Université de Montréal

Proteogenomic Characterization of 5-Azacytidine Effects on Acute Myeloid Leukemia Immunopeptidome

Par

Nandita Noronha

Programmes de Biologie Moléculaire

Faculté de Médecine

Thèse présentée à la Faculté de médecine

en vue de l'obtention du grade de Docteur

en biologie moléculaire,

option biologie des systèmes

Avril 2023

Université de Montréal

Faculté de médecine

Programmes de biologie moléculaire

Cette thèse intitulée

Proteogenomic Characterization of 5-azacytidine Effects on Acute Myeloid Leukemia Immunopeptidome

Présenté par

Noronha Nandita

A été évalué(e) par un jury composé des personnes suivantes

Brian Wilhem
Président-rapporteur

Claude Perreault
Directeur de recherche

Tarik Möröy
Membre du jury

Vanessa Dumeaux
Examinateur externe

Résumé

La 5-azacytidine (AZA) est un médicament approuvé pour le traitement des leucémies myéloïdes aiguës des patients qui ne sont pas éligibles à une greffe de cellules souches hématopoïétiques. Bien que l'AZA est augmenté significativement le pronostic des patients, le mécanisme d'action précis de l'AZA demeure nébuleux. En plus de son activité d'hypométhylation, il a été montré que l'AZA a aussi des effets immunologiques. Des études précédentes suggèrent que ces réponses immunitaires sont causées par des modifications du répertoire de peptides présentés par le CMH-I (MAPs), dont l'expression de MAPs dérivés de rétroéléments endogènes (EREs) et des cancer-testis antigens (CTAs). Ces gènes sont généralement réprimés par la méthylation de l'ADN. Dans cette thèse, nous avons testé cette hypothèse à l'aide de séquençage à haut débit et de spectrométrie de masse appliqués à quatre lignées cellulaires d'AML différentes. Notre approche protéogénomique d'avant-garde a révélé que l'AZA induit la présentation de MAPs dérivés de CTAs, mais pas d'EREs, malgré le fait que ces deux groupes de séquences soient surexprimés au niveau transcriptomique. Ces résultats indiquent que les réponses des lymphocytes T observées chez les patients suite au traitement à l'AZA dépendent probablement des MAPs dérivés des CTAs, et non pas des EREs. Les EREs stimulés par l'AZA ont tout de même un impact sur la réponse immunitaire en formant des ARN double-brins menant à une activation de l'immunité innée. L'incorporation de l'AZA et l'inhibition subséquente de la DNMT2 mène cependant à des agrégats protéiques et à l'autophagie, qui dégrade les transcrits EREs et limite leur surexpression. Nous avons démontré que les effets immunologiques de l'AZA peuvent être amplifiés par un traitement combiné de l'AZA et d'inhibiteurs de l'autophagie. De plus, le travail contenu dans cette thèse a montré que bien qu'elles soient un modèle expérimental pratique, les lignées cellulaires ont des limitations et doivent être utilisés avec prudence. Des différences majeures ont été observées entre des lignées cellulaires supposément identiques provenant de fournisseurs établis. Nos analyses ont permis de démontrer quelle lignée cellulaire était la plus similaire à la lignée parentale. Ainsi, ce travail fourni des recommandations pour améliorer les lignes directrices d'utilisation des lignées cellulaires en recherche.

Mots-clés: Leucémie myéloïde aiguë, agents hypométhylants, immunothérapie, protéogénomiques, spectrométrie de masse, séquençage à haut débit, lignées cellulaires, reproductibilité de la recherche.

Abstract

5-azacytidine (AZA) is approved for the treatment of acute myeloid leukemia (AML) patients ineligible for hematopoietic cell transplantation. Although AZA treatment has substantially improved patient outcomes, there remains a lack of clear understanding of the mechanisms driving these responses. In addition to its hypomethylating activity, AZA has been shown to have immunological effects. Previous reports suggest that these immune responses occur due to alterations in the repertoire of MHC-I-associated peptides (MAPs), including the expression of MAPs deriving from endogenous retroelements (EREs) and cancer-testis antigens (CTAs). These genes are typically silenced by methylation. With this thesis, we aimed to test this hypothesis using high-coverage RNA sequencing and mass spectrometry in four different AML cell lines. Our state-of-the-art proteogenomic approach uncovered that AZA treatment induced MAPs deriving from CTAs, but not EREs, despite both being upregulated at the RNA level. This indicates that T-cell responses post-AZA treatment are more likely to be dependent on CTA- than ERE-derived MAP presentation. AZA-induced EREs produced at the RNA level still contributed to immune responses by forming double-stranded RNA leading to a state of viral mimicry. However, AZA incorporation into RNA and subsequent DNMT2-inhibition led to protein aggregation and autophagy responses. These responses were responsible for degrading EREs, which limited their upregulation. We further demonstrate that the immune effects of AZA can be enhanced by the combination of AZA with autophagy inhibitors. Additionally, the work in this thesis has shown that although a practical model, cell lines have their caveats and must be used with caution. This work has highlighted the grave discrepancies between supposedly identical cell lines supplied by established repositories. Moreover, our analyses determine which of the two is closer to the parental cell line. Finally, this work provides recommendations for improving the current guidelines for cell line-based research.

Keywords: Acute myeloid leukemia, hypomethylating agents, immunotherapy, proteogenomics, mass spectrometry, next generation sequencing, cell lines-based research, reproducibility of research

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AML	Acute myeloid leukemia
$2–5A_n$	ATP molecules bound by a 2’-5’ phosphodiester linkage
3D	Three dimensional
AIMs	AZA-induced IMmune
AML	Acute myeloid leukemia
APC	Antigen presenting cells
ASP	Aspartic acid
ATCC	American Type Culture Collection
ATGs	Autophagy-related genes
ATP	Adenosine triphosphate
AZA	5-Azacitidine
B2m	Beta-2 microglobulin
BPs	Base pairs
CHR	Chromosome
CMA	Chaperone-mediated autophagy
CPs	Constitutive proteasomes
CTAs	Cancer testis antigens
DAC	5-aza-2'-deoxycytidine
DCK	Deoxycytidine kinase
DCs	Dendritic cells
DEGs	Differentially expressed genes
DEMs	Differentially expressed MAPs
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
DNMTs	DNA methyltransferase enzymes
DOIs	DEMs of Interest
DRiPs	Defective ribosomal products
DSMZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen
dsRNA	Double stranded RNA
ENV	Envelope protein

ER	Endoplasmic reticulum
ERE	Endogenous retroelements
ERV	Endogenous retroviral
FAB	French, American, and British
FC	Fold change
FDA	Food and Drug Administration
GLU	Glutamic acid
GO	Gene-ontology
GSEA	Gene set enrichment analysis
HCT	Hematopoietic cell transplantation
HDAC	Histone deacetylase
HE-EREs	Highly expressed AZA-induced EREs
hENT	Human equilibrative and concentrative nucleoside transporters
HLA	Human leukocyte antigen
HMAs	Hypomethylating agents
HSC	Hematopoietic stem cell
HTC	Hotspots of transcriptomic changes
IFN	Interferon
IPs	Immunoproteasomes
IRF3	Interferon- regulatory factor 3
KFERQ	Lys-Phe-Glu-Arg-Gln
KZFPs	Krab-zinc fingers
LC	Liquid chromatography
LINEs	Long interspersed nuclear elements
LOH	Loss of heterozygosity
LTRs	Long terminal repeats
M/Z	Mass-to-charge
MAPs	MHC-I associated peptides
MAVS	Mitochondrial antiviral-signalling protein
MDS	Myelodysplastic syndromes
MHC-I	Major histocompatibility complex class I
mRNA	Messenger RNA

MS	Mass spectrometry
MSI+	Microsatellite instability
MSS	Microsatellite stable
MTORC1	Mechanistic target of rapamycin complex 1
NDPK	Nucleoside diphosphate kinase
NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NGS	Next generation sequencing
NIH	National institute of Health
NMPK	Nucleoside monophosphate kinase
NSCLC	Non-small cell lung cancer
OASes.	Oligoadenylate synthases
ORFs	Open reading frames
PAMPs	Pathogen-associated molecular patterns
PAS	Phagophore assembly site
PI3K	Phosphatidylinositol-3 kinase
PKR	Protein kinase R
POL	Polymerase
PRRs	Pathogen recognition receptors
RLRs	RIG-I-like receptors
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
RNase	Ribonucleases
rRNA	Ribosomal RNA
RT	Reverse transcriptase
SGI-110	Guadecitabine
SINEs	Short interspersed nuclear elements
STR	Short tandem repeats
STR	Short tandem repeat
TAAs	Tumor associated antigens
TAP	Transporter associated with antigen processing
TAs	Tumor antigens
TCR	T-cell receptors

TEs	Transposable elements
TLRs	Toll- like receptors
TNF	Tumor necrosis factor
TRIF	Toll/IL-1R domain-containing adaptor inducing IFN
tRNAs	Transfer RNAs
TSAs	Tumor specific antigens
UCK	Uridine kinase
UTRs	Untranslated regions
VAL	Valine
WHO	World Health Organization

To my "cheerleader" friends, mentors,

and parents for their unwavering support throughout this journey

Acknowledgments

First and foremost, I would like to express my deepest gratitude to Dr. Claude Perreault. Thank you for giving me this amazing opportunity to learn and grow in your laboratory and for believing in me! I feel fortunate to have you as a mentor. You have always encouraged me to achieve my best, pushed me whenever I've needed it, taught me to be a better scientist, to be bold and follow the data – even when it not what we had expected. I have learnt so much from you and your famous $I^2/R+C$ formula will stay with me forever. Moving forward, I'm sure every project I take on will be evaluated based on its **interest**, risk and cost!

Secondly, I will always feel indebted to Dr. Gregory Ehx. You are an amazing mentor, and I will be forever grateful that I had the opportunity to learn from the "The most talented Post-Doc/Professor". I know it always sounded overdramatic, but I have always truly meant it! Your talent speaks for itself, and I am happy that you could share that with me. I've enjoyed all our meetings, discussing data and planning our projects together. I will also always appreciate of how supportive you are. You gave me several opportunities to spearhead projects to learn and grow. You are one of the people who believe in collaborating and succeeding together, rather than alone and I will always value that quality in you.

The next shoutout goes to all my lab members without whom this PhD experience would NEVER be the same! My evil twin sisters, Lucyle and Justine who sat on either side of me and were there through the ups and downs. I will always cherish all the time we've spent together and the food and culture we got to experience and share. MaVi and Eralda, for those dinners and coffee chats and loooong conversations. Anca for sharing her beautiful, intelligent mind and the company for those swing dance nights. JD, for all the help in bioinformatics and his sense of humor and all the pranks we played – not to forget the infamous 'Wall of War'. Momo for looking out for me like I'm his lil sister. Juliette for her help when I was struggling with IF and Maxime for his autophagy help. Krystel and Marie-Pierre, for their help in MS data analysis and for whenever I needed help or advice for my presentations. Leslie for sharing your immunogenicity expertise and answering all my silly lil questions regarding them. Caro who has a heart of gold and was always willing to help and not to forget the pampering with 'em candies and chocolates!

I also have so many people from IRIC to thank for making this a wonderful experience. Jalila, for continuing to be a great mentor. I cannot account for all my successes without appreciating all the help and support that you have provided. Thank you for being so amazing! I also want to thank IRIC for giving a nourishing experience in the ability of being part of the student organizations, and mental health awareness program, and so on. Julie and Pascal for helping with all the nitty-gritties of the academic stuff and all the resources you have provided me with. I'm also grateful for all the collaborators and technicians from the MS, genomic, bioinformatics, flow cytometry and microscopy platforms.

I also want to extend my gratitude to the many friends and people who have been supportive at various moments in my life during the PhD. Melanie, Maria, Roger, Swati, Unain, Maude and Tariq thank you for being there. Tripti who was more friend than roommate – I will always appreciate your support and all the fun moments we had. Ragini who though miles apart has always been there for me. Anh-Khoi for being so supportive and encouraging for me to achieve my best. I appreciate y'all so much! My family in Toronto who I'm so grateful to have around me and spend time with. And last but not the least my amazing parents and brother. Thank you for being so supportive through this adventure in Montreal, providing me with help and advice whenever I needed and for being so understanding!

1 Chapter 1: Introduction

1.1 Acute myeloid leukemia

1.1.1 Overview

Acute myeloid leukemia (AML) is the second most common type of leukemia with overall survival of 32% [1]. It is typically a disease of older adults, with an incidence rate that increases with age. The median age of diagnosis of patients is approximately 68 years. AML is the most fatal type of leukemia. Compared with all subtypes of leukemias, AML has the lowest 5-year survival rate and accounts for the highest percentage (62%) of deaths [2].

AML is characterized by the rapid proliferation of immature blood cells leading to their accumulation in the blood and the bone marrow. The etiology of AML is rather heterogeneous and not well understood. AML can arise in patients previously exposed to treatments with DNA-damaging agents such as radiation, alkylating agents, and topoisomerase II inhibitors (also known as therapy-related AML). Patients with pre-existing hematological or genetic disorders also have an increased risk of developing AML. However, these cases, referred to as secondary AML, account only for a fraction of cases (10-30%) [3] and the majority of patients develop AML *de novo* without a history of hematological malignancies or known exposure to leukemogenic agents [4].

AML arises from a series of malignant transformations in hematopoietic stem cells (HSCs) or relatively early myeloid progenitor cells [5]. HSCs are multi-potent and self-renewing primitive cells that regenerate all possible blood cells via progenitors and precursors and maintain life-long hematopoiesis. In 2002, Gilliland proposed the two-hit model that stated that leukemias occur as

a consequence of cooperative mutations in HSCs or progenitors that confer proliferative advantage and mutations that impair hematopoietic differentiation [6]. This process leads to the formation of leukemic stem cells (LSCs) that are considered to be the cell of origin for this disease (Figure 1.1). These transformations include both cytogenetic and mutational events affecting normal hematopoietic growth and differentiation. Immature AML blasts thus formed, proliferate rapidly without differentiating, and affect normal functions of the blood. Additionally, epigenetic modifications alter gene expression contributing to the pathogenesis of AML. Thus, AML is a complex disease associated with phenotypic, genetic, and epigenetic heterogeneity.

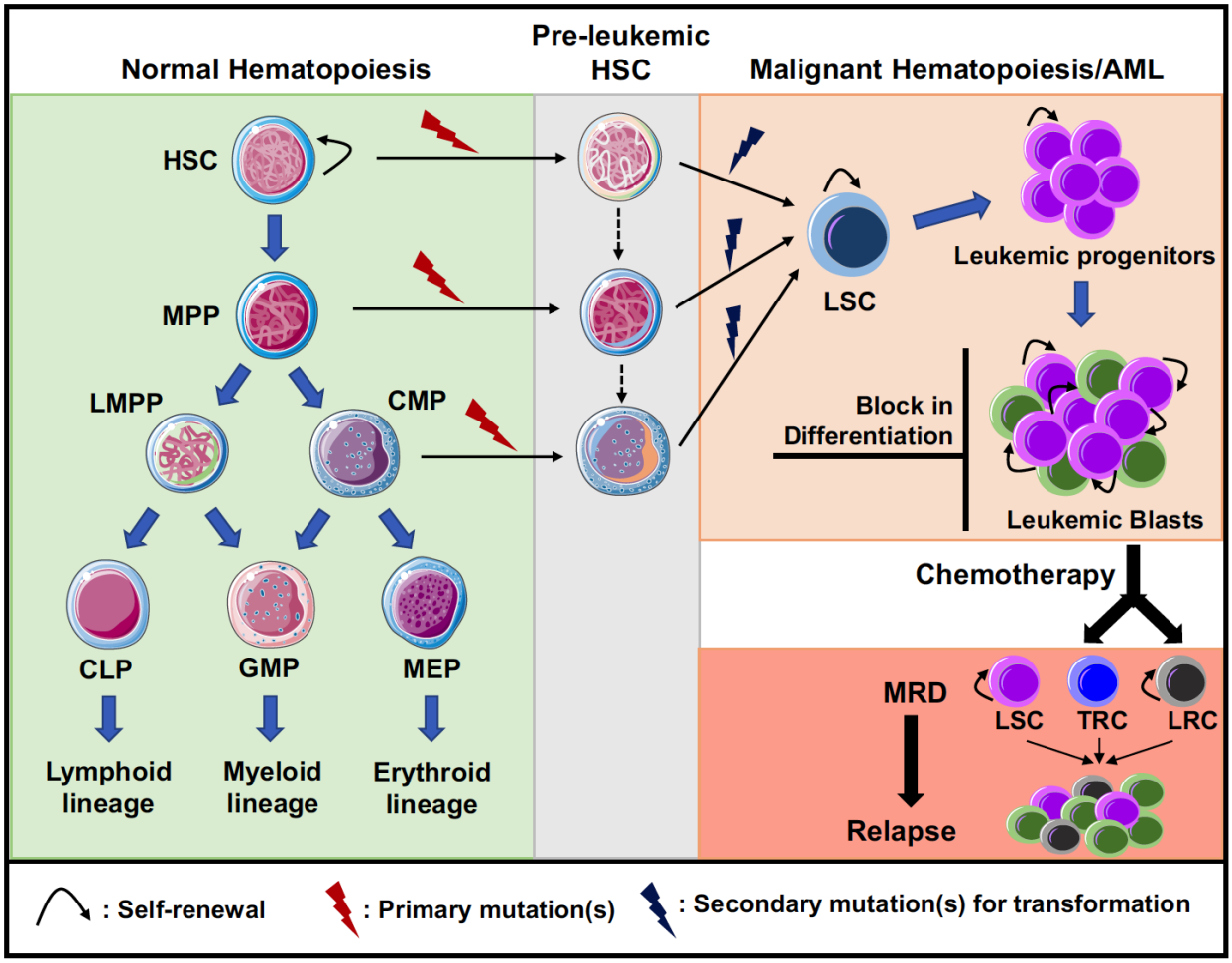

Figure 1.1. Comparison between normal and leukemic hematopoietic systems.

HSC, hematopoietic stem cell; MPP, multipotent progenitors; LMPP, lymphoid-primed MPPs; CMP, common myeloid progenitors; CLP, common lymphoid progenitor; GMP, granulocyte-macrophage progenitors; MEP, megakaryocyte-erythrocyte progenitor; B, LSC, leukemic stem cell, TRCs, (chemo)therapy-resistant cells; LRC, leukemia regenerating cell. Figure from Long et *al*., 2022 [7]

1.1.2 Classification of AML

Currently, two principal systems have been established and are commonly used for the classification of AML: The French, American, and British (FAB) system and the World Health Organization (WHO) system. The FAB system classifies AML based on morphology, cytochemistry, and degree of maturity of AML blasts [8]. Based on this, AML is divided into 8 sub-categories ranging from M0 to M7 (Table 1). However, this system does not account for the genetic and clinical heterogeneity of AML. In 1999, The WHO system of classification was introduced to include recent molecular biomarkers, genetic alterations, and immunophenotypic and clinical features of AML [9]. The WHO system includes 4 sub-groups of AML: AML with recurrent genetic abnormalities, AML with multilineage dysplasia, therapy-related AML and myelodysplastic syndromes (MDS), and those that do not categorize into any of these sub-groups (Table 2).

Table 1.1 FAB classification of AML

FAB subtype	Morphological classification
AML-M0	Undifferentiated acute myeloblastic leukemia
AML-M1	Acute myeloblastic leukemia with minimal maturation
AML-M2	Acute myeloblastic leukemia with maturation
AML-M3	Acute promyelocytic leukemia
AML-M4	Acute myelomonocytic leukemia
AML-M4 eos	Acute myelomonocytic leukemia with eosinophilia

AML-M5	Acute monocytic leukemia
AML-M6	Acute erythroid leukemia
AML-M7	Acute megakaryoblastic leukemia

Table adapted from Kumar et *al*., 2011 [8]

Table 1.2 WHO Classification of AML

AML and related neoplasms
AML with recurrent genetic abnormalities
AML with t(8;21)(q22q22.1); RUNX1-RUNX1T1
AML with inv(16)(p13.1q22) or t(16;16)(p13.1;q22);CBFB-MYH11
APL with PML-RARA
AML with t(9;11)(p21.3;q23.3); KMT2A-MLLT3
AML with t(6;9)(p23;q34.1); DEK-NUP214
AML with inv(3)(q21.3q26.2) or t(3;3)(q21.3;q26.2); GATA2, MECOM
AML (megakaryoblastic) with t(1;22)(p13.3;q13.1); RBM15-MKL1
Provisional entity: AML with BCR-ABL1
AML with mutated NPM1
AML with biallelic mutation of CEBPA
Provisional entity: AML with mutated RUNX1
AML with myelodysplasia-related changes
Therapy-related myeloid neoplasms
AML, not otherwise specified (NOS)
AML with minimal differentiation
AML without maturation
AML with maturation

Acute myelomonocytic leukemia
Acute monoblastic and monocytic leukemia
Pure erythroid leukemia
Acute megakaryoblastic leukemia
Acute basophilic leukemia
Acute panmyelosis with myelofibrosis
Myeloid sarcoma
Myeloid proliferations associated with Down syndrome
Transient abnormal myelopoiesis (TAM) associated with Down syndrome
Myeloid leukemia associated with Down syndrome

Table adapted from Hwang et *al*., 2020 [8]

1.1.3 Mutational landscape of AML

The advent of next-generation sequencing (NGS) has enabled a more comprehensive and precise analysis of the molecular landscape of AML. Moreover, NGS has emerged as a powerful tool in the diagnosis and prognosis of AML, especially in cases with normal karyotypes. Patients with normal karyotype represented approximately 50% of AML cases [8] and as the name suggests have no alterations in chromosomal number or structure. Therefore, cytogenetics alone cannot provide any information to predict prognosis or clinical relevance, and patients in this subgroup are associated with heterogeneous outcomes. NGS has enabled the identification of new markers for improving the risk stratification of AML patients to predict prognosis and overall patient survival (Table 1.3). These large-scale analyses have also uncovered that more than 90% of AML cases are associated with at least one mutation [10-12], with an average of 13 genes with mutations per AML sample [12]. These mutations were initially classified into two broadly defined groups

based on the two-hit model (Figure 1.2) [6]. The first class (class I) comprises mutations in genes involved in signaling pathways that offer a proliferative and/or survival advantage to hematopoietic progenitor cells [8]. Common class I mutations include mutations in FLT3, RAS, cKIT, and CBL [13]. The second class (class II) comprises mutations affecting transcription factors and components of the cell cycle that alter the differentiation of myeloid precursors [14]. Class II

Table 1.3 Risk Classification of AML Based on Genetics

Risk category	Genetic abnormality
Favorable	t(8;21)(q22;q22.1)/RUNX1::RUNX1T1 inv(16)(p13.1q22) or t(16;16)(p13.1;q22)/ CBFB::MYH11 Mutated NPM1without FLT3-ITD bZIP in-frame mutated CEBPA
Intermediate	Mutated NPM1with FLT3-ITD Wild-type NPM1 with FLT3-ITD (without adverse-risk genetic lesions) t(9;11)(p21.3;q23.3)/MLLT3::KMT2A Cytogenetic and/or molecular abnormalities not classified as favorable or adverse
Adverse	t(6;9)(p23.3;q34.1)/DEK::NUP214 t(v;11q23.3)/KMT2A-rearranged t(9;22)(q34.1;q11.2)/BCR::ABL1 t(8;16)(p11.2;p13.3)/KAT6A::CREBBP inv(3)(q21.3q26.2) or t(3;3)(q21.3;q26.2)/ GATA2, MECOM(EVI1) t(3q26.2;v)/MECOM(EVI1)-rearranged −5 or del(5q); −7; −17/abn(17p) Complex karyotype, monosomal karyotype Mutated ASXL1, BCOR, EZH2, RUNX1, SF3B1, SRSF2, STAG2, U2AF1, and/or ZRSR2 Mutated TP53

Table from Döhner et al., 2022 [15]

mutations include mutations in RUNX, MLL, CEBPα, and genetic re-arrangements such as PML-RARa, CBFb-MYH11, and RUNX1-RUNX1. Class I and class II mutations frequently co-occur,

supporting the two-hit model. However, ~40% of mutations in AML cannot be classified based on these two categories and several of these mutations have been associated with epigenetic alterations such as DNA methylation and histone modifications [13, 16, 17]. Examples of these include mutations in genes such as DNMT3, TET2, IDH1, IDH2, ASXL1, and WT1. Further, genetic alterations such as PML-RAR are tightly linked to epigenetic modifications contributing to leukemogenesis [18]. It is becoming increasingly clear that in addition to mutations, epigenetic dysregulation plays a major role in the development of AML. Furthermore, mutations in key epigenetic regulators, such as DNMT3a, TET2, and ASXL1, have been linked to a poor prognosis, highlighting the critical importance of maintaining the integrity of the epigenetic landscape in normal blood cells.

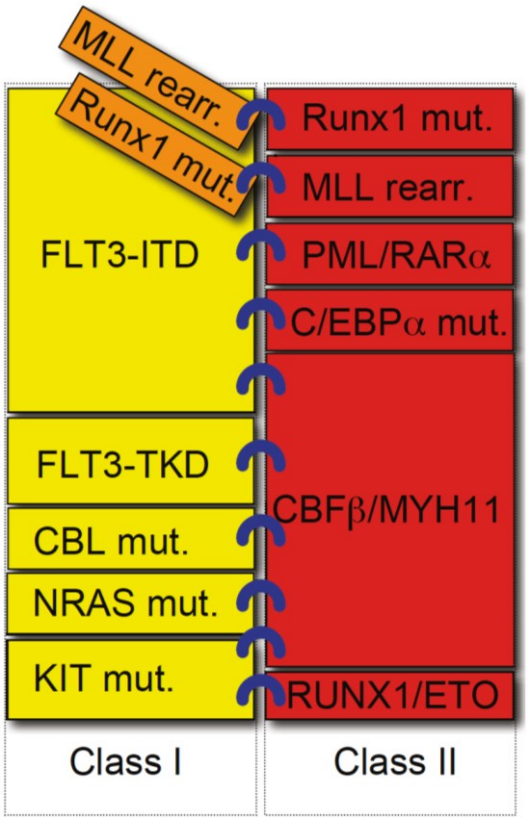

Figure 1.2. The two-hit model of leukemogenesis

The two-hit model proposes the collaboration of Class I mutations (in yellow) affecting signaling pathways providing a proliferative advantage and Class II mutations (in red) affecting transcription factors that hinder normal hematopoietic differentiation. The combinations of mutations are depicted as blue rings. Figure obtained from Takahashi et *al*., 2011[16]

1.1.4 Conventional therapy

Despite the advances in understanding the mutational and cytogenetic heterogeneity of AML, the current mainstay for patients remains almost unchanged for the last four decades. The objective of anti-AML treatment is to control and if possible, eliminate the disease. This is accomplished by initiating anti-AML therapy to achieve complete remission, followed by post-remission therapy to achieve long-term disease control and prevent relapse. Standard therapy consists of intensive induction therapy, which is a combination of 7 days of cytarabine, with 3 days of initial anthracycline treatment ("7 + 3" regimen). Patients are initially assessed for their eligibility (age and medical fitness) for intensive chemotherapy. These typically include patients belonging to favorable and intermediate-risk subgroups. Patients unfit for standard therapy receive alternative chemotherapy options such as Gilteritinib, Ivosidenib, and Enasidenib [15, 19]. Post-remission therapy involves additional chemotherapy (intermediate or high dose cytarabine) or allogeneic hematopoietic cell transplantation (HCT) [15].

1.2 Role of DNA methylation in AML

1.2.1 DNA methylation and associated enzymes

In mammalian cells, DNA methylation is an important epigenetic mechanism in which a methyl group is transferred onto the C5 position of cytosine residues leading to the formation of 5-methylcytosine residues [20]. DNA methylation suppresses gene expression both directly by inhibiting the binding of transcription factors and indirectly, by recruiting proteins involved in gene repression. These frequently occur at genomic regions rich in CpG dinucleotides called CpG

islands, predominantly found in the promoters of genes but can also be found in gene bodies and introns [21]. Interestingly, while DNA methylation in the vicinity of transcription start sites is associated with the silencing of gene expression, methylation in gene bodies is associated with enhanced gene expression and impacts gene splicing [22].

Methylation is carried out by a family of DNA methyltransferase enzymes (DNMTs) that comprise five members including canonical DNMT1, DNMT3A, and DNMT3B as well as non-canonical members, DNMT2 and DNMT3L. All members use a similar catalytic mechanism and cofactor S-adenosylmethionine as a methyl donor to carry out methylation [23]. DNMT3A and DNMT3B, are known as *de novo* DNMTs, as they can bind to unmethylated DNA and establish new methylation patterns to previously unmethylated DNA. DNMT1, on the other hand, requires the presence of hemimethylated DNA sequences for its binding and methylation of DNA. Thus, DNMT1 is also known as maintenance DNMT. DNMT3L lacks a catalytic motif present in other DNMTs and thus shows no methylation activity of its own. Instead, it acts as a cofactor and associates with DNMT3A and DNMT3B enhancing their methyltransferase activity. Finally, DNMT2 – initially thought to have DNA methyltransferase activity due to similarities in the sequence and structure to DNMT family members – was found to methylate tRNA [24]. Targets of DNMT2-mediated methylation are mainly transfer RNAs (tRNAs) coding for aspartic acid ($tRNA^{asp}$), glutamic acid ($tRNA^{glu}$), and valine ($tRNA^{val}$) [25]. While methylation of tRNAs protects them from misfolding and degradation (impacting protein synthesis rate and fidelity) [25, 26], DNA methylation has been shown to play an important role in genomic stability [27]. Alterations in DNA methylation have been shown to cooperate with genetic events leading to tumor initiation and progression [28].

1.2.2 Aberrant DNA methylation - A hallmark of AML

Recent research has established that in addition to transcription factors and signaling pathways, the regulation of DNA methylation is a critical factor in hematopoiesis. Several genes initially methylated in HSCs and early progenitors are selectively demethylated in a lineage-specific manner [29]. The resulting methylation patterns are critical in maintaining the HSC self-renewal capacity, as well as dictating hematopoietic differentiation and commitments to myeloid versus lymphoid cell fates [30, 31]. Given the importance of genomic methylation in maintaining normal hematopoiesis, it isn't surprising that aberrant DNA methylation is characteristic of several hematological malignancies [32]. Of note, several commonly reoccurring mutations in AML are associated with functional outcomes that disrupt genomic methylation. For instance, mutations in DNMT3A are observed in ~20% of AML patients; with an even higher prevalence in NK AML (~30-35%) [33, 34]. The most frequent mutation in DNMT3A occurs within the catalytic domain of the enzyme (R882H). This leads to the formation of dominant negative isoforms and the disruption of DNMT3A's methylating activity leading to gene hypomethylation [34]. Other important examples are mutations in TET2, IDH1, IDH2, and WT1. TET2 is a DNA demethylase enzyme that converts 5-methylcytosine to 5-hydroxymethylcytosine through its dioxygenase activity. Loss-of-function mutations in TET2 occur in ~6-27% AML [35] and are also associated with hypermethylation of DNA [36]. Further, AML patients with wild-type TET2 can still be subjected to enzymatic inactivation by mutated IDH enzymes. IDH is an important enzyme in the Krebs cycle and mutations in this gene occurs in ~20% of AML [37]. Mutant IDH enzymes produce 2-hydroxyglutarate, an oncometabolite that inhibits TET2 activity. Lastly, WT1 is a zinc-finger transcription factor involved in hematopoiesis and tumor suppression and directly interacts with and recruits TET2 to target genes to induce expression. Mutational profiling and in-depth

downstream functional investigation have uncovered that although these mutations occur as mutually exclusive events in AML patients, they share similar DNA hypermethylation patterns [38, 39]. This suggests that the shared downstream effects of aberrant hypermethylation significantly contribute to AML pathogenesis. Interestingly, DNA methylation profiling experiments have revealed that the majority of AML subgroups are associated with hypermethylated gene signatures compared to normal bone marrow cells [40]. Taken together, despite genetic and mutational heterogeneity across subgroups, aberrant methylation appears to be a common feature of AML.

1.3 Hypomethylating agents as a therapeutic option for AML

Given the significance of abnormal methylation in hematological malignancies, hypomethylating agents (HMAs) constitute a promising tool for the treatment of AML. HMAs are DNMT inhibitors that result in the hypomethylation of the DNA. The first generation of HMAs were pyrimidine nucleoside analogs and was originally developed as classical cytostatic agents [41]. At high concentrations, their DNA incorporation leads to the induction of DNA damage response and cytotoxic effects. Such doses when administered were found to be too toxic to patients without substantial antitumor benefits. However, when administered at low concentrations, these nucleoside analogs bind DNMT enzymes irreversibly and ultimately lead to their proteasomal degradation. Thus, by acting as suicide substrates for DNMTs, HMAs inhibit the function of DNMTs and consequently lead to DNA hypomethylation in subsequent replication cycles.

1.3.1 Cellular uptake and intracellular metabolism of HMAs

5-Azacitidine (AZA) and 5-aza-2'-deoxycytidine (DAC), are two widely used HMAs used in the treatment of myeloid malignancies (Figure 1.3). Both cytidine analogs enter cells by human equilibrative and concentrative nucleoside transporters (hENT/SLC29A and hCNT/SLC28A, respectively) and the SLC15 and SLC22 transporter families [42, 43]. Upon cellular uptake, AZA and DAC undergo different metabolic reactions to become their active form (Figure 1.4). This includes a series of ATP-dependent phosphorylation reactions, first by uridine kinase (UCK) or

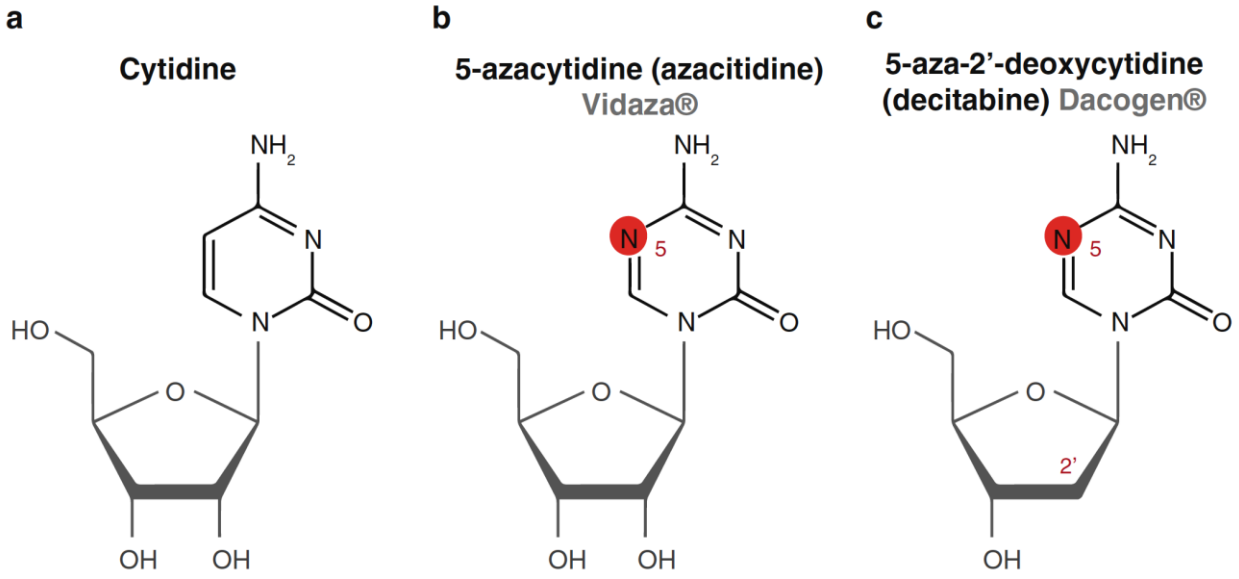

Figure 1.3. Chemical structures of cytidine and its analogs

The chemical structures of cytidine nucleoside (a) and 5-azacytidine (b) and decitabine (c) are shown, with the sugar moieties indicated in grey. Chemical modifications between cytidine nucleosides and azanucleosides are highlighted in red. Figure obtained from Diesch et *al*., 2016 [44].

deoxycytidine kinase (DCK) for AZA and DAC respectively, leading to their monophosphate forms of the analogs. This is followed by subsequent phosphorylation by nucleoside monophosphate kinase (NMPK) and nucleoside diphosphate kinase (NDPK) enzymes leading to

the formation of triphosphate forms of the nucleosides (Figure 1.4). While this enables DAC to incorporate directly into replicating DNA, only a small fraction (10-20%) of the triphosphate form of AZA enters DNA after multistep conversion reactions by ribonucleotide. The remainder (80-90%) of the AZA triphosphate forms enters RNA, leading to disruption of rRNA processing, tRNA demethylation, and ultimately inhibiting mRNA and protein synthesis [45-48].

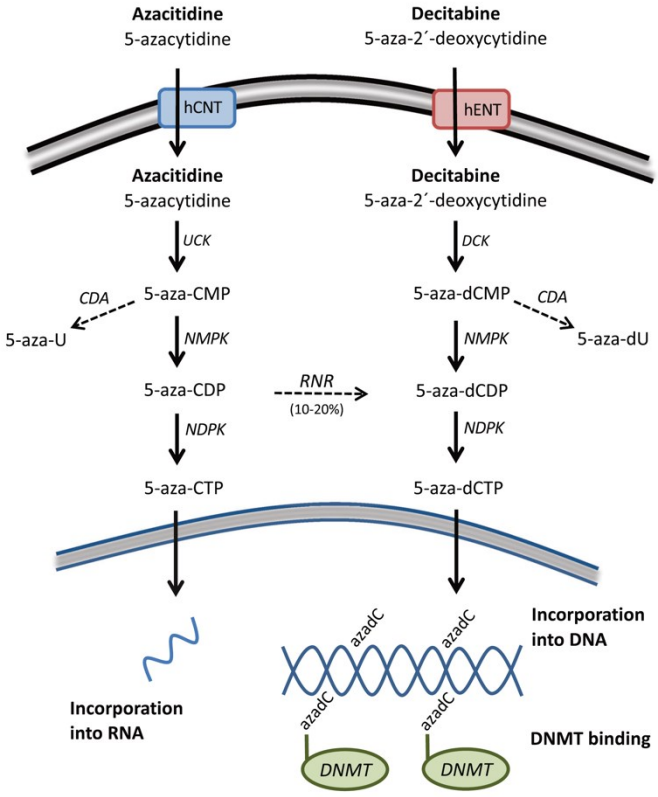

Figure 1.4. Cellular uptake and metabolism of AZA and DAC

CDP, cytidine diphosphate; CMP, cytidine monophosphate; hCNT, human concentrative nucleoside transporter; hENT, human equilibrative nucleoside transporter; NDPK, nucleoside diphosphate kinase; NMPK, nucleoside monophosphate kinase; RNR ribonucleotide reductase. Figure obtained from Stomper et *al*., 2021 [49].

1.3.2 Clinical administration of HMAs in patients

AZA and DAC have been tested in phase III clinical trials for the treatment of AML and have been shown to have similar response rates in AML [50]. Further meta-analysis uncovered that only AZA significantly improves the overall survival of patients compared to conventional care regimens [51, 52]. Therefore, AZA is currently FDA-approved as a first-line treatment in AML patients unfit for HCT transplantation [53]. Standard AZA treatment consists of repetitive cycles of 75 mg/m^2 injected daily subcutaneously for 7 consecutive days every 28 days [54]. However, the dynamics of AZA response are slow, and demethylation occurs gradually. Moreover, the reversibility of methylation necessitates several cycles of AZA or prolonged exposure to promote maximal hypomethylation [55]. Hence, AZA treatment is recommended to be continued for at least 4-6 cycles. Apart from improving overall survival, AZA administration has been associated with noteworthy improvements in other clinically relevant outcomes such as reductions in the need for transfusions, hospitalization, and intravenous antimicrobial use, and improvements in quality of life [55].

1.3.3 Predictive Molecular Biomarkers of Response to HMAs

Although AZA treatment has proved to be an essential tool in the treatment of AML patients, only ~50% of patients successfully respond [56]. Moreover, the response appears transient, with most responders ultimately becoming resistant within 2 years. Regrettably, even after numerous attempts to comprehend it, the precise biological mechanism of action remains obscure. Several studies have tried to identify the relationship between genomic methylation in patients and HMA treatment response. However, no significant differences in baseline methylation were observed among patients achieving complete or partial response compared to non-responders [57-59]. Moreover, given the association of mutations in genes coding for epigenetic modifiers

with aberrant genomic methylation in AML, several groups have tried to identify predictive clinical biomarkers, but no unambiguous evidence has emerged. While some studies suggested that mutations in TET2 could be used as a predictive marker for AZA response [60, 61], this was not reproducible in other studies and remains questionable [62-64]. Retrospective reports on DNMT3A mutations have shown that they can be predictive of better clinical response in previously untreated AML patients receiving HMAs [65, 66], but not in cohorts including relapsed and refractory patients [65]. However, these data could not be validated by other groups, while some groups found an opposite trend with DNMT3A mutants associated with a poorer prognosis and shorter overall survival [62, 64, 67, 68].

Chromatin reorganization has also been studied in response to HMA treatment. Experiments using Hi-C and whole-genome sequencing have uncovered that aberrant genomic methylation in AML is associated with alterations in the 3D chromatin structure, compared to healthy donors [69]. Interestingly, such aberrant 3D chromatin structure in AML can be reverted, at least partially to a more normal-like chromatin structure upon treatment with HMAs [69]. In another study, chromatin reorganization was shown to influence HMA treatment responses in AML [70]. RNA 5-methylcytosine transferases were found to interact with different molecular partners in AZA-sensitive and resistant AML cells leading to the formation of AZA-sensitive' or 'AZA-resistant' chromatin structures, respectively. Other studies have demonstrated that resistance is associated with cell cycle quiescence since increased quiescent leukemic progenitor cells were observed in non-responders than responders [71, 72]. A few studies have also demonstrated that induction of the autophagy pathway was found to be protective against AZA-mediated effects and an increase in basal autophagy was observed in AZA-resistant leukemic cells [73-75]. Finally, pharmacogenomic studies have reported that mutations or RNA expression in

genes coding for transporters and enzymes required for cellular uptake and metabolism of azanucleosides are associated with HMA resistance [76]. Reduced hENT1 and DCK expression were observed in non-responders and relapsed patients respectively [77]. Contrarily, the use of a novel mass spectrometry (MS) method has shown that non-responders do not have any defects in uptake and metabolism and both responders and non-responders have comparable levels of intracellular azanucleosides [78]. However, in comparison to responders, non-responders had higher levels of AZA incorporated into RNA than DNA, leading to a decreased demethylation of the genome. Thus, HMA resistance is heterogenous, and multiple factors are involved in the generation of HMA resistance.

1.3.4 New Avenues to Improve HMA treatment response

Several strategies have been employed to improve treatment responses towards HMAs, including the generation of new and improved HMAs as well as combination therapies to circumvent HMA resistance. Guadecitabine (SGI-110), a second-generation HMA whose active form has a longer half-life than intravenous decitabine and is resistant to degradation by cytidine deaminase (enzyme known to degrade decitabine and promotes resistance) was generated for the treatment of patients with myeloid malignancies. However, the Phase III ASTRAL-1 trial with SGI-110 showed that it was not superior to the treatment of choice (AZA/DAC/low-dose cytarabine) for complete remission rate and overall survival of previously untreated AML patients unfit for intensive chemotherapy [79]. CC-486 is another generation of HMA which is an oral form of AZA, generated to simplify the drug administration and dose adaptation. Phase III QUAZAR AML-001 trial showed that the new oral formulation significantly improved both median overall survival and relapse-free survival compared to placebo in older AML patients in remission after chemotherapy [80].

Other strategies involve combination therapies with drugs inhibiting histone deacetylase (HDAC) with the biological rationale to sensitize myeloid malignancies by completely releasing epigenetic silencing mediated by both DNMTs and HDACs. HDAC inhibitors such as entinostat, valproic acid, or vorinostat have been tested in multiple phase II clinical trials, however, no benefit was observed compared to HMAs alone in AML/MDS patients [81-84]. On the other hand, combination therapy with BCL-2 inhibitor Venetoclax has proved to be a promising therapeutic option. Phase III clinical trials with AZA combined with Venetoclax in previously untreated AML patients unfit for intensive chemotherapy demonstrated improved median overall survival. Moreover, patients receiving combination therapy had more rapid and durable responses as well as an increase in platelet transfusion independence [85]. The combination is now approved for the treatment of newly diagnosed AML patients unfit for intensive chemotherapy.

1.4 Immunological effects of HMAs

In recent years, there is a growing body of evidence suggesting that in addition to their demethylating activity, HMAs possess immunomodulatory activity. Consequently, these effects could potentially contribute at least in part to HMA's antitumor activity. In non-myeloid cancers such as non-small cell lung cancer (NSCLC), data from clinical trials [86] showed that a fraction of patients showed augmented and durable responses after receiving immune checkpoint treatment post-AZA therapy [87]. Preclinical studies using ovarian and melanoma mouse models have further focused on the synergy between checkpoint inhibitors and HMAs, showing that the combination increases overall survival compared to the use of checkpoint inhibitors alone [88, 89]. These data suggest that AZA improves the outcome of patients by sensitizing tumor cells for immune recognition, possibly by inducing immune activation pathways reversing immune evasion, or both. Accordingly, studies in breast and colorectal cancer mouse models show that

HMA treatment is associated with T-cell infiltration into the tumor microenvironment [90, 91]. Further, the anti-tumor benefit of HMAs in colorectal cancer was lost when $CD8^+$ T cells were depleted [91]. This suggests that the anti-tumor effects of AZA are dependent on T-cells, possibly via cytotoxic T-cell mediated recognition and elimination of tumors. Following this direction, some studies have shown that T-cell recognition is driven by HMA's ability to increase tumor immunogenicity by increasing the expression of genes coding for tumor antigens (TAs) [92-94]. Additionally, several studies have demonstrated that HMA's immunomodulatory responses could be explained by the expression of immune-related genes [87, 95-97]. Induction of gene expression could occur due to the release of epigenetic silencing mediated by DNA methylation [98]. In a study by Li *et al.*, gene expression analysis in a panel of 63 cancer cell lines (breast, colorectal and ovarian cancer) demonstrated that AZA induced a common subset of genes involved in immune responses called 'AZA-induced IMmune' (AIMs) gene set [97]. Following this study, two parallel studies in 2015 by Chiappinelli *et al.* and Roulois *et al.*, investigated the mechanism of AZA-induced immune responses in ovarian and colorectal cancer models. Both studies highlight that AZA induces the delayed expression of host-derived double-stranded RNA (dsRNA) and induction of innate immune response pathways [88, 99]. Moreover, host endogenous retroelements (EREs) are silenced by methylation and are known to form dsRNA due to their repetitive nature [100, 101]. Taken together, the literature suggests that AZA has several pleiotropic effects together contributing to its immunomodulatory activity. These concepts have been further described in detail in the following two sub-sections.

1.5 Endogenous retroelements (EREs)

Transposable elements (TEs) are highly repetitive sequences, first identified in maize by Barbara McClintock for their ability to mobilize and replicate "selfishly" within the host genome

[102]. TEs can be divided into two classes based on their mechanism of transposition (Figure 1.5). Class I or retrotransposons are TEs that use a "copy-and-paste" mechanism using an RNA intermediate [103], while class II TEs that use a DNA intermediate to mobilize using a "cut-and-paste" mechanism or "peel-and-paste" mechanism (in the case of the *Helitrons*, a TE family that uses a rolling-circle model for transposition) [104, 105]. Currently, all DNA transposons are inactivated in mammals and have lost the ability to mobilize (except for the *Piggybac* family in bats) [106]. These represent a minor fraction of the human genome (~3%). On the contrary, retrotransposons – often referred to as EREs – are the most abundant TEs in humans, occupying nearly half of the genome. EREs are thought to have originated millions of years ago from integration events in the host genome by foreign parasitic elements, predominantly deriving from retroviruses [107]. EREs rely on reverse transcriptase (RT) enzymes to reverse transcribe RNA intermediates to DNA and then insert them into target DNA sites.

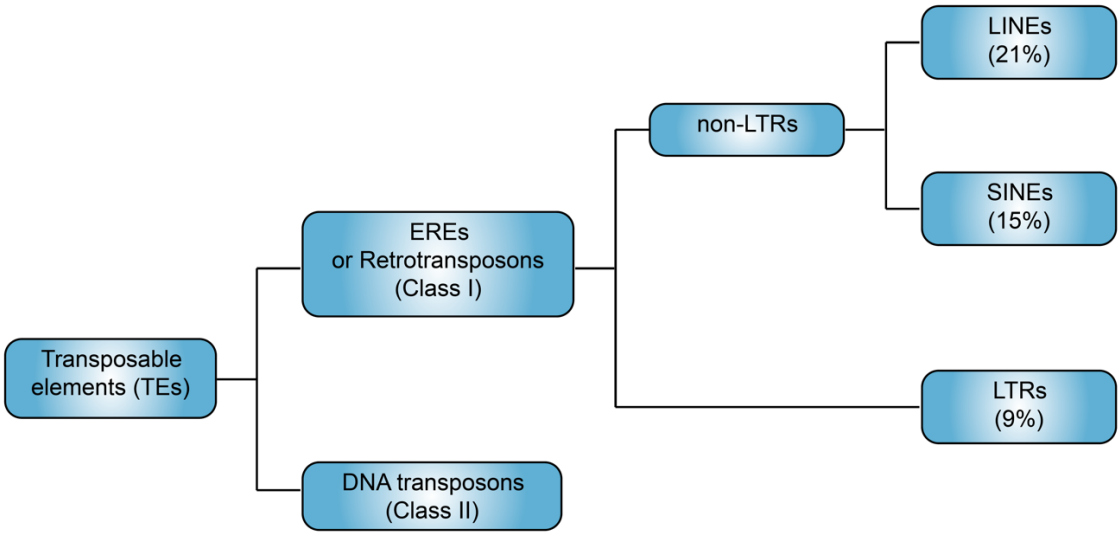

Figure 1.5. Classification and Organization of TEs in humans

Repetitive elements in the human genome can be broadly categorized based on their mechanism of transposition and presence or absence of specific domains such as long terminal repeats (LTRs).

Percentage of the genome represented by each of the ERE classes are indicated below. LINEs, Long interspersed nuclear elements; SINEs, Short interspersed nuclear elements

Although these repeat elements were often dismissed as "junk" or "non-coding" DNA, these elements are currently receiving an increasing amount of appreciation for their contributions to vital cellular processes. In some cases, EREs have become domesticated – a process known as exaptation. Exaptation refers to the ability to gain entirely new traits or perform new functions during evolution. Thus, although these sequences were once known to be detrimental due to their "selfish" purposes of replication, some of these have co-evolved with the human genome to provide new functions. Examples of this phenomenon include CENP-B, a protein involved in centromere assembly during cell division, and several endogenous retroviral (ERV) encoded Env proteins such as syncytins and ERV3, present in the placenta and important for embryonic development [108, 109].

1.5.1 Classification of EREs

EREs can be further classified based on the presence of flanking long terminal repeats (LTRs) (Figure 1.5). In humans, the majority of EREs lack LTRs, together referred to as non-LTR elements. Non-LTR elements occupy 36% of the host genome and consist of long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs) [110]. While no ERE can be genuinely autonomous since they depend on host machineries such as RNA polymerases, ribosomes and tRNAs (for LTRs [111]) to be replicated, EREs that encode their own enzymes for transposition are referred to as autonomous. LINEs and LTRs are considered autonomous EREs. LTRs are structurally related to retroviral genomic sequences and contain open reading frames (ORFs) that code for an RT enzyme (*pol*) and RNA binding proteins (*gag*). The *pol* gene encodes

for a multifunctional polypeptide chain with RT, protease, RNase H, as well as integrase activities. The *gag* gene encodes for a structural capsid-like protein, inside which the reverse transcription reaction takes place. Some LTRs families, often referred to as ERVs, contain an additional ORF coding for an envelope protein (Env). In the majority of cases, the ORFs between the LTRs bear several frameshift mutations or premature stop codons, rendering them non-functional. In some cases, these elements exist only as solo-LTRs, a product of unequal homologous recombination between LTRs leading to the excision of the ORFs [112]. These LTRs are thus non-autonomous and dependent on autonomous EREs for their transposition. While LTRs use their flanking terminal repeats for transcription initiation and termination, non-LTRs use promoters and enhancers in 5'-untranslated regions (UTRs) of their own gene or of nearby host genes to initiate their transcription. In general, LINEs contain an internal promoter for RNA polymerase II, followed by one or two ORFs and a short 3' UTR that terminates in a polyadenylation signal and poly-A rich tail. LINE-1, the most studied LINE family, consists of ORFs coding for an RNA-binding protein (coded by ORF1) and a bifunctional polypeptide with RT and endonuclease activities (coded by ORF2). SINEs have a much simpler sequence compared to LINEs and lack any coding potential. They only contain an internal promoter for RNA polymerase III (RNA pol III) followed by a poly-A rich sequence instead of a polyadenylation signal. Thus, SINEs are non-autonomous and depend on LINEs for their replication and transposition. SINEs are the only EREs that use RNA pol III for their transcription and are ancestrally related to tRNA, 5S rRNA, and 7SL RNAs [113]. Some SINEs can combine to form complex structures such as dimers or sometimes even trimers [114]. These include the *Alu* family, which represents the majority of SINE families in humans [115].

Finally, EREs are not randomly distributed within the genome. Contrarily, since ERE transposition is target-specific, they have a preference for certain genomic features or compartments. For instance, L1 is enriched at AT-rich genomic regions, while Alu elements have been identified to be located in GC-rich regions [116]. Moreover, for the successful persistence of these parasitic elements, there must be a fine balance between disrupting essential host gene functions and thriving in the genome. While only some can form replication-competent viruses or transpose within the genome, the majority of these elements are inactivated or suppressed by the host.

1.5.2 Host defenses to control EREs

In response to their expansion and detrimental effects, the host has evolved to develop several defense strategies involving both transcriptional and post-transcriptional mechanisms. These have been detailed below:

Transcriptional repression of EREs: As described earlier, DNA methylation is an epigenetic mechanism for silencing gene expression. However, nearly three decades ago, DNA methylation was posited by Bestor to be evolved as the primary mechanism for the control of transposons [117, 118]. This was supported by observations in prokaryotes that used DNA methylation as a defense mechanism to prevent the integration of exogenous DNA in the host genome [118]. More recent research by Peter Jones' group has further explored this hypothesis and validated the existence of a tight interplay between DNA methylation, ERE levels, and genome expansion across organisms [119]. Invertebrates with fairly smaller genome sizes, low levels of ERE, and higher proportions of coding sequences had lower CpG representation than expected. This was contrary to organisms such as fish and amphibians with intermediate-sized genome sizes and relatively more ERE levels with higher levels of CpG dinucleotides. Accordingly, a striking

negative correlation was observed between genome size and CpG representation, and a positive correlation between ERE and genome size. This data suggests that ERE insertion into host DNA leads to an increase in genome size and along with it, an increase in CpG sites available for methylation for its repression. Further, cytosine methylation is known to be inherently mutagenic [120]. Early experiments in bacteria have well established this phenomenon, showing that methylated DNA can accumulate up to 21-fold higher levels of spontaneous mutations compared to unmethylated DNA [121]. This can be attributed to the fact that methylated cytosines are prone to undergo deamination [122]. This leads to the formation of erroneous thymine bases, followed by DNA mismatch repair pathways leading to C-to-T and G-to-A transition mutations. For vertebrate genomes, the estimated time predicted for such CG > AT transitions to take place is 450 million years [123]. In the case of EREs, the accumulation of mutations over time could eventually negatively impact their mobility and render them non-functional. This suggests that, evolutionarily, the long-term consequence of methylation is to reduce genome expansion by reducing CpG content and possibly unfavor ERE transposition. This is supported by data demonstrating that the mechanisms of ERE repression differ based on the age of genomic integration of ERE elements [124]. Evolutionarily younger LTRs have higher CpG densities and are predominantly suppressed by DNA methylation. These EREs are sensitive to HMA treatment and are upregulated upon treatment. Intermediate-age LTRs, on the other hand, had lower CpG densities and were less sensitive to HMA treatment. This is in line with previous bodies of research demonstrating cytosine loss in ERE sequences over evolutionary time. Interestingly, several studies have proposed an arms race between EREs and components of the host defense system. These involve fast-evolving transcription factors, mainly krab-zinc fingers (KZFPs) – the majority of which act as transcriptional repressors [125, 126]. KZFPs have been shown to rapidly co-evolve

with EREs to maintain their suppression. KZFPs have zinc finger domains that allow DNA binding and can recruit co-factor KAP1 (TRIM28) through their KRAB domains. KAP1 is responsible for the recruitment of downstream epigenetic remodeling factors including histone methyltransferase enzymes. Supporting the arms race theory, published data demonstrates that intermediate-age LTRs are regulated by histone modifications, mainly by H3K9 methylation instead of DNA methylation. Analyses of the chromatin state of EREs across cell types in human cells further reveal that the majority of KZFP ERE targets are associated with heterochromatin-associated H3K9 methylation marks [127]. In summary, transcriptional control involves suppression of ERE expression primarily by DNA methylation that eventually leads to mutational inactivation and co-evolution of transcriptional factors to maintain repression.

Post-transcriptional control of EREs: EREs that escape this first line of transcriptional suppression are further controlled by post-transcriptional mechanisms. Dysregulation of cellular and epigenetic pathways in cancer and autoimmune disorders frequently leads to the re-expression of ERE transcripts. ERE-derived transcripts can assemble into dsRNA that resemble retroviral nucleic acids. In humans, host sensors recognize these ERE transcripts as 'foreign' nucleic acids and trigger interferon signaling and anti-viral signaling pathways, a phenomenon commonly known as viral mimicry [88, 99]. Further, autophagy is additionally responsible for regulating EREs, by degrading ERE sequences and thus controlling their transposition [128].

Key players involved in the anti-viral defense response and autophagy are described in detail in the following sub-sections.

1.5.3 dsRNA and innate immune system

The recognition of pathogens and foreign nucleic acids is a key mechanism as the first line of defense. This system includes the detection of evolutionarily conserved sequences on pathogens called pathogen-associated molecular patterns (PAMPs). PAMPs are detected by host germline-encoded pathogen recognition receptors (PRRs). Once recognized, this triggers a cascade of signaling pathways leading to inflammatory responses, and the release of cytokines and chemokines that are subsequently responsible for the elimination of the pathogens. In the case of an infection with viruses, viral dsRNA often accumulates intracellularly during viral replication. The host system has evolved to build sensors that recognize dsRNA as PAMPs. These include RIG-I-like receptors (RLRs), Toll-like receptors (TLRs), protein kinase R (PKR), and oligoadenylate synthases (OASes) (Figure 1.6).

RLRs and TLR3: RLRs are RNA helicases that act as cytosolic dsRNA sensors. They include RIG-I, MDA-5, and LGP2. Upon activation, both RIG-I and MDA-5 form filaments across the length of the dsRNA and share downstream signaling pathways by the recruitment and activation of the adaptor protein, mitochondrial antiviral-signaling protein (MAVS). LGP2 does not directly regulate the antiviral signaling cascade but is thought to regulate the activity of the former two RLRs. RIG-I and MDA-5 play non-redundant functions as they detect distinct types of dsRNA based on their differential binding preferences of dsRNA length and end structure. MDA-5 detects larger dsRNAs (~500–1,000 base pairs (bps)), while RIG-I detects smaller dsRNAs (~22–500 bp) and dsRNAs with a 5’-triphosphate group (5’-ppp) or 5’-diphosphate (5’-pp). In host cells, during the normal processing of mRNA, nascent transcripts undergo processing in the nucleus to cleave the 5’-ppp. Many viruses replicate in the nucleus and hence retain the 5’ppp, allowing dsRNA sensors such as RIG-I to distinguish between host and viral RNA species.

TLR3 is an endosomal dsRNA sensor and requires an acidic environment for its activation. TLR3 is present on the endosomal membrane of cells and prefers binding dsRNAs with lengths of 40-50 bps. Upon activation form dimers on a single dsRNA molecule leading to the recruitment of adaptor molecule Toll/IL-1R domain-containing adaptor inducing IFN (TRIF). Both RLR and TLR3 activation culminates in the activation of type I interferon signaling via activation of transcription factors interferon-regulatory factor 3 (IRF3) and nuclear factor kappa-light-chain-enhancer of activated B cells (NF- κB). TLR3 detects dsRNA released from viruses-infected cells or dying cells via endocytosis. Thus, TLR3 is considered a cell-extrinsic sensor, contrary to RLRs that detect dsRNAs originating within the same cell (cell intrinsic sensor).

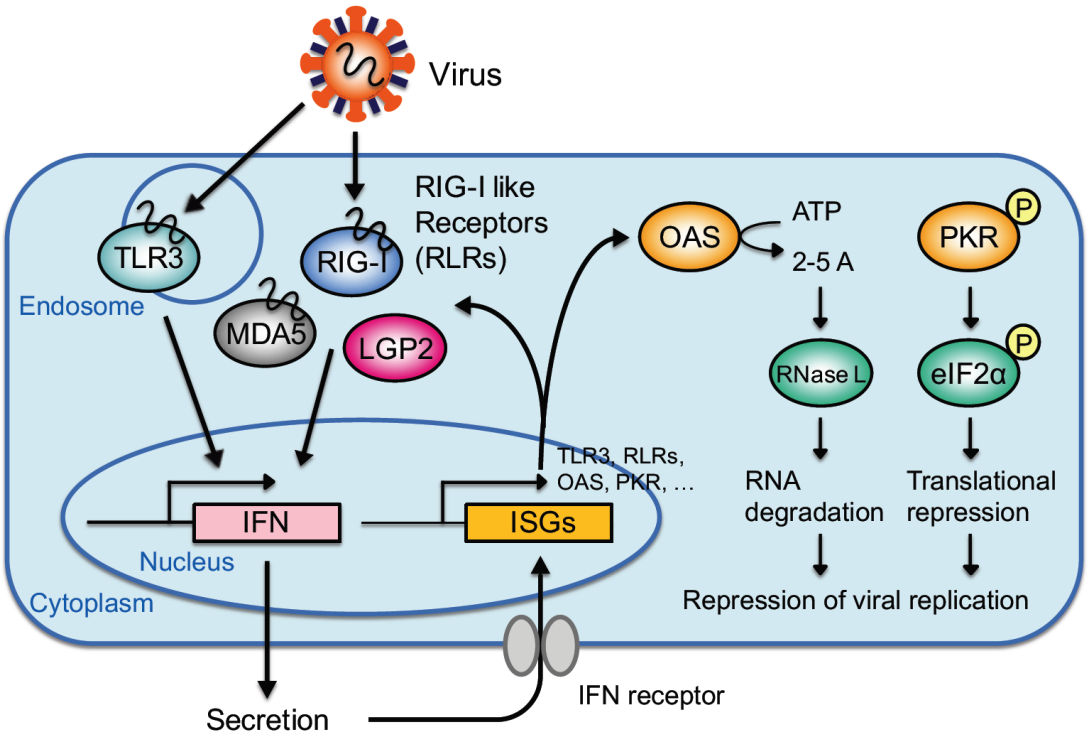

Figure 1.6 Cellular host responses against anti-viral dsRNA

Cellular hosts recognize dsRNA using cytoplasmic retinoic acid-inducible gene I (RIG-I)-like receptors (RLRs) or endosomal toll-like receptor 3 (TLR3). RLRs include RIG-I, MDA5, and LGP2 receptors. Activated TLR3 or RLRs culminate in similar downstream pathways leading to interferon (IFN) production. Secreted IFN results in the expression of IFN-stimulated genes

(ISGs). Expression of ISGs such as oligoadenylate synthetase (OAS) and protein kinase R (PKR) activates signaling pathways that ultimately lead to RNA degradation or translational repression, respectively. Figure obtained from Takahashi et *al*., 2020 [129].

PKR and OASes: PKR is an IFN-inducible dsRNA-dependent protein kinase. The binding of dsRNA to PKR leads to a change in its conformation, dimerization, and autophosphorylation of PKR resulting in the formation of an active kinase. The main substrate of PKR is EIF2α, a negative regulator of protein synthesis. Phosphorylation of EIF2α results in the global shutdown of translation and cell growth inhibition. PKR has also been shown to be involved in pro-inflammatory signaling pathways by acting as an upstream kinase in the activation of NF-κB [130]. A minimum dsRNA length of ~33 bp is required for PKR binding and activation.

Similar to PKRs, OASes are IFN inducible and are involved in the inhibition of translation. There are 4 human isoforms of OASes. These include OAS1, OAS2, OAS3, and OASL, the latter being an inactive form of OASes. Upon binding to dsRNA, OASes undergo conformation changes leading to the synthesis of 2′-5′-linked oligomers of adenosines (dimers up to 30-mers) which are ATP molecules bound by a 2'-5' phosphodiester linkage (2–5A_n). 2–5A_n are secondary messengers that activate an endonuclease, ribonuclease L (RNase L). Activated RNase L cleaves RNA, both cellular and viral leading to global translation shutdown, inhibition of cell growth, and viral replication.

Thus, depending on the type of dsRNA sensor detected, different pathways are activated with the common goal to inhibit viral replication. Activation of RLRs and TLR3 receptors leads to antiviral signaling and inflammation responses, while PKR and OASes activation leads to inhibition of protein synthesis and cell growth. The induction of ERE by AZA involves similar

signaling cascades, where key players such as OAS, TLR3, MDA5, and RIG-I have been shown to play a major role [88, 99, 131].

1.5.4 Autophagy

Autophagy is a highly conserved catabolic pathway and is important for cellular homeostasis and survival. The term autophagy literally translates to "self" and "eating" in Latin and involves the degradation of long-lived proteins as well as unwanted or harmful materials such as damaged organelles and protein aggregates. One of the most typical roles of autophagy is to provide nutrients to cells upon cellular stress. Thus, nutrient deprivation (ex. amino acids, glucose) or growth factor starvation (ex. insulin, IL-3) triggers autophagy leading to the degradation and recycling of cellular components [132]. This facilitates the provision of building blocks to synthesize new materials and provide energy in the cell. Additionally, autophagy may act as a protective mechanism in various pathologies such as cancer, aging, and neurodegenerative disorders such as Alzheimer's and Parkinson's as well as against pathogenic organisms including bacteria and viruses as well as EREs as a defense mechanism [128, 133]. In mammalian cells, there are primarily three types of autophagy detailed below.

Macroautophagy: The most well-researched type of autophagy is macroautophagy which is often referred to as the classical form of autophagy. Macroautophagy is largely non-selective and performs bulk degradation of cellular components. This process begins with an initiation phase wherein an induction complex comprising of proteins coded by autophagy-related genes (ATGs) is recruited to a subcellular location called a phagophore assembly site (PAS). The initiation complex is also known as the ULK complex and consists of ULK1, ULK2, ATG13, FIP200, and ATG101 (figure 1.7). This step is highly regulated by the mechanistic target of rapamycin complex 1 (MTORC1). During nutrient sufficiency, mTORC1 directly blocks the autophagy via a

phosphorylation-dependent inhibition of ULK1 and ATG13. Under conditions of nutrient starvation, mTORC1 is inactivated and disassociates from the induction complex. This causes the dephosphorylation of ULK1 and ATG13, and macroautophagy is induced. The induction complex is followed by the nucleation of the membrane to form a cup-like structure known as a phagophore. Nucleation is facilitated by the ULK-dependent phosphorylation of a class III phosphatidylinositol-3 kinase (PI3K) complex (consisting of a class III PI3K, Beclin-1, VPS34, VPS15, and ATG14). This step is regulated by a ubiquitination complex controlled by deubiquitinating peptidases USP10 and USP13. USP10 and USP13 negatively regulate the activity of the PI3K complex by regulating the degradation of Beclin-1. After successful nucleation, the phagophore is gradually elongated until it ultimately expands and becomes spherical encompassing a portion of the cytosol along with the cytoplasmic constituents. This leads to the formation of a double-membraned vesicle called the autophagosome. Finally, the autophagosome fuses with the lysosome and with it the lysosomal enzymes that facilitate the degradation of cellular components and release into the cytosol for further reuse.

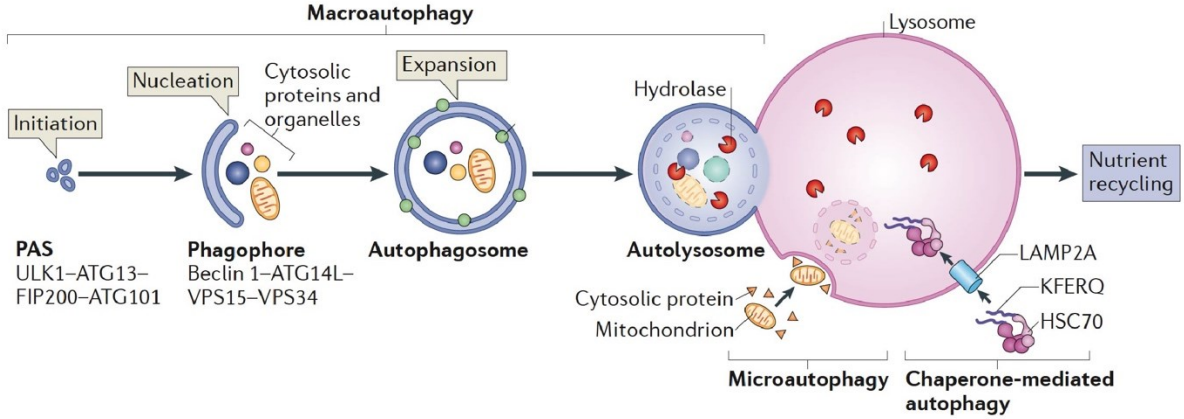

Figure 1.7 Overview of the three types of autophagy in mammalian cells

Macroautophagy involves the formation of cytosolic double-membrane vesicles at the phagophore assembly site (PAS) and involves the ULK complex (ULK1, ULK2, ATG13, FIP200, and

ATG101). This is followed by a nucleation step facilitated by a class III phosphatidylinositol-3 kinase (PI3K) complex (PI3K, Beclin-1, VPS34, VPS1,5, and ATG14). This then expands to form autophagosomes. Substrates are sequestered within and transported to the lysosome during the fusion of the autophagosomes and lysosomes (autolysosomes) where they are degraded by hydrolases. In microautophagy, cytoplasmic cargos are taken up directly via the invagination of the lysosomal membrane. Finally, Chaperone-mediated autophagy involves the translocation of specific proteins with pentapeptide motif KFERQ with the chaperone, HSC70 and proteins directly enter the lysosome via the LAMP2A receptor. Figure adapted with permission from Kaur et *al*., 2015 [134].

Microautophagy: Microautophagy is a process in which portions of the cytosol and its constituents are taken up directly by lysosomes, vacuoles, or late endosomes (figure 1.7). This direct engulfment occurs via the invagination or deformation of the organelles' membrane. Like macroautophagy, microautophagy can be induced by nutrient starvation and rapamycin and has been demonstrated to uptake cytosolic components such as protein aggregates and organelles including mitochondria, peroxisomes, and nuclear fragments [135]. However, with the limited tools to study microautophagy, there is relatively little known about this process since its discovery, and the field is still growing.

Chaperone-mediated autophagy (CMA): Unlike macroautophagy and microautophagy, CMA is a highly selective form of autophagy. Only proteins with a pentapeptide C-terminal sequence, Lys-Phe-Glu-Arg-Gln (KFERQ) are targeted to undergo CMA (figure 1.7) [136]. Upon induction, cytosolic chaperones are recruited to peptide sequences containing the KFERQ consensus motif. Target proteins are then translocated onto the lysosomal membrane, where they are unfolded and transported into the lysosomal lumen, followed by their degradation. CMA has been shown to account for the degradation of ~30% of cytosolic proteins during nutrient starvation [137]. Interestingly, the majority of aminoacyl-tRNA synthases contain these motifs suggesting

that these enzymes are a substrate for degradation by CMA [138]. This suggests that under the context of nutrient starvation in cells, CMA degradation of aminoacyl-tRNA synthases could be physiologically relevant in slowing down the rate of protein synthesis. Other substrates targeted for CMA include transcription factors, proteasome subunits, and vesicular trafficking proteins, among others [139].

1.6 Antigen processing and presentation in the MHC-I pathway

As discussed in the above few sections, the innate immune system acts as the first line of defense with very broad specificity to foreign pathogen invasion. We now discuss the MHC-I pathway, a part of the adaptive immune system which has a more targeted recognition of foreign pathogens and other "non-self"-derived antigens.

1.6.1 The MHC-I immunopeptidome

$CD8^+$ T-cells have access to cellular information by surveilling peptides deriving from intracellular proteins presented by MHC-I molecules on the surface of cells. These peptides are collectively known as the immunopeptidome [140]. During the life span of a cell, MHC-I-associated peptides (MAPs) are normally generated from the host own's proteins involved in housekeeping functions [141]. However, during viral infections or in the case of cancerous cells, peptides deriving from viral or tumor-specific proteins are presented to T-cells. Thus, the MHC-I pathway is crucial in triggering T-cells to quickly eliminate infected or malignant cells. Three important concepts are discussed below giving a brief overview of the MHC-I pathway.

MHC-I complexes: MHC-I molecules are heterodimers consisting of a heavy chain and a constant Beta-2 microglobulin (B2m) chain and bind to 8-11 amino acid long peptides. In humans,

the MHC-I is also known as the human leukocyte antigen (HLA) system. The heavy chain, encoded by highly polymorphic HLA genes (HLA-A, HLA-B, or HLA-C), is present on the short arm of chromosome 6 (chr 6p21). After synthesis, nascent MHC-I polypeptides are folded and assembled in the lumen of the endoplasmic reticulum (ER). Peptides are translocated from the cytosol to the ER via a transporter associated with antigen processing (TAP) proteins, where they are trimmed and loaded onto MHC-I molecules. This is performed by the peptide loading complex, a multi-subunit machinery present in the ER enabling the final assembly of MHC-I complexes with the B2M and the peptide. MHC-I-peptide complexes are then trafficked to the Golgi apparatus and are finally transported to the plasma membrane where they are ready to interact with circulating $CD8^+$ T-cells (Figure 1.8).

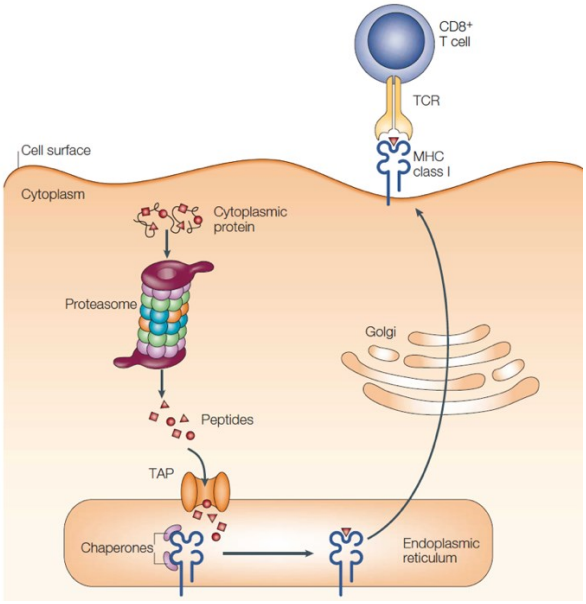

Figure 1.8 Antigen processing and presentation by MHC-I molecules

Intracellular proteins are degraded by the proteasome into peptides. The peptides are then translocated via the transporter for antigen processing (TAP) into the lumen of the endoplasmic reticulum (ER) where they bind to MHC class I molecules. Peptide binding is supported by chaperones and is essential for the correct folding of MHC class I molecules. Finally, MHC-peptide complexes are released from the ER and transported to the plasma membrane, where the

peptide can interact with T-cells via T-cell receptors (TCRs). Figure adapted from Yewdell et *al*., 2003 [142] with permission.

Peptide sources: Peptides generating MAPs arise from the degradation of two main categories of source proteins – older proteins undergoing normal turnover (retirees) and defective ribosomal products (DRiPs). DRiPs are nascent polypeptides that do not reach their mature conformational states or are misfolded as a result of translational infidelity, tRNA insufficiency, or errors in post-translational processes. This leads to their rapid degradation either right after or during translation (Figure 1.9). DRiPs are, therefore, not represented at the level of the proteome, but are observed at the immunopeptidomic levels. DRiPs represent ~30% of total proteins that undergo proteasomal degradation with half-lives of several minutes, whereas retirees are more stable proteins with half-lives of ~46h. Interestingly, DRiPs but not retirees are a major source of MAPs [143-145]. Thus, predominantly, newly synthesized peptides are represented at the level of the immunopeptidome allowing T-cells to surveil the latest "current events" of the cell, including the presence of a replicating virus.

Peptide generation: Peptides are generated by ubiquitylation and degradation of intracellular proteins by proteasomes. Constitutive proteasomes (CPs) are the main type of proteosomes expressed by all eukaryotes pivotal for several cellular processes, including cell cycle, stress response, apoptosis, differentiation, and other regulatory processes. CPs contain a 19S regulatory particle, required for binding polyubiquitylated substrates, and a core 20S proteolytic particle mediating peptide hydrolysis. The proteolytic activity of the 20S core particle is catalyzed by 3 subunits – PSMB5 (cleaves after hydrophobic residues; chymotrypsin-like activity), PSMB6 (cleaves after acidic residues; caspase-like activity), and PSMB7 (cleaves after basic residues; trypsin-like activity) [147]. In addition to CPs, exposure to IFN-γ or TNFα leads to the generation of a special form of the proteasome, known as immunoproteasomes (IPs). While the 19S particle

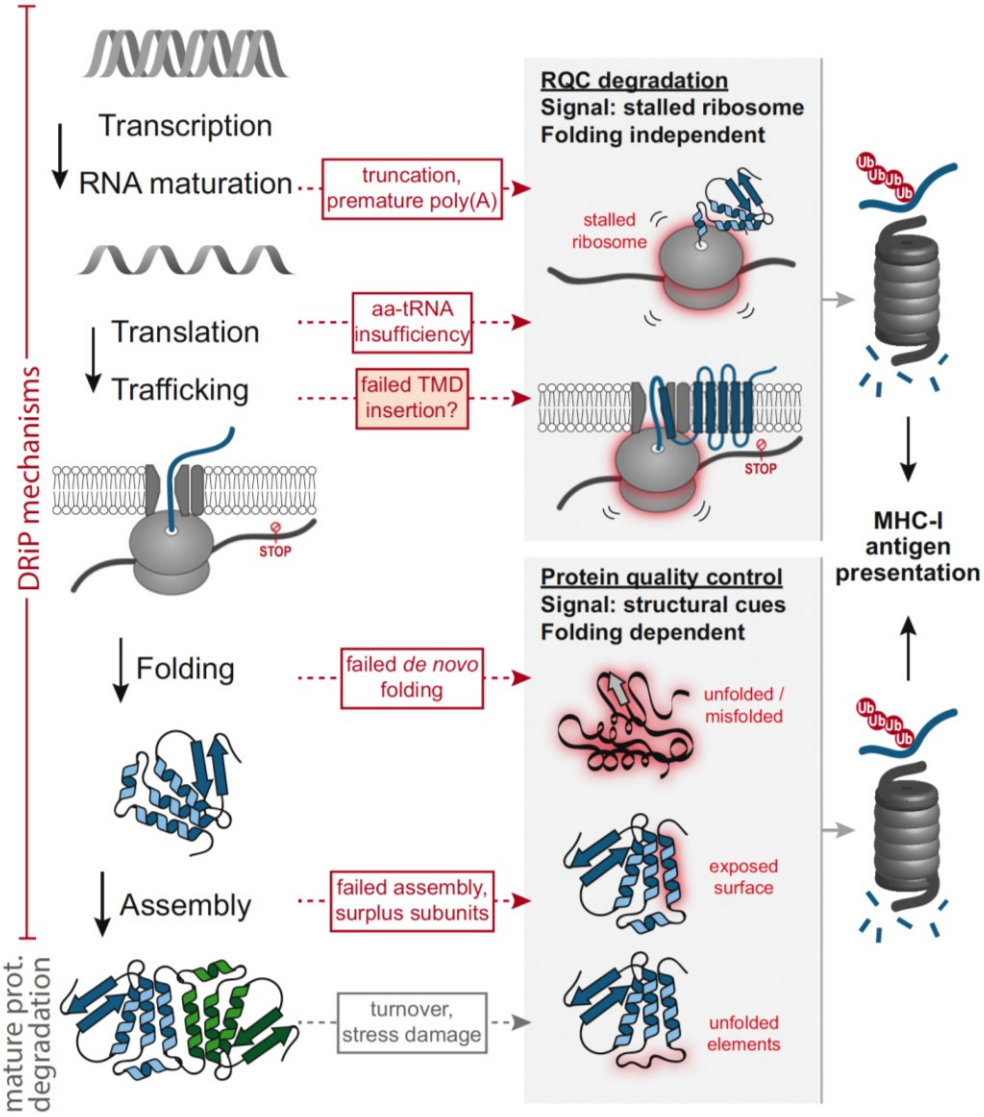

Figure 1.9 MAP generation during protein synthesis and turnover

The process of protein biosynthesis (left) and turnover is highly regulated both ribosome-associated quality control (RQC) and protein quality control are responsible for MAPs. During the translation of nascent polypeptides, errors or lack of amino acids leads to ribosome stalling and defective ribosomal products (DRiPs) generation and degradation in a folding-independent manner (Grey top panel). During protein maturation, the absence of proper assembly of subunits or protein misfolding can further trigger protein degradation (Grey lower panel). Figure adapted from Trentini et *al*., 2020 [146] with permission.

is shared between the two proteasomes, the 20S catalytic subunits in IPs are replaced by PSMB8, PSMB9, and PSMB10. PSMB8 and PSMB10 have similar proteolytic activities to their CP homologs PSMB5 and PSMB7, respectively), while PSMB9 has chymotryptic-like activity rather than caspase-like activity as observed in PSMB6 [147]. Due to these modifications, IPs and CPs are endowed with differences in catalytic properties and cleavage specificities (Figure 1.10). Accordingly, IPs preferentially cleave peptides more after C-terminal hydrophobic residues and basic residues, and less after acidic residues [148]. Finally, IPs appear to have increased chymotryptic than tryptic activity [149]. As a result of these differences, IPs and CPs generate different sets of MAP repertoires [150]. Moreover, IPs are especially important in influencing both the quantity and quality of the peptide repertoire. Not only does the absence of IPs shows a decreased abundance and diversity of MAPs, but also drastically affects T-cells responses in viral infections [151, 152]. IP deficiency is considered a feature of non-small cell lung cancer and is frequently associated with poor outcomes [153]. Further, IP upregulation in breast cancer was associated with better overall survival and increases in $CD8^+$ T-cell infiltration [154]. Thus, MAPs derived from immunoproteasomes are important for protection from viruses, and tumors.

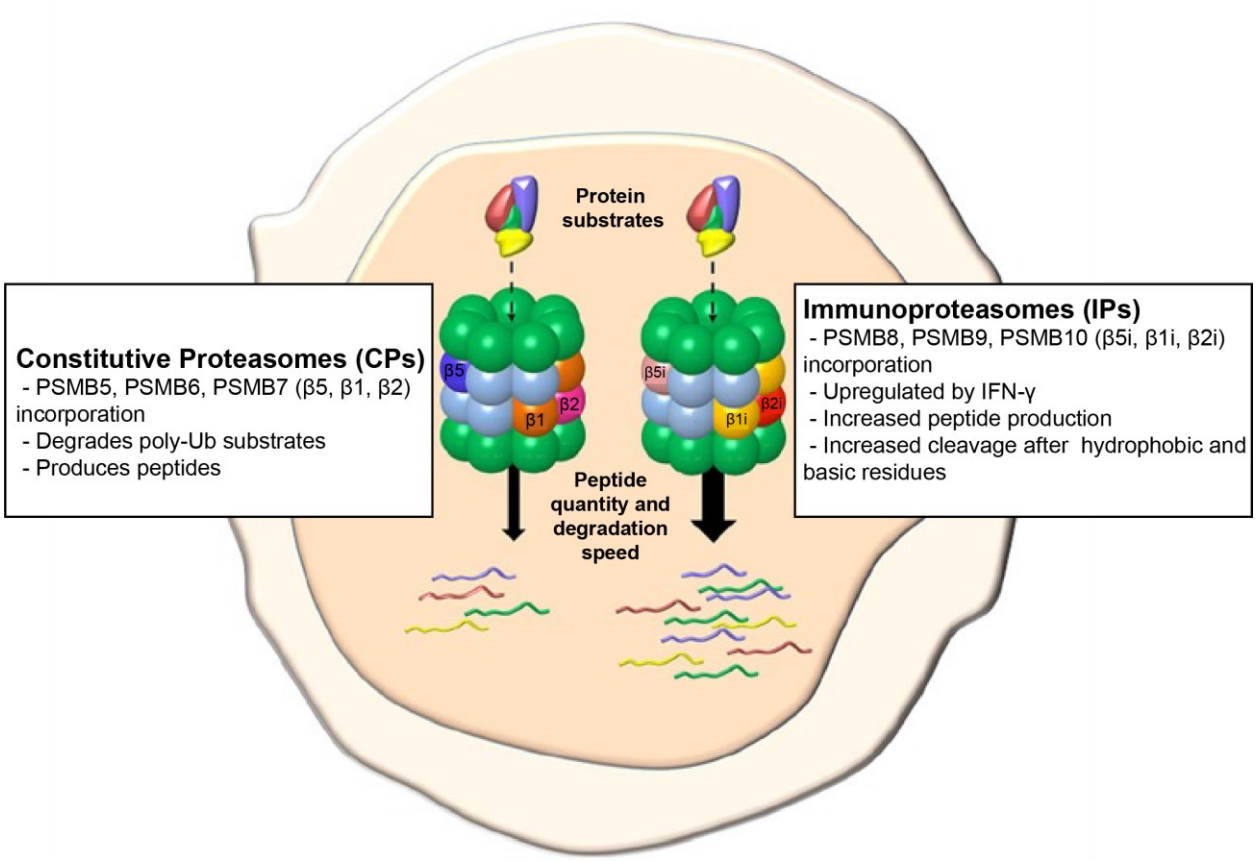

Figure 1.10 Schematic representation of differences between Constitutive Proteasomes and Immunoproteasomes

Constitutive proteasomes (left) and Immunoproteasomes (right) differ in 20S catalytic subunits. The proteolytic subunits of the CPs are PSMB5 (also known as β5), PSMB6 (also known as β1), and PSMB7 (also known as β2), and of IP are PSMB8 (also known as β5i), PSMB9 (also known as β1i) and PSMB10 (also known as β2i). These differences endow IPs with an increase in preference for cleavage after hydrophobic and basic residues, and increased chymotrypsin-like activity compared to CPs. Figure adapted from Zanker et *al*., 2014 [155] with permission.

1.6.2 Leveraging TAs to improve cancer therapy

Immunotherapy has transformed the landscape of cancer therapy. In comparison to previous standards of care, immunotherapy has brought a significant improvement in overall survival and quality of life for cancer patients. The concept of harnessing the immune system to eliminate cancer cells is not a recent or novel one. In the nineteenth century, William Coley – often referred to as the "Father of cancer immunotherapy" – demonstrated that the injection of extracts from heat-inactivated bacteria achieved favorable responses in cancer patients [156]. However, the absence of valid scientific explanations for Coley's results, and with the development of radiotherapy, this treatment regimen did not gain traction and was not pursued further [157]. It was only in the early twentieth century that this concept was revisited and the relevance of T-cells in anti-tumor immune responses was recognized [158, 159]. Compelling evidence demonstrated that infiltration of T-cells in tumors can lead to tumor regression in melanoma and some other cancers [160-163]. However, not all tumor types are associated with T-cell infiltration and tumor regression. This brings attention to the nature of MAPs presented by tumor cells and investigation of the tumor MAP repertoire is important for two reasons. First, the global immunopeptidome is a representation of the internal processes occurring in tumor cells and accounts for what is displayed to $CD8^+$ T-cells. Secondly, tumor cells are considered poor "antigen presenters" as they lack the necessary co-stimulation markers to efficiently prime T-cells [164]. Identifying MAPs that can serve as TAs is of critical importance as they can be leveraged as cancer vaccines and improve anti-tumor-specific T-cell responses.

Where to look for TAs candidates? Until recently, the focus on identifying TAs was solely on MAPs deriving from annotated protein-coding exon regions. These included regions that are over-expressed in tumors compared to normal tissues (tumor-associated antigens; TAAs) or

that are mutated or aberrantly expressed in tumor cells (Tumor-specific antigens; TSAs). Research from our lab was pioneering in establishing that allegedly non-coding genomic regions serve as an important source of TAs [165]. Some of the non-canonical sources generating TAs include UTRs, intergenic and intronic regions, and EREs. These have been identified and validated by ribosequencing and MS in several cancers including, leukemia, melanoma, ovarian, and colorectal cancers by our and other research groups [166-170]. Of note, pre-clinical mice studies show that TSAs deriving from EREs are highly immunogenic [165], probably due to the high sequence similarity of EREs with viruses. Naïve mice had similar $CD8^+$ T cell frequencies specific to ERE-derived TSAs as viral peptides and immunization with these EREs showed long-term tumor protection. Most importantly, these are shared across patients and/or tumor types making them attractive vaccine candidates that could be used to target a sizeable fraction of patients [167, 170-172].

Another class of TAs worth mentioning is the cancer-testis antigens (CTAs). Although deriving from annotated protein-coding exons, these have been identified in several cancers and serve as an attractive source of TAs candidates. Depending on their expression in tumor and normal tissues, CTAs can be classified as TAAs or TSAs. Generally, CTA gene expression is silenced in normal cells, except for germline cells such as the testis and placenta. Germline cells are considered immune privileged as they lack MHC-I expression. Frequently due to epigenetic dysregulation, CTAs are aberrantly overexpressed in cancers such as melanoma, bladder cancer, and others, including AML [173, 174]. Thus, although CTA genes are expressed by germline cells, CTA-derived MAPs cannot be presented by MHC-I molecules to the $CD8^+$ T-cells. This makes MAPs deriving from CTAs attractive targets for immunotherapy [175]. There are over 70 CTA gene families with over 240 genes and considerable efforts have been made to identify tumor-

specific CTA candidates for cancer vaccines [176]. Several clinical trials targeting CTA-derived TAs as anti-tumor therapy that have been FDA-approved are currently ongoing or have been completed. Although these trials are early phase I/II trials, some showed promising results with considerable safety, longer median free survival, or good clinical responses associated with CTA-specific immunological responses [177, 178].

How do we identify TAs candidates? Some research groups rely on reverse immunology using exome or transcriptomic data and MHC-I predicting algorithms to identify TAs; however, the majority of these predictions are false positives and cannot be further validated [179, 180]. MS, on the other hand, remains the best method for the identification and validation of MAPs and their amino acid sequences. Briefly, MAPs are immunoprecipitated using an MHC-I specific antibody, followed by dissociation from the MHC-I subunits and separation on liquid chromatography (LC) columns before being introduced into the mass spectrometer. Before entering the mass spectrometer, peptides are electrically charged (ionized) and transferred into the gas phase in a process called electrospray ionization. It's important to note that the efficiency of ionization can vary by several orders of magnitudes between peptides [181]. Therefore, while the abundance of a single peptide can be compared across samples, the abundance of two different peptides cannot be compared accurately. Once in the gas phase, peptides can be detected and sorted based on their mass-to-charge (m/z) ratio. The separated peptides are then measured, and their m/z and relative abundance are stored by a data system. The spectrum of peptides – the m/z ratio plotted against the relative abundance of the peptide – eluted at this point is called an MS1 spectrum. These peptides can then be further fragmented by colliding them with inert gases such as nitrogen or argon. Peptides are fragmented at the peptide bond between the amino acids and the resulting spectrum is called MS2 or MS/MS spectrum. Thus, the MS1 spectra reveal the mass of the intact

peptide and the MS2 spectra determine the mass of the amino acid fragment ions. Since the masses of amino acids are already known and well established, a theoretical MS2 spectra database can be built based on the MS1 information. Unlike RNA sequencing wherein *de novo* sequencing is used to construct the transcript sequences, MS uses theoretical databases to identify peptide sequences. Databases searches typically query for the best peptide in the theoretical databases that can explain the observed MS2 spectra. This feature of the MS comes with an important caveat – only those peptide sequences within the database can be identified. Therefore, the databases used for identification must be complete, without being not too large and unmanageable. This is especially important for the search for MAPs deriving from non-canonical genomic regions. Protein databases typically used for MS searches such as Uniprot lack non-coding sequences, and therefore cannot be used to identify such MAPs. One possible solution is to build databases using a proteogenomic approach and create MAP sequences corresponding to RNA transcripts expressed by the samples. This approach allows for building an ideally sized, but relevant database for assigning MS2 spectra to peptides as accurately as possible. Although MS-based methods remain the best methods to identify MAPs, this technique is limited by the requirement of large quantities of samples for analyses. Further, they may be inaccessible to many research groups due to the high cost of MS instruments as well as a need for experts and technicians specialized in the field. This could be circumvented by collaboration with research groups with MS expertise. Further, the use of cell lines as a model system instead of primary tumor samples can overcome the restriction of the sample size required for MS analyses.

1.7 Cell lines as an experimental model in biological research

Cell lines are a valuable tool that have been used for decades in biological research and have enabled progress in several scientific fields, including biotechnology and pharmacology. Cell lines are typically generated by immortalizing cells derived from primary tissues or tumors. Thus, cell lines can proliferate indefinitely *in vitro*, provided they have a suitable nutrient medium and sufficient space to grow. Cell lines serve as an appropriate model for researchers to perform experiments in a controlled manner without having the restriction of sample sizes observed for primary samples. They have been used to study physiological, and pathophysiological conditions, and differentiation processes, as well as elucidate mechanisms underlying tumor cells. Moreover, cell lines make an ideal model for performing large-scale drug screening assays, identification of drug targets, drug metabolism, and its effects. Finally, cell lines have played an important role generation in vaccines and gene therapies. Despite their practicality and ease of culture *in vitro*, cell lines come with their fair share of limitations. First, cell lines are an oversimplistic experimental model and may lack physiological relevance and not be representative of pathways occurring in actual tissues and tumors. Secondly, cell lines are prone to genetic instability leading to genetic drift and phenotypical differences. This is known to be especially true for cell lines with microsatellite instability (MSI+) as they have dysfunctional damage repair mechanisms, leading to the accumulation of genomic mutations [182]. Finally, handling cell lines without adequate caution can inadvertently lead to cell line cross-contamination, misidentification, or mycoplasma contamination. In fact, studies have reported that 15% of cell lines from labs across North America, Europe, and Asia were misidentified or contaminated [183]. Further, a recent study with HeLa cell lines procured from across 13 labs worldwide demonstrated significant variances and divergence after several passages leading to inconsistent scientific results [184]. Several agencies are

addressing this issue by setting up guidelines to ensure reliability and reproducibility. For instance, the National Institute of Health (NIH) issued guidelines that require the authentication of cell lines for researchers applying for grants [185]. Similarly, several scientific journals have added a strict mandate for cell line authentication for the submission of research articles [186]. The International Journal of Cancer has been in the lead in this matter and published several articles on the misrepresentation of cell lines as a caution to the scientific community and provided resources to improve the situation [187-190]. Currently, cell lines are primarily authenticated using genotype-based methods, with short tandem repeats (STR) profiling as the gold standard. STR profiling is a convenient and inexpensive method wherein 8-10 highly polymorphic loci are amplified with commercially available primers. The sizes of amplified loci differ across individuals and thus the resulting profile produced is unique and characteristic of any particular individual. Thus, STR profiles can be used to compare with references in databases to validate cell lines' identity and detect cross-contamination or misidentification.

In summary, the use of cell lines in research is critical but requires careful consideration when used. By ensuring and adhering to strict guidelines and good laboratory practices, researchers can ensure the responsible and effective use of cell lines to advance scientific research. Finally, to ensure the physiological relevance and increase confidence when data using cell lines are generated, validation in other models including patient or primary samples is important.

1.8 References

1. *Surveillance, Epidemiology, and End Results (SEER) Program Populations (2013–2019)*. Available from: https://seer.cancer.gov/statfacts/html/amyl.html.
2. Howlader, N., et al., *SEER cancer statistics review, 1975–2018.* National Cancer Institute, 2021.
3. Shallis, R.M., et al., *Epidemiology of acute myeloid leukemia: Recent progress and enduring challenges.* Blood reviews, 2019. **36**: p. 70-87.
4. Cheson, B.D., et al., *Revised recommendations of the international working group for diagnosis, standardization of response criteria, treatment outcomes, and reporting standards for therapeutic trials in acute myeloid leukemia.* Journal of clinical oncology, 2003. **21**(24): p. 4642-4649.
5. Bonnet, D. and J.E. Dick, *Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell.* Nature medicine, 1997. **3**(7): p. 730-737.
6. Kelly, L.M. and D.G. Gilliland, *Genetics of myeloid leukemias.* Annual review of genomics and human genetics, 2002. **3**(1): p. 179-198.
7. Long, N.A., et al., *Acute myeloid leukemia stem cells: Origin, characteristics, and clinical implications.* Stem Cell Reviews and Reports, 2022. **18**(4): p. 1211-1226.
8. Kumar, C.C., *Genetic abnormalities and challenges in the treatment of acute myeloid leukemia.* Genes & cancer, 2011. **2**(2): p. 95-107.
9. Vardiman, J.W., N.L. Harris, and R.D. Brunning, *The World Health Organization (WHO) classification of the myeloid neoplasms.* Blood, The Journal of the American Society of Hematology, 2002. **100**(7): p. 2292-2302.
10. Papaemmanuil, E., et al., *Genomic classification and prognosis in acute myeloid leukemia.* New England Journal of Medicine, 2016. **374**(23): p. 2209-2221.
11. Morita, K., et al., *Clearance of somatic mutations at remission and the risk of relapse in acute myeloid leukemia.* Journal of Clinical Oncology, 2018. **36**(18): p. 1788.
12. Network, C.G.A.R., *Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia.* New England Journal of Medicine, 2013. **368**(22): p. 2059-2074.
13. Lagunas-Rangel, F.A., et al., *Acute myeloid leukemia—genetic alterations and their clinical prognosis.* International journal of hematology-oncology and stem cell research, 2017. **11**(4): p. 328.
14. Grimwade, D., A. Ivey, and B.J. Huntly, *Molecular landscape of acute myeloid leukemia in younger adults and its clinical relevance.* Blood, The Journal of the American Society of Hematology, 2016. **127**(1): p. 29-41.
15. Döhner, H., et al., *Diagnosis and management of AML in adults: 2022 recommendations from an international expert panel on behalf of the ELN.* Blood, The Journal of the American Society of Hematology, 2022. **140**(12): p. 1345-1377.
16. Takahashi, S., *Current findings for recurring mutations in acute myeloid leukemia.* Journal of hematology & oncology, 2011. **4**(1): p. 1-11.
17. Chen, S.-J., Y. Shen, and Z. Chen, *A panoramic view of acute myeloid leukemia.* Nature genetics, 2013. **45**(6): p. 586-587.
18. Di Croce, L., et al., *Methyltransferase recruitment and DNA hypermethylation of target promoters by an oncogenic transcription factor.* Science, 2002. **295**(5557): p. 1079-1082.

19. Döhner, H., et al., *Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel.* Blood, The Journal of the American Society of Hematology, 2017. **129**(4): p. 424-447.
20. Robertson, K.D., *DNA methylation and human disease.* Nature Reviews Genetics, 2005. **6**(8): p. 597-610.
21. Weber, M., et al., *Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells.* Nature genetics, 2005. **37**(8): p. 853-862.
22. Jones, P.A., *Functions of DNA methylation: islands, start sites, gene bodies and beyond.* Nature Reviews Genetics, 2012. **13**(7): p. 484-492.
23. Moore, L.D., T. Le, and G. Fan, *DNA methylation and its basic function.* Neuropsychopharmacology, 2013. **38**(1): p. 23-38.
24. Goll, M.G., et al., *Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2.* Science, 2006. **311**(5759): p. 395-398.
25. Schaefer, M., et al., *RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage.* Genes & development, 2010. **24**(15): p. 1590-1595.
26. Tuorto, F., et al., *The tRNA methyltransferase Dnmt2 is required for accurate polypeptide synthesis during haematopoiesis.* The EMBO journal, 2015. **34**(18): p. 2350-2362.
27. Gros, C., et al., *DNA methylation inhibitors in cancer: recent and future approaches.* Biochimie, 2012. **94**(11): p. 2280-2296.
28. Baylin, S.B. and P.A. Jones, *A decade of exploring the cancer epigenome—biological and translational implications.* Nature Reviews Cancer, 2011. **11**(10): p. 726-734.
29. Ji, H., et al., *Comprehensive methylome map of lineage commitment from haematopoietic progenitors.* Nature, 2010. **467**(7313): p. 338-342.
30. Bröske, A.-M., et al., *DNA methylation protects hematopoietic stem cell multipotency from myeloerythroid restriction.* Nature genetics, 2009. **41**(11): p. 1207-1215.
31. Trowbridge, J.J., et al., *DNA methyltransferase 1 is essential for and uniquely regulates hematopoietic stem and progenitor cells.* Cell stem cell, 2009. **5**(4): p. 442-449.
32. Jeong, M. and M.A. Goodell, *New answers to old questions from genome-wide maps of DNA methylation in hematopoietic cells.* Experimental hematology, 2014. **42**(8): p. 609-617.
33. Park, D.J., et al., *Characteristics of DNMT3A mutations in acute myeloid leukemia.* Blood research, 2020. **55**(1): p. 17.
34. Russler-Germain, D.A., et al., *The R882H DNMT3A mutation associated with AML dominantly inhibits wild-type DNMT3A by blocking its ability to form active tetramers.* Cancer cell, 2014. **25**(4): p. 442-454.
35. Wang, R., X. Gao, and L. Yu, *The prognostic impact of tet oncogene family member 2 mutations in patients with acute myeloid leukemia: a systematic-review and meta-analysis.* BMC cancer, 2019. **19**(1): p. 1-11.
36. Tulstrup, M., et al., *TET2 mutations are associated with hypermethylation at key regulatory enhancers in normal and malignant hematopoiesis.* Nature Communications, 2021. **12**(1): p. 6061.
37. Beat, A., *Study Offers Insights on Prognostic Significance of IDH Mutations Across Age Groups in AML.* 2020.
38. Wang, Y., et al., *WT1 recruits TET2 to regulate its target gene expression and suppress leukemia cell proliferation.* Molecular cell, 2015. **57**(4): p. 662-673.

39. Rampal, R., et al., *DNA hydroxymethylation profiling reveals that WT1 mutations result in loss of TET2 function in acute myeloid leukemia.* Cell reports, 2014. **9**(5): p. 1841-1855.
40. Figueroa, M.E., et al., *DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia.* Cancer Cell, 2010. **17**(1): p. 13-27.
41. Šorm, F., et al., *5-Azacytidine, a new, highly effective cancerostatic.* Experientia, 1964. **20**: p. 202-203.
42. Rius, M., et al., *Human concentrative nucleoside transporter 1-mediated uptake of 5-azacytidine enhances DNA demethylation.* Molecular cancer therapeutics, 2009. **8**(1): p. 225-231.
43. Ueda, K., M. Hosokawa, and S. Iwakawa, *Cellular uptake of decitabine by equilibrative nucleoside transporters in HCT116 cells.* Biological and Pharmaceutical Bulletin, 2015. **38**(8): p. 1113-1119.
44. Diesch, J., et al., *A clinical-molecular update on azanucleoside-based therapy for the treatment of hematologic cancers.* Clinical epigenetics, 2016. **8**: p. 1-11.
45. Čihák, A., J. Weiss, and H. Pitot, *Characterization of polyribosomes and maturation of ribosomal RNA in hepatoma cells treated with 5-azacytidine.* Cancer Research, 1974. **34**(11): p. 3003-3009.
46. Lu, L.-J.W. and K. Randerath, *Mechanism of 5-azacytidine-induced transfer RNA cytosine-5-methyltransferase deficiency.* Cancer Research, 1980. **40**(8_Part_1): p. 2701-2705.
47. Reichman, M. and S. Penman, *The mechanism of inhibition of protein synthesis by 5-azacytidine in HeLa cells.* Biochimica et Biophysica Acta (BBA)-Nucleic Acids and Protein Synthesis, 1973. **324**(2): p. 282-289.
48. Weiss, J.W. and H.C. Pitot, *Inhibition of ribosomal ribonucleic acid maturation by 5-azacytidine and 8-azaguanine in Novikoff hepatoma cells.* Archives of biochemistry and biophysics, 1974. **160**(1): p. 119-129.
49. Stomper, J., et al., *Hypomethylating agents (HMA) for the treatment of acute myeloid leukemia and myelodysplastic syndromes: Mechanisms of resistance and novel HMA-based therapies.* Leukemia, 2021. **35**(7): p. 1873-1889.
50. Ma, J. and Z. Ge, *Comparison between decitabine and azacitidine for patients with acute myeloid leukemia and higher-risk myelodysplastic syndrome: a systematic review and network meta-analysis.* Frontiers in Pharmacology, 2021: p. 1919.
51. Kantarjian, H.M., et al., *Multicenter, randomized, open-label, phase III trial of decitabine versus patient choice, with physician advice, of either supportive care or low-dose cytarabine for the treatment of older patients with newly diagnosed acute myeloid leukemia.* J Clin Oncol, 2012. **30**(21): p. 2670-7.
52. Fenaux, P., et al., *Efficacy of azacitidine compared with that of conventional care regimens in the treatment of higher-risk myelodysplastic syndromes: a randomised, open-label, phase III study.* Lancet Oncol, 2009. **10**(3): p. 223-32.
53. Malik, P. and A.F. Cashen, *Decitabine in the treatment of acute myeloid leukemia in elderly patients.* Cancer Manag Res, 2014. **6**: p. 53-61.
54. Wells, R., et al., *Optimizing outcomes with azacitidine: recommendations from Canadian centres of excellence.* Current Oncology, 2014. **21**(1): p. 44-50.
55. Silverman, L.R., et al., *Continued azacitidine therapy beyond time of first response improves quality of response in patients with higher-risk myelodysplastic syndromes.* Cancer, 2011. **117**(12): p. 2697-2702.

56. Shapiro, R.M. and A. Lazo-Langner, *Systematic review of azacitidine regimens in myelodysplastic syndrome and acute myeloid leukemia.* BMC hematology, 2018. **18**: p. 1-9.
57. Fandy, T.E., et al., *Early epigenetic changes and DNA damage do not predict clinical response in an overlapping schedule of 5-azacytidine and entinostat in patients with myeloid malignancies.* Blood, The Journal of the American Society of Hematology, 2009. **114**(13): p. 2764-2773.
58. Shen, L., et al., *DNA methylation predicts survival and response to therapy in patients with myelodysplastic syndromes.* J Clin Oncol, 2010. **28**(4): p. 605-613.
59. Soriano, A.O., et al., *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, The Journal of the American Society of Hematology, 2007. **110**(7): p. 2302-2308.
60. Bejar, R., et al., *TET2 mutations predict response to hypomethylating agents in myelodysplastic syndrome patients.* Blood, The Journal of the American Society of Hematology, 2014. **124**(17): p. 2705-2712.
61. Itzykson, R., et al., *Impact of TET2 mutations on response rate to azacitidine in myelodysplastic syndromes and low blast count acute myeloid leukemias.* Leukemia, 2011. **25**(7): p. 1147-1152.
62. Jung, S.-H., et al., *Somatic mutations predict outcomes of hypomethylating therapy in patients with myelodysplastic syndrome.* Oncotarget, 2016. **7**(34): p. 55264.
63. Traina, F., et al., *Impact of molecular mutations on treatment response to DNMT inhibitors in myelodysplasia and related neoplasms.* Leukemia, 2014. **28**(1): p. 78-87.
64. Kuendgen, A., et al., *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.* Oncotarget, 2018. **9**(45): p. 27882.
65. Coombs, C.C., et al., *Mutational correlates of response to hypomethylating agent therapy in acute myeloid leukemia.* Haematologica, 2016. **101**(11): p. e457.
66. Metzeler, K.H., et al., *DNMT3A mutations and response to the hypomethylating agent decitabine in acute myeloid leukemia.* Leukemia, 2012. **26**(5): p. 1106-1107.
67. Hiller, J.K., et al., *Evaluating the impact of genetic and epigenetic aberrations on survival and response in acute myeloid leukemia patients receiving epigenetic therapy.* Annals of hematology, 2017. **96**: p. 559-565.
68. Tobiasson, M., et al., *Mutations in histone modulators are associated with prolonged survival during azacitidine therapy.* Oncotarget, 2016. **7**(16): p. 22103.
69. Xu, J., et al., *Subtype-specific 3D genome alteration in acute myeloid leukaemia.* Nature, 2022: p. 1-12.
70. Cheng, J.X., et al., *RNA cytosine methylation and methyltransferases mediate chromatin organization and 5-azacytidine response and resistance in leukaemia.* Nat Commun, 2018. **9**(1): p. 1163.
71. Ali, A., et al., *Granulomonocytic progenitors are key target cells of azacytidine in higher risk myelodysplastic syndromes and acute myeloid leukemia.* Leukemia, 2018. **32**(8): p. 1856-1860.
72. Unnikrishnan, A., et al., *Integrative Genomics Identifies the Molecular Basis of Resistance to Azacitidine Therapy in Myelodysplastic Syndromes.* Cell Rep, 2017. **20**(3): p. 572-585.

73. Cluzeau, T., et al., *Induction of Autophagic Cell Death Circumvents Azacitidine-Resistance In Myelodysplastic Syndrome-Derived Cell Lines*. 2010, American Society of Hematology.
74. Cluzeau, T., et al., *Azacitidine-resistant SKM1 myeloid cells are defective for AZA-induced mitochondrial apoptosis and autophagy.* Cell Cycle, 2011. **10**(14): p. 2339-2343.
75. Romano, A., et al., *Proteomic analysis reveals autophagy as pro-survival pathway elicited by long-term exposure with 5-azacitidine in high-risk myelodysplasia.* Frontiers in Pharmacology, 2017. **8**: p. 204.
76. Qin, T., et al., *Mechanisms of resistance to 5-aza-2'-deoxycytidine in human cancer cell lines.* Blood, The Journal of the American Society of Hematology, 2009. **113**(3): p. 659-667.
77. Wu, P., et al., *The hENT1 and DCK genes underlie the decitabine response in patients with myelodysplastic syndrome.* Leukemia Research, 2015. **39**(2): p. 216-220.
78. Unnikrishnan, A., et al., *AZA-MS: a novel multiparameter mass spectrometry method to determine the intracellular dynamics of azacitidine therapy in vivo.* Leukemia, 2018. **32**(4): p. 900-910.
79. Fenaux, P., et al., *S879 Results of ASTRAL-1 study, a phase 3 randomized trial of guadecitabine (G) vs treatment choice (TC) in treatment naïve acute myeloid leukemia (TN-AML) not eligible for intensive chemotherapy (IC).* HemaSphere, 2019. **3**(S1): p. 394-395.
80. Wei, A.H., et al., *Oral azacitidine maintenance therapy for acute myeloid leukemia in first remission.* New England Journal of Medicine, 2020. **383**(26): p. 2526-2537.
81. Craddock, C.F., et al., *Outcome of Azacitidine Therapy in Acute Myeloid Leukemia Is not Improved by Concurrent Vorinostat Therapy but Is Predicted by a Diagnostic Molecular SignatureCombined Vorinostat and Azacitidine in AML.* Clinical Cancer Research, 2017. **23**(21): p. 6430-6440.
82. Issa, J.P., et al., *Results of phase 2 randomized study of low-dose decitabine with or without valproic acid in patients with myelodysplastic syndrome and acute myelogenous leukemia.* Cancer, 2015. **121**(4): p. 556-561.
83. Lübbert, M., et al., *Valproate and retinoic acid in combination with decitabine in elderly nonfit patients with acute myeloid leukemia: results of a multicenter, randomized, 2× 2, phase II trial.* Journal of Clinical Oncology, 2020. **38**(3): p. 257-270.
84. Prebet, T., et al., *Prolonged administration of azacitidine with or without entinostat for myelodysplastic syndrome and acute myeloid leukemia with myelodysplasia-related changes: results of the US Leukemia Intergroup trial E1905.* Journal of clinical oncology, 2014. **32**(12): p. 1242.
85. DiNardo, C.D., et al., *Azacitidine and venetoclax in previously untreated acute myeloid leukemia.* New England Journal of Medicine, 2020. **383**(7): p. 617-629.
86. Juergens, R.A., et al., *Combination Epigenetic Therapy Has Efficacy in Patients with Refractory Advanced Non–Small Cell Lung CancerCombinatorial Epigenetic Therapy for Lung Cancer.* Cancer discovery, 2011. **1**(7): p. 598-607.
87. Wrangle, J., et al., *Alterations of immune response of non-small cell lung cancer with azacytidine.* Oncotarget, 2013. **4**(11): p. 2067.
88. Chiappinelli, K.B., et al., *Inhibiting DNA Methylation Causes an Interferon Response in Cancer via dsRNA Including Endogenous Retroviruses.* Cell, 2015. **162**(5): p. 974-86.

89. Wang, L., et al., *Decitabine Enhances Lymphocyte Migration and Function and Synergizes with CTLA-4 Blockade in a Murine Ovarian Cancer ModelDecitabine Synergizes with CTLA-4 Blockade.* Cancer immunology research, 2015. **3**(9): p. 1030-1041.
90. Luo, N., et al., *DNA methyltransferase inhibition upregulates MHC-I to potentiate cytotoxic T lymphocyte responses in breast cancer.* Nat Commun, 2018. **9**(1): p. 248.
91. Yau, H.L., et al., *DNA hypomethylating agents increase activation and cytolytic activity of CD8+ T cells.* Molecular Cell, 2021. **81**(7): p. 1469-1483. e8.
92. Gang, A.O., et al., *5-Azacytidine treatment sensitizes tumor cells to T-cell mediated cytotoxicity and modulates NK cells in patients with myeloid malignancies.* Blood Cancer J, 2014. **4**: p. e197.
93. Luo, N., et al., *DNA methyltransferase inhibition upregulates MHC-I to potentiate cytotoxic T lymphocyte responses in breast cancer.* Nature communications, 2018. **9**(1): p. 1-11.
94. Srivastava, P., et al., *Induction of cancer testis antigen expression in circulating acute myeloid leukemia blasts following hypomethylating agent monotherapy.* Oncotarget, 2016. **7**(11): p. 12840.
95. Karpf, A.R., et al., *Limited gene activation in tumor and normal epithelial cells treated with the DNA methyltransferase inhibitor 5-aza-2'-deoxycytidine.* Molecular pharmacology, 2004. **65**(1): p. 18-27.
96. Karpf, A.R., et al., *Inhibition of DNA methyltransferase stimulates the expression of signal transducer and activator of transcription 1, 2, and 3 genes in colon tumor cells.* Proceedings of the National Academy of Sciences, 1999. **96**(24): p. 14007-14012.
97. Li, H., et al., *Immune regulation by low doses of the DNA methyltransferase inhibitor 5-azacitidine in common human epithelial cancers.* Oncotarget, 2014. **5**(3): p. 587.
98. Héninger, E., T.E. Krueger, and J.M. Lang, *Augmenting antitumor immune responses with epigenetic modifying agents.* Frontiers in immunology, 2015. **6**: p. 29.
99. Roulois, D., et al., *DNA-Demethylating Agents Target Colorectal Cancer Cells by Inducing Viral Mimicry by Endogenous Transcripts.* Cell, 2015. **162**(5): p. 961-73.
100. Schulz, W.A., C. Steinhoff, and A.R. Florl, *Methylation of endogenous human retroelements in health and disease.* Curr Top Microbiol Immunol, 2006. **310**: p. 211-50.
101. Dewannieux, M. and T. Heidmann, *Endogenous retroviruses: acquisition, amplification and taming of genome invaders.* Current opinion in virology, 2013. **3**(6): p. 646-656.
102. McClintock, B. *Controlling elements and the gene*. in *Cold Spring Harbor symposia on quantitative biology*. 1956. Cold Spring Harbor Laboratory Press.
103. Boeke, J.D., et al., *Ty elements transpose through an RNA intermediate.* Cell, 1985. **40**(3): p. 491-500.
104. Grabundzija, I., et al., *A Helitron transposon reconstructed from bats reveals a novel mechanism of genome shuffling in eukaryotes.* Nature communications, 2016. **7**(1): p. 10716.
105. Vos, J.C., I. De Baere, and R. Plasterk, *Transposase is the only nematode protein required for in vitro transposition of Tc1.* Genes & development, 1996. **10**(6): p. 755-761.
106. Ray, D.A., et al., *Multiple waves of recent DNA transposon activity in the bat, Myotis lucifugus.* Genome Research, 2008. **18**(5): p. 717-728.
107. Vabret, N., N. Bhardwaj, and B.D. Greenbaum, *Sequence-specific sensing of nucleic acids.* Trends in immunology, 2017. **38**(1): p. 53-65.

108. Casola, C., D. Hucks, and C. Feschotte, *Convergent domestication of pogo-like transposases into centromere-binding proteins in fission yeast and mammals.* Molecular biology and evolution, 2008. **25**(1): p. 29-41.
109. Lavialle, C., et al., *Paleovirology of 'syncytins', retroviral env genes exapted for a role in placentation.* Philosophical Transactions of the Royal Society B: Biological Sciences, 2013. **368**(1626): p. 20120507.
110. Ishak, C.A., M. Classon, and D.D. De Carvalho, *Deregulation of retroelements as an emerging therapeutic opportunity in cancer.* Trends in cancer, 2018. **4**(8): p. 583-597.
111. Cullen, H. and A.J. Schorn, *Endogenous retroviruses walk a fine line between priming and silencing.* Viruses, 2020. **12**(8): p. 792.
112. Vitte, C. and O. Panaud, *Formation of solo-LTRs through unequal homologous recombination counterbalances amplifications of LTR retrotransposons in rice Oryza sativa L.* Molecular Biology and Evolution, 2003. **20**(4): p. 528-540.
113. Pelissier, T., et al., *Synthesis and processing of tRNA-related SINE transcripts in Arabidopsis thaliana.* Nucleic acids research, 2004. **32**(13): p. 3957-3966.
114. Kramerov, D.A. and N.S. Vassetzky, *SINEs.* Wiley Interdisciplinary Reviews: RNA, 2011. **2**(6): p. 772-786.
115. Zhang, X.-O., H. Pratt, and Z. Weng, *Investigating the potential roles of SINEs in the human genome.* Annual Review of Genomics and Human Genetics, 2021. **22**: p. 199-218.
116. Sultana, T., et al., *The landscape of L1 retrotransposons in the human genome is shaped by pre-insertion sequence biases and post-insertion selection.* Molecular cell, 2019. **74**(3): p. 555-570. e7.
117. Yoder, J.A., C.P. Walsh, and T.H. Bestor, *Cytosine methylation and the ecology of intragenomic parasites.* Trends in genetics, 1997. **13**(8): p. 335-340.
118. Bestor, T., *DNA methylation: evolution of a bacterial immune function into a regulator of gene expression and genome structure in higher eukaryotes.* Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 1990. **326**(1235): p. 179-187.
119. Zhou, W., et al., *DNA methylation enables transposable element-driven genome expansion.* Proceedings of the National Academy of Sciences, 2020. **117**(32): p. 19359-19366.
120. Coulondre, C., et al., *Molecular basis of base substitution hotspots in Escherichia coli.* Nature, 1978. **274**(5673): p. 775-780.
121. Zhang, X. and C. Mathews, *Effect of DNA cytosine methylation upon deamination-induced mutagenesis in a natural target sequence in duplex DNA.* Journal of Biological Chemistry, 1994. **269**(10): p. 7066-7069.
122. Duncan, B.K. and J.H. Miller, *Mutagenic deamination of cytosine residues in DNA.* Nature, 1980. **287**(5782): p. 560-561.
123. Cooper, D.N. and M. Krawczak, *Cytosine methylation and the fate of CpG dinucleotides in vertebrate genomes.* Human genetics, 1989. **83**: p. 181-188.
124. Ohtani, H., et al., *Switching roles for DNA and histone methylation depend on evolutionary ages of human endogenous retroviruses.* Genome research, 2018. **28**(8): p. 1147-1157.
125. Ecco, G., et al., *Transposable elements and their KRAB-ZFP controllers regulate gene expression in adult tissues.* Developmental cell, 2016. **36**(6): p. 611-623.
126. Jacobs, F.M., et al., *An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons.* Nature, 2014. **516**(7530): p. 242-245.
127. Imbeault, M., P.-Y. Helleboid, and D. Trono, *KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks.* Nature, 2017. **543**(7646): p. 550-554.

128. Guo, H., et al., *Autophagy supports genomic stability by degrading retrotransposon RNA.* Nature communications, 2014. **5**(1): p. 1-11.
129. Takahashi, T. and K. Ui-Tei, *Mutual regulation of RNA silencing and the IFN response as an antiviral defense system in mammalian cells.* International Journal of Molecular Sciences, 2020. **21**(4): p. 1348.
130. Zamanian-Daryoush, M., et al., *NF-κB activation by double-stranded-RNA-activated protein kinase (PKR) is mediated through NF-κB-inducing kinase and IκB kinase.* Molecular and cellular biology, 2000. **20**(4): p. 1278-1290.
131. Banerjee, S., et al., *OAS-RNase L innate immune pathway mediates the cytotoxicity of a DNA-demethylating drug.* Proceedings of the National Academy of Sciences, 2019. **116**(11): p. 5071-5076.
132. Lum, J.J., R.J. DeBerardinis, and C.B. Thompson, *Autophagy in metazoans: cell survival in the land of plenty.* Nature reviews Molecular cell biology, 2005. **6**(6): p. 439-448.
133. Shintani, T. and D.J. Klionsky, *Autophagy in health and disease: a double-edged sword.* Science, 2004. **306**(5698): p. 990-995.
134. Kaur, J. and J. Debnath, *Autophagy at the crossroads of catabolism and anabolism.* Nature reviews Molecular cell biology, 2015. **16**(8): p. 461-472.
135. Mijaljica, D., M. Prescott, and R.J. Devenish, *Microautophagy in mammalian cells: revisiting a 40-year-old conundrum.* Autophagy, 2011. **7**(7): p. 673-682.
136. Dice, J.F., *Peptide sequences that target cytosolic proteins for lysosomal proteolysis.* Trends in biochemical sciences, 1990. **15**(8): p. 305-309.
137. Chiang, H. and J.F. Dice, *Peptide sequences that target proteins for enhanced degradation during serum withdrawal.* Journal of Biological Chemistry, 1988. **263**(14): p. 6797-6805.
138. Dice, J.F., *Chaperone-mediated autophagy.* Autophagy, 2007. **3**(4): p. 295-299.
139. Arias, E. and A.M. Cuervo, *Chaperone-mediated autophagy in protein quality control.* Current opinion in cell biology, 2011. **23**(2): p. 184-189.
140. Istrail, S., et al., *Comparative immunopeptidomics of humans and their pathogens.* Proceedings of the National Academy of Sciences, 2004. **101**(36): p. 13268-13272.
141. Pearson, H., et al., *MHC class I-associated peptides derive from selective regions of the human genome.* J Clin Invest, 2016. **126**(12): p. 4690-4701.
142. Yewdell, J.W., E. Reits, and J. Neefjes, *Making sense of mass destruction: quantitating MHC class I antigen presentation.* Nature Reviews Immunology, 2003. **3**(12): p. 952-961.
143. Croft, N.P., et al., *Kinetics of antigen expression and epitope presentation during virus infection.* PLoS pathogens, 2013. **9**(1): p. e1003129.
144. Wu, T., et al., *Quantification of epitope abundance reveals the effect of direct and cross-presentation on influenza CTL responses.* Nature communications, 2019. **10**(1): p. 2846.
145. Schubert, U., et al., *Rapid degradation of a large fraction of newly synthesized proteins by proteasomes.* Nature, 2000. **404**(6779): p. 770-774.
146. Trentini, D.B., et al., *Role for ribosome-associated quality control in sampling proteins for MHC class I-mediated antigen presentation.* Proceedings of the National Academy of Sciences, 2020. **117**(8): p. 4099-4108.
147. Kim, S., et al., *Evaluation of Immunoproteasome-Specific Proteolytic Activity Using Fluorogenic Peptide Substrates.* Immune Network, 2022. **22**(3).
148. Gaczynska, M., K.L. Rock, and A.L. Goldberg, *γ-Interferon and expression of MHC genes regulate peptide hydrolysis by proteasomes.* Nature, 1993. **365**(6443): p. 264-267.

149. Mishto, M., et al., *Proteasome isoforms exhibit only quantitative differences in cleavage and epitope generation.* Eur J Immunol, 2014. **44**(12): p. 3508-21.
150. Toes, R., et al., *Discrete cleavage motifs of constitutive and immunoproteasomes revealed by quantitative analysis of cleavage products.* The Journal of experimental medicine, 2001. **194**(1): p. 1-12.
151. de Verteuil, D., et al., *Deletion of immunoproteasome subunits imprints on the transcriptome and has a broad impact on peptides presented by major histocompatibility complex I molecules.* Molecular & Cellular Proteomics, 2010. **9**(9): p. 2034-2047.
152. Hutchinson, S., et al., *A dominant role for the immunoproteasome in CD8+ T cell responses to murine cytomegalovirus.* PloS one, 2011. **6**(2): p. e14646.
153. Tripathi, S.C., et al., *Immunoproteasome deficiency is a feature of non-small cell lung cancer with a mesenchymal phenotype and is associated with a poor outcome.* Proceedings of the National Academy of Sciences, 2016. **113**(11): p. E1555-E1564.
154. Rouette, A., et al., *Expression of immunoproteasome genes is regulated by cell-intrinsic and–extrinsic factors in human cancers.* Scientific reports, 2016. **6**(1): p. 34019.
155. Zanker, D. and W. Chen, *Standard and immunoproteasomes show similar peptide degradation specificities.* European journal of immunology, 2014. **44**(12): p. 3500-3503.
156. Decker, W.K. and A. Safdar, *Bioimmunoadjuvants for the treatment of neoplastic and infectious disease: Coley's legacy revisited.* Cytokine & growth factor reviews, 2009. **20**(4): p. 271-281.
157. Decker, W.K., et al., *Cancer immunotherapy: historical perspective of a clinical revolution and emerging preclinical animal models.* Frontiers in immunology, 2017. **8**: p. 829.
158. Rosenstein, M., et al., *In vitro growth of murine T cells. VI. Accelerated skin graft rejection caused by adoptively transferred cells expanded in T cell growth factor.* Journal of immunology (Baltimore, Md.: 1950), 1981. **127**(2): p. 566-571.
159. Rosenberg, S.A., *A Journey in Science: Immersion in the search for effective cancer immunotherapies.* Molecular Medicine, 2021. **27**(1): p. 63.
160. Klapper, J.A., et al., *High-dose interleukin-2 for the treatment of metastatic renal cell carcinoma: a retrospective analysis of response and survival in patients treated in the surgery branch at the National Cancer Institute between 1986 and 2006.* Cancer, 2008. **113**(2): p. 293-301.
161. Smith, F.O., et al., *Treatment of metastatic melanoma using interleukin-2 alone or in conjunction with vaccines.* Clinical Cancer Research, 2008. **14**(17): p. 5610-5618.
162. Atkins, M.B., et al., *High-dose recombinant interleukin 2 therapy for patients with metastatic melanoma: analysis of 270 patients treated between 1985 and 1993.* Journal of clinical oncology, 1999. **17**(7): p. 2105-2105.
163. Rosenberg, S.A., et al., *Durability of complete responses in patients with metastatic cancer treated with high-dose interleukin-2: identification of the antigens mediating response.* Annals of surgery, 1998. **228**(3): p. 307.
164. Apavaloaei, A., et al., *The origin and immune recognition of tumor-specific antigens.* Cancers, 2020. **12**(9): p. 2607.
165. Laumont, C.M., et al., *Noncoding regions are the main source of targetable tumor-specific antigens.* Sci Transl Med, 2018. **10**(470).
166. Ehx, G., et al., *Atypical acute myeloid leukemia-specific transcripts generate shared and immunogenic MHC class-I-associated epitopes.* Immunity, 2021. **54**(4): p. 737-752.e10.

167. Zhao, Q., et al., *Proteogenomics Uncovers a Vast Repertoire of Shared Tumor-Specific Antigens in Ovarian CancerThe Global Landscape of Ovarian Cancer–Specific Antigens.* Cancer immunology research, 2020. **8**(4): p. 544-555.
168. Cleyle, J., et al., *Immunopeptidomic analyses of colorectal cancers with and without microsatellite instability.* Molecular & Cellular Proteomics, 2022. **21**(5).
169. Ruiz Cuevas, M.V., et al., *BamQuery: a proteogenomic tool for the genome-wide exploration of the immunopeptidome.* bioRxiv, 2022: p. 2022.10.07.510944.
170. Chong, C., et al., *Integrated proteogenomic deep sequencing and analytics accurately identify non-canonical peptides in tumor immunopeptidomes.* Nature communications, 2020. **11**(1): p. 1293.
171. Apavaloaei, A., et al., *Induced pluripotent stem cells display a distinct set of MHC I-associated peptides shared by human cancers.* Cell Reports, 2022. **40**(7): p. 111241.
172. Ehx, G., et al., *Atypical acute myeloid leukemia-specific transcripts generate shared and immunogenic MHC class-I-associated epitopes.* Immunity, 2021. **54**(4): p. 737-752. e10.
173. Scanlan, M.J., A.J. G Simpson, and L.J. Old, *The cancer/testis genes: review, standardization, and commentary.* Cancer immunity, 2004. **4**(1).
174. Ries, R.E., et al., *Evaluation of Cancer Testis Antigens (CTAs) As Immuno-Therapeutic Targets in Pediatric AML*. 2017, American Society of Hematology Washington, DC.
175. Mahdevar, E., et al., *Immunoprotective effect of an in silico designed multiepitope cancer vaccine with BORIS cancer-testis antigen target in a murine mammary carcinoma model.* Scientific Reports, 2021. **11**(1): p. 23121.
176. Almeida, L.G., et al., *CTdatabase: a knowledge-base of high-throughput and curated data on cancer-testis antigens.* Nucleic acids research, 2009. **37**(suppl_1): p. D816-D819.
177. Hubbard, J.M., et al., *Safety and activity of PolyPEPI1018 combined with maintenance therapy in metastatic colorectal cancer: an open-label, multicenter, phase 1b study.* Clinical Cancer Research: an Official Journal of the American Association for Cancer Research, 2022.
178. Kono, K., et al., *Vaccination with multiple peptides derived from novel cancer-testis antigens can induce specific T-cell responses and clinical responses in advanced esophageal cancer.* Cancer science, 2009. **100**(8): p. 1502-1509.
179. Löffler, M.W., et al., *Multi-omics discovery of exome-derived neoantigens in hepatocellular carcinoma.* Genome medicine, 2019. **11**(1): p. 1-16.
180. Capietto, A.-H., S. Jhunjhunwala, and L. Delamarre, *Characterizing neoantigens for personalized cancer immunotherapy.* Current opinion in immunology, 2017. **46**: p. 58-65.
181. Pappireddi, N., L. Martin, and M. Wühr, *A review on quantitative multiplexed proteomics.* ChemBioChem, 2019. **20**(10): p. 1210-1224.
182. Korch, C., et al., *DNA profiling analysis of endometrial and ovarian cell lines reveals misidentification, redundancy and contamination.* Gynecologic oncology, 2012. **127**(1): p. 241-248.
183. Lorsch, J.R., F.S. Collins, and J. Lippincott-Schwartz, *Fixing problems with cell lines.* Science, 2014. **346**(6216): p. 1452-1453.
184. Liu, Y., et al., *Multi-omic measurements of heterogeneity in HeLa cells across laboratories.* Nature biotechnology, 2019. **37**(3): p. 314-322.
185. *NIH Guide Notice NOT-OD-08-017* https://grants.nih.gov/grants/guide/notice-files/not-OD-08-017.html. 2007.

186. Fusenig, N.E., et al., *The need for a worldwide consensus for cell line authentication: experience implementing a mandatory requirement at the International Journal of Cancer.* PLoS biology, 2017. **15**(4): p. e2001438.
187. MacLeod, R.A., et al., *Widespread intraspecies cross-contamination of human tumor cell lines arising at source.* International Journal of Cancer, 1999. **83**(4): p. 555-563.
188. Lacroix, M., *Persistent use of "false" cell lines.* International journal of cancer, 2008. **122**(1): p. 1-4.
189. Capes-Davis, A., et al., *Match criteria for human cell line authentication: where do we draw the line?* International journal of cancer, 2013. **132**(11): p. 2510-2519.
190. Visconti, P., et al., *Short tandem repeat profiling for the authentication of cancer stem-like cells.* International Journal of Cancer, 2021. **148**(6): p. 1489-1498.

2 Chapter 2: Major multilevel molecular divergence between THP-1 cells from different biorepositories

Nandita Noronha[1,†], Grégory Ehx[1, †], Marie-Christine Meunier[3], Jean-Philippe Laverdure[1], Catherine Thériault[1] and Claude Perreault[1,4,*].

[1]Institute for Research in Immunology and Cancer (IRIC), Université de Montréal, Montreal, Quebec H3C 3J7, Canada. [3]HLA Laboratory, Hôpital Maisonneuve-Rosemont, Montreal, Quebec H1T 2M4, Canada. [4]Department of Medicine, Université de Montréal, Montreal, Quebec H3C 3J7, Canada.

[†]these authors participated equally to the present work

[*]Correspondence: claude.perreault@umontreal.ca

Keywords: Acute Myeloid Leukemia, DNA Copy Number Variations, Histocompatibility, Reproducibility of Results, THP-1 cells, Transcriptomics

The manuscript in this chapter is published in *International Journal of Cancer* 2020 2020 Oct 1;147(7):2000-2006.

doi: 10.1002/ijc.32967.

2.1 Author contributions

NN, GE and CT performed the experiments. MCM performed and analyzed the HLA genotyping. NN, GE and JPL performed the bioinformatic analyses. NN, GE and CP designed the study, analyzed the data, interpreted the results and wrote the article. All authors edited the manuscript.

Contributions per figures:

NN: Generated and analysed data for Figure 2.1, Figure 2.2a (bottom panel), 2.2b, 2.2e, Suppl. figures S2.2, S2.3, S2.5.

GE: Generated and analysed data for Figure 2.2a (top panel), 2.2c-d, Suppl. figure S2.4.

MCM: Generated and analysed data for Suppl. figure S2.1.

2.2 Novelty and Impact

THP-1 cells are widely used, mainly for studying myeloid leukemias. We performed multi-omic analyses on THP-1 cells obtained from two major repositories. The two THP-1 cell lines presented numerous genomic, transcriptomic and proteomic discrepancies that have pervasive effects, namely on genes instrumental in leukemogenesis. This indicates that the two THP-1 cell lines are not the same entity and have undergone biologically important genetic drift which can be underestimated by analyses of a limited number of STR.

2.3 Abstract

The THP-1 cell line is broadly used as a model for acute myeloid leukemia (AML) with *MLL* fusion and to study monocyte differentiation and function. We studied THP-1 cells obtained from two major biorepositories. The two cell lines were closely related with a percentage match of short tandem repeat (STR) profiles ranging from 93.75% to 100%, depending on the algorithm used. Nevertheless, we found that the two cell lines presented discordant HLA type, cytogenetic aberrations and AML-related gene expression (including critical targets of *MLL* fusion). These discrepancies resulted mainly from loss of heterozygosity (LOH) involving five chromosomal regions. In view of their aberrant expression of key "leukemia" genes (e.g., *LIN28B*, *MEIS1* and *SPARC*), we argue that one of the THP-1 cell lines may not be a reliable model for studying leukemia. Their defective expression of HLA molecules and abnormal adhesion properties is also a caveat for studies of antigen presentation. In a more general perspective, our findings show that seemingly minor discrepancies in STR profiles among cell lines may be the sign of major genetic drift, of sufficient magnitude to affect the reliability of cell line-based research.

2.4 Introduction

AML is a malignant hematological disorder characterized by the proliferation of non-functional hematopoietic cells. More than 70% of infant AML cases bear a chromosomal translocation involving the *MLL* gene[1], whose fusion product is instrumental in leukemogenesis[2]. Established in 1980 from the blood of a child with AML, THP-1 is one of the most widely used cell lines to study the biology of AML with *MLL* fusion as well as monocyte function[3]. Highly appreciated for its versatility, this cell line has been used in over 10,000 publications.

The scientific community is currently going through a well documented 'crisis' of reproducibility as more than 70% of researchers fail to reproduce another scientist's experiments, and more than half failed to reproduce their own[4]. It is well recognized that cross-contamination and phenotypic drift of cells in culture can generate irreproducible or misleading data[5-9]. Cell lines with microsatellite instability (MSI+), in particular have been shown to develop alterations in microsatellite loci and drift with passages, due to their defective mismatch repair genes[10]. This biological diversity clearly contributes to the reproducibility crisis and is disquieting considering that human cancer cell lines are the workhorse of cancer research. The substantial biological diversity among genetically unstable cancer cell lines within and between labs has been extensively reported[11, 12]. However, little is known about the possible diversity of cell lines between reference biorepositories. Here, we compared THP-1 cells, a cell line considered as relatively stable[13], obtained from two different major biorepositories. We found that they were not the same entity and have undergone genetic drift with major and pervasive functional effects.

2.5 Material and methods

Cell lines

THP-1 cell lines (RRID: CVCL_0006) were freshly purchased from the American Type Culture Collection (ATCC) (TIB-202, ATCC, Manassas, US-VA) and the Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ) (ACC16, DSMZ, Braunschweig, Germany) for the purpose of the present study. Cells were maintained in RPMI 1640 (Gibco, NY-US, 11875-093) containing L-glutamine and supplemented with 20% heat inactivated fetal bovine serum (FBS, Gibco 12483) and 1% penicillin-streptomycin (10,000 U/mL, Gibco 15140-122). All cultures were free of mycoplasma contamination, as verified by absence of RNA-Seq reads (see in next sections) mapping to four different mycoplasma genomes (analysis made as reported previously[14]). Cell line authentication was performed by STR profiling. DNA was extracted from both cell lines and submitted for STR profiling to ATCC's cell authentication service (July 2019). STR profiling was performed based on 13 Combined DNA Index System (CODIS: D18S51, D21S11, TH01, D3S1358, FGA, TPOX, D8S1179, vWA, CSF1PO, D16S539, D7S820, D13S317, D5S818) loci plus amelogenin (for gender determination), Penta E, Penta D, D2S1338 and D19S433 (results reported in supplemental data and other data can be made available upon reasonable request).

HLA genotyping of THP-1 cell lines

Cells were genotyped at 6 HLA genes: HLA-A, HLA-B, HLA-C and HLA-DRB1, HLA-DQB1, HLA-DPB1. Typing was performed by next generation sequencing using NGSgo® commercial kit (GenDx, Netherlands) on the MiSeq platform (Illumina, USA). Paired-end

sequences were analyzed using NGSengine® software v2.12.0 (GenDx) and the IPD-IMGT/HLA database release 3.34.0.

RNA sequencing

After thawing, THP-1 cells (ATCC and DSMZ) were expanded for one week to recover from cryopreservation. Three replicate expansions were performed for sample collection with the following sequences: seeding at 0.2 million/mL, 48h of expansion, dilution 1:2 with fresh medium, 24h of expansion and collection while in the log phase (final cell density = 0.5 million /ml). Replicates 1 and 3 were collected one week after thawing and replicate 2 was collected two weeks after thawing.

Total RNA was extracted using TRIzol® (Life Technologies, Carlsbad, Canada) according to the manufacturer's recommendations. RNA samples were purified using QIAGEN Mini RNeasy kit following manufacturer's instructions. Presence of contaminants were assessed with nanodrop using 260/280 and 260/230 ratios. RNA quantification was performed using QuBit (ABI) and 1µg of total RNA was used for library preparation. Sample quality was assessed with Bioanalyzer Nano (Agilent Technologies) and all samples had RNA Integrity Numbers above 8.8. Transcriptomic libraries were prepared with the KAPA mRNA HyperPrep Kit (KAPA, Cat no. KR1352). Libraries were quantified with the QuBit and BioAnalyzer (average size ~380 base-pairs). All libraries were diluted to 10 nM and normalized by quantitative PCR using the KAPA library quantification kit (KAPA; Cat no. KK4973). Libraries were pooled to equimolar concentrations and sequencing was performed with the Illumina Nextseq500 on half a flow cell of the Nextseq High Output v2 (75 cycles) kit using 4 nM of the pooled libraries. Around 40 M single-end reads passing filter were generated per sample. Library preparation and sequencing were performed at the genomics platform of Institute for Research in Immunology and Cancer.

Transcriptomic analysis

Adapters were trimmed using trimmomatic 0.35 and transcript quantification was performed using Kallisto v0.43.0 with the --single -l 300 -s 50 --rf-stranded (GRCh38.88). Analysis of differentially expressed genes (DEGs) was done in R3.5.1. Raw read counts have been converted to counts per million (cpm), normalized relative to the library size and lowly expressed genes were filtered out by keeping genes with cpm >1 in at least 2 samples using edgeR and limma. This was followed by voom transformations and linear modelling using limma's lmfit. Finally, moderated t-statistics were computed with eBayes. Genes with p-values < 0.05, fold change (FC) > 2 and false discovery rate < 0.01 were considered significantly differentially expressed. Gene ontology and biological pathways annotations were made with DAVIDv6.8 (https://david.ncifcrf.gov). Functional annotations with p-value < 0.05 were considered significant. Gene set enrichment analysis (GSEA) was performed with fgsea package in R[15]. A pre-ranked gene list was generated by ranking expressed genes obtained from limma-voom on the moderated t-statistics. Hallmark and MLL-fusion target gene sets were obtained from MSigDB database and gene sets for chromosomal positions were generated based on genes expressed in hotspots of transcriptomic changes (HTCs, see next paragraph). Heatmap of AML-related genes was generated by using aheatmap function in R to depict median z-scores of kallisto transcript per million values of genes among the top 300 in AML-related GeneCard list (see supplemental data) or having FC > 8.

Integrative chromosome plots

Start sites of each gene were retrieved from BioMart (https://useast.ensembl.org/biomart) and were used to plot the log2(FC) of DEGs, obtained from limma-voom analyses, along each chromosome. The normalized average log2(FC) was computed based on log2(FC) from the entire

list of expressed genes generated by limma-voom analysis. These genes were sorted according to their genomic position and each chromosome was sliced into 25 equal regions. For each region, the average log2(FC) has been computed. Each average has then been multiplied by the number of upregulated genes (if average > 0) or downregulated genes (if average < 0) and finally divided by the total number of genes of the region. Normalized average log2(FC) were plotted at the center position of each of the 25 regions. Based on the empirical observation that the normalized average log2(FC) was higher than 0.45 at the LOH position in p arm of chromosome 6, HTC regions were established when more than two (sub 25) contiguous regions had normalized average log2(FC) above 0.45 or below -0.45. The positions of LOH and allelic ratio aberrations were retrieved from the Deletions and table2 outputs of eSNP-karyotyping. Plots were built with Graphpad v7.00.

eSNP-karyotyping

Detection of chromosomal aberrations using RNA-seq data was performed using eSNP karyotyping as previously described[16].

MHC-I expression analysis

Quantification of MHC class I, HLA-A, B, C surface expression was performed using the QIFIKIT bead-based flow cytometric assay (Dako, Agilent Technologies, Santa-Clara, US-CA, K0078) according to the manufacturer's recommendations. Briefly, triplicates of THP-1 ATCC and DSMZ samples were incubated first with an Fc-receptor blocking reagent (anti-CD32; Becton Dickinson (BD), Bedford, US-MA, 552930) for 10 minutes at room temperature, followed by incubation with either with HLA-ABC-FITC (Invitrogen, Carlsbad, Canada, 14-9983-82) or Mouse IgG2a κ Isotype Control (BD 553454) for 30 min at 4°C. Secondary staining was performed using FITC-conjugated antibody provided in the kit. Analysis was made by flow cytometry (BD

FACS Canto II) and number of HLA molecules was determined by interpolation on the standard curve of the bead populations.

For quantification of HLA-A surface expression, triplicates of THP-1 ATCC and DSMZ samples were incubated with anti-CD32 and then either with anti-HLA-A2-PE (BD 558570) or isotype IgG2b,k-PE (BD 559529) for 20 min at 4°C. MFI of samples were analyzed using flow cytometry and data were analyzed with FlowJo software v10.5.3 (Tree Star Inc., Ashland, US-OR). Analyses were made with Graphpad v7.00.

Morphological assessment of ATCC and DSMZ cells

Cells were seeded at 0.2 million cells per ml in 6 well plates and imaged 48 hours after seeding using Leica DMIRB microscope and QCapture software.

Data availability

Raw RNA-Seq data are openly available through Gene Expression Omnibus (GEO GSE130985). The datasets supporting the conclusions of this article are included within this article and its additional files.

2.6 Results and discussion

THP-1 cells, freshly purchased from ATCC or DSMZ, were submitted for STR profiling to the ATCC cell line authentication service. STR profiling, considered gold standard for cell line authentication testing, is based on examining the varying number of polymorphic repeats and comparing them with size standards [17]. The percentage match between two cell lines can be influenced by the algorithm used [8]. In the present case, percentage match was calculated based

on 8 core STR loci plus amelogenin (Table S2.1). ATCC authentication service used the Masters algorithm (= number of shared alleles between query sample and ATCC database profile / total number of alleles in the ATCC database profile) which yielded a 100% match [(15 / 15) × 100 = 100%]. However, the Tanabe algorithm [= (2 × number of shared alleles) / (number of alleles in sample#1 and number of alleles in sample#2)] yielded a 93.75% match: [(2 × 15) / (15 + 17)] × 100 = 93.75%. The discrepancy highlighted by the Tanabe algorithm is due to the fact that the THP-1 cell line from ATCC shows loss of two STR alleles (D13S317 and vWA loci). In accordance with this, the ATCC and DSMZ reference STR profiles of the THP-1 cell line found in the Cellosaurus database [18] show slight differences. Hence, THP-1 cells from ATCC and DSMZ are two different cell lines, one of them presenting at least a minor genetic drift. Additionally, based on STR profiling data from Cellosaurus database, we found that of 12 THP-1 sources reported, 6 had exact STR match with the ATCC profile observed here while 2 had an exact match with the DSMZ profile (Table S2.2). In total, we identified at least 4 variants of the THP-1 cell line based on the profiling data (Table S2.2). In our case, since THP-1 cells are microsatellite stable, we observed that modified STR profile was exclusively due to loss of STR alleles; in contrast to MSI+ cell lines in which alterations in STR profiles are due to both occurrence and loss of new and existing alleles, respectively [19]. We next sought to investigate the extent of the divergence between the two ATCC and DSMZ cell lines as well as its possible biological consequences. To address this, we first examined the HLA type of each cell line using both next-generation DNA sequencing and PCR-SSP. While the HLA type of ATCC cells was identical to the one found in a previous report (HLA-A*02:01/02:01 and HLA-B*15:11/15:11) [20], DSMZ cells were heterozygous for both HLA-A and -B alleles (HLA-A*02:01/24:02 and HLA-B*15:11/35:01) (Table S2.3 and Figure S2.1). To confirm this observation, we performed

RNA-sequencing of both cell lines and determined their HLA type with the Optitype software [21]. We also performed this analysis on publicly available RNA-Seq data on DSMZ THP-1 cells [22]. Concordant results were obtained with all methods: at both *HLA-A* and *HLA-B* loci THP-1 cells from ATCC had a single allele whereas DSMZ THP-1 cells had two. We concluded that ATCC cells have undergone LOH before integration or during maintenance in this repository. Interestingly, in the original study describing the establishment of THP-1, the HLA type was based on serological typing and reported as HLA-A2, A9, B5 (nomenclature of now-obsolete HLA typing system) [3]. The apparent discrepancy between this HLA serotyping and the HLA genotype of DSMZ THP-1 can be explained by the lower resolution of serological typing and HLA antisera cross-reactivity [23]. Indeed, HLA-B5 antisera has been shown to be cross-reactive to HLA-B35 and HLA-B15 (HLA-B alleles identified in DSMZ THP-1)[20, 24]. Moreover, based on DNA sequencing and HLA-A9 antisera reactivity, A*24:02 is now classified under HLA-A9 antigen family (along with HLA A*23:01 and A*24:03) [25]. Finally, A*24:02 being the most common *HLA-A* allele in Japan, its presence in DSMZ cells is consistent with the reported ethnic origin of THP-1 cells [3]. Altogether, these observations suggest that DSMZ THP-1 cells are more similar to the original cells described in 1980 than ATCC's.

Using eSNP-karyotyping [16], we confirmed LOH of the p arm of chromosome 6 (location of the HLA complex) of ATCC cells (Figure 2.1A, top panel), consistent with a reported karyotypic analysis of ATCC THP-1 [26]. Unexpectedly, our eSNP-karyotyping further provided evidence of LOH specific to ATCC cells in chromosomes 10p, 12p, 13q and 20p. Analysis of allelic ratios (major/minor SNPs), which highlights uneven chromosome copy numbers (Figure S2.2), showed similar patterns in cell lines for most chromosomes, with major differences present in chromosomes 2p, 6p, 9p and 11 (Figure 2.1A, bottom panel). We next performed a differential

gene expression analysis to evaluate the potential impact of these genomic discrepancies. This revealed that the two cell lines had dramatically different transcriptomes as evidenced by principal component analysis (PCA) and volcano plot (2627 differentially expressed genes (DEGs)) (Figure S2.3A and Table S2.4). These alterations likely have pervasive impact on numerous biological processes, as evidenced by gene ontology (GO) term analysis (Figure 2.1B and Table S2.5-2.6). They also modify cell phenotype since DSMZ cells expressed 3-4 fold more MHC-I molecules (either total HLA or HLA-A*02, an allele shared by both cell lines) whereas ATCC cells formed more numerous aggregates, in agreement with their overexpression of genes linked to cell adhesion GO terms (Figure 2.1C and S2.3B).

To establish the causative relationship between the cytogenetic and transcriptomic alterations, we combined our eSNP and DEG data for each chromosome (Figure 2.2A (top panel), Figure S2.4 and Tables S2.7-8). This revealed that in specific chromosomes (8, 9, 10, 12 and X), more than 20% of expressed genes were DEGs (respectively 40.2%, 24.7%, 50.1%, 23.1% and 37.3%). To evaluate the connection between genomic location and gene expression, we computed a normalized average log2(Fold Change (FC)) (see supplementary methods) along each chromosome. This allowed us to identify several HTCs between both cell lines. Strikingly, the positions of these HTCs correlated well with those of LOH or allelic ratio aberrations and their normalized average log2(FC) was either in agreement with eSNP analysis (e.g. the specific LOH in chr6p and chr12p of ATCC cells resulted in overexpression of these regions by DSMZ cells) or provided further insights into it (e.g. copy number of chromosomes 8 and 10 could be higher in ATCC and DSMZ cells, respectively). Finally, we collected the list of genes located in the HTCs (chr1:q24.3-q42.13, chr6:p25.3-p22.1, chr8, chr9:q12-q21.31, chr10, chr12:p13.33-p12.1, chr20:p13-p11.21 and chrX) and compared their expression using the GSEA tool (Figure 2.2A

(bottom panel), Figure S2.5 and Table S2.9). As controls, we examined various biological processes with the HALLMARK genes sets, which were constructed independently of gene chromosomal location [27] in contrast to our HTCs (Figure 2. 2B). This showed that every HTC gene set was significantly differentially expressed, in contrast with controls. Altogether, these results show that transcriptomic and phenotypic differences are closely linked to differential chromosomal aberrations between both cell lines.

Finally, we tested whether inter-cell line divergences could impact the study of genes typically associated with AML. We extracted the full list of AML-related genes from the GeneCards database and queried each one of them in our list of DEGs. This showed that 378/2646 AML-related genes were among the DEGs, 84 being either differentially expressed more than 8-fold or among the top 300 genes of the list (Figure 2. 2C and Table S2.10). Among the four DEGs with the highest log2(FC) (>7) were *LIN28B*, *MEIS1* and *SPARC*, whose leukemic function has been documented in THP-1 cells [28-30]. Importantly, *MEIS1* is systematically overexpressed in AMLs harboring *MLL* fusions, where its expression is promoted by the fusion proteins and is implicated in leukemogenesis [31]. We therefore examined the expression of *MLL* fusion junction in both cell lines and found no difference of expression (Figure 2. 2D). GSEA of *MLL* fusion targets [32] were also not significant (Figure 2. 2E (top two panels)). However, a gene set of *MLL* fusion targets established based on THP-1 cells [28] was significantly overexpressed in ATCC cells, showing that results obtained by studies of *MLL* fusion in THP-1 could differ based on their biorepository of origin (Figure 2. 2E (bottom panel) and Table S2.11).

In summary, we demonstrate that THP-1 cells obtained from two different well-established biorepositories are in fact THP-1 variants that have undergone genetic drift and present important molecular and phenotypic differences. We demonstrate that these alterations can have deleterious

effects on the reproducibility and conclusions of studies using this cell line as model. The genetic heterogeneity and the unstable nature of tumor samples and cell lines, and their influence on tumor evolution and phenotypic variability, have been well reported in the genetically unstable HeLa and MCF-7 cells [11, 12]. Specifically, MCF-7 cells from the same repository cultured simultaneously by two different labs quickly showed genetic differences and phenotypic variability [33]. In the present study, although THP-1 cells are considered as genetically stable, we found that freshly acquired cells from two different well-established cell banks (which are commonly assumed to provide rigorously identical cell lines across the world) are not the same entity due to genetic drift. This emphasizes the crucial necessity of clearly naming the source of each cell line used in research articles. Finally, our data also demonstrate that seemingly minor discrepancies in standard STR profiles (8 loci and amelogenin) should alert to the need of more in-depth evaluation because they may be the sign of extensive genetic drift with dramatic functional consequences. The primary purpose of STR testing in research was to evaluate cell line cross-contaminations. For this purpose, testing 8-16 STR loci is sufficient. However, this is much less effective for detection of the genetic drift in a given cell line. Increasing the number of STR loci tested (from 8-16 to 80-160 loci [34, 35]) can substantially improve the accuracy of genetic drift (and LOH in particular) detection [8]. Evaluation of the functional importance of genetic drift may then be achieved by functional assays and RNA-sequencing.

2.7 Acknowledgments

We thank the following members of IRIC core facilities for technical assistance: S. Boissel from the genomic platform and P. Gendron from the bioinformatic platform. This work was supported by The Leukemia & Lymphoma Society of Canada.

2.8 Conflict of interest

The authors declare that they have no conflict of interest.

2.9 Figures

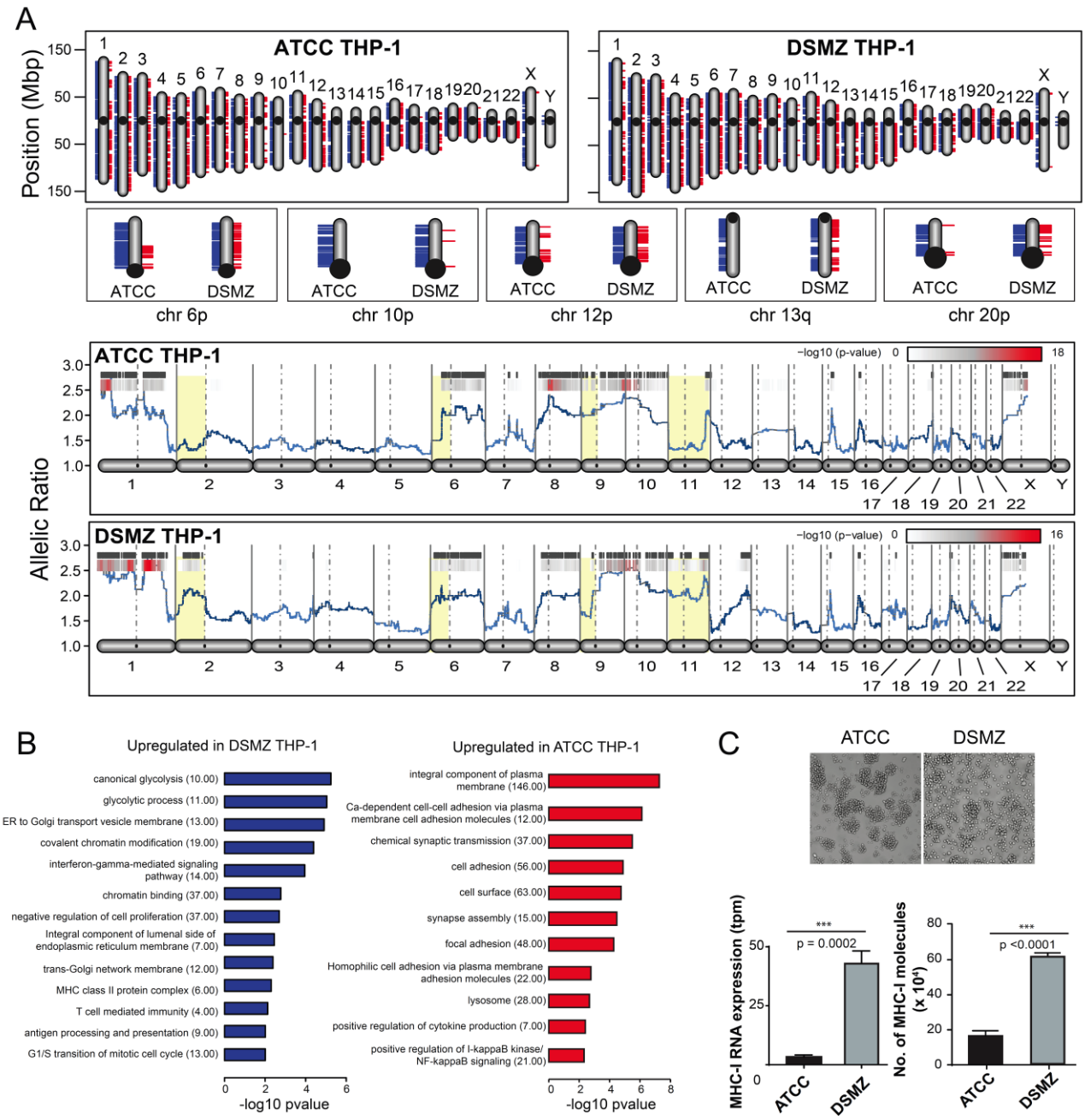

Figure 2.1 ATCC and DSMZ THP-1 cells show different cytogenetic aberrations and transcriptomic profiles

a. Top panel: LOH analysis by eSNP-karyotyping. Blue=homozygous and red=heterozygous SNPs. Bottom panel: Allelic ratio analysis (differences between ATCC and DSMZ cells are highlighted in light yellow). **b.** DAVID annotation of upregulated DEGs in DSMZ and ATCC cells. **c.** Top panel: Light microscopy images after 48h of seeding at identical cell densities. Bottom panel: MHC-I expression at RNA level from limma-voom analysis (left) and protein level by flow cytometry (right: total HLA-ABC molecule number assessed by bead-based assay). Statistical significance was determined using unpaired *t*-test.

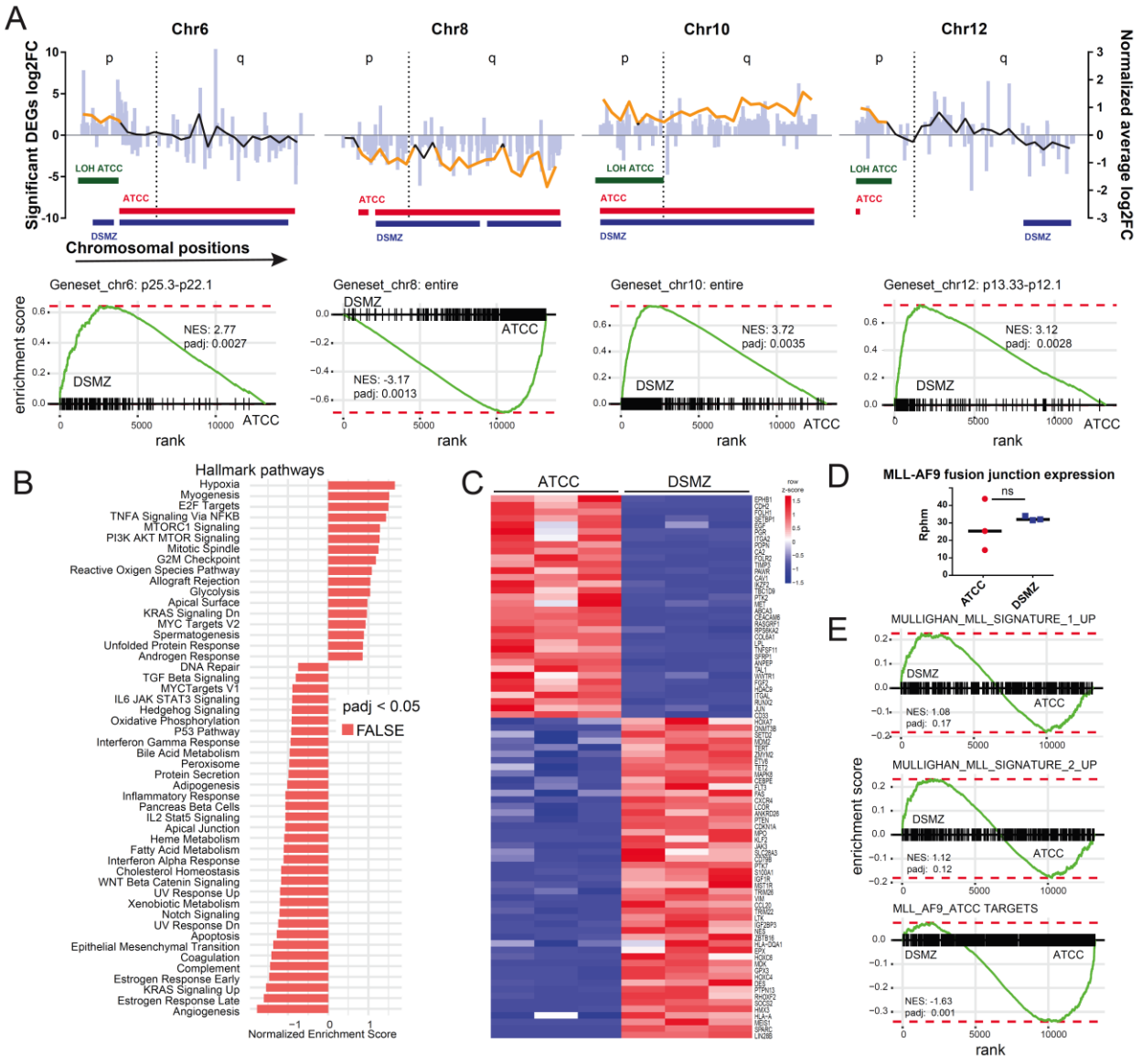

Figure 2.2 Divergence in cytogenetic aberrations leads to major transcriptomic differences between ATCC and DSMZ cells

a. Top panel: Integrative chromosome plots depicting log2(FC) of significant DEGs (light blue bars) at their gene start position (y-axis) and normalized average log2(FC) in RNA expression (black lines). HTCs are indicated in orange. Green bars represent LOH regions in indicated cells, while horizontal red and blue bars represent aneuploidy regions (eSNP-karyotyping data). Bottom panel: GSEA of genes located in the indicated HTCs. **b.** Normalized enrichment scores for hallmark gene sets evaluating enrichment of biological processes. **c.** Heatmap of AML-related genes among the DEGs. Genes are sorted by their log2FC. **d.** MLL-AF9 fusion junction expression in ATCC and DSMZ cells (count of fusion reads, normalized to the total number of reads); Rphm,

reads per hundred million. Statistical significance was determined using unpaired *t*-test; ns, non-significant **e.** GSEA of gene sets associated with MLL fusion AML, established from studies using patient specimens (top two panels) or THP-1 cells (bottom panel); NES, normalized enrichment score; padj, false discovery rate adjusted p-value.

2.10 Supplementary Figures and Tables

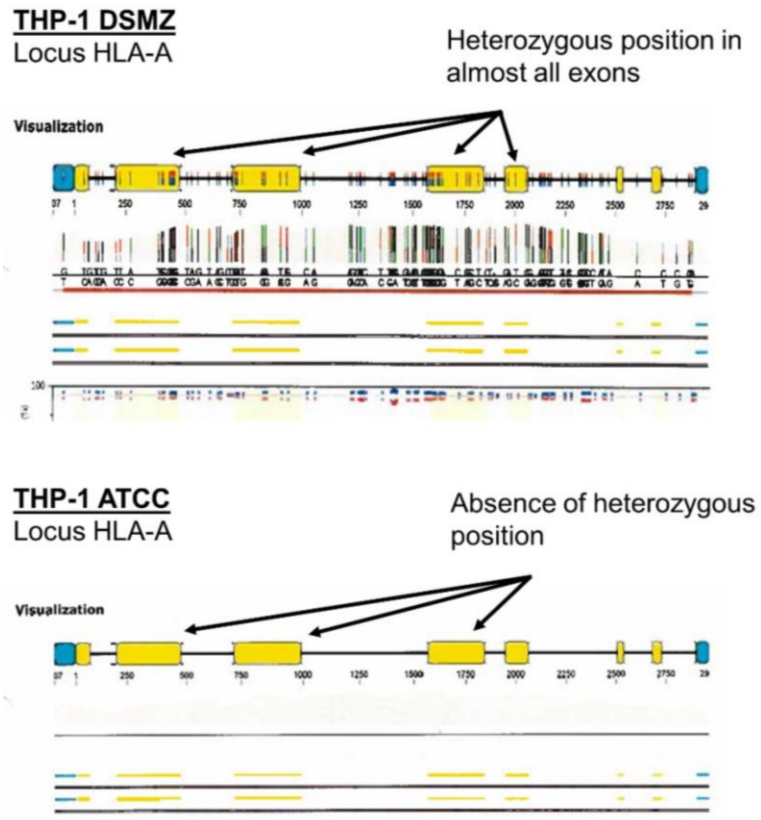

Figure S 2.1 Visualization of HLA-A locus following next generation sequencing

THP-1 DMSZ cells show heterozygous positions in almost all exons. On the contrary, loss of heterozygosity (LOH) of ATCC THP-1 cell line can be observed in almost all exons.

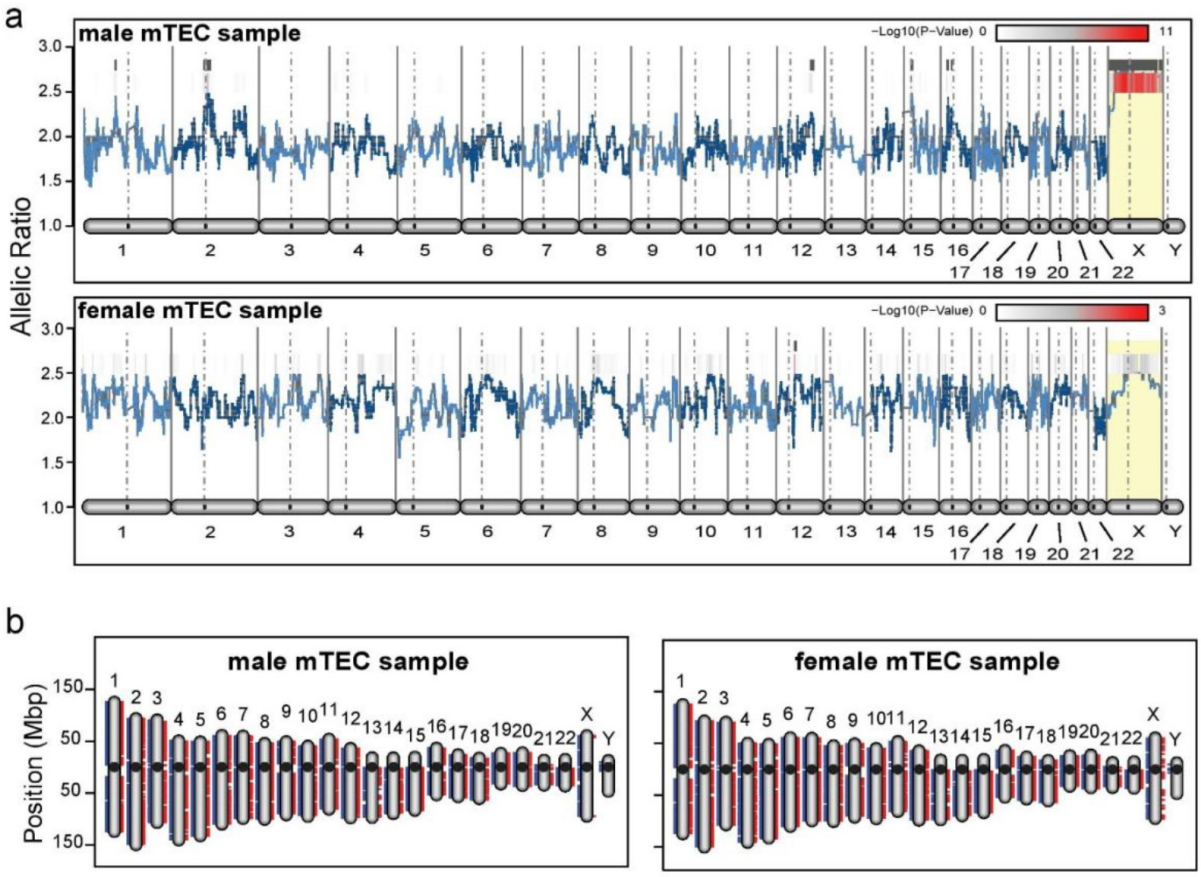

Figure S 2.2 Illustrative eSNP-karyotyping analysis of two normal healthy RNA-Seq samples available in our lab (medullary thymic epithelial cells, mTECs) obtained either from a male or female donor

a. Allelic ratio plots show that a single copy of chromosome X results in the significant ($p<0.01$, black vertical bars) detection of chromosomal aberrations. b. the absence of heterozygous SNPs (red horizontal bars along chromosomes) confirms the presence of a single copy of chromosome X in male donor.

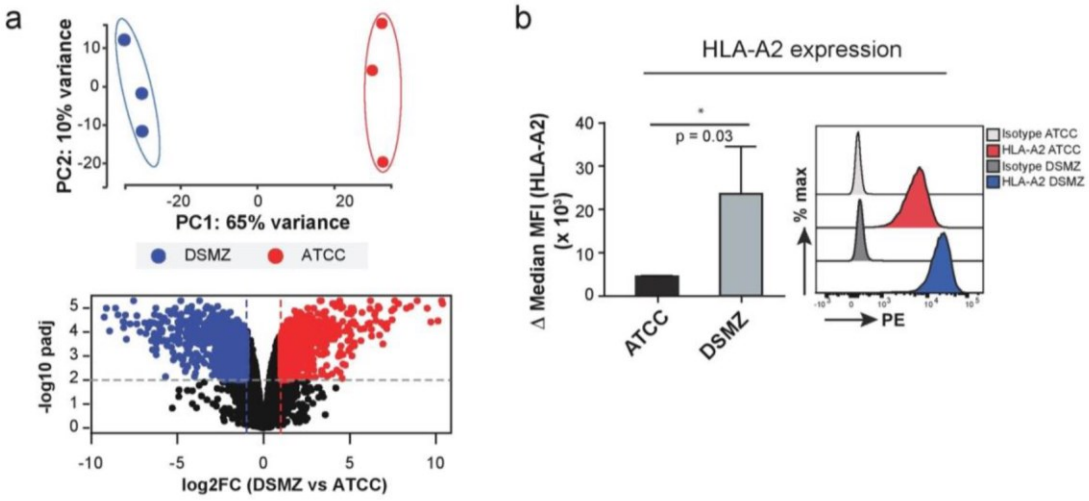

Figure S 2.3

a. PCA (top panel) and volcano plots (bottom panel) depiction of transcriptomic data. Colored dots in volcano plot represent DEGs with fold change higher than 2 or lower than -2 and adjusted p-value < 0.01; PC, principal component. b. Mean fluorescence intensity (MFI) of HLA-A02 surface expression. Statistical significance was determined using unpaired t-test.

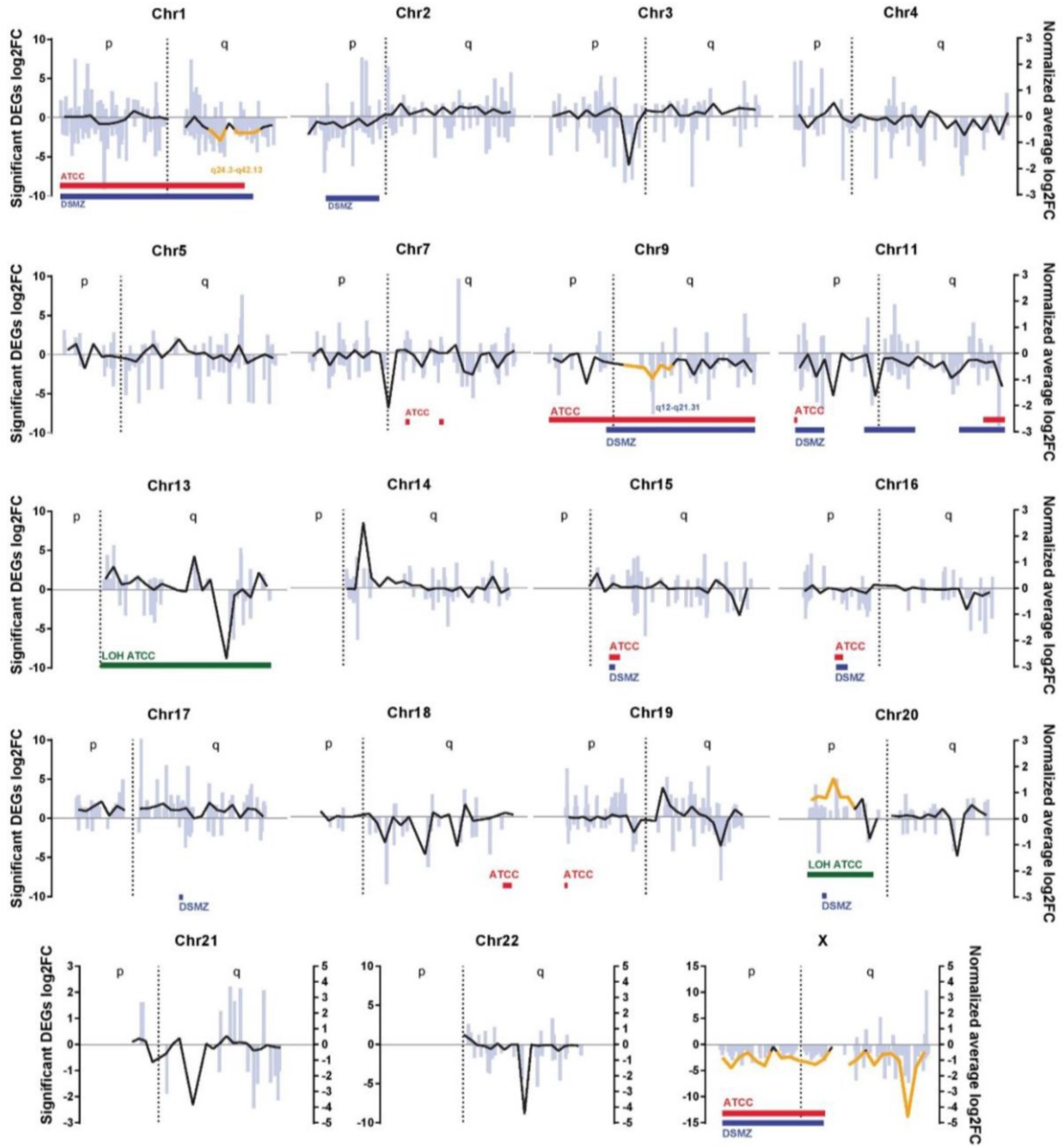

Figure S 2.4 Integrative plots of chromosomes not showed in figure 2

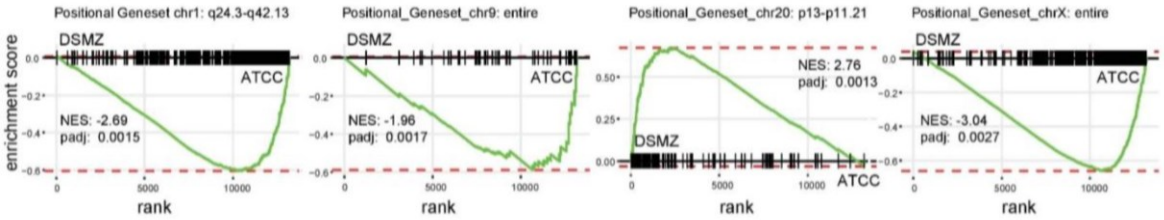

Figure S 2.5 GSEA of HTCs (from figure S4) not showed in figure 2

Table S 2.1 Short Tandem repeats (STR) profiling of ATCC and DSMZ cell lines, performed by ATCC

Locus	Database profile: THP-1		Query profile: ATCC		Query profile: DSMZ	
TH01	8	9.3	8	9.3	8	9.3
D5S818	11	12	11	12	11	12
D13S317	13		13		8	13
D7S820	10		10		10	
D16S539	11	12	11	12	11	12
CSF1PO	11	13	11	13	11	13
Amelogenin	X	Y	X	Y	X	Y
vWA	16		16		16	17
TPOX	8	11	8	11	8	11
Shared alleles between query and database:			15		15	
Total alleles in the database profile:			15		15	
Percent match:			100		100	

Table S 2.2 THP-1 variants based on STR profiling data available from Cellosaurus database

Cell line sources	ATCC/DSMZ	Additional details
AddexBio	ATCC	
CLS	ATCC	
Cosmic-CLP	ATCC	
ECACC	ATCC	
KCLB	ATCC	
PubMed=25877200	ATCC	
RCB	DSMZ	
TKG	DSMZ	
CCRID	unknown	homozygous vWA
JCRB	unknown	varied TH01 loci
ICLC	ATCC?	no information of D13S317 loci
PubMed=11416159	ATCC?	no information of D13S317 loci

Table S 2.3 HLA typing of ATCC and DSMZ THP1 cells

ATCC THP-1		
	Allele 1	*Allele 2*
HLA- A	02:01:01G	
HLA-B	15:11:01G	
HLA-C	03:03:01G	
DRB1	01:01:01G	15:01:01G
DQB1	05:01:01G	06:02:01G
DPB1	02:01:02G	04:02:01G

DSMZ THP-1		
	Allele 1	*Allele 2*
HLA- A	02:01:01G	24:02:01G
HLA-B	15:11:01G	35:01:01G
HLA-C	03:03:01G	03:03:01G
DRB1	01:01:01G	15:01:01G
DQB1	05:01:01G	06:02:01G
DPB1	02:01:02G	04:02:01G

Due to size constraints, the following three tables are not presented in the document. The files are available as an Excel file online at https://onlinelibrary.wiley.com/doi/full/10.1002/ijc.32967

Table S 2.4 Differentially expressed genes between DSMZ and ATCC, related to figure 1

Table S 2.5 GO term and KEGG pathway enrichment analysis of DEG, related to figure 1

Table S 2.6 GO term and KEGG pathway enrichment analysis of DEG, related to figure 1

Table S 2.7 Chromosomal positions of differential loss of heterozygozity (LOH).

Chromosome	Region		Cell line with LOH	notes
	start	end		
6	0	29829643	ATCC	
10	0	39800000	ATCC	
12	0	21516950	ATCC	4 hetero in ATCC and 78 in DSMZ
13	22600000	114364328	ATCC	
20	0	23364952	ATCC	3 hetero in ATCC and 63 in DSMZ

These regions were chosen either when heterozygous positions were absent while present in the other cell line or when the ratio of the number of heterozygous positions was higher than 10 between both cell lines (indicated in notes column).

Table S 2.8 Regions with aberrant allelic ratio (p<0.01 in table2 output of eSNP-Karyotyping)

Chromosome	ATCC		DSMZ	
	start	end	start	end
1	0	212285673	0	222479373
2			25233798	86022839
3			195952600	195984145
6	31888435	166326158	11579297	26878743
6			31781043	160040705
7	73096012	77024584		
7	100118557	101092793		
8	11849072	18056608	23566245	91068347
8	23260791	145055585	97727075	145055987
9	0	137611586	38549464	137582088
10	0	133797422	0	133797422
11	0	216290	0	18484005
11	120217578	134247121	43853198	76926636
11			105025886	134152192
12	0	1955534	102337750	131938173
15	28717938	35939697	28820037	31377308
16	15131600	19702997	16379153	21979439
17			45507108	46555327
18	74510926	78139949		
19	0	620392	58417419	58549990
20			5497596	6000572
X	2906947	79362608	2488411	77682737

Due to size constraints, the following three tables are not presented in the document. The files are available as an Excel file online at https://onlinelibrary.wiley.com/doi/full/10.1002/ijc.32967

Table S 2.9 . Positional gene sets used for GSEA analysis, related to figure 2

Table S 2.10 List of AML-related genes retrieved from GeneCards database (https://www.genecards.org)

Table S 2.11 MLL-AF9 target genes described in THP-1 cells by Prange KHM et al. Oncogene 2017. 8;36(23):3346-3356 [Table S4]

2.11 References

1. Krivtsov, A.V. and S.A. Armstrong, *MLL translocations, histone modifications and leukaemia stem-cell development.* Nat Rev Cancer, 2007. **7**(11): p. 823-33.
2. Barabe, F., et al., *Modeling the initiation and progression of human acute leukemia in mice.* Science, 2007. **316**(5824): p. 600-4.
3. Tsuchiya, S., et al., *Establishment and characterization of a human acute monocytic leukemia cell line (THP-1).* Int J Cancer, 1980. **26**(2): p. 171-6.
4. Baker, M., *1,500 scientists lift the lid on reproducibility.* Nature, 2016. **533**(7604): p. 452-4.
5. Fusenig, N.E., et al., *The need for a worldwide consensus for cell line authentication: Experience implementing a mandatory requirement at the International Journal of Cancer.* PLoS Biol, 2017. **15**(4): p. e2001438.
6. MacLeod, R.A., et al., *Widespread intraspecies cross-contamination of human tumor cell lines arising at source.* Int J Cancer, 1999. **83**(4): p. 555-63.
7. Capes-Davis, A., et al., *Check your cultures! A list of cross-contaminated or misidentified cell lines.* Int J Cancer, 2010. **127**(1): p. 1-8.
8. Capes-Davis, A., et al., *Match criteria for human cell line authentication: where do we draw the line?* Int J Cancer, 2013. **132**(11): p. 2510-9.
9. Drexler, H.G., et al., *False and mycoplasma-contaminated leukemia-lymphoma cell lines: time for a reappraisal.* Int J Cancer, 2017. **140**(5): p. 1209-1214.
10. Korch, C., et al., *DNA profiling analysis of endometrial and ovarian cell lines reveals misidentification, redundancy and contamination.* Gynecol Oncol, 2012. **127**(1): p. 241-8.
11. Liu, Y., et al., *Multi-omic measurements of heterogeneity in HeLa cells across laboratories.* Nat Biotechnol, 2019. **37**(3): p. 314-322.
12. Ben-David, U., et al., *Genetic and transcriptional evolution alters cancer cell line drug response.* Nature, 2018. **560**(7718): p. 325-330.
13. Bosshart, H. and M. Heinzelmann, *THP-1 cells as a model for human monocytes.* Ann Transl Med, 2016. **4**(21): p. 438.
14. Olarerin-George, A.O. and J.B. Hogenesch, *Assessing the prevalence of mycoplasma contamination in cell culture via a survey of NCBI's RNA-seq archive.* Nucleic Acids Res, 2015. **43**(5): p. 2535-42.
15. Sergushichev, A.A., *An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation.* bioRxiv, 2016: p. 060012.
16. Weissbein, U., et al., *Analysis of chromosomal aberrations and recombination by allelic bias in RNA-Seq.* Nat Commun, 2016. **7**: p. 12144.
17. Butler, J.M., *Genetics and genomics of core short tandem repeat loci used in human identity testing.* J Forensic Sci, 2006. **51**(2): p. 253-65.
18. Bairoch, A., *The Cellosaurus, a Cell-Line Knowledge Resource.* J Biomol Tech, 2018. **29**(2): p. 25-38.
19. Parson, W., et al., *Cancer cell line identification by short tandem repeat profiling: power and limitations.* Faseb j, 2005. **19**(3): p. 434-6.
20. Battle, R., et al., *Molecular characterisation of the monocytic cell line THP-1 demonstrates a discrepancy with the documented HLA type.* Int J Cancer, 2013. **132**(1): p. 246-7.
21. Szolek, A., et al., *OptiType: precision HLA typing from next-generation sequencing data.* Bioinformatics, 2014. **30**(23): p. 3310-6.

22. Marzotto, M., et al., *Arnica montana Stimulates Extracellular Matrix Gene Expression in a Macrophage Cell Line Differentiated to Wound-Healing Phenotype.* PLoS One, 2016. **11**(11): p. e0166340.
23. Howell, W.M., V. Carter, and B. Clark, *The HLA system: immunobiology, HLA typing, antibody screening and crossmatching techniques.* J Clin Pathol, 2010. **63**(5): p. 387-90.
24. Hein, J., et al., *Low resolution DNA typing of the HLA-B5 cross-reactive group by nested PCR-SSP.* Tissue Antigens, 1995. **45**(1): p. 27-35.
25. Fussell, H., et al., *HLA-A9 antibodies and epitopes.* Tissue Antigens, 1996. **47**(4): p. 307-12.
26. Odero, M.D., et al., *Cytogenetic and molecular analysis of the acute monocytic leukemia cell line THP-1 with an MLL-AF9 translocation.* Genes Chromosomes Cancer, 2000. **29**(4): p. 333-8.
27. Liberzon, A., et al., *The Molecular Signatures Database (MSigDB) hallmark gene set collection.* Cell Syst, 2015. **1**(6): p. 417-425.
28. Prange, K.H.M., et al., *MLL-AF9 and MLL-AF4 oncofusion proteins bind a distinct enhancer repertoire and target the RUNX1 program in 11q23 acute myeloid leukemia.* Oncogene, 2017. **36**(23): p. 3346-3356.
29. Alachkar, H., et al., *SPARC promotes leukemic cell growth and predicts acute myeloid leukemia outcome.* J Clin Invest, 2014. **124**(4): p. 1512-24.
30. Zhou, J., et al., *Inhibition of LIN28B impairs leukemia cell growth and metabolism in acute myeloid leukemia.* J Hematol Oncol, 2017. **10**(1): p. 138.
31. Kumar, A.R., et al., *A role for MEIS1 in MLL-fusion gene leukemia.* Blood, 2009. **113**(8): p. 1756-8.
32. Mullighan, C.G., et al., *Pediatric acute myeloid leukemia with NPM1 mutations is characterized by a gene expression profile with dysregulated HOX gene expression distinct from MLL-rearranged leukemias.* Leukemia, 2007. **21**(9): p. 2000-9.
33. Kleensang, A., et al., *Genetic variability in a frozen batch of MCF-7 cells invisible in routine authentication affecting cell function.* Scientific Reports, 2016. **6**(1): p. 28994.
34. Mizutani, M., et al., *Analysis of 168 short tandem repeat loci in the Japanese population, using a screening set for human genetic mapping.* J Hum Genet, 2001. **46**(8): p. 448-55.
35. Weaver, D.A., et al., *Localization of tumor suppressor gene candidates by cytogenetic and short tandem repeat analyses in tumorigenic human bronchial epithelial cells.* Carcinogenesis, 2000. **21**(2): p. 205-11.

3 Chapter 3: Autophagy degrades immunogenic endogenous retroelements induced by 5-azacytidine in acute myeloid leukemia

Nandita Noronha[1], Chantal Durette[1], Bianca E Silva[2], Justine Courtois[2], Juliette Humeau[1], Allan Sauvat[3], Marie-Pierre Hardy[1], Krystel Vincent[1], Jean-Philippe Laverdure[1], Joël Lanoix[1], Frédéric Baron[2], Pierre Thibault[1,4], Claude Perreault[1,4], Gregory Ehx[1,2,4,5].

[1] IRIC, University of Montreal, Montreal, Canada. [2] GIGA-I3: Hematology, University of Liege, Liege, Belgium, [3] Equipe labellisée par la Ligue contre le Cancer, Université de Paris, Sorbonne Université, Inserm U1138, Institut Universitaire de France, Paris, France. [4] Senior authors. [5] Lead contact

***Correspondence**: g.ehx@uliege.be (G.E.)

Keywords: azacitidine, acute myeloid leukemia, EREs, autophagy, MHC-I epitopes

The manuscript in this chapter is currently under revision in *Leukemia.*

3.1 Author contributions

N.N., G.E., and C.P. designed the study. N.N. and G.E. performed all the cell culture experiments, bioinformatic analyses, and data interpretation. G.E. built databases for mass spectrometry analysis and performed main bioinformatics analysis on differentially expressed MAPs and RNA-seq from AML leucegene cohort. C.D. and J.L. performed immunoprecipitation and mass spectrometry experiments. J.C. and B.E.S contributed to cell culture experiments and performed functional assays. J.H. and A.S. provided bioinformatics assistance for the analysis of immunofluorescence data. M.-P.H., K.V., J-P.L., F.B., P.T., and C.P., contributed to the analysis and interpretation of data. N.N., G.E. and C.P. wrote the manuscript and the final manuscript was edited and approved by all authors.

Contributions per figures:

NN: Generated and analysed data for Figure 3.1B, 3.2A-3.2H, 3.2J, 3.3A, 3.6A-C, S3.2, S3.4-3.5, S3.7, S3.8 A-B. Generated data for figure 3.4A-E, 3.4G-H, 3.5D, S3.8C.

GE: Analysed and generated figures 3.1A, 3.2I, 3.4A-E, 3.4G-H, S3.1A, S3.8C. Generated and analysed data for figure 3.3B-D, 3.4I, 3.5B-C, 3.6D-F, S3.1B-C, S3.3, S3.6B

BES: Generated and analysed data for Figure 3.4F, S3.6A

JC: Generated and analysed data for Figure 3.5A

3.2 Abstract

The hypomethylating agent 5-azacytidine (AZA) is the first-line induction therapy for acute myeloid leukemia (AML) patients unsuitable for intensive chemotherapy. Evidence suggests that the anti-tumor effect of AZA results partly from T-cell cytotoxic responses against MHC-I-associated peptides (MAPs) whose expression is induced by hypomethylation. We analyzed the impact of AZA on the transcriptome and MAP repertoire of four AML cell lines and validated salient findings in the transcriptome of 437 primary AML samples. We demonstrate that AZA caused pleiotropic changes in AML cells via perturbation of transcription, translation, and protein degradation. Overall, 1,364 MAPs were upregulated in AZA-treated cells, including several cancer-testis antigens. Increased MAP abundance was due to the upregulation of corresponding transcripts in a minority of cases and post-translational events in most cases. Furthermore, AZA-induced hypomethylation increased the abundance of numerous transcripts, of which 38% were endogenous retroelements (EREs). Upregulated ERE transcripts triggered innate immune responses but were degraded by autophagy and not processed into MAPs. Autophagy resulted from the formation of protein aggregates caused by AZA-dependent inhibition of DNMT2, a tRNA-methyl transferase enzyme. We found that autophagy inhibition had a synergistic effect with AZA on AML cell proliferation and survival, increased ERE levels and triggered pro-inflammatory responses. Finally, autophagy gene signatures were associated with a lower abundance of $CD8^+$ T-cell markers in AML patients expressing high levels of EREs. Altogether, this work demonstrates that the impact of AZA is regulated at several levels and provides a rationale for assessing the synergy between AZA and autophagy inhibitors in AML treatment.

3.3 Introduction

Acute myeloid leukemia (AML) is the most common acute leukemia in adults, with an overall 5-year survival below 30%. Standard therapy involves intensive chemotherapy with a '7+3' regimen of cytarabine and anthracycline. Although AML is a heterogeneous disease, aberrant genomic methylation (hypermethylation in particular [1, 2]) is a hallmark of AML blasts. Therefore, hypomethylating agents (HMAs) such as 5-azacytidine (azacitidine, AZA) and 5-aza-2′-deoxycytidine (decitabine, DAC) are used as first-line induction therapy for AML patients unsuitable for intensive chemotherapy [3]. AZA is also used in maintenance therapy for fit patients without an *FLT3* mutation [3]. However, only 18–47% of patients respond to these therapies, stressing the need to improve therapy efficacy, possibly by combining them with other pharmacologic agents [4, 5].

AZA and DAC are cytidine nucleoside analogs that incorporate into genomic DNA during replication in the prophase of mitosis [6]. High concentrations of AZA and DAC exert cytotoxic effects by inducing DNA double-strand breaks. However, at low concentrations, they act as suicide substrates for DNA methyltransferases (DNMTs) 1 and 3, leading to their degradation and the DNA demethylation of daughter cells. AZA differs from DAC by its ability to incorporate into RNA and DNA, thus inhibiting DNMT2, a transfer RNA (tRNA) methyltransferase [7]. While both HMAs have similar response rates in AML [8], only AZA significantly improves overall survival compared with conventional care regimens in phase III randomized trials [9, 10]. Therefore, AZA is currently FDA-approved as a first-line treatment in AML [11].

In addition to their cytotoxic and demethylating effects, HMAs may mediate anti-leukemic activities by promoting effector immune cells to recognize malignant blasts. Specifically, HMAs enhance the expression of transcripts coding for cancer-testis antigens (CTAs) [12, 13]. CTA genes

are normally silenced by genomic methylation and code for antigens deemed immunogenic because they are not expressed in normal MHC-positive somatic cells [14]. Accordingly, some studies have shown that HMAs promote cytotoxic T-cell (CTL) activity [15], while others demonstrate a specific cytotoxic activity against CTAs [13, 16, 17]. These studies suggest that HMA-induced CTAs can promote anti-leukemic CD8 T-cell reactions by generating immunogenic MHC-I-associated peptides (MAPs).

Along with CTAs, HMAs promote the expression of endogenous retroelements (EREs) [18, 19]. EREs are highly repetitive sequences that are remnants of transposable elements incorporated into the human genome millions of years ago [20]. EREs represent ~45% of the human genome and can be separated into LINEs and SINEs (long and short interspersed elements, respectively) and LTRs (long terminal repeats), the latter of which includes endogenous retroviruses (ERVs). EREs are epigenetically silenced mainly by genomic methylation in normal somatic cells [21] and dysregulated ERE expression is associated with several pathologic conditions, including autoimmunity, inflammatory disorders, aging, and cancer [22]. HMA-induced ERE overexpression in solid cancers leads to viral mimicry and concomitant innate immune response [23, 24]. Moreover, we and others have demonstrated that in addition to being expressed, EREs are capable of being presented by MHC-I molecules and serve as immunogenic tumor antigens, notably in AML [25-28]. While EREs are perfect candidates for generating immunogenic MAPs following HMA treatment, there is a lack of robust evidence to support that HMAs enhance their MAP presentation (and subsequent CTL responses) in AML.

T cells recognize MAPs, not transcripts. Hence, available transcriptomic studies do not allow inferences on the MAP repertoire of AZA-treated cells. Herein, we sought to directly evaluate AZA's capacity to enhance the presentation of CTAs and ERE-derived MAPs in AML.

We discovered that AZA promotes the expression of multiple CTA-derived, but not ERE-derived MAPs. Mechanistically, we propose that the lack of ERE-derived MAPs is due to AZA-induced autophagy of ERE transcripts.

3.4 Materials and methods

Cell culture

THP-1, OCI-AML-3, MOLM-13, and SKM-1 cell lines were freshly purchased from the Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ) for the current study. THP-1, MOLM-13, and SKM-1 cells were maintained in RPMI 1640 (Gibco, NY-US, 11875-093) containing L-glutamine and supplemented with 10% heat-inactivated fetal bovine serum (FBS, Gibco 12483) and 1% penicillin-streptomycin (10,000 U/mL, Gibco 15140-122). OCI-AML-3 cells were maintained in MEM alpha (Gibco, NY-US, 12571063) containing L-glutamine and nucleotides supplemented with 10% heat-inactivated fetal bovine serum (FBS, Gibco 12483) and 1% penicillin-streptomycin (10,000 U/mL, Gibco 15140-122).

Cell line treatments

For AZA treatments, cell lines were treated daily with 0.25 μM or 0.5 μM of AZA (Sigma Aldrich A2385) for 72h, followed by removal of the drug (replacement of the medium) at time points as indicated in the results section. For spautin-1 dose responses, cell lines were treated with spautin-1 (Millipore Sigma SML0440) or 0.1% DMSO control for four days, and spautin-1 or DMSO was replenished when fresh media was added. For co-treatment experiments with AZA and spautin-1 or rapamycin (kind gift from Guy Sauvageau's lab), cell lines were treated for four days with either spautin-1 (5 μM), rapamycin (0.5 μM), or 0.1% DMSO in presence of 0.5 μM of AZA. AZA was added daily for 72h, followed by discontinuation for 24h. Levels of genomic 5-methylcytosine

after AZA treatment were measured by ELISA with the MethylFlash Global DNA Methylation Kit (Epigentex, P-1030).

Flow cytometry

Cells (~1×10^5 cells/sample) were collected and washed 1X with PBS (Sigma P3813) before fixation/permeabilization with either the FOXP3/Transcription Factor Staining Buffer Set (eBioscience) and staining with anti-DNMT1-PE (EPR3522, Abcam) or the Fix Buffer I (Becton Dickinson, BD) followed by Phosflow Perm Buffer III (BD) and staining with anti-H2γX-AF647 (pS139, BD). All staining steps were performed at 4ºC for 30 min in the dark, and cells were pre-incubated with Fc receptor blocking antibody (BD Pharmingen 552930) for 10 min before incubation with antibodies of interest. Data were acquired on a FACS Canto II (BD).

For protein aggregate detection, 1×10^5 cells/sample were washed 3X with PBS and then fixed/permeabilized with the Cytofix/Cytoperm kit (BD) according to the manufacturer's instructions. Cells were then washed 3X with PBS and resuspended in 250 µL of assay buffer (ENZO #51035) supplemented with Proteostat Aggresome detection dye (ENZO #51035) diluted 1:10,000. Cells were analyzed with a FACS Canto II (BD) after 30 min of staining without additional washes.

Autophagy activity was measured using an autophagy assay kit (Abcam, ab139484) according to the manufacturer's protocol. In brief, cells (~2×10^5/sample) were cultured for 24h in various concentrations of AZA, DAC, or rapamycin in the presence of 120 µM of chloroquine (to accumulate autophagic granules and enable their detection). Cells were collected by centrifugation and washed with assay buffer before being resuspended in 250 µL of culture medium containing 5% FBS and mixed with 250 µL of diluted green staining solution. Cells were incubated for 30

min at 37ºC in the dark and washed with assay buffer. Relative autophagy activities were measured using a Cytoflex flow cytometer (Beckman Coulter).

Immunoproteasome activity

Immunoproteasome activity was assessed on THP-1 or OCI-AML-3 cells (~5×10^5 cells/sample) lysed in 1 mL of lysis buffer (50 mM Tris-HCl, 2 mM DTT, 5 mM $MgCl_2$, 10% (v/v) glycerol, 2 mM ATP, and 0.05% (v/v) digitonin). The assay was performed immediately after lysis with the Immunoproteasome Activity Fluorometric Assay Kit I (Ubiquitin-Proteasome Biotechnologies, TX-US, J4160) according to the manufacturer's instructions. Fluorescence was detected using the TriStar² LB 942 microplate reader (Berthold Technologies GmbH & Co.KG).

Real-time PCR

Quantitative real-time PCR was performed for candidate ERE and dsRNA-induced interferon genes using validated Universal ProbeLibrary assays (Roche) on the Viia7 Real-time PCR system (Applied Biosystems). Relative target mRNA levels were normalized to GAPDH and ACTIN and analyzed using Expression Suite software v1.1 (ThermoFisher).

Immunofluorescence experiments

THP-1 cells were attached on chambered slides (iBidi 80826) with poly-L-lysine and fixed using ice-cold methanol for 15 min at -20ºC, washed three times with PBS, and incubated with saturation buffer (5% BSA-PBS) for 1h. Primary antibody was added (1:200, anti-dsRNA, clone J2, SCICONS) and incubated overnight at 4ºC. Cells were washed three times for 15 minutes with PBS on a shaker, followed by incubation with secondary antibodies (1:2000 goat anti-mouse IgG AlexaFluor 594 Invitrogen A-11020) at room temperature for 1h and washed three times for 10 minutes with PBS. Next, cells were incubated with DAPI containing PBS, and slides were stored

at 4ºC in the dark for at least three days before confocal analyses. Confocal analyses were performed with a Zeiss LSM700 confocal microscope, and images were quantified using EBImage package on R. Transfected cell lines with Poly(I:C) were used as positive controls. No unspecific staining was observed with secondary antibodies alone.

Library preparation and RNA sequencing

Total RNA was isolated using TRIzol (Thermo Scientific) followed by RNeasy purification (Qiagen). RNA was quantified using Qubit (Thermo Scientific), and quality was assessed with the 2100 Bioanalyzer (Agilent Technologies). Transcriptome libraries were generated using the KAPA RNA HyperPrep (Roche) using a poly-A selection (Thermo Scientific). Sequencing was performed on the Illumina NextSeq500, obtaining around 120M paired-end reads per sample (60M clusters) for AZA vs. control in AML cell lines and 30M single-end reads per sample for spautin-1+AZA validation RNA-seq experiment.

Gene expression analyses

All transcript expression (canonical genes and EREs) quantifications were performed with kallisto v0.43.0 [29] with default parameters. Kallisto's transcript-level count estimates were converted into gene-level counts using the R package tximport. EdgeR was used to normalize counts using the TMM algorithm and output count-per-million (cpm) values. Differential gene expression analyses were conducted in R3.6.1, as reported previously [30]. In brief, raw read counts were converted to cpm, normalized relative to library size, and lowly expressed genes were filtered by keeping genes with cpm >1 in at least two samples using edgeR 3.26.8 [31] and limma 3.40.6 [32]. Subsequently, voom transformations and linear modeling using limma's lmfit were performed. Moderated t-statistics were then computed with eBayes. Genes with false-discovery rates ≤0.05

and $-1 \geq \log_2(FC) \geq 1$ were considered significantly differentially expressed. For differential gene expression analyses performed on AZA-treated cell lines, a unique paired analysis comparing AZA-treated vs. control cells was performed.

Gene ontology and biological pathway annotations were performed with DAVID v6.8 (https://david.ncifcrf.gov). Functional annotations with a p-value <0.05 were considered significant. Gene set enrichment analysis (GSEA) was performed with the fgsea package in R [33]. A pre-ranked gene list was generated by ranking expressed genes obtained from limma-voom on moderated t-statistics. Gene sets were obtained either from the HALLMARK or REACTOME matrix, downloaded from the MSigDB database. The enrichment analysis for REACTOME gene sets among genes significantly upregulated in the AZA+Spautin-1 vs. AZA-only cells was performed with the Reactome FI module in Cytoscape v3.7.2 [34].

Database generation for mass spectrometry identifications

To build databases to analyse MAPs originating from any region of the genome (annotated protein-coding exons, introns, EREs, ncRNAs, intergenic regions, etc.) and including MAPs deriving from mutations present in the genome of the analyzed cell line, we adopted an alignment-free proteogenomic approach. We built two personalized, non-overlapping proteomes, canonical and non-canonical, and concatenated them to perform MS identifications. All scripts and usage instructions for the pipeline can be found on Zenodo (doi 10.5281/zenodo.7096388).

Personalized canonical proteomes

RNA-Seq reads were trimmed using Trimmomatic v0.35 and aligned to GRCh38.88 using STAR v2.5.1b [35] running with default parameters except for --alignSJoverhangMin, --alignMatesGapMax, --alignIntronMax, --quantMode and --alignSJstitchMismatchNmax

parameters for which default values were replaced by 10, 200,000, 200,000, TranscriptomeSAM and "5 -1 5 5", respectively, to generate bam files. Single-base mutations with a minimum alternate count setting of 5 were identified using freeBayes 1.0.2-16-gd466dde [36]. Transcript expression was quantified in transcripts per million (tpm) with kallisto v0.43.0 with default parameters. Finally, we used pyGeno [37] to insert high-quality sample-specific single-base mutations (freeBayes quality >20) into the reference exome and export sample-specific sequences of known proteins generated by expressed transcripts (tpm >0) to generate fasta files of personalized canonical proteomes.

Personalized non-canonical proteomes

Step 1. We built consensus genomes and transcriptomes (including only genomic regions covered by RNA-seq reads and single-nucleotide polymorphisms as ambiguous nucleotides) from STAR-generated bam files of each sample (per replicate per cell line). This was performed with the reference genome and transcriptome as input of the samtools [38] and bcftools suites [39]: 'samtools mpileup -C50 -uf reference.fasta sample.bam | bcftools call -c | vcfutils.pl vcf2fq | gzip >> consensus.fastq.gz'. The consensus genomes and transcriptomes were then chopped into k-mers (of 24, 27, 30, or 33 nucleotide lengths, corresponding to the length of MAPs: 8–11 amino acids) with a homemade python script, and k-mers containing consensus nucleotides (R|Y|M|K|W|S|B|D|H|V|N) were disambiguated (A|T|C|G) using a homemade python script. These k-mers were then reverse-complemented, and all k-mers (non-ambiguous and disambiguated, originals and reverse-complemented) were assembled in a single database generated with Jellyfish v2.2.3 [40].

Step 2. The fastq files of each sample (per replicate per cell line) were used to generate k-mer libraries of either 24, 27, 30, or 33 nucleotides in length containing k-mers present at least

twice per sample. This was performed with Jellyfish: 'jellyfish count -L 2 -m <length> -F 2 -s 1G -o sample.jf' on trimmed forward and reverse-complemented (with the fastx_reverse_complement function of the FASTX-Toolkit v0.0.14) reverse fastq files. Next, the k-mer databases were combined into single databases per cell line (four databases were obtained eventually, one per MAP length) by keeping only those k-mers with three occurrences in three different samples (out of six samples: three controls and three AZA-treated). This was performed with a script ('joinCounts') obtained from the DE-kupl pipeline [41]: 'joinCounts -r 3 -a 3 <fastq files>'. This allowed us to retain k-mers that most likely generate MAPs (since high RNA expression is a robust predictor of MAP generation [28, 42]).

Step 3. The k-mers generated in step 2 were queried in the k-mer databases generated in step 1. This allowed us to discard consensus artifacts such as exon-intron junctions, wrong SNP calling, false intron coverage and to filter k-mers on their minimum occurrence and inter-sample sharing. The query was performed with the 'jellyfish query -i' command.

Step 4. The personalized canonical proteomes were chopped into peptide k-mers (8, 9, 10, or 11 amino acids) using a homemade python script.

Step 5. The resulting k-mers from step 3 were translated into their peptide sequence with a homemade python script. Peptide k-mers containing stops were removed (with awk), and peptide k-mers generated in step 4 were removed from this list to prevent overlaps between the canonical and non-canonical proteome.

Step 6. The non-canonical peptides were tested for their capacity to bind HLA alleles of their respective cell line (determined with Optitype [43]) with either MHC flurry 1.4.0 [44] or netMHCpan 4.0 [45] for alleles not supported by MHC flurry. Predictions were made with the

epitopepredict module [46] to handle MHC flurry and NetMHCpan. Peptides with a percentile rank ≤2% were kept for further processing.

Step 7. Since leucine and isoleucine variants are not distinguishable by standard MS approaches, we inspected the list of non-canonical peptides and discarded those for which an existing variant (MHC binder as well) was flagged as canonical. Next, short peptides with sequences completely included in the sequence of longer peptides were discarded (awk) from the list, and peptide sequences were used to generate a fasta file, eventually concatenated with the personalized canonical proteome to generate the final MS databases.

MHC-I peptide isolation by immunoprecipitation

W6/32 antibodies (BioXcell) were incubated in PBS for 60 min at room temperature with PureProteome protein A magnetic beads (Millipore) at a ratio of 1 mg of antibody per 1 mL of slurry. Antibodies were covalently cross-linked to magnetic beads using dimethylpimelidate as described [47]. The beads were stored at 4ºC in PBS (pH 7.2) and 0.02% NaN_3. Frozen cell pellets (118–135×10^6 cells/pellet) were thawed and resuspended in 0.4 mL PBS (pH 7.2) and solubilized with 1 mL of detergent buffer containing PBS (pH 7.2) and 1% (w/v) CHAPS (Sigma) supplemented with a protease inhibitor cocktail (Sigma). Cell pellets were incubated for 60 min with tumbling at 4ºC and then spun at 16,600g for 20 min at 4ºC. Supernatants were transferred into new tubes containing 1 mg of W6/32 antibody covalently-cross-linked protein A magnetic beads per sample and incubated with tumbling for 20h at 4ºC. Samples were placed on a magnet to recover bound MHC-I complexes to magnetic beads. Magnetic beads were first washed with 8× 1 mL PBS, then with 1× 1 mL of 0.1X PBS, and finally with 1× 1 mL of H_2O. MHC-I complexes were eluted from the magnetic beads by acidic treatment using 0.2% formic acid (FA). To remove residual magnetic beads, eluates were transferred into 2.0 mL Costar mL Spin-X centrifuge tube

filters (0.45 mm, Corning) and spun for 5 minutes at 855g. Filtrates containing peptides were separated from MHC-I subunits (HLA molecules and β-2 macroglobulin) using homemade stage tips packed with two 1 mm diameter octadecyl (C-18) solid phase extraction disks (EMPORE). Stage tips were pre-washed with methanol, then with 80% acetonitrile (ACN) in 0.1% trifluoroacetic acid (TFA), followed by 0.1% TFA, and finally with 1% TFA. Samples were loaded onto stage tips and washed with 1% TFA, followed by 0.1% TFA. Peptides were eluted with 30% ACN in 0.1% TFA, dried using vacuum centrifugation, and then stored at -20ºC until MS analysis.

TMT labeling

MHC-I peptide extracts were reconstituted in 200 µL of 200 mM HEPES buffer (pH 8.2). TMT0-126 reagents or TMT6-plex (Thermo Fisher Scientific) were dissolved in 40 µL of anhydrous ACN (Sigma-Aldrich), and 5 µL of 0.02 mg/ µL was added to the peptides. The solutions were gently mixed and incubated for 90 min without agitation at room temperature before the reactions were quenched by hydroxylamine (Thermo Fisher Scientific). Samples were desalted on homemade C18 stage tips and dried down.

Mass spectrometry analyses

Dried peptide extracts were resuspended in 4% FA (EMD Millipore) and loaded on a custom C18 analytical column (20 cm × 150 mm i.d. packed with C18 Jupiter Phenomenex) with a 106-min gradient from 0% to 30% ACN (0.2% FA) and a 600 nL/min flow rate on an EasynLC II system. Samples were analyzed with an Exploris mass spectrometer (Thermo Fisher Scientific) in positive ion mode with the source at 2.8 kV. Each full MS spectrum, acquired with 240,000 resolution, was followed by MS/MS spectra, where the most abundant multiply charged ions were selected for

MS/MS sequencing with a resolution of 30,000, 100% normalized automatic gain control, injection time of 700 ms, and collisional energy of 36%.

Identification of MAPs and differential MAP analyses

Liquid chromatography (LC)-MS/MS (LC-MS/MS) data were searched against respective cell line-specific databases using PeaksXPro. For peptide identification, no enzyme was selected, and tolerance was set at 10 ppm and 0.01 Da for precursor and fragment ions, respectively. The occurrences of oxidation (M) and deamidation (NQ) were set as variable modifications. Following peptide identification, we used a modified target decoy approach built-in PEAKS and applied a sample-specific threshold on the PEAKS score to ensure a false discovery rate of 5%, calculated as the ratio between the number of decoy hits and the number of target hits above the score threshold. Binding affinities to the sample's HLA alleles were predicted with NetMHCpan 4.1b [48], and only 8 to 11-amino-acid-long peptides with a rank eluted ligand threshold $\leq 2\%$ were used for further annotation; these filtering steps were performed with MAPDP software [49]. Intensities of all modifications for a single peptide were summed, and peptides containing too many missing values were eliminated by keeping peptides quantified in two out of three replicates of at least one condition. Next, VSN normalization was performed, which was the best available normalization method based on analyses with NormalyzerDE [50]. Imputation for missing values was performed by Perseus with width of 0.3 and downshift of 1, and MAPs with p-values <0.05 and fold-changes (FC)>2 were considered significantly differentially expressed using limma analysis. MAPs exclusively detected in one condition were defined by having valid values from all three biological replicates in one condition while no values in the other condition.

Biotype attribution to identified MAPs

BamQuery [51] was used to annotate if MAPs derived from protein-coding, EREs, or other non-coding regions. CTAs were annotated using [52].

Bioinformatic analyses performed on MAPs

Amino acid compositions, aromaticity, and GRAVY indexes were assessed with the ProtParam module of Biopython. The RNA expression of each MAP was obtained using BamQuery [51].

Quantification and statistical analysis

Unless indicated otherwise, all statistical tests comparing two conditions were performed using the Mann–Whitney U test. All correlations were assessed with the Pearson correlation coefficient. Unless mentioned otherwise, all boxes in boxplots represent the median, 25^{th}, and 75^{th} percentiles, and whiskers extend to the 10th and 90th percentiles. Unless mentioned otherwise, all bar plots represent the average with standard deviation (SD). Plots and statistical tests were mainly performed with GraphPad Prism v9.1.1. For all statistical tests, **** refers to $p< 0.0001$, *** refers to $p< 0.001$, ** refers to $p< 0.01$, and * refers to $p< 0.05$.

3.5 Results

3.5.1 Low-dose AZA leads to delayed, transient ERE expression in AML cell lines.

To investigate the effects of AZA on the immunopeptidome of AML, we selected four AML cell lines (THP-1, MOLM-13, SKM-1, and OCI-AML-3) belonging to aggressive FAB types (M4/M5) and together covering different but frequent mutational statuses (MLL-AF9, FLT3-ITD, TET2 (L1418fs), and NPM1c+DNMT3A (R882C), respectively). As we wished for our immunopeptidomic analyses to reveal the effects of AZA independently of cytotoxic activity, we established a protocol that allowed DNMT1 degradation and genomic DNA demethylation without affecting viability or inducing DNA damage responses. The most desirable AZA doses were 0.25 μM (for MOLM-13 and SKM-1) and 0.5 μM (for THP-1 and OCI-AML-3) as they sufficiently reduced cell growth, genome methylation, and DNMT1 expression without inducing cytotoxic effects (Figure 3.1A, Figure S3.1).

Since previous studies in ovarian and colorectal cancer demonstrated that AZA leads to delayed ERE expression [23, 24], we sought to identify the time point post-AZA discontinuation with the highest ERE expression. Using THP-1 as a model, we performed RNA sequencing (RNA-seq) every 48h from day 3 to 11 post-AZA discontinuation. We then quantified the expression of multiple ERE transcripts upregulated by AZA [24]. We also assessed the expression of genes involved in dsRNA-induced interferon signaling in response to AZA-induced ERE expression [23] and observed that, together with ERE transcripts, they reached maximum expression around day 5 (72h after the last AZA treatment; Figure 3.1B). qPCR analyses further validated that the effect of AZA on ERE and dsRNA-induced immune response genes was maximal on day 4 (48h after the last AZA treatment; Figure S3.2). Thus, we treated the four AML cell lines with these optimal

AZA doses, administered three times at 24h intervals (0h, 24h, and 48h), and harvested the cells on day 4 to perform RNA-seq and mass spectrometry (MS) analyses.

3.5.2 Proteogenomic characterization of the immunopeptidome

We have previously demonstrated that a significant fraction of the AML immunopeptidome derives from non-exonic genomic regions such as introns, EREs, or intergenic regions [28]. Typically, MS search engines rely on reference protein databases (such as Uniprot) to match individual acquired MS/MS spectra to a peptide sequence. However, these databases do not contain non-exonic sequences, and building a personalized database containing all genomic sequences would generate an unmanageable database for search engines. We, therefore, used a proteogenomic approach to create MS databases containing only MAP sequences corresponding to the RNA transcripts expressed by our cell lines. These customized cell-line-specific MS databases had manageable sizes for the PEAKS search engine (Figure 3.2A and Figure S3.3).

Precisely, the RNA-seq reads were first chopped into shorter sequences (k-mers, of 24–33 nucleotides) corresponding to the different possible MAP lengths (8–11 residues). These k-mers were then filtered based on inter-sample sharing and abundance since abundant transcripts have higher chances of generating MAPs [28, 42] . K-mers possibly deriving from sequencing artifacts or lowly abundant polymorphisms were removed. The resulting k-mers were then *in silico* translated into peptide sequences, and those present in the canonical proteome were discarded. Finally, peptides were evaluated for their predicted capacity to bind MHC-I allotypes of their respective cell line, and binders were concatenated with the canonical proteome to generate MS databases ranging between 250 and 350 Mb (sizes tolerated by MS search engines [28]).

In parallel to MS, we performed a differential expression analysis of annotated protein-coding genes and ~4.2×10^6 ERE regions reported in the Repeatmasker annotations [53]. Differential abundance analyses were also performed on MS data to correlate transcript expression and MAP presentation. These analyses revealed that the number of elements differentially expressed by AZA-treated cells varied across cell lines, with THP-1 being the most sensitive (Figure 3.2B, Table S3.1). Overall, the proportion of differentially expressed MAPs (DEMs) was five times greater than that of differentially expressed transcripts (DEGs): 6–23% vs. 1.9–4.69%, respectively. Notably, the number of DEMs per cell line correlated almost perfectly with the number of DEGs rather than the total number of MAPs per cell line, suggesting that transcriptomic alterations were reflected in the immunopeptidome (Figure 3.2C). However, the directionality of differential expression differed for MAPs and transcripts. While most DEGs (>70%) were upregulated by AZA, this was not the case for DEMs (Figures 3.2D-E, Table S3.2). This means that as with other drugs [54], changes in the immunopeptidome post-AZA treatment result from differential mRNA expression and post-translational events.

3.5.3 AZA-induced EREs do not generate MAPs but trigger innate immune responses

Next, we focused on variations in CTA and ERE expressions at the mRNA and MAP levels. As expected, we observed a striking upregulation of ERE transcripts, representing ~38% of upregulated DEGs (Figure 3.2 F-G, Table S3.3). In contrast, EREs represented only 0.22% of upregulated DEMs, meaning that AZA-induced ERE transcripts were not processed adequately for MHC-I presentation. Due to the delay between AZA treatment and ERE induction (Figure 3.1B), we repeated our immunopeptidomic analyses at a later time point (day 7) on THP-1 (the cell line with the highest AZA-induced EREs) and did not observe higher ERE MAP presentation

(Figure S3.4). In contrast with EREs, we observed lower discrepancies in proportions of CTA-derived elements among upregulated DEGs and DEMs, and they were still observed on day 7. Among the upregulated DEMs, 152 (~10%) were AZA-specific (i.e., presented by all three AZA-treated replicates but undetected in control cells; Figure S3.5A). MAPs induced *de novo* by AZA contained CTAs but not ERE MAPs (Figure S3.5B). We conclude that at the immunopeptidomic level, AZA upregulates the expression of CTA MAPs (some of which are AZA-specific) but not ERE MAPs. The latter point was striking, considering the dramatic upregulation of ERE DEGs (38% of all DEGs) following AZA treatment.

Because EREs are remnants of ancient viruses, they can induce innate immune responses triggered by double-stranded RNA (dsRNA). Accordingly, we investigated whether AZA-induced EREs would trigger such responses. Gene ontology (GO) analysis performed on upregulated DEGs revealed that multiple innate immune responses and inflammatory responses were triggered in AZA-treated cells (Figure 3.2H). Specifically, we observed that OAS1, OAS2, OAS3, GBP1, and RIG-I – five genes instrumental in anti-dsRNA responses – were expressed at higher levels in AZA-treated cells than in controls (Figure 3.2I). Accordingly, using microscopy, we observed greater amounts of dsRNA in AZA-treated cells than in controls (Figure 3.2J). These data suggest that the dramatic upregulation of ERE transcripts induced by AZA leads to dsRNA formation, thereby triggering innate anti-viral immune responses.

3.5.4 AZA-induced EREs correlate with innate immune responses in primary AML

In cancer cells, innate immune responses benefit the host because they can initiate cancer cell apoptosis and increase their adjuvanticity [55]. We, therefore, asked two questions: i) what is the profile of EREs induced by AZA, and ii) can this profile be detected in primary AML samples? We first noted that ERE induction by AZA in AML cell lines was not random. Using ERE

distribution in the genome as a reference, we observed that AZA selectively upregulated two classes of EREs: LINEs and LTRs (Figure 3.3A). The fact that repression of SINEs depends mainly on histone methylation rather than DNA methylation [56] can explain why AZA did not induce SINE expression.

Using the previously published RNA-seq data of the Leucegene cohort (primary AML samples from 437 patients), we quantified the expression of the 506 ERE transcripts significantly upregulated by AZA in our AML cell lines. Patients were segregated based on their cumulative expression of these EREs, and patients expressing above-median levels were compared with those expressing below-median levels. DEGs and GO analyses revealed that high levels of AZA-induced EREs were associated with upregulated defense responses against viruses (Figure 3.3B). Furthermore, ERE expression levels significantly correlated with the expression of two critical dsRNA response regulators: RIG-I and MDA5 (Figure 3.3C), supporting the notion that AZA-induced EREs trigger innate immune responses *in vitro* and *in vivo*.

To complement our previous analysis, we performed GO analyses on genes downregulated by AML patients expressing high levels of AZA-induced EREs. This showed that multiple pathways controlling proliferation were downregulated, suggesting that high ERE expression (and anti-dsRNA response) could impact the growth of AML blasts (Figure 3.3D). Unexpectedly, we also observed that many GO terms related to protein degradation/catabolism and autophagy were significantly downregulated in these patients. Since ERE RNAs can trigger autophagy [57] and be degraded by the autophagic process [58], we hypothesized that enhanced autophagy in low-ERE expressing blasts could protect them from the deleterious effects that EREs have on their proliferation. This idea was explored in-depth in our next series of experiments.

3.5.5 AZA molds the immunopeptidome and induces protein aggregation through DNMT2 inhibition

We next sought to explore AZA-induced changes in the immunopeptidome for two reasons: First, because of the disconnect between the number of ERE DEGs and ERE DEMs (Figure 3.2F) and second, because our initial MAPs of interest (CTAs and EREs) represented only a fraction of AZA-induced MAPs (Figure 3.2F). To this end, we assessed the global impact of transcriptomic variations on the immunopeptidome using BamQuery, a computational tool that quantifies the RNA expression of any MAP of interest, including those derived from non-annotated genomic regions [51]. Most AZA-altered DEMs displayed no change at the RNA level (Figure 3.4A). Nevertheless, among DEMs coded by DEGs, RNA upregulation strongly associated with convergent upregulation of the corresponding DEMs (Figure 3.4A). This was not the case for downregulated transcripts. Moreover, fold changes in RNAs generating upregulated DEMs were significantly higher than those in downregulated DEMs (Figure 3.4B). We conclude that transcript upregulation has a modest but genuine impact on the abundance of corresponding DEMs.

We next focused on DEMs whose source RNA fold-change did not explain their immunopeptidomic fold-change (DEMs of Interest, DOIs). To gain insights into co- or post-translational events conducing to immunopeptidomic alterations in AZA-treated cells, we started by analyzing the residue composition of DOIs. We found that upregulated DOIs contained more polar residues than downregulated DOIs (Figure 3.4C). Accordingly, these MAPs presented a lower overall hydrophobicity (Figure 3.4D). Hydrophobic residues are the preferential cleavage sites of proteasomes, particularly immunoproteasomes [59]. MAP generation by constitutive proteasomes depends mainly on their tryptic and chymotryptic-like activities, and chymotryptic-like activity is further amplified in immunoproteasomes [60, 61]. Hence, we assessed how protease

activity contributes to the immunopeptidome by examining the C-terminal residue of each DOI and observed that upregulated DOIs derived more frequently from tryptic cleavage than downregulated DOIs (Figure 3.4E). Accordingly, AZA treatment significantly reduced immunoproteasome activity (Figure 3.4F, Figure S3.6A).

Typically, alterations in proteasomal activity are associated with disrupted protein homeostasis [62, 63]. To investigate whether alterations in protein homeostasis were responsible for perturbed proteasomal activity in AZA-treated cells, we examined the residue composition of proteins that generated DOIs. Assuming that proteins generating multiple DOIs were degraded more actively than those generating a single DOI, we correlated the number of generated DOIs for each protein with the frequency of each residue in the considered protein. This analysis showed that aspartic acid (Asp) and glycine (Gly) had the strongest positive correlation with the number of upregulated DOIs (Table S3.4). Interestingly, the tRNAs of Asp and Gly are stabilized by DNMT2, a tRNA-methyl transferase enzyme inhibited by AZA [64, 65]. A targeted analysis comparing the frequency of Asp, Gly, and valine (Val, the third amino acid whose tRNA is methylated by DNMT2) revealed that proteins generating more than three upregulated DOIs presented significantly higher cumulative frequencies of Asp, Gly, and Val than those generating less than three upregulated DOIs (Figure 3.4G). Demethylated tRNAs are susceptible to ribonuclease cleavage and fragmentation [66]. We, therefore, hypothesized that AZA-mediated DNMT2 inhibition results in an insufficiency of Asp, Gly, and Val tRNAs and leads to a decrease in protein synthesis, ribosomal stalling, and consequent protein aggregate generation during translation of proteins rich in the aforementioned residues. Accordingly, AZA-specific peptides (technically the most upregulated DOIs) were generated from proteins with a significantly lower proportion of aromatic residues, a feature often associated with less efficient protein folding

(Figure 3.4H) [67]. Finally, we experimentally quantified protein aggregates in AZA-treated cells and discovered that AZA induced protein aggregate accumulation in a dose-dependent manner (Figure 3.4I, Figure S3.6B). Importantly, DAC, a hypomethylating drug that does not inhibit DNMT2, did not induce the formation of protein aggregates.

3.5.6 Autophagy degrades AZA-induced EREs

Given the inverse correlation between autophagy-related GO terms and EREs in AML patients (Figure 3.3B) and the generation of protein aggregates by AZA (Figure 3.4I), we evaluated whether AZA induced autophagy. Twenty-four hours of treatment with AZA resulted in a dose-dependent induction of autophagy (Figure 3.5A). Interestingly, this was not observed with DAC, suggesting that autophagy induction is dependent on protein aggregates generation resulting from DNMT2 inhibition. As EREs were previously suggested to trigger autophagy [57], we evaluated whether DAC induced the same EREs as AZA. Using data reported by Pappalardi *et al*. [68], we observed that DAC expressed AZA-induced EREs in a dose-dependent manner (Figure 3.5B). This suggests that the autophagy induced in AZA-treated (but not in DAC-treated) cells results from DNMT2 inhibition rather than ERE induction. Nevertheless, we observed an inverse correlation between EREs and two well-established autophagy markers, ATG3 and SQSTM1, which was already observed in AML patients at diagnosis (Figure 3.5C). The latter finding suggests that autophagy does not need to be induced by AZA to degrade EREs.

Next, we sought to verify this autophagy-dependent degradation of EREs. We treated THP-1 cells for 72h with AZA alone or AZA combined with an autophagy inducer (rapamycin) or inhibitor (spautin-1). Ninety-six hours after the initiation of the treatment, the levels of dsRNAs were examined by fluorescence microscopy. As shown in Figure 3.5D, autophagy inhibition significantly increased the levels of dsRNAs compared to AZA alone, while autophagy activation

decreased them. Altogether, these results demonstrate that autophagy contributes to the degradation of EREs.

3.5.7 Autophagy inhibition synergizes with AZA and could increase AML immunogenicity

Finally, we examined whether inhibiting autophagy would augment the anti-AML effect of AZA. THP-1 and OCI-AML-3 cells were treated for 72h with AZA and/or spautin-1 (which inhibits autophagy and suppresses the unfolded protein response [69]), and their survival and cell counts were evaluated 24h after discontinuing treatment. A synergistic effect between AZA and spautin-1 was observed for proliferation and cell death (Figure 3.6A, Figure S3.7A). While spautin-1 alone reduced proliferation, it did not kill the cells (Figure S3.7B, C), suggesting that autophagy acts as a survival mechanism upon AZA treatment.

Next, we investigated the molecular effects of autophagy inhibition combined with AZA. THP-1 cells treated with low AZA doses and/or IC50 concentrations of spautin-1 were analyzed by RNA-seq, and as expected, we observed that spautin-1 alone tended to increase the levels of all three ERE families (Figure 3.6B). Furthermore, ERE levels tended to be higher (except for LINEs) when spautin-1 was combined with AZA. In contrast, AZA-induced EREs in these cells were at the same levels as in control cells (Figure S3.8B), possibly due to the drastic inhibition of proliferation (Figure S3.8A) mediated by the combined treatment (AZA's DNA demethylating activity relies on active cell division for AZA to integrate into genomic DNA). Nevertheless, a pathway enrichment analysis on genes significantly upregulated in AZA+Spautin-1 vs. AZA-treated cells revealed the enrichment of multiple pathways related to antigen presentation, antiviral mechanisms, and noncanonical NF-κB signaling (Figure S3.8C). Targeted gene set enrichment

analysis (GSEA) analyses further validated the enrichment of these three processes (Figure 3.6C), supporting the rationale of inhibiting autophagy to improve AZA immune effects.

Given that autophagy degrades EREs, we wondered if autophagy could prevent AML blasts from being recognized by CTLs. Therefore, we segregated the 437 Leucegene AML patients based on two parameters. The first parameter was the expression of CD8A and CD8B transcripts, a reliable marker of CTL abundance in RNA-seq datasets [70]. Second, the count of highly expressed AZA-induced EREs (HE-EREs), i.e., the number of EREs whose expression is above their median RNA expression across all patients having a non-null expression of the given ERE (a metric aimed at reflecting the density and diversity of epitopes possibly presented by leukemic blasts [28]) (Figure 3.6D). A GSEA comparing $CD8^{high}$ vs. $CD8^{low}$ patients within ERE^{high} and ERE^{low} groups revealed that the presence of CTLs was associated with the same processes in ERE^{high} and ERE^{low} patients, except for two gene sets related to DNA repair/proliferation, three related to metabolism, and two related to unfolded protein response and mTORC1 signaling (Figure 3.6E). As mTORC1 regulates autophagy, we performed additional GSEAs with four gene sets related to autophagy from the REACTOME database. We found that all four were inversely associated with the presence of CD8 T cells in ERE^{high} patients (two significantly), while the opposite was found for ERE^{low} patients (Figure 3.6F). Altogether, these results further support that autophagy prevents the generation of ERE-derived MAPs, thereby precluding AML blasts from being recognized by $CD8^{+}$ T cells.

3.6 Discussion

Due to their dual capacity to trigger innate immune responses and generate highly immunogenic MAPs [53], EREs represent attractive targets for developing new immunotherapeutic avenues [25, 71, 72]. Although AZA has been proposed to promote anti-tumor CTL responses through the induction of CTA MAPs, the contribution of ERE MAPs to such responses remains elusive. Aiming to unravel this contribution, we performed a thorough proteogenomic investigation to uncover changes in the MAP repertoire after AZA treatment. As expected, we identified CTAs upregulated at the transcriptomic and immunopeptidomic levels. In contrast, ERE MAP abundance remained unchanged after AZA treatment in the four cell lines examined, suggesting that T-cell-mediated responses post-AZA treatment are more likely due to the recognition of CTA-derived MAPs than EREs. A recent report analyzing the $CD8^+$ T-cell subsets targeting ERE-derived MAPs revealed no increase in ERE reactive T cells post-AZA treatment in myeloid hematological malignancies, further supporting our observations [73].

The virtual absence of ERE MAP induction by AZA was paradoxical. Indeed, AZA strongly induced ERE transcripts (Figure 3.2F), and the processing of numerous EREs should generate MAPs [53]. In AML patients, the basal ERE expression was positively associated with the expression of molecules involved in dsRNA detection and anti-viral immune responses. In a previous report, high ERE expression in primary AML cells was associated with a favorable prognosis [74]. In addition, Ohtani *et al.* demonstrated that clinical responses to AZA in myelodysplastic syndrome and AML were associated with the expression of a specific class of EREs inducing innate immune responses [75]. Therefore, elevated ERE expression certainly exerts a beneficial effect on patients' outcomes by inducing anti-dsRNA immune responses, and maximizing these responses should be pursued.

Notably, we observed that ERE levels (and associated innate immune responses) were inversely correlated to the expression of autophagy molecules in AML patients, and enhanced autophagy was found in our AZA-treated cells. Upon examination of the immunopeptidomic changes, we could attribute this latter observation to AZA's inhibition of DNMT2 activity. Indeed, previous studies have shown that tRNA methylation by DNMT2 is involved in Asp-tRNA codon fidelity, and its loss leads to the production of misfolded proteins [76]. DNMT2 inhibition is a property of AZA but not DAC. Accordingly, protein aggregation and concomitant autophagy responses were observed in AZA- but not DAC-treated AML cells. Autophagy is increasingly implicated in resistance to anti-cancer therapies, including resistance to AZA [77]. We propose that autophagy mitigates tumor immunogenicity by preventing ERE MAP presentation. Interestingly, an article exploring the immunopeptidomic effects of DAC in glioblastoma cell lines evidenced an induction of ERE MAPs following similar treatment conditions to ours [78]. Since DAC also induces ERE transcripts but does not induce autophagy, in contrast to AZA, we hypothesize that the autophagic process triggered by DNMT2 inhibition is responsible for the absence of ERE MAP induction following AZA treatment.

While little is known about the interplay between EREs (remnants of ancient viruses) and autophagy, it is well-reported that autophagy is a defense response against viruses. Following infection, autophagy is triggered by the signaling of pattern-recognition receptors (such as Toll-like and RIG-like receptors) [79] to sustain the presentation of intracellular source proteins by MHC-II molecules [80]. Thereby, autophagy inhibition reduces the presentation of viral MAPs by infected cells to $CD4^+$ T cells [81, 82]. In contrast, autophagy inhibition tends to increase MHC-I expression and the capacity to induce antiviral $CD8^+$ T cell responses [83, 84]. Furthermore, autophagy can limit RIG-I-dependent IFN production by disrupting its signaling cascade [85, 86].

While these studies point to a potential negative correlation between autophagy and MHC-I presentation, they do not provide a precise molecular mechanism explaining how AZA-induced autophagy could prevent the generation of ERE MAPs. While this question will need to be explored in future studies, we surmise that autophagy degrades ERE RNAs instead of degrading their translational product. Indeed, autophagy has been reported to target viral [87] and ERE [57] dsRNA to autophagosomes. Notably, this would explain why CTA MAPs (which do not result from dsRNA) were successfully presented at higher levels after AZA treatment.

In conclusion, our results demonstrate that AZA-induced autophagy mitigates the ERE-dependent immune effects of AZA. They suggest that autophagy inhibition could be a desirable therapeutic option to combine with AZA. Adding autophagy inhibitors to AZA could have three desirable consequences: (1) to increase the direct cytotoxicity of AZA by preventing AML adaptation to proteotoxic stress (in agreement with results from [88]), (2) to increase ERE transcripts abundance and the subsequent beneficial anti-dsRNA innate immune responses and (3) to improve ERE MAP presentation and thereby adaptive T-cell responses. Further investigations will nevertheless be needed to verify this last hypothesis.

3.7 Acknowledgements

The authors wish to thank Anca Apavaloaei, Jean-David Larouche, and Maria Virginia Ruiz Cuevas for valuable discussions and suggestions. We also want to thank Caroline Côté, Jalila Chagraoui, and Eugenie Goupil for technical assistance and advice regarding autophagy experiments and immunofluorescence analyses and Jeremy Zumer for advice regarding normalization methods of the mass spectrometry data. We are grateful to the IRIC genomics core

facility, flow cytometry platform, Christian Charbonneau from the Microscopy facility, and Patrick Gendron from the IRIC bioinformatics platform. This study was supported by grants from the Canadian Institutes of Health Research (FDN 148400) and the Canadian Cancer Society (#705604). N.N. is supported by doctoral studentships from the IRIC and the Fonds de Recherche du Québec – Santé. J.H. is supported by post-doctoral fellowships from the Cole Foundation and the Power Corporation of Canada. BES and JC are research fellows. G.E. is supported by post-doctoral fellowships from the IRIC, FRQS, The Cole Foundation, and the FNRS. FB senior research associate at the FNRS Belgium.

3.8 Figures

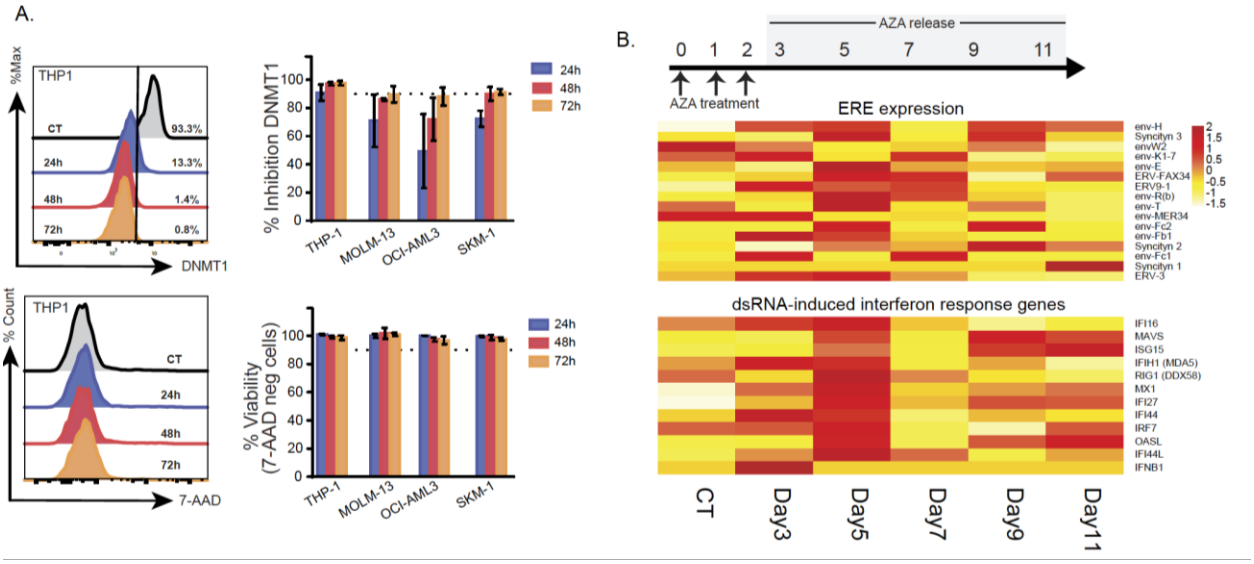

Figure 3.1- Low-dose AZA treatment leads to delayed, transient ERE and dsRNA-induced interferon gene expression in AML cell lines

A, Low AZA doses were added to four AML cell lines daily for three days (0.25 μM: MOLM-13 and SKM-1; 0.5 μM: THP-1 and OCI-AML-3) and DNMT1 inhibition (upper panel) and cell viability were monitored by flow cytometry using 7-AAD (lower panel). Dotted lines represent 90% DNMT1 inhibition/viability in the upper and lower panel respectively. The left panels depict representative histograms of THP-1 cells, while the right panels depict bar plots summarizing the percentage of expression/staining (as indicated in figures) of all four AML cells. Percentages were calculated by comparing AZA-treated cells to the control cells. **B**, Low AZA doses (0.5 μM) were added to THP-1 cells daily for three days, after which AZA treatment was released by washing and expanding cells in the absence of AZA. Cells were collected for RNA-seq at the time points highlighted in grey. Heatmap comparing RNA expression levels of EREs induced by AZA (upper panel) and genes involved in dsRNA-induced viral response (lower panel) identified in previous studies [23, 24] of control (day 3) and AZA-released cells at the time points indicated.

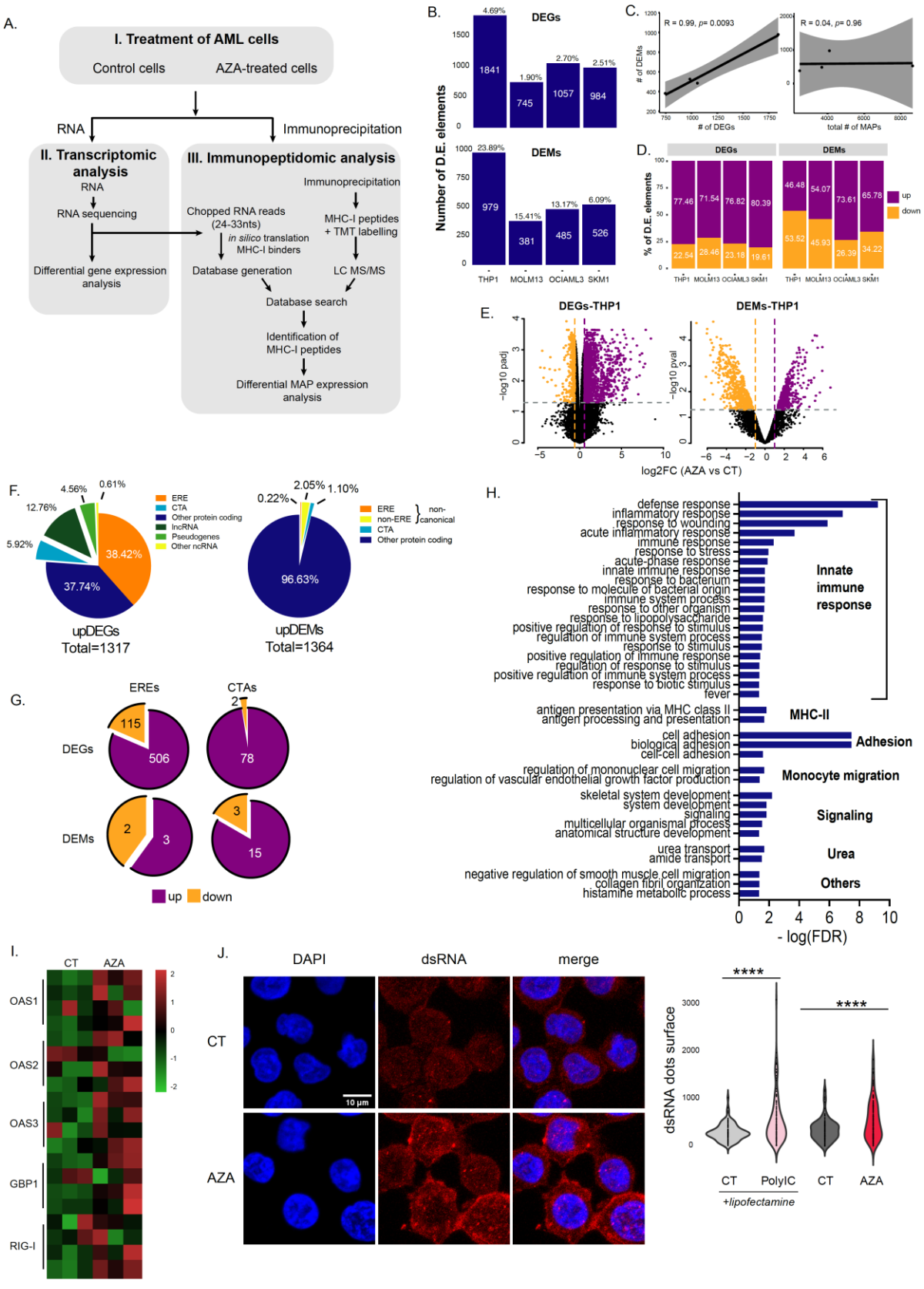
A.
I. Treatment of AML cells
Control cells
AZA-treated cells
RNA
Immunoprecipitation
II. Transcriptomic analysis
RNA
RNA sequencing
Differential gene expression analysis
III. Immunopeptidomic analysis
Immunoprecipitation
Chopped RNA reads (24-33nts)
in silico translation MHC-I binders
MHC-I peptides + TMT labelling
Database generation
LC MS/MS
Database search
Identification of MHC-I peptides
Differential MAP expression analysis
B.
DEGs
DEMs
Number of D.E. elements
4.69%
1841
1.90%
745
2.70%
1057
2.51%
984
23.89%
979
15.41%
381
13.17%
485
6.09%
526
THP1 MOLM13 OCIAML3 SKM1
C.
R = 0.99, p= 0.0093
# of DEMs
# of DEGs
R = 0.04, p= 0.96
total # of MAPs
D.
DEGs
DEMs
% of D.E. elements
77.46
71.54
76.82
80.39
22.54
28.46
23.18
19.61
46.48
54.07
73.61
65.78
53.52
45.93
26.39
34.22
up
down
E.
DEGs-THP1
DEMs-THP1
−log10 padj
−log10 pval
log2FC (AZA vs CT)
F.
12.76%
4.56%
0.61%
5.92%
38.42%
37.74%
ERE
CTA
Other protein coding
lncRNA
Pseudogenes
Other ncRNA
upDEGs
Total=1317
2.05%
1.10%
0.22%
96.63%
ERE
non-ERE
CTA
Other protein coding
non-canonical
upDEMs
Total=1364
G.
EREs
CTAs
DEGs
115
506
2
78
DEMs
2
3
3
15
up
down
H.
defense response
inflammatory response
response to wounding
acute inflammatory response
immune response
response to stress
acute-phase response
innate immune response
response to bacterium
response to molecule of bacterial origin
immune system process
response to other organism
response to lipopolysaccharide
positive regulation of response to stimulus
regulation of immune system process
response to stimulus
positive regulation of immune response
regulation of response to stimulus
positive regulation of immune system process
response to biotic stimulus
fever
Innate immune response
antigen presentation via MHC class II
antigen processing and presentation
MHC-II
cell adhesion
biological adhesion
cell-cell adhesion
Adhesion
regulation of mononuclear cell migration
regulation of vascular endothelial growth factor production
Monocyte migration
skeletal system development
system development
signaling
multicellular organismal process
anatomical structure development
Signaling
urea transport
amide transport
Urea
negative regulation of smooth muscle cell migration
collagen fibril organization
histamine metabolic process
Others
- log(FDR)
I.
CT
AZA
OAS1
OAS2
OAS3
GBP1
RIG-I
J.
DAPI
dsRNA
merge
CT
AZA
10 µm
dsRNA dots surface
****
****
CT
PolyIC
CT
AZA
+lipofectamine

Figure 3.2 Proteogenomic characterization of AZA-mediated changes shows upregulation of MAPs derived from CTA but not from EREs

A, Schematic representation of the study design for RNA-seq and MS analyses. **B**, The total number of differentially expressed (D.E.) genes and MAPs (DEGs and DEMs, respectively) varies across cell lines. The numbers above the bars indicate the percentage of total genes or MAPs that were DEGs or DEMs, respectively. **C**, Pearson correlation between the number of DEMs and DEGs (left panel) or DEMs and MAPs (right panel). Each dot corresponds to a cell line. **D**, Percentage of D.E. elements up- or downregulated across cell lines. **E**, Representative volcano plots of DEGs and DEMs between AZA (violet) and untreated (gold) THP-1 cells. **F**, Pie charts depicting the percentage of biotypes of upregulated transcripts (left) and MAPs (right). The total number of upregulated D.E. elements is indicated below the pie charts. **G**, Pie charts depicting the number of up- and downregulated EREs (left) and CTAs (right) belonging to DEGs (upper panel) or DEMs (lower panel) fractions. **H**, Histogram depicting GO-term analysis of the most significantly enriched biological processes associated with upregulated DEGs across all cell lines after AZA treatment. **I**, Heatmap of genes involved in anti-dsRNA responses. **J**, Representative images (left) and quantification (right) of dsRNA signals in THP-1 cells from microscopy images (two independent experiments). THP-1 cells transfected with 10 μg/ml polyinosinic:polycytidylic acid using lipofectamine (PolyIC+lipofectamine) were used as a positive control and were compared with cells treated with lipofectamine alone (CT+lipofectamine); (unpaired t-test; **** $p < 0.0001$).

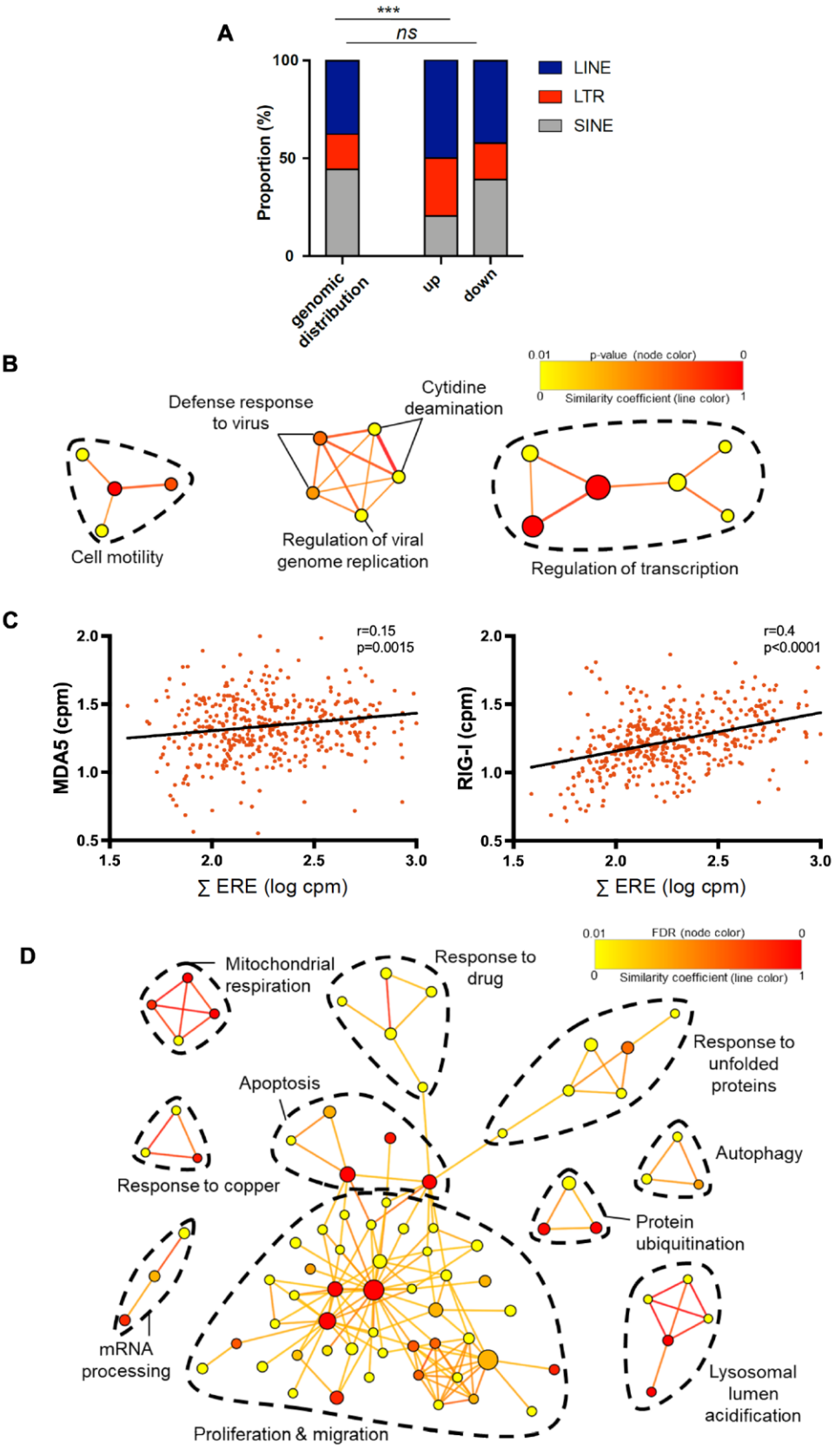
A
***
ns
100
50
0
Proportion (%)
LINE
LTR
SINE
genomic distribution
up
down
B
0.01
p-value (node color)
0
0
Similarity coefficient (line color)
1
Defense response to virus
Cytidine deamination
Cell motility
Regulation of viral genome replication
Regulation of transcription
C
r=0.15
p=0.0015
MDA5 (cpm)
∑ ERE (log cpm)
r=0.4
p<0.0001
RIG-I (cpm)
∑ ERE (log cpm)
D
0.01
FDR (node color)
0
0
Similarity coefficient (line color)
1
Mitochondrial respiration
Response to drug
Response to unfolded proteins
Apoptosis
Response to copper
Autophagy
Protein ubiquitination
mRNA processing
Proliferation & migration
Lysosomal lumen acidification

Figure 3.3 AZA-induced EREs trigger innate immune responses

A, Stacked bar plots of ERE group distribution at the genomic and transcriptomic levels for up- and downregulated AZA-altered ERE sequences. **B**, Network analysis of GO-terms in AML patients (Leucegene cohort; n=437) expressing high levels of EREs induced by AZA in our four cell lines. **C**, Pearson correlation between genes involved in anti-dsRNA responses (MDA-5 and RIG-I) and AZA-induced EREs in AML patients. **D**, Network analysis of GO-terms enriched in patients expressing low levels of AZA-induced EREs. In B, and D, the line color reflects the similarity coefficient between connected nodes. Node color reflects the false discovery rate (FDR) of the enrichment. Node size is proportional to gene set size.

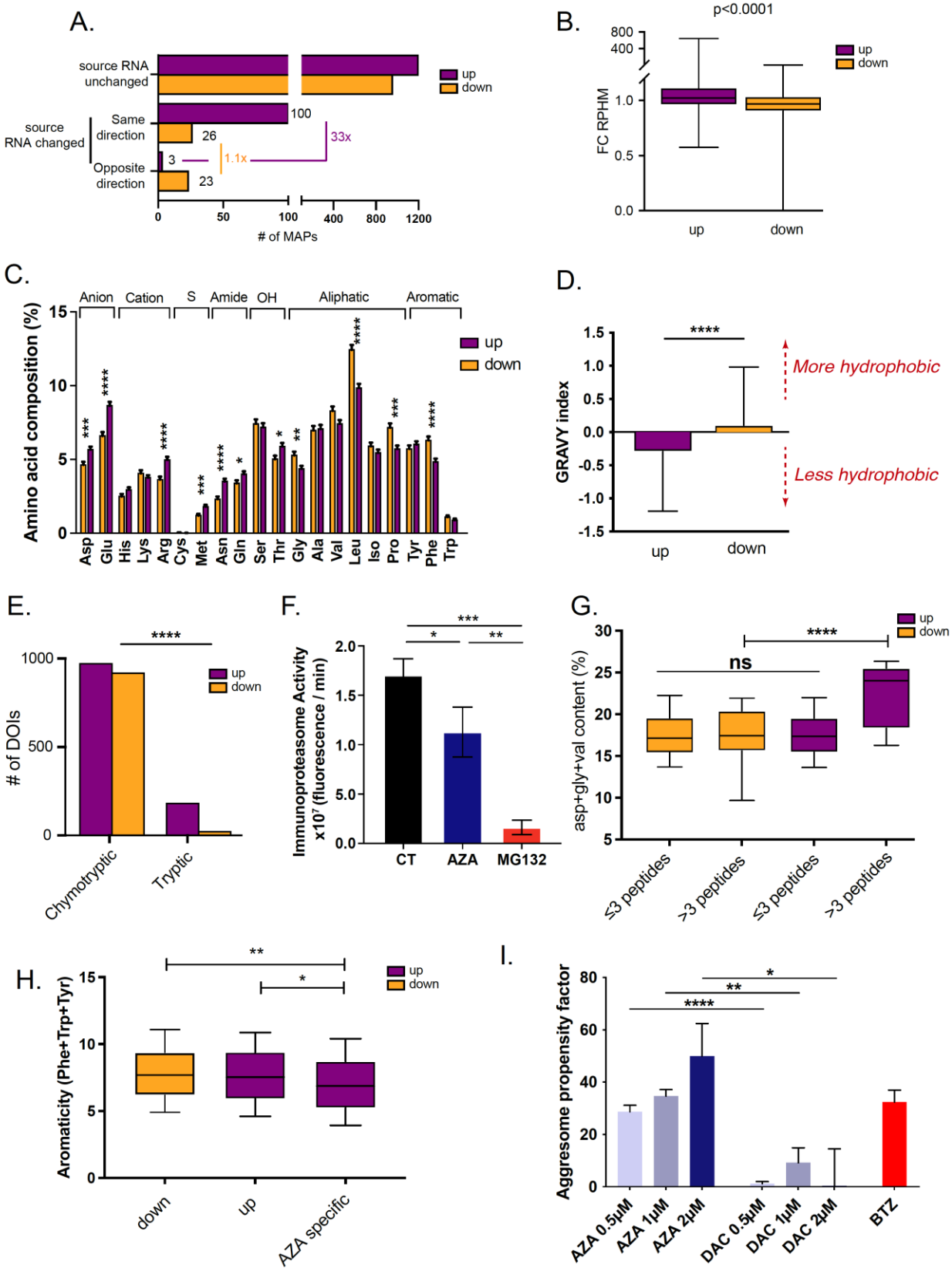
A.
source RNA unchanged
source RNA changed
Same direction
Opposite direction
100
26
3
23
33x
1.1x
0
50
100
400
800
1200
# of MAPs
up
down
B.
p<0.0001
FC RPHM
800
400
1.0
0.5
0.0
up
down
C.
Anion
Cation
S
Amide
OH
Aliphatic
Aromatic
Amino acid composition (%)
Asp
Glu
His
Lys
Arg
Cys
Met
Asn
Gln
Ser
Thr
Gly
Ala
Val
Leu
Iso
Pro
Tyr
Phe
Trp
D.
****
GRAVY index
More hydrophobic
Less hydrophobic
E.
# of DOIs
Chymotryptic
Tryptic
F.
Immunoproteasome Activity x10^7 (fluorescence / min)
CT
AZA
MG132
G.
asp+gly+val content (%)
ns
≤3 peptides
>3 peptides
H.
Aromaticity (Phe+Trp+Tyr)
down
up
AZA specific
I.
Aggresome propensity factor
AZA 0.5μM
AZA 1μM
AZA 2μM
DAC 0.5μM
DAC 1μM
DAC 2μM
BTZ

Figure 3.4 AZA molds the immunopeptidome through DNMT2 inhibition

A, Bar plots depicting the number of up- and downregulated DEMs with no change in source RNA expression or being altered in the same or opposite direction as their source RNA. **B**, Fold changes in RPHM expression of source transcripts generating up- and downregulated DEMs. **C**, Amino-acid composition of up- and downregulated DOIs. **D**, Hydrophobicity of up- and downregulated DOIs assessed by their GRAVY index. Scores >0 reflect higher hydrophobicity. **E**, Number of DOIs associated with chymotryptic or tryptic activities based on their C-termini amino acid composition (Fisher's exact test). **F**, Immunoproteasome activity in OCI-AML-3 cells after AZA treatment. MG132, a proteasome inhibitor, was used as a negative control (unpaired t-test). **G**, Proportion of DNMT2-target amino acids (glycine, valine, and aspartic acid) in proteins having generated MAPs among the up- or downregulated DEM fractions. **H**, Aromaticity (frequency of Phe, Trp, and Tyr residues) of proteins having generated AZA-specific (identified only in AZA condition) MAPs or up- and downregulated DEMs. **I**, Quantification of protein aggregates induced with increasing concentrations of AZA and DAC in OCI-AML-3 cells. Bortezomib (BTZ) was used as a positive control (unpaired t-test; **** $p< 0.0001$, *** $p< 0.001$, ** $p< 0.01$, * $p< 0.05$).

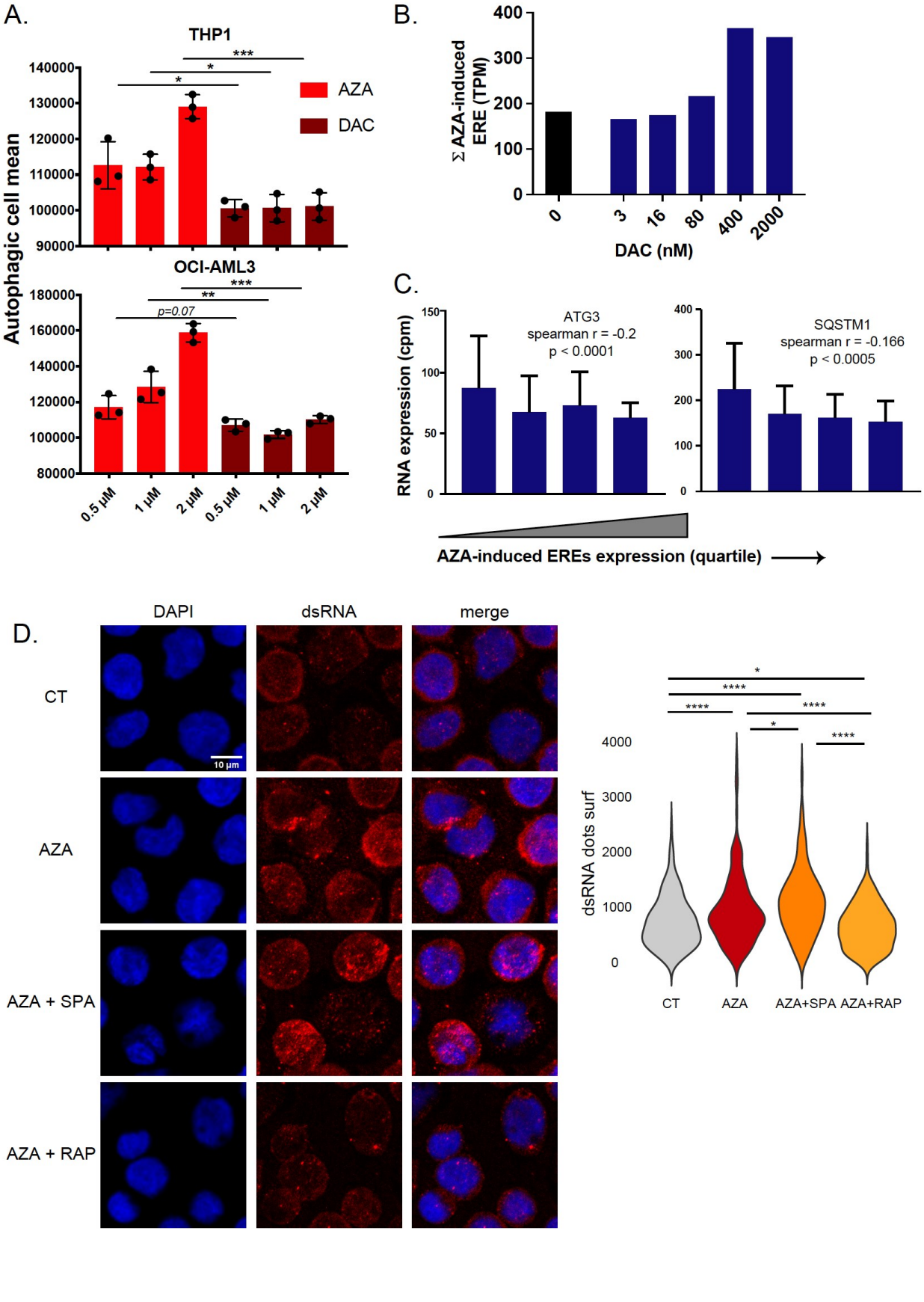
A.
THP1
OCI-AML3
Autophagic cell mean
AZA
DAC
p=0.07
0.5 µM
1 µM
2 µM
B.
Σ AZA-induced ERE (TPM)
DAC (nM)
C.
ATG3
spearman r = -0.2
p < 0.0001
SQSTM1
spearman r = -0.166
p < 0.0005
RNA expression (cpm)
AZA-induced EREs expression (quartile)
D.
DAPI
dsRNA
merge
CT
AZA
AZA + SPA
AZA + RAP
10 µm
dsRNA dots surf
AZA+SPA
AZA+RAP

Figure 3.5 Autophagy degrades AZA-induced EREs

A, THP-1 and OCI-AML-3 cells undergoing autophagy were assessed with increasing doses of AZA and DAC by flow cytometry using specific autophagy detection fluorescent probes. **B**, EREs induced by AZA in our cell lines were quantified in published RNA-seq data of DAC-treated THP-1 cells [89]. **C**, Bar plots of mean RNA expression (in cpm) of key autophagy genes (ATG3 and SQSTM1) in Leucegene AML patients segregated into quartiles based on AZA-induced ERE expression. Spearman correlations were computed without this segregation. **D**, Representative images (left) and quantification (right) of dsRNA signals measured from microscopy images. CT: 0.1% DMSO treated with lipofectamine; AZA+SPA: AZA and spautin-1; AZA+RAP: AZA and rapamycin (unpaired t-test; * $p< 0.05$, **** $p< 0.0001$

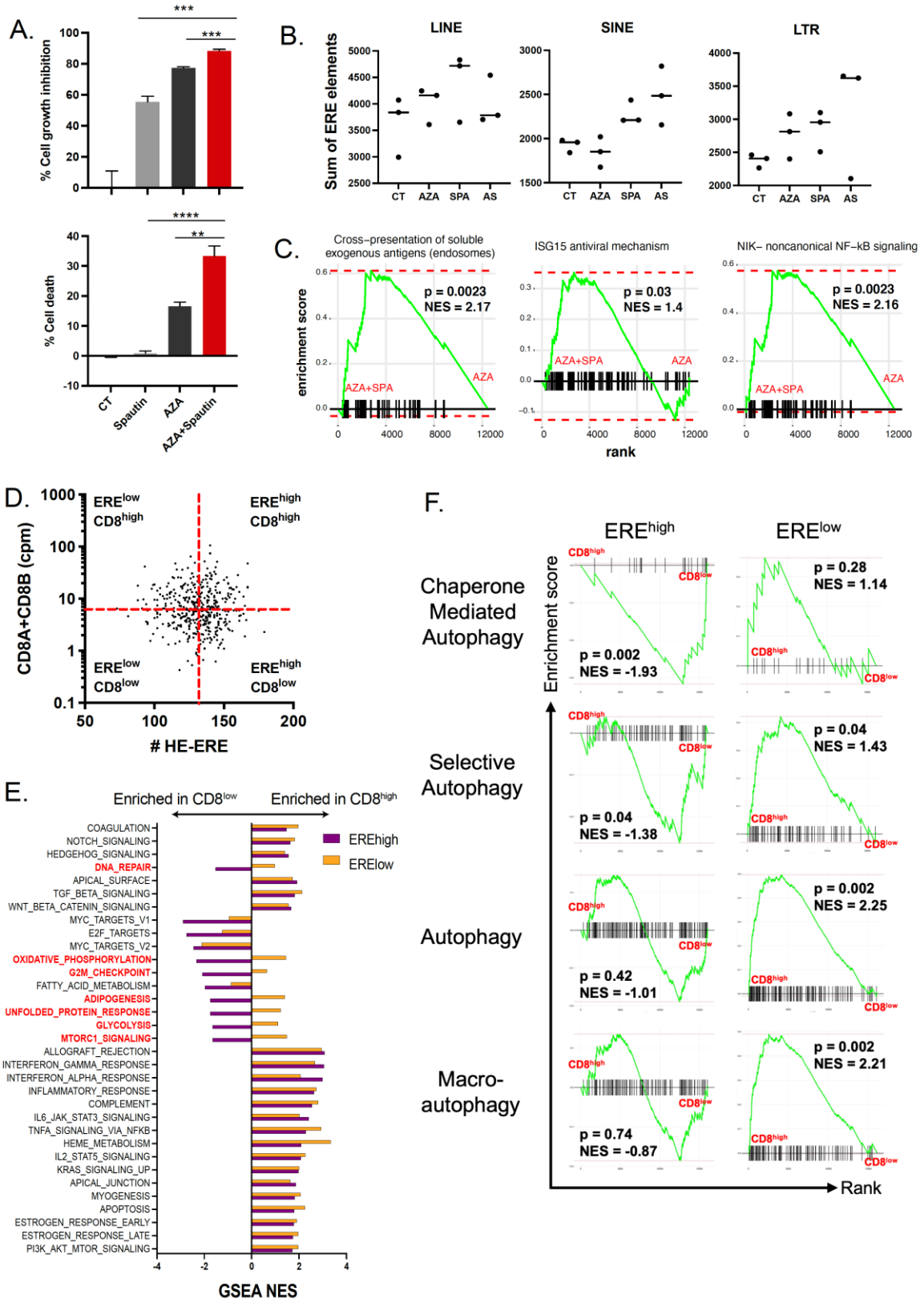
A.
% Cell growth inhibition
% Cell death
CT
Spautin
AZA
AZA+Spautin
B.
LINE
SINE
LTR
Sum of ERE elements
CT AZA SPA AS
C.
Cross–presentation of soluble exogenous antigens (endosomes)
ISG15 antiviral mechanism
NIK– noncanonical NF–kB signaling
enrichment score
p = 0.0023
NES = 2.17
p = 0.03
NES = 1.4
p = 0.0023
NES = 2.16
AZA+SPA
AZA
rank
D.
ERElow CD8high
EREhigh CD8high
ERElow CD8low
EREhigh CD8low
CD8A+CD8B (cpm)
HE-ERE
E.
Enriched in CD8low
Enriched in CD8high
EREhigh
ERElow
COAGULATION
NOTCH_SIGNALING
HEDGEHOG_SIGNALING
DNA_REPAIR
APICAL_SURFACE
TGF_BETA_SIGNALING
WNT_BETA_CATENIN_SIGNALING
MYC_TARGETS_V1
E2F_TARGETS
MYC_TARGETS_V2
OXIDATIVE_PHOSPHORYLATION
G2M_CHECKPOINT
FATTY_ACID_METABOLISM
ADIPOGENESIS
UNFOLDED_PROTEIN_RESPONSE
GLYCOLYSIS
MTORC1_SIGNALING
ALLOGRAFT_REJECTION
INTERFERON_GAMMA_RESPONSE
INTERFERON_ALPHA_RESPONSE
INFLAMMATORY_RESPONSE
COMPLEMENT
IL6_JAK_STAT3_SIGNALING
TNFA_SIGNALING_VIA_NFKB
HEME_METABOLISM
IL2_STAT5_SIGNALING
KRAS_SIGNALING_UP
APICAL_JUNCTION
MYOGENESIS
APOPTOSIS
ESTROGEN_RESPONSE_EARLY
ESTROGEN_RESPONSE_LATE
PI3K_AKT_MTOR_SIGNALING
GSEA NES
F.
EREhigh
ERElow
Chaperone Mediated Autophagy
Selective Autophagy
Autophagy
Macro-autophagy
Enrichment score
p = 0.002
NES = -1.93
p = 0.28
NES = 1.14
p = 0.04
NES = -1.38
p = 0.04
NES = 1.43
p = 0.42
NES = -1.01
p = 0.002
NES = 2.25
p = 0.74
NES = -0.87
p = 0.002
NES = 2.21
Rank

Figure 3.6 Autophagy inhibition synergizes with AZA and could increase AML immunogenicity

A, Cell growth inhibition and cell death (7-AAD) of THP-1 cells treated either with AZA (1 μM), spautin-1 (5 μM), or both. Control cells were treated with 0.1% DMSO; two independent experiments (unpaired t-test; **** $p < 0.0001$, *** $p < 0.001$, ** $p < 0.01$). **B**, Sum of ERE transcripts (in cpm) separated into LINE, SINE, and LTR families in THP-1 cells, treated as indicated. CT: 0.1% DMSO; SPA: spautin-1; AS: AZA and spautin-1. Horizontal bars represent median values for each condition **C**, GSEA comparison of indicated gene sets between THP-1 cells treated with AZA or AZA + spautin-1. NES, normalized enrichment score. **D**, Scatterplots of Leucegene AML patients based on their RNA expression of CD8A+CD8B (cpm) vs. their count of highly expressed AZA-induced EREs (HE-EREs: # of AZA-induced EREs whose expression is above their median expression across all patients). **E**, Bar plots indicating normalized enrichment scores for hallmark gene sets between $CD8^{high}$ vs. $CD8^{low}$ Leucegene patients within ERE^{high} and ERE^{low} groups (defined in D panel). **F**, GSEA of the indicated REACTOME gene sets in the indicated comparisons. NES, normalized enrichment score.

3.9 Supplementary Figures and Tables

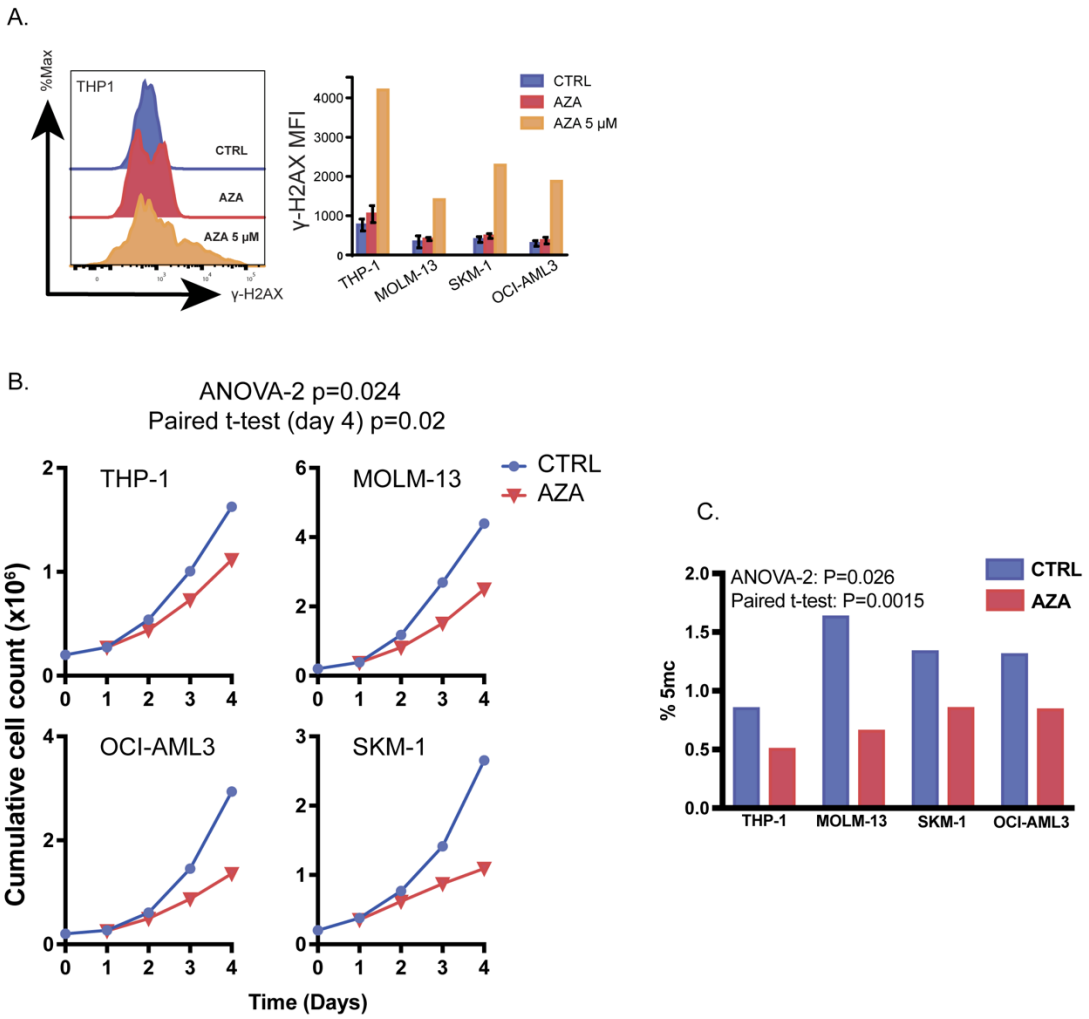

Figure S 3.1 Low AZA treatment reduces DNA methylation and cell growth without inducing major cytotoxic effects from DNA damage in AML cell lines

A, Low AZA doses were added to four AML cell lines daily for three days (0.25 µM: MOLM-13 and SKM-1; 0.5 µM: THP-1 and OCI-AML-3), and the formation of DNA double-strand breaks was monitored by flow cytometry by measuring histone H2AX phosphorylation. The left panels depict representative histograms of THP-1 cells, while the right panels depict bar plots summarizing the percentage of expression of all four AML cells. Percentages were calculated by comparing AZA-treated cells to control cells. A high AZA dose (5 µM) was used as a positive control for double-strand break formation.

B, Cell growth of four AML cell lines was monitored after AZA treatment by counting 7-AAD negative cells via flow cytometry

C, 5-methylcytosine levels measured by ELISA with the MethylFlash Global DNA Methylation Kit after AZA treatment in AML cell lines.

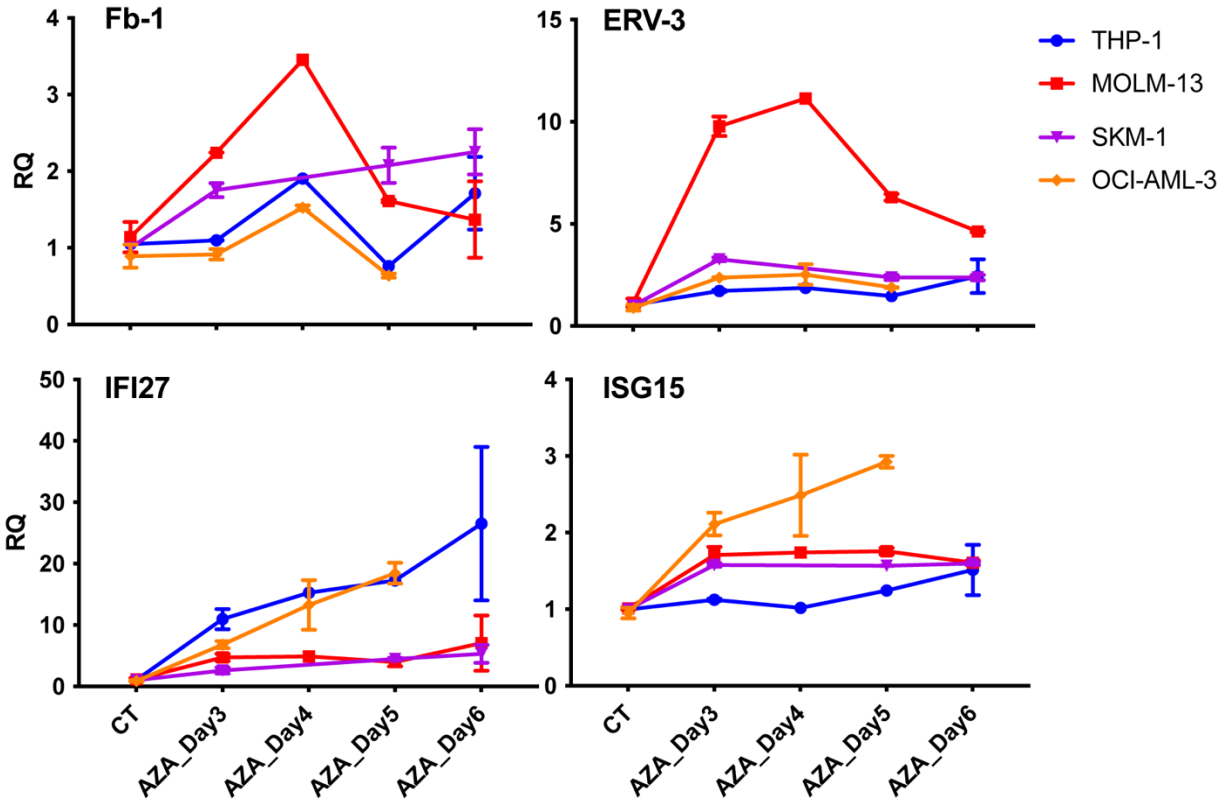

Figure S 3.2 Low AZA treatment leads to delayed, transient ERE and dsRNA-induced pro-inflammatory gene expression in AML cell lines

Relative quantification levels of ERE (upper panel) and dsRNA-induced interferon (lower panel) gene candidates in four AML cell lines monitored by qPCR after AZA treatment for three days, followed by AZA discontinuation.

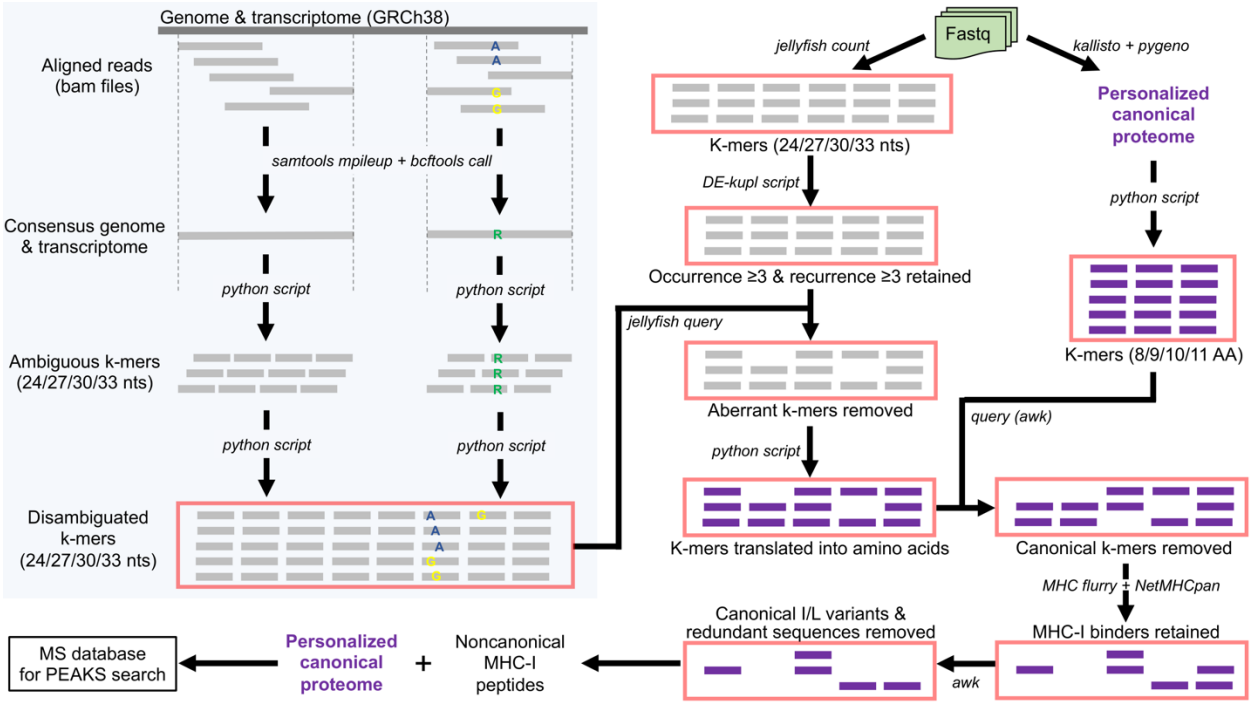

Figure S 3.3 Detailed proteo-genomic pipeline used for database generation for MS analyses

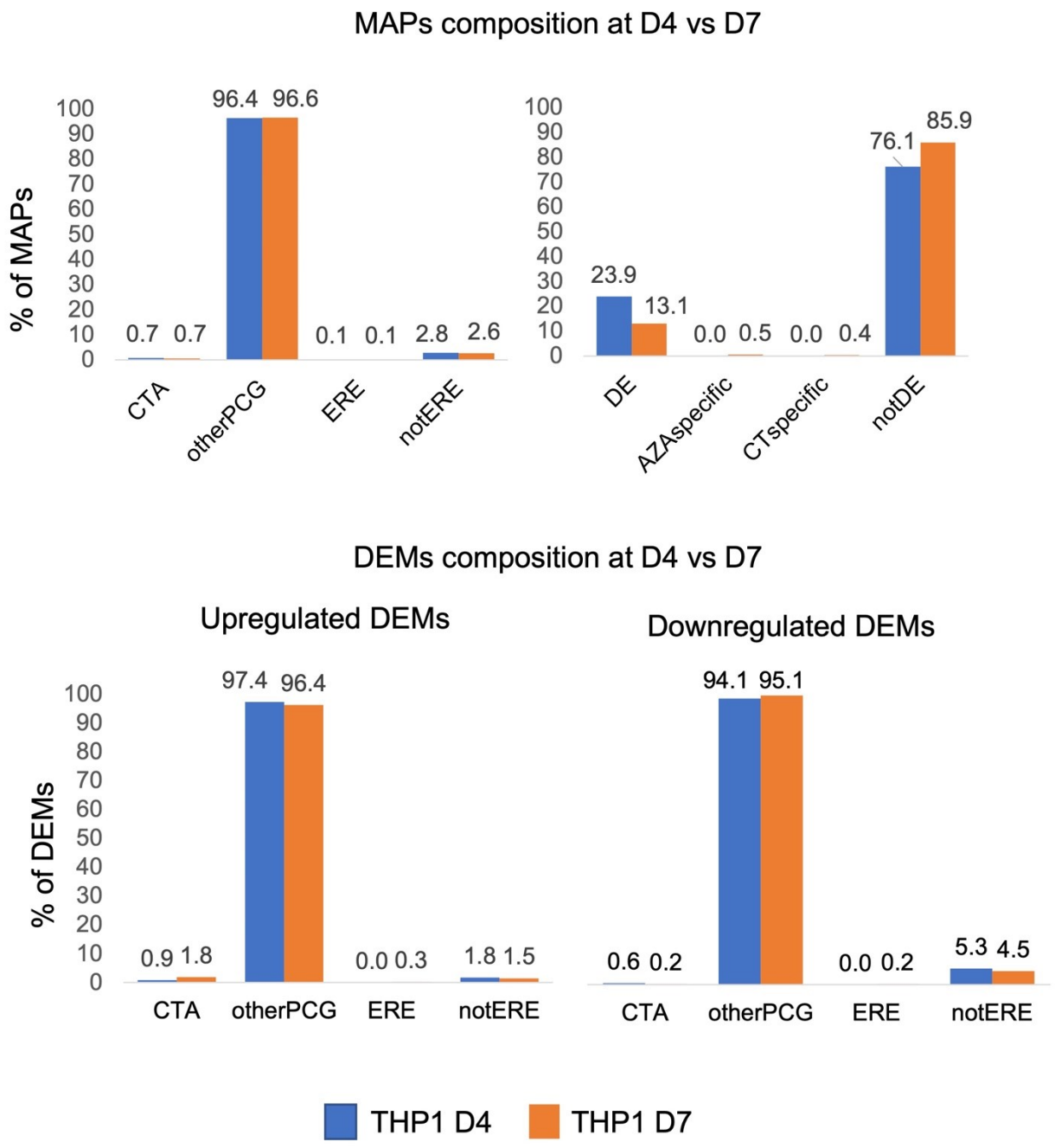

Figure S 3.4 Immunopeptidomic analyses at a later time-point reveal no increase in ERE-derived DEMs

Comparison of MAP (upper panels) and DEM (lower panels) composition on days 4 and 7. OtherPCG: other protein coding genes, DE: differentially expressed.

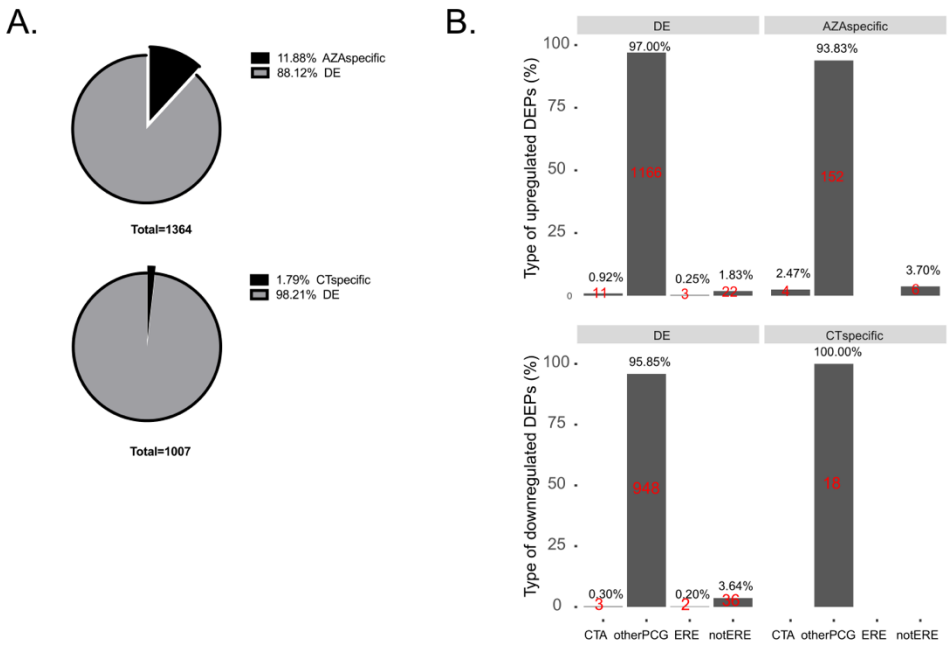

Figure S 3.5 MAPs presented de novo after treatment derived from CTAs rather than EREs

A, Pie charts indicating the proportion of new MAPs previously unidentified on untreated cells (AZA specific) or MAPs unidentified after AZA treatment (CT specific) and differentially expressed for up- (upper panel) and downregulated DEMs (lower panel). **B**, Bar plots indicating DEM composition according to biotypes for up- (upper panel) and downregulated DEMs (lower panel). OtherPCG: other protein coding genes, DE: differentially expressed.

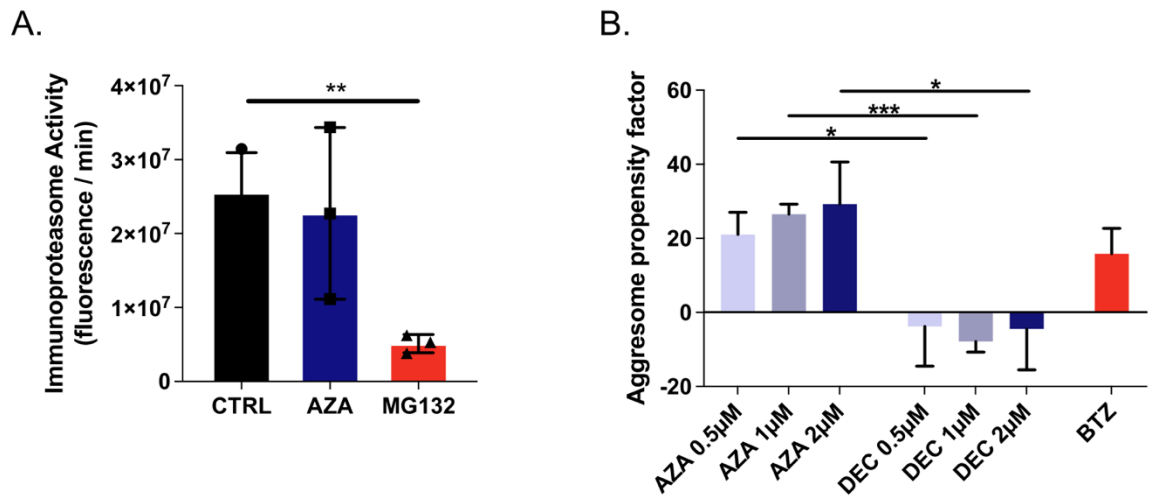

Figure S 3.6 AZA molds the immunopeptidome through DNMT2 inhibition

A, Immunoproteasome activity monitored in THP-1 cells after AZA treatment. MG132, a proteasome inhibitor, was used as a negative control (unpaired t-test; ** $p< 0.01$) **B**, Quantification of protein aggregates induced with increasing AZA and DAC concentrations in THP-1 cells. Bortezomib (BTZ) was used as a positive control (unpaired t-test; *** $p< 0.001$, * $p< 0.05$).

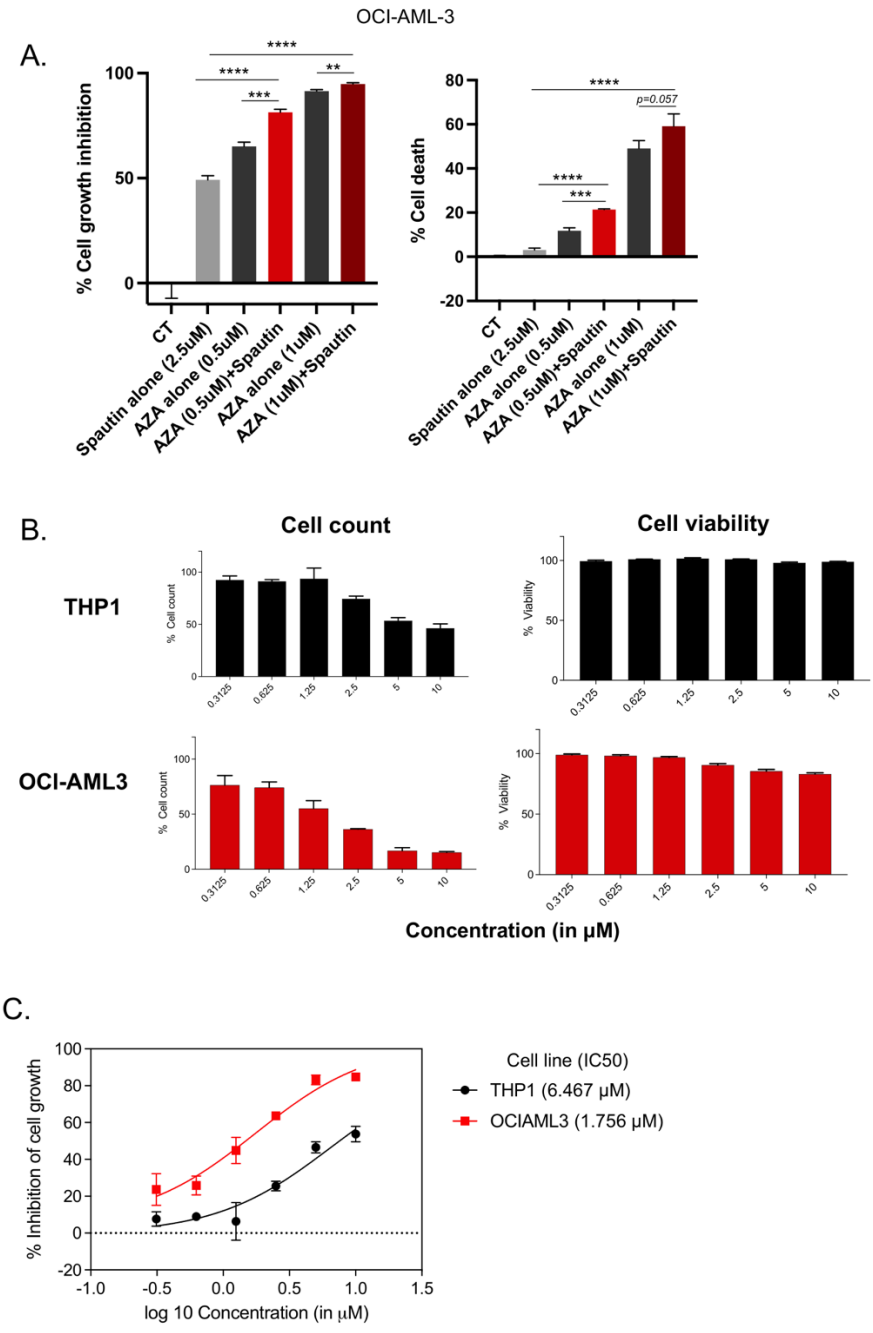

Figure S 3.7 Autophagy inhibition synergizes with AZA, and spautin-1 treatment alone does not induce cell death

A, Cell growth inhibition and cell death of OCI-AML-3 cells treated with either increasing concentrations of AZA, spautin-1, or both, monitored with 7-AAD via flow cytometry. Control cells were OCI-AML-3 cells treated with 0.1% DMSO (two independent experiments; unpaired t-test; **** p< 0.0001, *** p< 0.001, ** p< 0.01). **B**, Viable cell counts after treatment with increasing concentrations of spautin-1 in THP-1 and OCI-AML-3 compared to DMSO-treated

control cells using 7-AAD via flow cytometry. **C**, Dose-response curves and IC50 values generated from B.

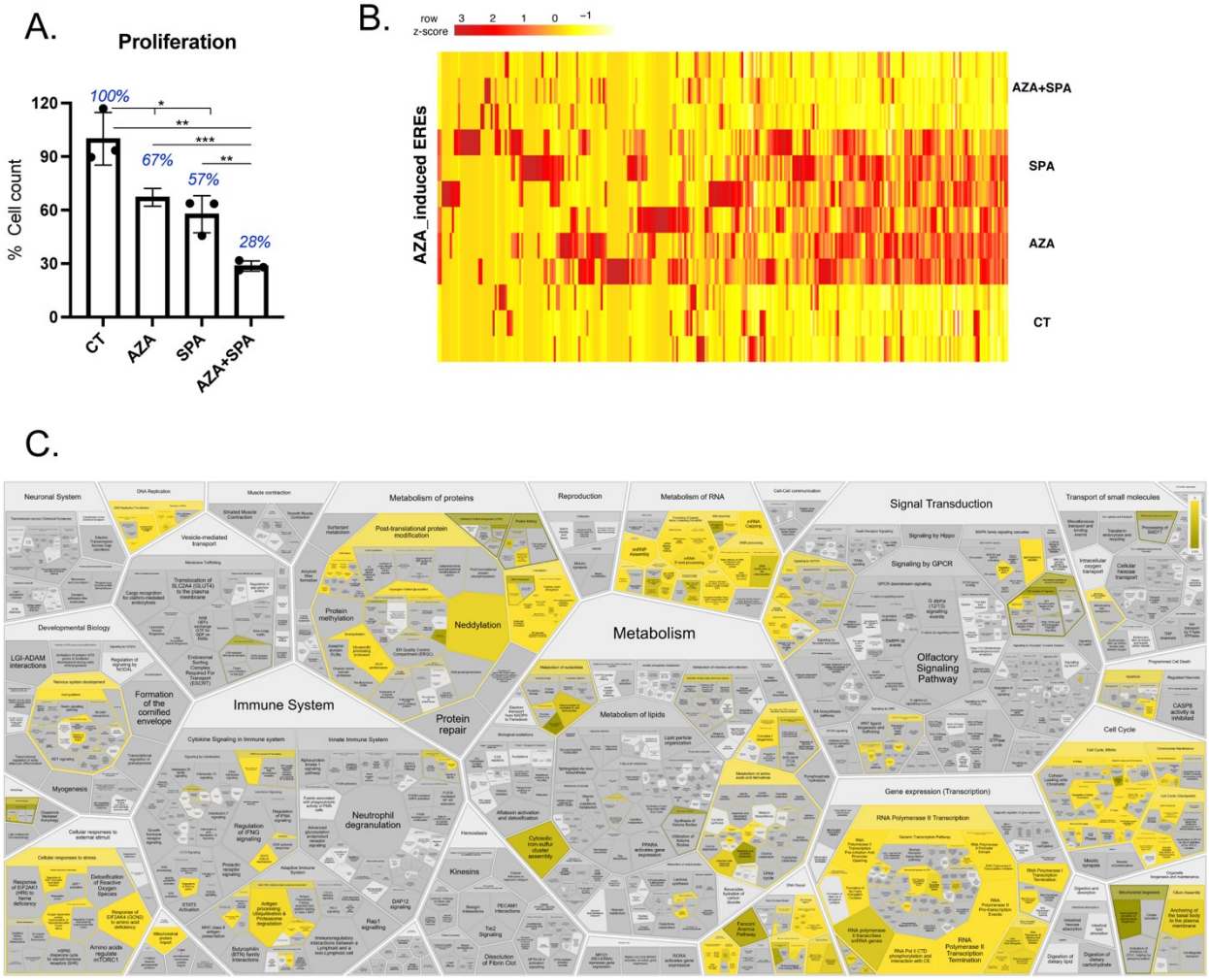

Figure S 3.8 Autophagy inhibition synergizes with AZA and might increase AML immunogenicity

A, Cell growth inhibition with either AZA, spautin-1, or both, monitored with 7-AAD via flow cytometry in THP-1 cells (unpaired t-test; ** $p < 0.01$, *$p < 0.05$) **B**, Heatmap of AZA-induced ERE expression (in cpm) in AZA, spautin-1, AZA+spautin-1, and DMSO-treated THP-1 cells. **C**, Pathway enrichment analysis using Cytoscape software to identify biological pathways significantly enriched in AZA- and spautin-1-treated cells compared to cells treated with AZA alone. Significant pathways are highlighted in yellow.

Due to size constraints, the following three tables are not presented in the document. The files are available as an Excel file online at

https://www.biorxiv.org/content/10.1101/2022.12.02.518683v1.supplementary-material

Table S 3.1 Differentially expressed genes identified by limma voom analysis performed per cell line

Table S 3.2 Differentially expressed MAPs identified by limma analysis performed per cell line

Table S 3.3 Differentially expressed genes identified by paired limma analysis identifying commonly altered genes by across cell lines

Table S 3.4 Correlation analysis of number of DOIs generated for each protein with the frequency of each amino acid residue in the considered protein

	up		*down*	
AA	r	p	r	p
asp	0,13	****	0,02	NS
gly	0,08	****	-0,03	NS
tyr	0,06	***	0,01	NS
glu	0,05	**	0	NS
asn	0,04	*	0,09	***
met	0,03	*	0,05	**
lys	0,02	NS	-0,05	*
phe	0,02	NS	0,04	*
ala	0,01	NS	-0,03	NS
arg	0	NS	-0,02	NS
val	0	NS	0	NS
gln	0	NS	0,04	NS
ser	-0,02	NS	-0,04	NS
iso	-0,04	*	0,04	NS
cys	-0,06	***	-0,01	NS
leu	-0,06	***	-0,02	NS
his	-0,07	****	0,09	***
pro	-0,07	****	-0,03	NS
trp	-0,07	****	0	NS
thr	-0,09	****	0,06	**

3.10 References

1. Schoofs, T., W.E. Berdel, and C. Muller-Tidow, *Origins of aberrant DNA methylation in acute myeloid leukemia.* Leukemia, 2014. **28**(1): p. 1-14.
2. Kroeger, H., et al., *Aberrant CpG island methylation in acute myeloid leukemia is accentuated at relapse.* Blood, The Journal of the American Society of Hematology, 2008. **112**(4): p. 1366-1373.
3. Döhner, H., et al., *Diagnosis and management of AML in adults: 2022 recommendations from an international expert panel on behalf of the ELN.* Blood, The Journal of the American Society of Hematology, 2022. **140**(12): p. 1345-1377.
4. Šimoničová, K., et al., *Different mechanisms of drug resistance to hypomethylating agents in the treatment of myelodysplastic syndromes and acute myeloid leukemia.* Drug Resist Updat, 2022. **61**: p. 100805.
5. DiNardo, C.D., et al., *Azacitidine and Venetoclax in Previously Untreated Acute Myeloid Leukemia.* N Engl J Med, 2020. **383**(7): p. 617-629.
6. Jasielec, J., V. Saloura, and L.A. Godley, *The mechanistic role of DNA methylation in myeloid leukemogenesis.* Leukemia, 2014. **28**(9): p. 1765-73.
7. Schaefer, M., et al., *Azacytidine inhibits RNA methylation at DNMT2 target sites in human cancer cell lines.* Cancer Res, 2009. **69**(20): p. 8127-32.
8. Ma, J. and Z. Ge, *Comparison between decitabine and azacitidine for patients with acute myeloid leukemia and higher-risk myelodysplastic syndrome: a systematic review and network meta-analysis.* Frontiers in Pharmacology, 2021: p. 1919.
9. Kantarjian, H.M., et al., *Multicenter, randomized, open-label, phase III trial of decitabine versus patient choice, with physician advice, of either supportive care or low-dose cytarabine for the treatment of older patients with newly diagnosed acute myeloid leukemia.* J Clin Oncol, 2012. **30**(21): p. 2670-7.
10. Fenaux, P., et al., *Efficacy of azacitidine compared with that of conventional care regimens in the treatment of higher-risk myelodysplastic syndromes: a randomised, open-label, phase III study.* Lancet Oncol, 2009. **10**(3): p. 223-32.
11. Malik, P. and A.F. Cashen, *Decitabine in the treatment of acute myeloid leukemia in elderly patients.* Cancer Manag Res, 2014. **6**: p. 53-61.
12. Srivastava, P., et al., *Induction of cancer testis antigen expression in circulating acute myeloid leukemia blasts following hypomethylating agent monotherapy.* Oncotarget, 2016. **7**(11): p. 12840-56.
13. Matsushita, M., et al., *Identification of Novel HLA-A*24:02-Restricted Epitope Derived from a Homeobox Protein Expressed in Hematological Malignancies.* PLoS One, 2016. **11**(1): p. e0146371.
14. Ehx, G. and C. Perreault, *Discovery and characterization of actionable tumor antigens.* Genome Medicine, 2019. **11**(1): p. 29.
15. Luo, N., et al., *DNA methyltransferase inhibition upregulates MHC-I to potentiate cytotoxic T lymphocyte responses in breast cancer.* Nat Commun, 2018. **9**(1): p. 248.
16. Gang, A.O., et al., *5-Azacytidine treatment sensitizes tumor cells to T-cell mediated cytotoxicity and modulates NK cells in patients with myeloid malignancies.* Blood Cancer J, 2014. **4**(3): p. e197.

17. Goodyear, O.C., et al., *Azacitidine augments expansion of regulatory T cells after allogeneic stem cell transplantation in patients with acute myeloid leukemia (AML).* Blood, 2012. **119**(14): p. 3361-9.
18. Nahas, M.R., et al., *Hypomethylating agent alters the immune microenvironment in acute myeloid leukaemia (AML) and enhances the immunogenicity of a dendritic cell/AML vaccine.* British journal of haematology, 2019. **185**(4): p. 679-690.
19. Stone, M.L., et al., *Epigenetic therapy activates type I interferon signaling in murine ovarian cancer to reduce immunosuppression and tumor burden.* Proceedings of the National Academy of Sciences, 2017. **114**(51): p. E10981-E10990.
20. Dewannieux, M. and T. Heidmann, *Endogenous retroviruses: acquisition, amplification and taming of genome invaders.* Current opinion in virology, 2013. **3**(6): p. 646-656.
21. Schulz, W.A., C. Steinhoff, and A.R. Florl, *Methylation of endogenous human retroelements in health and disease.* Curr Top Microbiol Immunol, 2006. **310**: p. 211-50.
22. Payer, L.M. and K.H. Burns, *Transposable elements in human genetic disease.* Nature Reviews Genetics, 2019. **20**(12): p. 760-772.
23. Chiappinelli, K.B., et al., *Inhibiting DNA Methylation Causes an Interferon Response in Cancer via dsRNA Including Endogenous Retroviruses.* Cell, 2015. **162**(5): p. 974-86.
24. Roulois, D., et al., *DNA-Demethylating Agents Target Colorectal Cancer Cells by Inducing Viral Mimicry by Endogenous Transcripts.* Cell, 2015. **162**(5): p. 961-73.
25. Laumont, C.M., et al., *Noncoding regions are the main source of targetable tumor-specific antigens.* Sci Transl Med, 2018. **10**(470).
26. Smith, C.C., et al., *Endogenous retroviral signatures predict immunotherapy response in clear cell renal cell carcinoma.* The Journal of clinical investigation, 2019. **128**(11): p. 4804-4820.
27. Mullins, C.S. and M. Linnebacher, *Endogenous retrovirus sequences as a novel class of tumor-specific antigens: an example of HERV-H env encoding strong CTL epitopes.* Cancer Immunology, Immunotherapy, 2012. **61**(7): p. 1093-1100.
28. Ehx, G., et al., *Atypical acute myeloid leukemia-specific transcripts generate shared and immunogenic MHC class-I-associated epitopes.* Immunity, 2021. **54**(4): p. 737-752.e10.
29. Bray, N.L., et al., *Near-optimal probabilistic RNA-seq quantification.* Nature Biotechnology, 2016. **34**(5): p. 525-527.
30. Noronha, N., et al., *Major multilevel molecular divergence between THP-1 cells from different biorepositories.* Int J Cancer, 2020. **147**(7): p. 2000-2006.
31. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.* Bioinformatics, 2010. **26**(1): p. 139-40.
32. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing and microarray studies.* Nucleic Acids Res, 2015. **43**(7): p. e47.
33. Sergushichev, A.A., *An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation.* bioRxiv, 2016: p. 060012.
34. Shannon, P., et al., *Cytoscape: a software environment for integrated models of biomolecular interaction networks.* Genome Res, 2003. **13**(11): p. 2498-504.
35. Dobin, A., et al., *STAR: ultrafast universal RNA-seq aligner.* Bioinformatics, 2013. **29**(1): p. 15-21.
36. Garrison, E. and G. Marth, *Haplotype-based variant detection from short-read sequencing.* arXiv: Genomics, 2012.

37. Daouda, T., C. Perreault, and S. Lemieux, *pyGeno: A Python package for precision medicine and proteogenomics.* F1000Res, 2016. **5**: p. 381.
38. Li, H., et al., *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-9.
39. Li, H., *A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data.* Bioinformatics, 2011. **27**(21): p. 2987-93.
40. Marcais, G. and C. Kingsford, *A fast, lock-free approach for efficient parallel counting of occurrences of k-mers.* Bioinformatics, 2011. **27**(6): p. 764-70.
41. Audoux, J., et al., *DE-kupl: exhaustive capture of biological variation in RNA-seq data through k-mer decomposition.* Genome Biology, 2017. **18**(1): p. 243.
42. Pearson, H., et al., *MHC class I-associated peptides derive from selective regions of the human genome.* J Clin Invest, 2016. **126**(12): p. 4690-4701.
43. Szolek, A., et al., *OptiType: precision HLA typing from next-generation sequencing data.* Bioinformatics, 2014. **30**(23): p. 3310-6.
44. O'Donnell, T.J., A. Rubinsteyn, and U. Laserson, *MHCflurry 2.0: Improved Pan-Allele Prediction of MHC Class I-Presented Peptides by Incorporating Antigen Processing.* Cell Syst, 2020. **11**(1): p. 42-48.e7.
45. Reynisson, B., et al., *NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data.* Nucleic Acids Res, 2020. **48**(W1): p. W449-w454.
46. Farrell, D., *epitopepredict: A tool for integrated MHC binding prediction.* bioRxiv, 2021: p. 2021.02.05.429892.
47. Lamoliatte, F., et al., *Uncovering the SUMOylation and ubiquitylation crosstalk in human cells using sequential peptide immunopurification.* Nature communications, 2017. **8**(1): p. 1-11.
48. Reynisson, B., et al., *NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data.* Nucleic acids research, 2020. **48**(W1): p. W449-W454.
49. Courcelles, M., et al., *MAPDP: a cloud-based computational platform for immunopeptidomics analyses.* Journal of proteome research, 2020. **19**(4): p. 1873-1881.
50. Willforss, J., A. Chawade, and F. Levander, *NormalyzerDE: online tool for improved normalization of omics expression data and high-sensitivity differential expression analysis.* Journal of proteome research, 2018. **18**(2): p. 732-740.
51. Ruiz Cuevas, M.V., et al., *BamQuery: a proteogenomic tool for the genome-wide exploration of the immunopeptidome.* bioRxiv, 2022: p. 2022.10.07.510944.
52. Almeida, L.G., et al., *CTdatabase: a knowledge-base of high-throughput and curated data on cancer-testis antigens.* Nucleic acids research, 2009. **37**(suppl_1): p. D816-D819.
53. Larouche, J.D., et al., *Widespread and tissue-specific expression of endogenous retroelements in human somatic tissues.* Genome Med, 2020. **12**(1): p. 40.
54. Caron, E., et al., *The MHC I immunopeptidome conveys to the cell surface an integrative view of cellular regulation.* Molecular systems biology, 2011. **7**(1): p. 533.
55. Bloy, N., et al., *Immunogenic stress and death of cancer cells: contribution of antigenicity vs adjuvanticity to immunosurveillance.* Immunological reviews, 2017. **280**(1): p. 165-174.
56. Varshney, D., et al., *SINE transcription by RNA polymerase III is suppressed by histone methylation but not by DNA methylation.* Nature communications, 2015. **6**(1): p. 1-12.

57. Paget, M., et al., *Stress granules are shock absorbers that prevent excessive innate immune responses to dsRNA.* BioRxiv, 2022: p. 2021.04. 26.441141.
58. Guo, H., et al., *Autophagy supports genomic stability by degrading retrotransposon RNA.* Nature communications, 2014. **5**(1): p. 1-11.
59. Gaczynska, M., K.L. Rock, and A.L. Goldberg, *Gamma-interferon and expression of MHC genes regulate peptide hydrolysis by proteasomes.* Nature, 1993. **365**(6443): p. 264-7.
60. Mishto, M., et al., *Proteasome isoforms exhibit only quantitative differences in cleavage and epitope generation.* Eur J Immunol, 2014. **44**(12): p. 3508-21.
61. Javitt, A., et al., *Pro-inflammatory Cytokines Alter the Immunopeptidome Landscape by Modulation of HLA-B Expression.* Front Immunol, 2019. **10**: p. 141.
62. Li, J. and M. Hochstrasser, *Microautophagy regulates proteasome homeostasis.* Curr Genet, 2020. **66**(4): p. 683-687.
63. Li, Y., S. Li, and H. Wu, *Ubiquitination-Proteasome System (UPS) and Autophagy Two Main Protein Degradation Machineries in Response to Cell Stress.* Cells, 2022. **11**(5).
64. Goll, M.G., et al., *Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2.* Science, 2006. **311**(5759): p. 395-8.
65. Schaefer, M., et al., *RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage.* Genes Dev, 2010. **24**(15): p. 1590-5.
66. Schaefer, M., et al., *RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage.* Genes & development, 2010. **24**(15): p. 1590-1595.
67. Eilers, M., et al., *Internal packing of helical membrane proteins.* Proc Natl Acad Sci U S A, 2000. **97**(11): p. 5796-801.
68. Pappalardi, M.B., et al., *Discovery of a first-in-class reversible DNMT1-selective inhibitor with improved tolerability and efficacy in acute myeloid leukemia.* Nat Cancer, 2021. **2**(10): p. 1002-1017.
69. Kunimasa, K., et al., *Spautin-1 inhibits mitochondrial complex I and leads to suppression of the unfolded protein response and cell survival during glucose starvation.* Sci Rep, 2022. **12**(1): p. 11533.
70. Danaher, P., et al., *Gene expression markers of Tumor Infiltrating Leukocytes.* J Immunother Cancer, 2017. **5**: p. 18.
71. Ehx, G., et al., *Atypical acute myeloid leukemia-specific transcripts generate shared and immunogenic MHC class-I-associated epitopes.* Immunity, 2021. **54**(4): p. 737-752. e10.
72. Zhao, Q., et al., *Proteogenomics Uncovers a Vast Repertoire of Shared Tumor-Specific Antigens in Ovarian CancerThe Global Landscape of Ovarian Cancer–Specific Antigens.* Cancer immunology research, 2020. **8**(4): p. 544-555.
73. Saini, S.K., et al., *Human endogenous retroviruses form a reservoir of T cell targets in hematological cancers.* Nature communications, 2020. **11**(1): p. 1-14.
74. Onishi-Seebacher, M., et al., *Repeat to gene expression ratios in leukemic blast cells can stratify risk prediction in acute myeloid leukemia.* BMC Medical Genomics, 2021. **14**(1): p. 1-20.
75. Ohtani, H., et al., *Activation of a subset of evolutionarily young transposable elements and innate immunity are linked to clinical responses to 5-azacytidine.* Cancer research, 2020. **80**(12): p. 2441-2450.
76. Tuorto, F., et al., *The tRNA methyltransferase Dnmt2 is required for accurate polypeptide synthesis during haematopoiesis.* The EMBO journal, 2015. **34**(18): p. 2350-2362.

77. Romano, A., et al., *Proteomic Analysis Reveals Autophagy as Pro-Survival Pathway Elicited by Long-Term Exposure with 5-Azacitidine in High-Risk Myelodysplasia.* Front Pharmacol, 2017. **8**: p. 204.
78. Kong, Y., et al., *Transposable element expression in tumors is associated with immune infiltration and increased antigenicity.* Nat Commun, 2019. **10**(1): p. 5228.
79. Choi, Y., J.W. Bowman, and J.U. Jung, *Autophagy during viral infection - a double-edged sword.* Nat Rev Microbiol, 2018. **16**(6): p. 341-354.
80. Dengjel, J., et al., *Autophagy promotes MHC class II presentation of peptides from intracellular source proteins.* Proc Natl Acad Sci U S A, 2005. **102**(22): p. 7922-7.
81. Paludan, C., et al., *Endogenous MHC class II processing of a viral nuclear antigen after autophagy.* Science, 2005. **307**(5709): p. 593-6.
82. Blanchet, F.P., et al., *Human immunodeficiency virus-1 inhibition of immunoamphisomes in dendritic cells impairs early innate and adaptive immune responses.* Immunity, 2010. **32**(5): p. 654-69.
83. Loi, M., et al., *Macroautophagy Proteins Control MHC Class I Levels on Dendritic Cells and Shape Anti-viral CD8(+) T Cell Responses.* Cell Rep, 2016. **15**(5): p. 1076-1087.
84. Parekh, V.V., et al., *Autophagy-related protein Vps34 controls the homeostasis and function of antigen cross-presenting CD8α(+) dendritic cells.* Proc Natl Acad Sci U S A, 2017. **114**(31): p. E6371-e6380.
85. Jounai, N., et al., *The Atg5 Atg12 conjugate associates with innate antiviral immune responses.* Proc Natl Acad Sci U S A, 2007. **104**(35): p. 14050-5.
86. Lei, Y., et al., *The mitochondrial proteins NLRX1 and TUFM form a complex that regulates type I interferon and autophagy.* Immunity, 2012. **36**(6): p. 933-46.
87. Xu, C., et al., *The Autophagy Cargo Receptor SQSTM1 Inhibits Infectious Bursal Disease Virus Infection through Selective Autophagic Degradation of Double-Stranded Viral RNA.* Viruses, 2021. **13**(12).
88. Nawrocki, S.T., et al., *The novel autophagy inhibitor ROC-325 augments the antileukemic activity of azacitidine.* Leukemia, 2019. **33**(12): p. 2971-2974.
89. Pappalardi, M.B., et al., *Discovery of a first-in-class reversible DNMT1-selective inhibitor with improved tolerability and efficacy in acute myeloid leukemia.* Nature cancer, 2021. **2**(10): p. 1002-1017.

4 Chapter 4: Discussion

4.1 Proceed with Caution: The Advantages and Pitfalls of Working with Cell lines

Cell lines are an important model in biological sciences and have been instrumental in the advancement of research. They have short doubling times, are easy to culture and store long-term giving unlimited access to sample material. Further, there are several databases with extensive information on cell lines, including mutation profiles, karyotyping, drug sensitivity, RNA expression, gene dependency, and so on. Thus, researchers have access to in-depth knowledge of their study model beforehand. This not only benefits researchers to efficiently plan experiments but also generate new hypotheses based on current data. However, cell lines have been long associated with the irreproducibility of data based on how they have been handled or the source of the cell lines [1]. Cell lines, especially MSI+, can easily undergo genotypic and phenotypic drift leading to variability in the data. Finally, cell lines can be easily cross-contaminated with other cell lines when handled without caution. Researchers can opt to generate their own cell lines in-house from primary tissues, or purchase them from cell repositories, or borrow them from labs that previously possess the cell lines. The latter is the most frequent means of acquiring cell lines due to its convenience and low cost. However, the reliability of the cell lines acquired by this method is highly questionable as there may be inadequate information or documentation on the duration of passage, culture media used, and so on.

In this thesis, we studied THP1 cells – a microsatellite stable (MSS) cell line considered to be fairly stable – and found huge levels of genetic drift between source cell repositories, ATCC, and DSMZ. We found that THP1 cells obtained from ATCC were associated with LOHs on several

chromosomal arms leading to differences in physical appearance, HLA loss, and transcriptomic changes leading to variable conclusions depending on the cell line used. This is especially disquieting since both repositories are trusted to be vigorous and thus cell lines received from either are expected to be identical.

4.1.1 Recommendations to current guidelines for cell line-based research

Our study shows that the current STR profiling technique is not sufficient for the accurate authentication of cell lines. The current method uses amplification of only 8 polymorphic loci in STR profiling. Although this limited number of loci is sufficient to detect cross-contamination of cell lines, it is not extensive enough to accurately identify genetic drifts occurring within a particular cell line. We, therefore, recommend increasing the number of loci used for amplification to 80-160 loci. Further, we recommend the use Tanabe algorithm, instead of the Masters algorithm used by ATCC, to compute a score of the percentage of match between cell lines. As indicated in section 2.6, the Masters algorithm only takes the number of alleles of the reference cell line to compute the match score, while the Tanabe algorithm takes the total number of alleles of both cell lines into consideration. Thus, although the simplicity of the Masters algorithm allows for the detection of cross-contamination of cell lines, the latter is more powerful in detecting genetic drifts across cell lines that arise from the gain or loss of alleles. Finally, our study advocates for the importance of following strict laboratory practices to ensure that cell lines do not undergo genetic drift. This includes proper documentation of cell passages, and no over-passaging or over-diluting the cell lines. We also emphasize the importance of acquiring cell lines from valid sources and reporting the origin while publishing or submitting data to public data repositories. In the case of THP1 cells, we recommend the use of the DSMZ counterpart for leukemia and immunopeptidomic-related studies.

4.1.2 Cell lines as a model to investigate AZA-induced changes

In this study, we opted to use cell lines as a model to investigate the effects of AZA on the immunopeptidome because their use permits access to large sample quantities required for MS analyses. Following recommendations based on our previous work, we used cell lines newly purchased from DSMZ and cultured them ensuring good laboratory practices. Cell lines were never cultured longer than 4-6 weeks and were diluted in a timely and controlled manner to prevent genetic drift and ensure reproducibility. Further, our studies were performed on multiple cell lines to provide a broader perspective of AZA-induced changes, rather than identify changes specific to a single cell line. Cell lines were selected to represent the mutational and cytogenetic heterogeneity observed in AML patients, allowing us to investigate their contributions to AZA effects on the transcriptome and immunopeptidome. Cell lines also differed in HLA types allowing us to identify a diverse range of MAPs induced by AZA. Finally, to ensure the physiological relevance of our findings, we validated key results in external datasets using transcriptomic data from AML patient specimens.

4.2 The Elusive Connection: Discrepancies Between Mutations, Transcriptome, and Immunopeptidome

As described in section 1.3.3, there has been no coherent association between specific gene mutations and favorable patient outcomes with AZA treatment. Consistent with this, we didn't observe any apparent association between mutations associated with either hyper- or hypomethylating function and levels of transcriptomic changes in cell lines. For instance, THP1 has an MLL-AF9 translocation that is associated with hypomethylation patterns [2]. However, we observed that THP1 was the most "sensitive" in terms of AZA-induced transcriptomic changes – with almost two-fold higher D.E. elements compared to other cell lines (Figure 3.2B). This could

be explained by the near-tetraploid nature which was a feature only observed in THP1. Despite such differential transcriptomic sensitivities across cell lines, we observed a consistent upregulation of transcriptional events in all cell lines (Figure 3.2D, E). Importantly this congruent unidirectionality was lost at the level of the immunopeptidome (Figure 3.2D, E). Similar findings were observed in other studies, including glioblastoma cell lines treated with HMAs [3, 4]. In glioblastoma, the majority of alterations at the transcriptomic level were derived from their upregulation. However, this trend was completely lost at the protein level and was only observed at the immunopeptidome level when accompanied by an HMA-induced increase in the HLA molecules themselves. Taken together, these data are indicative of the contribution of post-transcriptomic changes and other factors leading to the disconnect between the transcriptome and immunopeptidome. It is also worth mentioning that HMAs have been previously well-reported to increase HLA levels due to their demethylation activity. However, unlike most solid cancers that undergo the downregulation of MHC-I molecules as an immune escape mechanism [5, 6], this is rarely observed in AML [7-10]. Leukemic cells express similar levels of MHC-I molecules compared to their normal counterparts [8], and normal granulocytes themselves have very high basal levels of MHC-I surface expression compared to other normal cells [11]. This pre-existing high MHC-I surface level could explain why we do not observe any further increases after AZA treatment, unlike what was observed in solid cancers.

4.3 From the Archives to the Clinic: Revisiting CTAs for AML Therapy

Regardless of the absence of a global increase in MAPs after AZA treatment, there was a significant increase in the proportion of MAPs deriving from CTA gene products (Figure 3.2 G). This was expected as genes coding for CTAs are regulated by methylation and HMA treatment has previously been shown to induce MAPs deriving from CTAs, both in hematological and non-

hematological cancers [4, 12, 13]. As described in detail in section 1.6.2, CTAs have been a well-described source of TAs that are mostly expressed in cancers, but not in most normal tissues. Moreover, some phase I/II clinical trials targeting CTAs as TAs as cancer treatments show promising results. In fact, studies have shown that patients treated with HMAs had an increase in CTA-specific $CD8^+$ T-cells [14, 15], suggesting that HMA-induced CTA-derived MAPs are responsible for the elimination of leukemic blasts by $CD8^+$ T-cells. Based on these studies, research groups have tested the combination of HMAs with CTAs vaccines to enhance the efficacy of tumor killing by HMAs. Notably, two phase-I clinical trials were approved to investigate the efficiency of this combination. The first trial was designed to use dendritic cells pulsed with CTA peptides – MAGE-A1, MAGE-A3, and NY-ESO-1 – as the cancer vaccine (NCT01483274). Unfortunately, this trial did not begin as recruitment was withdrawn (due to a patient population barrier to the study). The second trial used a combination of NY-ESO-1, PRAME, MAGE-A3, and WT-1 as a multi-peptide (NPMW-peptide vaccine) and was administered subcutaneously with an adjuvant in 5 MDS/AML patients along with the AZA treatment (NCT02750995). NY-ESO-1, PRAME, and MAGE-A3 are CTAs that have been previously shown to be re-expressed at the RNA level upon HMA treatment [14, 16]. WT-1, a TAA that is overexpressed in several hematological malignancies, was additionally added to the peptide combination [17]. In this trial, although there were no serious adverse events, neither vaccine-specific immune response nor clinical benefit was observed [18]. Unfortunately, there is not much information available regarding why this trial failed. First, the lack of vaccine-specific immune response indicates that vaccination was not efficient, and the choice and design of the vaccine should be reviewed. Here, long peptides (25-29 amino acids) were used for vaccination. Long peptides contain multiple epitopes that can be processed in the antigen-presenting cells (APCs) to generate multiple MAPs

and therefore have a broader coverage of HLAs. It is possible that NPMW-peptides were not processed efficiently and failed to bind with MHC molecules. Other vaccination strategies, such as DC-loaded peptides would be worth pursuing. Secondly, the design of the clinical trial could be an additional factor contributing to the failure of the trial. The peptide vaccination was administered after patients received six cycles of AZA treatment. However, previous studies investigating CTA re-expression demonstrated that increases are observed as early as 5-10 days post-HMA treatment. Therefore, it is crucial to determine the kinetics of CTA expression six months after AZA treatment or consider a treatment regimen administering the peptide vaccination early on during AZA treatment. Finally, another consideration could be to validate the CTA re-expression upon AZA treatment and use it as a patient selection criterion for peptide vaccination. Nevertheless, AZA-induced CTA-derived MAPs in AML warrant the need for additional and improved trials to assess the benefit of peptide vaccination for AML patients.

4.4 Busting the Myth: The Lack of EREs at the MAP Level

One of the most interesting and unexpected findings in our study was the lack of EREs upregulated at the MAP level. There are several reasons supporting the postulation that ERE-derived TAs would contribute to AZA-mediated immunological responses. Firstly, ERE-derived MAPs are an important source of TSA candidates; they are absent in normal cells and have been shown to be highly immunogenic. Secondly, as discussed in section 1.5.2, their genomic sequences are mainly regulated and silenced by DNA methylation. Upon treatment of DNA demethylating agents, their RNA expression has been demonstrated to be upregulated both *in vitro* and *in vivo*. Finally, translation products of EREs are known to be unstable, making them perfect candidates for DRiPs generation and therefore MAPs. Based on this evidence, we wished to validate if AZA indeed induces the expression of AZA-induced MAPs using transcriptomic and MS analysis. The

strength of using a proteogenomic approach lies in the direct identification of MAPs by MS rather than the use of prediction algorithms that are associated with very low accuracy and high false positive identifications. However, there are limitations associated with MS that are important to be considered (section 1.6.2). Another important consideration is that peptide sequences must be included in the reference databases for peptides to be identified. Therefore, to circumvent as many of these limitations and increase the confidence of our study, we performed experiments in triplicates using large quantities of samples (~100 million cells per replicate) to maximize the chances of identifying peptides by MS. Further, using deep-coverage RNA sequencing data and HLA binding prediction algorithms we created cell line-specific databases. These databases thus represented all possible peptide coding sequences that could be translated and processed to be presented at the MAP level. Finally, EREs are highly unstable and are expected to be presented at the immunopeptidome as soon as they are transcribed. Therefore, choosing the time point where we observed the highest expression of EREs would be crucial to enable their detection at the MAP level. In our study, we observed that AZA induced maximum ERE expression after 72 hours of treatment followed by drug release (Figure 3.1B, S3.2). Interestingly, the increase in ERE expression was delayed and transient as reported in other solid cancers, although with different kinetics in AML cells. This suggests that some mechanisms of ERE regulation by AZA are shared across cancer types. However, despite these large levels of ERE induction at the RNA level in AML cells, we did not identify any ERE-derived MAPs. We also verified the possibility of a delayed presentation of ERE-derived MAPs and did not identify any at a later time point (Figure S3.4). Taken together, our data suggest that it is unlikely that AZA-mediated immunological responses would be driven by T-cell recognition of ERE-derived MAPs. Concurrent with our findings, a study performed on patients with hematological cancers showed no significant

increases in T-cell populations targeting ERE-derived MAPs after AZA treatment [19]. Nevertheless, ERE expression at the RNA level had important biological implications. Due to their repetitive nature, EREs formed dsRNA that culminated in a state of viral mimicry. This was demonstrated by an upregulation of genes associated with biological functions and pathways involved in inflammatory and innate immune responses, which was accompanied by an accumulation of cellular dsRNA (Figure 3.2H-J). Both ERE expression and viral mimicry have been described to contribute to cellular stress and cell death [20-22]. A report by Sistigu *et al* has demonstrated that DNA-damaging drugs can elicit type I interferon signatures leading to tumor elimination *in vivo* and that such drug-induced viral mimicry is a hallmark of successful chemotherapy [23]. In the case of AML, high basal ERE expression in patients alone was able to predict better clinical responses. The AZA-induced EREs identified in our study were not only relevant in our cell lines but have physiological relevance as they were associated with genes involved in immune responses in AML patients' specimens (Figure 3.3B-D). Thus, we believe that a subset of EREs, notably these AZA-induced EREs, are key in triggering immune responses. In line with this, Ohtani *et al* also found that specific ERE classes play a role in driving clinical responses in patients with hematologic malignancies post-AZA treatment. These ERE classes were evolutionarily "younger" in nature and are more potent in triggering immune responses. Based on these findings, we could speculate that AZA-induced EREs identified in this study would likely have an evolutionarily younger age and hence were capable of generating immune responses. Altogether our data show that although AZA-induced EREs did not generate MAPs, they play an important role at the RNA level culminating in anti-viral-like responses.

4.5 Unlocking the Potential: Leveraging AZA's Anti-viral Responses for AML Treatment

Anti-viral-like responses are known to increase the immunogenicity of cancer cells. Interferon signaling as a result of viral mimicry can enhance the antigenicity of cells by increasing the expression of the genes coding for proteins involved in antigen processing and presentation and MHC-I molecules themselves [20, 24]. However, antigenicity alone is not sufficient to successfully trigger T-cell responses. T-cells require co-stimulatory responses provided by APCs, essentially dendritic cells (DCs). This facilitates T-cell priming and provides co-stimulatory cytokines to allow the expansion of T-cells, thus enhancing the adjuvanticity of tumor or infected cells. Adjuvanticity can be described as a secondary discriminatory mechanism employed by the immune system to provide further context to recognize 'self' from 'non-self' [25]. In the case of viral or bacterial infections, this is accomplished by signals from PAMPs such as LPS, and viral nucleic acids, such as dsRNA. These danger signals activate the innate immune system by allowing the maturation of DCs and T-cell stimulation. In the context of cancer cells, inflammatory signaling cascades triggered by viral mimicry contribute to an increase in adjuvanticity [26]. Interestingly, there was a significant expansion of viral-specific T-cells in hematological cancer patients post-AZA treatment [19]. Although the authors did not comment on this finding, we can speculate that AZA-induced viral mimicry could be responsible for such T-cell expansion. This finding could be further explored as a therapeutic tool to improve AZA therapies. A study in melanoma explored the cross-reactivity of T-cells against viral antigens and TAs and demonstrated that vaccination against TAs with sequences homologous to viral peptides provides tumor protection in mice [27]. Therefore, it may be worth investigating peptide homologies between AZA-induced peptides with viral epitopes and their ability to cross-react effectively. This could

be considered as a method to pre-screen peptides as we could expect these peptides to be more successful in vaccination. Further, as viral-specific T-cells are expanded after AZA treatment, patients can be selected on serotype status pre-treatment, and respective T-cells can be targeted to engineer high-affinity T-cell receptors (TCR) against specific TAs. TCR gene therapy has been successfully established for targeting WT1, a TAA in AML and MDS patients, and has shown encouraging results in phase I/II clinical trials [28].

4.6 Bringing RNA Effects of AZA to the Forefront: A Fresh Perspective on an Overlooked Mechanism

While AZA is primarily known for its effects on DNA demethylation, it is now well-established that AZA plays a major role in RNA demethylation. As mentioned in section 1.3.1, owing to their chemical properties, AZA can incorporate into both RNA and DNA following cellular uptake. Importantly, the majority (80-90%) of AZA is incorporated RNA, whereas only a minor fraction (10-20%) becomes a part of DNA. AZA incorporation in RNA leads to DNMT2 inhibition and consequently RNA demethylation. The main targets of AZA-mediated demethylation are mainly tRNAs and some other RNAs [29, 30]. In the case of tRNAs, DNMT2-mediated methylation is crucial to maintain the stability of specific tRNAs, notably $tRNA^{asp}$, $tRNA^{glu}$, and $tRNA^{val}$. As methylated tRNAs are essential to ensure protein fidelity and proper protein folding, AZA is implicated in disrupting these processes. Finally, AZA incorporation into RNA can directly impact mRNA stability itself [31]. Consequently, AZA's RNA affects several processes, including a reduction in cellular metabolic rates and protein synthesis. These effects are AZA-specific and are observed much more robustly compared to DAC [32, 33]. AZA-mediated RNA effects can thus partly explain the disconnect observed between the transcriptomic and immunopeptidomic changes. Moreover, an in-depth analysis of our immunopeptidomic data was

indicative of AZA's DNMT2 inhibition. We observed that proteins generating multiple DOIs after AZA treatment had significantly higher proportions of DNMT2-dependent amino acids (Figure 3.4G). Based on this, we believe that during protein translation, tRNA insufficiency of these amino acids leads to ribosome stalling, and consequently DRiPs generation and MAP presentation. To further confirm this hypothesis, we can perform amino acid sequence analysis of source proteins generating multiple DOIs and that of the DOIs themselves to identify the locations of Asp, Glu, and Val (DNMT2-dependent amino acids). This will provide useful information about how and where in the amino acid sequence, ribosome stalling probably took place. This could be leveraged for designing RNA vaccine targets by generating tailormade TA-coding sequences that enhance DRiPs generation after AZA treatment and thus, promoting the presentation of such TAs.

4.7 Amplifying the Anti-Leukemic Effects of AZA through Autophagy Inhibition

Apart from leading to enhanced DRiPs generation, tRNAs lacking DNMT2-dependent methylation are prone to codon infidelity leading to inaccurate protein synthesis. In particular, loss of DNMT2-mediated methylation leads to codon misrecognition by $tRNA^{asp}$ and $tRNA^{glu}$ [34]. Consequently, $tRNA^{glu}$ erroneously decodes Asp as Glu leading to a significant increase in Asp > Glu amino acid substitutions. In line with this, our immunopeptidomic data show higher levels of Glu compared to Asp in the upregulated DOIs compared to downregulated (Figure 3.4C). Finally, codon infidelity can lead to protein misfolding causing protein aggregation. This triggers cellular processes, including unfolded protein response responses and autophagy that are responsible for the clearance of protein aggregates. In our study, we found that autophagy induction was AZA-specific and not observed with DAC-treated cells. This indicates that autophagy is mediated by AZA's RNA effects, rather than its DNA demethylating effects. Ironically, AZA-induced

autophagy can abrogate the extent to which EREs are expressed since autophagy is responsible for viral clearance as well as ERE degradation [35, 36]. This could explain the absence of ERE-derived MAPs after AZA treatment. To test this hypothesis, we determined the levels of EREs when cells were treated with AZA combined with an autophagy inhibitor, Spautin-1. Spautin-1 is a novel compound previously shown to be effective in leukemia and acts by preventing the formation of autophagy initiation complexes by inhibiting the deubiquitinating activity of USP10 and USP13 [37, 38]. As expected, we observed that autophagy inhibition alone increases the levels of intracellular EREs (Figure 3.6B). The combination of Spautin-1 and AZA also followed a similar trend of higher ERE levels, except for LINEs (Figure 3.6B). However, there were no increases in AZA-induced EREs after the combination treatment (Figure S3.8B). We found that Spautin-1 and AZA had additive effects leading to a drastic drop in the proliferation rate of cells (Figure 3.6A, Figure S3.7, 8A). As the demethylation of the genome by AZA is dependent on actively proliferating cells, transcriptomic upregulation mediated by AZA was possibly inhibited. This could explain the absence of AZA-induced EREs when AZA was combined with Spautin-1. Therefore, for future directions, it is crucial to identify a treatment regimen for a combination treatment that involves autophagy inhibition while permitting cells to actively proliferate. This can be achieved by pre-treating cells with autophagy inhibitors to drop the basal levels of autophagy and then proceed with AZA treatment of cells. Transcriptomic and immunopeptidomic analysis on these treated cells will be crucial to understanding if we can identify AZA-induced EREs and ERE-derived MAPs. Alternatively, we could hypothesize that performing immunopeptidomic analysis on DAC-treated cells would indicate whether, in the absence of autophagy, ERE-derived MAPs are presented. In line with this, a study using DAC-treated glioblastoma cells has indeed shown that DAC induces the presentation of ERE-derived MAPs [39]. Taken together, AZA-

mediated RNA effects induce autophagy responses that counteract the extent of ERE-mediated immune effects of AZA.

4.8 AZA's hidden talents: Inhibition of the immunoproteasome

Another unexpected finding in our study was that AZA treatment decreases the IP activity of AML cells. As described in section 1.6.1, IPs are a specialized form of CPs that are expressed in the presence of IFN-γ or TNFα. IPs and CPs differ from each other in their composition of the core 20S proteolytic particle (Figure 1.10), endowing them with differential cleavage and catalytic properties. Notably, IPs have increased chymotryptic than tryptic activity and have cleavage preferences for peptides more specifically after C-terminal hydrophobic residues and basic residues, than acidic residues. Thus, IPs and CPs generate diverse MAPs thereby, shaping the immunopeptide differently. Our immunopeptidomic analyses revealed that AZA-induced DOIs derived frequently from tryptic peptides (Figure 3.4E), indicating a reduced contribution of IP to MAP changes. Consistent with this observation, we found a reduction in IP activity post-AZA treatment (Figure 3.4F, Figure S3.6A). Of note, IP activity has been previously shown to have a pro-survival effect in leukemia and IP inhibition leads to cell death in AML cells [40]. This suggests that the mechanism of cell death induced by AZA could be via its inhibition of IP activity. This can be validated if a synergic effect is obtained upon combining AZA with IP inhibitors. Alternatively, cells with higher expression of IP subunits might be more resistant to AZA. Preliminary analysis to test this hypothesis could be performed using transcriptomic data from primary AML patients post-AZA treatment, although this technique will not be able to determine the contributions of IP activity, itself.

4.9 Limitations of the study

As with most research studies, the current study is subject to limitations. First, we used cell lines as a model system to investigate transcriptomic and immunopeptidomic changes. The use of patient specimens pre- and post-AZA treatment would have been of higher physiological relevance. However, patient specimens are limited in their sample quantities, especially for MS studies. One potential solution to overcome this could be the use of mouse models to expand these primary leukemic cells. However, the selective proliferation of some clones can alter the clonal architecture of AML and may not be truly representative of the patients [41]. Moreover, there are potential problems in accurately distinguishing human from mouse peptides with current HLA-peptide binding prediction algorithms. Another possibility is the use of organoids for IP analysis. This is a field that is recently being explored and has been performed with successful results for colorectal cancer [42]. Secondly, our research is limited in focusing solely on the impact of AZA on AML cells. Given the importance of T-cells in the elimination of leukemic cells [43, 44], the role of hypomethylation of the genome in immune players is important to be considered as well. Finally, the combination of BCL-2 inhibitor, Ventoclax, and AZA has received successful results in the phase 3 VIALE-A study (NCT02993523) [45] and is now the newly FDA-approved treatment for patients with intensive chemotherapy-ineligible newly diagnosed AML. Therefore, studying the immunopeptidome of AZA in combination with Ventoclax is relevant and should be pursued.

4.10 Conclusion

The work presented in this thesis highlights the benefits and the challenges associated with using cell lines in scientific research. While cell lines offer a convenient model system for scientific investigations, their misidentification can lead to erroneous conclusions and hinder the progress of

research. Through our study of THP-1 cell lines obtained from different repositories, we have demonstrated the existence of numerous differences, including variations in morphology, expression of key leukemia-related genes, and LOHs in several genomic regions that affected their HLA types. Based on this work, we recommend that THP1 cells from DSMZ are more suitable for AML-based research, antigen presentation, and immunopeptidomic studies. Our findings also highlight the need for greater scrutiny and validation of cell lines used in research. To address this issue, we recommend an increase in the number of loci currently used for STR profiling, which can provide a more robust and reliable method for confirming the identity of cell lines.

Further, through this thesis, we applied a proteogenomic approach to characterize AZA-mediated transcriptome and immunopeptidomic changes in AML cells. Our findings support previous studies that have reported an upregulation of CTA-derived MAPs following AZA treatment. However, we also discovered that AZA treatment failed to upregulate ERE-derived MAPs, despite observing an increase in their RNA levels. Nevertheless, EREs played an important role in inducing immune responses and viral mimicry, which may contribute to the adjuvanticity of leukemic cells. We also identified that AZA treatment led to an increase in protein aggregation, resulting in the induction of autophagy responses due to their RNA-mediated effects. Our data suggest that AZA-induced autophagy was responsible for limiting the extent of EREs upregulation, which could explain the lack of ERE-derived MAP presentation. Overall, our findings suggest that CTAs, rather than ERE-derived MAPs may be more useful as cancer vaccine targets in AZA-alone treatments. Furthermore, our results suggest that inhibiting autophagy may enhance the presentation of ERE-derived MAPs and the anti-leukemic effects of AZA. Our study highlights the importance of further investigating the potential for combination therapies of AZA with autophagy inhibitors to enhance the immunogenicity of AML cells.

4.11 References

1. Liu, Y., et al., *Multi-omic measurements of heterogeneity in HeLa cells across laboratories.* Nature biotechnology, 2019. **37**(3): p. 314-322.
2. Figueroa, M.E., et al., *DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia.* Cancer Cell, 2010. **17**(1): p. 13-27.
3. Caron, E., et al., *The MHC I immunopeptidome conveys to the cell surface an integrative view of cellular regulation.* Molecular systems biology, 2011. **7**(1): p. 533.
4. Shraibman, B., et al., *Human Leukocyte Antigen (HLA) Peptides Derived from Tumor Antigens Induced by Inhibition of DNA Methylation for Development of Drug-facilitated Immunotherapy.* Mol Cell Proteomics, 2016. **15**(9): p. 3058-70.
5. Rooney, M.S., et al., *Molecular and genetic properties of tumors associated with local immune cytolytic activity.* Cell, 2015. **160**(1-2): p. 48-61.
6. Hazini, A., K. Fisher, and L. Seymour, *Deregulation of HLA-I in cancer and its central importance for immunotherapy.* Journal for Immunotherapy of Cancer, 2021. **9**(8).
7. Wetzler, M., et al., *HLA class I antigen cell surface expression is preserved on acute myeloid leukemia blasts at diagnosis and at relapse.* Leukemia, 2001. **15**(1): p. 128-133.
8. Berlin, C., et al., *Mapping the HLA ligandome landscape of acute myeloid leukemia: a targeted approach toward peptide-based immunotherapy.* Leukemia, 2015. **29**(3): p. 647-659.
9. Brouwer, R.E., et al., *Loss or downregulation of HLA class I expression at the allelic level in acute leukemia is infrequent but functionally relevant, and can be restored by interferon.* Human immunology, 2002. **63**(3): p. 200-210.
10. Masuda, K., et al., *Loss or down-regulation of HLA class I expression at the allelic level in freshly isolated leukemic blasts.* Cancer science, 2007. **98**(1): p. 102-108.
11. Boegel, S., et al., *HLA and proteasome expression body map.* BMC medical genomics, 2018. **11**(1): p. 1-12.
12. Goodyear, O.C., et al., *Azacitidine augments expansion of regulatory T cells after allogeneic stem cell transplantation in patients with acute myeloid leukemia (AML).* Blood, The Journal of the American Society of Hematology, 2012. **119**(14): p. 3361-3369.
13. Bauer, J., et al., *Mass Spectrometry-Based Immunopeptidome Analysis of Acute Myeloid Leukemia Cells Under Decitabine Treatment Delineates Induced Presentation of Cancer/Testis Antigens on HLA Class I Molecules.* Blood, 2018. **132**(Supplement 1): p. 5223-5223.
14. Srivastava, P., et al., *Induction of cancer testis antigen expression in circulating acute myeloid leukemia blasts following hypomethylating agent monotherapy.* Oncotarget, 2016. **7**(11): p. 12840.
15. Gang, A.O., et al., *5-Azacytidine treatment sensitizes tumor cells to T-cell mediated cytotoxicity and modulates NK cells in patients with myeloid malignancies.* Blood Cancer J, 2014. **4**: p. e197.
16. Atanackovic, D., et al., *Cancer-testis antigen expression and its epigenetic modulation in acute myeloid leukemia.* American journal of hematology, 2011. **86**(11): p. 918-922.
17. Oka, Y., et al., *WT1 peptide vaccine for the treatment of cancer.* Current opinion in immunology, 2008. **20**(2): p. 211-220.

18. Holmberg-Thydén, S., et al., *Therapeutic Cancer Vaccination Targeting Shared Tumor Associated Antigens in Combination with Azacitidine for High Risk Myelodysplastic Syndrome-a Phase I Clinical Trial.* Blood, 2020. **136**: p. 23-24.
19. Saini, S.K., et al., *Human endogenous retroviruses form a reservoir of T cell targets in hematological cancers.* Nature communications, 2020. **11**(1): p. 1-14.
20. Kroemer, G., et al., *Immunogenic cell stress and death.* Nature immunology, 2022. **23**(4): p. 487-500.
21. Grundy, E.E., N. Diab, and K.B. Chiappinelli, *Transposable element regulation and expression in cancer.* The FEBS journal, 2022. **289**(5): p. 1160-1179.
22. Xie, W., C. Liang, and J.A. Birchler, *Inhibition of RNA interference and modulation of transposable element expression by cell death in Drosophila.* Genetics, 2011. **188**(4): p. 823-834.
23. Sistigu, A., et al., *Cancer cell–autonomous contribution of type I interferon signaling to the efficacy of chemotherapy.* Nature medicine, 2014. **20**(11): p. 1301-1309.
24. Chiappinelli, K.B., et al., *Inhibiting DNA Methylation Causes an Interferon Response in Cancer via dsRNA Including Endogenous Retroviruses.* Cell, 2015. **162**(5): p. 974-86.
25. Patel, S.A. and A.J. Minn, *Combination cancer therapy with immune checkpoint blockade: mechanisms and strategies.* Immunity, 2018. **48**(3): p. 417-433.
26. Gamrekelashvili, J., et al., *Primary sterile necrotic cells fail to cross-prime CD8+ T cells.* Oncoimmunology, 2012. **1**(7): p. 1017-1026.
27. Chiaro, J., et al., *Viral molecular mimicry influences the antitumor immune response in murine and human melanoma.* medRxiv, 2020: p. 2020.09. 09.20191171.
28. Chapuis, A.G., et al., *T cell receptor gene therapy targeting WT1 prevents acute myeloid leukemia relapse post-transplant.* Nature medicine, 2019. **25**(7): p. 1064-1072.
29. Lu, L.-J.W. and K. Randerath, *Mechanism of 5-azacytidine-induced transfer RNA cytosine-5-methyltransferase deficiency.* Cancer Research, 1980. **40**(8_Part_1): p. 2701-2705.
30. Lu, L.W., et al., *Drug effects on nucleic acid modification. I. A specific effect of 5-azacytidine on mammalian transfer RNA methylation invivo.* Biochemical and Biophysical Research Communications, 1976. **68**(4): p. 1094-1101.
31. Aimiuwu, J., et al., *RNA-dependent inhibition of ribonucleotide reductase is a major pathway for 5-azacytidine activity in acute myeloid leukemia.* Blood, The Journal of the American Society of Hematology, 2012. **119**(22): p. 5229-5238.
32. Schaefer, M., et al., *Azacytidine inhibits RNA methylation at DNMT2 target sites in human cancer cell lines.* Cancer Res, 2009. **69**(20): p. 8127-32.
33. Hollenbach, P.W., et al., *A comparison of azacitidine and decitabine activities in acute myeloid leukemia cell lines.* PloS one, 2010. **5**(2): p. e9001.
34. Tuorto, F., et al., *The tRNA methyltransferase Dnmt2 is required for accurate polypeptide synthesis during haematopoiesis.* The EMBO journal, 2015. **34**(18): p. 2350-2362.
35. Xu, C., et al., *The Autophagy Cargo Receptor SQSTM1 Inhibits Infectious Bursal Disease Virus Infection through Selective Autophagic Degradation of Double-Stranded Viral RNA.* Viruses, 2021. **13**(12).
36. Paget, M., et al., *Stress granules are shock absorbers that prevent excessive innate immune responses to dsRNA.* BioRxiv, 2022: p. 2021.04. 26.441141.
37. Shao, S., et al., *Spautin-1, a novel autophagy inhibitor, enhances imatinib-induced apoptosis in chronic myeloid leukemia.* International journal of oncology, 2014. **44**(5): p. 1661-1668.

38. Liu, J., et al., *Beclin1 controls the levels of p53 by regulating the deubiquitination activity of USP10 and USP13.* Cell, 2011. **147**(1): p. 223-234.
39. Kong, Y., et al., *Transposable element expression in tumors is associated with immune infiltration and increased antigenicity.* Nat Commun, 2019. **10**(1): p. 5228.
40. Rouette, A., et al., *Expression of immunoproteasome genes is regulated by cell-intrinsic and–extrinsic factors in human cancers.* Scientific reports, 2016. **6**(1): p. 34019.
41. Kawashima, N., et al., *Comparison of clonal architecture between primary and immunodeficient mouse-engrafted acute myeloid leukemia cells.* Nature Communications, 2022. **13**(1): p. 1624.
42. Newey, A., et al., *Immunopeptidomics of colorectal cancer organoids reveals a sparse HLA class I neoantigen landscape and no increase in neoantigens with interferon or MEK-inhibitor treatment.* Journal for immunotherapy of cancer, 2019. **7**: p. 1-15.
43. Yau, H.L., et al., *DNA hypomethylating agents increase activation and cytolytic activity of CD8+ T cells.* Molecular Cell, 2021. **81**(7): p. 1469-1483. e8.
44. Grimm, J., et al., *Azacitidine-induced reconstitution of the bone marrow T cell repertoire is associated with superior survival in AML patients.* Blood Cancer Journal, 2022. **12**(1): p. 19.
45. DiNardo, C.D., et al., *Azacitidine and venetoclax in previously untreated acute myeloid leukemia.* New England Journal of Medicine, 2020. **383**(7): p. 617-629.