2.3 Screen

The screen viral supernatant and infection were prepared similarly to the method detailed above. To achieve 1000x representation of the library considering a 10% infection efficiency

7.3×10^7 cells were transduced in triplicates with the hEPI9 library. After puromycin treatment half of each replicate was treated with 0.075 µM Azacitidine every two days for 21 days followed by gDNA extraction and library preparation as previously described (Fellmann et al., 2011). The analysis of the sequencing results was performed using the R package EdgeR (Robinson, McCarthy and Smyth, 2010) as described (Dai et al., 2014). Gene level analysis was performed with the function Roast (D. Wu et al., 2010)and the results ranked on a gene-by-gene level.

6.2.4 Dose-response

Suspension cells were seeded at the appropriate concentration (see Table 6.7 in 24-well non-treated plates before adding the corresponding amount of either Azacitidine, Decitabine, Ara-C or ABT-199 and the corresponding control H2O or DMSO. All drugs were prepared at an intermediary concentration and protected from light. Lastly a stock aliquot was never used twice or kept for more than a month at -80°C to avoid degradation and limit variability of treatment effect. An aliquot of each well was taken after four days of incubation at 37°C for cytometry analysis.

6.2.5 Competitive growth assay

For the pilote experiment for the competitive growth assay, cell concentration was measured using the cell counter. Then the overall number of cells was taken and spun to remove media and resuspended in the corresponding volume in fresh media before mixing parental and daughter cell lines at either 1:1 or 9:1 ratios. Each point was seeded in duplicate in a 24-well non-treated plate. Then days 1, 2, 3, 4 ,5 ,7 and 8, an aliquot was collected and replaced by the same volume of fresh media, for GFP flow cytometry measurement.
For the competitive growth assay with Azacitidine treatment, cells

were prepared the same way as before except half of the mix was treated with Azacitidine (final concentration 4µM) prior to dispensing the duplicates. Flow cytometry measurement were done at days 1, 3, 6 and 8. In treated wells cell aliquot volume was replaced by the same volume of treated media.

6.2.6 Flow Cytometry

All flow cytometry's measurements were done on a LSR Fortessa cytometer (BD, Franklin Lakes, New Jersey), either using tubes or 96-well plate.

GFP positive measurement

Cells were directly analyzed on the cytometer. The gating strategy was straight forward, first selecting full cells based on side and forward scatters, then excluding doublets based on the side scatter, before gating on positive cells in B530-A **Figure 6.1A**. The percentage of GFP positive among parental gate was used as read out.

Viability assay

Cell viability was assessed by flow cytometry of cells stained with 1µg/mL DAPI (4',6-Diamidino-2-phenylindole dihydrochloride) (Thermo Scientific), and 100µM MitoTracker® Red CMXRos (Thermo Scientific). 200µL of cells were transferred to a 96-well round bottom plate, Dapi and MitoTracker were premixed with water before adding to the wells. The mix was protected from light in a cardboard box at room temperature during the 30min incubation prior to measurement on the cytometer. The gating strategy is illustrated in **Figure 6.1B**, first small debris are excluded based on the side and forward scatter, then cells are divided in three gates on a V450-A (Dapi) and G610-A (MitoTracker) plot : dead cells (positive to Dapi), live cells (negative to Dapi, positive to MitoTracker), apoptotic cells (negative for both) . The gating was done using FlowJo v9. The percentage of live cells was used for statistical analysis (ANOVA test) using GraphPad Prism software (version 6).

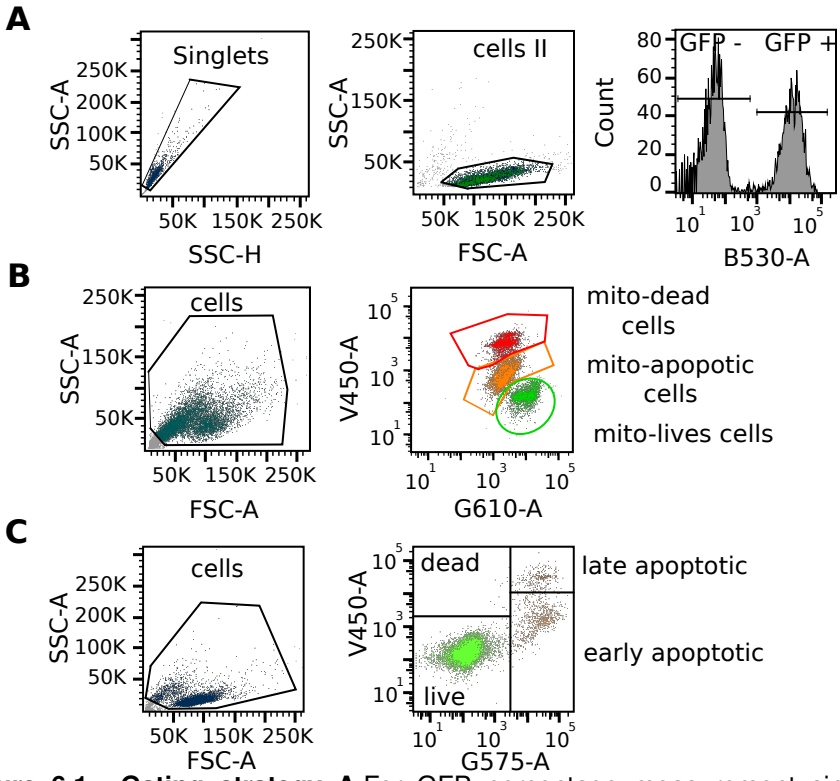


Figure 6.1: Gating strategy A. For GFP percentage measurement singlet were selected first, then by size were selected live cells and finally cells were separated between GFP negative and GFP positive based on signal in the green chanel (B530). **B.**For viability assay analysed by Dapi and MitoTracker® Red CMXRos. Debris were excluded using size scatters then cells were divided between live cells (Dapi-/mito+), apoptotic cells (positive for both) and dead cells (dapi+/mito-). **C.** For apoptosis assay, debris were selected then cells were divided between live cells (double negative), early apoptotic cells (annexinV+/Dapi-), late apoptotic cells (annexinV+/Dapi+) and dead cells (annexinV-/Dapi+).

Apoptosis assay

The percentage of apoptotic cells was determined by flow cytometry of cells stained with AnnexinV-PE conjugated antibody (BD, 1:200 dilution) and 1µg/mL DAPI. 200µL of cells were transferred to a 96-well round bottom plate, Dapi and AnnexinV-PE were premixed with water before adding to the wells. The mix was protected from light in a cardboard box at room temperature during the 15min incubation prior to measurement using plate adaptor on the LSR Fortessa cytometer (BD, Franklin Lakes, New Jersey). The gating strategy is illustrated in **Figure 6.1C**, first small debris are excluded based on the side and forward scatter, then cells are divided in four gates on a V450-A (Dapi) and G575-A (annexinV-PE) plot : dead cells (positive to Dapi only), live cells (negative for both stain), early apoptotic cells (negative to Dapi and positive to AnnexinV), and late apoptotic cells (positive for both) . The gating was done using FlowLogic. Statistical analysis (ANOVA test) was done using GraphPad Prism software (version 6).

6.3 Bioinformatic analysis

6.3.1 External data sources

ChIP-seq data The analysis was done using ChIP-seq data from Douet et al., 2017, original data can be found in GEO under accession number GSE58175. Detail method of analysis can be found in Douet et al., 2017. The original data were cleaned and trimmed, before being aligned on human genome (GRCh37/hg19). The enrichment analysis was done using MACS2 software B. Zhang et al., 2009.

Repeats annotation and sequences The list of repeats position and characterisation was obtained from RepeatMasker Smit et al., 1995.
LINES-1 consensus sequences were obtained from Khan, Smit and Boissinot, 2006.

Blood cells expression data The expression data of *BAZ1A* and *macroH2A1* were obtained from Bagger, Kinalis and Rapin, 2019. Precisely Human Normal Haematopoiesis are cells are from GSE42519, Human AML cells and MDS cells are from: GSE13159 , GSE15434 , GSE61804 , GSE14468 , The Cancer Genome Atlas (TCGA).

6.3.2 Association analysis with ChIP-seq data

Analysis of association between macroH2As and repeats In R, the overlap between regions (peaks or DNA elements) was calculated using regioneR version 1.20.1,**GelRegioneR:Tests**, an R package based on permutation tests performed 10000 iterations using a OverlapPermTest and a circularRandomizeRegions to evaluate the specificity of the association. Results were normalised using z-score, then the matrix was generated. These functions are compiled in a new package to be published soon.

Analysis of local overlaps between macroH2As and LINE-1 The permutation tests were performed with 10000 iterations to evaluate the association between macroH2A1s on the different families of LINEs using numOverlaps and randomizeRegions with a step of 1000bp over a window of10000bp on both side of the start position, before generating the variation of the z-score in the vicinity of the

original region set using localZScore. For visualisation baseR plot was used.

6.3.3 Mapping reads to a LINEs consensus genome

MacroH2A1 ChIP-seq (GSE58175) reads were re-analysed using a pipeline for mapping it on repeat specific areas, this pipeline was based and modified starting on a method published in X. Sun et al., 2018. **Figure 3.5 A** shows the schematics of the pipeline used. Burrows-Wheeler Aligner (BWA) was used for the initial alignment on hg19 (Li and Durbin, 2009). Reads with more than 3 mismatches, with a secondary alignment score over 48 and with soft clipping based were filtered out. All reads overlapping with the BED file containing LINE-1 annotated positions were extracted and re-aligned with BWA MEM on consensus sequences of LINEs-1. Reads were sorted by the left coordinate to recreate a BAM file and its index file using SAMtools. KaryoplotR was used for visualisation (Gel and Serra, 2017).

Material

6.4 Antibodies

Target	Source	WB dilution	IP dilution
BAZ1A	Bethyl laboratories (A301-318A)	1:2000	N/A
histone H3	Abcam(ab1791)	1:5000	2µg/mL
Histone H3K9me3	Abcam(ab8898)	N/A	1µg/mL
IgG	Abcam(ab46540)	N/A	
macroH2A1	home-made (Buschbeck et al., 2009)	1:1000	3µg/mL
macroH2A2	home-made (Buschbeck et al., 2009)	1:1000	5µg/mL
macroH2A1.1	home-made (Sporn et al., 2009)		N/A
macroH2A1.2	home-made (Sporn et al., 2009)		N/A

6.5 Cell lines

name	source
MOLM-13	Leibniz Institute DSMZ-GmbH (ACC 554)
SKK-1	Hans Drexler (DSMZ)
HEK293T	ATCC
HepG2	ATCC (HB-8065)
HL-60	DSMZ (ACC3)

6.6 Oligonucleotides

name	sequence	origin	use
5SrRNA F	TCTACGGCCATACCACCCTGA	(Shen et al., 2013)	ChIP-qPCR
5SrRNA R	GCCTACAGCACCCGGTATTCC	(Shen et al., 2013)	ChIP-qPCR
BAZ1A F	AGTTGTGCTGTGACGGGTAG		RT-qPCR
BAZ1A R	GTAAGCGCGAACGATGGGTAA		RT-qPCR
DXZ4 F	GCCTACGTCACGCAGGAAG		ChIP-qPCR
DXZ4 R	TATGTTTGGGCAGGAAGATCG		ChIP-qPCR
ERV-9 LTR F	TTACAAAGCAGTATTGCTGCCCGC	(Montoya Durango et al., 2009)	ChIP-qPCR
ERV-9 LTR R	AGAGGAAGGAATGCCTCTTGCAGT	(Montoya Durango et al., 2009)	ChIP-qPCR
GAPDH F	cggctactagcggttttacg		RT-qPCR
GAPDH R	gctgcgggctcaatttatag		RT-qPCR
GRXCR1 F	CCTCTGGTTTGGACAATGAATAG		ChIP-qPCR
GRXCR1 R	ATGGGGCAAGAAAATGAAAG		ChIP-qPCR

L1 3'UTR F	AATGAGATCACATGGACACAGGAAG	(Montoya Durango et al., 2009)	ChIP-qPCR
L1 3'UTR R	TGTATACATGTGCCATGCTGGTGC	(Montoya Durango et al., 2009)	ChIP-qPCR
L1 5'UTR exp F	ACGGAATCTCGCTGATTGCTA	(Guo et al., 2014)	RT-qPCR
L1 5'UTR exp R	AAGCAAGCCTGGGCAATG	(Guo et al., 2014)	RT-qPCR
L1 5'UTR F	AGCCTAACTGGGAGGCACCC	(Montoya Durango et al., 2009)	ChIP-qPCR, RT-qPCR
L1 5'UTR R	GATGATGGTGATGTACAGATGGG	(Montoya Durango et al., 2009)	ChIP-qPCR, RT-qPCR
L1HS (ORF1) F (Vazquez et al., 2019)	CGATCAACTGGAAGAAAGGG RT-qPCR	(Vazquez et al., 2019)	RT-qPCR
L1HS (ORF1) R	AGGTACACCAATCAGACGTAG		

L1HS F	ACAGCTTTGAAGAGAGCAGTGGTT	(Liu et al., 2018)	ChIP-qPCR
L1HS R	AGTCTGCCCGTTCTCAGATCT	(Liu et al., 2018)	ChIP-qPCR
L1-ORF1-1 F	TAACCAATACAGAGAAGTGC	(Montoya Durango et al., 2009)	RT-qPCR
L1-ORF1-1 R	GATAATATCCTGCAGAGTGT	(Montoya Durango et al., 2009)	RT-qPCR
L1PB3 F	CCATCTTACCCCTGCAAGAA		ChIP-qPCR
L1PB3 R	ACCCAGTAGTGGGATTGCTG		ChIP-qPCR
LAMA5 F	CAGGCGTGTCTTTAGGATGTAG		ChIP-qPCR
LAMA5 R	TGCCACTGACGTGATGATATG		ChIP-qPCR
LINEs F	TCCTAATGCTATCCCTCCC		ChIP-qPCR
LINEs R	CACCGCATGTTCTCACTCAT		ChIP-qPCR
MACROH2A1 F	GGCAGGAACGGTTTTCCAAAG		RT-qPCR
MACROH2A1 R	CTGTCATTGCTCAGCCTAGTT		RT-qPCR
MACROH2A1.1 R	ttcacccgacaaaacactgac		RT-qPCR
MACROH2A1.1/1.2 F	GAGTTCCAGGACAGCTTCCAC		RT-qPCR
MACROH2A1.2 R	TCCTTGGCCAGAAGCTGAAC		RT-qPCR
NAA15 F	TTGGCACGTTTATGGCCTTCT		RT-qPCR
NAA15 R	CGTTTCCCTGTAACCCTCAAGA		RT-qPCR
RPLP0 F	catcaacgggtacaaacgag		RT-qPCR

RPLP0 R	agatggatcagccaagaagg	(Zeng et al., 2009)	RT-qPCR
SAT2 F	CTGCACTACCTGAAGAGGAC	(Zeng et al., 2009)	ChIP-qPCR
SAT2 R	GATGGTTCAACACTCTTACA	(Zeng et al., 2009)	ChIP-qPCR
SATa F	TCATTCCCACAAACTGCGTTG	(Zeng et al., 2009)	ChIP-qPCR
SATa R	TCCAACGAAGGCCACAAGA	(Zeng et al., 2009)	ChIP-qPCR
ZUB-SEQ-GF	TCCTGCTGGAGTTCGTGACC		sequencing
ZUB-SEQ-PR	ATGCTCCAGACTGCCTT		sequencing

6.7 Plasmids with short hairpin

name	internal ref	gene ID	sh ID	short hairpin sequence
control *renilla*	590			
BAZ1A sh1	664	11177	BAZ1A.1432	AGAGGACAATGTTGCTAATAAA
BAZ1A sh2	665	11177	BAZ1A.5649	GCAGGTATTAAGTCCAAGTTTA
MACROH2A1 sh1	659	9555	H2AFY.455	CCCAGAAGAAGCCTGTATCTAA
MACROH2A1 sh2	703	9555	H2AFY.1097	CCACCTTCAGTTTAAAAGAAAA
NAA15 sh1	X	74838	Naa15.4634	GCAGGTTAGTGATGCAAAGATA

6.8 Plasmides

name	reference	use
psPax2	addgene 12260	packaging plasmid
pCMV-VSV-G	addgene 8454	packaging plasmid
pcSGEP	modified from addgene 111170	shRNA stable integration

6.9 Reagents

Name	Reference	Provider	Stock concentration	storage
Polyethylenimine				
Ligthcycler 480 SYBR Green I Master	ROC 00075	Roche	5X	-20°

RPMI (Roswell Park Memorial Institute) 1640 Medium, no glutamine	31870-025	Gibco	1X	4°
DMEM (Dulbecco's Modified Eagle Medium), high glucose, no glutamine	11960-044	Gibco	1X	4°
FBS	xx	Gibco	1X	-20°
penicillin-streptomycin	15140-122	Gibco	10.000 U/ml	-20°
L-Glutamine (200 mM)	25030-024	Gibco	100X	-20°
Trypsin-EDTA (0.05%), phenol red	25300-054	Gibco	1X	-20°
Azacytidine	A2385-100MG	Sigma-Aldrich	4.2mM	-80° once resu-pended , powder at -20°
5-Aza-2-deoxycytidine (Decitabine)	A3656-5MG	Sigma-Aldrich	10mM	-80° once resu-pended , powder at 4°

Ara-C hydrochloride	C6645-100MG	Sigma-Aldrich	10mM	-20° once resuspended , powder at 4°
DMSO	XX	Sigma-Aldrich	1X	RT

6.10 Kits

Name	Reference	Provider	storage
Maxwell® RSC simplyRNA Cells Kit	AS1390	Promega	RT, except DNase aliquoted at -20°
First Strand cDNA Synthesis kit	K1612	ThermoScientific	-20°
ChIP DNA Clean and Concentrator (Capped Columns)	D5205	Zymo Research	RT
Pierce™ BCA Protein Assay Kit	23227	ThermoFisher	RT (except BSA dilution at -20ºC
RNeasy® Mini kit	74104	Qiagen	RT
QIAquick Gel Extraction Kit	28706X4	Qiagen	RT

6.11 Disposable labware

name	reference	provider	used name
cell culture			
Nunclon Delta cultureware	142475	Thermo Scientfic	treated culture-ware
Nunc™ Non-Treated cultureware	150239	Thermo Scientfic	non-treated culture-ware
Nunc™ EasYFlask™ Cell Culture Flasks	156367	Thermo Scientfic	Flask
microbiology			
VWR® Culture Tubes, Plastic, with Dual-Position Caps, 14mL	60818-667	VWR	bacteria tube
molecular biology			
Microwin 96-well plate			
1.5 ml Bioruptor® Pico Microtubes with Caps	C30010016	Diagenode	sonication 1.5ml tube

Bibliography

Abdel-Wahab, Omar, Mazhar Adli, Lindsay M. LaFave, Jie Gao, Todd Hricik, Alan H. Shih, Suveg Pandey et al. 2012. “ASXL1 Mutations Promote Myeloid Transformation through Loss of PRC2-Mediated Gene Repression”. *Cancer Cell* 22, no. 2 (August): 180–193. ISSN: 15356108.

Abdel-Wahab, Omar, Ann Mullally, Cyrus Hedvat, Guillermo Garcia-Manero, Jay Patel, Martha Wadleigh, Sebastien Malinge et al. 2009. “Genetic characterization of TET1, TET2, and TET3 alterations in myeloid malignancies”. *Blood* 114, no. 1 (July): 144–147. ISSN: 00064971.

Adema, Vera, Laura Palomo, Andrea Toma, Olivier Kosmider, Francisco Fuster-Tormo, Rocío Benito, Rocío Salgado et al. 2020. *Distinct mutational pattern of myelodysplastic syndromes with and without 5q– treated with lenalidomide,* 4, May.

Agelopoulos, Marios, and Dimitris Thanos. 2006. “Epigenetic determination of a cell-specific gene expression program by ATF-2 and the histone variant macroH2A”. *The EMBO Journal* 25, no. 20 (October): 4843–4853. ISSN: 0261-4189.

Aimiuwu, Josephine, Hongyan Wang, Ping Chen, Zhiliang Xie, Jiang Wang, Shujun Liu, Rebecca Klisovic et al. 2012. “RNA-dependent inhibition of ribonucleotide reductase is a major pathway for 5-azacytidine activity in acute myeloid leukemia”. *Blood* 119, no. 22 (June): 5229–5238. ISSN: 00064971.

Allfrey, V. G., R Faulkner and A E Mirsky. 1964. “ACETYLATION AND METHYLATION OF HISTONES AND THEIR POSSIBLE ROLE IN THE REGULATION OF RNA SYNTHESIS”. *Proceedings of the National Academy of Sciences* 51, no. 5 (May): 786–794. ISSN: 0027-8424.

Angelov, Dimitar, Annie Molla, Pierre Yves Perche, Fabienne Hans, Jacques Côté, Saadi Khochbin, Philippe Bouvet and Stefan Dimitrov. 2003. “The histone variant macroH2A interferes with transcription factor binding and SWI/SNF nucleosome remodeling”. *Molecular Cell* 11, no. 4 (April): 1033–1041. ISSN: 10972765.

Arand, Julia, David Spieler, Tommy Karius, Miguel R. Branco, Daniela Meilinger, Alexander Meissner, Thomas Jenuwein et al. 2012. “In Vivo Control of CpG and Non-CpG DNA Methylation by DNA Methyltransferases”. Edited by Dirk Schübeler. *PLoS Genetics* 8, no. 6 (June): e1002750. ISSN: 1553-7404.

Arber, Daniel A, Attilio Orazi, Robert Hasserjian, Jürgen Thiele, Michael J Borowitz, Michelle M Le Beau, Clara D Bloomfield, Mario Cazzola and James W Vardiman. 2016. “The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia”. *Blood* 127, no. 20 (May): 2391–2405. ISSN: 0006-4971.

Arnesen, T., D. Gromyko, F. Pendino, A. Ryningen, J. E. Varhaug and J. R. Lillehaug. 2006. “Induction of apoptosis in human cells by RNAi-mediated knockdown of hARD1 and NATH, components of the protein N-α-acetyltransferase complex”. *Oncogene* 25, no. 31 (July): 4350–4360. ISSN: 09509232.

Audia, James E., and Robert M. Campbell. 2016. “Histone modifications and cancer”. *Cold Spring Harbor Perspectives in Biology* 8 (4): 1–32. ISSN: 19430264.

Awada, Hassan, Bicky Thapa and Valeria Visconte. 2020. “The Genomics of Myelodysplastic Syndromes: Origins of Disease Evolution, Biological Pathways, and Prognostic Implications”. *Cells* 9, no. 11 (November): 2512. ISSN: 2073-4409.

Bagger, Frederik Otzen, Savvas Kinalis and Nicolas Rapin. 2019. “BloodSpot: a database of healthy and malignant haematopoiesis updated with purified and single cell mRNA sequencing profiles”. *Nucleic Acids Research* 47, no. D1 (January): D881–D885. ISSN: 0305-1048.

Ball, Brian J., Christopher A. Famulare, Eytan M. Stein, Martin S. Tallman, Andriy Derkach, Mikhail Roshal, Saar I. Gill et al. 2020. “Venetoclax and hypomethylating agents (HMAs) induce high response rates in MDS, including patients after HMA therapy failure”. *Blood Advances* 4, no. 13 (July): 2866–2870. ISSN: 24739537.

Barrero, María J., Borja Sese, Mercè Martí and Juan Carlos Izpisua Belmonte. 2013. “Macro histone variants are critical for the differentiation of human pluripotent cells”. *Journal of Biological Chemistry* 288, no. 22 (May): 16110–16116. ISSN: 00219258.

Beck, Christine R, José Luis Garcia-Perez, Richard M Badge and John V Moran. 2011. “LINE-1 elements in structural variation and disease.” *Annual review of genomics and human genetics* 12:187–215. ISSN: 1545-293X.

Bejar, Rafael, Ross Levine and Benjamin L. Ebert. 2011. *Unraveling the molecular pathophysiology of myelodysplastic syndromes,* 5, February.

Bejar, Rafael, Allegra Lord, Kristen Stevenson, Michal Bar-Natan, Albert Pérez-Ladaga, Jacques Zaneveld, Hui Wang et al. 2014. “TET2 mutations predict response to hypomethylating agents in myelodysplastic syndrome patients”. *Blood* 124, no. 17 (October): 2705–2712. ISSN: 15280020.

Bejar, Rafael, Kristen E. Stevenson, Bennett A. Caughey, Omar Abdel-Wahab, David P. Steensma, Naomi Galili, Azra Raza et al. 2012. “Validation of a prognostic model and the impact of mutations in patients with lower-risk myelodysplastic syndromes”. *Journal of Clinical Oncology* 30, no. 27 (September): 3376–3382. ISSN: 0732183X.

Bennett, J. M., D. Catovsky, M. T. Daniel, G. Flandrin, D. A. G. Galton, H. R. Gralnick and C. Sultan. 1982. “Proposals for the classification of the myelodysplastic syndromes”. *British Journal of Haematology* 51, no. 2 (June): 189–199. ISSN: 0007-1048.

Berdasco, María, and Manel Esteller. 2019. *Clinical epigenetics: seizing opportunities for translation,* 2, February.

Bereshchenko, Oxana, Oriana Lo Re, Fedor Nikulenkov, Sara Flamini, Jana Kotaskova, Tommaso Mazza, Marguerite Marie Le Pannérer et al. 2019. “Deficiency and haploinsufficiency of histone macroH2A1.1 in mice recapitulate hematopoietic defects of human myelodysplastic syndrome”. *Clinical Epigenetics* 11, no. 1 (August): 121. ISSN: 18687083.

Bersanelli, Matteo, Erica Travaglino, Manja Meggendorfer, Tommaso Matteuzzi, Claudia Sala, Ettore Mosca, Chiara Chiereghin et al. 2021. “Classification and Personalized Prognostic Assessment on the Basis of Clinical and Genomic Features in Myelodysplastic Syndromes”. *Journal of Clinical Oncology* 39, no. 11 (April): 1223–1233. ISSN: 0732-183X.

Black, Ben E., Melissa A. Brock, Sabrina Bédard, Virgil L. Woods and Don W. Cleveland. 2007. “An epigenetic mark generated by the incorporation of CENP-A into centromeric nucleosomes”. *Proceedings of the National Academy of Sciences of the United States of America* 104, no. 12 (March): 5008–5013. ISSN: 00278424.

Blum, William, Rebecca B. Klisovic, Bjoern Hackanson, Zhongfa Liu, Shujun Liu, Hollie Devine, Tamara Vukosavljevic et al. 2007. “Phase I study of decitabine alone or in combination with valproic acid in acute myeloid leukemia”. *Journal of Clinical Oncology* 25, no. 25 (September): 3884–3891. ISSN: 0732183X.

Borghesan, Michela, Caterina Fusilli, Francesca Rappa, Concetta Panebianco, Giovanni Rizzo, Jude A. Oben, Gianluigi Mazzoccoli et al. 2016. “DNA Hypomethylation and histone variant macroH2A1 synergistically attenuate chemotherapy-induced senescence to promote hepatocellular carcinoma progression”. *Cancer Research* 76, no. 3 (February): 594–606. ISSN: 15387445.

Borrow, Julian, Vincent P Stanton, J Michael Andresen, Reinhard Becher, Frederick G Behm3, Raju S K Chaganti4, Curt I Civin et al. 1996. *The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses a putative acetyltransferase to the CREB-binding protein.* Technical report.

Bowen, David T, and Eva Hellstrom-Lindberg. 2001. *Best supportive care for the anaemia of myelodysplasia: inclusion of recombinant erythropoietin therapy?* Technical report.

Brahma, Sandipan, Maheshi I. Udugama, Jongseong Kim, Arjan Hada, Saurabh K. Bhardwaj, Solomon G. Hailu, Tae Hee Lee and Blaine Bartholomew. 2017. "INO80 exchanges H2A.Z for H2A by translocating on DNA proximal to histone dimers". *Nature Communications* 8, no. 1 (June): 1–12. ISSN: 20411723.

Buschbeck, Marcus, and Sandra B Hake. 2017. "Variants of core histones and their roles in development , stem cells and cancer". *Nature Publishing Group* 18 (5): 299–314. ISSN: 1471-0072.

Buschbeck, Marcus, Iris Uribesalgo, Indra Wibowo, Pau Rué, David Martin, Arantxa Gutierrez, Lluís Morey, Roderic Guigó, Hernán López-Schier and Luciano Di Croce. 2009. "The histone variant macroH2A is an epigenetic regulator of key developmental genes". *Nature Structural & Molecular Biology* 16, no. 10 (October): 1074–1079. ISSN: 1545-9993.

Calvert, Andrea E., Alexandra Chalastanis, Yongfei Wu, Lisa A. Hurley, Fotini M. Kouri, Yingtao Bi, Maureen Kachman et al. 2017. "Cancer-Associated IDH1 Promotes Growth and Resistance to Targeted Therapies in the Absence of Mutation". *Cell Reports* 19, no. 9 (May): 1858–1873. ISSN: 22111247.

Campos, Eric I., Jeffrey Fillingham, Guohong Li, Haiyan Zheng, Philipp Voigt, Wei Hung W. Kuo, Harshika Seepany et al. 2010. "The program for processing newly synthesized histones H3.1 and H4". *Nature Structural and Molecular Biology* 17, no. 11 (November): 1343–1351. ISSN: 15459993.

Cao, Ru, Liangjun Wang, Hengbin Wang, Li Xia, Hediye Erdjument-Bromage, Paul Tempst, Richard S. Jones and Yi Zhang. 2002. “Role of histone H3 lysine 27 methylation in polycomb-group silencing”. *Science* 298, no. 5595 (November): 1039–1043. ISSN: 00368075.

Cao, Ru, and Yi Zhang. 2004. *The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3,* 2, April.

Castro-Diaz, Nathaly, Gabriela Ecco, Andrea Coluccio, Adamandia Kapopoulou, Benyamin Yazdanpanah, Marc Friedli, Julien Duc, Suk Min Jang, Priscilla Turelli and Didier Trono. 2014. “Evolutionally dynamic L1 regulation in embryonic stem cells”. *Genes and Development* 28, no. 13 (July): 1397–1409. ISSN: 15495477.

Cazzola, Mario, Matteo G Della Porta and Luca Malcovati. 2008. “Clinical Relevance of Anemia and Transfusion Iron Overload in Myelodysplastic Syndromes”. *Hematology* 2008, no. 1 (January): 166–175. ISSN: 1520-4391.

Chadwick, Brian P. 2008. “DXZ4 chromatin adopts an opposing conformation to that of the surrounding chromosome and acquires a novel inactive X-specific role involving CTCF and antisense transcripts”. *Genome Research* 18, no. 8 (August): 1259–1269. ISSN: 10889051.

Chakravarthy, Srinivas, Sampath Kumar Y. Gundimella, Cecile Caron, Pierre-Yves Perche, John R. Pehrson, Saadi Khochbin and Karolin Luger. 2005. “Structural Characterization of the Histone Variant macroH2A”. *Molecular and Cellular Biology* 25, no. 17 (September): 7616–7624. ISSN: 0270-7306.

Chamseddine, Ali N., Elias Jabbour, Hagop M. Kantarjian, Zachary S. Bohannan and Guillermo Garcia-Manero. 2016. *Unraveling Myelodysplastic Syndromes: Current Knowledge and Future Directions,* 1, January.

Chang, Evelyn Y, Helder Ferreira ——, Joanna Somers ——, Dmitri A Nusinow, Tom Owen-Hughes and Geeta J Narlikar. 2008. “MacroH2A Allows ATP-Dependent Chromatin Remodeling by SWI/SNF and ACF Complexes but Specifically Reduces Recruitment of SWI/SNF †”.

Changolkar, Lakshmi N., Geetika Singh and John R. Pehrson. 2008. “macroH2A1-Dependent Silencing of Endogenous Murine Leukemia Viruses”. *Molecular and Cellular Biology* 28 (6): 2059–2065. ISSN: 0270-7306.

Chen, Hongshan, Penelope D. Ruiz, Wendy M. McKimpson, Leonid Novikov, Richard N. Kitsis and Matthew J. Gamble. 2015. “MacroH2A1 and ATM Play Opposing Roles in Paracrine Senescence and the Senescence-Associated Secretory Phenotype”. *Molecular Cell* 59, no. 5 (September): 719–731. ISSN: 10972765.

Chen, Yimin, Shuyi Chen, Jielun Lu, Danyun Yuan, Lang He, Pengfei Qin, Huo Tan and Lihua Xu. 2021. “MicroRNA-363-3p promote the development of acute myeloid leukemia with RUNX1 mutation by targeting SPRYD4 and FNDC3B”. *Medicine* 100, no. 18 (May): e25807. ISSN: 15365964.

Cluzeau, Thomas, Guillaume Robert, Nicolas Mounier, Jean Michel Karsenti, Maeva Dufies, Alexandre Puissant, Arnaud Jacquel et al. 2012. “BCL2L10 is a predictive factor for resistance to Azacitidine in MDS and AML patients”. *Oncotarget* 3, no. 4 (April): 490–501. ISSN: 1949-2553.

Comiskey, Daniel F., Matías Montes, Safiya Khurshid, Ravi K. Singh and Dawn S. Chandler. 2020. “SRSF2 regulation of MDM2 reveals splicing as a therapeutic vulnerability of the p53 pathway”. *Molecular Cancer Research* 18, no. 2 (February): 194–203. ISSN: 15573125.

Cook, Adam J.L., Zachary A. Gurard-Levin, Isabelle Vassias and Geneviève Almouzni. 2011. “A Specific Function for the Histone Chaperone NASP to Fine-Tune a Reservoir of Soluble H3-H4 in the Histone Supply Chain”. *Molecular Cell* 44, no. 6 (December): 918–927. ISSN: 10972765.

Corujo, David, and Marcus Buschbeck. 2018. “Post-Translational Modifications of H2A Histone Variants and Their Role in Cancer”. *Cancers* 10, no. 3 (February). ISSN: 2072-6694.

Costanzi, Carl, and John R. Pehrson. 2001. “MACROH2A2, a New Member of the MACROH2A Core Histone Family”. *Journal of Biological Chemistry* 276, no. 24 (June): 21776–21784. ISSN: 00219258.

Creppe, Catherine, Peggy Janich, Neus Cantariño, Marc Noguera, Vanesa Valero, Eva Musulén, Julien Douet et al. 2012. “MacroH2A1 Regulates the Balance between Self-Renewal and Differentiation Commitment in Embryonic and Adult Stem Cells”. *Molecular and Cellular Biology* 32, no. 8 (April): 1442–1452. ISSN: 0270-7306.

Creyghton, Menno P., Albert W. Cheng, G. Grant Welstead, Tristan Kooistra, Bryce W. Carey, Eveline J. Steine, Jacob Hanna et al. 2010. “Histone H3K27ac separates active from poised enhancers and predicts developmental state”. *Proceedings of the National Academy of Sciences of the United States of America* 107, no. 50 (December): 21931–21936. ISSN: 00278424.

Dai, Zhiyin, Julie M. Sheridan, Linden J. Gearing, Darcy L. Moore, Shian Su, Sam Wormald, Stephen Wilcox et al. 2014. “edgeR: a versatile tool for the analysis of shRNA-seq and CRISPR-Cas9 genetic screens”. *F1000Research* 3 (October): 95. ISSN: 2046-1402.

Damaraju, Vijaya L., Delores Mowles, Sylvia Yao, Amy Ng, James D. Young, Carol E. Cass and Zeen Tong. 2012. “Role of human nucleoside transporters in the uptake and cytotoxicity of azacitidine and decitabine”. *Nucleosides, Nucleotides and Nucleic Acids* 31, no. 3 (March): 236–255. ISSN: 15257770.

Dardenne, Etienne, Sandra Pierredon, Keltouma Driouch, Lise Gratadou, Magali Lacroix-Triki, Micaela Polay Espinoza, Eleonora Zonta et al. 2012. “Splicing switch of an epigenetic regulator by RNA helicases promotes tumor-cell invasiveness”. *Nature Structural and Molecular Biology* 19, no. 11 (September): 1139–1146. ISSN: 15459985.

De Barrios, Oriol, Balázs Gyorffy, María Jesús Fernández-Aceñero, Ester Sánchez-Tilló, Lidia Sánchez-Moral, Laura Siles, Anna Esteve-Arenys et al. 2017. “ZEB1-induced tumourigenesis requires senescence inhibition via activation of DKK1/mutant p53/Mdm2/CtBP and repression of macroH2A1”. *Gut* 66, no. 4 (April): 666–682. ISSN: 0017-5749.

Di Nardo, C. D., E. Jabbour, F. Ravandi, K. Takahashi, N. Daver, M. Routbort, K. P. Patel et al. 2016. *IDH1 and IDH2 mutations in myelodysplastic syndromes and role in disease progression,* 4, April.

Diesch, Jeannine, Marguerite-Marie Le Pannérer, René Winkler, Raquel Casquero, Matthias Muhar, Mark van der Garde, Michael Maher et al. 2021. “Inhibition of CBP synergizes with the RNA-dependent mechanisms of Azacitidine by limiting protein synthesis”. *Nature Communications.*

Diesch, Jeannine, Anabel Zwick, Anne Kathrin Garz, Anna Palau, Marcus Buschbeck and Katharina S. Götze. 2016. “A clinical-molecular update on azanucleoside-based therapy for the treatment of hematologic cancers”. *Clinical Epigenetics* 8, no. 1 (December): 71. ISSN: 18687083.

Douet, Julien, David Corujo, Roberto Malinverni, Justine Renauld, Viola Sansoni, Melanija Posavec Marjanović, Neus Cantariño et al. 2017. “MacroH2A histone variants maintain nuclear organization and heterochromatin architecture”. *Journal of Cell Science* 130, no. 9 (March): 1570–1582. ISSN: 14779137.

Douet-Guilbert, N., E. De Braekeleer, A. Basinko, A. Herry, N. Gueganic, C. Bovo, K. Trillet et al. 2012. *Molecular characterization of deletions of the long arm of chromosome 5 (del(5q)) in 94 MDS/AML patients,* 7, July.

Doyen, Cécile-Marie, Woojin An, Dimitar Angelov, Vladimir Bondarenko, Flore Mietton, Vassily M. Studitsky, Ali Hamiche, Robert G. Roeder, Philippe Bouvet and Stefan Dimitrov. 2006. “Mechanism of Polymerase II Transcription Repression by the Histone Variant macroH2A”. *Molecular and Cellular Biology* 26, no. 3 (February): 1156–1164. ISSN: 0270-7306.

Ebert, Benjamin L. 2011. “Molecular Dissection of the 5q Deletion in Myelodysplastic Syndrome”. *Seminars in Oncology* 38, no. 5 (October): 621–626. ISSN: 00937754.

Ernst, Jason, and Manolis Kellis. 2010. “Discovery and characterization of chromatin states for systematic annotation of the human genome”. *Nature Biotechnology* 28, no. 8 (August): 817–825. ISSN: 1087-0156.

Fellmann, Christof, Thomas Hoffmann, Vaishali Sridhar, Barbara Hopfgartner, Matthias Muhar, Mareike Roth, Dan Yu Lai et al. 2013. “An optimized microRNA backbone for effective single-copy RNAi”. *Cell Reports* 5, no. 6 (December): 1704–1713. ISSN: 22111247.

Fellmann, Christof, Johannes Zuber, Katherine McJunkin, Kenneth Chang, Colin D. Malone, Ross A. Dickins, Qikai Xu et al. 2011. “Functional Identification of Optimized RNAi Triggers Using a Massively Parallel Sensor Assay”. *Molecular Cell* 41, no. 6 (March): 733–746. ISSN: 10972765.

Figueroa, Maria E., Omar Abdel-Wahab, Chao Lu, Patrick S. Ward, Jay Patel, Alan Shih, Yushan Li et al. 2010. “Leukemic IDH1 and IDH2 Mutations Result in a Hypermethylation Phenotype, Disrupt TET2 Function, and Impair Hematopoietic Differentiation”. *Cancer Cell* 18, no. 6 (December): 553–567. ISSN: 15356108.

Fliedner, T. M., D. Graessle, C. Paulsen and K. Reimers. 2002. *Structure and function of bone marrow hemopoiesis: Mechanisms of response to ionizing radiation exposure,* 4.

Florl, A. R., R. Löwer, B. J. Schmitz-Dräger and W. A. Schulz. 1999. "DNA methylation and expression of LINE-1 and HERV-K provirus sequences in urothelial and renal cell carcinomas". *British Journal of Cancer* 80, no. 9 (July): 1312–1321. ISSN: 0007-0920.

Fried, V, and JR Pehrson. 1992. "MacroH2A, a core histone containing a large nonhistone region". *Science* 257 (5075): 1398–1400. ISSN: 0036-8075.

Gamble, Matthew J., Kristine M. Frizzell, Christine Yang, Raga Krishnakumar and W. Lee Kraus. 2010. "The histone variant macroH2A1 marks repressed autosomal chromatin, but protects a subset of its target genes from silencing". *Genes & Development* 24, no. 1 (January): 21–32. ISSN: 0890-9369.

Gascón, Pere, Andriy Krendyukov, Nicola Mathieson and Matti Aapro. 2019. *Epoetin alfa for the treatment of myelodysplastic syndrome-related anemia: A review of clinical data, clinical guidelines, and treatment protocols,* June.

Gawlitza, Anja L., Johanna Speith, Jenny Rinke, Roman Sajzew, Elena K. Müller, Vivien Schäfer, Andreas Hochhaus and Thomas Ernst. 2019. "5-Azacytidine modulates CpG methylation levels of EZH2 and NOTCH1 in myelodysplastic syndromes". *Journal of Cancer Research and Clinical Oncology* 145, no. 11 (November): 2835–2843. ISSN: 14321335.

Gel, Bernat, Anna Díez-Villanueva, Eduard Serra, Marcus Buschbeck, Miguel A. Peinado and Roberto Malinverni. 2015. "regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests". *Bioinformatics* 32, no. 2 (September): btv562. ISSN: 1367-4803.

Gel, Bernat, and Eduard Serra. 2017. "KaryoploteR: An R/Bioconductor package to plot customizable genomes displaying arbitrary data". *Bioinformatics* 33, no. 19 (October): 3088–3090. ISSN: 14602059.

Germing, Ulrich, Carlo Aul, Charlotte M. Niemeyer, Rainer Haas and John M. Bennett. 2008. *Epidemiology, classification and prognosis of adults and children with myelodysplastic syndromes,* 9, September.

Gerstung, Moritz, Andrea Pellagatti, Luca Malcovati, Aristoteles Giagounidis, Matteo G. Della Porta, Martin Jädersten, Hamid Dolatshad et al. 2015. "Combining gene mutation with gene expression data improves outcome prediction in myelodysplastic syndromes". *Nature Communications* 6, no. 1 (January): 1–11. ISSN: 20411723.

Giagounidis, Aristoteles, Ghulam J. Mufti, Pierre Fenaux, Ulrich Germing, Alan List and Kyle J. MacBeth. 2014. "Lenalidomide as a disease-modifying agent in patients with del(5q) myelodysplastic syndromes: linking mechanism of action to clinical outcomes". *Annals of Hematology* 93, no. 1 (January): 1–11. ISSN: 0939-5555.

Giaimo, Benedetto Daniele, Francesca Ferrante, Andreas Herchenröther, Sandra B. Hake and Tilman Borggrefe. 2019. *The histone variant H2A.Z in gene regulation,* 1, June.

Gnyszka, Agnieszka, Zenon Jastrzebski and Sylwia Flis. 2013. "DNA methyltransferase inhibitors and their emerging role in epigenetic therapy of cancer." *Anticancer research* 33, no. 8 (August): 2989–96. ISSN: 1791-7530.

Greenberg, Peter L., Heinz Tuechler, Julie Schanz, Guillermo Sanz, Guillermo Garcia-Manero, Francesc Solé, John M. Bennett et al. 2012. "Revised international prognostic scoring system for myelodysplastic syndromes". *Blood* 120, no. 12 (September): 2454–2465. ISSN: 00064971.

Gu, Xiaorong, Rita Tohme, Benjamin Tomlinson, Nneha Sakre, Metis Hasipek, Lisa Durkin, Caroline Schuerger et al. 2021. "Decitabine- and 5-azacytidine resistance emerges from adaptive responses of the pyrimidine metabolism network". *Leukemia* 35, no. 4 (April): 1023–1036. ISSN: 0887-6924.

Guberovic, Iva, Sarah Hurtado-Bagès, Ciro Rivera-Casas, Gunnar Knobloch, Roberto Malinverni, Vanesa Valero, Michelle M. Leger et al. 2021. "Evolution of a histone variant involved in compartmental regulation of NAD metabolism". *Nature Structural & Molecular Biology.*

Gujar, Hemant, Daniel J. Weisenberger and Gangning Liang. 2019. "The roles of human DNA methyltransferases and their isoforms in shaping the epigenome". *Genes* 10 (2). ISSN: 20734425.

Guler, Gulfem Dilek, Charles Albert Tindell, Robert Pitti, Catherine Wilson, Katrina Nichols, Tommy KaiWai Cheung, Hyo Jin Kim et al. 2017. "Repression of Stress-Induced LINE-1 Expression Protects Cancer Cell Subpopulations from Lethal Drug Exposure". *Cancer Cell* 32, no. 2 (August): 221–237. ISSN: 18783686.

Guo, Huishan, Maneka Chitiprolu, David Gagnon, Lingrui Meng, Carol Perez-Iratxeta, Diane Lagace and Derrick Gibbings. 2014. "Autophagy supports genomic stability by degrading retrotransposon RNA". *Nature Communications* 5, no. 1 (December): 5276. ISSN: 2041-1723.

Haase, Detlef, Ulrich Germing, Julie Schanz, Michael Pfeilstöcker, Thomas Nösslinger, Barbara Hildebrandt, Andrea Kundgen et al. 2007. "New insights into the prognostic impact of the karyotype in MDS and correlation with subtypes: Evidence from a core dataset of 2124 patients". *Blood* 110, no. 13 (December): 4385–4395. ISSN: 00064971.

Haferlach, T, Y Nagata, V Grossmann, Y Okuno, U Bacher, G Nagae, S Schnittger et al. 2014. "Landscape of genetic lesions in 944 patients with myelodysplastic syndromes". *Leukemia* 28:241–247.

Hanahan, Douglas, and Robert A. Weinberg. 2011. *Hallmarks of cancer: The next generation,* 5.

Hasan, Nesrin, and Nita Ahuja. 2019. "The Emerging Roles of ATP-Dependent Chromatin Remodeling Complexes in Pancreatic Cancer". *Cancers* 11, no. 12 (November): 1859. ISSN: 2072-6694.

Hauer, Michael H., and Susan M. Gasser. 2017. *Chromatin and nucleosome dynamics in DNA damage and repair,* 22, November.

Hebbes, T. R., A. W. Thorne and C. Crane-Robinson. 1988. “A direct link between core histone acetylation and transcriptionally active chromatin.” *The EMBO Journal* 7, no. 5 (May): 1395–1402. ISSN: 02614189.

Hernández-Muñoz, Inmaculada, Anders H. Lund, Petra Van Der Stoop, Erwin Boutsma, Inhua Muijrers, Els Verhoeven, Dmitri A. Nusinow, Barbara Panning, York Marahrens and Maarten Van Lohuizen. 2005. “Stable X chromosome inactivation involves the PRC1 Polycomb complex and requires histone MACROH2A1 and the CULLIN3/SPOP ubiquitin E3 ligase”. *Proceedings of the National Academy of Sciences of the United States of America* 102, no. 21 (May): 7635–7640. ISSN: 00278424.

Hildebrand, Erica M., and Job Dekker. 2020. “Mechanisms and Functions of Chromosome Compartmentalization”. *Trends in Biochemical Sciences* 45, no. 5 (May): 385–396. ISSN: 09680004.

Hirai, Hisamaru, Yukio Kobayashi, Hiroyuki Mano, Koichi Hagiwara, Yoshiro Maru, Mitsuhiro Omine, Hideaki Mizoguchi, Junji Nishida and Fumimaro Takaku. 1987. “A point mutation at codon 13 of the N-ras oncogene in myelodysplastic syndrome”. *Nature* 327 (6121): 430–432. ISSN: 00280836.

Hodge, Dayle Q., Jihong Cui, Matthew J. Gamble and Wenjun Guo. 2018. “Histone Variant MacroH2A1 Plays an Isoform-Specific Role in Suppressing Epithelial-Mesenchymal Transition”. *Scientific Reports* 8, no. 1 (December): 841. ISSN: 2045-2322.

Hollenbach, Paul W., Aaron N. Nguyen, Helen Brady, Michelle Williams, Yuhong Ning, Normand Richard, Leslie Krushel, Sharon L. Aukerman, Carla Heise and Kyle J. MacBeth. 2010. “A comparison of azacitidine and decitabine activities in acute myeloid leukemia cell lines”. *PLoS ONE* 5, no. 2 (February). ISSN: 19326203.

Hongs, Lin, Gary P Schrothp, Harry R Matthew&, Peter Yaus and E Morton Bradburysliii. 1993. *Studies of the DNA binding properties of histone H4 amino terminus. Thermal denaturation studies reveal that acetylation markedly reduces the binding constant of the H4 â€œtailâ€ to DNA.* Technical report 1.

Hsu, Chen Jen, Oliver Meers, Marcus Buschbeck and Florian H. Heidel. 2021. *The role of macroh2a histone variants in cancer,* 12, June.

Hsu, Chen Jen, Tina M. Schnoeder, Patricia Arreba-Tutusaus, Maximilian Lassi, Raffaele Teperino, Stefan Dimitrov, Marcus Buschbeck and Florian H. Heidel. 2019. "Functional Impact of MacroH2A Histone Variants in Acute Myeloid Leukemia". *Blood* 134, no. Supplement_1 (November): 1250–1250. ISSN: 0006-4971.

Jabbour, Elias, Jean-Pierre Issa, Guillermo Garcia-Manero and Hagop Kantarjian. 2008. "Evolution of decitabine development". *Cancer* 112, no. 11 (June): 2341–2351. ISSN: 0008543X.

Jädersten, Martin, Leonie Saft, Alexander Smith, Austin Kulasekararaj, Sabine Pomplun, Gudrun Göhring, Anette Hedlund et al. 2011. "TP53 mutations in low-risk myelodysplastic syndromes with del(5q) predict disease progression". *Journal of Clinical Oncology* 29, no. 15 (May): 1971–1979. ISSN: 0732183X.

Jankowska, Anna M., Hadrian Szpurka, Ramon V. Tiu, Hideki Makishima, Manuel Afable, Jungwon Huh, Christine L. O'Keefe, Rebecca Ganetzky, Michael A. McDevitt and Jaroslaw P. Maciejewski. 2009. "Loss of heterozygosity 4q24 and TET2 mutations associated with myelodysplastic/myeloproliferative neoplasms". *Blood* 113, no. 25 (June): 6403–6410. ISSN: 00064971.

Janotka, Luboš, Lucia Messingerová, Kristína Šimoničová, Helena Kavcová, Katarína Elefantová, Zdena Sulová and Albert Breier. 2021. "Changes in Apoptotic Pathways in MOLM-13 Cell Lines after Induction of Resistance to Hypomethylating Agents". *International Journal of Molecular Sciences* 22, no. 4 (February): 2076. ISSN: 1422-0067.

Jansz, Natasha. 2019. *DNA methylation dynamics at transposable elements in mammals,* 6, December.

Ji, Feng, Chenyu Zhao, Bin Wang, Yan Tang, Zhigang Miao and Yongxiang Wang. 2018. "The role of 5-hydroxymethylcytosine in mitochondria after ischemic stroke". *Journal of Neuroscience Research* 96, no. 10 (October): 1717–1726. ISSN: 10974547.

Kadoch, Cigall, and Gerald R. Crabtree. 2015. *Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics,* 5, June.

Kaminskas, Edvardas, Ann Farrell, Sophia Abraham, Amy Baird, Li Shan Hsieh, Shwu Luan Lee, John K. Leighton et al. 2005. "Approval summary: Azacitidine for treatment of myelodysplastic syndrome subtypes". *Clinical Cancer Research* 11, no. 10 (May): 3604–3608. ISSN: 10780432.

Kantarjian, Hagop, Yasuhiro Oki, Guillermo Garcia-Manero, Xuelin Huang, Susan O'Brien, Jorge Cortes, Stefan Faderl et al. 2007. "Results of a randomized study of 3 schedules of low-dose decitabine in higher-risk myelodysplastic syndrome and chronic myelomonocytic leukemia". *Blood* 109, no. 1 (January): 52–57. ISSN: 00064971.

Kapoor, Avnish, Matthew S. Goldberg, Lara K. Cumberland, Kajan Ratnakumar, Miguel F. Segura, Patrick O. Emanuel, Silvia Menendez et al. 2010. "The histone variant macroH2A suppresses melanoma progression through regulation of CDK8". *Nature* 468, no. 7327 (December): 1105–1109. ISSN: 0028-0836.

Kazazian, Haigh, and John V Moran. 1998. *The impact of L 1 retrotransposons on the human genome.* Technical report.

Khan, Hameed, Arian Smit and Stéphane Boissinot. 2006. "Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates." *Genome research* 16, no. 1 (January): 78–87. ISSN: 1088-9051.

Kim, Jin-Man, Yonghwan Shin, Sunyoung Lee, Mi Yeong Kim, Vasu Punj, Hong-In Shin, Kyunghwan Kim et al. 2018. "MacroH2A1.2 inhibits prostate cancer-induced osteoclastogenesis through cooperation with HP1α and H1.2". *Oncogene* 37 (June). ISSN: 0950-9232.

Kim, Kyuryung, Silvia Park, Hayoung Choi, Hye Joung Kim, Yong Rim Kwon, Daeun Ryu, Myungshin Kim, Tae Min Kim and Yoo Jin Kim. 2020. “Gene expression signatures associated with sensitivity to azacitidine in myelodysplastic syndromes”. *Scientific Reports* 10, no. 1 (December). ISSN: 20452322.

Kim, T., M. S. Tyndel, H. J. Kim, J. S. Ahn, S. H. Choi, H. J. Park, Y. K. Kim et al. 2017. “The clonal origins of leukemic progression of myelodysplasia”. *Leukemia* 31, no. 9 (September): 1928–1935. ISSN: 14765551.

Kokavec, Juraj, Tomas Zikmund, Filipp Savvulidi, Vojtech Kulvait, Winfried Edelmann, Arthur I. Skoultchi and Tomas Stopka. 2017. “The ISWI ATPase Smarca5 (Snf2h) Is Required for Proliferation and Differentiation of Hematopoietic Stem and Progenitor Cells”. *STEM CELLS* 35, no. 6 (June): 1614–1623. ISSN: 10665099.

Komrokji, Rami S., Eric Padron, Benjamin L. Ebert and Alan F. List. 2013. *Deletion 5q MDS: Molecular and therapeutic implications,* 4, December.

Konkel, Miriam K., Jerilyn A. Walker and Mark A. Batzer. 2010. “LINEs and SINEs of primate evolution”. *Evolutionary Anthropology* 19 (6): 236–249. ISSN: 10601538.

Kosmider, Olivier, Véronique Gelsi-Boyer, Meyling Cheok, Sophie Grabar, Véronique Della-Valle, Françoise Picard, Franck Viguié et al. 2009. “TET2 mutation is an independent favorable prognostic factor in myelodysplastic syndromes (MDSs)”. *Blood* 114, no. 15 (October): 3285–3291. ISSN: 00064971.

Kozlowski, Marek, David Corujo, Michael Hothorn, Iva Guberovic, Imke K Mandemaker, Charlotte Blessing, Judith Sporn et al. 2018. “MacroH2A histone variants limit chromatin plasticity through two distinct mechanisms.” *EMBO reports* (September): e44445. ISSN: 1469-3178.

Kuendgen, Andrea, Catharina Müller-Thomas, Michael Lauseker, Torsten Haferlach, Petra Urbaniak, Thomas Schroeder, Carolin Brings et al. 2018. *Efficacy of azacitidine is independent of molecular and clinical characteristics-an analysis of 128 patients with myelodysplastic syndromes or acute myeloid leukemia and a review of the literature,* 45, June.

Kulasekararaj, Austin G., Alexander E. Smith, Syed A. Mian, Azim M. Mohamedali, Pramila Krishnamurthy, Nicholas C. Lea, Joop Gäken et al. 2013. "¡i¿TP53¡/i¿ mutations in myelodysplastic syndrome are strongly correlated with aberrations of chromosome 5, and correlate with adverse prognosis". *British Journal of Haematology* 160, no. 5 (March): 660–672. ISSN: 00071048.

Kunimoto, Hiroyoshi, and Hideaki Nakajima. 2021. "TET2: A cornerstone in normal and malignant hematopoiesis". *Cancer Science* 112, no. 1 (January): 31–40. ISSN: 1347-9032.

Kustatscher, Georg, Michael Hothorn, Céline Pugieux, Klaus Scheffzek and Andreas G Ladurner. 2005. "Splicing regulates NAD metabolite binding to histone macroH2A".

Kweon, Soo Mi, Yibu Chen, Eugene Moon, Kotryna Kvederaviciutė, Saulius Klimasauskas and Douglas E. Feldman. 2019. "An Adversarial DNA N6-Methyladenine-Sensor Network Preserves Polycomb Silencing". *Molecular Cell* 74, no. 6 (June): 1138–1147. ISSN: 10974164.

Lachner, M., D. O'Carroll, S. Rea, K. Mechtler and T. Jenuwein. 2001. "Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins". *Nature* 410, no. 6824 (March): 116–120. ISSN: 00280836.

Lacoste, Nicolas, Adam Woolfe, Hiroaki Tachiwana, Ana Villar Garea, Teresa Barth, Sylvain Cantaloube, Hitoshi Kurumizaka, Axel Imhof and Geneviève Almouzni. 2014. "Mislocalization of the Centromeric Histone Variant CenH3/CENP-A in Human Cells Depends on the Chaperone DAXX". *Molecular Cell* 53, no. 4 (February): 631–644. ISSN: 10972765.

Lavigne, Matthieu D, Giannis Vatsellas, Alexander Polyzos, Evangelia Mantouvalou, George Sianidis, Ioannis Maraziotis, Marios Agelopoulos and Dimitris Thanos. 2015. “Composite macroH2A/NRF-1 Nucleosomes Suppress Noise and Generate Robustness in Gene Expression.” *Cell reports* 11, no. 7 (May): 1090–101. ISSN: 2211-1247.

Lee, Chia Hsien, Wen Hong Kuo, Chen Ching Lin, Yen Jen Oyang, Hsuan Cheng Huang and Hsueh Fen Juan. 2013. “MicroRNA-regulated protein-protein interaction networks and their functions in breast cancer”. *International Journal of Molecular Sciences* 14 (6): 11560–11606. ISSN: 14220067.

Leeb, Martin, Diego Pasini, Maria Novatchkova, Markus Jaritz, Kristian Helin and Anton Wutz. 2010. “Polycomb complexes act redundantly to repress genomic repeats and genes”. *Genes and Development* 24, no. 3 (February): 265–276. ISSN: 08909369.

Lei, Shaorong, Jianhong Long and Jiaguang Li. 2014. “MacroH2A suppresses the proliferation of the B16 melanoma cell line”. *Molecular Medicine Reports* 10, no. 4 (October): 1845–1850. ISSN: 17913004.

Lewis, Peter W., Simon J. Elsaesser, Kyung Min Noh, Sonja C. Stadler and C. David Allis. 2010. “Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres”. *Proceedings of the National Academy of Sciences of the United States of America* 107, no. 32 (August): 14075–14080. ISSN: 00278424.

Ley, Timothy J., Li Ding, Matthew J. Walter, Michael D. McLellan, Tamara Lamprecht, David E. Larson, Cyriac Kandoth et al. 2010. “¡i¿DNMT3A¡/i¿ Mutations in Acute Myeloid Leukemia”. *New England Journal of Medicine* 363, no. 25 (December): 2424–2433. ISSN: 0028-4793.

Li, Heng, and Richard Durbin. 2009. “Fast and accurate short read alignment with Burrows-Wheeler transform”. *Bioinformatics* 25, no. 14 (July): 1754–1760. ISSN: 13674803.

Liggett, L. Alexander, and Vijay G. Sankaran. 2020. *Unraveling Hematopoiesis through the Lens of Genomics,* 6, September.

Lindsley, R. Coleman, Brenton G. Mar, Emanuele Mazzola, Peter V. Grauman, Sarah Shareef, Steven L. Allen, Arnaud Pigneux et al. 2015. "Acute myeloid leukemia ontogeny is defined by distinct somatic mutations". *Blood* 125 (9): 1367–1376. ISSN: 15280020.

Liu, Nian, Cameron H Lee, Tomek Swigut, Edward Grow, Bo Gu, Michael C Bassik and Joanna Wysocka. 2018. "Selective silencing of euchromatic L1s revealed by genome-wide screens for L1 regulators". *Nature* 553, no. 7687 (January): 228–232. ISSN: 0028-0836.

Lo Re, Oriana, Julien Douet, Marcus Buschbeck, Caterina Fusilli, Valerio Pazienza, Concetta Panebianco, Carlo Castruccio Castracani, Tommaso Mazza, Giovanni Li Volti and Manlio Vinciguerra. 2018. "Histone variant macroH2A1 rewires carbohydrate and lipid metabolism of hepatocellular carcinoma cells towards cancer stem cells", ISSN: 1559-2308.

Long, Hannah K., Sara L. Prescott and Joanna Wysocka. 2016. "Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution". *Cell* 167, no. 5 (November): 1170–1187. ISSN: 0092-8674.

Lu, J. Yuyang, Lei Chang, Tong Li, Ting Wang, Yafei Yin, Ge Zhan, Xue Han et al. 2021. "Homotypic clustering of L1 and B1/Alu repeats compartmentalizes the 3D genome". *Cell Research* 31, no. 6 (June): 613–630. ISSN: 17487838.

Lübbert, Michael, Stefan Suciu, Liliana Baila, Björn Hans Rüter, Uwe Platzbecker, Aristoteles Giagounidis, Dominik Selleslag et al. 2011. "Low-dose decitabine versus best supportive care in elderly patients with intermediate- or high-risk myelodysplastic syndrome (MDS) ineligible for intensive chemotherapy: Final results of the randomized phase III study of the european organisation for research and treatment of cancer leukemia group and the German MDS study group". *Journal of Clinical Oncology* 29, no. 15 (May): 1987–1996. ISSN: 0732183X.

Luger, Karolin, Armin W. Mäder, Robin K. Richmond, David F. Sargent and Timothy J. Richmond. 1997. “Crystal structure of the nucleosome core particle at 2.8 Å resolution”. *Nature* 389 (6648): 251–260. ISSN: 00280836.

Lund, K., P. D. Adams and M. Copland. 2014. *EZH2 in normal and malignant hematopoiesis,* 1.

Maher, Michael, Jeannine Diesch, Marguerite-Marie Le Pannérer and Marcus Buschbeck. 2021. “Epigenetics in a Spectrum of Myeloid Diseases and Its Exploitation for Therapy”. *Cancers* 13, no. 7 (April): 1746. ISSN: 2072-6694.

Mahfouz, Reda Z., Ania Jankowska, Quteba Ebrahem, Xiaorong Gu, Valeria Visconte, Ali Tabarroki, Pramod Terse et al. 2013. “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, no. 4 (February): 938–948. ISSN: 10780432.

Malcovati, Luca, Ulrich Germing, Andrea Kuendgen, Matteo G. Della Porta, Cristiana Pascutto, Rosangela Invernizzi, Aristoteles Giagounidis et al. 2007. “Time-dependent prognostic scoring system for predicting survival and leukemic evolution in myelodysplastic syndromes”. *Journal of Clinical Oncology* 25, no. 23 (August): 3503–3510. ISSN: 0732183X.

Marfella, Concetta G A, and Anthony N Imbalzano. 2007. “The Chd family of chromatin remodelers”. *Mutation Research* 618:30–40.

Masaki, So, Shun Ikeda, Asuka Hata, Yusuke Shiozawa, Ayana Kon, Seishi Ogawa, Kenji Suzuki, Fumihiko Hakuno, Shin-Ichiro Takahashi and Naoyuki Kataoka. 2019. “Myelodysplastic Syndrome-Associated SRSF2 Mutations Cause Splicing Changes by Altering Binding Motif Sequences”. *Frontiers in Genetics* 10, no. MAR (April). ISSN: 1664-8021.

McClintock, B. 1951. “Chromosome organization and genic expression.” *Cold Spring Harbor symposia on quantitative biology* 16:13–47. ISSN: 00917451.

McClintock, Barbara. 1956. “Controlling elements and the gene.” *Cold Spring Harbor symposia on quantitative biology* 21:197–216. ISSN: 00917451.

Medeiros, B. C., A. T. Fathi, C. D. DiNardo, D. A. Pollyea, S. M. Chan and R. Swords. 2017. *Isocitrate dehydrogenase mutations in myeloid malignancies,* 2, February.

Misteli, Tom. 2020. *The Self-Organizing Genome: Principles of Genome Architecture and Function,* 1, October.

Montoya Durango, Diego E, Yongqing Liu, Ivo Teneng, Ted Kalbfleisch, Mary E Lacy, Marlene C Steffen and Kenneth S Ramos. 2009. “Epigenetic control of mammalian LINE-1 retrotransposon by retinoblastoma proteins”. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 665, nos. 1-2 (June): 20–28. ISSN: 00275107.

Moreland, John L., Apostol Gramada, Oleksandr V. Buzko, Qing Zhang and Philip E. Bourne. 2005. “The Molecular Biology Toolkit (MBT): A modular platform fro developing molecular visualization applications”. *BMC Bioinformatics* 6, no. 1 (February): 21. ISSN: 14712105.

Müller-Kuller, Uta, Mania Ackermann, Stephan Kolodziej, Christian Brendel, Jessica Fritsch, Nico Lachmann, Hana Kunkel et al. 2015. “A minimal ubiquitous chromatin opening element (UCOE) effectively prevents silencing of juxtaposed heterologous promoters by epigenetic remodeling in multipotent and pluripotent stem cells”. *Nucleic Acids Research* 43, no. 3 (February): 1577–1592. ISSN: 1362-4962.

Nakatani, Y., D. Ray-Gallet, J. P. Quivy, H. Tagami and G. Almouzni. 2004. “Two distinct nucleosome assembly pathways: Dependent or independent of DNA synthesis promoted by histone H3.1 and H3.3 complexes”. In *Cold Spring Harbor Symposia on Quantitative Biology,* 69:273–280. Cold Spring Harb Symp Quant Biol.

Ni, Kai, and Kathrin Muegge. 2021. “LSH catalyzes ATP-driven exchange of histone variants macroH2A1 and macroH2A2”. *Nucleic Acids Research,* no. 1 (July). ISSN: 0305-1048.

Nikoloski, Gorica, Saskia M.C. Langemeijer, Roland P. Kuiper, Ruth Knops, Marion Massop, Evelyn R.L.T.M. Tönnissen, Adrian Van Der Heijden et al. 2010. “Somatic mutations of the histone methyltransferase gene EZH2 in myelodysplastic syndromes”. *Nature Genetics* 42, no. 8 (August): 665–667. ISSN: 10614036.

Okada, S, H Nakauchi, K Nagayoshi, S Nishikawa, Y Miura and T Suda. 1992. “In vivo and in vitro stem cell function of c-kit- and Sca-1-positive murine hematopoietic cells”. *Blood* 80, no. 12 (December): 3044–3050. ISSN: 0006-4971.

Orkin, Stuart H., and Leonard I. Zon. 2008. *Hematopoiesis: An Evolving Paradigm for Stem Cell Biology,* 4, February.

Palau, Anna, Mar Mallo, Laura Palomo, Ines Rodríguez-Hernández, Jeannine Diesch, Diana Campos, Isabel Granada et al. 2017. “Immunophenotypic, cytogenetic, and mutational characterization of cell lines derived from myelodysplastic syndrome patients after progression to acute myeloid leukemia”. *Genes, Chromosomes and Cancer* 56, no. 3 (March): 243–252. ISSN: 10452257.

Papaemmanuil, Elli, Moritz Gerstung, Luca Malcovati, Sudhir Tauro, Gunes Gundem, Peter Van Loo, Chris J. Yoon et al. 2013. “Clinical and biological implications of driver mutations in myelodysplastic syndromes”. *Blood* 122, no. 22 (November): 3616–3627. ISSN: 00064971.

Park, Suk Youl, and Jeong Sun Kim. 2020. *A short guide to histone deacetylases including recent progress on class II enzymes,* 2, February.

Pasque, Vincent, Astrid Gillich, Nigel Garrett and John B Gurdon. 2011. “Histone variant macroH2A confers resistance to nuclear reprogramming”. *The EMBO Journal* 30, no. 12 (June): 2373–2387. ISSN: 02614189.

Pellagatti, Andrea, Richard N. Armstrong, Violetta Steeples, Eshita Sharma, Emmanouela Repapi, Shalini Singh, Andrea Sanchi et al. 2018. “Impact of spliceosome mutations on RNA splicing in myelodysplasia: Dysregulated genes/pathways and clinical associations”. *Blood* 132, no. 12 (September): 1225–1240. ISSN: 15280020.

Pliatska, Maria, Maria Kapasa, Antonis Kokkalis, Alexander Polyzos and Dimitris Thanos. 2018. “The Histone Variant MacroH2A Blocks Cellular Reprogramming by Inhibiting Mesenchymal-to-Epithelial Transition.” *Molecular and cellular biology* 38, no. 10 (May): 00669–17. ISSN: 1098-5549.

Poot, Raymond A., Graham Dellaire, Bastian B. Hülsmann, Margaret A. Grimaldi, Davide F.V. Corona, Peter B. Becker, Wendy A. Bickmore and Patrick D. Varga-Weisz. 2000. “HuCHRAC, a human ISWI chromatin remodelling complex contains hACF1 and two novel histone-fold proteins”. *EMBO Journal* 19, no. 13 (July): 3377–3387. ISSN: 02614189.

Posavec Marjanović, Melanija, Sarah Hurtado-Bagès, Maximilian Lassi, Vanesa Valero, Roberto Malinverni, Hélène Delage, Miriam Navarro et al. 2017. “MacroH2A1.1 regulates mitochondrial respiration by limiting nuclear NAD(+) consumption.” *Nature structural & molecular biology* 24, no. 11 (November): 902–910. ISSN: 1545-9985.

Prats-Martín, Concepción, Sergio Burillo-Sanz, Rosario M. Morales-Camacho, Olga Pérez-López, Milagros Suito, Maria T. Vargas, Teresa Caballero-Velázquez et al. 2020. “¡i¿ASXL1¡/i¿ mutation as a surrogate marker in acute myeloid leukemia with myelodysplasia-related changes and normal karyotype”. *Cancer Medicine* 9, no. 11 (June): 3637–3646. ISSN: 2045-7634.

Qin, Taichun, Ryan Castoro, Samih El Ahdab, Jaroslav Jelinek, Xiaodan Wang, Jiali Si, Jingmin Shu et al. 2011. “Mechanisms of Resistance to Decitabine in the Myelodysplastic Syndrome”. Edited by S. K. Batra. *PLoS ONE* 6, no. 8 (August): e23372. ISSN: 1932-6203.

Racki, Lisa R., Janet G. Yang, Nariman Naber, Peretz D. Partensky, Ashley Acevedo, Thomas J. Purcell, Roger Cooke, Yifan Cheng and Geeta J. Narlikar. 2009. “The chromatin remodeller ACF acts as a dimeric motor to space nucleosomes”. *Nature* 462, no. 7276 (December): 1016–1021. ISSN: 00280836.

Rada-Iglesias, Alvaro, Ruchi Bajpai, Tomek Swigut, Samantha A. Brugmann, Ryan A. Flynn and Joanna Wysocka. 2011. “A unique chromatin signature uncovers early developmental enhancers in humans”. *Nature* 470, no. 7333 (February): 279–285. ISSN: 00280836.

Rangasamy, Danny. 2013. “Distinctive patterns of epigenetic marks are associated with promoter regions of mouse LINE-1 and LTR retrotransposons”. *Mobile DNA* 4, no. 1 (December): 27. ISSN: 1759-8753.

Robbez-Masson, Luisa, Christopher H.C. Tie, Lucia Conde, Hale Tunbak, Connor Husovsky, Iva A. Tchasovnikarova, Richard T. Timms, Javier Herrero, Paul J. Lehner and Helen M. Rowe. 2018. “The hush complex cooperates with trim28 to repress young retrotransposons and new genes”. *Genome Research* 28 (6): 836–845. ISSN: 15495469.

Robinson, M. D., D. J. McCarthy and G. K. Smyth. 2010. “edgeR: a Bioconductor package for differential expression analysis of digital gene expression data”. *Bioinformatics* 26, no. 1 (January): 139–140. ISSN: 1367-4803.

Rocquain, Julien, Nadine Carbuccia, Virginie Trouplin, Stéphane Raynaud, Anne Murati, Meyer Nezri, Zoulika Tadrist et al. 2010. “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, no. 1 (December): 401. ISSN: 1471-2407.

Rogakou, Emmy P., Duane R. Pilch, Ann H. Orr, Vessela S. Ivanova and William M. Bonner. 1998. “DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139”. *Journal of Biological Chemistry* 273, no. 10 (March): 5858–5868. ISSN: 00219258.

Rose, Alexander S., and Peter W. Hildebrand. 2015. “NGL Viewer: A web application for molecular visualization”. *Nucleic Acids Research* 43 (W1): W576–W579. ISSN: 13624962.

Runge, John S., Jesse R. Raab and Terry Magnuson. 2018. “Identification of two distinct classes of the human INO80 complex genome-wide”. *G3: Genes, Genomes, Genetics* 8, no. 4 (April): 1095–1102. ISSN: 21601836.

Saber, Wael, and Mary M. Horowitz. 2016. “Transplantation for myelodysplastic syndromes: Who, when, and which conditioning regimens”. *Hematology (United States)* 2016 (1): 478–484. ISSN: 15204383.

Saksouk, Nehmé, Teresa K Barth, Celine Ziegler-Birling, Nelly Olova, Agnieszka Nowak, Elodie Rey, Julio Mateos-Langerak et al. 2014. “Article Redundant Mechanisms to Form Silent Chromatin at Pericentromeric Regions Rely on BEND3 and DNA Methylation”. *Molecular Cell* 56:580–594.

Satake, Noriko, Yasushi Ishida, Yoshiko Otoh, Shin Ichi Hinohara, Hirofumi Kobayashi, Akiko Sakashita, Nobuo Maseki and Yasuhiko Kaneko. 1997. “Novel MLL-CBP fusion transcript in therapy-related chronic myelomonocytic leukemia with a t(11;16) (q23;p13) chromosome translocation”. *Genes Chromosomes and Cancer* 20, no. 1 (September): 60–63. ISSN: 10452257.

Schanz, Julie, Christian Steidl, Christa Fonatsch, Michael Pfeilstöcker, Thomas Nösslinger, Heinz Tuechler, Peter Valent et al. 2011. “Coalesced multicentric analysis of 2,351 patients with myelodysplastic syndromes indicates an underestimation of poor-risk cytogenetics of myelodysplastic syndromes in the International Prognostic Scoring System”. *Journal of Clinical Oncology* 29, no. 15 (May): 1963–1970. ISSN: 0732183X.

Schieber, Michael, Christian Marinaccio, Lyndsey C. Bolanos, Wendy D. Haffey, Kenneth D. Greis, Daniel T. Starczynowski and John D. Crispino. 2020. “FBXO11 is a candidate tumor suppressor in the leukemic transformation of myelodysplastic syndrome”. *Blood Cancer Journal* 10, no. 10 (October): 98. ISSN: 2044-5385.

Shaban, Haitham A., Roman Barth, Ludmila Recoules and Kerstin Bystricky. 2020. “Hi-D: Nanoscale mapping of nuclear dynamics in single living cells”. *Genome Biology* 21, no. 1 (April): 95. ISSN: 1474760X.

Shen, Meili, Tingting Zhou, Wenbing Xie, Te Ling, Qiaoyun Zhu, Le Zong, Guoliang Lyu, Qianqian Gao, Feixiong Zhang and Wei Tao. 2013. “The Chromatin Remodeling Factor CSB Recruits Histone Acetyltransferase PCAF to rRNA Gene Promoters in Active State for Transcription Initiation”. Edited by Mary Bryk. *PLoS ONE* 8, no. 5 (May): e62668. ISSN: 1932-6203.

Shukron, Ofir, Vladimir Vainstein, Andrea Kündgen, Ulrich Germing and Zvia Agur. 2012. “Analyzing transformation of myelodysplastic syndrome to secondary acute myeloid leukemia using a large patient database”. *American Journal of Hematology* 87, no. 9 (September): 853–860. ISSN: 03618609.

Siamishi, Iliana, Norimasa Iwanami, Thomas Clapes, Eirini Trompouki, Connor P. O’Meara and Thomas Boehm. 2020. “Lymphocyte-Specific Function of the DNA Polymerase Epsilon Subunit Pole3 Revealed by Neomorphic Alleles”. *Cell Reports* 31, no. 11 (June): 107756. ISSN: 22111247.

Sims, Robert J., Scott Millhouse, Chi Fu Chen, Brian A. Lewis, Hediye Erdjument-Bromage, Paul Tempst, James L. Manley and Danny Reinberg. 2007. “Recognition of Trimethylated Histone H3 Lysine 4 Facilitates the Recruitment of Transcription Postinitiation Factors and Pre-mRNA Splicing”. *Molecular Cell* 28, no. 4 (November): 665–676. ISSN: 10972765.

Sloan, Cricket A., Esther T. Chan, Jean M. Davidson, Venkat S. Malladi, J. Seth Strattan, Benjamin C. Hitz, Idan Gabdank et al. 2016. “ENCODE data at the ENCODE portal”. *Nucleic Acids Research* 44, no. D1 (January): D726–D732. ISSN: 13624962.

Smit, AFA, R Hubley and P. Green. 0. *RepeatMasker Open-4.0.*

Smit, Arian F.A., Gábor Tóth, Arthur D. Riggs and Jerzy Jurka. 1995. “Ancestral, Mammalian-wide Subfamilies of LINE-1 Repetitive Sequences”. *Journal of Molecular Biology* 246, no. 3 (February): 401–417. ISSN: 0022-2836.

Soriano, Andres O., Hui Yang, Stefan Faderl, Zeev Estrov, Francis Giles, Farhad Ravandi, Jorge Cortes et al. 2007. “Safety and clinical activity of the combination of 5-azacytidine, valproic acid, and all-trans retinoic acid in acute myeloid leukemia and myelodysplastic syndrome”. *Blood* 110, no. 7 (October): 2302–2308. ISSN: 00064971.

Souers, Andrew J, Joel D Leverson, Erwin R Boghaert, Scott L Ackler, Nathaniel D Catron, Jun Chen, Brian D Dayton et al. 2013. “ABT-199, a potent and selective BCL-2 inhibitor, achieves antitumor activity while sparing platelets.” *Nature medicine* 19, no. 2 (February): 202–8. ISSN: 1546-170X.

Sperling, Adam S., Christopher J. Gibson and Benjamin L. Ebert. 2017. *The genetics of myelodysplastic syndrome: From clonal haematopoiesis to secondary leukaemia,* 1, January.

Sporn, J. C., G. Kustatscher, T. Hothorn, M. Collado, M. Serrano, T. Muley, P. Schnabel and A. G. Ladurner. 2009. “Histone macroH2A isoforms predict the risk of lung cancer recurrence”. *Oncogene* 28, no. 38 (September): 3423–3428. ISSN: 09509232.

Sripayap, Piyanuch, Tadashi Nagai, Mitsuyo Uesawa, Hiroyuki Kobayashi, Tomonori Tsukahara, Ken Ohmine, Kazuo Muroi and Keiya Ozawa. 2014. “Mechanisms of resistance to azacitidine in human leukemia cell lines”. *Experimental Hematology* 42 (4). ISSN: 18732399.

Strahl, Brian D., and C. David Allis. 2000. *The language of covalent histone modifications,* 6765, January.

Sun, Xiaoji, Xuya Wang, Zuojian Tang, Mark Grivainis, David Kahler, Chi Yun, Paolo Mita, David Fenyö and Jef D Boeke. 2018. “Transcription factor profiling reveals molecular choreography and key regulators of human retrotransposon expression”. *Proceedings of the National Academy of Sciences* 115, no. 24 (June): E5526–E5535. ISSN: 0027-8424.

Sun, Zhen, Dan Filipescu, Joshua Andrade, Alexandre Gaspar-Maia, Beatrix Ueberheide and Emily Bernstein. 2018. “Transcription-associated histone pruning demarcates macroH2A chromatin domains”. *Nature Structural & Molecular Biology* 25, no. 10 (October): 958–970. ISSN: 1545-9993.

Taddei, A., C. Maison, D. Roche and G. Almouzni. 2001. “Reversible disruption of pericentric heterochromatin and centromere function by inhibiting deacetylases”. *Nature Cell Biology* 3, no. 2 (January): 114–120. ISSN: 14657392.

Tang, Zuojian, Jared P Steranka, Sisi Ma, Mark Grivainis, Nemanja Rodić, Cheng Ran Lisa Huang, Ie-Ming Shih et al. 2017. “Human transposon insertion profiling: Analysis, visualization and identification of somatic LINE-1 insertions in ovarian cancer.” *Proceedings of the National Academy of Sciences of the United States of America* 114, no. 5 (January): E733–E740. ISSN: 1091-6490.

Thol, Felicitas, Frederik Damm, Andrea Lüdeking, Claudia Winschel, Katharina Wagner, Michael Morgan, Haiyang Yun et al. 2011. “Incidence and prognostic influence of DNMT3A mutations in acute myeloid leukemia”. *Journal of Clinical Oncology* 29, no. 21 (July): 2889–2896. ISSN: 0732183X.

Timinszky, Gyula, and Andreas G Ladurner. 2014. “PARP1 and CBP lose their footing in cancer”. *Nature Structural & Molecular Biology* 21, no. 11 (November): 947–948. ISSN: 1545-9993.

Ueda, Kumiko, Mika Hosokawa and Seigo Iwakawa. 2015. “Cellular uptake of decitabine by equilibrative nucleoside transporters in HCT116 cells”. *Biological and Pharmaceutical Bulletin* 38, no. 8 (August): 1113–1119. ISSN: 13475215.

Valencia, A., E. Masala, A. Rossi, A. Martino, A. Sanna, F. Buchi, F. Canzian et al. 2014. “Expression of nucleoside-metabolizing enzymes in myelodysplastic syndromes and modulation of response to azacitidine”. *Leukemia* 28 (3): 621–628. ISSN: 14765551.

Van Zant, Gary, and Ying Liang. 2012. “Concise Review: Hematopoietic Stem Cell Aging, Life Span, and Transplantation”. *STEM CELLS Translational Medicine* 1, no. 9 (September): 651–657. ISSN: 2157-6564.

Vardiman, James W., Jüergen Thiele, Daniel A. Arber, Richard D. Brunning, Michael J. Borowitz, Anna Porwit, Nancy Lee Harris et al. 2009. *The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: Rationale and important changes,* 5, July.

Vazquez, Berta N, Joshua K Thackray, Nicolas G Simonet, Sanjay Chahar, Noriko Kane-Goldsmith, Simon J Newkirk, Suman Lee et al. 2019. “SIRT7 mediates L1 elements transcriptional repression and their association with the nuclear lamina”. *Nucleic Acids Research* 47, no. 15 (September): 7870–7885. ISSN: 0305-1048.

Veselý, J., A. Čihák and F. Šorm. 1968. “Biochemical mechanism of drug resistance-VII. Inhibition of orotic acid metabolism by 5-azacytidine in leukemic mice sensitive and resistant to 5-azacytidine”. *Biochemical Pharmacology* 17 (4): 519–524. ISSN: 00062952.

Vieira-Silva, Tânia Soraia, Sara Monteiro-Reis, Daniela Barros-Silva, João Ramalho-Carvalho, Inês Graça, Isa Carneiro, Ana Teresa Martins et al. 2019. “Histone variant MacroH2A1 is downregulated in prostate cancer and influences malignant cell phenotype”. *Cancer Cell International* 19, no. 1 (December): 112. ISSN: 1475-2867.

Vos, Dick de, and Wendy van Overveld. 2005. *Decitabine: A historical review of the development of an epigenetic drug,* 13, December.

Waddington, C H. 1942. “The epigenotype”. *Endeavour* 1:18–20.

———. 1957. *he Strategy of the Genes: A Discussion of Some Aspects of Theoretical Biology.* Edited by London: Allen & Unwin.

Wagstaff, Bradley J., Miriam Barnerβoi and Astrid M. Roy-Engel. 2011. "Evolutionary Conservation of the Functional Modularity of Primate and Murine LINE-1 Elements". Edited by Gil Ast. *PLoS ONE* 6, no. 5 (May): e19672. ISSN: 1932-6203.

Walsh, C. P., J. R. Chaillet and T. H. Bestor. 1998. *Transcription of IAP endogenous retroviruses is constrained by cytosine methylation [4],* 2.

Walter, M. J., L. Ding, D. Shen, J. Shao, M. Grillot, M. McLellan, R. Fulton et al. 2011. "Recurrent DNMT3A mutations in patients with myelodysplastic syndromes". *Leukemia* 25, no. 7 (July): 1153–1158. ISSN: 08876924.

Wang, Diane C, William Wang, Linlin Zhang and Xiangdong Wang. 2018. "A tour of 3D genome with a focus on CTCF". *Cell & Developmental Biology.*

Wang, Xiaodong, Jian Lu, Jisen Cao, Bozhao Ma, Chao Gao and Feng Qi. 2018. "MicroRNA-18a promotes hepatocellular carcinoma proliferation, migration, and invasion by targeting Bcl2L10". *OncoTargets and Therapy* 11:7919–7934. ISSN: 11786930.

Wang, Zi, Pan Wang, Yanan Li, Hongling Peng, Yu Zhu, Narla Mohandas and Jing Liu. 2021. *Interplay between cofactors and transcription factors in hematopoiesis and hematological malignancies,* 1, December.

Wapenaar, Hannah, and Frank J. Dekker. 2016. *Histone acetyltransferases: challenges in targeting bi-substrate enzymes,* 1, May.

Wei, Yi, Lanlan Yu, Josephine Bowen, Martin A. Gorovsky and C. David Allis. 1999. "Phosphorylation of histone H3 is required for proper chromosome condensation and segregation". *Cell* 97, no. 1 (April): 99–109. ISSN: 00928674.

Welch, John S., Allegra A. Petti, Christopher A. Miller, Catrina C. Fronick, Michelle O'Laughlin, Robert S. Fulton, Richard K. Wilson et al. 2016. "TP53 and Decitabine in Acute Myeloid Leukemia and Myelodysplastic Syndromes ". *New England Journal of Medicine* 375, no. 21 (November): 2023–2036. ISSN: 0028-4793.

Wolf, D., and V. Rotter. 1985. "Major deletions in the gene encoding the p53 tumor antigen cause lack of p53 expression in HL-60 cells". *Proceedings of the National Academy of Sciences of the United States of America* 82 (3): 790–794. ISSN: 00278424.

Wu, Di, Di Hu, Hao Chen, Guoming Shi, Irfete S. Fetahu, Feizhen Wu, Kimberlie Rabidou et al. 2018. "Glucose-regulated phosphorylation of TET2 by AMPK reveals a pathway linking diabetes to cancer". *Nature* 559, no. 7715 (July): 637–641. ISSN: 14764687.

Wu, Di, Elgene Lim, François Vaillant, Marie-Liesse Asselin-Labat, Jane E. Visvader and Gordon K. Smyth. 2010. "ROAST: rotation gene set tests for complex microarray experiments". *Bioinformatics* 26, no. 17 (September): 2176–2182. ISSN: 1460-2059.

Wu, Hong, Wenchao Wang, Feiyang Liu, Ellen E Weisberg, Bei Tian, Yongfei Chen, Binhua Li et al. 2014. "Discovery of a Potent, Covalent BTK Inhibitor for B-Cell Lymphoma." *ACS chemical biology* (February). ISSN: 1554-8937.

Wu, Xiaoji, and Yi Zhang. 2017. *TET-mediated active DNA demethylation: Mechanism, function and beyond,* 9, September.

Xu, Wei, Hui Yang, Ying Liu, Ying Yang, Ping Wang, Se Hee Kim, Shinsuke Ito et al. 2011. "Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of α-ketoglutarate-dependent dioxygenases". *Cancer Cell* 19, no. 1 (January): 17–30. ISSN: 18783686.

Yang, Zhenyu, Yangli Zhou and Lei Zhong. 2021. *Discovery of BAZ1A bromodomain inhibitors with the aid of virtual screening and activity evaluation,* February.

Zeng, Weihua, Jessica C. de Greef, Yen-Yun Chen, Richard Chien, Xiangduo Kong, Heather C. Gregson, Sara T. Winokur et al. 2009. "Specific Loss of Histone H3 Lysine 9 Trimethylation and HP1γ/Cohesin Binding at D4Z4 Repeats Is Associated with Facioscapulohumeral Dystrophy (FSHD)". Edited by Anne C. Ferguson-Smith. *PLoS Genetics* 5, no. 7 (July): e1000559. ISSN: 1553-7404.

Zhang, Bin, Jufang Chang, Ming Fu, Jie Huang, Rakesh Kashyap, Ezequiel Salavaggione, Sanjay Jain et al. 2009. "Dosage Effects of Cohesin Regulatory Factor PDS5 on Mammalian Development: Implications for Cohesinopathies". Edited by Ellen A. A. Nollen. *PLoS ONE* 4, no. 5 (May): e5232. ISSN: 1932-6203.

Zhang, Jixiang, Peng Tan, Lei Guo, Jing Gong, Jingjing Ma, Jia Li, Minjung Lee et al. 2019. "p53-dependent autophagic degradation of TET2 modulates cancer therapeutic resistance". *Oncogene* 38, no. 11 (March): 1905–1919. ISSN: 14765594.

Zhang, Ling, Kathy L. McGraw, David A. Sallman and Alan F. List. 2017. *The role of p53 in myelodysplastic syndromes and acute myeloid leukemia: molecular aspects and clinical implications,* 8, August.

Zhang, Yi, H.-H. Ng, Hediye Erdjument-Bromage, Paul Tempst, Adrian Bird and Danny Reinberg. 1999. "Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation". *Genes & Development* 13, no. 15 (August): 1924–1935. ISSN: 0890-9369.

Zhao, Yingming, and Benjamin A. Garcia. 2015. "Comprehensive catalog of currently documented histone modifications". *Cold Spring Harbor Perspectives in Biology* 7, no. 9 (September): a025064. ISSN: 19430264.