PhD degree in Systems Medicine (curriculum in Molecular Oncology)

European School of Molecular Medicine (SEMM),

University of Milan and University of Naples “Federico II”

A multi-OMICs approach to dissect aberrant epigenetic mechanisms in Triple Negative Breast Cancer

Settore disciplinare: BIO/10

Giulia Robusti

Tutor: Dr. Tiziana Bonaldi

European Institute of Oncology (IEO), Milan, Italy

PhD Coordinator: Prof. Saverio Minucci

Anno accademico 2022-2023

Table of Contents

List of abbreviations

ac	Acetylation
ACN	Acetonitrile
AML	Acute Myeloid Leukemia
BC	Breast Cancer
BET	Bromodomain and Extra-Terminal domain
BL1	Basal-like 1
BL2	Basal-like 2
ChIP	Chromatin Immunoprecipitation
ChIP-seq	ChIP-sequencing
CPM	Count Per Million
CPTAC	Clinical Proteomic Tumor Analysis Consortium
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CRISPRa	CRISPR activation
CRISPRi	CRISPR interference
dCas9	Deactivated Cas9
DDA	Data Dependent Acquisition
DIA	Data Independent Acquisition
DNMT	DNA Methyltransferase
DNMTi	DNA Methyltransferase inhibitors
ER	Estrogen Receptor
ESI	Electrospray Ionization
FAD	Flavin Adenine Dinucleotide
FDA	Food and Drug Administration
FDR	False Discovery Rate
FFPE	Fresh Frozen Paraffin Embedded
FUCCI	Fluorescent Ubiquitination-based Cell Cycle Indicator
GO	Gene Ontology
GSEA	Gene Set Enrichment analysis
HATs	Histone Acetyltransferases
HCD	Higher Energy Collisional Dissociation
HDACi	HDAC inhibitors
HDACs	Histone Deacetylases
HDMs	Histone Demethylases
HER2	Epidermal Growth Factor receptor 2
HME	Histone Modifying enzymes
HMTs	Histone Methyltransferases
HP1	Heterochromatic protein 1
HPLC	High Performance Liquid Chromatography
iBAQ	intensity-Based Absolute Quantification
IGF2BP3	Insulin-like growth factor-2 RNA-binding protein 3
IHC	Immunohistochemistry
IM	Immunomodulatory subtype
JmJC	Jumonji C domain-containing proteins
K	Lysine
KMTs	Lysine methyltransferases
KRAB	Krüppel Associated Box

L/H	Light to Heavy ratio
LAR	Luminal Androgen Receptor subtype
LC	Liquid Chromatography
LFQ	Label Free Quantification
lincRNAs	Long intergenic non-coding RNAs
M	Mesenchymal and claudin-low subtype
m/z	Mass to charge ratio
MALDI	Matrix-assisted laser desorption ionization
me	Methylation
me1	Mono-methylation
me2	Di-methylation
me3	Tri-methylation
miRNA	MicroRNA
MS	Mass Spectrometry
MSL	Mesenchymal stem–like subtype
NGS	Next Generation sequencing
OCT	Optimal Cutting Temperature
PADI4	Peptidyl arginine deiminase 4
PCA	Principal Component analysis
PCoA	Principal Coordinates Analysis
PHD	Plant homeodomain
PIC	Phenyl-isocyanate
PR	Progesterone receptor
PRC2	Polycomb Repressive Complex 2
PRMTs	Arginine methyltransferases
PRO	Propionic anhydride
PTMs	Post Translational Modifications
Q	Quadrupole
R	Arginine
RA	Relative Abundance
RIN	RNA Integrity Number
RNAi	RNA interference
RNA-seq	RNA-sequencing
SAM	S-adenyl methionine
SE	Super-enhancer
sgRNA	Single guide RNA
shRNA	Short hairpin RNAs
SILAC	Stable Isotope Labelling by Amino acids in Cell culture
siRNA	Small interfering RNAs
snoRNA	Small nucleolar RNAs
TCGA	The Cancer Genome Atlas
TES	Transcription End Site
TFA	Trifluoroacetic acid
TFs	Transcription Factors
TILs	Tumor-infiltrating lymphocytes
TN	Triple Negative
TNBC	Triple Negative Breast Cancer

TOF	Time of Flight
TPM	Transcript per Million
TSS	Transcription Start Site
WB	Western Blot
X-ChIP	Crosslinking ChIP
XICs	EXtracted Ion Chromatograms
Δm	Delta mass

Figure Index

Table Index

1. ABSTRACT

In recent years, experimental evidence has shown that epigenetic changes including DNA methylation and histone post-translational modifications (PTMs) play a role in cancer development and progression. Histone PTMs have been largely studied in the context of cancer and have been described as hallmarks of cancer and markers of diagnosis and prognosis. In addition, the histone modifying enzymes involved in the recognition, deposition, and removal of histone PTMs are often altered and overexpressed in cancer. Therefore, the identification of these enzymes as potential targets for epigenetic therapies is promising for the development of specific treatments.

In this project, we focused on the molecular characterization of epigenetic mechanisms in Triple Negative breast cancer (TNBC), which is the most aggressive type of breast cancer, characterized by a high proliferation rate, and associated with poorer patient outcomes compared to other breast cancer subtypes, particularly Luminal A subtype. Because of the absence of specific molecular markers, there are no targeted therapies available for TNBC and only few treatment options are in use, including standard chemotherapy, surgery, and radiation therapy. Therefore, there is an urgent need to identify specific markers for this type of tumor as well as novel potential targets.

Preliminary data obtained by our group, by mass spectrometry (MS), highlighted an epigenetic signature distinguishing the TNBC subtype from the other breast cancer subtypes. We confirmed the same pattern in an additional set of samples and we selected three of the identified changes (H3K4me2, H3K9me3, and H4K20me3) for a follow-up mechanistic investigation, using two parallel approaches.

We first performed a multi-OMICs analysis to investigate the genomic distribution of H3K4me2 in TNBCs and its effects at the expression level. By integration of ChIP-seq and RNA-seq data from a set of fresh frozen clinical samples we found that the localization of H3K4me2 at promoters and super-enhancer regions results in higher levels of expression and seems to drive expression of genes associated to the TNBC phenotype. By the integration of data from clinical samples and the model breast cancer cell line MDA-MB-231 we identified a set of genes that are candidate TN-specific genes whose expression is regulated by the unique presence of H3K4me2 in their cis-regulatory regions. We then took advantage of a CRISPRi epigenetic tool to genetically manipulate the expression of these genes through the LSD1-mediated removal of H3K4me2 at these regulatory regions. The

result obtained revealed a direct correlation between the presence of H3K4me2 at promoters and gene expression.

Then, we investigated the possible upstream causes of histone PTM deposition focusing on H3K9me3 modification, which we found to be overexpressed in TN versus Luminal A tumors. By correlating the levels of histone PTMs in TNBC samples with the expression levels of the histone modifying enzymes by RNA-seq analysis, we found a strong correlation between the increase in H3K9me3 levels and the over-expression of the H3K9-selective methyltransferases SETDB1, SUV39H1, and SUV39H2 in TN subtype. Upon pharmacological and genetic manipulation of these methyltransferases, we identified a strong effect at phenotypic level. In addition, the knockdown of SUV39H1 and SUV39H2, but not SETDB1, revealed a decrease in H3K9me3 levels paralleled by an increase in H4K20me3 levels, suggesting a potential mechanism of crosstalk. Hence, we hypothesized that upon depletion of SUV39H1 and SUV39H2 a re-localization of both H3K9me3 and H4K20me3 may occur. To assess this hypothesis, we used immunofluorescence to analyze the localization of H3K9me3 and H4K20me3 on mitotic chromosomes and correlated these data with ChIP-seq analysis of these two epigenetic marks in a set of Luminal A and TN cell lines. At basal condition, we observed a more pronounced localization of these histone marks at centromeric regions in Luminal A cells compared to TN. Interestingly the same analysis carried out in TN cell lines depleted for SUV39H1 and SUV39H2 revealed a re-localization at centromeres of the two PTMs, which acquired a distribution much more similar to that of Luminal A cell lines. Additional experiments are needed to better characterize the molecular mechanism at the base of this potential crosstalk.

Overall the results obtained indicate that our integrative approach in the analysis of aberrant epigenetic features, based on the quantitative MS analysis, and the integration of ChIP-seq and RNA-seq data can suggest novel molecular targets in the TNBC subtype, which we studied through pharmacological and genetic modulation.

2. INTRODUCTION

2.1 Epigenetics

The term “Epigenetics” was coined in 1942 by C.H. Waddington (1), who tried to interpret the connection between genotype and phenotype by referring to all those changes occurring in addition to those reported by traditional genetics. Indeed, epigenetics includes a complex set of processes that result in heritable phenotypic changes without changes in the DNA sequence. Epigenetic processes include DNA methylation, histone post-translational modifications (PTMs), chromatin remodelers, and non-coding RNAs, such as microRNAs (2). DNA methylation is the most studied epigenetic mark and it involves the transfer of a methyl-group to the C5 position of cytosine in order to form 5-methylcytosine. DNA methylation regulates gene expression by recruiting proteins for gene silencing or by inhibiting the binding of transcription factors (TFs) to DNA. In particular, methylation is catalyzed by a family of DNA methyltransferases known as DNMTs that transfer a methyl group from S-adenyl methionine (SAM) to cytosine residues (3). Histone PTMs are found in distinct regions of the genome, especially at gene promoters and gene bodies, where their presence is linked to transcriptional activity. Chromatin remodelers are versatile ATP-dependent enzymes that catalyze a broad range of chromatin remodeling reactions and are involved in nucleosome positioning and allow accessibility of DNA to various regulatory factors, such as transcription factors and components of DNA replication (4). In addition, numerous studies have suggested the role of non-coding RNA presence on chromatin and their role in the regulation of chromatin structure and status (5). All these features can change during the cell life and are influenced by the external environment, determining biological signals that can affect the phenotype. However, failure to properly maintain these epigenetic marks can result in altered cell homeostasis and disease, including cancer, where they are often linked to inappropriate expression or silencing of oncogenes and tumor suppressor genes.

In the next paragraphs, I will provide a global overview about histone PTMs, histone modifying enzymes, and the role of epigenetic alteration in the context of cancer. In addition, I will describe the possible strategies to investigate the molecular mechanisms underlying aberrant epigenetic features in the context of cancer.

2.2 Histones and histone post-translational modifications

Histones are small (11-16kDa) basic proteins, rich in positively charged arginine (R) and lysine (K) residues, which interact with the DNA phospho-negative groups. They are characterized by a C-terminal globular domain and an N-terminal tail that protrudes outside the nucleosome. In the nucleus of eukaryotic cells, histones bind to DNA to form the nucleosome (6), the basic unit of the chromatin. In the nucleosome, around 147 bp of DNA are wrapped around the so called "core histone octamer", composed by two copies of histone H2A and H2B and a dimer of histone H3 and histone H4. In addition, a linker histone H1 stabilizes the chromatin to form higher order chromatin structures by binding the nucleosome and protecting the 80 bp-long linker DNA that is present between the nucleosomes and which is more subjected to nuclease digestion (7). Nucleosomes are present along the genome to form a filament characterized by different levels of compactness with the highest degree of condensation found in the metaphasic chromosome (**Figure 1**). Based on nucleosome density, chromatin compaction, and DNA accessibility, two types of chromatin exist: euchromatin and heterochromatin. Euchromatin is more accessible and is associated with transcriptionally active regions, while heterochromatin is associated with gene silencing and more compact chromatin. Constitutive heterochromatin is always condensed and characterized by non-coding, highly repetitive regions such as pericentric and sub-telomeric regions, while facultative heterochromatin is characterized by genomic regions, which can switch between transcriptionally active and inactive chromatin states during development and differentiation (8).

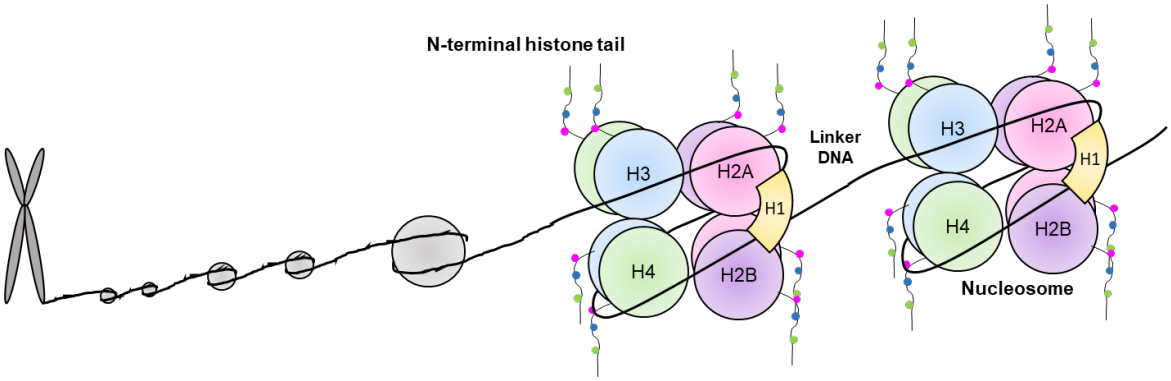


Figure 1: Nucleosome structure and core histone octamer

The nucleosome, the basic unit of the chromatin, is characterized by 147 bp of DNA wrapped around the histone octamer composed by two copies of histone H2A and H2B and a tetramer of histone H3 and H4. Linker DNA is present between nucleosome and histone tails protrude from the core histone octamer. Histone H1 may interact with exit and entry sites of nucleosomal DNA and linker DNA.

2.2.1 Histone post-translational modifications

Histones are characterized by a number of PTMs that are mainly present at their N-terminal tail (**Figure 2**). Histone PTMs were reported on over 60 different residues either by specific antibodies or by mass spectrometry (9) and represent not only an epigenetic mark involved in many cellular processes and the DNA repair but are also important for the maintenance of chromatin structure and compaction (10). Histone PTMs include methylation, acylations (of which acetylation is the most abundant), phosphorylation, SUMOylation, ubiquitylation, ADP-ribosylation, deamination, as well as other less common modifications such as crotonylation, and butyrylation (10-12). Methylation (me) and acetylation (ac) are the most common and most studied histone PTMs, especially in the context of diseases. Acetylation of histone tails is associated with transcription activation, as the addition of the acetyl group neutralizes the positively charged $–NH_2$ group of lysines and reduces the electrostatic interaction between histones and DNA, resulting in a more open structure that is more accessible to the transcriptional machinery (13). Acetylation is generally localized at regulatory regions of active genes, promoters, and enhancer regions, with a pattern that is conserved from yeast to human. In addition, levels of histone acetylation are often proportional to levels of gene expression (14). Differently from acetylation, histone methylation mainly occurs on the side chains of lysine and arginine residues without altering the charge of histone proteins, and can cause transcriptional activation or repression depending on the location and the type of residue (10-12). Moreover, there is an added level of complexity to keep in consideration, since lysine residues can be mono- (me1), di- (me2), or tri-(me3) methylated and arginine residues can be mono-methylated and symmetrically or asymmetrically di-methylated (11).

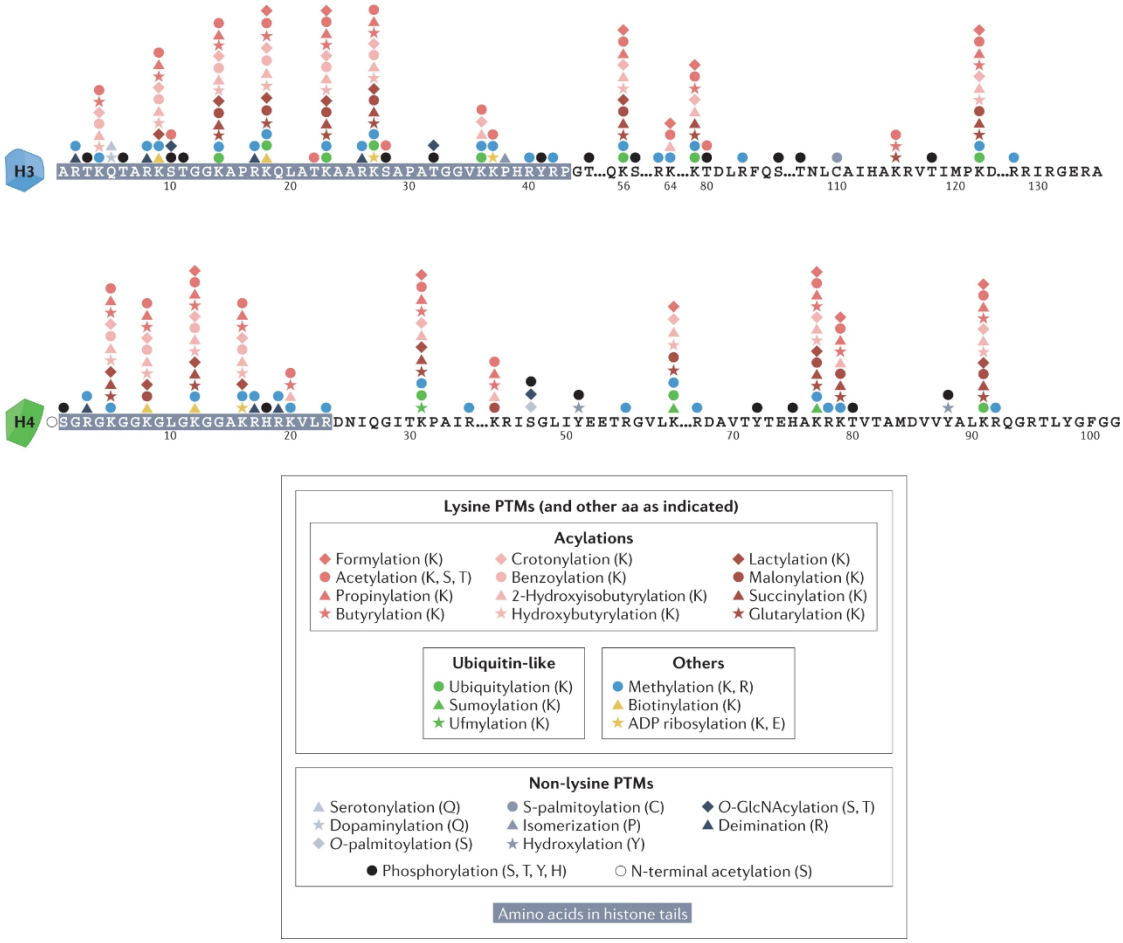


Figure 2: Histone PTMs reported for histone H3 and histone H4

Amino acid sequences of histones H3 and H4 are displayed with amino acids located within the histone tails highlighted with a grey background. The different PTMs are displayed by different symbols which are explained in the legend. This figure was adapted from (15).

In general, H3K4, H3K36, and H3K79 methylation is associated with transcriptional activation, while H3K9, H3K27, and H4K20 methylation is linked to transcriptional repression (**Figure 3**). More specifically, H3K4me3 is one of the best characterized histone PTMs and localizes at transcription start sites (TSS) in promoter regions of the most active genes in eukaryotic cells. Therefore, H3K4me3 levels closely correlate with active transcription (16). In addition to H3K4me3, the sequence nearby the TSS is also broadly enriched for H3K4me2, which is also present at distal gene regulatory elements, better known as enhancers. The major characterized marks at these regions are H3K4me1 and H3K27ac. H3K79 methylation seems to slightly overlap with H3K4 methylation marks and to be associated as well with active transcription (17). Inactive promoters are characterized by the presence of H3K27me3 or H3K9me3, which are instead generally distributed at gene bodies. H3K27me3 is a hallmark of transcriptional repression associated with facultative heterochromatin (18), while H3K9me3 is a marker of constitutive heterochromatin as it is

enriched at transcriptionally silent regions and repetitive regions of the genome. H4K20me3 is also associated with transcriptional repression and is generally located at repetitive regions (19). Finally, H3K36 methylation is associated with actively transcribed regions, and is associated with transcriptional elongation (17).

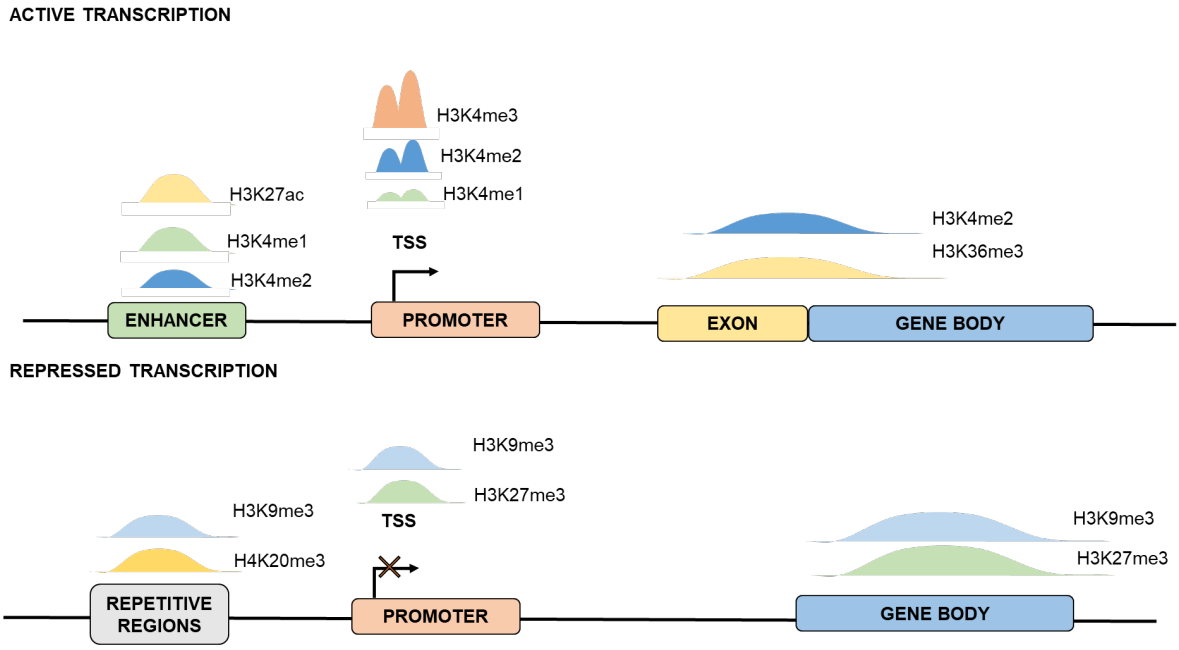


Figure 3: Genome-wide distribution of the most studied histone PTMs in association with gene activation and gene silencing

Histone PTMs of lysine residues of histone H3 and histone H4 are displayed at different genomic regions including promoters, enhancers, exons, gene body and repetitive regions.

2.2.2 Histone variants

In addition to canonical forms, variants of core and linker histone are also present and contribute to chromatin structure and regulation of gene expression (20). The "canonical" histones H1, H2A, H2B, H3, and H4 are encoded by multiple gene copies in higher eukaryotes, are highly and exclusively expressed in the S phase of the cell cycle, and are placed in the chromatin in a replication-dependent manner. However, a smaller fraction of the histone pool includes histone "variants" that are slightly or very different from the canonical counterparts and are expressed throughout the cell cycle (20,21). In humans, variants of histone H2A, H3, and linker histone H1 family are the most prominent. Eight variants of histone H3 have been identified and include H3.1, H3.2, and H3.3 (which are also the most abundant), testis specific variants (H3T and H3.5), primate specific variants (H3.X and H3.Y) and CENP-A, which plays an essential role in cell division as it localizes

at the centromeres (20). Histones H2A and H2B have respectively 20 and 17 variants. Among them, H2A.Z, macroH2A, and γ-H2A.X, a sensitive molecular marker of DNA damage, are the most characterized. Even though histone H4 was believed to exist as a single histone sequence, a single variant, H4G, has been recently described (22). Linker histone H1 is characterized by seven somatic variants, histones H1.1 to H1.5, which are replication-dependent, and histones H1.0 and H1x, which are transcribed throughout the cell cycle (23). Histone H1 variants play a role in regulating nuclear functions by binding to the nucleosome in different ways and can create distinct higher-order chromatin structures. Variants not only have different expression patterns compared to the standard histones but also possess a distinct set of PTMs and are distributed differently throughout the genome, resulting in diverse chromatin arrangements. For example, the presence of H3.3 on gene promoters is associated with active transcription, while CENP-A is essential in cell division processes. Moreover, anomalies in the deposition of these variants have been observed in various pathological conditions, including cancer (20).

2.3 Writers, erasers and readers of histone PTMs

Histone "writers", "erasers", and "readers" are key players in the dynamic regulation of chromatin structure and gene expression. Histone "writers" include a set of enzymes, which are involved in the deposition of histone PTMs, while histone "erasers" are enzymes responsible for removing the histone modifications. These enzymes are better known as histone modifying enzymes (HMEs) and considering acetylation and methylation the most studied histone PTMs, HMEs responsible for deposition and removal of these modifications have been extensively studied (24). Histone acetyltransferases (HATs) and histone deacetylases (HDACs) are the enzymes responsible for acetylation and deacetylation and show low specificity, as they are able to bind to multiple histone sites and catalyze the acetylation of many other non-histone proteins. On the contrary, methyltransferases (HMTs) and demethylases (HDMs) are responsible for methylation of histone H3 and H4 tails, and differently from acetylation, they show higher specificity, as they bind only to specific residues (24). They usually modify one single lysine on a single histone and result in either activation or repression of transcription.

In addition, histone PTMs are specifically recognized by "reader" proteins that translate the epigenetic marks into functional outcomes. These reader proteins have specialized domains, such as bromodomains, chromodomains, or Plant homeodomain (PHD) finger domains, that

can recognize and interact with specific histone modifications (25). By binding to the modified histones, reader proteins recruit other factors, such as transcriptional co-activators or co-repressors, to regulate gene expression and chromatin structure (25). Together, histone writers, erasers, and readers form a complex regulatory network that controls the accessibility of DNA and influences gene expression patterns. Their precise interplay ensures the maintenance of cellular identity, proper development, and response to environmental signals.

2.3.1 Writers

HATs are responsible for the transfer of an acetyl group from the cofactor acetyl-CoA to the nitrogen of a lysine side chain and for this reason, they are also known as lysine acetyltransferases (26). In general, HATs mediate different biological processes, such as DNA repair and cell cycle progression (26). Three major HAT families have been described in humans: GNAT (HAT1, GCN5, PCAF), MYST (Tip60, MOF, MOZ, MORF, HBO1), and p300/CBP (27). In general, these enzymes modify more than one lysine with limited specificity. This is not the case of HMTs which instead recognize specific residues. Two types of HMTs have been recognized: lysine methyltransferases (KMTs) (**Figure 4**) and arginine methyltransferases (PRMTs) (24). KMT subfamilies containing SET domains include the SUV39, SET1, SET2, EZ, RIZ, SMYD, and SUV4-20 families and SET7/9 and SET8 (28,29), while there is only one non-SET domain, DOT1L. In addition, 9 different PRMTs have been described and have been divided into three groups (Types I, II, and III) based on asymmetric, symmetric di-methylation, and mono-methylation (30). The SUV39 HMT family was the first to be identified and is named after the first mammalian KMT identified as the homolog of *D. Melanogaster* Su(var)3-9 and is involved in the specific methylation of the heterochromatic mark H3K9 (31). Among the most studied HMT, EZH2 has been described to mediate the deposition of H3K27me3 as part of the Polycomb Repressive Complex 2 (PRC2). The SET1 family is involved in H3K4 methylation. Members of this family are conserved from yeast to mammals and include MLL1, MLL2, MLL3, MLL4, SET1A, and SET1B. In particular, MLL proteins are part of complexes that contain specific cofactors like WDR5 or Menin which play an important role in the recognition of methylated sites (32). Other methyltransferases include the SET2 family, mainly involved in methylation of H3K36 and the SMYD family, involved in methylation of H3K4 residues. For methylation of histone H4, SET8 is the only methyltransferase involved in mono-methylation of H4K20, while the SUV4-20 family, which includes SUV4-

20H1 and SUV4-20H2, contributes to di- and tri-methylation of H4K20 (33). Finally, DOT1L is the only methyltransferase known to methylate the globular domain of H3K79 and has a role in the regulation of many cellular processes such as transcription, development, proliferation, and differentiation of normal cells (32,34). The main methyltransferases of lysine residues are shown in **Figure 4**.

2.3.2 Erasers

HDACs are typically present within larger, co-repressor multi-protein complexes and regulate the expression and activity of many proteins by removal of acetyl groups and by preventing the transcription of genes, as they are recruited at promoter regions (35,36). Until now, 18 HDACs have been identified in mammalian cells and have been divided into four main classes based on sequence similarities: Class I (HDAC1, 2, 3, and 8), Class II (HDAC4, 5, 6, 7, 9, and 10), Class III or Sirtuins (SIRT1, 2, 3, 4, 5, 6, and 7), and Class IV (HDAC11) (37). Classes I, II, and IV share common features, in particular the zinc dependent mechanisms for their enzymatic activity, whereas class III enzymes have a different structure and depend on NAD^+. Lysine demethylases are enzymes that play a crucial role in the regulation of gene expression by removing methyl groups from lysine residues in histone proteins. KDMs specific for one or more residues have been identified (**Figure 4**), and have been categorized into two main families: KDM1, which is characterized by an amine oxidase domain, and Jumonji C domain-containing proteins (JmJC). The KDM1 family is composed by two members KDM1A (LSD1), and KDM1B (LSD2), while the second and largest family of KDM includes KDM2, KDM3, KDM4, KDM5, KDM6, and KDM7 (38). Instead, loss of arginine methylation occurs mainly through PADI4 (peptidyl arginine deiminase 4), which is not considered as a demethylase, as it causes conversion of arginine in citrulline (39). Only one arginine demethylase, JMJD6, has been identified with activity *in vivo* (40).

The KDM1 family is involved in demethylation of mono- and di-methylated residues. LSD1 is probably the most extensively studied demethylase, as it was the first to be identified in the context of epigenetic regulation (41,42). Through its flavin adenine dinucleotide (FAD)-dependent amine oxidase activity, it is involved in demethylation of mono- and di-methylated H3K4 and H3K9, resulting respectively in gene silencing or gene expression. Various transcription factors have been reported to recruit LSD1 to gene promoters as well as enhancers, leading to H3K4 demethylation (42). Moreover, it has been recently found to demethylate mono-methylated and di-methylated H4K20 and to contribute to the balance of

several other methylated lysine residues of histone H3 such as H3K27, H3K36, and H3K79, as well as non-histone proteins, such as p53 (41). KDM1B is another FAD-dependent amino oxidase homolog, which specifically targets H3K4me1 and H3K4me2 and mainly associates with the gene body regions of actively transcribed genes (43). Among the JmJC domain-containing proteins, around 20 are lysine demethylases (44). Differently from KDM1, this family of enzymes is capable of removing also trimethylations (45). The KDM2 subfamily, which includes KDM2A and KDM2B, is involved in the specific demethylation of H3K36me2 (46), while KDM2B can also demethylate H3K4me3 (47). KDM3A and KDM3B are specific demethylases of H3K9me1 and H3K9me2. The KDM4 family is involved in the activation of gene expression and specifically demethylates H3K9me2, H3K9me3, H3K36me2, and H3K36me3. The KDM5 subfamily has catalytic activity towards H3K4me2 and H3K4me3. H3K27me3 is instead modulated by the KDM6 family. KDM7B(PHF8) is involved in demethylation of H4K20 (38).

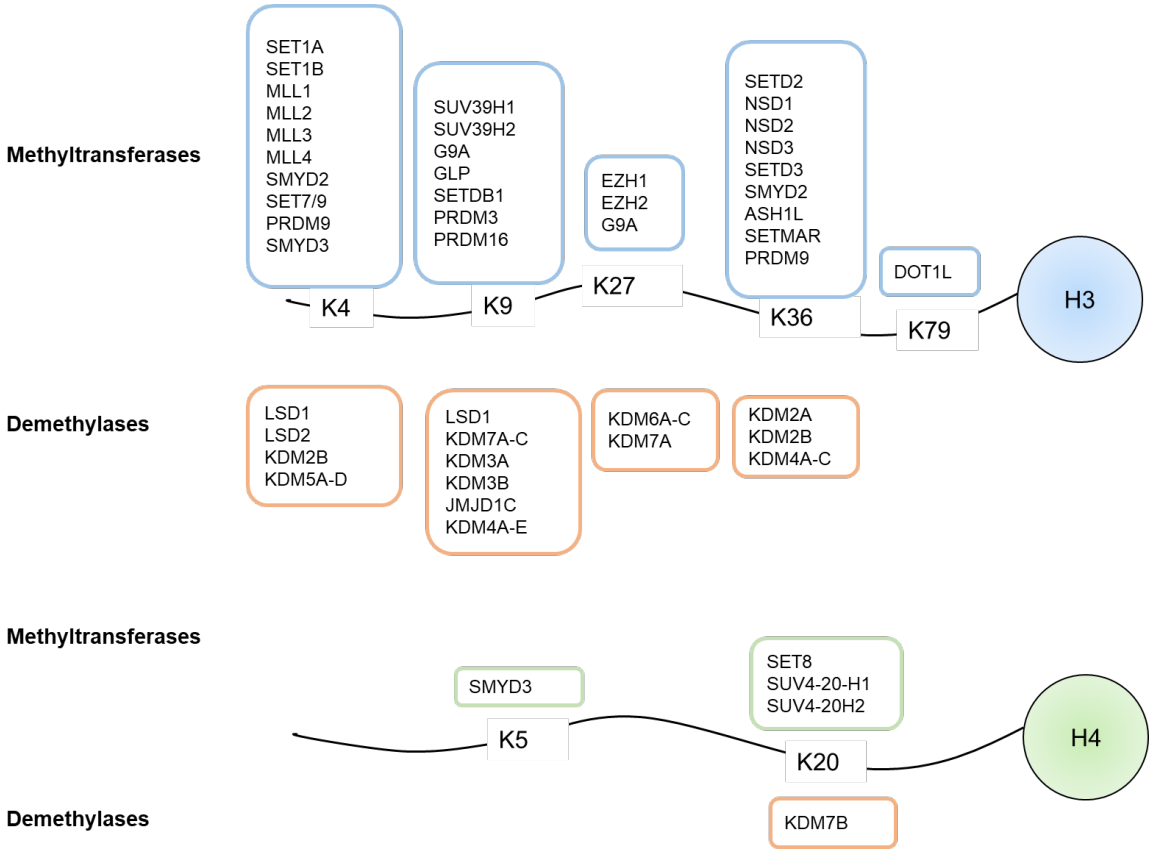


Figure 4: Histone methyltransferases and histone demethylases for specific lysine residues of histone H3 and histone H4

The figure displays the histone methyltransferases and demethylases for the most common methylated lysines of histone H3 and for histone H4.

2.3.3 Readers

One of the first domains to be identified in histone readers was the bromodomain. By recognition of acetylated residues, bromodomains facilitate the recruitment of other protein complexes, such as transcription factors and chromatin remodelers, to specific genomic regions like gene promoters and enhancers, leading to the activation of gene transcription (48). One of the most studied bromodomain-containing family is the group of BET (bromodomain and extra-terminal domain), which employs the bromodomain to bind the modified histone and regulate transcription by interaction with the transcriptional machinery. In particular, BET proteins bind to histone at acetylated lysine regions located in regions occupied by multiple enhancers (49). Chromodomains, instead, are responsible for the recognition and binding of methylated lysine residues on histone tails. They are often found in proteins involved in gene silencing, such as PRC2, which plays a critical role in maintaining cellular identity and developmental processes (50). The chromodomain of HP1 binds H3K9me3 at heterochromatic regions (51). PHD fingers are versatile histone readers that can recognize different histone modifications, most of them bind mainly unmodified or methylated states of H3K4. PHD finger domains are frequently found in proteins associated with transcriptional regulation and chromatin remodeling (52). The Tudor domain binds to methylated arginine residues on histone tails and plays a role in chromatin organization and DNA repair. Tudor domain-containing proteins are involved in diverse cellular processes, including transcriptional regulation, RNA processing, and epigenetic inheritance (50).

2.4 Histone PTM alterations and cancer

Even though cancer has conventionally been described as a genetic disorder, it is now recognized that the development and progression of cancer also involve epigenetic changes in addition to genetic alterations. Indeed, disruption of epigenetic processes can lead to altered gene function resulting in cellular neoplastic transformation, and cancer initiation. The first epigenetic changes observed in cancer identified significant alterations in DNA methylation with for instance genome-wide hypo-methylation and a site-specific CpG island promoter hyper-methylation (53). In addition, histone PTMs have been shown to also contribute to neoplastic pathologies and different PTM or combinations have emerged to play roles in different cancer types as general hallmarks, diagnostic, and prognostic markers. Because of their involvement in cancer, altered HMEs have raised significant interest in the field of targeted therapeutic approaches. These enzymes play an essential role in modulating gene expression patterns by modifying histones, thereby influencing gene expression and

cellular functions. The identification of HMEs as potential targets for epigenetic therapy is a promising avenue for the treatment of several diseases and in particular cancer treatment. The field of epi-drugs holds the potential to revolutionize cancer treatments, offering therapies that address the unique epigenetic signatures of individual patients, ultimately improving treatment efficacy and minimizing adverse effects.

2.4.1 Histone PTMs as hallmarks of cancer and their role in cancer diagnosis

In 2005, Fraga et al. proposed the loss of H4K20me3 and H4K16ac as epigenetic hallmarks of cancer, as their global decrease was found in different tissues, cell lines, and various primary tumors using immune-detection, high-performance capillary electrophoresis, and mass spectrometry. Moreover, they found that these PTM alterations occur early during tumorigenesis and become more accentuated at advanced stages (54). Loss of acetylation was correlated with reduced recruitment at repeated sequences of the acetyltransferases MOZ, MOF, and MORF while not finding any direct correlation with the histone modifying enzymes responsible for the deposition of H4K20me3, Fraga and colleagues hypothesized that the loss of this mark at DNA repetitive sequences could derive from a global loss of DNA methylation at these regions. After this breakthrough study, which demonstrated for the first time the association between alterations in histone PTMs and cancer, increasing interest led to further studies with the aim to investigate the link between histone PTMs and cancer development. For instance, a more recent study by our group showed a general decrease of the H3K14ac mark in different tumor types compared to normal tissues. Importantly, this histone mark also decreased in cancer cell populations that were laser microdissected from breast cancer tumors, suggesting the specificity of this change in tumor cell populations and thus representing a potential novel epigenetic hallmark of cancer (55). In addition, other PTMs seem instead to be cancer-specific or subtype-specific. For example, decreased levels of H3K18me1 were found in Luminal A breast cancer, and a decrease of H3K27me1 was detected in ovarian cancer when compared to normal tissues (55). An increase in H3K9 methylation, leading to aberrant gene silencing, has been found in various form of cancer including gastric (56), triple negative breast cancer, ovarian, and prostate cancer (55). Finally, a general loss of H3K27me3 is also found in several cancer types compared to normal tissues (57).

Histone PTM profiling has been also considered in the context of cancer diagnosis. While histone levels and their PTMs in tumor tissues could be potential cancer biomarkers, their applicability is limited because of invasive biopsies. While tissue tumor biomarkers can be

useful in advanced phases of the disease, to stage the tumor, and to define prognosis, they are not applicable to early diagnosis or to analyze tumors located in hard-to-reach areas like the brain. Instead, easily obtainable fluids such as blood and urine could be used to identify biomarkers for early diagnosis, and to date, blood is the most extensively used source of material for biomarker discovery. Although there are only few studies, there is increasing evidence that the detection of histone PTMs in circulating nucleosomes from patient sera is representative of the tumor epigenetic landscape. Thus, histone PTM detection in circulating nucleosomes could be considered as potential diagnostic markers. (58). Nucleosomes are released in the bloodstream from cancer tissues as a consequence of cell death and apoptosis. In particular, due to higher cellular turnover, circulating nucleosomes are present in higher levels in sera from cancer patients compared with healthy samples generally due to increased cell death induced by chemotherapy (59). For instance, higher nucleosome levels have been detected in sera from cancer patients with colorectal, lung, breast, ovarian, renal, prostate cancer, and lymphoma when compared to healthy or benign disease samples (12,59). However, relying only on the bulk amount of circulating nucleosomes does not allow to distinguish between tumor patients and individuals with benign inflammatory conditions (59). A more promising method is based on the quantification of specific histone PTMs and their combinations. In fact, growing evidence indicates that circulating nucleosomes might mirror the histone PTMs found within tumor cells including DNA methylation (60) and combinations of histone marks (58). So far, several histone PTMs, such as H4K16ac, H4K20me3, H3K9me3, and H3K27me3 have been identified and studied in circulating nucleosomes. For instance, in most of cancer patient sera, lower levels of H3K9me3 and H3K27me3 were detected, while an increase was detected in sera of patients who suffered from liver tumor (58,61,62). This shows that histone acetylation and methylation may vary among different types of cancers at early stages of disease and these differences could be detected as well in circulating nucleosomes.

2.4.2 Histone PTMs as biomarkers for cancer prognosis

Histone PTMs have been largely investigated as potential prognostic biomarkers. Their detection in patient tissues, mainly through immunohistochemistry (IHC), showed a correlation of several PTMs with cancer development, progression, histological grading, and response to therapy. In general, altered global levels of acetylation and methylation of histones H3 and H4 correlate with tumor progression and clinical outcome (12,63). Overall, hypo-acetylation has been identified as a marker of cancer progression (64) and has also

been found to be a potential prognostic factor in many types of solid tumors (65). For example, decreased levels of H4K12ac are linked to the progression of lung cancer (66), and increasing tumor grade in colon (67) and breast cancer (68). Histone methylation has also been linked with tumor grade, progression, and poorer prognosis. By comparison with healthy tissues moderate to low levels of H3K4me2 were reported in prostate, kidney, breast, pancreas, liver, and lung cancers with a correlation with worse patient outcome and poor prognosis (11). Low levels of H3K9me3 were observed in lung and prostate cancer (69,70). However, a global increase in H3K9 methylation correlates with poor prognosis and tumor recurrence in other tumor types (71), underlying the fact that PTM changes might have different effects depending on the tumor type. Low levels of H3K18ac are linked with poor survival and cancer recurrence and play an important role in driving the progression of many types of solid tumors including prostate, kidney, lung, breast, colon, and pancreatic cancer (64) but loss of this mark is associated with a better prognosis in glioblastoma (72). Similarly, reduced levels of H4R3me2 have been correlated with poor survival and tumor size in breast cancer (68) and tumor grade in prostate cancer (63), while higher levels seem to correlate with poor survival and recurrence in hepatocellular carcinoma (73). Furthermore, a decrease in H3K27me3 was associated with breast, ovarian, pancreatic, colon, kidney, and lung cancer, while an increase in the same mark correlates with tumor progression in prostate, liver, and esophageal cancer (57,71). In addition to the extensively studied histone PTMs, less-explored histone modifications are under investigation for their potential role as tumor suppressors and oncogenes. Notably, phosphorylation, ADP-ribosylation, and ubiquitination have been studied for their role in regulation of tumorigenesis. These modifications seem to play crucial roles in DNA damage detection, carcinogenesis, the regulation of transcription, and chromatin remodeling (74).

2.4.3 The role of histone modifying enzymes in cancer and their correlation with prognosis

HMEs have been extensively investigated as potential therapeutic targets, as they have been identified to be among the proteins with the highest mutation rates in cancer (75). Indeed, based on TCGA data, 20% to 90% of tumors showed to harbor at least one mutation in an HME, with the most affected tumors being bladder cancer, esophageal cancer, and melanoma, where around 80% of patients carry at least one mutation (including missense mutations, deletions, and amplifications) (55).

Deregulation of HAT activity is a consequence of the overexpression, mutations, and amplification of HAT genes, which was identified in many cancers, especially those of epithelial (breast, gastric, ovarian, lung, and colorectal) and hematological (leukemia and lymphomas) origin (76). These enzymes can play a role both as tumor suppressor genes or oncogenes, depending on the specific type of cancer and the type of mutation (76). This is the case, for instance, of altered p300 and CBP, which are generally considered as tumor suppressor genes (77) but were reported to correlate with cancer initiation and progression in small cell lung cancer with poor prognosis (78). Moreover, a decrease of expression of MOF, responsible for the deposition of H4K16ac, correlates with poor prognosis and a higher risk of metastasis in several types of cancer, such as breast, ovarian, colorectal, renal, gastric cancer, and hepatocellular carcinoma (76). However, changes in acetylation levels might not only derive from altered activity of HATs but increased recruitment of deacetylases could also result in loss of acetylation. In the context of cancer, HDACs are commonly found to be overexpressed and have been extensively studied in various tumor types, such as gastric, breast, pancreatic, hepatocellular, colorectal, lung, and prostate cancer where it associates with a poor prognosis (36). This is also the reason why several inhibitors have been developed to target directly these enzymes.

HMTs and HDMs have been also extensively studied for their potential roles in cancer. Generally, demethylases appear to be upregulated, while the levels of methyltransferases may vary more widely, depending on the type and grade of the tumor, in comparison to normal tissue (55). In correlation with the increase in H3K9 methylation observed in many solid tumors, higher expression levels of SUV39H1 and SUV39H2 were observed in several cancers and reported to correlate with poor prognosis and shorter survival time. In addition, mice carrying deletions of Suv39h1 and Suv39h2 showed genomic instability, associated with a substantial loss of H3K9 methylation, and were described to be more prone to develop cancer (79). MLL proteins have been largely studied in cancer, particularly in leukemia, where rearrangements and fusions of MLL genes are often present (80). However, increasing evidence is showing their importance also in solid tumors such as bladder and breast cancer (81,82). EZH2 is generally overexpressed in a wide range of cancers such as breast, prostate, lung cancer, and glioma and correlates with metastasis, tumor progression, and poor survival, resulting in silencing of tumor suppressor genes (83). In prostate and esophageal cancers, enhanced EZH2 levels are in line with an increase in H3K27 methylation (70,84). However, in other cases, no significant association has been found between EZH2 levels and loss of H3K27me3 (57). Low levels of the H4K20 methyltransferases, SUV4-20H1 and SUV4-20H2, have been correlated with cell invasiveness and tumorigenesis, however, also

in this case, a direct correlation with levels of H4K20me3 was not found. Additional examples of methyltransferases with role in cancer include the SMYD family (85). For instance, SMYD2 and SMYD3, are considered oncogenes in many tumoral contexts and increased expression correlates with poor prognosis (85).

Among the demethylases, LSD1 activity was described to contribute to several phases of cancer, from initiation to progression and metastasis formation as well as recurrence after therapy (86). Since overexpression of LSD1 has been reported in several cancer types (86) and correlated with poor overall survival, several inhibitors have been developed. KDM4 and KDM5 families are probably the most studied demethylases in the context of cancer. There is strong evidence on the role of KDM4 subfamily as proto-oncogene, and the overexpression of KDM4 family members has been described in prostate cancer where it was shown to stimulate androgen receptor-dependent growth. Upregulated levels of KDM4A-B have been described in ERα-positive breast cancer cells, while triple-negative breast cancer cells were reported to have high levels of KDM4C (87). In contrast to other KDM subfamilies that act on multiple lysine residues, the KDM5 subfamily specifically exhibits catalytic activity on H3K4me3/me2 mark and there is emerging evidence of pathological consequences of KDM5 deregulation in various types of cancer. For instance, in a study in collaboration with Di Nisio et al., we studied the role of two different isoforms of KDM5B: a canonical isoform (PLU-1) and a truncated form at the N-terminal (KDM5B-NTT), which is catalytically inactive, displays a dominant negative activity, and is associated with increased levels of H3K4me3 (88) (also *see paragraph 5.10*).

2.4.4 Epigenetic therapies for cancer treatment

Unlike genetic mutations, epigenetic changes are reversible, and can be potentially reverted through the so-called “epigenetic therapy”. Epigenetic drugs or “Epi-drugs” are chemical compounds that have been generated to alter DNA and chromatin structure by targeting the writers, erasers, and readers of altered epigenetic features (89). Until now, several epi-drugs have been developed and are undergoing different clinical trial phases while others have been approved for clinical use by the Food and Drug Administration (FDA). The first generation of epi-drugs included DNA methyltransferase inhibitors (DNMTi) and HDAC inhibitors (HDACi). However, because these drugs are considered “broad reprogrammers”, which lead to generalized epigenetic changes, they result in low substrate specificity and are associated with toxic effects. Therefore, a second generation of epi-drugs has been

developed upon a more in depth evaluation of global cytotoxic effects (90). Recently, a third generation of epi-drugs has been developed and it includes both inhibitors of histone methyltransferases and demethylases and inhibitors of reader domains (91).

Several epi-drugs have already demonstrated clinical benefits. Among the DNMTi, Azecitidine and Decitabine are the most used and they have been approved by the FDA to treat some types of cancers like leukemia and myelodysplastic syndromes (92-94). Azacitidine works by decreasing abnormal DNA methylation and exhibits antitumor activity by preventing cancer development, while Decitabine, a deoxycytidine analog, is an irreversible DNMT1 inhibitor that promotes apoptosis by influencing the cell cycle. However, both these drugs are highly toxic and harmful to the human body at high concentrations and patients have to be exposed to low doses for long time (95).

To date, HDACi have been extensively investigated and their anti-tumor effects are dependent on their capability to maintain proper tumor suppressor gene expression. For cancer treatment, Vorinostat and Panobinostat have been FDA approved. Vorinostat directly inhibits the activities of HDAC1, HDAC2, HDAC3, and HDAC6 by reversibly binding to Zn^{2+} at the active site of the enzyme resulting in antitumor activity. Panobinostat is another pan-HDACi that has been approved mainly for the treatment of multiple myeloma where it causes apoptosis of cancer cells (95). Currently, the development of many HDACi is still in progress, with a lot of limitations, such as off-target effects, low selectivity, and low efficacy. Indeed, HDACi are highly associated with side effects such as hematological toxicity, cardiac arrhythmia, and weight loss. Currently, additional HDACi are undergoing clinical trials, and recent findings have further supported the concept of targeting histone acetylation readers, including the BET proteins (96).

Regarding methylation, the discovery and development of small molecule inhibitors of HMTs and KMTs are ongoing. Because HMTs are more specific than HDACs, these molecules could potentially be associated with fewer side effects. Several compounds including inhibitors of DOT1L, EZH2, and LSD1 have been designed, have undergone preclinical testing, and some of them have been quickly introduced in clinical trials. The discovery of drugs targeting HMTs heavily depends on the cofactor binding pocket, which possesses structural features that facilitate interaction with small molecule inhibitors (96). Examples of HMT inhibitors include Pinometostat and Tazemetostat, which specifically act as selective inhibitors of DOT1L and EZH2, respectively. The finding that DOT1L plays a key role in the development of MLL-rearranged leukemia led to intense research in the field of drug discovery, in order to design and synthesize inhibitors to target DOT1L activity,

dependent on the interaction with MLL fusion proteins (96). Because of it is overexpression in several tumors, also EZH2 has been the target of intense drug development efforts and Tazemetostat has been FDA-approved for the treatment of sarcoma (97) and refractory follicular lymphoma (98).

Inhibitors of histone JmJC demethylases are also under development: these inhibitors are based on the block of the enzyme catalytic domain and inhibit the enzyme activity by chelating the Fe (II) active site. However, because of the similarity between KDM active site pockets, it has been difficult to achieve selectiveness. Instead, the pharmacological inhibition of LSD1 with small molecules has shown particular effect on proliferation, invasion, and migration of cancer cells and LSD1 has become an evolving clinical target for anticancer therapy. Many LSD1 inhibitors have been identified and have been mostly proposed for the treatment of Leukemia (99). For instance, the inhibitor DDP-38003 was developed by the Experimental Therapeutic Unit of the IFOM-IEO Campus and was shown to induce myeloid differentiation in acute myeloid leukemia (AML) by suppressing the GSE1 protein, a poorly explored LSD1 interactor. These findings contributed to additional information on the mechanisms of action of this drug in the context of leukemia beyond their established effects in inhibiting lysine histone demethylase activity (100).

The effectiveness of epigenetic therapy has mainly been observed in cases of hematological cancers (91), primarily using DNMTi and HDACi, but studies are also ongoing in the context of solid tumors. Moreover, in recent years, novel technologies have been developed for gene editing or epigenetic changes, an example is the CRISPR/dCas9 (clustered regularly interspaced short palindromic repeats) approach, which has been shown to be the most promising technology for epigenetic editing (101).

2.5 Epi-proteomics approaches to study histone PTMs

The increasing evidence of the role of epigenetics in cancer biology has encouraged the investigation of epigenetic marks in both cancer cell lines and clinical samples (102). The analysis of histone PTMs in the context of patient-derived tissues has been traditionally performed with antibody-based techniques such as ELISA, immunoblots, immunofluorescence, and IHC. However, all these techniques share several limitations that include the possibility to measure just one or few modifications at a time, the need for specific and reliable antibodies and the problem of poor signal, cross-reactivity, namely the recognition of other modifications in addition to the one of interest, and epitope masking,

the masking of a modification when another modification is present in a nearby residue. As an alternative to traditional antibody, mass spectrometry has become the most suitable method for the analysis of histone PTMs, as it allows an unbiased, comprehensive, and quantitative investigation of histone marks and variants. Mass spectrometry has been widely used to analyze histone PTMs in cell lines and more recently several methods have been developed for the analysis of patient-derived samples, which include tissue and liquid biopsies. In recent years, the optimization of these protocols by our group allowed the analysis in Formalin Fixed Paraffin Embedded tissues (FFPE), fresh frozen, and optimal cutting temperature (OCT) frozen samples. Thanks to these developments, this "epi-proteomic" strategy has allowed the investigation of epigenetic marks in pathological conditions, including cancer.

2.5.1 Mass spectrometry: basic concepts and data acquisition in shotgun proteomics

Mass spectrometry is an analytical tool that allows the determination of the mass to charge ratio (m/z) of different types of molecules, including particular peptides and proteins. MS-based proteomics has become the method of choice not only for the analysis of complex protein samples and therefore for total proteome analysis, but also for the analysis of protein PTMs. PTMs are identified and localized at specific residues by a delta mass (Δm), which is calculated as the difference between the theoretical and the experimentally measured masses of peptides/proteins. Therefore, theoretically, any PTM alone or in combination can be identified and quantified in a single run without needing *a priori* information on the modification site (103).

The mass spectrometer is generally coupled to a liquid chromatography (LC) setup, which allows separation of complex peptide mixtures through a reversed phase high performance liquid chromatography (HPLC). Because the mass spectrometer concept is based on the use of electromagnetic fields, the analytes must be charged and transformed into a gas phase (104). Therefore, a sample introduction system, better known as ion source, in which peptides get ionized in the gas phase, is essential. However, because peptides and proteins are polar and nonvolatile compounds and their transfer into a gas phase might cause their destruction, the so called "soft ionization techniques" were developed. They include matrix-assisted laser desorption ionization (MALDI) (105) or electrospray ionization (ESI) (106), which is usually connected online with the HPLC. Moreover, in order to increase peptide concentration and sensitivity, nano-ESI sources can be used with the option of nl/min flow rates. In ESI, peptides are dissolved in an aqueous buffer containing 0.1% formic acid and

are eluted through a chromatographic column with increasing concentration of an organic solvent, acetonitrile, which separates molecules based on hydrophobicity. Peptides are then ionized and electrostatically dispersed thanks to the high voltage applied (2-6 kV) as well as the high temperature provided by a heated capillary. The result is an aerosol, containing positively charged droplets, with a high proton amount, which need to be de-solvated prior to enter the instrument (**Figure 5**). Generally, an uncharged gas, like nitrogen, is used to favor the liquid nebulization prior to MS analysis. The central part of the MS instrument is the mass analyzer, which can store ions and separate them based on m/z. Several mass analyzers have been developed. The most common are the Orbitrap, the quadrupole (Q), and the time of flight (TOF). Although these mass analyzers differ in the working method, they all select a single m/z species from a mixture of peptide ions generated by the source and fragment it to obtain an MS/MS spectrum (MS2), which is generally acquired with a higher energy collisional dissociation (HCD) fragmentation technique. The most used acquisition method is the data dependent acquisition "DDA" mode, where the most intense precursors in a certain time frame are selected and fragmented, with generation of the MS2 spectra that provide information on the amino acid sequence.

The most used approach for analysis of a mixture of proteins is the so called "shotgun proteomics". Shotgun proteomic workflows are based on the analysis of digested proteins (typically by trypsin) instead of intact proteins (**Figure 5**), as examining peptides rather than proteins offers several advantages: MS has a higher sensitivity for small molecules; peptides are eluted more easily than proteins through chromatography, and are also more easily fragmented at MS/MS level. (107). To identify a peptide, its sequence information obtained through fragmentation is crucial. Several MS/MS databases have been created for protein identification starting from peptides, all following a similar approach: each MS/MS spectrum is compared to a database containing theoretical fragmentation patterns for peptides from the searched database. User-defined criteria, such as mass tolerance, the choice of proteolytic enzyme, and post-translational modifications, are applied to narrow down the pool of candidate peptides. The result is a list of fragment ion spectra matched to the peptide, ordered by a search score that measures the similarity between the experimental and the theoretical spectrum. The peptide sequence with the highest score is selected for further analysis. In the case of histone PTMs, the database used for the search can either be the same as that used for proteomics investigations or a smaller database containing only histones. PTMs data analysis follows a similar approach with the additional step involving searching for specific modifications of interest. However, it is not feasible to search for all possible modifications simultaneously, because this would lead to increased search times

and higher rates of false positives, as the number of variable modifications in the database search increases (108). To address this challenge, the RAW data are analyzed through multiple parallel searches, using different combinations of variable modifications (109). Despite the availability of various software to assist in histone PTM analysis and quantification, this task still heavily relies on visual inspection and validation of the RAW data and, consequently, is predominantly carried out manually.

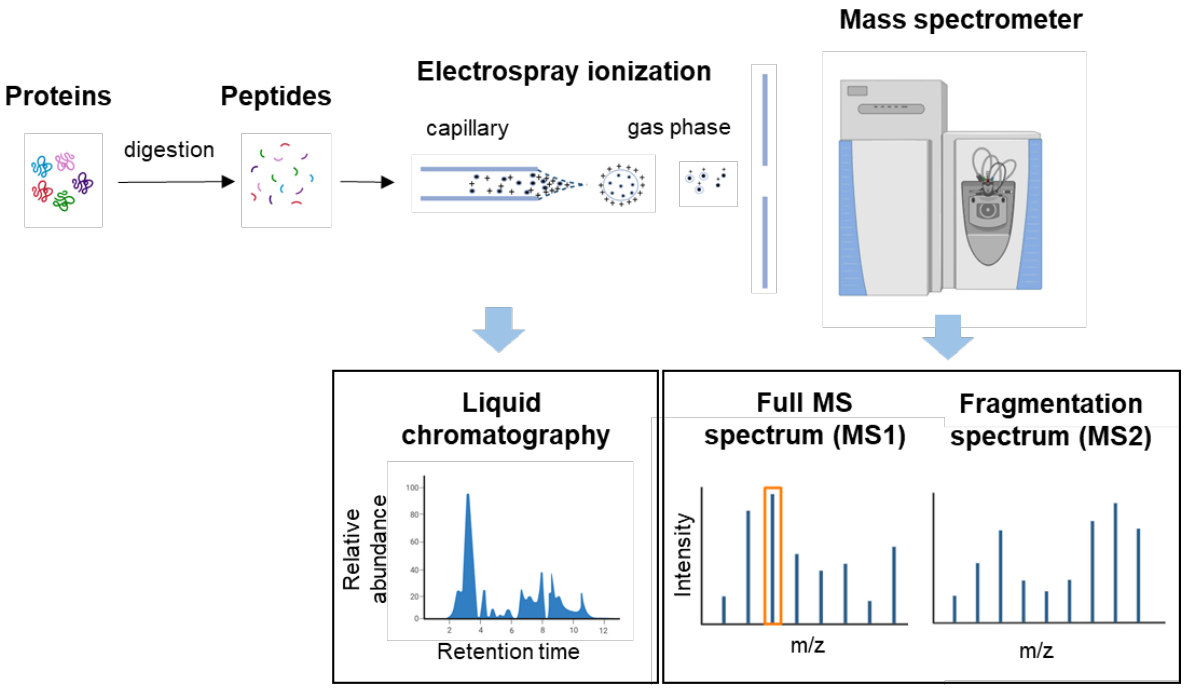


Figure 5: LC-MS workflow for shotgun proteomics.
Proteins are digested into peptides and separated by high performance liquid chromatography. Peptides are then ionized and analyzed by mass spectrometry, generating a full MS spectrum (MS1). The most intense precursor ions are then fragmented in the MS2 spectrum. Created with BioRender.com

2.5.2 Mass Spectrometry to analyze histone PTMs and quantification strategies

Three different MS methods can be used for the analysis of histone PTMs: "top-down", "middle-down" and "bottom-up" (**Figure 6**) (110).

In "top down" approaches, whole histones are separated through chromatography, ionized, and directly MS-analyzed, giving information on the complete panel of histone isoforms in a sample, as well as their stoichiometry. Even though this method allows the detection of a comprehensive combination of PTMs, as the number of modifications detected increases

exponentially, analysis of data remains challenging since discrimination of the modification would require very sophisticated software. In “middle-down” methods, a digestion step is performed in order to have long histone peptides of a >5 kDa size, which comprise most of the N-terminal tails of core histones. This type of digestion is generally performed with proteases that cut at residues that occur with low frequency, such as aspartate or glutamic acid. AspN digestion results in the histone H4 peptide 1-24, while GluC generates the histone H3 peptide 1-50. These types of digestion allow the detection of most of the characterized H3 and H4 PTMs and allow the study of combinatorial marks. However, both top down and middle down methods involve lower sensitivity and demanding data analysis. In addition, the analysis of isobaric species, namely the peptides carrying the same modifications but on different residues, is challenging, as in most of the cases they are not possible to distinguish. Therefore, these approaches are used only by few specialized laboratories and no applications have been reported in the context of clinical samples, until now.

The most employed approach is the so called “bottom-up” MS analysis, which is based on the analysis of relatively small peptides of a length of 5 to 20 amino acids, which result from protein digestion and which is applicable to clinical samples. In this case, trypsin is the main protease used for digestion. However, because trypsin cuts proteins at lysine and arginine residues, in the case of histones, whose sequence is rich in these amino acids, too short peptides of inconsistent length are produced. These peptides are not detectable by the mass spectrometer and therefore difficult to accurately quantify. To resolve this issue, two main approaches can be used: the digestion with the Arg-C protease, which cleaves at the C-terminus of arginine residues only, but can be achieved only in in-solution protocols, or an “ArgC-like” digestion, which involves the use of trypsin after a step of chemical derivatization of lysine residues, which blocks the cut at lysine residues, making possible to have longer peptides of appropriate length for MS analysis with the shorter peptide of 5 amino acids. Lysine residues are chemically acylated through deuterated acetic anhydride or more commonly propionic anhydride (PRO). This approach is also useful to discriminate between isobaric peptides. In addition, a second step of derivatization after digestion, with addition of PRO or phenyl-isocianate (PIC) is usually performed, to increase the retention time and detectability of short hydrophilic peptides. With this method, it is possible to detect up to four co-occurring modifications. Even though this method is characterized by the limitation of losing information about distant marks, several advantages are present including established sample processing for histone extraction and the availability of established protocols also for the analysis of clinical samples.

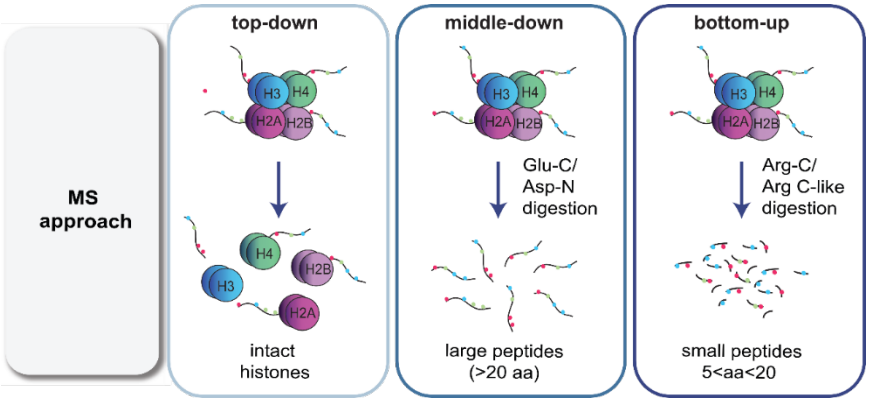


Figure 6: MS methods for analysis of histone PTMs

Different MS-methods used for histone PTMs analysis. In the 'top-down' strategy, intact histones are chromatographically separated and subjected to direct MS analysis. In the middle-down' method, peptides >20 amino acids in length are generated by digestion with the Glu-C or Asp-N enzymes. In the 'bottom-up' approach, histones are digested into shorter peptides, typically ranging from 5 to 20 amino acids in length, using enzymes like the Arg-C protease. Adapted from (110).

Because histone PTMs are present in most of the samples, it is essential to be able to accurately quantify their abundance. The methods for quantification of histone PTMs involve the extraction of the peaks, namely eXtracted Ion Chromatograms (XICs), matching the m/z value and chromatography retention time of the histone peptides from the chromatographic profile. XICs can be obtained manually or through a dedicated software called EpiProfile (111). Two main strategies can be used to quantify histone PTMs: label free quantification, where XICs can be directly compared across samples that were acquired separately, and isotope-based methods which are characterized by the use of internal standards, resulting in improved quantification accuracy. Labeled histones obtained by cultivating one or more cell lines in media containing heavy-isotope labeled amino acids with the Stable Isotope Labelling by Amino acids in Cell culture (SILAC) strategy (112,113) can be used as internal standard. Alternatively, synthetic isotope labeled peptides like Protein-AQUA™ peptides carrying the most common histones H3, H4, and H2A PTMs can be used, with the advantage of allowing an absolute quantitation (114). The use of an internal standard usually reduces experimental variability compared with label-free strategies and offers higher flexibility and it is particularly useful for the analysis of large clinical samples datasets (**Figure 7**).

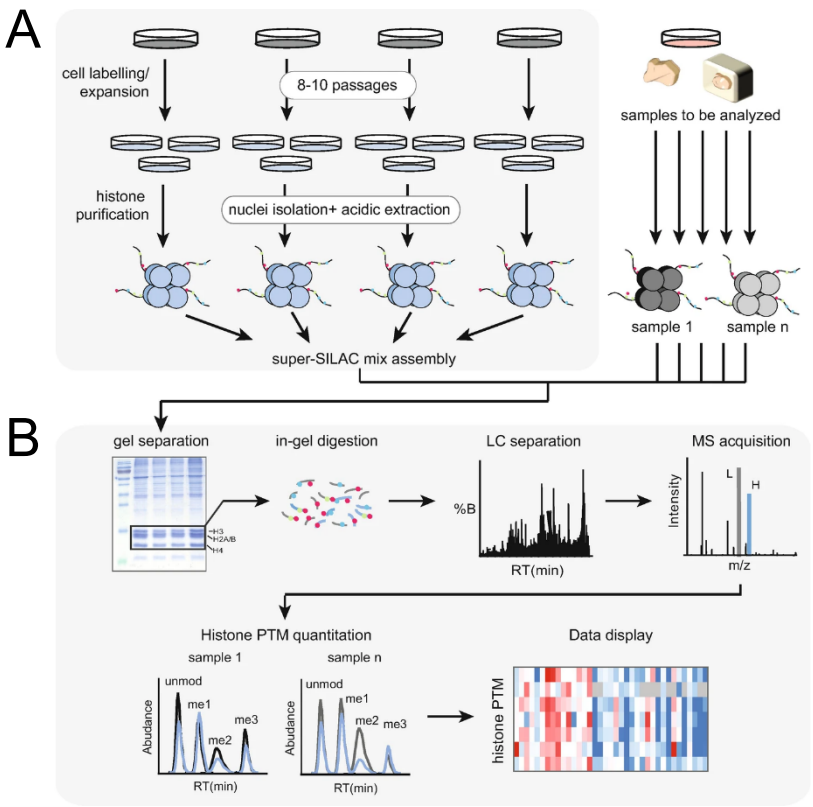


Figure 7: Representation of super-SILAC strategy applied to the analysis of histone PTMs

(A) Generation of super-SILAC mix by growing different cell lines in medium containing heavy amino acids. Super-SILAC is then mixed 1 to 1 with the unlabeled sample to be analyzed. (B)Application of super-SILAC strategy for MS quantitation of histone PTMs. After in-gel digestion peptides are separated by liquid chromatography and acquired in the mass spectrometer. The light and heavy form of the modified peptides, which elute at the same retention time, are quantitated and data are displayed as an heatmap based on L/H ratios (L: Light, H: Heavy). Adapted from (113).

2.5.3 Sample preparation for analysis of histone PTMs from clinical samples

Several methods have been developed to extract histones from cell lines, however, the high amount of starting material needed and the number of steps involved make it more difficult to apply it to patient-derived tissues. For these samples, histone extraction protocols have been optimized and adapted to different type of storage conditions and to low number of cells.

Patient-derived tissues are collected in hospital biobanks and can be stored in three different ways, including FFPE, fresh frozen and OCT-frozen tissues. Fresh frozen sample storage is the optimal condition for not only histone analysis, as this avoids potential contaminants for MS-analysis but also for other applications like nucleic acid extraction, however, these samples are not very easy to section and the cut might also affect the morphology. The OCT

tissues are a better option for cutting, as they are embedded in a cryopreservation medium composed of polyvinyl alcohol, and polyethylene glycol, which allows morphological preservation. The OCT compound is, however, a strong contaminant for all downstream applications. Fresh frozen samples are present in limited amount as most of clinical samples are stored as FFPE whereby samples are fixed in formaldehyde and embedded in paraffin. A specific protocol called PAT-H-MS (115) was developed by our group in 2017, specifically to allow the extraction of histone and subsequent MS analysis of PTMs from FFPE tissues.

Finally, histones can be enriched from blood, which represents a particularly attractive source of non-invasive molecular biomarkers. A protocol to isolate circulating nucleosomes from the serum of cancer patients has been proposed as a non-invasive method to analyze histone PTMs by MS (58). Despite the presence of histones in these preparations has been demonstrated by MS, a systematic processing technique has not been defined and validated yet. A summary of the protocols for sample preparation prior to MS analysis is shown in **Figure 8**.

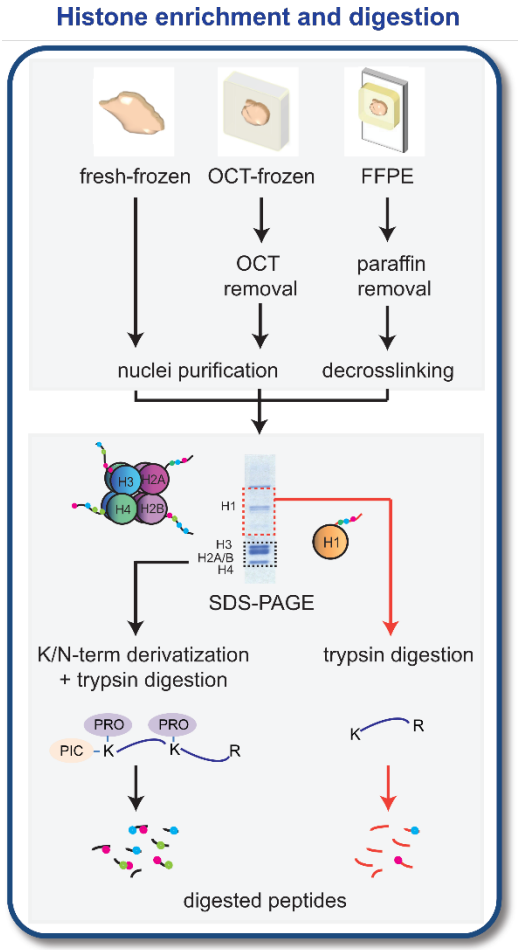


Figure 8: Sample preparation for the analysis of patient-derived tissues by MS

Workflow for histone enrichment from fresh frozen, OCT-frozen, and FFPE samples, which are separated by SDS-PAGE and in-gel digested. Adapted from (116).

2.6 Strategies for the investigation of the molecular mechanisms linked to aberrant epigenetic features

Studying histone PTM aberrations in the context of cancer can not only uncover epigenetic biomarkers for patient stratification but also suggest potential aberrant mechanisms targetable for therapy. The altered molecular mechanisms associated with epigenetic signature identified in cell lines and clinical samples can be investigated at two different levels (**Figure 9**):

1. Downstream: investigating downstream molecular mechanisms allows to unravel the effects caused by the aberrant deposition of histone PTMs on chromatin. In this case, a comprehensive analysis using a multi-OMICs approach, based on the integration of ChIP-seq, RNA-seq, and proteomic data can be used. The integration of data could allow gaining an overview on the distribution of these modifications at genome-wide and locus-specific level as well as identifying the effect of their deposition on the expression of specific genes associated with the tumor phenotype;

2. Upstream: studying the upstream mechanisms implies the analysis of the causes of altered histone PTMs deposition. Such causes may include altered expression or activity of histone modifying enzymes, histone readers or other chromatin-associated factors, which bind to chromatin and other histone PTMs, such as histone remodelers or other factors like intermediate metabolites.

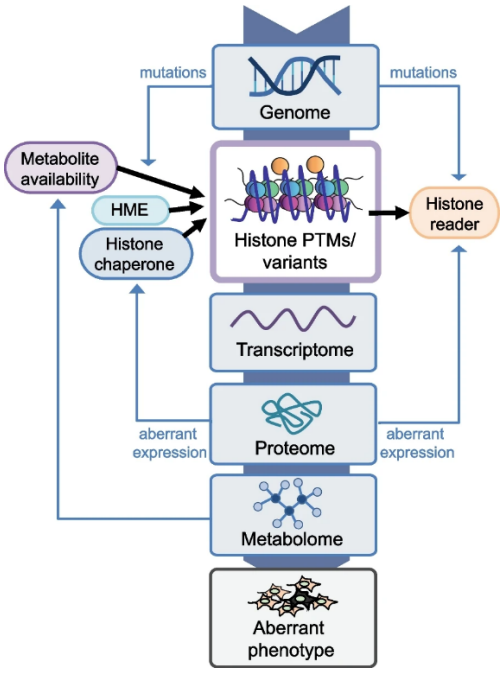


Figure 9 (previous page): Molecular mechanisms underlying histone PTMs and histone variants deposition

Histone PTMs and histone variants regulate gene expression leading to alterations at RNA, protein, and metabolite levels leading to abnormal phenotypes. At the same time, the protein levels of histone-modifying enzymes, histone chaperones and readers, as well as intermediate metabolites can contribute to the modulation of the levels of histone PTMs and variants. Taken from (116).

2.6.1 Investigation of downstream mechanisms through multi-OMICs approaches

Even though proteomics is comprehensive and quantitative, it only gives information on global modification levels, without providing information about the distribution of these modifications on the genome and the effect that the deposition of these modifications produces. To obtain locus-specific information, MS data can be combined with ChIP-seq analysis performed with antibodies specific for the histone PTMs of interest. Then, these data can be integrated with transcriptomic and proteomic data, to generate a global analysis with information about regulation of gene expression.

ChIP-seq is based on chromatin immunoprecipitation followed by high-throughput DNA sequencing and has emerged as a widely employed technique to map the genomic locations of chromatin-binding proteins and histone modifications in cell lines and tissues (117). The application of ChIP-seq to histone PTMs allows the mapping of individual histone modifications across the entire genome, offering valuable insights into cell identity and disease states. In a typical ChIP-seq workflow proteins like histones or transcription factors (TFs) are cross-linked to genomic DNA, allowing to capture a snapshot of interactions between these proteins and DNA. Next, the chromatin is sheared by sonication into smaller fragments typically ranging from 200 to 600 base pairs (bp) in length for TFs or around 250/300 bp for histone PTM analysis. Antibodies that specifically target the histone modification or TF of interest are then used to immune-precipitate the protein-DNA complexes. After the crosslinks are reversed, the DNA is purified and subjected to high-throughput sequencing.

In addition, because histone PTMs localize at promoters, enhancers, or gene bodies, the effects on gene expression can be observed by integrating data with transcriptomic information, which involves evaluating the complete set of RNA transcripts and their abundance in single cells or cell populations. (118). Transcriptomics is based on the use of either microarrays or high-throughput sequencing approaches (RNA-seq) to quantify transcripts. RNA-seq is more comprehensive than microarrays, as it allows the detection of new transcripts, enables more precise quantification and identification of transcripts with

low expression levels, and facilitates simultaneously the discovery of novel genes, exons, and transcript isoforms (119). In addition, the continuous advancements in transcriptomic research have revolutionized modern medicine, enabling its application in disease diagnosis and prognosis. Very large transcriptomic datasets are now available in databases such as the Cancer Genome Atlas (TCGA) (120) and these public repositories are a valuable source of information, especially in the context of clinical samples. Proteomic data (see *paragraph 2.5.1)*, can be also integrated in multi-OMICs workflows. Global proteomics focuses on studying the entire set of proteins expressed in a sample. In addition, also in the case of proteomic data public repositories such as the Clinical Proteomic Tumor Analysis Consortium CPTAC (121) are present. Therefore, integrating proteomic data with other OMICs data could potentially lead to the identification of specific protein biomarkers, signaling pathways, and cellular processes that are affected by altered histone PTMs (122). A remarkable example of a multi-OMICs approach applied to the investigation of epigenetic features was reported in the context of Alzheimer disease where integration of transcriptomic, proteomic, and epigenomic data identified altered molecular pathways involved in this type of neurological disease (123). In the context of cancer integrative analysis it was mainly performed on ChIP-seq and RNA-seq data for instance in ovarian and breast cancer (124,125).

2.6.2 Upstream mechanisms

Based on the complexity of the mechanisms involved in the regulation of epigenetic changes, the investigation of the causes of deposition of aberrant histone PTMs is not an easy task. In the simplest scenario, alterations in histone PTMs are a result of the abnormal activity of HMEs. However, a direct correlation between histone PTMs and respective HMEs is not always present. For instance, even though for K16ac a correlation with specific HAT has been found, this is not the case for H4K20me3 levels. Additional causes might include the presence of altered multi subunit complexes to which HMEs belong, as well as the interdependence between histone PTMs and DNA methylation levels. Other mechanisms include changes in expression or mutations in histone chaperones or non-enzymatic reactions that can be mediated by chemically reactive metabolites, which can modify histones (126,127). Various methods can be employed to target and modulate modifications at genome wide and genomic specific regions. Two types of modulation can be used, genetic manipulation including established techniques like RNA interference and novel technologies like CRISPR-dCas9, and pharmacological inhibition.

Since the initial application of RNA interference (RNAi) in mammalian cells, the use of short hairpin RNAs (shRNAs) for targeted gene silencing has become a standard technique. This involves introducing shRNA precursors into cells through plasmid and viral vector systems, allowing efficient processing by the RNAi pathway, resulting in potent gene knockdown. Briefly, shRNAs spontaneously fold into hairpin structures that are recognized by the cellular RNAi machinery and are processed to form active small interfering RNAs (siRNAs). The advantages of using shRNA-based technologies include the ability of long-term gene silencing and the infection of challenging cell lines using viral vectors. Moreover, the availability of inducible promoters allows a temporal control of the expression of the shRNA, thus minimizing cellular toxicity (128).

A more recent technique involves the use of CRISPR interference (CRISPRi) based on the use of CRISPR-dCas9 as an innovative epigenetic editing tool (129). CRISPRi is characterized by a catalytically deactivated or dead Cas9 (dCas9) which is not able to cleave the DNA strand but is used to direct a synthetic small guide RNA (sgRNA) to the target sequence to block transcription initiation and elongation. To increase the efficiency of CRISPRi in eukaryotic cells, the dCas9 could also be fused with several domains of eukaryotic transcriptional repressors or activators (CRISPRa). Generally, the dCas9 can be fused with various epigenetic effectors, including catalytic domains of histone-modifying enzymes such as histone methyltransferases, DNA methyltransferases, demethylases, and acetyltransferases (130). The Krüppel Associated Box (KRAB) repressor is among the most common repressors fused with dCas9. KRAB domain achieves repression by collaborating with the KAP1 co-repressor complex and is linked to the deposition of H3K9me3 resulting in gene silencing (129) . In addition, other examples of epigenetic effectors include p300 and LSD1.

Additionally, the use of specific inhibitors offers an alternative strategy. Histone modifying enzymes, which might affect the levels of histone PTMs, can be targeted by the use of several inhibitors. These drugs are commercially available and target both writers, erasers and readers of histone marks, as described in the *paragraph 2.4.4.*

2.7 Breast cancer: general features and definition of molecular subtypes

Breast cancer (BC) is the most diagnosed cancer in women, with 2,261,419 cases reported by the World Health Organization in 2020, and it accounts for most of female cancer related deaths, with 684,996 cases reported in the same year (131). Breast cancer is a complex and

heterogeneous disease, which has been largely studied and classified in different subtypes based on histology, morphology, and molecular features. For prognostic prediction and treatment decision making, breast tumors are classified based on IHC, by testing estrogen receptor (ER), progesterone receptor (PR), Epidermal Growth Factor receptor 2 (HER2), and the status of the proliferation marker Ki67 (132).

ER is an important diagnostic marker with significant high expression in around 70-75% of invasive breast carcinomas (133). The diagnosis of ER altered levels is particularly relevant in terms of therapy selection and it is also a predictive factor of better clinical outcome. The measurement of ER expression is mandatory in the clinical practice in order to identify which patients will benefit from endocrine therapy mainly by the use of estrogen modulators, estrogen receptor down-regulators, or third-generation aromatase inhibitors (134). PR-positivity is present in around 50% of ER-positive cases and less present in ER-negative tumors. PR expression is directly regulated by ER and physiological values of PR are informative of the functional ER pathway (135). In general, high expression of PR is positively associated with increased overall survival and better prognosis, while lower levels are generally associated with a more aggressive disease, as well as poorer prognosis (136). Thus, favorable treatment of breast cancer patients highly depends on the assessment of PR expression (137). HER2 expression accounts for around 15-25% of breast cancer cases and has been described as one of the earliest events during breast carcinogenesis (138). HER2 amplification is often present and causes an increased activation of proto-oncogenic signaling pathways, resulting in uncontrolled cell growth. Finally, the Ki67 antigen is a cellular marker of proliferation and is strictly correlated with tumor aggressiveness, response to therapy, and relapse (139). It is also considered as a prognostic marker, as its overexpression was reported to be associated with poorer clinical outcomes and poorer survival rates of breast cancer patients (140).

Tumor molecular characterization enables to categorize patients in order to provide the most suitable therapy, which can include hormone therapy or specific targeted therapies, such as anti-HER2 drugs. In addition to pathological assessment, gene expression profiles such as the PAM50 classification through microarrays, qPCR, and RNA-seq have been included to assist the choice of the best treatment.

2.7.1 Breast Cancer Subtypes

Different classifications have been applied to breast cancer. Four main subtypes have been defined based on histology and IHC: Luminal A, Luminal B, HER2 type, and Triple Negative (TN) (**Figure 10**). Perou et al. and others (141,142) described five molecular intrinsic subtypes, based on the analysis of microarray gene expression data: Luminal A, Luminal B, HER2-enriched, Basal-like, and Normal Breast-like. Among these subtypes, the normal breast-like subtype was omitted, as it is thought to represent sample contamination by normal mammary glands and not reproducibly defined (143). Additionally, the intrinsic subtype claudin-low breast cancer, which seems to display a TNBC phenotype, was reported (144). Based on this classification, in 2009, Parker et al. developed a 50-gene signature for subtype identification, better known as PAM50, that can reliably classify these different subtypes (145).

Luminal A tumors are the most diagnosed type of breast cancer, accounting for around 60-70% of breast cancer cases. This type of breast cancer is characterized by high positivity to ER and/or PR, absence of HER2, and low expression of Ki67 (less than 20%). From the clinical point of view, these tumors are in general characterized by a good prognostic outcome with low incidence of relapse and high survival rate, as they are generally slow growing and low grade. The main therapy used is based on hormone therapy. Tamoxifen and aromatase inhibitors are used with good response, while a lower benefit is observed with the use of chemotherapy (133).

Luminal B tumors represent 10-20% of cases and show positivity to ER with a lower degree of PR expression, can be positive or negative to HER2, and show high proliferation rates (more than 20%). These tumors are characterized by a higher grade and poorer prognosis compared to Luminal A with a faster growth. These tumors may benefit from hormone therapy together with classical chemotherapy (135).

The HER2 positive group accounts for 10-15% of breast cancer cases and is characterized by increased expression of HER2 and absence of ER and PR. This type of tumor is more aggressive than Luminal subtypes and is characterized by a fast growth and a poor prognosis. HER2 tumor prognosis has much improved after the introduction of anti-HER2 treatments (e.g. Trastuzumab), in addition to chemotherapy. Proliferation is generally high with a Ki67 > 30% (135).

TNBC is a heterogeneous group of breast cancers, which is characterized by the absence of ER, PR, and HER2, and is therefore called Triple Negative. TNBC accounts for 15–25% of all breast cancers and is characterized by high levels of the proliferation marker Ki67. It is a

very aggressive tumor often characterized by early relapse, generally within three years after chemotherapy, and is generally diagnosed at advanced stages. Most of TNBC tumors overlap with the basal-like intrinsic subtype with around 75% of cases. Because the TN subtype is very heterogeneous, it is further classified in several additional subtypes. In 2011, Lehman et al. described six distinct subtypes based on gene expression profiles (146). This classification includes two basal-like subtypes BL1, (associated with cell cycle progression and proliferation genes), and BL2 (enriched for growth factor signaling and cell motility genes); a mesenchymal and claudin-low subtype, (M, enriched for genes related to cell motility, angiogenesis, and extracellular matrix); a mesenchymal stem–like subtype, (MSL, exhibiting a gene expression profile with high mesenchymal and stem cell-like features); an immunomodulatory subtype, (IM, enriched for immune-related genes and lymphocyte infiltration); and a luminal androgen receptor subtype (LAR, which exhibits androgen receptor signaling and a luminal-like gene expression profile). In addition, this classification identified differential sensitivity to chemotherapy and targeted agents. By Gene Set Enrichment Analysis (GSEA), the top canonical pathways associated with each TNBC subtype were characterized. BL1 subtype resulted to be more sensitive to genotoxic agents such as taxanes (e.g. Paclitaxel and Docetaxel), as expression of genes associated with DNA damage response and proliferation (i.e AURKA, AURKB, ATR/BRCA) was detected. The BL2 subtype is characterized by expression of EGFR, MET, NGF, Wnt/β-catenin, and IGF-1R pathways, and potential targeted therapeutic drugs proposed include mTOR inhibitors and growth factor inhibitors (e.g. Lapatinib, Gefitinib, and Cetuximab). The M subtype exhibits tissue characteristics resembling to sarcoma or squamous epithelial cells, characterized by expression of the TGF-β pathway, and it is usually characterized by resistance to chemotherapeutic drugs. As a result, patients with the M subtype might be prescribed mTOR inhibitors or drugs that target the epithelial-mesenchymal transition, while the use of PI3K inhibitors and antiangiogenic drugs have been proposed for the MSL subtype. In the case of IM subtype, the use of immune checkpoint (e.g. PD1, PDL1, CTLA-4) inhibitors was proposed. Lastly, the LAR subtype of breast cancer displays significant overexpression of the androgen receptor, and anti-AR therapy is a recommended treatment option (147). Clearly, there is a major need to better understand the molecular basis of TNBC and to develop effective treatments for this aggressive type of breast cancer, for which chemotherapy remains the primary treatment option. More extensive molecular analyses of TNBCs are required to understand the complexity of the disease and to identify molecular drivers that can be therapeutically targeted.

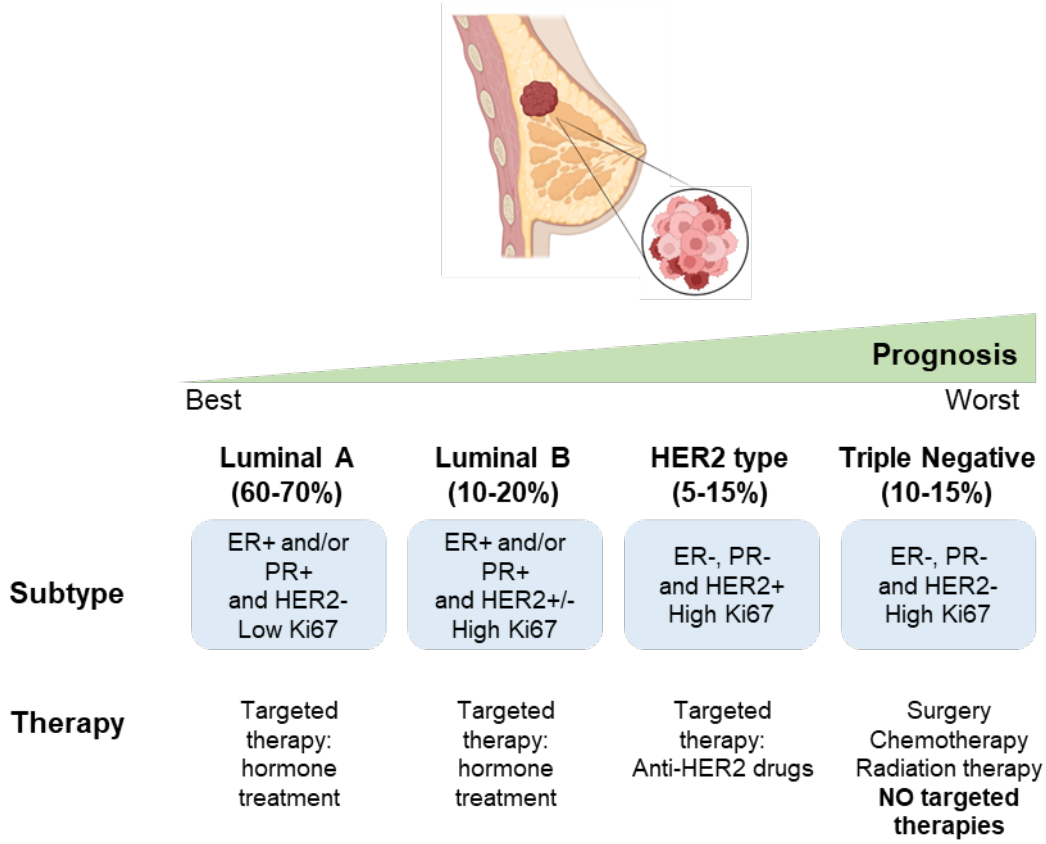


Figure 10: Molecular subtypes of breast cancer

The different breast cancer subtypes are displayed with prevalence percentages and with subtype specific characteristics. ER: estrogen receptor, PR: progesterone receptor, HER2: Epidermal growth factor receptor 2. Ki67: proliferation marker. Therapies are indicated for each specific subtype. Created with BioRender.com

2.7.2 Epigenetics and breast cancer

In recent years, several studies have reported the role of epigenetic alteration in breast cancer tumorigenesis including effect on stemness and drug resistance. DNA methylation, histone PTMs, nucleosome remodeling, and RNA-mediated gene targeting are known to modulate a series of molecular cellular and biological processes involved in breast cancer tumorigenesis (148).

The function of DNA methylation in breast cancer has been extensively studied, and it involves gene silencing via hyper-methylation and gene activation through hypo-methylation. Hyper-methylation of CpG islands is found in poorly differentiated breast tumors, while lower methylation levels are observed in well-differentiated tumors (149). Generally, in breast cancer, abnormal methylation patterns lead to changes in gene expression, contributing to the development of specific clinic-pathological characteristics. Moreover, extensive research has documented that TNBC tumors are characterized by

genome-wide hypo-methylation (150), when compared to other breast cancer subtypes. Additionally, a separate study revealed a significant correlation between the methylation status and the response to chemotherapy, as well as resistance in TNBC (151).

Concerning histone PTMs, a comprehensive study conducted by Elsheik et al. (68), by IHC on nearly 900 samples, documented varying levels of histone PTMs in breast cancer. This study included differential levels of lysine acetylation (H3K9ac, H3K18ac, and H4K12ac), lysine methylation (H3K4me2 and H4K20me3), and arginine methylation (H4R3me2), which were observed in poorer prognostic subtypes, including basal-like and HER2 tumors and could represent potential prognostic markers for breast cancer. In addition, it has been recently demonstrated that depletion of H3K27me3 enhances TNBC resistance of cancer cells to chemotherapy, while maintenance of H3K27me3 methylation inhibits the transition to a drug-tolerant state, and delays tumor recurrence *in vivo* (152).

2.8 Mass spectrometry profiling of histone PTMs reveals an epigenetic signature distinguishing TNBC from other breast cancer subtypes

Exploiting the MS methods developed during the last years, our group analyzed histone PTMs in breast cancer patient-derived samples. Histone H3 and H4 lysine methylations and acetylations were profiled in >100 breast cancer samples, which were available as FFPE, OCT, or fresh frozen tissues (**Figure 11**). The sample cohort comprised normal breast tissues and the tumors belonging to the four breast cancer molecular subtypes, namely Luminal A, Luminal B, HER2-positive and TNBC. This analysis revealed differences in tumor compared to normal tissues as previously described with for instance loss of H4K20me3 (55). In addition, differences between breast cancer subtypes were found both in this dataset and in previous analyses. For instance, differences between TNBC and other breast cancer subtypes were reported, with more differences observed when compared with Luminal A tumors. Examples include a decrease of H3K27me3 and an increase of H3K9me3 (55,115,153).

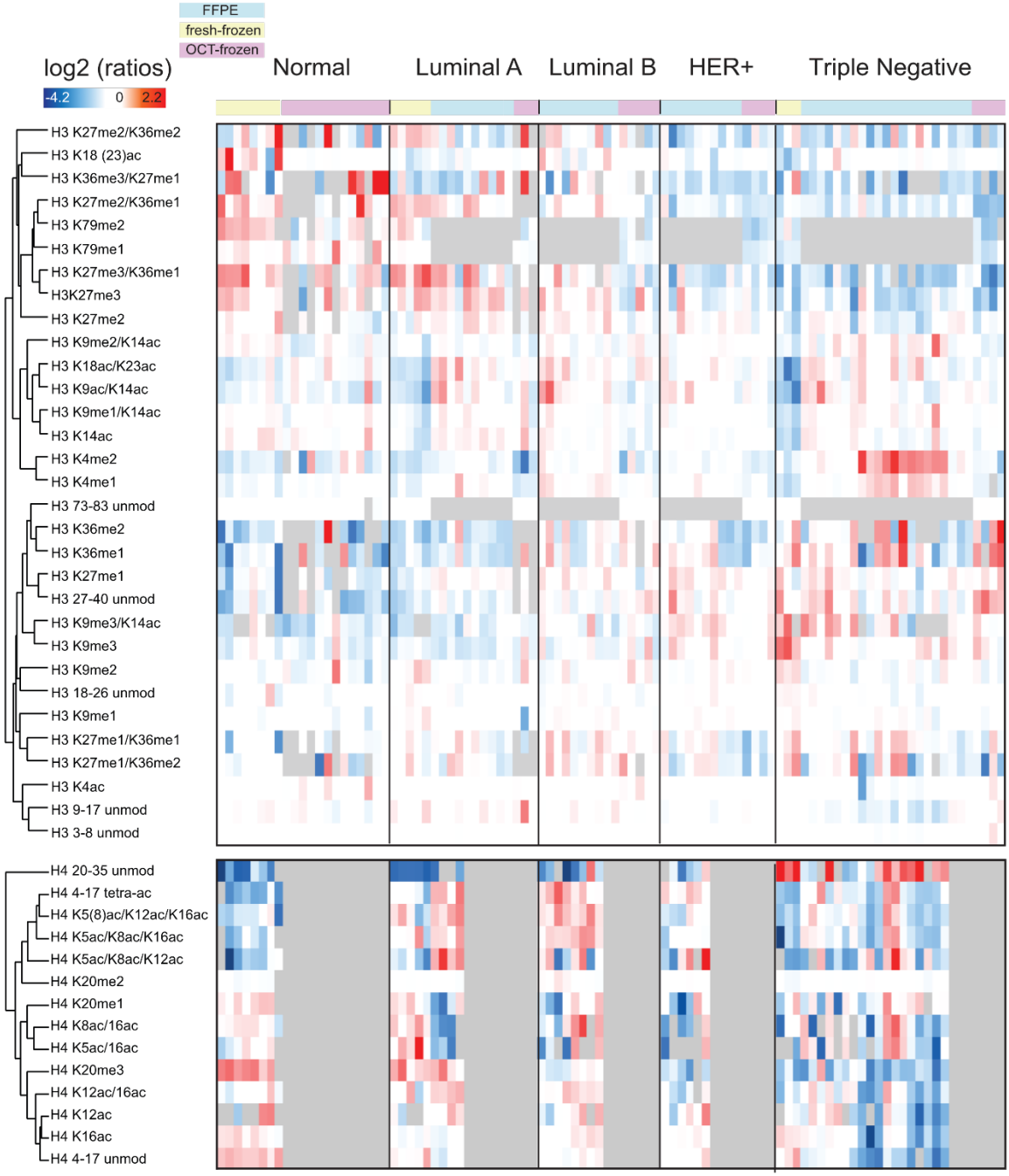


Figure 11: MS-based profiling of histone PTMs in breast cancer subtypes and normal breast tissues

Heatmap display of log2(L/H ratios) (L: breast biopsies, and H: super-SILAC internal standard) showing differentially modified histone peptides. Grey: not quantified. L/H relative abundance ratios were log2 transformed and normalized over the average value across the samples.

The patterns of histone PTMs identified in TNBC were very different from the other subtypes, especially when compared to Luminal A samples, which exhibited a PTM pattern more similar to normal tissues. This finding is in line with the notion that Luminal A samples are less aggressive compared to other subtypes. Even though many peptides were identified in all samples, it was not possible to quantify all modifications due to the type of storage conditions and technical limitations. A principal component analysis (PCA) of histone PTMs profiles further showed a separation of TNBC from the other breast cancer subtypes (**Figure 12A**). Thanks to this analysis, it was possible to identify an epigenetic signature distinguishing TNBC from other subtypes, which includes increased levels of H3K4me2, H3K9me3, and unmodified H4K20, along with decreased levels of H3K27me3, H3K79me, and H4K20me3 (**Figure 12B**). These are the preliminary results from which my Ph.D. project originated. Based on these findings, we selected the most interesting PTMs, namely H3K4me2, H3K9me3, and H4K20me3 for further investigation in order to explore the molecular mechanisms underlying the observed epigenetic changes in TNBC.

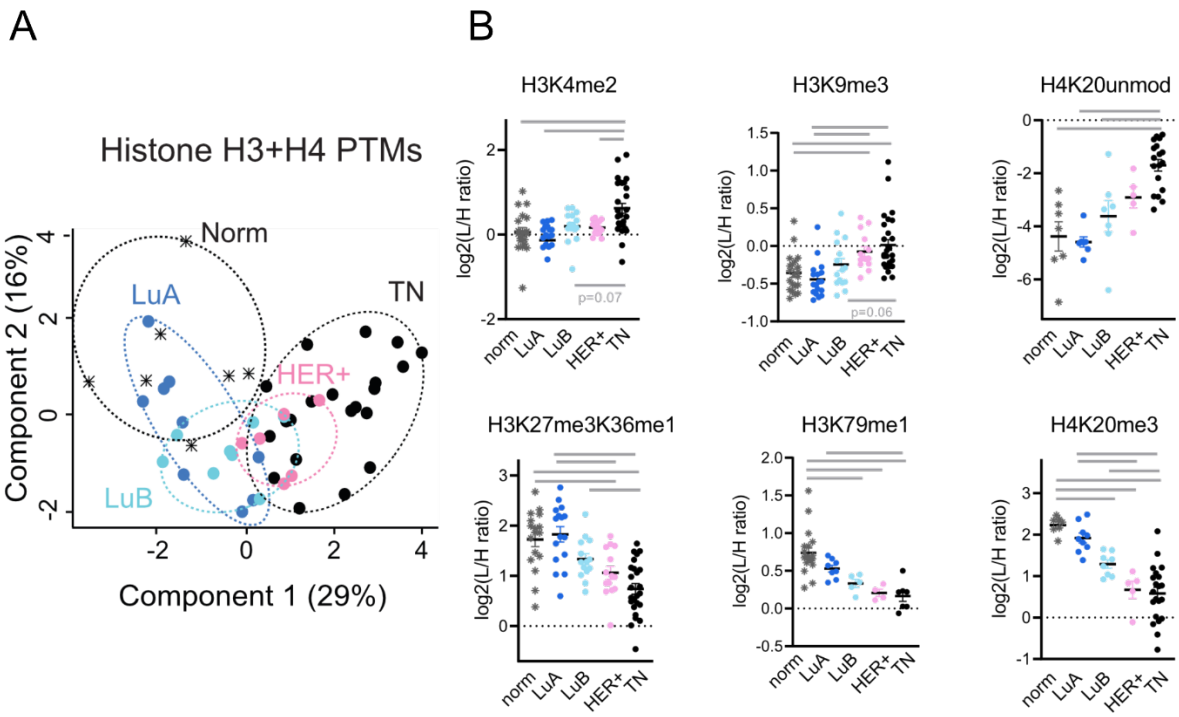


Figure 12: Epigenetic signature identified in TNBC subtype

A) PCA analysis based on the histone PTMs detected for histone H3 and H4 in the different breast cancer subtypes. (B) Epigenetic signature identified in TNBC. Changes of specific histone marks are shown with significant differences in TNBC compared to different breast cancer subtypes and shown as log2 transformed L/H ratios. Bars show p-value < 0.05 by one-way ANOVA. LuA: Luminal A-like, LuB: Luminal B-like, TN: Triple Negative, HER+: HER-positive, norm: Normal tissue.

3. AIM OF THE PROJECT

In recent years, the role of epigenetics has emerged to have a role in cancer initiation development, and progression. In particular, aberrant histone post-translational modifications have emerged as critical players in cancer development and progression and alteration in the levels of some of them have been recognized either as hallmarks of cancer (e.g. H4K20me3, H4K16ac, and H3K14ac) or as prognostic and predictive markers in various cancer types. In line with this, the histone modifying enzymes that catalyze the deposition and removal of histone PTMs are often altered or mutated in cancer and have been explored as potential therapeutic targets for cancer treatment. In this context, studying histone PTMs and altered epigenetic molecular mechanisms in primary cancer patient samples may provide potential molecular information to define epigenetic biomarkers for patient stratification and to suggest potential novel therapeutic targets. In my Ph.D. project, I studied epigenetic changes in one of the most aggressive breast cancer subtypes, TNBC, for which at the moment specific molecular markers and targeted treatments have not been identified.

The preliminary data, previously obtained in our laboratory and based on the MS-profiling of > 100 breast cancer samples, revealed an epigenetic signature distinguishing TNBC from the other breast cancer subtypes consisting in increased levels of H3K4me2, H3K4me3, H3K9me3, and H4K20un paralleled by a decrease in H4K20me3 levels and H3K27me3 levels.

Based on these data, the aim of my Ph.D. project was to investigate the molecular mechanisms underlying three specific PTMs, namely H3K4me2, H3K9me3, and H4K20me3, whose altered levels were distinguishing TNBC from the other BC subtypes.

First, to identify possible downstream effects of histone PTM changes, we dissected possible molecular mechanisms through a multi-OMICs analysis which we performed by integrating ChIP-seq and RNA-seq data in breast clinical samples and breast cancer cell lines. Second, with the purpose to detect some of the possible causes of deposition of aberrant histone PTMs we based our analysis on the study of altered histone modifying enzymes involved in the deposition of histone PTMs.

4. MATERIALS AND METHODS

4.1 Tissue Specimens

Breast fresh frozen cancer tissues were obtained from the Biobank for Translational Medicine Unit (B4MED) of the European Institute of Oncology. The collection of the samples by the Biobank, in the presence of patient consent, was approved by the Ethical Committee of the European Institute of Oncology on 6th June 2011 for research purposes, without any further approval required by the Ethical Committee (154). All the samples were collected and snap-frozen in liquid nitrogen and stored at -80°C before processing. Breast subtypes were defined by immunohistochemistry; Luminal A: ER (estrogen receptor) and/or PR (progesterone receptor) (+), Triple Negative: ER, PR, and HER2(−), irrespective of Ki67 score. For the TNBC subtype, samples with relapse within 3 years from chemotherapy are reported as EV+. Samples were selected with a tumor cellularity > 60%, to minimize contamination from non-tumoral cells. For histone PTM analysis, samples with cellularity > 40% were in some cases taken into consideration. The cellularity was evaluated by Hematoxylin & Eosin staining. All the samples were cut into small pieces of around 20-30-mg weight, for downstream analysis, as indicated in **Table 1**.

Sample	ALIQUOT ID	GROUP and relapse	Infiltrating carcinoma %	Inflammatory cells %	Connective tissue %	Normal tissue %	ChIP-seq	RNA-seq	MS proteome
TN1	15-B2-00516-02	TN	55	5	40				
TN2	15-B1-00599-02 A	TN, EV- 2020	70	5	25				
TN3	15-B2-00510-02	TN	70	5	25		x		
TN4	15-B1-00599-02 B	TN, EV- 2020	70	5	25		x		x
TN5	15-B1-00577-03	TN	60	10	30				
TN6	15-B1-00569-03	TN	60	5	35				
TN7	16-B2-00124-02	TN, EV-, 10-2020	90		10		x	x	x
TN8	16-B1-00272-02 A	TN, EV-, 07-2019	80		20				
TN9	16-B1-00272-02 B	TN, EV-, 07-2019	80		20		x	x	x
TN10	16-B2-00334-02 A	TN, EV-, 11-2020	55		45				
TN11	16-B2-00334-02 B	TN, EV-, 11-2020.	60		40				
TN12	17-B2-00216-02	TN, EV-, 01-2021	80		20		x	x	
TN13	17-B2-00247-02	TN, EV-, 08-2020	65		35				x
TN14	15-B1-00625-02	TN, EV+, 2017	65	5	30		x	x	x
TN15	16-B1-00132-02	TN, EV+, 05-2020	60-70		30		x		x
TN16	17-B1-00022-02	TN, EV+, 02-2018	65		35				x
TN17	17-B2-00267-02	TN EV+, 2019	40						
TN18	15-B1-00446-02	TN EV+, 03-2016	60				x	x	x
TN19	16-B1-00397-02	TN EV+, 02-2018	60				x	x	x
LU1	15-B2-00456-02	LUMA	45		50	5			
LU2	15-B1-00598-02	LUMA	60		35	5			
LU3	15-B1-00575-02	LUMA	70		25	5	x	x	x
LU4	15-B1-00337-02	LUMA	90		10		x		
LU5	15-B2-00302-02	LUMA	50		50				
LU6	15-B1-00667-02	LUMA	70		30		x	x	x
LU7	15-B2-00583-01	LUMA	75				x		x
LU8	16-B2-00273-02	LUMA	60				x		x
LU9	16-B2-00293-02	LUMA	50						
LU10	15-B2-00584-01	LUMA	75				x	x	x
LU11	15-B2-00377-03	LUMA	40-50						
LU12	15-B2-00388-02	LUMA	40-50						
LU13	16-B2-00259-02	LUMA	40						
LU14	15-B2-00521-02	LUMA	80				x		

Table 1: List of breast cancer clinical samples analyzed in this study

4.2 Cell culture and inhibitors

Breast cancer cell lines were grown in the corresponding media, as indicated in **Table 2**. All the media were supplemented with penicillin (100 U/ml) and streptomycin (100 mg/ml). Cells were grown at 37°C in a 5% CO_2 humidified atmosphere and passaged at 80-90% confluence. OICR-9429 and OTS186935 were purchased from MedChemExpress and re-suspended in DMSO.

Model	Cell line	Growth medium
Breast cancer (Triple Negative)	MDA-MB-231	DMEM with stable Glutamine +10% NA FBS
	MDA-MB-436	DMEM/HAM's F12 (1:1) with stable Glutamine +10% SA FBS
	MDA-MB-468	DMEM/HAM's F12 (1:1) with stable Glutamine + 10% SA FBS
Breast cancer (Luminal-A)	MCF7	DMEM with stable Glutamine +10% NA FBS
	ZR-75-1	RPMI1640 +10% SA FBS +2mM L-Glutamine+1mM Sodium Pyruvate+10mM Hepes
	T47-D	DMEM with stable Glutamine +10% SA FBS
Embryonic kidney	293T	DMEM with stable Glutamine +10% SA FBS
Breast cancer FUCCI	MDA-MB-231FUCCI	DMEM with stable Glutamine +10% NA FBS
Breast cancer dCas9	MDA-MB-231dCas9LSD1	DMEM with stable Glutamine +10% TET free serum

Table 2: Cell lines and cell culture media

DMEM= Dulbecco modified Eagle medium (Euroclone Spa, ECM0103L); FBS= Fetal Bovine Serum, NA=North American origin (FBS_NA, Fisher Scientific, 12389802), Ham F12 (Gibco), SA= South American fetal bovine serum (FBS_SA Microgem), RPMI1640, TET FREE (Tetraclyin free Serum (Euroclone Spa, ECS0182L). MDA-MB-231FUCCI cells were provided by the group of prof. Saverio Minucci.

4.3 Histone enrichment from fresh frozen samples and cell lines

Histones were enriched from fresh frozen samples and cell lines with the "low-input" protocol, as previously described (155). Briefly, 20 to 30 mg of fresh frozen tissue were resuspended in Nuclei Isolation Buffer (PBS 1x, 0.1% Triton X-100, protease inhibitors: 0.5 mM PMSF, 5 µM Leupeptin, 5 µM Aprotinin, 5 mM Na-butyrate), minced with sterilized scissors and homogenized in a 1 ml Dounce homogenizer (1 ml Tissue Grinder, Dounce, Wheaton). The sample was filtered through a 100 µm cell strainer (Falcon, 352360). Cell pellets (0.5-1x10^6) were lysed in 1 ml of Nuclei Isolation Buffer and resuspended vigorously, to break the plasma membrane. Next, nuclei were pelleted with a 15-min centrifugation at 2300 x g at 4°C and lysed with 40-100 µl of Nuclei Isolation Buffer supplemented with 0.1% SDS to break the nuclear membrane. To degrade DNA, 250 U of Benzonase (Sigma-Aldrich)

were added and activated with an incubation at 37°C for 2 min. To evaluate the purity of histones, 1/10 of the extract was loaded on a 17 % SDS-PAGE gel, together with 1, 0.5, and 0.25 µg of recombinant histone H3.1 (New England BioLabs). Before gel loading, the LDS sample buffer (Invitrogen, Thermo Fisher Scientific) supplemented with 10 mM DTT (Dithiothreitol; VWR 441496P) was added to the samples and incubated 5 min at 95°C. Histone quantification was performed after Coomassie staining (InstantBlue Coomassie Protein Stain, Abcam ab119211). 3 µg of histone octamer were mixed with an equal amount of heavy-isotope labeled histones (super-SILAC mix), which was generated from heavy-labeled breast cancer cell lines as previously described (112), and used as the internal standard for quantification.

4.4 "PRO-PIC" in-gel digestion prior to mass spectrometry analysis of histone PTMs

After gel loading and staining with Coomassie, histone bands were excised and subsequently cut into 1 mm^3 pieces for in-gel digestion, as previously described (156). Gel bands were de-stained in 50% acetonitrile (ACN, Carlo Erba) and dried in 100% ACN. Unmodified and mono-methylated lysines were chemically acylated with propionic anhydride (PRO, Sigma-Aldrich), 1 M ammonium bicarbonate (Ambic, Merck Life Science S.R.L.), and saturated sodium propionate ($C_3H_5NaO_2$, Merck Life Science S.R.L.) as catalyzer of the reaction. Samples were incubated at 37 °C for 4 h. The gel bands were then washed and dehydrated in ddH_2O, 50% ACN, and 100% ACN serially. Histones were digested by the addition of 100 ng/µl trypsin (Promega) solution, diluted in basic pH digestion buffer Ambic 50 mM, and incubated overnight at 37°C. After digestion, the peptides were extracted by the addition of 100% ACN. Peptides were dried in a Vacuum concentrator (SpeedVac, nConcentrator 5301, Eppendorf) to a final volume of 5 µl. Then the peptides were derivatized at the N-terminus at 37 °C by chemical modification with phenyl isocyanate (PIC, Sigma-Aldrich), buffered to pH 8.5 (with the reaction carried out for 1h 30min), in presence of 1M triethylammonium bicarbonate (Sigma Aldrich); this further modification allows increasing the hydrophobicity of short hydrophilic peptides (156). The reaction was stopped by adding 1% trifluoroacetic acid (TFA; Life Technology). Finally, the extracted peptides were desalted and concentrated with the use of C18 stage tips reversed phase micro-columns (157) C18 micro-columns are based on the hydrophobic interaction between the peptides and the C18 resin (Empore Octadecyl C18 47-mm Extraction Disks, Model 2215, 3M). After activation of the resin with 100% methanol, a wash with Buffer B (ACN80%, 0.1% FA), and equilibration with Buffer A (0.1% FA), peptides were loaded in Buffer A. Before MS

analysis, the samples were eluted in the elution buffer (50% ACN in 0.1%FA), concentrated, and resuspended in 1%TFA.

4.5 In-gel digestion for Histone H1 analysis and nuclear proteins analysis

Nuclear extract from clinical samples obtained with the low input protocol were processed for in gel digestion. First the band for histone H1 was cut by cutting a large band (20-45 kDa). The in gel digestion protocol was performed as previously described (158). Briefly gel bands were cut in 1 mm^3 cubes and de-stained in destaining buffer (25mM ammonium bicarbonate, ethanol 50% pH 8). Then, the gel pieces were dehydrated in ethanol 100%. After 1h reduction at 56 °C (10mMDTT in 50mM Ambic) and alkylation (55mM iodacetamide in Ambic 50mM) for 45 min at 25 °C, gel pieces were sequentially washed with digestion buffer (Ambic 50 mM), EtOH100%, and dried in SpeedVac. Digestion was performed overnight at 37°C with trypsin solution (12.5ng/µl trypsin in 50mM Ambic). The next day peptides were extracted by addition of extraction buffer (3%TFA in 30% ACN) with 2 cycles of extraction and 2 step of dehydration in ACN 100%. Samples were dried in a SpeedVac until 10-20% of the original volume. Prior to MS analysis, peptides were purified through a C18 stage tip (as described above), eluted in ACN80%, 0.1%FA, dried in SpeedVac and resuspended in 0.1% FA in water.

4.6 In solution digestion for total proteome analysis

20 mg of breast cancer fresh frozen samples were lysed in 100 µl of lysis buffer (0.1% Triton, 0.1%SDS, 1X PBS, protease inhibitors) and proteins were precipitated in acetone overnight. The in-solution digestion was performed using the iST Sample Preparation Kit (PreOmics). Briefly, 50 µl of LYSE buffer were added to 100 µg of proteins in order to reduce, alkylate and denature proteins with incubation for 10 minutes at 95°C. Proteins were digested with 50 µl DIGEST (mix of trypsin and Lys-C) at 37 °C with an overnight incubation. Next, the reaction was acidified and stopped with 100 µl of STOP solution. Samples were loaded on a C18 resin cartridge and washed. Peptides were then eluted from the cartridge by adding 120 µl of ELUTION solution and dried in a SpeedVac (Concentrator 5301, Eppendorf). High pH (HpH) Reversed-Phase fractionation (Pierce) was performed to reduce the sample complexity prior to LC/MS analysis. Dried peptides were re-suspended in 0.1% TFA and loaded on the HpH cartridge. According to a principle similar to C18 stage tips, the peptides were eluted with increasing percentages of organic solvent (ACN + 0.1% Triethylamine) for separation in 8 fractions. The 8 eluted fractions were then dried in a SpeedVac and re-suspended in LC-LOAD solution and sonicated in an ultrasonic bath (Elmasonic S10, Elma)

for 10 minutes. Samples were then loaded for LC-MS/MS acquisition. Fraction 7 and 8 were pooled prior to MS analysis.

4.7 Liquid Chromatography Tandem mass spectrometry analysis (LC-MS/MS)

Reversed-phase chromatography was used to separate peptide mixtures on a high-performance liquid chromatography (HPLC) system (EASY-nLC 1200, Thermo Fisher Scientific), through an EASY-Spray column (Thermo Fisher Scientific), 25-cm long (inner diameter 75 μm, PepMap C18, 2 μm particles) connected online to a Q Exactive HF or a Q Exactive Plus mass spectrometer (Thermo Fisher Scientific). The solvent used were: Solvent A (0.1% formic acid in H_2O) and Solvent B (80% ACN, 0.1% FA). Histone peptides were re-suspended in 1% TFA solution and injected at a flow rate of 250 nl/min and separated with a 55-min linear gradient of 10-45% solvent B. For histone H1 analysis peptides were injected in an aqueous 1% TFA solution and were separated with a 95 min 3%–60% gradient of solvent B (85 min 3–35%, 10 min 35–60%), at a flow rate of 250 nL/min. For total proteome MS data acquisition, peptides were separated with a 55-minute 3%-60% gradient of solvent B at a flow rate of 300 nl/min (50 min 3-30%, 5 min 30-60%). The Q Exactive instruments were operated in the data-dependent acquisition (DDA) mode. Survey full scan MS spectra (m/z 300–1350) were analyzed in the Orbitrap detector with a resolution of 60,000–70,000 at m/z 200. The 10-12 most intense peptide ions with charge state 2 were sequentially isolated to a target value for MS1 of $3x10^6$ and fragmented by Higher Energy Collision Dissociation (HDC) with a normalized collision energy setting of 28%. The maximum allowed ion accumulation times were 20 ms for full scans and 80 ms for MS/MS, and the target value for MS/MS was set to $1x10^5$. The dynamic exclusion time was set to 10-20 s, and the standard mass spectrometric conditions for all experiments were as follows: spray voltage of 1.7 kV, no sheath and auxiliary gas fow.

4.8 MS raw data analysis

Acquired MS RAW data for histone were mainly analyzed by the use of EpiProfile 2.0 software (PRO-PIC version). This is a universal and accurate platform for the quantification of histone marks via LC-MS/MS (111). In addition, the RAW data were analyzed using the integrated MaxQuant software version 1.5.2.8 (159), which generates peak list and identifies proteins using the Andromeda search engine (160). The Uniprot HUMAN_ histones 1502 database was used for histone peptide identification and analyzed as previously described

(161). Briefly, the enzyme specificity was set to Arg-C because of the use of an "ArgC-like" digestion. The estimated False Discovery Rate (FDR) of all peptide identifications was set at a maximum of 1%. The mass tolerance was set to 6 ppm for precursor and fragment ions. One missed cleavage was allowed, and the minimum peptide length was set to 5 amino acids. Variable modifications for the PRO-PIC protocol included lysine propionylation (+ 56.0262 Da), monomethylation with propionylation (+ 70.0419, which correspond to the sum of monomethylation and propionylation), dimethylation (+ 28.0313 Da), trimethylation (+ 42.0797 Da), acetylation (+42.04695 Da) and N-terminal PIC labelling (+ 119.0371 Da). Peptides with Andromeda score less than 40 and localization probability score less than 0.75 were removed. The histone PTM analysis was then performed manually using the QualBrowser software version 4.0.27.13 (Thermo Fisher Scientific), based on peptide identification, retention times and elution patterns of isobaric peptides. For each histone modified peptide, after extraction of the area under the curve (AUC) at specific retention times, the percentage relative abundance (%RA) was estimated by dividing the AUC of each modified peptide for the sum of the areas corresponding to all the observed forms of that peptide and multiplying by 100. Light/Heavy (L/H) ratios of %RA were then calculated and log2 transformed. Acquired MS RAW data for histone H1 and total proteome analysis were analyzed using the integrated MaxQuant software v.1.6.2.3, which performed peak list generation and protein identification using the Andromeda search engine. The Uniprot HUMAN_1802 was used for peptide identification. Enzyme specificity was set to trypsin, and two missed cleavages were allowed. Methionine oxidation and N-terminal acetylation were included as variable modifications, and the FDR was set to 1%, both at the protein and peptide level. The label-free software MaxLFQ was activated as well as the "match between runs" feature (match from and to, matching time window = 2 min). Analysis of Histone H1 was performed by normalization of LFQ values on average of all the values. The protein groups output from MaxQuant was analyzed using Perseus by filtering out values detected as "potential contaminants", "only identified by site", and "reverse" and by selection of 75% of valid values in at least one group. LFQ values were plotted in a volcano plot by t-test with significant p-value< 0.05. Statistical analysis was performed with Perseus software version 1.6.15.0 for both histone PTMs analysis and global proteomic analysis.

4.9 Chromatin Immunoprecipitation (ChIP)

The protocol for crosslinking chromatin immunoprecipitation (X-ChIP) was adapted and modified from (162) and the Abcam X-ChIP protocol (163). Chromatin was prepared from $5x10^7$ to $7x10^7$ cells at 80% confluence, or 25–30 mg of fresh frozen tissues in PBS with protease inhibitors (0.5 mM PMSF, 5 μM Leupeptin, 5 μM Aprotinin, 5 mM Na-butyrate). The tissue was cut into small pieces using scissors and homogenized in a Dounce homogenizer. The samples were fixed in 1% v/v formaldehyde (Sigma, F8775) in PBS and incubated at room temperature for 10 minutes to stabilize protein-DNA interaction. Formaldehyde was quenched by the addition of 125 mM Glycine (Serva, SE39004) for 5 min. The samples were centrifuged at 13000 rpm for 5 minutes at room temperature, washed twice in PBS, and then resuspended in ChIP SDS lysis buffer with protease inhibitors (100 mM NaCl, 50 mM Tris-Cl pH 8.1, 5 mM EDTA pH 8, 0.2% NaN3, and 0.5% SDS in water). Nuclei from cell lines were pelleted for 6 minutes at 1200 rpm, while nuclei from fresh-frozen tissues were pelleted at 3750 rpm at 4°C for 5 minutes. All the samples were resuspended in 3 ml (cell lines) and 1ml (fresh frozen tissues) ice-cold IP Buffer (100mM Tris-Cl pH 8.6, 100mM NaCl, 5mM EDTA pH 8.0, 0.2 % NaN_3, 5% (v/v) Triton 100X, water) for sonication. The chromatin was sonicated with a Branson Digital Sonifier 250, with a 3 mm microtip for fresh frozen samples and a large Tip for cell lines, at 30% Amplitude, for 15 second on and 15 second off, for 8 to 12 cycles (depending on the type of sample), in order to achieve a length of around 200-300 bp. The sheared chromatin was quantified by Bradford assay and 400 μg were incubated overnight at 4 °C with antibody and IgG as negative control (*see paragraph 4.20*). 1% (for cell lines) and 2% (for fresh frozen tissue) of the input chromatin was collected prior to the addition of the antibody and stored at -20 °C. The immunoprecipitation was performed by adding Dynabeads Protein G (Invitrogen, 10004D), previously washed in IP buffer, for 3 h on a rotating wheel at 4 °C, followed by two washes in ChIP Wash buffer 150 mM (1% Triton 100, 0.1% SDS, 150 mM NaCl, 2mM EDTA pH8.0, 20mM Tris-Cl pH 8.0, water), one wash in Wash buffer 500 mM (1% Triton 100, 0.1% SDS, 500 mM NaCl, 2mM EDTA pH8.0, 20mM Tris-Cl pH 8.0, water) (to remove the unspecific background), and one wash in TE 1x (to remove salt residues). The DNA was eluted and de-crosslinked with SDS 2% in TE1x and Proteinase K (Merck Life Science S.R.L.), overnight in thermomixer at 65 °C at 1100 rpm. The de-crosslinking step was also carried out for the input samples. DNA fragments were purified with the Qiagen PCR purification kit (QIAquick PCR Purification Kit, Qiagen, 28106), according to the manufacturer protocol. Half of the immunoprecipitated chromatin was instead processed for Western Blot and

prepared to check immunoprecipitation efficiency. In this case a Flow Through (FT) 1% in case of cell lines and 2% in case of tissues was also processed. Briefly after the second wash in Wash Buffer 150mM, a 1X (250mM Tris-Cl pH 8.8, 2% SDS,0.5M Beta-mercaptoethanol in water) and 2X (500mM Tris-Cl pH 8.8, 4% SDS, 1M Beta-mercaptoethanol in water) de-cross-linking solutions were added respectively to IP/IgG and input/FT samples for decrosslinking, elution, and protein denaturation. LDS sample buffer (Invitrogen) supplemented with 10mM DTT (DTT – Dithiothreitol; VWR 441496P) was added and the samples were loaded on a pre-cast 4-12% gel (Nupage Novex 4-12%, Life Technologies Italia). For validation of ChIP prior to sequencing, a ChIP-qPCR was performed: 1 µl of DNA from ChIP and input samples was mixed with 10 µl of SYBR green Master Mix (Applied Biosystems, 4385614) and 0.5 µM of each primer mix in a final reaction volume of 20 µl. Real time PCR was performed in two technical replicates (n=2) and quantitative PCR amplifications were carried out in the Viia7 (BioRad), using the following conditions: 1) 95 °C for 2 minutes, 2) 40 cycles at 95 °C for 10 seconds and 60 °C for 30 seconds. Percentage of IP enrichment compared to the input was calculated as % of input x $2^{\wedge(\Delta CT)}$, where ΔCT= Ct input - Ct IP.

4.10 ChIP Sequencing (ChIP-seq) and analysis

ChIP-sequencing was performed in collaboration with the Genomic Unit at IEO and data were analyzed by the bioinformatician of our group A.Vai. DNA libraries were prepared from 5 to 10 ng of input and IP DNA through an in-house protocol (164) and sequenced on a NovaSeq 6000 (Illumina) instrument with a read length of 50bp PE (Paired End) and a sequencing depth of 50 million reads per sample for H3K4me2 and H3K4me3 ChIP, and 80 million reads for H3K9me3 and H4K20me3 ChIP-experiments. The increased number of reads for H3K9me3 and H4K20me3 is derived from the fact that these modifications are more spread in the genome and located at repetitive regions, so a higher percentage of reads might be lost in the alignment. Reads were mapped to the hg38 version of the reference human genome assembly using Bowtie (165) keeping only uniquely mapping reads and allowing for no more than two mismatches per read. The options “best and strata” were enabled. Specifying these options makes Bowtie report only the alignment with the lowest number of mismatches. The maximum insert size for valid paired-end alignments was set to 700bp. To identify H3K4me2 enriched regions, MACS2 was employed (166), enabling the –BROAD option and setting both the q-value and the BROAD-CUTOFF parameter to 0.001. Luminal A only peaks were defined as those peaks overlapping in four or more Luminal A samples and with no overlap with any of the TN samples. Analogously, TN only peaks were

defined as those peaks overlapping in four or more TN samples and with no overlap with any of the Luminal A samples. Common peaks are those peaks found in at least four Luminal A samples and in at least four TN samples. All these overlaps were performed using BEDOPS (167). Genomic annotations of peaks were performed using BEDTools (168) and using the RefSeqCurated (hg38) genome annotation. To resolve the ambiguity of peaks overlapping with multiple genomic features, we assigned the peak to the feature with the highest coverage. For H3K4me3, the peaks were called with MACS2 using the narrow mode and setting 0.001 as q-value cut-off. For H3K9me3 and H4K20me3, peaks were called using SICER2 (169) with the following parameters: --bin_size=200bp, --gap_size=600bp, --effective_genome_faction=0.8 and q-value of 0.01. The peaks list of all the modifications were filtered to remove blacklisted regions before proceeding with downstream analysis (170).

The list of breast cancer super-enhancers (SE) was retrieved from (125,171) and from SEdb (172). SE hg19 genomic coordinates were lifted to hg38 using the UCSC liftOver tool (125). SE retrieved from (125) already provided a Super-enhancer-gene association, while for the others super-enhancers, corresponding genes were assigned if super-enhancers were no more than 50kb away from gene promoter (defined as TSS ± 1kb).

Centromeric regions coordinates were retrieved from https://github.com/marbl/CHM13 which refers to the telomere-to-telomere genome assembly (173) and corresponding sequences were extracted using an in-house R script. Reads were mapped directly against repetitive sequences using Bowtie, allowing for no more than two mismatches per read and keeping only one alignment per read. Both the INPUT and IP reads were mapped. To assess if reads mapped to repetitive sequences more than by chance, we computed a p-value with the Poisson distribution using as lambda parameter the number of INPUT reads mapped.

4.11 Reads density at promoters and SE and visualization of peaks

To identify differences in the reads density distribution either at promoters or at SE, we used featureCounts (174). Counts were normalized using CPM (counts per million), and regions with very low counts removed with the edgeR function filterByExpr with default parameters, and used as input for the multi-dimensional scaling (MDS) plot, produced with the plotMDS function implemented in edgeR. To visualize peaks, the BigWig files were uploaded to the IGV tool version 2.16.0. Big Wigs were generated starting from BAM files with the bamCoverage function of deepTools with the following parameters: --ignoreDuplicates --

centerReads –bin-size 10bp --normalizeUsing RPGC (reads per genomic content) --effectiveGenomeSize 2701495761.

4.12 RNA extraction and retro-transcription

RNA was extracted from both cell lines and fresh frozen samples after homogenization of the tissue through a 2 ml syringe, by using the Quick-RNA™MiniPrep kit (Zymo Research), according to the manufacturer protocol. RNA quality was evaluated on the Agilent 2100 Bioanalyzer prior to sequencing with a RNA Integrity Number (RIN) equal or above 7. A total of 500 ng RNA was reverse-transcribed into cDNA using the One script plus cDNA synthesis kit (Abm). RT-qPCR analysis was performed in 20 µL of final reaction volume with addition of Fast SYBR green Master Mix (Applied biosystems, Thermo Fisher Scientific), retro-transcribed cDNA, and 0.5 µM of primer mix. All the quantitative PCR amplifications were performed in the Viia7 PCR machine (BioRad) with the following protocol: 1) 95 °C for 2 minutes, 2) 40 cycles at 95 °C for 10 seconds and 60 °C for 30 seconds. Expression of target genes was calculated with normalization on GAPDH or βACTIN (reference genes) with the $2^{-\Delta\Delta CT}$ method, where ΔCT = CT(target gene) - CT(reference gene) and ΔΔCT = ΔCT(target sample) - ΔCT(a reference sample) (175). The primers used are listed in **Table 3**.

GENE	Forward	Reverse	Application
SUV39H1	GTCATGGAGTACGTGGGAGAG	CCTGACGGTCGTAG ATCTGG	RT-qPCR
SUV39H2	TGGGGTGTAAAGACCCTTGTG	ATTCC CTTGTTGTCATAGAAC	RT-qPCR
SETDB1	AGGAACTTCGGCATTTCATCG	TGTCCCGGTATTGTAGTCCCA	RT-qPCR
GAPDH	GAGTCAACGGATTTGGTCGT	GACAAGCTTCCCGTTCTCAG.	RT-qPCR
β-ACTIN	CTCTTCCAGCCTTCCTTCCT	AGCACTGTGTTGGCGTACAG	RT-qPCR
SLAMF7	CCTTCATCTGCGTTGCCAGGAA	GGAGGAATCTGGGTCATCAGCA	RT-qPCR
PLA2G4A	5ACTGCACAATGCCCTTTACC	CGGGAGCCATAAAAGTACCA	RT-qPCR
IGF2BP3	TCGTGACCAGACACCTGATGAG	GGTGCTGCTTTACCTGAGTCAG	RT-qPCR
ANLN	CAGACAGTTCCATCCAAGGGAG	CTTGACAACGCTCTCCAAAGCG	RT-qPCR
MSLN	CCTGAGGACATTCGCAAGTGGA	CTTCCCTTCACAAAGCGGTCGA	RT-qPCR
PRSS3	GCTGAGTGTAAAGCCTCCTACC	CAACTCCTTGGAGCTGTCCGTT	RT-qPCR
GAPDH	TTCGCTCTCTGCTCCTCCTG	CCTAGCCTCCCGGGTTTCTC	ChIP-qPCR
β-ACTIN	AGCCGGCCTTGCACATG	GGCTATTCTCGCAGCTCACCA	ChIP-qPCR
Gene desert	ACTAGAAAACACTGAGTTGCTTGATT	CCCTGCTTTGCTCTACCAAC	ChIP-qPCR
ITGB4	GAGCTCTTAGCCCGAGTCATC	GCTAAGCAGAGAACGGATGC	ChIP-qPCR

Table 3: List of primers for RT-qPCR and ChIP-PCR

4.13 RNA sequencing and analysis

RNA sequencing (RNA-seq) from internal samples was performed in collaboration with the Genomic Unit at IEO and the analysis was performed by A.Vai. Libraries were prepared with the TruSeq RNA Sample Preparation v2 kit (RS-122-2002, Illumina), starting from 100 ng of total RNA for clinical samples and 500 ng of total RNA for breast cancer cell lines. Sequencing was performed using a NovaSeq 6000 (Illumina) instrument with a read length of 50bp PE and a sequencing depth of 35 million reads per sample for breast cancer cell lines and a read length of 150bp PE and a sequencing depth of 80 million reads per sample for breast clinical samples.

In the case of TCGA data, the RNA-seq count table from TCGA-BRCA dataset was used for the integrative analysis. Genes with an exon length < 200 bp and mitochondrial genes were removed from the count table. The annotation used was Gencode version 38. Significant differentially expressed genes were identified using edgeR with a FDR cutoff of 0.05 and an absolute fold change greater than 1. The matched RNA-seq data were analyzed with the RNentropy method and differentially expressed genes were defined as described in (176). Distribution of RNA quantification values for genes bearing H3K4me2 were plotted based on transcript per million (TPM) values in promoters, intronic and intergenic regions and computed by Wilcoxon test.

4.14 Cell transfection and infection of breast cancer cell lines with lentiviral vectors

Lentiviral transductions were performed in a Biosafety level 2 laboratory (BSL-2). Lentiviral constructs were transiently transfected in 293T cells by using the calcium phosphate transfection method together with the packaging plasmid pCMV-DR8.91 and the envelope plasmid pMD2G-VSVG. One 10mm plate of 293T was used to infect a well of a 6 well plate or two 293T 10 mm plates were used to infect one 10 mm plate of breast cancer cell lines. After 12 hours of transfection, the supernatant was removed and 6 ml of fresh medium were added, in order to remove calcium precipitates and concentrate the virus. After 48h-72h from transfection, the supernatant containing the virus was harvested and used to infect MDA-MB-231 or MCF7 cell lines with the addition of Polybrene 1X, a cationic polymer which neutralizes the charge repulsion between virions and the cell surface, resulting in increased infection efficiency. Infected cells were then selected by incubation with different

antibiotics, depending on the resistance carried by the vector or by FACS sorting, when a fluorescent mark was present (e.g mCherry or BFP).

4.15 CRISPRi: generation of stable cell line expressing dCas9

To generate CRISPRi stable cell line, MDA-MB-231 cell line was first transduced with lentivirus expressing Tet-On 3G rtTA transactivator with Blasticidin resistance (pLVX-Tet3G Blasticidin, Addgene, #128061) and selected for 10 days with addition of 10 µg/µl Blasticidin to cell media. Next, cells were transduced with dCas9-LSD1_mCherry (pHR_TRE3G-LSD1-dCas9-P2A-mCherry Addgene #138462). Because this type of vector is inducible, doxycycline (Thermo Fisher Scientific) was added following infection (1 µg/ml for 48 h) and cells expressing mCherry were FACS sorted by definition of the mCherry gate.

4.16 CRISPRi: sgRNA design and vector cloning

For sgRNAs targeting regions with H3K4me2, target regions were chosen by selecting two regions within the signal of H3K4me2 in the promoter region. gRNA sequences were designed with the CRISPOR Tool (177) taking into account the specificity and off-target scores. sgRNA were cloned into the Addgene Vector pKLV2-U6gRNA5(BbsI)-PGKpuro2ABFP-W (Addgene, #67974) characterized by the presence of the BFP marker. The vector was linearized by digestion with BbsI enzyme (NEB, R3539S) and purified with the Qiaquick Gel Extraction Kit (Qiagen Spa). Briefly, 1 µl of 100 µM Top strand oligo, 1 µl of 100 µM Bottom strand oligo, 1 µl 10x T4 ligation buffer (NEB), 0.5 ul T4 PNK (NEB M0201) were mixed on ice with nuclease free water to reach a final volume of 10 µl. The reaction was performed as follows: 1) 37 °C for 30 minutes, 2) 95 °C for 5 minutes 3) ramp down to 25 °C (at 0.1°C/second). Ligation was performed after a serial dilution of oligos. First dilution (142 fmol/µl): 139 µl EB buffer + 2 µl 10 µM ds-oligo, and second dilution (7.1 pmol/µl): 57 µl EB buffer + 3 µl 1st dilution. The ligation was performed at RT for 2h after mixing the following reagents on ice: 20 ng/ µl linearized lentiviral vector (=3.7 fmol), 7.1 fmol/µl ds-oligo 2 µl (=14.2 fmol), 1 µl 10x ligase buffer (NEB M0202S), 1 µl T4 ligase (NEB M0202S) to a final volume of 10 µl by addition of water. The ligated vector was transformed in STBL3 competent bacteria and DNA was purified with the Plasmid DNA purification kit (NucleoSpin Plasmid Macherey-Nagel). The purified DNA was sequenced by Sanger Sequencing to assess the presence of the sgRNA sequence. The list of oligos for gRNA synthesis is provided in **Table 4**. The negative control sgRNAs Scramble1 and Scramble2 were previously published (124). MDA-MB-231 dCas9-LSD1mCherry cells

were plated at 0.5 x 10^6 per well in 6 well. 24h-48h after plating, each well was transduced with lentiviruses containing sgRNAs, which were packaged in 293T cells. To maximize dCas9 and sgRNA expression, doxycycline was added to induce expression of dCas9 and cells expressing higher fluorescence for both mCherry and BFP were FACS sorted 72h post-induction.

Guide	Target	Forward	Reverse
gRNA1	SLAMF7	CACC**GGCCTCTGGCTCAACTCGGG**	AAAC**CCCGAGTTGAGCCAGAGGCC**
gRNA 2	SLAMF7	CACC**GGTCTGAGTAATACCTAAGA**	AAAC**TCTTAGGTATTACTCAGACC**
gRNA 3	PLA2G4A	CACC**GTAGCATTCAAAACGGCGGTG**	AAAC**CACCGCCGTTTTGAATGCTA**C
gRNA 4	PLA2G4A	CACC**GTCTCTCGCTAAAGATCGCT**	AAAC**AGCGATCTTTAGCGAGAGAC**
gRNA 5	IGF2BP3	CACC**GGAACGCGTGCGTAAACGCT**	AAAC**AGCGTTTACGCACGCGTTCC**
gRNA 6	IGF2BP3	CACC**GTGGAACGCGTGCGTAAACGC**	AAAC**GCGTTTACGCACGCGTTCCA**C
gRNA S1	Scramble1	CACC**GCTGATCTATCGCGGTCGTC**	AAAC**GACGACCGCGATAGATCAGC**
gRNA S2	Scramble2	CACC**GAACAGTCGCGTTTGCGACT**	AAAC**AGTCGCAAACGCGACTGTTC**

Table 4: List of oligos for sgRNA synthesis

4.17 Design of plasmids with shRNA for knockdown of selected methyltransferases

Human shRNA lentiviral plasmids, carrying puromycin resistance, for knockdown of the H3K9 methyltransferases SETDB1, SUV39H1 and SUV39H2 were purchased from Sigma as MISSION shRNA Bacterial Glycerol Stock (SHCLNG-NM_003173, SHCLNG-NM_024670, SHCLNG-NM_012432):

- sh1#SETDB1 (TRCN0000276169),
- sh3#SETDB1(TRCN0000148112),
- sh5#SUV39H1 (TRCN0000275372),
- sh6#SUV39H1 (TRCN0000275322),
- sh9#SUV39H2(TRCN0000006938),
- sh12#SUV39H2(TRCN0000011057).

A control vector pLKO.1 Scramble with resistance to Puromycin was transformed into STBL3 competent cells. MDA-MB-231 were plated in a 10 mm plate and after 24h were transduced with lentiviruses for sh1, sh3, sh5, sh6, sh9, sh12, and scramble. As an additional control also the MCF7 cell line was transduced with lentivirus for the scramble vector. Selection was performed for 48h by addition of puromycin, 2.5 µg/mL for MDA-MB-231

and 1.5 µg/mL for MCF7. RT-qPCR analysis and MS analysis data shown were performed 48h after Puromycin selection.

4.18 Immunostaining of unfixed chromosome spreads

The protocol for immunostaining of unfixed chromosome spreads was adapted from (178). Cells in exponential growth with a confluence of 70/80% were blocked in metaphase by adding 0.1 µg/ml Colcemid, which works as an inhibitor of the mitotic spindle by depolymerization of microtubules. After incubation for 2-3 h at 37°C in a 5% CO_2 humidified atmosphere, mitotic cells were harvested by shake-off, a wash in PBS, and a 1 min wash in trypsin. Cells were pelleted at 1200 rpm for 5 min at 4°C. After 2 washes in PBS, the cells were counted and re-suspended in the hypotonic solution KCl 75 mM and incubated at RT in a rotating wheel for 10 min. Approximately $1x10^5$ cells were spun onto ethanol washed glass slides at 1800 rpm for 10 min with a medium acceleration with a Shandon Cytospin 4. Washes steps were performed for 10 min in KCM buffer (120mM KCl, 20 mM NaCl, 10 mM Tris-HCl pH 8.0, 0.5 mM EDTA, 0.1% Triton, water) at RT. Primary and secondary antibodies used are listed in **Table 5**. After one wash in KCM buffer, immunostaining with primary antibodies diluted in KCM supplemented with 1% BSA was performed for 1 hour at 4°C. After a wash in KCM buffer, a mix of secondary antibodies conjugated with different fluorophores was added and was incubated for 1 h at 4°C. Slides were washed in KCM buffer and fixed in 4% (v/v) formaldehyde in KCM for 10 minutes and rinsed in deionized water. After a wash in KCM buffer, staining with DAPI (Sigma-Aldrich Merck) was performed at 1: 1000 dilution in 1% BSA KCM for 15 minutes at 4°C and slides were mounted in a drop of Mowiol-Dabco mounting medium (Sigma-Aldrich Merck) and let to dry overnight at RT. Slides were stored at 4°C prior to acquisition.

4.19 Image Acquisition and Analysis

Evos Fluorescence Microscope was used for taking pictures of cells in culture. Chromosome images were acquired with a Leica DM6 B MultiFluo with HC PL APO 100X/1.40 objective lens with a Andor Zyla VSC-04470 sCMOS Camera with a pixel size of 6.5 µm and a resolution of 2048 x 2048 (4.2 Mpixel). Images were acquired with fluorescent filters for DAPI, Cy3 and Alexa488 and analyzed by a custom-made Fiji/ImageJ macro. In all staining conditions CENP-A staining was used as a reference for peri-centromeric regions with use of the Huang and the Otsu algorithms to automatically define the thresholds for H3K9me3 and H4K20me3 signals inside the chromosome area and inside the centromere area. The

Total Raw Intensive density of each mark at centromere was normalized on the Total Raw Intensive density in the chromosome area defined by DAPI staining.

4.20 Antibodies

The antibodies used for different applications are listed in **Table 5** with specific targets and concentrations.

Target	Antibody	ChIP:(µg of antibody: µg of chromatin)	Chromosome staining	WB
H3K4me2	H3K4me2 antibody (ab7766, Abcam)	1:100		1 µg/ml
H3K4me3	H3K4me3 antibody (ab8580, Abcam)	1:50		1 µg/ml
H3K9me3	H3K9me3 (C15410193, Diagenode)	1:50	1:250	1:2000
H4K20me3	H4K20me3 antibody (ab9053, Abcam)	1:50	1:200	1 µg/ml
H3 total	H3 Antibody (ab1791, Abcam)			1:5000
CENP-A	CENP-A antibody (ab13939, Abcam)		1:250	
IgG	Merck Life Science S.R.L., normal rabbit IgG (12-370)	1:100 or 1:50		
dCas9	Anti-CRISPR-Cas9 antibody [7A9-3A3] (ab191468, Abcam)			1 µg/ml
Alexa Fluor® 488	AffiniPure Donkey Anti-Rabbit IgG Jackson (ImmunoResearch)		1:200	
Cy™3	AffiniPure Donkey Anti-Mouse IgG Jackson (ImmunoResearch)		1:200	

Table 5: List of antibodies and adopted dilutions depending on experiment

4.21 Cell synchronization for cell cycle analysis assay

To arrest the cells in the G2/M phase, sub-confluent MCF7, and MDA-MB-231 were incubated for 24 h in growing media supplemented with 3.5 mM Thymidine (Merck Life Science S.R.L.). Cells were then released for 3 h and incubated for 20 h in a medium containing 100 ng/ml Nocodazole (Merck Life Science S.R.L.). Cell-cycle profiles of synchronized cell lines were verified by flow cytometry with Propidium Iodide staining. The percentage of cells in the different phases of the cell cycle was calculated with FlowJo 10.8.1 Software.

4.22 Phenotypic assays on BC cell lines

Clonogenic assay was performed as described in (179). Cells were seeded in 6-well plates, at 500 cells per well and let to grow for 14 days or 4000 cells were seeded in p60 and let to grow for 8 days. Cells were fixed in methanol and stained with crystal violet solution. Colonies were counted by colony count tool on ImageJ Fiji.

Proliferation assay was performed in 96 well plate from time 0h to 120h after seeding 2500 cells/well with Cell Titer Glo (Promega) assay. Measurements were made according to

manufacturer instructions. Briefly, an equal volume of Cell Titer-Glo reagent was added directly to the wells containing media. Plates were incubated at room temperature for 30 minutes and luminescence was measured with Glomax Explorer

For migration assay cells were re-suspended at a cell concentration of $5x10^5$ cells/ml and 70µl were seeded in each well of a 2µm insert in Ibidi plate. After 24h the insert was removed. Migration was determined by wound-healing assay at 0, 12, and 24 h after seeding. Migration was defined bas wound area normalized on time 0h with the formula: $(time_0-time_X)/time_0$ normalized on scramble control. Migration was calculated by definition of wound closure area by ImageJ Fiji. Pictures were taken at Evos at 2x and 4x magnification.

4.23 Western blot analysis

Western Blot analysis was performed starting from nuclear or total cell extracts. Protein amount was quantified with the Bradford assay or with the Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific). Equal protein amount was separated through an SDS-Page gel electrophoresis and transferred on a PVDF membrane (Immobilon-P, Merck Millipore), activated in methanol and transferred by wet-transfer method. Unspecific signal was blocked with BSA 10% in TBS with 0.1% Tween 20 for 1h at RT. The primary antibody was incubated overnight at 4°C or for 1 h in the case of normalizers. After 3 washes in TBS with addition of 0.1% Tween 20, the membrane was incubated for one hour in HRP-conjugated secondary antibody (Cell Signaling Technology). Next, proteins were detected by enhanced chemiluminescence ECL solution 1: 1 v/v solution A: Solution B (ECL, Bio-Rad) after a 5 min incubation.

4.24 Flow cytometry and cell sorting

Flow cytometry analysis was performed by FACSCelesta by the use of a FACSDiva Software v8.01.1 and analysed using FlowJo 10.8.1 Software.

Cell cycle analysis was performed by Propidium Iodide (PI) staining. PI is a fluorescent intercalating dye of nucleic acid which bind in proportion to quantity of DNA within a cell. Consequently, cells in the S phase, containing more DNA than those in G1, will absorb a higher amount of the dye and exhibit a brighter fluorescence until their DNA content doubles. Cells in the G2 phase will appear twice as bright as cells in G1. Briefly, 0.5-1 x 10^6 cells were harvested, washed with 1% BSA in PBS and re-suspended in 250 µl PBS. To allow DNA staining with PI, cells were fixed with 750 µl of EtOH added dropwise while vortexing, and then they were incubated on ice for at least 30 min. Fixed cells were

then washed with 1ml of 1% BSA in PBS and after centrifugation at 3000 rpm for 5 min were re-suspended in 500 µl of PI 50µg/ml diluted 1:8 in PBS and 250 µg/ml of RNase and incubated overnight at 4°C.

Because of lentiviral infection, evaluation of mCherry and BFP expression was performed only after cell fixation. Briefly, cells were pelleted at 3000 rpm for 5 minutes at 4°C and resuspended in 1% BSA in PBS. Then cells were pelleted at 3000 rpm for 10 min and resuspend in 250 µl cold PBS. Cells were fixed by the addition of 250 µl formaldehyde 2% in PBS (v/v). After 20 min incubation on ice, cells were pelleted at 3000 rpm for 10 min, washed in 1 ml BSA 1% in PBS and resuspended in 500µl PBS. Samples were kept at 4°C prior to acquisition.

Sorting was performed on sorterMelody and FACSAria with definition of gates by the use of a negative control, namely the wild type cell line not expressing the fluorescence. For sorting of cells expressing dCas9mCherry and/or sgRNA-BFP, prior to sorting, cells were re-suspended in sorting medium (Medium + 2% FBS + 1% P/S + 0.3% v/v Gentamicin 50mg/ml) at 5 x 10^6 per ml. High protein concentrations can disrupt sort stream formation therefore a low amount of serum (2% FBS) is sufficient to keep the cells alive for the duration of the sorting procedure. Then cells were filtered with 100µm cell strainer and dispensed in sterile FACS tubes. Sorted cells were collected in collection medium (Medium + 33% FBS + 3% P/S + 0.3% Gentamicin) and kept in culture. For sorting of MDA-MB-231FUCCI cells, prior to sorting, cells were re-suspended in sorting medium (PBS + 2% FBS + 1% P/S) at 5 x 10^6 per ml. Sorted cells were collected in PBS and pelleted for downstream analysis.

4.25 Statistical analysis

Statistical analysis was performed by GraphPad Prism version 9.3.1. t-test was performed for comparison of two samples with significant p-value<0.05. One-way ANOVA test was used for comparison of three or more samples with significant p-value<0.05. The type of statistical analysis performed is indicated in the figure legend in the *Results section.*

5. RESULTS

5.1 Mass spectrometry profiling of histone PTMs in breast cancer clinical samples reveals an epigenetic pattern in Triple Negative breast cancer

The preliminary data obtained by epigenetic profiling of histone PTMs in different breast cancer subtypes displayed significant changes in the levels of some histone PTMs in the TNBC subtype.

For the investigation of upstream and downstream mechanisms underlying the deposition of these PTMs, we selected a new set of fresh-frozen TNBC and Luminal A patient-derived tissues available at the Biobank for Translational Medicine Unit of the European Institute of Oncology. A total of 32 samples were selected and analyzed, including 13 Luminal A samples and 19 TNBC samples. All the samples were processed with the low input protocol adapted for fresh frozen tissues (*see Materials and methods section 4.3*) (155) and histones were proteolytically digested with the in-gel PRO-PIC protocol, with the addition of the histone-focused super-SILAC mix (112). LC-MS/MS analysis was performed and MS raw data were processed using the EpiProfile 2.0 software (PRO-PIC version) for the accurate quantification of histone marks (111), in addition, some data were manually validated with the Thermo Xcalibur QualBrowser version 4.0.27.13 (Thermo Fisher Scientific).

The heatmap-display shown in **Figure 13** illustrates the different modified peptides for histone H3 and histone H4 quantified as log2 L/H ratios of percentage relative abundances (%RA) of each modified peptide in the TN and Luminal A samples. By MS analysis, it was possible to quantify most of the modified peptides of histone H3 and histone H4. Even though in some cases it was not possible to quantify all the peptides in all the samples, generally due to their very low abundance or instrument sensitivity, overall, the analysis showed a clear different pattern between Luminal A and TN samples. In addition, a PCA (**Figure 14A**) performed on peptides belonging to both histone H3 and histone H4 indicated a clear distinction between TN and Luminal A samples, underlining a difference in the epigenetic profiles between these two subtypes, in line with what we observed in the preliminary data. The PCA also clearly showed a higher heterogeneity of TN tumors compared to Luminal A samples. Moreover, upon comparative analysis of individual modifications in the two subtypes, we identified several significant changes in the histone PTM levels (**Figure 14B**).

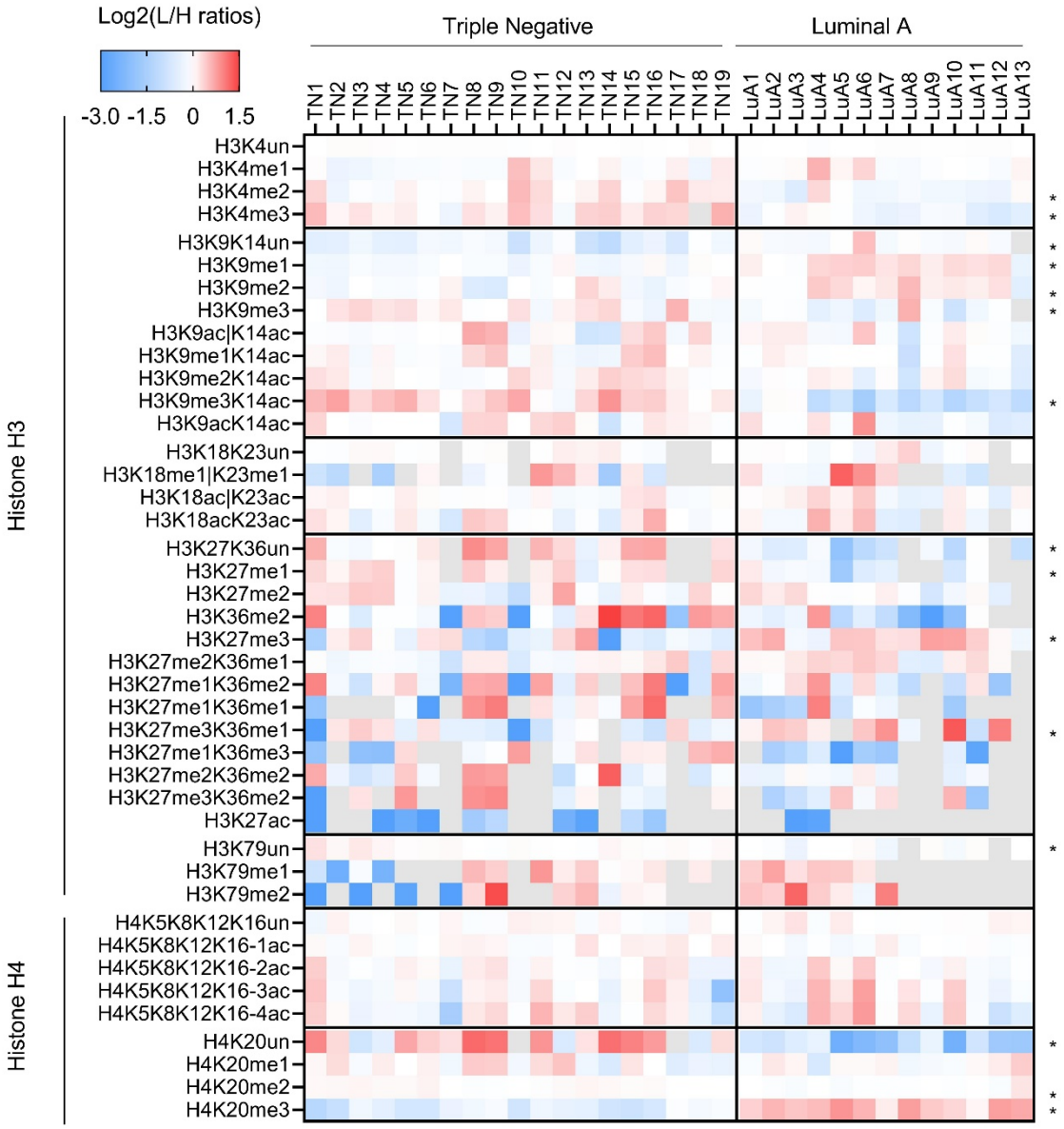


Figure 13: Profiling of histone modifications in breast cancer Triple Negative and Luminal A fresh frozen samples

*MS-based profiling of histone PTMs in Triple Negative (TN) and Luminal A (LuA) samples. Heatmap display of Log2 (L/H ratios) (L: breast cancer biopsies, and H: super-SILAC internal standard) showing differentially modified histone peptides. L/H relative abundance ratios were Log2 transformed and normalized over the average value across the samples. Red and blue color indicate respectively increased and decreased levels of each modified peptide. Grey: not quantified. * indicates significant changes observed.*

The changes identified (**Figure 15**) clearly reproduce the epigenetic signature observed in the preliminary results, with a loss of H4K20me3 in TN samples, paralleled by an increase of H4K20un being the most significant change. This is not surprising, as this histone mark has been defined to be a hallmark of cancer and to correlate with poorer prognosis in breast cancer samples. In particular, lower levels of H4K20me3 have been found by IHC in breast

cancer samples compared with normal tissues and correlate with decreased disease-free survival (180).

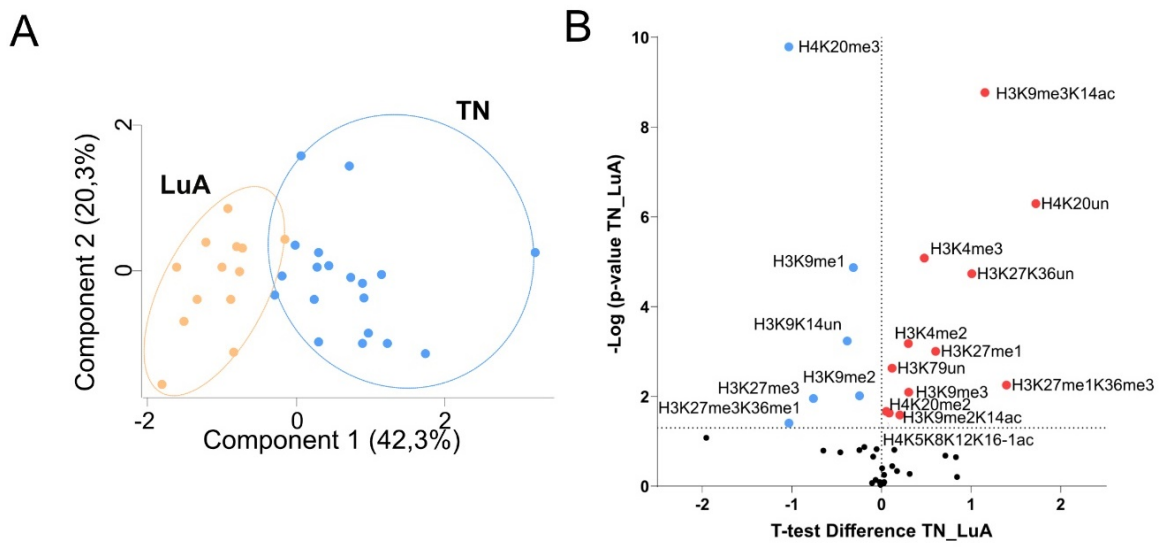


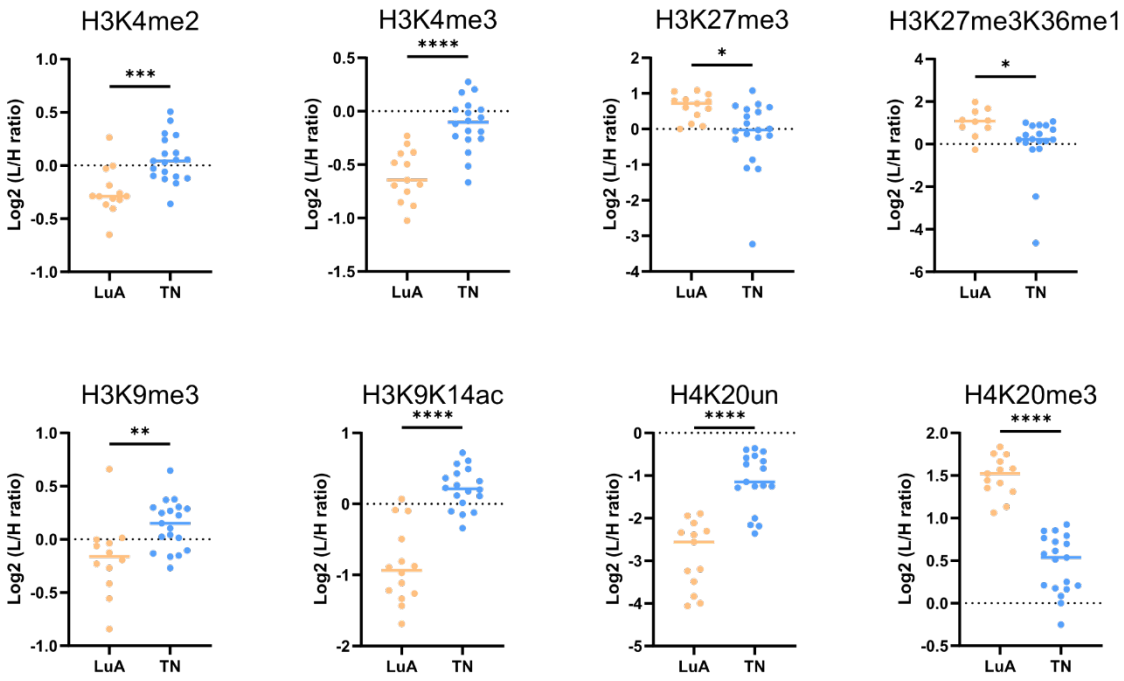
Figure 14: Comparative analysis of histone PTM profiles, from quantitative MS-based epi-proteomic analysis of TNBC and Luminal A samples

(A) PCA analysis of modified peptides identified and quantified for histone H3 and histone H4 in Luminal A (LuA) and Triple Negative (TN) samples. (B) Volcano-plot displaying histone modifications whose level is significantly increased (red) or decreased (blue) in Triple Negative, compared to Luminal A tumors with p-value <0.05 calculated by Student t-test.

Additional changes observed by the quantification by MS included an increase in the level of the repressive mark H3K9me3 both alone or in combination with K14ac in the TNBC subtype. This result was not only observed in this dataset and preliminary results, but also in previous studies carried out by our group where higher H3K9me3 levels were observed in the TN subtype and in other breast tumors compared to normal tissues (55). Loss of H3K27me3 alone, or in combination with K36me1, was also detected. In particular, loss of H3K27me3 has been previously reported to be a predictor of poor outcome not only in breast cancer, but also in other cancer types, including ovarian and pancreatic cancer (57). Moreover, the loss of this modification might result in expression of oncogenes and its involvement in chemo-resistance in TNBC has been described (152). However, a decrease in H3K27me3 levels in TNBC might depend on the high proliferation rates which characterize TNBC tumors, as previous studies have attributed changes of H3K27me3 at least partially to different proliferation rates (55). Moreover, thanks to the improvement of our MS methods by optimization of the digestion and derivatization steps (*see Material and Methods paragraph 4.4*), it was also possible to quantify, together with an increase of

H3K4me2 levels, an increase of H3K4me3 in the TN subtype versus the Luminal A one. This piece of data is very interesting, in light of the fact that these two modifications are functionally linked to active transcription and might be involved in expression of oncogenes and activation of relevant pathways in TNBC. Hence, our result suggests that the increase in H3K4 methylation levels might correlate with the TNBC phenotype, with H3K4me2 and H3K4me3 working as potential prognostic biomarkers for TNBC tumors.

A



B

Histone PTMs	TN	LuA
H3K4me2	↑	↓
H3K4me3	↑	↓
H3K9me3	↑	↓
H3K9me3K14ac	↑	↓
H3K27me3	↓	↑
H3K27me3K36me1	↓	↑
H4K20un	↑	↓
H4K20me3	↓	↑

Figure 15: TNBC-specific epigenetic signature emerging from the epi-proteomic analysis

Changes of specific histone marks are shown with significant differences in Triple Negative (TN) compared to Luminal A (LuA) subtype and shown as Log2 transformed L/H ratios. Bars show p-value (<0.05, **<0.01, ***<0.001, ****<0.0001) calculated by t-test. (B) Summary of results as in A. The red arrow indicates an increase; the blue arrow indicates a decrease.*

5.2 Quantitative MS-based analysis of histone H1 variants

Histone H1 represent the most divergent family of histones with a large number of positively charged residues in the globular domain which are important for nucleosome binding and which are conserved among species and individual variants (158). Instead, the N- and the C- terminal tails are less conserved among histone H1 variants and it is therefore possible to distinguish them by MS. In our lab we have developed and validated a MS-based method which allows quantification of histone H1 variants, based on the use of a rather standard in-gel digestion protocol of proteins by trypsin, followed by LC-MS/MS (158). More in detail, for the quantification of histone H1 variants, we used a label-free quantification (LFQ) method known as MaxLFQ (181). This approach is based on the extraction of ion intensities and is present into the freely accessible MaxQuant computational proteomic platform (159). Unlike labeling techniques, label-free quantification methods are straight forward and suitable for the analysis of any type of samples, including those obtained from patients. Therefore, for a more explorative analysis, in addition to the analysis of histone PTMs, we also performed mass spectrometry-based analysis of histone H1 variant levels in some of the clinical samples available (11 Luminal A and 18 TN). A total of 20 µg of nuclear extracts were loaded on a 4-12% PAA gel and a large band (20- 45 kDa) of around the size of histone H1 was excised for in-gel digestion with trypsin for each sample as illustrated in **Figure 16A**. The in-gel digestion was performed in order to enrich histone H1 in comparison to other proteins, at the same time allowing maintaining a sufficient number of background proteins for proper normalization of protein abundance in the sample. By looking at the intensity-based absolute quantification (iBAQ) score in comparison with the other proteins present in the band cut from the gel, which is indicative of the most abundant proteins identified in a sample, we detected as expected histone H1 variants among the most abundant proteins (**Figure 16B**). To compare histone H1 variant levels across samples, we used the LFQ values, which did not highlight any substantial changes other than an increase in histone H1.5 and H1.1 levels in TN compared with Luminal A samples **(Figure 16C-D**). While H1.0 variant is cell cycle-independent and is highly expressed in normal differentiated cells, H1.1-H1.5 are cell cycle-dependent and might be altered in tumor cells (182). Therefore, the higher levels of H1.1 and H1.5 observed in TN tumors compared to Luminal A could likely be the results of the higher proliferation rates typical of this tumor subtype, compared to the less aggressive Luminal A subtype.

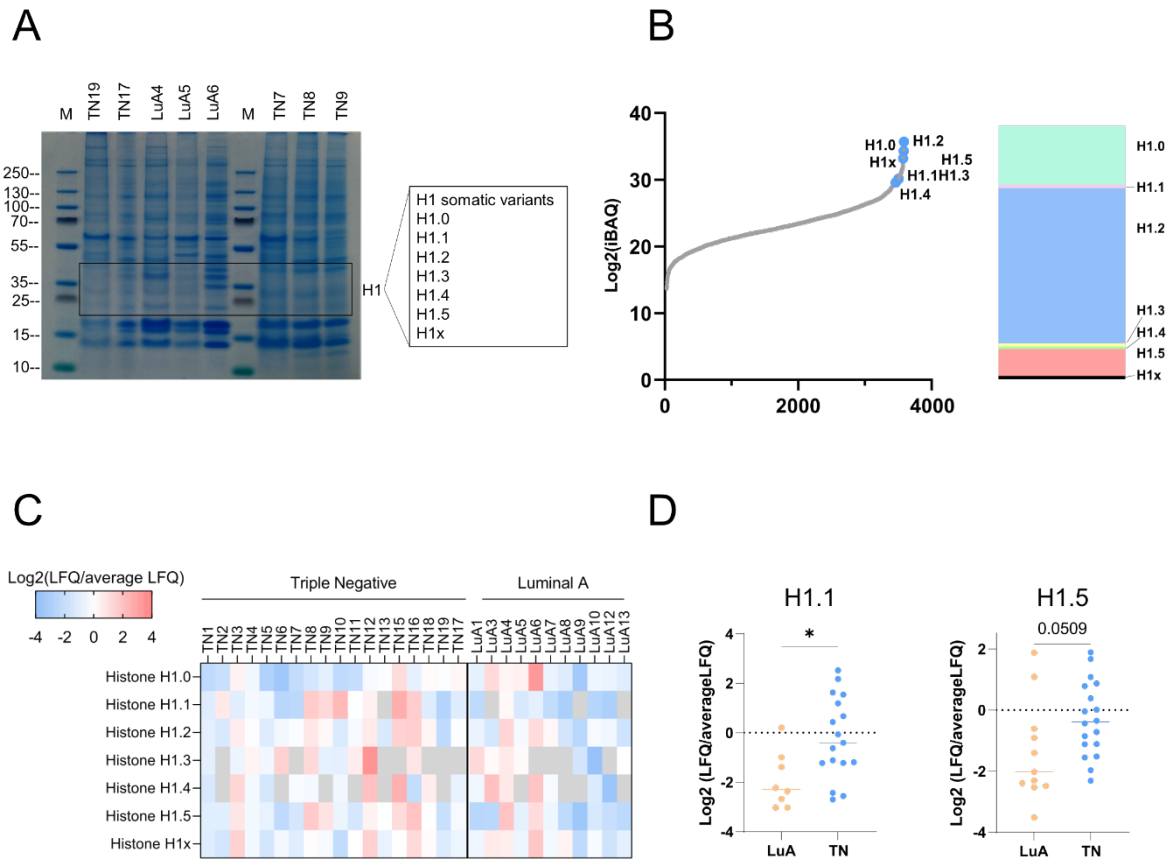


Figure 16: Analysis of histone H1 variants in BC and comparison between BC subtypes, by LC-MS/MS label free quantification

*(A) 4-12% PAA gel loaded with 20 µg of nuclear extract. The box indicates the band cut for each sample containing histone H1 variants. (B)Proteins ranked by intensity-based absolute quantification (iBAQ) scores with histone H1 variants highlighted and most abundant histone H1 variants displayed in the histogram as percentage. (C) Heatmap display of LFQ values normalized on LFQ average across samples. (D)Histone variants significantly increased in TN compared to Luminal A samples calculated by t-test, * p-value <0.05.*

5.3 Investigation of histone PTM changes in different phases of cell cycle

There is vast literature about changes of bulk histone PTM levels during cell cycle (183,184); since TNBC is characterized by higher proliferation rates compared to the Luminal subtype, one possibility we wanted to explore was that some of the observed changes in the histone PTM levels between Luminal A and TN samples could be due to different cell proliferation rates. A previous study carried out in our group (55) showed a negative correlation of H3K27me3 levels and the Ki67 proliferation index. In addition, a H3K27me3 decrease was observed in cells blocked in G2/M phase. Thanks to the improvement of our MS methods, in order to validate our previous results and further verify correlation of increased levels of histone PTMs at the G2/M phase in comparison to asynchronous cells, we used two strategies:

1) the treatment of breast cancer cells with nocodazole to synchronize them in the G2/M phase of the cell cycle;
2) the isolation of cells in G2/M phase by sorting, by exploiting the FUCCI (Fluorescent Ubiquitination-based Cell Cycle Indicator) system, an innovative lentiviral system that is characterized by a set of fluorescent probes, which enable the visualization of cell cycle progression in living cells and the selection of cells in a specific phase of cell cycle (185).

We selected two breast cancer cell lines, MDA-MB-231 as TN and MCF7 as Luminal A for the synchronization by nocodazole and enrichment of cells blocked in G2/M phase. Nocodazole is commonly used to block cells in G2/M phase as it interferes with microtubule polymerization and de-polymerization dynamics, preventing the proper formation of the mitotic spindle. Upon cell synchronization, the percentage of cells in each phase of the cell cycle was analyzed by propidium iodide staining by FACS and by comparison with an asynchronous control. Asynchronous cells displayed a normal cell cycle distribution, with the majority of the cells in G1 phase. Instead, upon block with nocodazole, the majority of MDA-MB-231 and MCF7 cells were synchronized in G2/M phase (**Figure 17**).

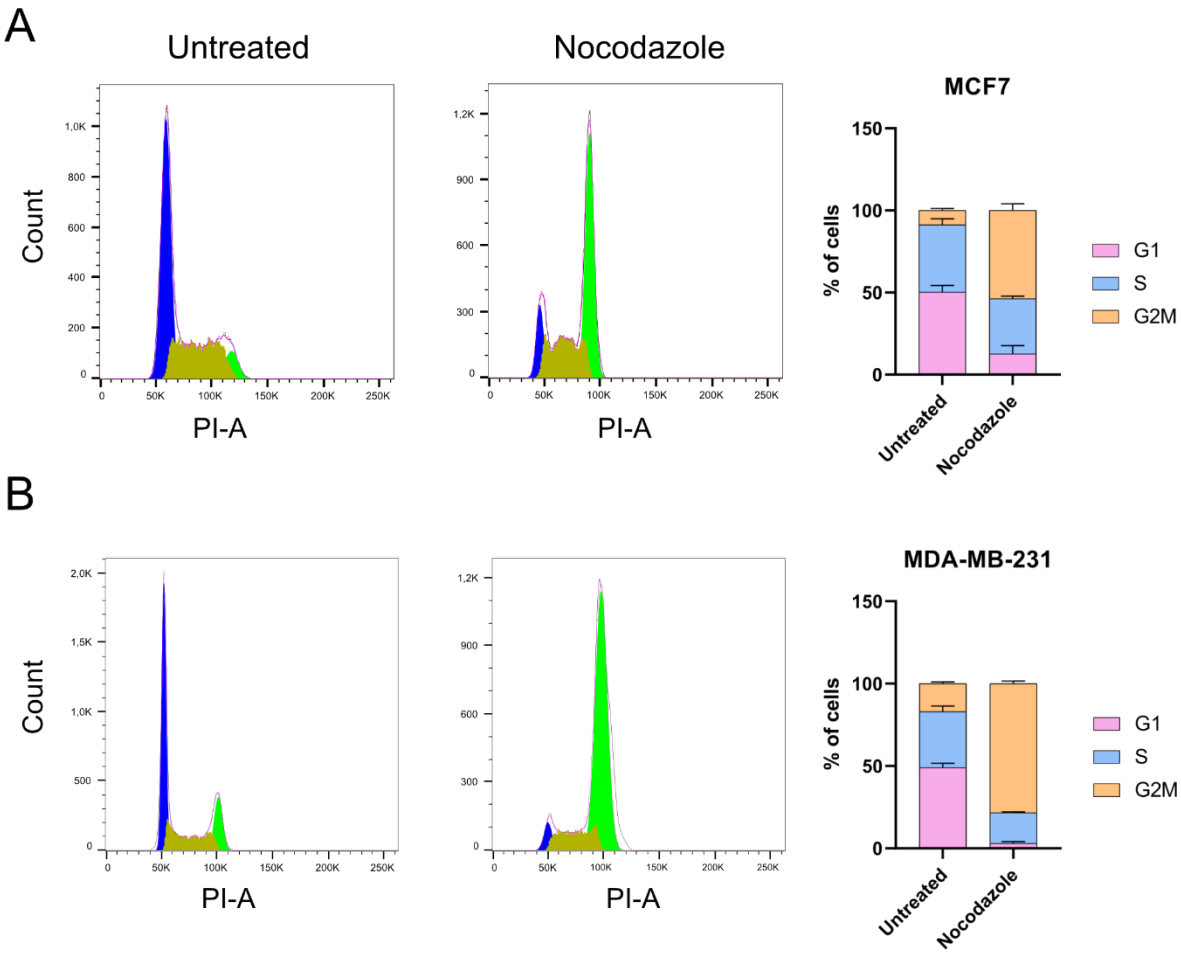


Figure 17: Cell cycle analysis of synchronized cells

(A) MCF7 cell line untreated (asynchronous cells), and blocked in G2/M phase by inhibition with nocodazole. (B) MDA-MB-231 display of cell cycle of untreated cells and cells blocked in G2/M by nocodazole. G1 phase is indicated in blue, S phase is indicated in yellow, G2/M is indicated in green. The histogram displays the different cell cycle phases in percentage analyzed by FlowJo. n=3.

As an additional analysis on the histone PTMs dependence on cell-cycle phases, we decided to profile histone marks across cell cycle progression in the MDA-MB-231FUCCI cell line which was kindly given by the group of prof. Saverio Minucci at IEO. The FUCCI system is based on the use of two different fluorescent proteins, a red fluorescent protein, mCherry that is linked to the protein Cdt1, which is abundant in G1 phase cells, and a green fluorescent protein, mVenus fused to Geminin, a protein that is absent in G1 and accumulates through S and G2/M phases. By profiling the color of the fluorescence emitted by individual cells, it is therefore possible to distinguish whether a cell is in G1 (red), S (green), or G2/M (both red and green) phases and then sort them (**Figure 18**) based on expression of mCherry and mVenus, respectively. Upon MS-based analysis of histones extracted from differently sorted cells, it was possible to carry out a more specific analysis of histone PTMs by MS at G2/M phase of the cell cycle.

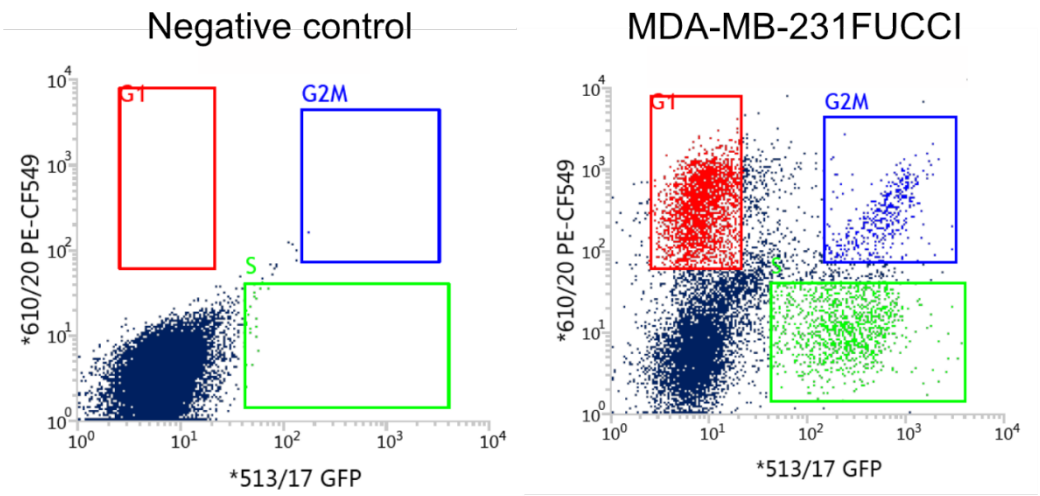


Figure 18: Sorting of MDA-MB-231FUCCI cycling cells

Dot-plots display the data for negative control (wild type cell line) and MDA-MB-231FUCCI in different phases of cell cycle, based on fluorescence analysis of specific markers. Cells in G1 are mCherry-positive and are highlighted in red, cells in S phase are mVenus-positive and are highlighted in green. G2/M cells are mCherry-positive and mVenus-positive and are highlighted in blue.

After confirming the block in G2/M phase for the cells synchronized with nocodazole, and subsequent sorting of MDA-MB-231 FUCCI cells, we performed MS analysis of histone level by comparison of the L/H ratios of histone PTMs or combinations of PTMs quantified in the cells synchronized in G2/M phase and those of asynchronous cells. In nocodazole-treated cells, overall, we identified similar changes in histone PTMs levels in G2/M phase in both MDA-MB-231 and MCF7 cell lines (**Figure 19**). We identified significant changes in methylation and acetylation levels occurring at H3K9 and K14 residues: more specifically, H3K9me1 and H3K9me2 were increased while the same modifications in combination with K14ac were significantly decreased. The peptide bearing H3K18K23un was increased in G2/M phase, opposite to decreased levels of H3K18acK23ac. No particular changes were observed at the H3K36 and H3K27 residues, where we only identified a decrease of H3K27K36un. In addition, acetylated lysine of histone H4, K5, K8, and K12 were overall decreased in G2/M phase, while the unmodified form of these residues was increased. Lower levels of H4K20un and H4K20me1 were also identified in G2/M phase compared to asynchronous cells.

The analysis of histone PTM changes in MDA-MB-231FUCCI cells in G2/M phase revealed a pattern very similar to that obtained by nocodazole block. For instance, the increase in H3K9me1 and H3K9me2 was comparable to the synchronized cells, suggesting that these modifications are cell cycle-dependent, as well as the levels of K18K23unmodified which

were increased and paralleled by a decrease in acetylation. In addition, the increased levels of unmodified K4, K5, K8, and K12 of histone H4 paralleled by decreased acetylation were also confirmed, as well as the decreased levels of H4K20un. From this analysis it was also possible to detect K27 and K36 methylation and acetylation levels, in particular lower levels of H3K27me3 were identified in G2/M phase, in line with our previous study (55).

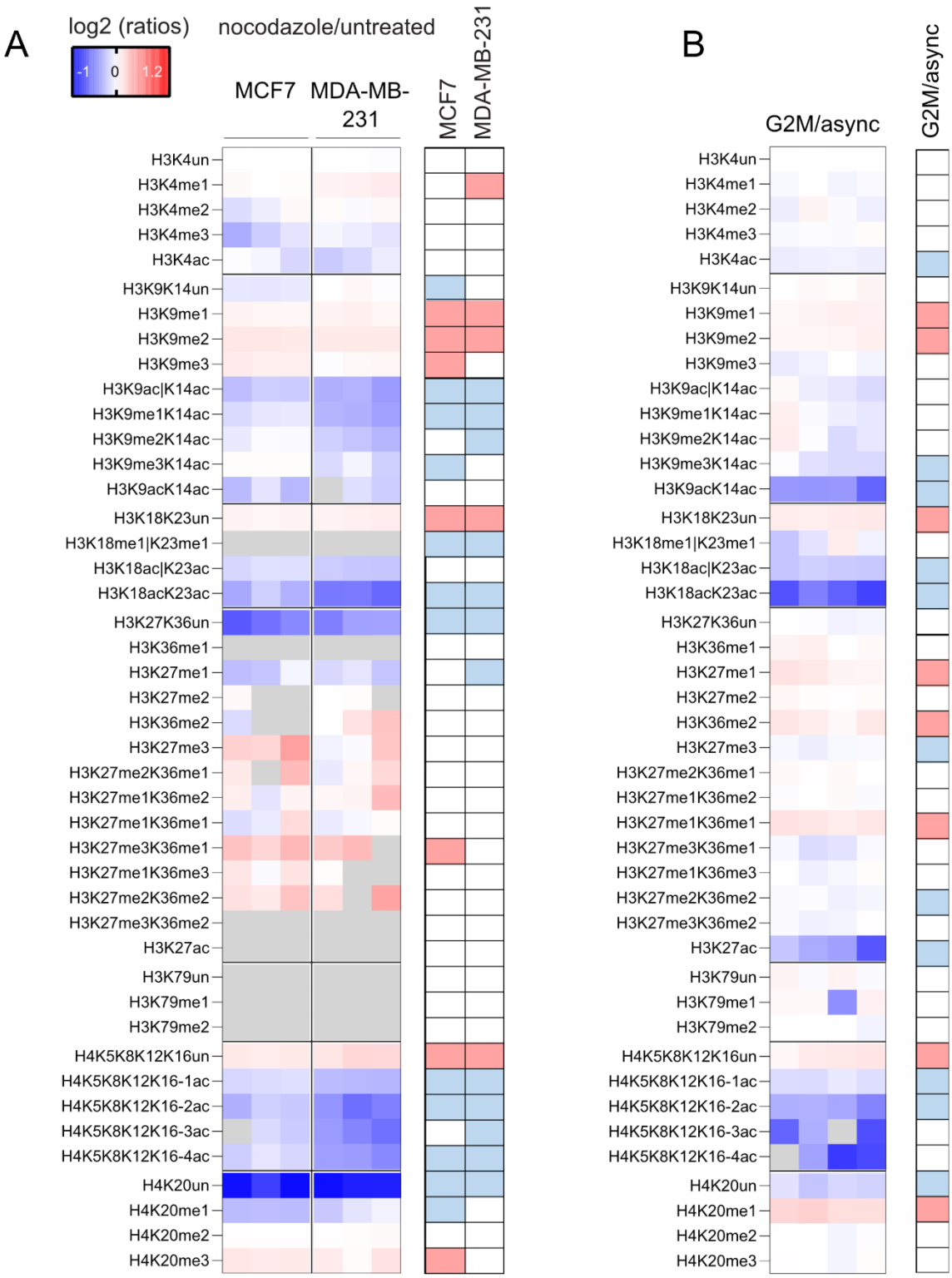


Figure 19 (previous page): Profiling of histone PTMs in breast cancer cells in G2/M phase
(A)Heatmap display of the log2-transformed ratios for specific histone H3 and histone H4 PTMs in MDA-MB-231 and MCF7 breast cancer cell lines. The L/H ratio values for nocodazole-treated cells (synchronized in G2/M phase) were normalized on the L/H ratio values obtained for untreated cells (asynchronous) n=3. (B) Heatmap display as in A for MDA-MB-231FUCCI cells sorted in G2/M phase. The L/H ratio values for G2/M were normalized on L/H ratio values quantified for asynchronous cells, n=4. t-test was performed to compare modified peptides in nocodazole-treated and asynchronous cells and G2/M-sorted cells with asynchronous MDA-MB-231FUCCI. The red color indicates a significant increase of PTMs in G2/M phase, while the blue color indicates a decrease, p-value < 0.05. The grey color indicates peptides that were not quantified.

Taken together, most of the histone PTMs that we have identified in the TNBC epigenetic signature, including H3K4me2/me3, H3K9me3, H4K20un, and H4K20me3, seem to be independent on the specific cell cycle phase, ruling out the possibility that the observed difference among subtype could be explained merely by differences in proliferation rates. More in detail, we found no specific changes for H3K4me2 and H3K4me3 levels at G2/M phase in both synchronized cells and MDA-MB-231FUCCI. In addition, levels of H3K9me3 alone or in combination with K14ac were also not changing and reduced in G2/M phase respectively in both conditions suggesting that the increase observed in TNBC tumors is not cell cycle-dependent. Levels of H4K20un were also significantly decreased in G2/M phase. On the contrary, even though the levels of H3K27me3 were not changing in nocodazole-treated cells, the levels of H3K27me3 in G2/M phase of MDA-MB-231FUCCI cells suggest that this modification is linked with G2/M phase and the lower levels observed in TNBC clinical samples might depend from the higher proliferation rates of this tumor.

5.4 A multi-OMICs analysis to investigate mechanisms downstream of the H3K4me2 and H3K4me3 increase in the Triple Negative subtype

Based on the results obtained from the epigenetic profiling of breast tumor samples, we set to investigate the molecular mechanisms linked to the deposition of H3K4me2 and H3K4me3 in BC. The selection of these two modifications was based on the following considerations:

1) The extent of change in the TNBC subtype compared to Luminal A, as well as in comparison with the other subtypes (as shown in the preliminary data **Figure 12B**). In the epigenetic changes observed, H3K4me2 and H3K4me3 show a significant increase in TNBC tumors, not only in comparison to Luminal A samples but also when compared to all the other subtypes.

2) The possibility of correlating the genomic distribution of modifications with activation or repression of gene expression. Previous research has established a direct association between H3K4me2 and H3K4me3 and positive transcription regulation, as these PTMs generally localize at transcription start sites of the regulated genes. Additionally, it has been demonstrated that H3K4me2 also localizes at enhancers (17) .

3) The independence from cell proliferation, based on our PTM-analysis across the cell cycle and previous literature (55). As a matter of fact, our MS based indicated no correlation between H3K4me2 and H3K4me3 and cell cycle, with no changes observed in G2/M phase of MDA-MB-231FUCCI. In addition, in both MDA-MB-231 and MCF7 cells synchronized in G2/M phase, both H3K4me2 and H3K4me3 were unchanged.

To study the downstream effects of H3K4me2 and H3K4me3 deposition and the impact on gene expression, we set up an experimental strategy (**Figure 20**) that combines our epi-proteomic data (which provides indication about bulk levels of the histone PTM of interest) with other epi-genomic approaches: a) ChIP-seq, which provide information on the locus-specific distribution of the histone marks of interest at genome wide level; b) transcriptomics using both experimental RNA-seq data from a subset of in house matched clinical samples and public repository (TCGA), to correlate presence/abundance of the mark with the expression of the associated genes. More in detail, from the primary samples analyzed by

MS, we selected 7 Luminal A and 8 TNBC fresh frozen specimens based on the most divergent levels of H3K4me2 and H3K4me3. Importantly, to minimize contamination from non-tumoral cells, we selected tissues from patients with a tumor cellularity of 60% or higher and with the higher levels of H3K4me2 and H3K4me3 identified by MS.

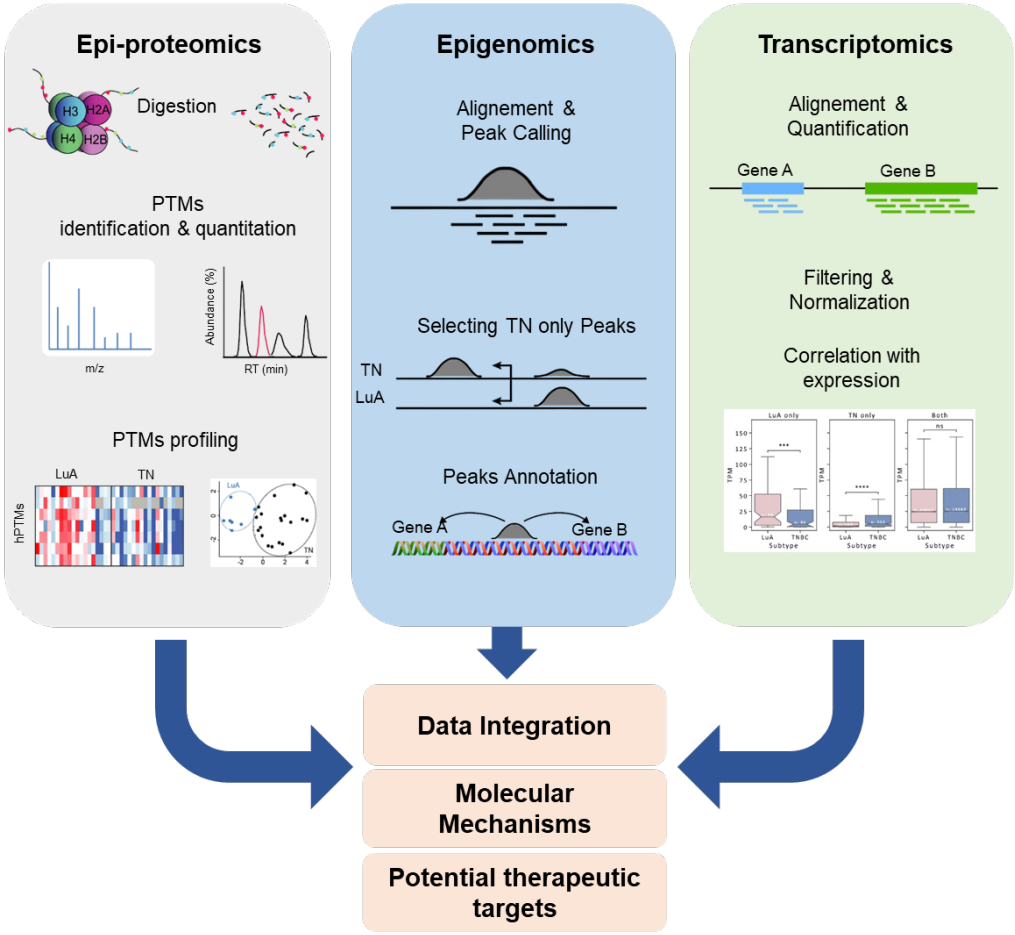


Figure 20: Multi-OMICs analysis workflow

Outline of the experimental strategy used for the multi-OMICs approach. The integrative analysis was performed starting from MS-profiling of histone PTMs, followed by integration of ChIP-seq and RNA-seq data with the purpose to analyze the molecular mechanisms involved and identify potential therapeutic targets.

5.4.1 Optimization of chromatin extraction from clinical samples for ChIP-seq analysis

In order to define the localization and the distribution of both H3K4me2 and H3K4me3 at genome wide and locus specific level, we carried out ChIP-seq experiments of H3K4me2 and H3K4me3.

First, we needed to optimize the chromatin extraction step, which was optimized for clinical samples (**Figure 21**). After crosslinking and chromatin extraction, a sonication step was performed to shear the DNA in order to obtain fragment of one or two nucleosomes as input for immune-affinity enrichment. Then, an immunoprecipitation with a ChIP-grade specific anti-H3K4me2 or H3K4me3 antibodies was performed overnight in parallel with an immunoprecipitation performed with IgGs, as a negative control. The efficiency of the immunoprecipitation was assessed by Western Blot (WB) against the histone PTM of interest and total histone H3, as normalizer and/or RT-qPCR for target regions for both H3K4me2 and H3K4me3. After de-cross-linking (*see Material and Methods paragraph 4.9*), input 2 %, IP, IgG, 2% flow through (FT) of the IP, and 2% FT of the IgG were loaded on a 4-12% PAA gel.

The WB in **Figure 21** for H3K4me2 (left panel) and for H3K4me3 (right panel) clearly show the enrichment of the modification in the IP by comparison with the input and the flow-through, where a depletion of the modification was also observed. After DNA purification, RT-qPCR was performed for target genes that harbor H3K4 methylation at promoter regions, namely the housekeeping genes GAPDH and β-ACTIN. Two negative controls were also included in the analysis, an intergenic (genedesert) and an intronic region (ITGB4). However, because ChIP from clinical samples is quite challenging and of poorer quality compared to cell lines, in RT-qPCR data some background can be observed, as denoted by a low enrichment for the modification also in the negative controls. Finally, samples were processed for library preparation and sequenced by next generation sequencing (NGS) on the Illumina Novaseq 6000 platform with a sequencing depth of 50 million reads, and raw data were analyzed by the bioinformatician of our group Alessandro Vai, (*see Material and Methods paragraph 4.10*).

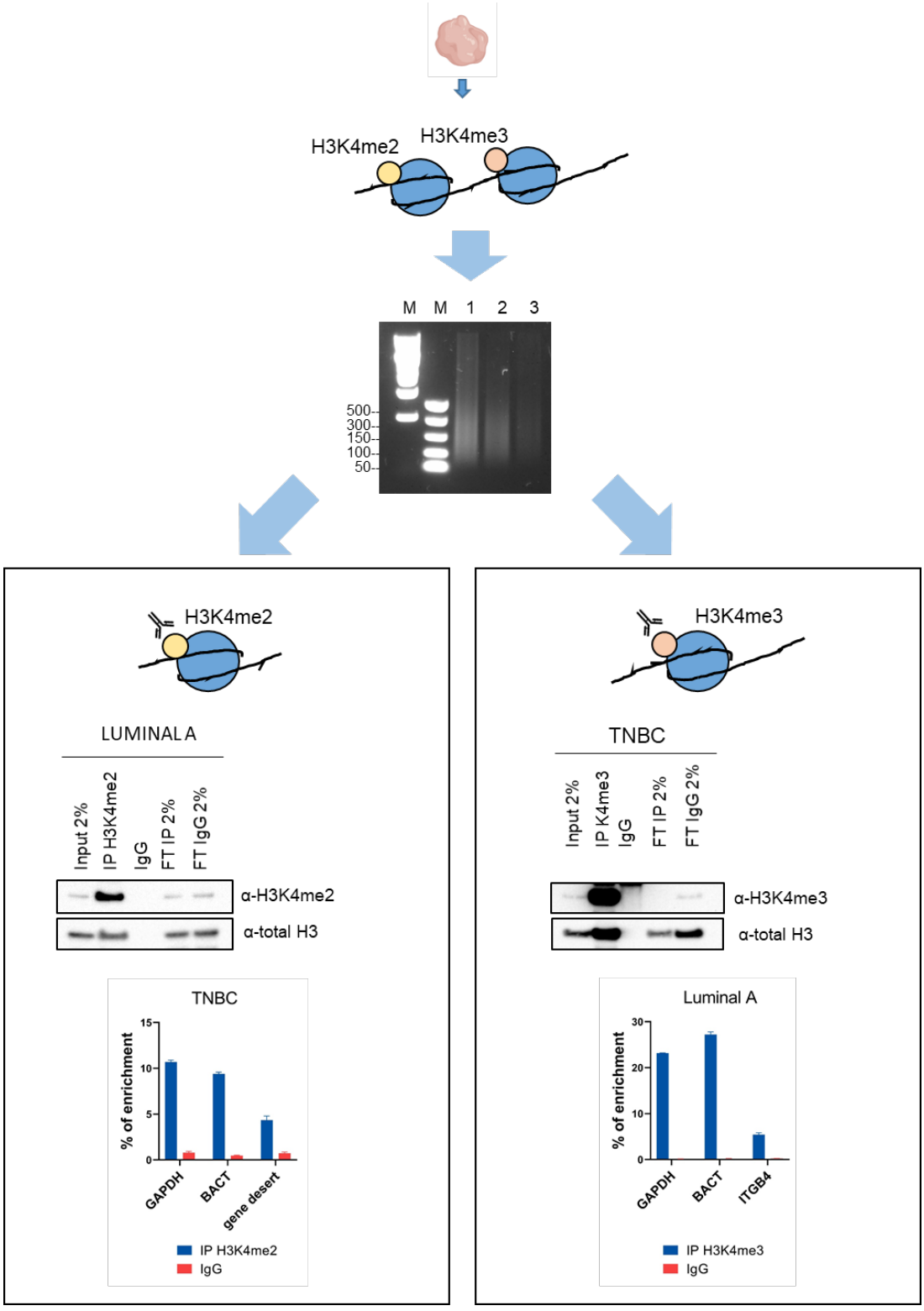


Figure 21 : Experimental workflow of Chromatin Immunoprecipitation

The workflow display shows sonication of chromatin on a 1% agarose gel with a sonication range between 100 and 300 bp for 3 different samples. Immunoprecipitation efficiency is displayed for both H3K4me2 (left panel) and H3K4me3 (right panel) for some explicative samples analyzed by Western Blot and ChIP-qPCR (BACT and GAPDH are shown as positive controls, ITGB4 and gene desert are shown as negative controls).

5.4.2 ChIP-seq analysis identified a unique distribution of H3K4me2 peaks in Triple Negative tumors

We employed H3K4me2 and H3K4me3 ChIP-seq data to define their global genomic distribution in TN and Luminal A samples. H3K4me2 was analyzed in 7 Luminal A samples and 8 TN samples, while H3K4me3 was analyzed in 3 Luminal A samples and 5 TN samples. First, a bioinformatic analysis was performed to define the distribution of peaks (**Figure 22**). For H3K4me2, we found a more heterogeneous distribution in Luminal A samples compared to TN. In both subtypes, 30% to 60% of peaks were localized at promoter regions. In addition, a good portion of H3K4me2 peaks was present at the level of gene bodies and intergenic regions. This was not the case for H3K4me3 distribution, whose localization was clearly present at promoters, with 60% to 90% of peaks in both subtypes, in agreement with the known presence of this modification around the TSS.

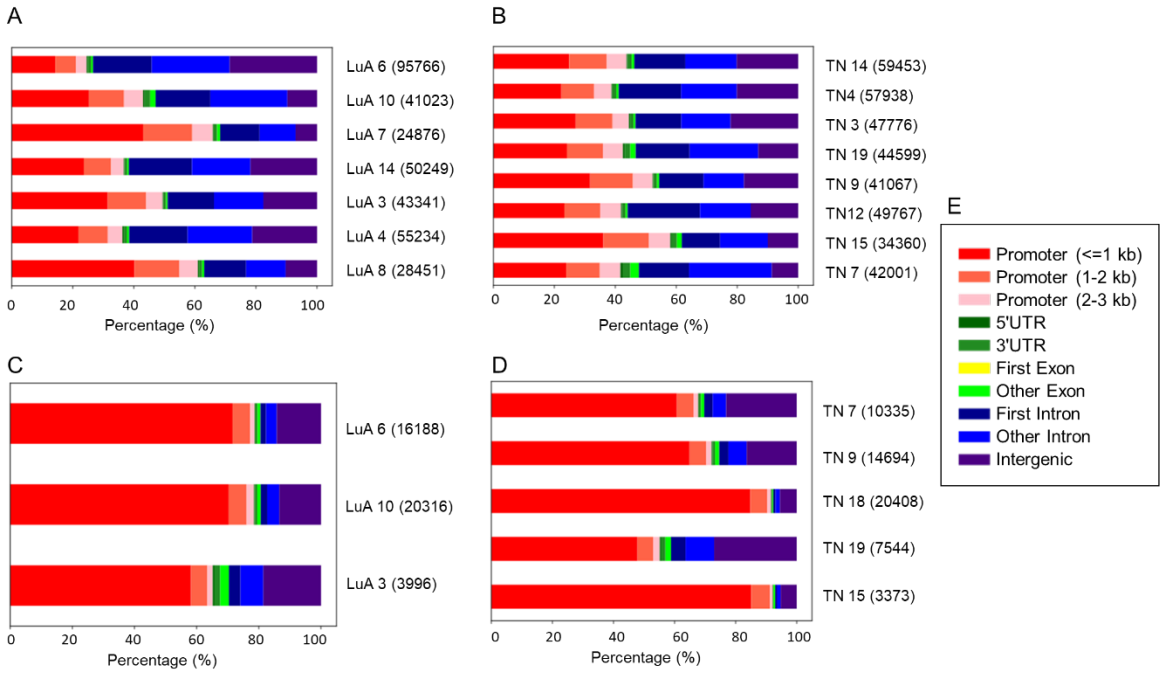


Figure 22: ChIP-seq analysis of H3K4me2 and H3K4me3 distribution in Luminal A and Triple Negative samples

(A) H3K4me2 peak distribution in Luminal A and (B)TN samples with indicated total number of peaks identified for each sample. (C)H3K4me3 peak distribution in Luminal A samples and (D) in TN samples. (E) Legend displaying different colors according to specific genomic regions.

Even though the study of H3K4me3 is more adaptable to a correlation with activation of expression due to a much more extensive literature linking this mark to active transcription of the associated gene, we decided to focus our mechanistic investigation on H3K4me2 for several reasons: first, the novelty, since there is paucity of studies centered on the assessment

of H3K4me2 role in transcription due to a more spread distribution in the genome compared to H3K4me3. Secondly, the discriminative potential of this mark; as a matter of fact, a Principal Coordinates Analysis (PCoA) of H3K4me2 read density (a multivariate statistical technique used to analyze and visualize similarities or dissimilarities among samples based on a distance or dissimilarity matrix) showed a clear separation between Luminal A and TN samples when the modification localizes at promoters (**Figure 23A**) and/or enhancers (**Figure 23B**). Instead, when carrying out the same type of analysis for H3K4me3 at promoter regions, no separation was observed and the two subtypes clustered together (**Figure 23C**). Consequently, H3K4me2 seems to emerge as a more efficient discriminatory epigenetic mark between these two subtypes.

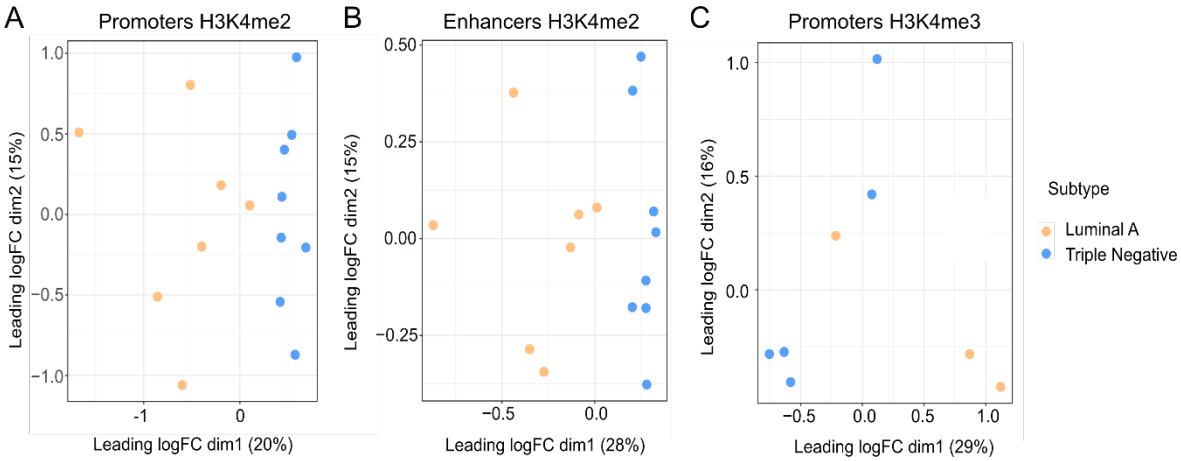


Figure 23: Principal Coordinates Analysis (PCoA)

(A) PCoA analysis of read density at promoter regions for H3K4me2 ChIP-seq data (B) PCoA analysis of read density at enhancer regions for H3K4me2 ChIP-seq data. (C) PCoA analysis of read density at promoter regions for H3K4me3 ChIP-seq data.

Next, in order to identify specific H3K4me2 regions in the TN samples, we defined common peaks, namely the peaks present in both subtypes, and unique peaks, namely peaks identified only in TN but not in Luminal A samples, and vice versa (**Figure 24**, and *material and methods paragraph 4.10*). Combined H3K4me2 ChIP-seq data showed that around 40% of the common peaks localized at promoter regions, while the other portion localized at other genomic regions, including intronic and intergenic regions. Instead, the H3K4me2 peaks unique for either the TN or the Luminal A subtypes had a lower localization at promoter regions and were mostly localized at gene bodies and intergenic regions. Interestingly, we found that most of the TN-unique peaks overlap with genomic regions that were previously annotated as super-enhancers and which we identified by alignment with H3K27ac published ChIP-seq data (125). Super-enhancers (SE) are a group of enhancers that are

densely packed together and are associated with a high level of transcriptional activity and the expression of key cell identity genes. They are often found near genes that play important roles in cell differentiation and development and are thought to have a significant impact on cell identity and function. Moreover, dysregulation of super-enhancers has been implicated in various diseases, including cancer and in particular TNBC (125).

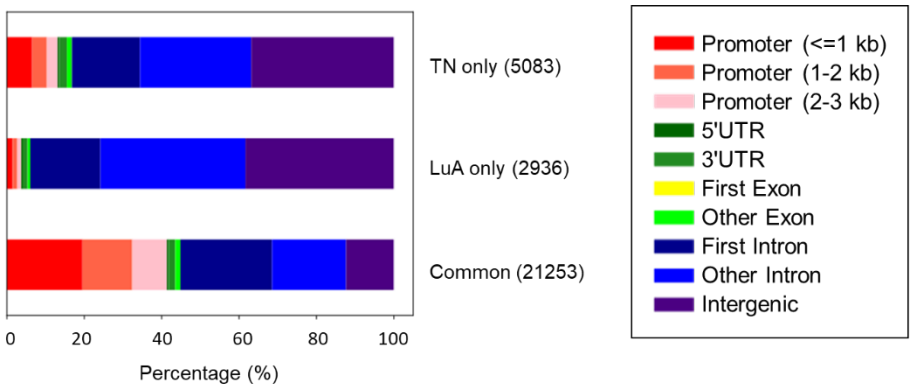


Figure 24: Combined analysis for H3K4me2 peaks in Luminal A and Triple Negative samples
Comparative analysis of common and unique peaks identified for H3K4me2 in TN and LuA samples. Common, and unique peaks for Luminal A and TN samples are displayed; association to different genomic regions indicates that the majority of unique peaks are located in intronic and intergenic regions.

5.4.3 Integration of epi-genomics and transcriptomic data indicates a correlation between H3K4me2 deposition and gene expression

To explore the potential correlation between H3K4me2 deposition at specific genomic regions and gene expression in the context of TNBC, we integrated our ChIP-seq data with RNA-seq data obtained from the same samples. RNA was extracted from 19 clinical samples (8 Luminal A and 11 TNBC). Prior to sequencing, a quality control of extracted RNA was performed using the Bioanalyzer tool (**Figure 25**). The RNA samples exhibited sufficient quality, as demonstrated by the presence of the ribosomal RNA 28S and 18S bands and we selected a set of samples for subsequent library preparation and RNA-seq analysis. However, given the relatively limited sample size, including RNA-seq data from 3 Luminal A samples and 6 TNBC samples displayed in the PCoA analysis (**Figure 25B**), and because of the heterogeneity of clinical samples, we also used RNA-seq data from a publicly available

source, the TCGA-BRCA breast cancer dataset (431 Liminal A samples and 180 TN samples).

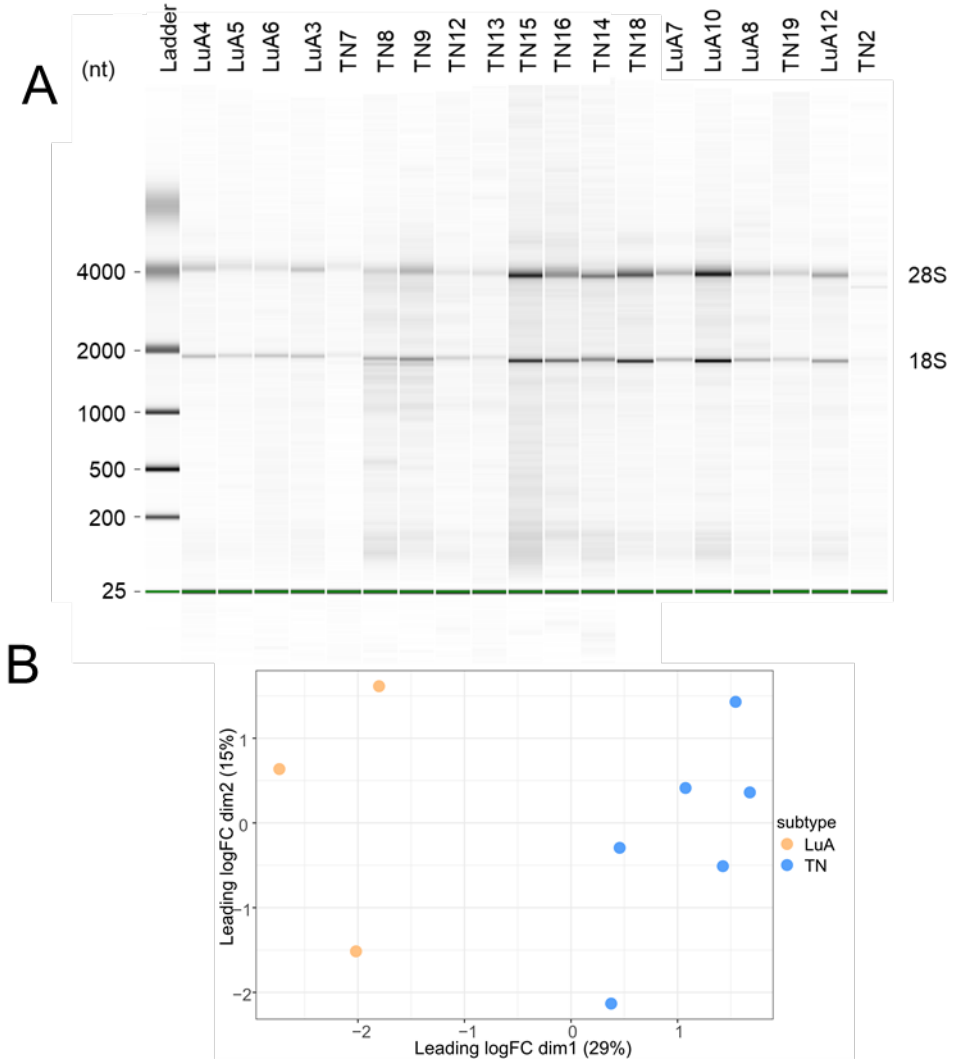


Figure 25: Bioanalyzer analysis and PCoA of RNA-seq from patient-derived samples

(A) Bioanalyzer gel show the quality assessment of total RNA extracted from different fresh frozen breast cancer samples with bands for ribosomal RNA 28S and 18S, highlighting the good quality of RNA. (B) PCoA analysis of RNA-seq data from patient samples selected based on RNA quality.

Gene expression levels were quantified as Transcripts Per Million (TPM), a normalized measure of gene expression that takes into account the length of the gene and the total number of mapped reads. The formula to calculate TPM requires normalizing the raw read counts by the gene length and the total number of mapped reads in the sample and provides an accurate representation of the relative abundance of different transcripts in a sample. The integration of matched RNA-seq data with ChIP-seq data allowed to unravel a clear correlation between H3K4me2 deposition at promoters and super-enhancer regions and the corresponding gene expression levels. A similar trend was also observed in the case of

integration with the TCGA dataset, where we also found a correlation between the presence of H3K4me2 at both promoter and super-enhancer regions and the respective gene expression level (**Figure 26**). Common H3K4me2 peaks corresponded to genes with the same expression in the two subtypes, while unique peaks correlated with higher expression in the respective subtype, both when we interrogated our matched experimental RNA-seq data and the transcriptomics data from TCGA. Interestingly, we also observed a higher increase in expression when H3K4me2 localizes at super-enhancers. This result was particularly evident in the case of matched data, while an overall trend could be inferred from the intersection of ChIP-seq with the data integrated from TCGA. The increase in gene expression in the presence of H3K4me2 is particularly marked when the modification is localized at super-enhancers, as overall higher TPM levels were observed compared to TPM levels in the promoters. This is not surprising, given that the role of super-enhancers in enhancing gene expression has been previously described (186).

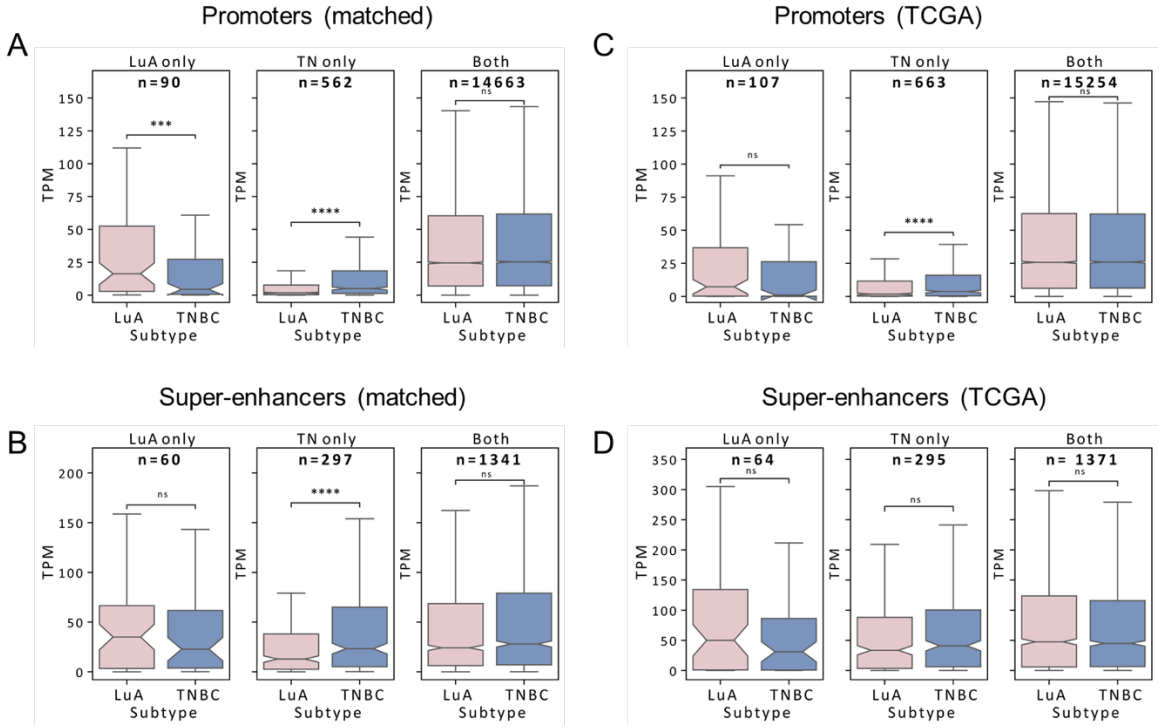


Figure 26: Integration of epi-genomic and transcriptomic data

*(A)Matched RNA-seq data for genes that harbor H3K4me2 peak in promoter regions and (B) super-enhancer regions. (C) TCGA data for genes that harbor H3K4me2 peak in promoter regions and (D) super-enhancers regions. Values were calculated as transcripts per million (TPM). ***p-value < 0.01, ****p-value < 0.0001 computed by Wilcoxon test.*

By integrating ChIP-seq and RNA-seq data, we identified a total of 809 genes with TN-unique peaks at promoter regions, of which 167 were also upregulated at the transcriptomic level in the TN subtype (**Figure 27A**). In addition, we identified 350 genes with TN-unique

peaks at super-enhancers, of which 55 genes were also upregulated at the transcriptomic level (**Figure 27B**). Interestingly, a GeneOntology (GO) analysis performed for these genes revealed the presence of GO terms associated with lymphocyte activation, immune system processes and signaling receptor activity, which could all contribute to TNBC phenotypic features (**Figure 27C-D**).

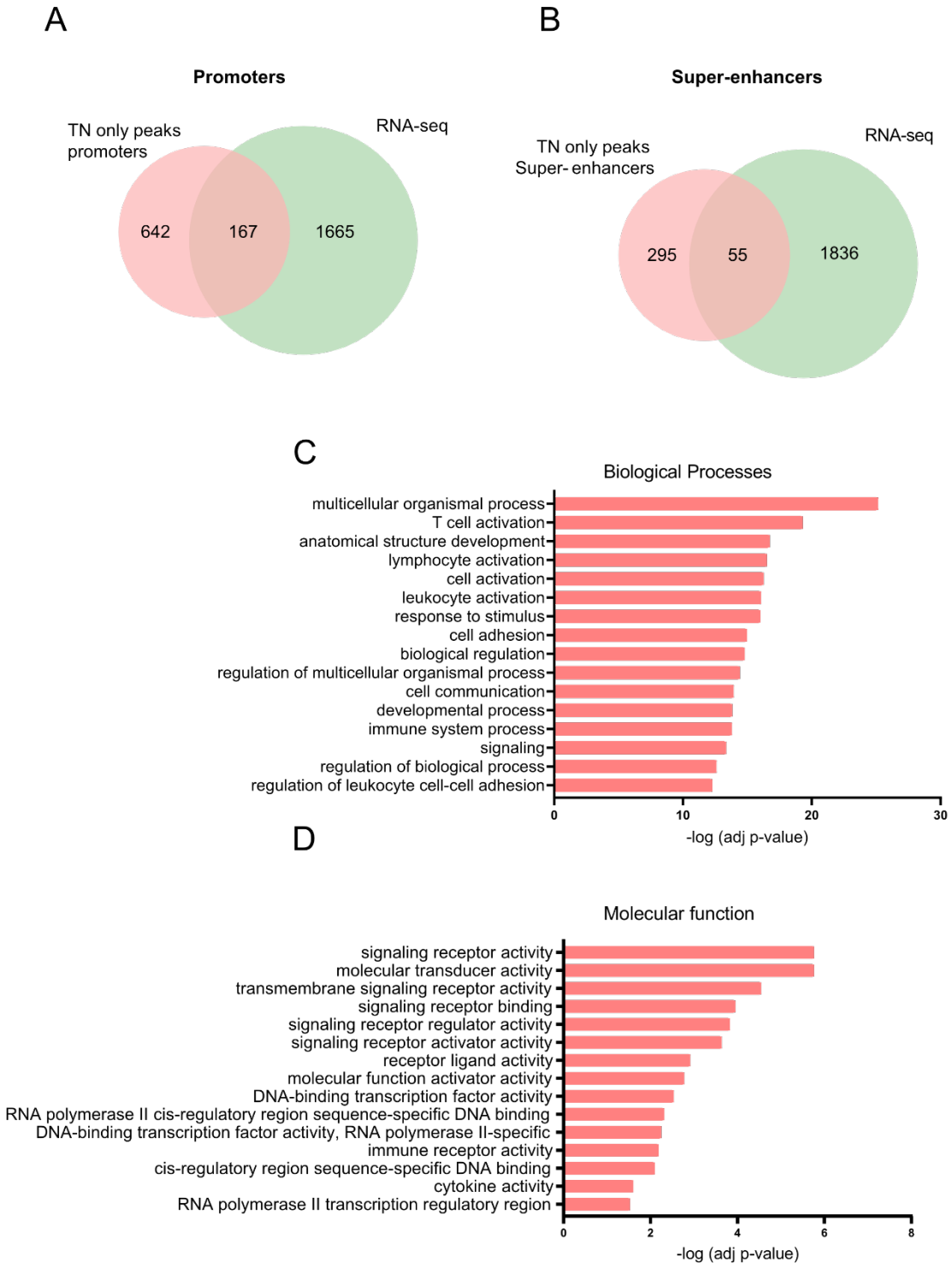


Figure 27: Selection of genes upregulated by H3K4me2 deposition

(A)Venn diagram of integrated data for identification of genes with peaks for H3K4me2 at promoters and (B)super-enhancer and upregulated at transcript level. (C) GeneOntology analysis was

performed for genes harboring H3K4me2 peaks at promoter and enhancer regions and upregulated in the TN-only subtype. The analysis shows both Biological processes and (D) molecular function performed by FDR 0.05 on g:Profiler. Most significant terms are displayed as - log of adjusted p-value.

This result is particularly intriguing in light of the fact that TNBC has previously been described to be a more immunogenic tumor compared to Luminal A. Indeed, the immune system seems to play a pivotal role in TNBC progression and response to treatment. Tumor-infiltrating lymphocytes (TILs), comprising cytotoxic T cells and natural killer cells, are integral components of the immune response against TNBC and presence of TILs is generally an indication of a better prognosis for this tumor subtype (187). TILs infiltrate the tumor microenvironment and recognize tumor-specific antigens, exerting antitumor effects. However, in TNBC, the immune system interaction with tumor cells is complex: TILs can delay tumor growth by inducing apoptosis and inhibiting angiogenesis. Conversely, the tumor microenvironment can suppress the immune response through checkpoint proteins, like PD-L1 or CTLA4 (187,188). Signaling receptor activity also holds significance in TNBC. For instance, dysregulation of signaling pathways, such as PI3K/AKT, was shown to contribute to tumor growth and invasion. Elevated receptor tyrosine kinase activity is also common in TNBC, with a role in promoting cell proliferation and survival (189).

Overall, the integrative analysis identified a list of genes which might contribute to TNBC aggressiveness. In addition, among the genes identified with H3K4me2 peaks we removed 47 genes that were shorter than 200 bp, such as microRNAs (miRNAs) and small nucleolar RNAs (snoRNA), and we also removed 18 genes corresponding to long intergenic non-coding RNAs (lincRNAs). Among the potential target genes listed in **Table 6**, we identified several genes to be involved in cell adhesion, migration and proliferation as well as to be involved in immune response and response to tumor microenvironment. Among the genes, we prioritized potential target genes to be validated and followed up based on the following criteria: 1) genes that are implicated in the development or progression of cancer and described in previous research in the context of breast cancer; b) genes that have higher fold change at expression level.

Cell Adhesion and tumor microenvironment	Signaling and Receptors		Immune Response	Enzymes and Metabolism	Transporters and Ion Channels	Cell Adhesion and differentiation	Poorly characterized
B3GAT1	ADGRF1	STAG3	AIM2	NPTX2	CA6	ANKRD45	HMSD
CHST4	BAIAP2L2	TMEM61	BTN1A1	NT5DC4	CACNA1B	DSC3	MID1
FOXL1	GAREM2	ZNF385C	FCER2	PDXK	CACNA1E	APBA2	MSLN
GALNT13	GPR12	KIF1A	FCRL2	PCSK1N	KCNS1	ECEL1	MUC15
CLDN10	GPR55	ALK	IDO2	PRAC2	LASTR	EDAR	TMEM71
CLDN14	GRIN2B	APCDD1L	PGLYRP4	PRSS3	SLC6A2	EMILIN3	ZNF571-AS1
MAL	SYT16	APCDD1L-DT	SILC1	SLC5A5	SLC10A4	FABP7	ACTG2
MATN4	UPK3B	KLHDC8A	SPIB	TMPRSS5	SLC13A5	PCSK1N	KIAA1755
OLFM2	USP6NL	PAX3	TREML2	HK3	SLC22A16	GABBR2	C6orf223
PCBP3	WNT5B	CARMIL2	TRIM2	HKDC1	SLC35F3	GABRE	CASC11
PLA2G4A	ZC4H2	CD70	GSDMC	HSD17B2	SLC22A16	GABRP	CMTM7
PLEKHB1	ZIC1	CHRDL2	HLA-DOB	KLK1	SLC13A5	SFRP5	KIAA1755
PLEKHG1	DLX6-AS1	DEFB1	IL12RB1	KLK6		SIRPG	LASTR
SEC14L4	LRP8	DEUP1	IL1R2	PTH2R		SLAMF7	SPAAR
SLC10A4	LY6D	DLX6	IL20RB	GLYATL2		SLAMF8	TCP11
SLC13A5	MFSD4A	S100B	IL21R	ALDH1A3		GCAWKR	
SOSTDC1	MGAM2	OTOF	IL2RA	C10orf90		SPHKAP	
THEMIS2	ZIC4	NKX2-5	ULBP3	COL9A1		SRD5A2	
TIGIT	DMRT3	RUNX3	APOBEC3A	DUOX2		KRT17	
TTC24	FOXD1	SIM1		FAM107A		KRT5	
ULBP3	FOXI3	SPIB		GPR12		KRT6A	
UPK3B	GRIN2B	S100A8		GPR55			
USP6NL	INA	CRLF1		IGF2BP1			
EGFR	KRT75	CXCL10		IGF2BP3			
	MEIOC	CXCL5		METTL7B			
				MIOX			
				NFIA-AS2			
				NPFFR1			

Table 6: List of genes with peaks at promoter regions and overexpression at transcriptomic level

5.5 Proteomic analysis of fresh frozen TNBC and Luminal A samples

As an added layer of information, we also performed a standard proteomic analysis of these clinical samples upon shotgun proteomics based on in-solution tryptic digestion prior to MS analysis of the same clinical samples analyzed by epi-proteomics, including 4 Luminal A samples and 8 TNBC samples. By label-free quantification (LFQ) method applied to the MS RAW data acquired on an Orbitrap type of instrument, we identified on average 4200 proteins per sample (**Figure 28A**). Despite the low number of samples, the two subtypes clearly separated by PCA analysis based on their quantitative proteomes (**Figure 28B**). After filtering, it was possible to quantify 3951 proteins in common; relative quantification of proteins in TN versus Luminal A reveal differentially expressed proteins, of which 221 proteins were significantly upregulated and 313 were significantly downregulated in the TN subtype versus the Luminal A, as shown in the volcano plot (**Figure 28C**). However, because of the limited number of samples analyzed, their overall inter-patient heterogeneity, the number of significantly regulated proteins was overall low; this limited the possibility to carry out the intersection of ChIP-seq data with the information of gene expression at the proteome level in addition to the transcriptome one. Further experiments devoted to a more in-depth proteomic analysis of a larger cohort to breast cancer patients will be devoted to address this issue. For instance, preliminary test carried out recently in our lab have shown

that introducing data independent acquisition (DIA) MS method for the analysis of proteomes are expected to significantly increase the proteome coverage also from low abundant samples (which were excluded in this first attempt).

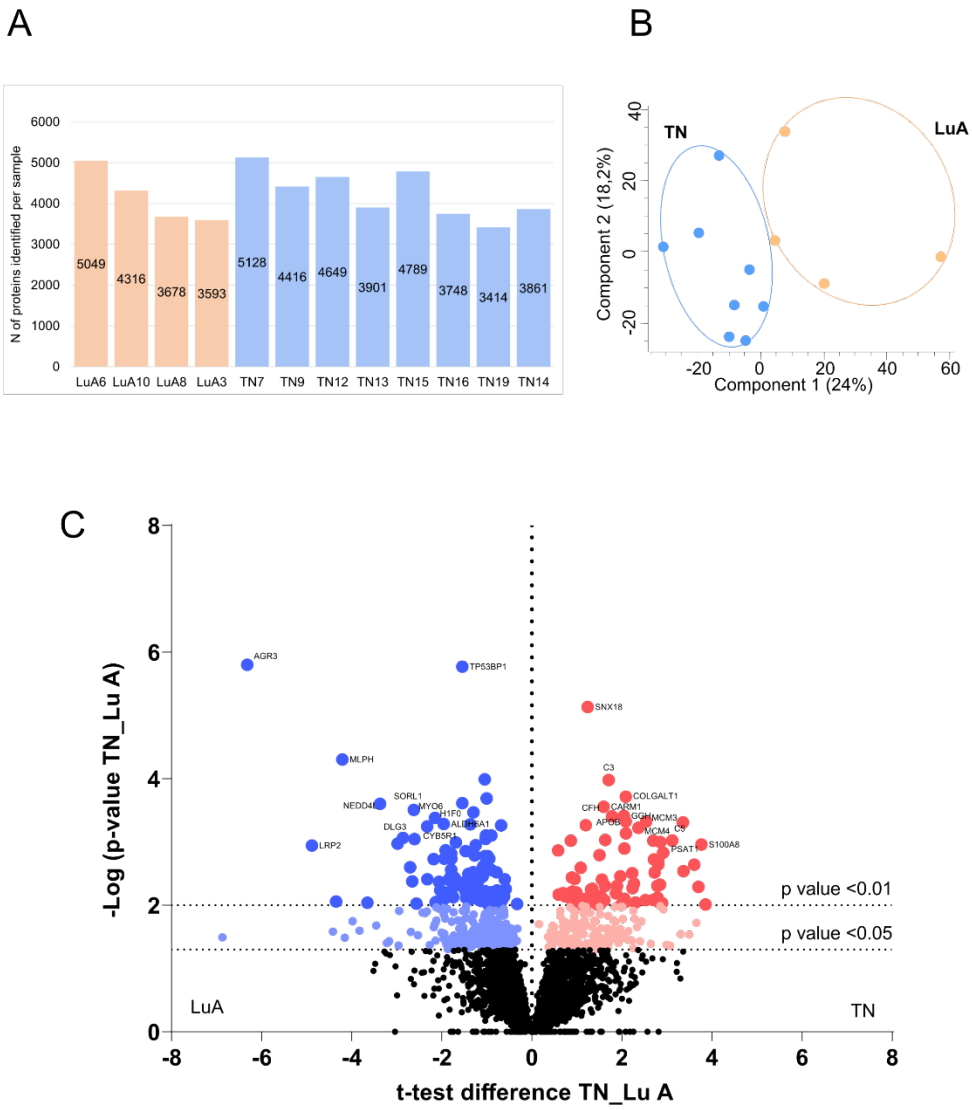


Figure 28: Global proteomic analysis of clinical samples

(A) Number of proteins identified per sample, Luminal A samples are displayed in orange and TN samples are displayed in blue. (B)PCA analysis performed on valid LFQ values for each sample. (C)Volcano plot displaying significantly upregulated (red) and downregulated proteins (blue) in TN compared to Luminal A samples defined by p-value<0.05. Most significant proteins are displayed with p-value <0.01.

5.6 Selection of model breast cancer cell lines for investigation of the impact of H3K4me2 modulation on breast cancer phenotypes

For the investigation of the molecular mechanisms both upstream and downstream of the increased levels of H3K4me2 in breast TNBC samples, we first set to identify the best cellular models which would reproduce epigenetic features similar to the primary clinical samples. However, the overall epigenetic pattern is known to vary significantly between cell lines and tissues, based on previous research from our group (161). In fact, even though specific differences among subtypes were found to be consistent across cell lines (e.g. H4K20me3), other histone PTMs are cell line dependent and therefore are representative of primary tumors only in some cases (161).

To select the ideal TNBC cells for investigating the connection between the presence of H3K4me2 at promoters and enhancers and the expression of genes linked to TNBC aggressiveness, we profiled histone PTMs in a panel of TN (MDA-MB-231, MDA-MB-436, and MDA-MB-468) and Luminal A breast cancer cell lines (MCF7, T47D, and ZR-751). As shown in **Figure 29**, the analysis of histone PTMs showed higher levels of H3K4me2 and H3K4me3 in the TN MDA-MB-231 cell line only, while the other two TN cell lines, MDA-MB-468 and MDA-MB-436, did not show significant changes in the levels of these modifications. Based on these data, we decided to select MDA-MB-231 as TN and MCF7 as Luminal A (displaying the lower levels of K4 methylation) as the best models for our downstream analysis, as they display the most divergent levels of H3K4me2 and H3K4me3.

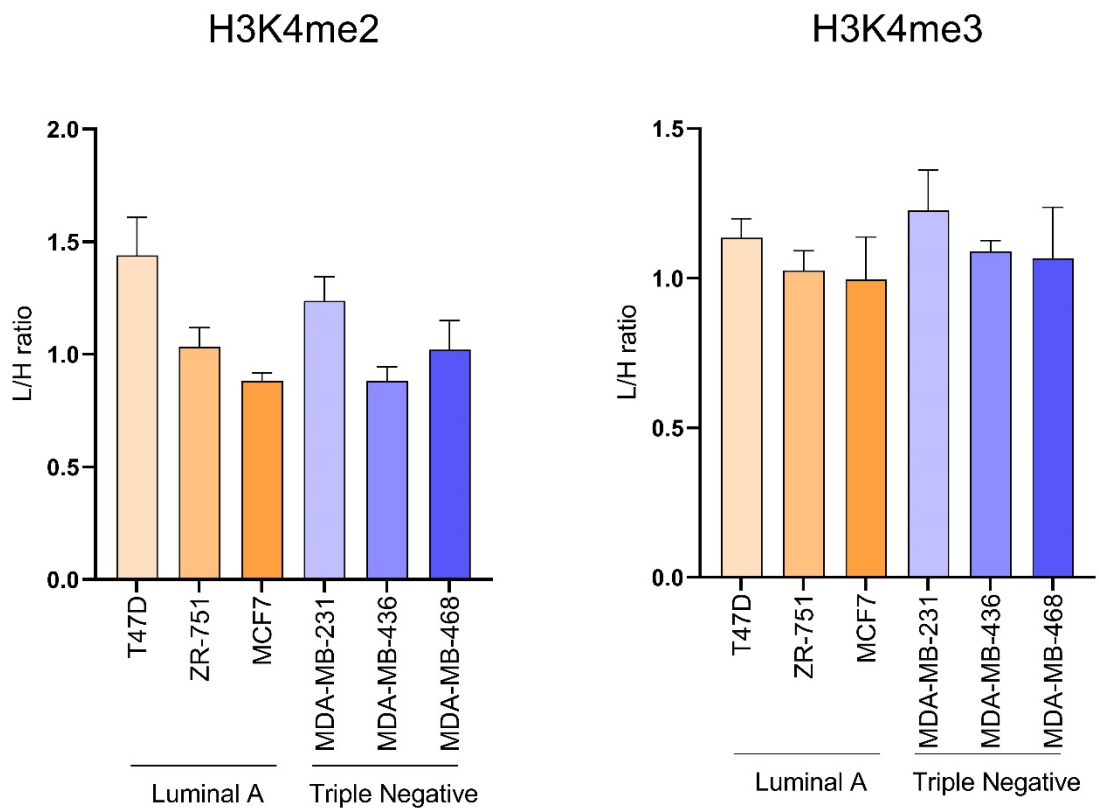


Figure 29: Histone PTM analysis in breast cancer cell lines

H3K4me2 and H3K4me3 levels are shown as L/H ratios of relative abundances in the Luminal A cell lines T47D, ZR-751, and MCF7 and in the Triple Negative cell lines MDA-MB-231, MDA-MB-436, and MDA-MB-468.

In addition, ChIP-seq analysis of the two cell lines selected, MDA-MB-231 and MCF7, showed a distribution similar to TN and Luminal A clinical samples, with the majority of the H3K4me2 peaks in common between the two subtypes localized at promoter regions (around 60%), while subtype-specific peaks were localized not only at promoters, but also at gene bodies and intergenic regions **(Figure 30A**). An example of peak distribution in TNBC clinical samples and TN breast cancer cell lines for promoter and super-enhancer regions is given for the genes SLAMF7 and MMP7 respectively (**Figure 30B-C**).

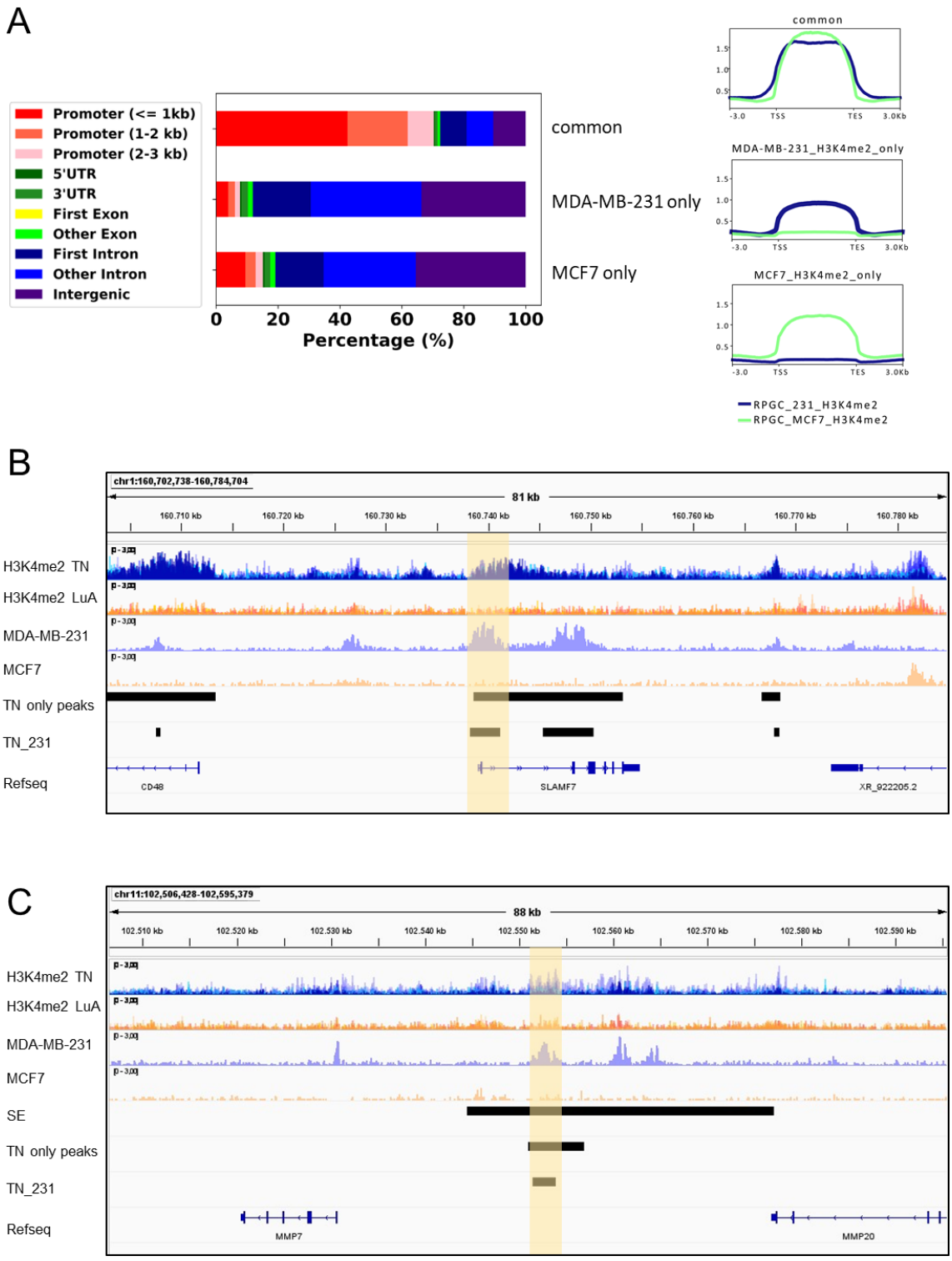


Figure 30: ChIP-seq analysis for H3K4me2 in breast cancer cell lines

(A) Comparative analysis of common and unique peaks identified for H3K4me2 in MDA-MB-231 and MCF7 cells. Association to different genomic regions indicates that the majority of unique peaks are located in intronic and intergenic regions. Signal at TSS and TES is displayed. (B) IGV browser display of H3K4me2 peak distribution at promoter of SLAMF7 gene and (C) H3K4me2 peak distribution at super-enhancer associated with MMP7 gene. Overlaid signal for H3K4me2 ChIP from 8 TNBC samples (blue color) and overlaid signal for H3K4me2 ChIP from 7 Luminal A samples (orange color) are displayed; MDA-MB-231 ChIP signal is displayed in blue; MCF7 ChIP signal is

displayed in orange. TN-unique peaks are shown in black as well as unique peaks confirmed in MDA-MB-231 cell line (TN_231). SE indicates super-enhancer regions. Selected genes are displayed with Refseq annotation hg38.

Next, in order to identify potential target genes for a follow up validation, we had to select the genes that had a TN-unique distribution and were expressed in our MDA-MB-231 TN cell line model. Overall, by integrating ChIP-seq data and RNA-seq data in TNBC clinical samples, we identified a total of 167 genes which were upregulated at transcript level and with H3K4me2 peak at promoter region. Of these genes, 51 were characterized by the presence of H3K4me2 in promoter regions also in MDA-MB-231 cells and half of them were expressed at transcript level. (**Figure 31A**). We performed the same type of data integration also for super-enhancer regions characterized by the presence of H3K4me2 peaks. Of the 55 genes upregulated at transcript level, 34 showed H3K4me2 in the same regions in MDA-MB-231 cells and 13 were also expressed at transcript level. These genes are displayed in **Figure 31B**, ranked by fold change. Because expression of genes could also depend on copy number variation, amplifications or gain of function mutations, we excluded genes that were characterized by these changes only in the TN subtype.

For instance, among the genes identified, ADGRF1 displayed presence of H3K4me2 peaks at both promoter and enhancer region. ADGRF1 is a member of the G-protein coupled receptor 2 family, predicted to be involved in G protein-coupled receptor signaling pathway and to act upstream of or within several biological processes. However, it is known to be characterized by high copy number variation in TNBC therefore its expression might not depend only on the presence of H3K4me2 (190).

Instead among possible interesting candidate genes to validate functionally, we identified SLAMF7, PLA2G4A, and IGF2BP3 with H3K4me2 peaks at promoter regions. SLAMF7 is an essential molecule for the interaction between immune cells and cancer cells and positive regulation for NK cell activation (191). PLA2G4A is a phospholipid metabolizing enzyme that releases free fatty acids and which contribute to the development of the tumor microenvironment, promoting immune evasion, angiogenesis, tumor growth, and invasiveness (192). IGF2BP3 (Insulin-like growth factor-2 messenger RNA-binding protein 3) has been reported to contribute to tumorigenesis in several human cancers and has been described to be involved in biological processes such as proliferation and migration (193), suggesting a role in promoting TNBC growth. We chose these genes to validate our hypothesis of a direct, selective regulation of gene expression in TNBC by H3K4me2 by the CRISPRi system.

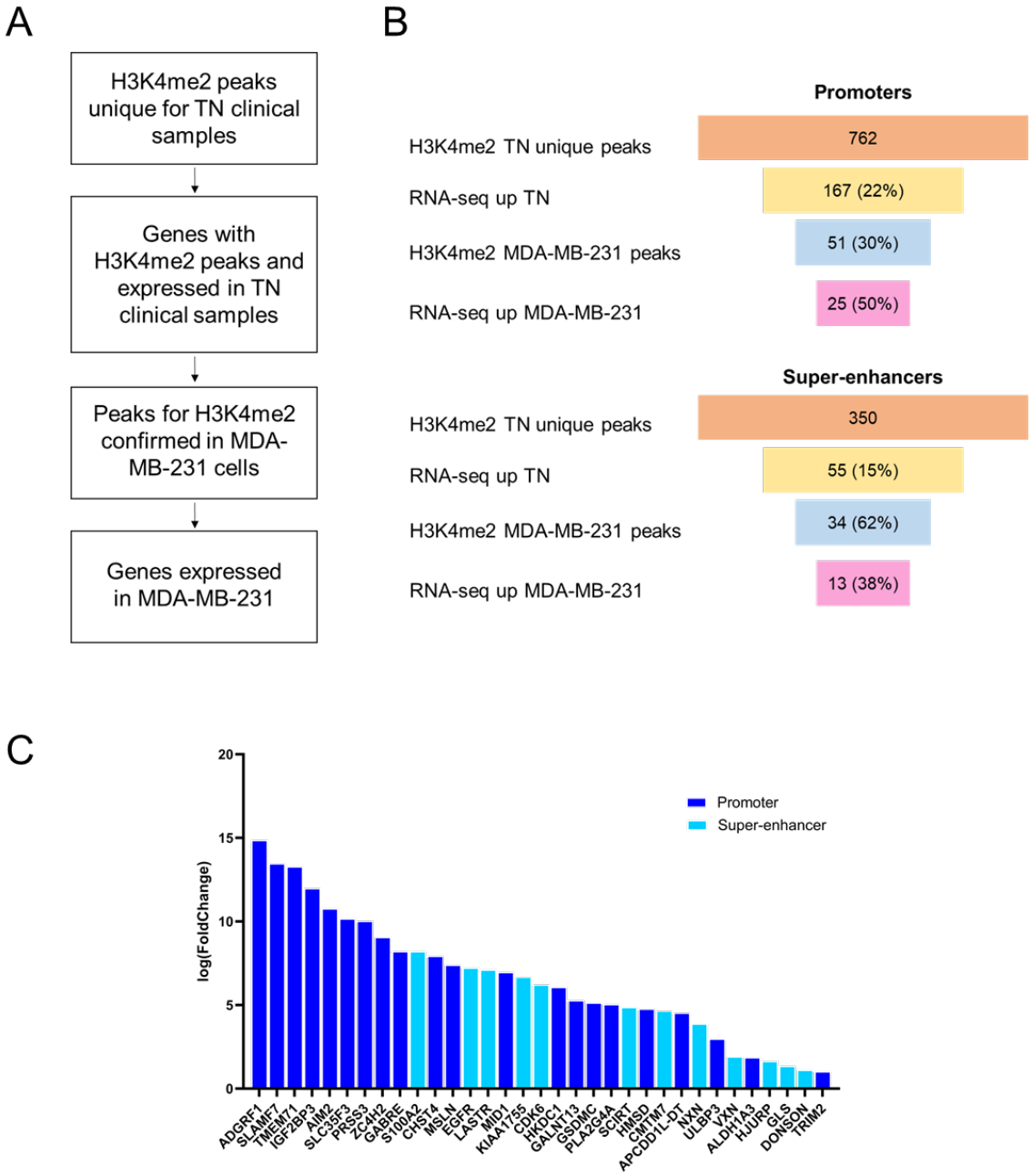


Figure 31: Selection of target genes in MDA-MB-231 cell line

(A)Workflow of data integration starting from genes identified in TNBC clinical samples with H3K4me2 peaks and upregulated at transcript level, followed by integration of ChIP-seq and RNA-seq data from MDA-MB-231 cells to identify potential targets. (B)Number of genes identified by integrative analysis for ChIP-seq and RNA-seq data in TNBC clinical samples and MDA-MB-231 cell line with H3K4me2 peaks at promoters and super-enhancers (C)Genes selected ranked by log Fold change in MDA-MB-231 cells. Genes with H3K4me2 peak in super-enhancers are highlighted in light blue, genes with peaks in promoters are highlighted in dark blue.

5.7 Setup of the CRISPRi system in MDA-MB-231 cell line

5.7.1 Generation of a stable cell line expressing dCas9 bound to LSD1 effector domain

To mechanistically prove the association between the presence of H3K4me2 and the expression of genes associated with TNBC aggressive phenotype, we used the CRISPRi technology which allows epigenetic editing and targeting of specific region of interest by the use of repressor domains specific for histone PTMs removal or deposition (130). Based on the data obtained from the integrative analysis, as a first strategy, we decided to study the localization of H3K4me2 at promoter regions, to systematically probe the functions of H3K4me2 at promoters in TNBC and determine the consequences on expression of identified target genes. For this purpose, we engineered MDA-MB-231 cell line to stably express a Cas9, deficient for nuclease activity, fused to the LSD1 effector domain (dCas9-LSD1) and tagged with the mCherry fluorescent marker (194). LSD1 is a demethylase specific for H3K4 methylation. The LSD1 effector domain induces local chromatin repression via demethylation of H3K4me2 and, when fused to dCas9, allows the use of the programmable properties of CRISPR to target and inhibit any genomic loci of interest, namely the regions occupied by H3K4me2 (**Figure 32A**). Because the dCas9-LSD1 vector that we selected was an inducible vector with the TRE3G promoter sequence but not the TET3G transactivase expression sequence, we first infected the cells with a Tet-On 3G vector for the stable expression of the Tet-On transactivation system (**Figure 32B**). Notably, target cells that express the Tet-On 3G transactivator protein and contain a gene of interest under the control of a TRE3G promoter, when bound by doxycycline, a tetracycline analog, result in a conformational change of the Tet-On 3G protein, resulting in the expression of the target gene, in this case dCas9-LSD1.

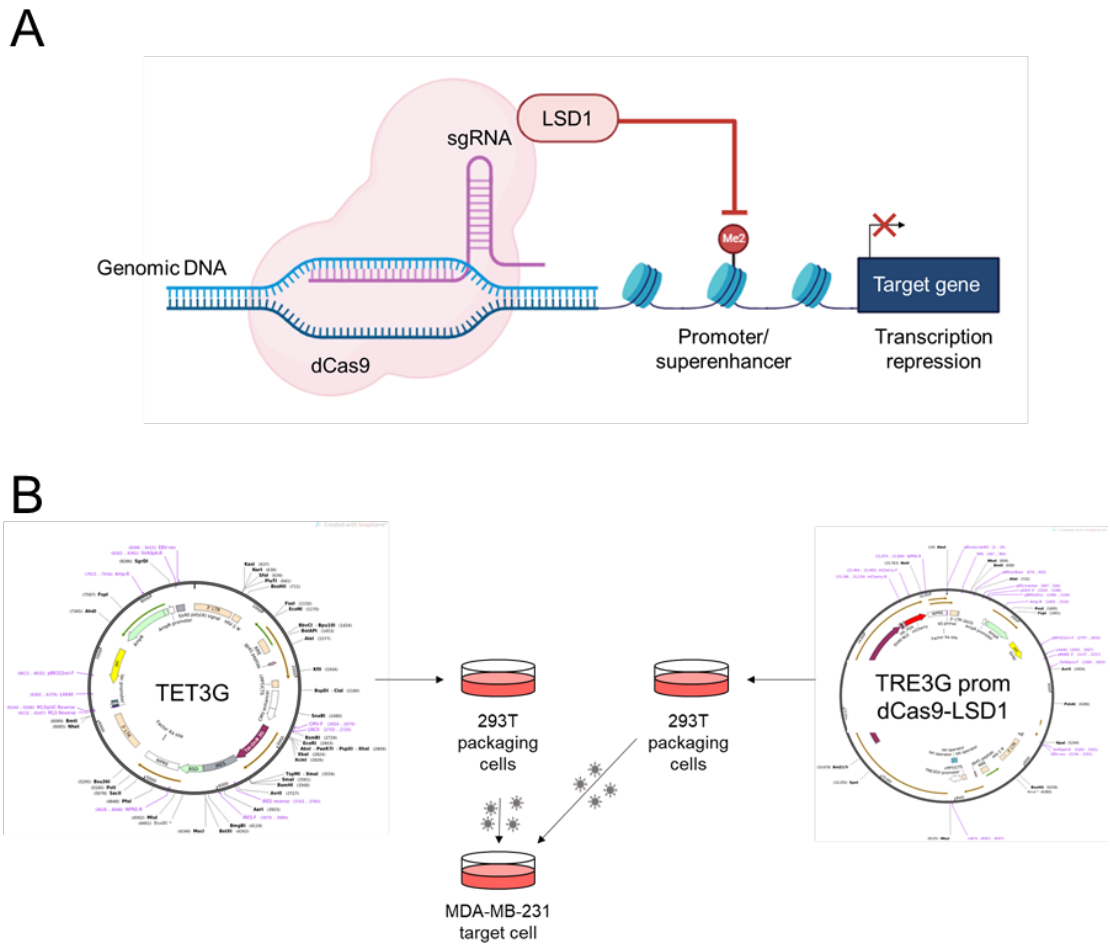


Figure 32: CRISPRi system and scheme of cell infection with the CRISPRi vector

(A)CRISPRi system based on the use of dCas9 bound to the LSD1 effector domain to target genomic regions characterized by the presence of H3K4me2. The removal of H3K4me2 from these regions would result in transcriptional repression. (B)Workflow of cell infection with TET3G transactivation system and dCas9LSD1 vector expressing the TRE3G promoter. Created with BioRender.com

After generation of the MDA-MB-231 cell line with the stable expression of the TET3G transactivase, we infected the cells with the dCas9LSD1 vector. After induction with doxycycline for at least 48h, we assessed the quality of infection by FACS analysis for the mCherry marker (**Figure 33A**). Even though the infection was successful in only around 40% of cells, probably because of the large size of the vector (around 16,400 bp), it was possible to select the MDA-MB-231dCas9LSD1-mCherry positive cells by sorting. In addition, after 48h induction, mCherry expression and dCas9 expression were assessed respectively at an Evos fluorescence microscope (**Figure 33B**) and by Western Blot (**Figure 33C**). We tested different concentrations of doxycycline ranging from 1 to 10 µg/ml, which

resulted in similar expression of dCas9. Therefore, we selected 1 µg/ml as the ideal concentration. In addition, release from doxycycline for 48 h resulted in reduced levels of expression of dCas9.

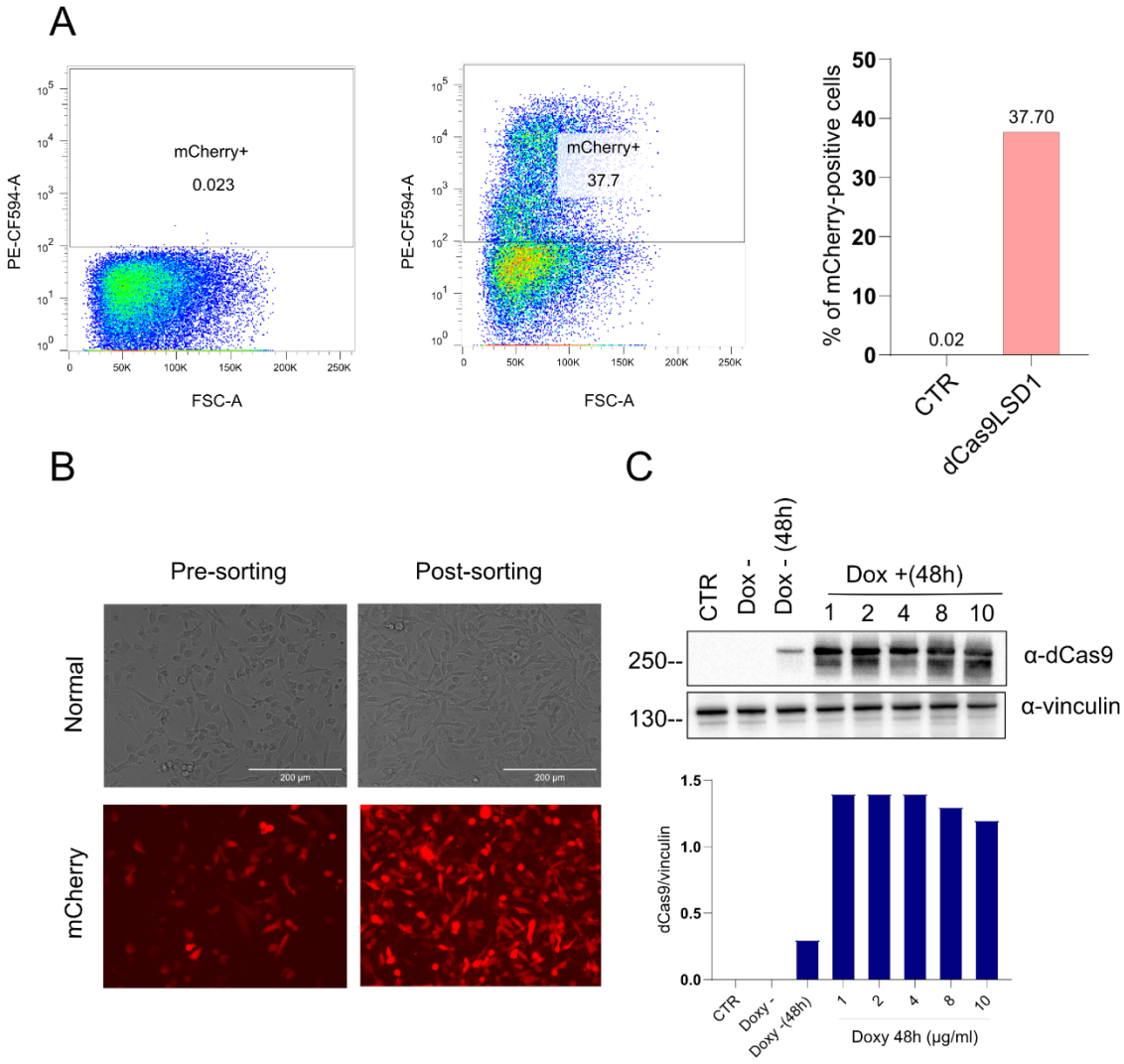


Figure 33: Validation of dCas9LSD1-mCherry vector expression in MDA-MB-231 cells

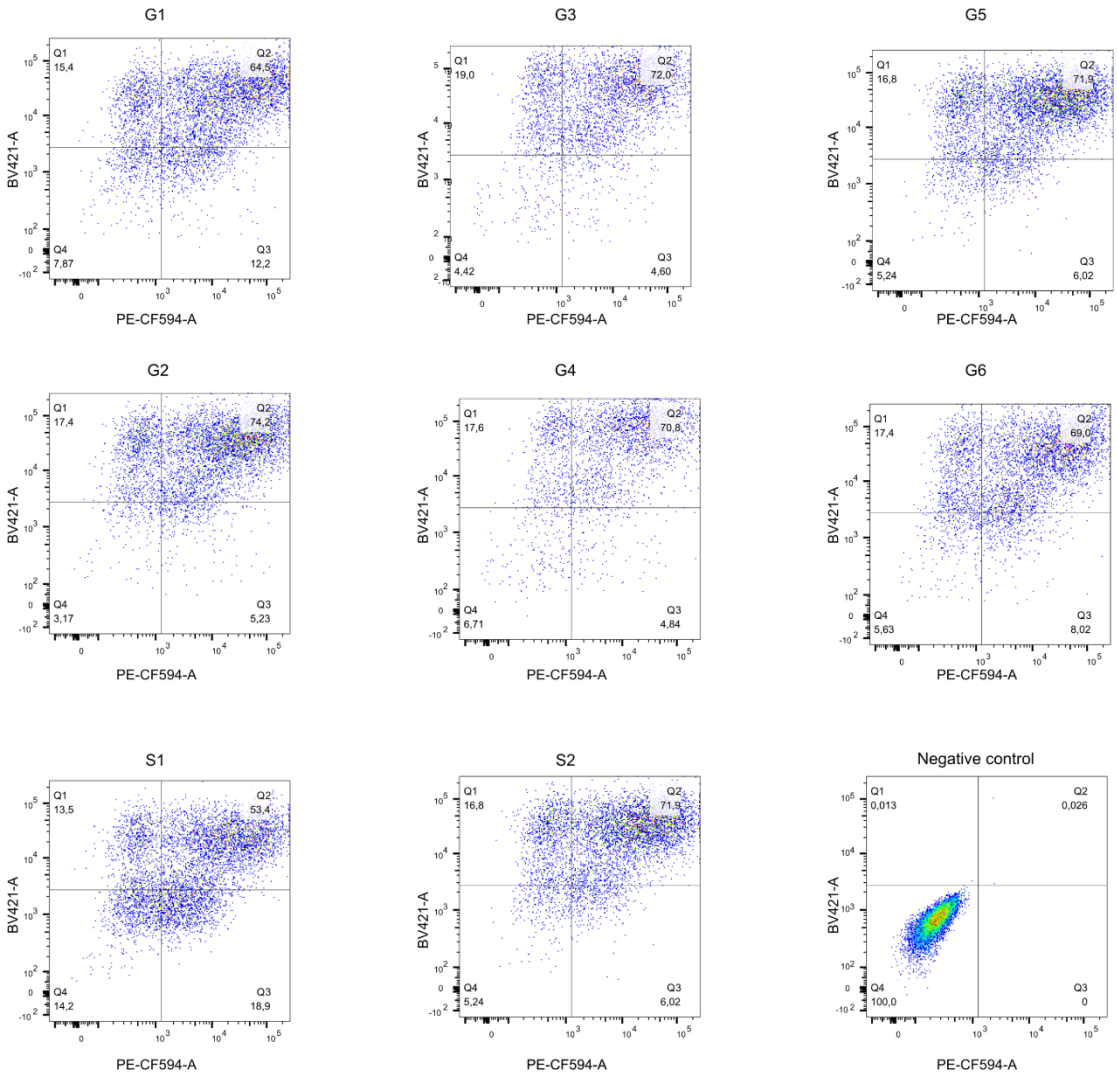
(A) FACS analysis with PE-CF594 laser for quantification of mCherry-positive (+) cells by comparison with wild-type as control (CTR). Selection of gate was performed on wild-type cell line. (B) Pre-sorting and post-sorting pictures of mCherry-positive cell lines taken at 20x magnification at Evos fluorescence microscope after addition of doxycycline for 48h. Scale bar=200µm. (C) Western Blot for expression of dCas9 vector by incubation with α-Cas9 antibody and α-vinculin antibody as normalizer.

5.7.2 Modulation of target genes with specific guide RNAs directed to the promoter regions occupied by H3K4me2

To test the link between H3K4me2 and gene expression when this mark is localized at promoter regions, we selected as target genes SLAMF7, PLA2G4A, and IGF2BP3 based on their role in cancer. For this experiment, each well received a different set of custom-designed guide RNAs (sgRNAs) to specifically inhibit one target per well by designing 2 sgRNA for each of the selected genes. The target region was selected at H3K4me2 peak summit, in promoter regions. sgRNA G1 and sgRNA G2 were designed to target SLAMF7 promoter region, sgRNA G3 and sgRNA G4 were designed to target PLA2G4A promoter region, and sgRNA G5 and sgRNA G6 were designed to target IGF2BP3 promoter region. As negative controls, we used two different non-targeting scrambled sgRNAs, S1 and S2. First, the sgRNA were cloned into a sgRNA vector with the BFP selection marker and puromycin resistance, which was digested with the BbsI enzyme. Ligated vectors were sequenced by Sanger sequencing to assess the presence of the target sequence and lack of mutations. Next, the MDA-MB-231dCas9LSD1 cell line was infected with the sgRNA-BFP vector and induced with doxycycline prior to FACS analysis to assess infection efficiency by selection of mCherry and BFP positive cells. 60% to 70% of cells were infected and positive to both mCherry and BFP (**Figure 34**).

Figure 34 (next page): FACS analysis of MDA-MB-231dCas9LSD1 infected with sgRNA-BFP vector

Percentage of mCherry-positive and BFP-positive MDA-MB-231 cells, assessed by flow cytometry analysis 48h upon Doxycycline induction. Dot plots display the data for each gRNABFP vector and a negative control (wild type cell line). Q1: BFP-positive and mCherry-negative cells, Q2: BFP-positive and mCherry-positive cells, Q3: mCherry-positive and BFP-negative cells, Q4: BFP-negative and mCherry-negative cells.



The cells were then FACS sorted and kept in culture in the absence of doxycycline. Pictures of cells after sorting are shown in **Figure 35**. The cells were induced for 72h to maximize dCas9 expression and pelleted for RNA extraction and RT-qPCR. The RT-qPCR analysis showed a reduction in gene expression upon epigenetic modulation by using all the sgRNA designed, indicating that the localization of H3K4me2 at these regions is linked to the expression of these genes. This result obtained on three representative target genes confirms our hypothesis that the presence of H3K4me2 at the promoters of the target genes identified through our multi-OMICs strategy positively regulates their expression. Currently, we are applying the same strategy to the genes presenting H3K4me2 peaks in their associated enhancers.

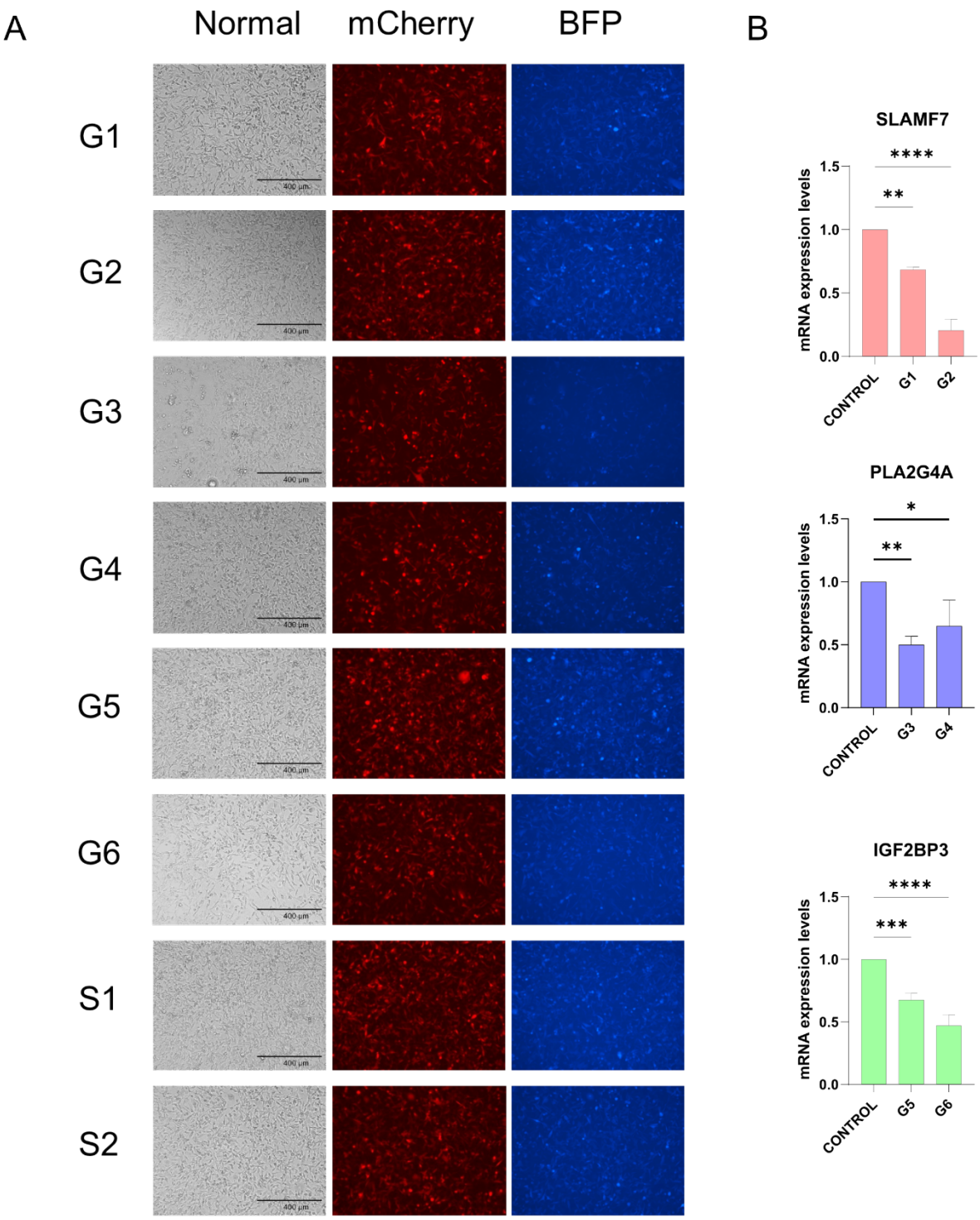


Figure 35: MDA-MB-231dCas9LSD1 infected with gRNABFP vector targeting SLAMF7, PLA2G4A, and IGF2BP3

(A) Images acquired at Evos fluorescence microscope at 10x magnification for mCherry, and BFP expression showing positive expression for both mCherry and BFP for all the gRNA. Scale bar=400µm. (B) RT-qPCR data showing mRNA expression levels for target genes normalized on control (geometric mean between scramble 1 and scramble2) and normalized on GAPDH and β-ACTIN (geometric mean).

5.8 Selection of methyltransferase inhibitors for modulation of H3K4me2 levels

The results obtained so far showed that the increase in H3K4me2 at specific genomic regions in TNBC favors the expression of a set of genes that may drive the TNBC aggressive phenotype. We therefore reason that decreasing the levels of H3K4me2, not only by CRISPRi, but by mean of other strategies, among which the pharmacological inhibition of the methyltransferases depositing this PTM, could affect the expression of the target genes identified. As a matter of fact, methyltransferase inhibitors could be exploited as a potential therapeutic strategy for modulating H3K4me2 levels, thus inhibiting the expression of TN-specific genes and ultimately phenotypic outcomes.

Several inhibitors have been developed to target the many methyltransferases responsible for the deposition of H3K4me2 and H3K4me3 (195) and we started testing some of them. Interesting preliminary results have been collected with the compound OICR-9429. This inhibitor functions as a potent antagonist of the WDR5 domain competing with the interaction between WDR5 and the K4-sepecific methyltransferase MLL by binding to the central peptide-binding pocket of WDR5; this results in the inhibition of histone H3K4 methylation. WDR5 inhibitors have been developed to disrupt the interaction between the MLL-WDR5 proteins and thus impair the transcription of oncogenic genes. These inhibitors are now emerging as promising new agents for treating cancer in preclinical studies. OICR-9429 has been proposed for treatment in a range of cancer types, not only of hematological origin, such as non-MLL-rearranged leukemia, but also for solid tumors including colon, pancreatic, prostate, and bladder cancer, also in combination with chemotherapy (195,196).

First, we assessed the effect of its use on the level of H3K4me2 by WB, upon treatment of MDA-MB-231 cells for 48h at 50 μM concentration (experimental conditions chosen based on available literature) (196,197). OICR-9429 reduced the level of H3K4me2, normalized on total histone H3 (**Figure 36A**). We then performed RT-qPCR (**Figure 36B**) of some of the target genes identified in our multi-OMICs analysis to assess the effect of the compound on gene expression. Remarkably, some of these genes displayed reduced expression upon OICR-9429 treatment when compared with DMSO as control, in line with the reduced H3K4me2 level. Therefore, deposition of H3K4me2 at specific genomic regions of these genes including IGF2BP3, PLA2G4A, PRSS3, MSLN, and ANLN, seem directly linked to an effect on expression levels.

Next, we tested the effect of the inhibitor on MDA-MB-231 cell viability of by Cell Titer Glo assay: we treated the cells for 48h with different concentrations of OICR-9429,

extrapolating an IC50 of 138 µM (**Figure 36C**). The highest effect on cell viability was measured at concentrations between 100 µM and 200 µM. In line with our hypothesis that H3K4me2 might have a role in TN phenotypic features, upon 48h treatment we also observed an effect of the drug on colony formation, which is an assay that measures the stemness property of cancer cells (**Figure 36D**). These preliminary results are very encouraging but need to be corroborated with the use of other inhibitors, and test the effect on the expression of additional target genes.

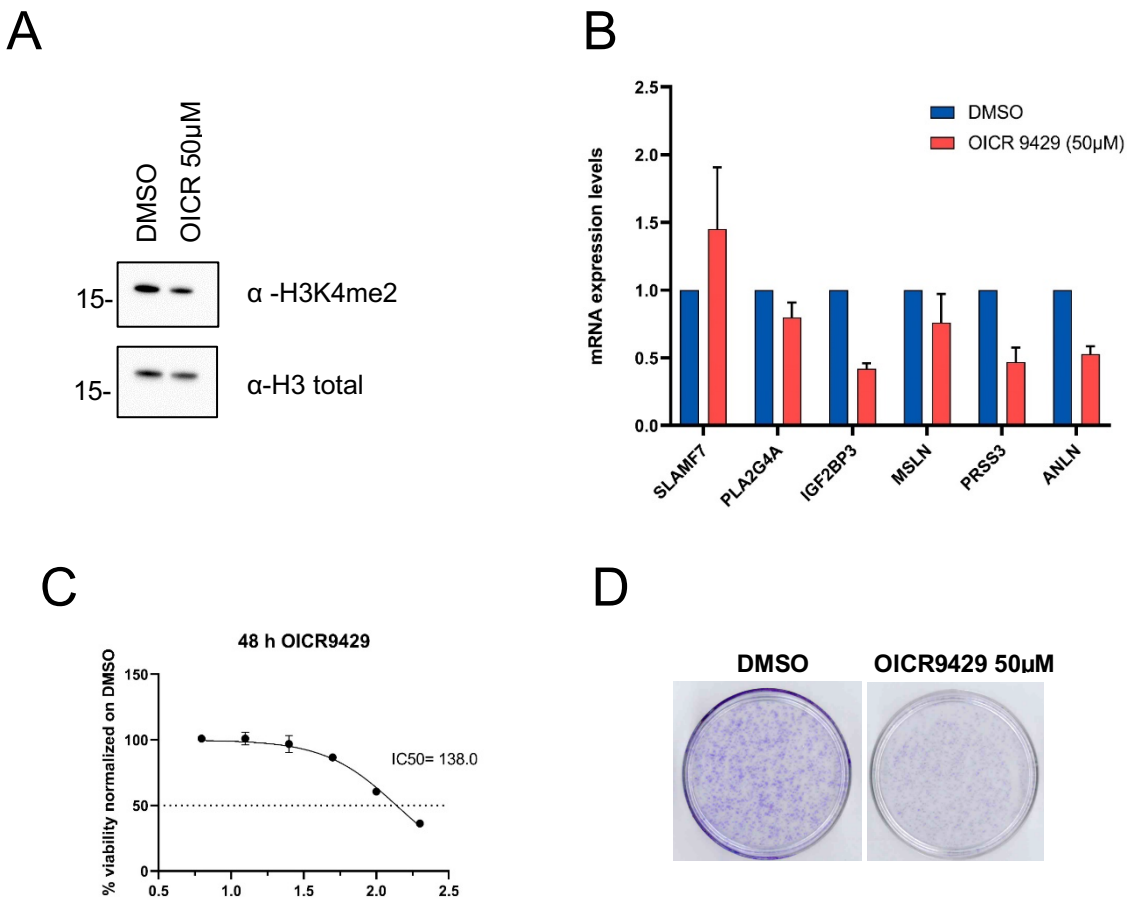


Figure 36: treatment of MDA-MB-231 cell line by OICR-9429 inhibitor

(A) Western Blot for H3K4me2 and total histone H3 as control in MDA-MB-231 cells. (B)RT-qPCR for expression of H3K4me2 target genes. Ct values were normalized on β-ACTIN. $2^{\wedge(\Delta\Delta Ct)}$ values were normalized on DMSO as control. (C) Cell Titer Glo after treatment for 48h with OICR-9429 at different concentrations (6.25, 12.5, 25, 50, 100 and 200 µM). The percentage of viability was calculated by normalization of values on DMSO and drug concentration was log-transformed for calculation of IC50. (D) Colony formation assay stained by crystal violet in DMSO as control and cells treated with OICR-9429 50 µM for 48h. Cells were kept in culture for 8 days.

5.9 Investigation of upstream mechanisms of subtype specific PTMs by investigation of histone modifying enzymes

In a simple cancer epigenetic scenario, the alterations in the levels of histone modifications derive from the abnormal expression, or function, of the epigenetic enzymes responsible for the addition or removal of these modifications. To explore whether variations in histone modifications observed in TN tumors compared to the other breast cancer subtypes are linked to the alteration of epigenetic enzymes, we analyzed their RNA expression levels in both our RNA-seq data and the data extrapolated from the public repository TCGA, including not only TNBC and Luminal A samples, but also the other breast cancer subtypes. The bubble diagram depicted in **Figure 37** (left panel) illustrates the changes in gene expression between TN tumors and Luminal A tumors (right panel). Globally, we identified a series of changes in histone modifying enzyme levels; such changes were more pronounced in the TCGA data than in our datasets, probably due to the small number of tested samples and the known heterogeneity of clinical specimens; however, their trend in our experimental transcriptomics confirmed overall the same changes observed in the TCGA data. We identified not only variable levels of histone methyltransferases and demethylases, but we also identified changes in lysine acetyltransferases (KATs), which were globally reduced in TNBC compared to all the other subtypes and a variable expression of histone deacetylases (HDAC).

We searched for possible correlation between expression of histone modifying enzymes and the changes observed in the levels of histone PTMs in the TNBC subtype but in most of the cases we did not find a direct correlation. For instance, this is the case of the enzymes involved in deposition and removal of H3K27me3. We identified an increase in EZH2 levels, a methyltransferase specific for the deposition of H3K27me3, a mark observed to be overexpressed in most of cancer types and a frequent prognostic indicator of poor survival (83). In this case, the levels observed were opposite to the levels of H3K27me3, which we found instead to be decreased in TNBC tumors. In addition, opposite to the increase in EZH2 levels, we also identified a decrease in expression of KDM6A, which is a H3K27me3 demethylase. Similarly, for H3K4me2 and H3K4me3, even though increased levels for KMT2B, a methyltransferase, were identified, other methyltransferases were unchanged or even decreased, and higher expression of demethylases was also observed (e.g. KDM1A, KDM1B, KDM5A, KDM5C). The levels of H4K20me3 were also not consistent with the levels of its specific methyltransferases. Indeed, KMT5B (SUV4-20H1) and KMT5C (SUV4-20H2), were highly expressed in TNBC samples compared to all the other subtypes, while the levels of the demethylase PHF2 were instead characterized by lower expression in

TNBC. This result suggests that other factors may contribute to changes in the level of histone PTMs, including enzyme mutations, higher turnover rates or the altered function of multi-subunit chromatin modifying complexes.

There was, however, one case where histone PTMs and HME levels showed a remarkable correlation. Indeed, in accordance with the increase of H3K9me3 in TN, we observed an increase in several specific H3K9me3 methyltransferases, namely SUV39H1, SUV39H2 (which was also found in the internally acquired data), and SETDB1. In addition, a decrease of H3K9-specific demethylases KDM3B and PHF2 was also identified.

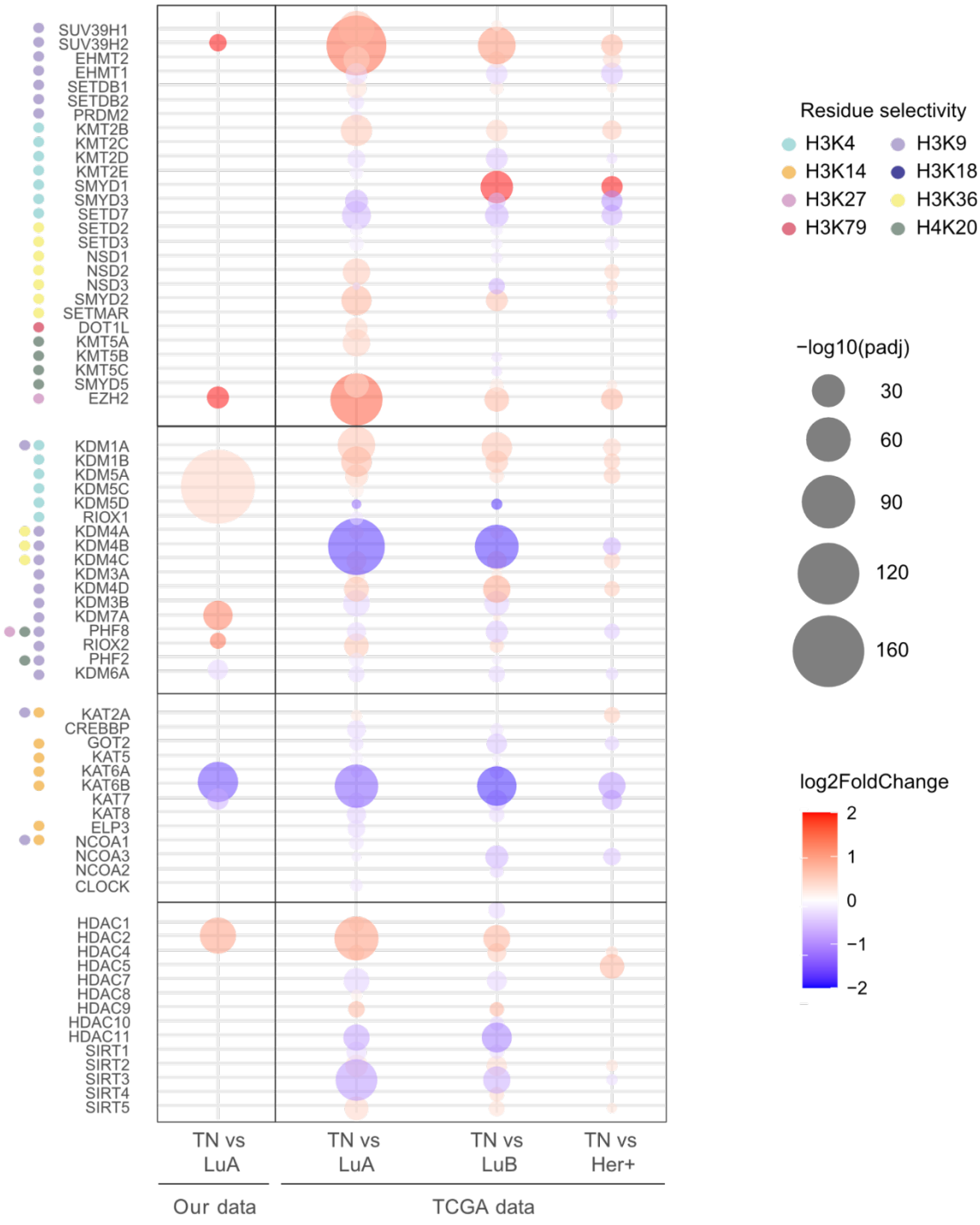


Figure 37 (previous page): Analysis of expression levels of epigenetic enzymes in breast tumor samples belonging to different molecular subtypes

Left panel: analysis of 3 Luminal A samples, and 6 TNBC samples. Right panel: analysis of gene expression from breast cancer samples from the TCGA repository. The color indicates an increase (red) or decrease (blue) in TNBC compared to the other subtypes. The size of the bubble indicates the significance of the change. Only enzymes with at least one significant change are shown. Residue selectivity is also displayed. TN: Triple Negative, LuA: Luminal A, LuB: Luminal B, Her+: Her2 positive breast cancer.

5.10 Investigation of upstream mechanisms of H3K4me2 and H3K4me3 levels in TNBC

The analysis presented in **Figure 37** did not reveal any correlation between the expression of epigenetic enzymes associated with H3K4 in different breast tumor subtypes and the increase in H3K4 methylation found in TNBCs. However, we discovered, in collaboration with the group of Prof. Negri at Sapienza University, that the levels of H3K4 methylation in breast cancer cell lines belonging to different subtypes may depend on the differential expression of an isoform of the H3K4-specific demethylase KDM5B. In a collaborative study carried out with their group, we experimentally characterized a predicted isoform, namely KDM5B-NTT, which is truncated at the N-terminal domain and is catalytically inactive, thus acting as a dominant negative (88). KDM5B-NTT was found at higher levels in the TN MDA-MB-231 cell line compared to the Luminal A MCF7 cell line and accumulates in breast cancer cell lines due to its higher protein stability compared to the canonical isoform, PLU-1. Within this study, we conducted MS analysis of H3K4 methylation levels upon overexpression of KDM5B-NTT isoform in MCF7 and MDA-MB-231 cell lines (**Figure 38**). After 24h from transfection, overexpression of KDM5B-NTT in the MCF7 cell line resulted in increased H3K4me3 levels, which reached a signal intensity similar to that observed in the MDA-MB-231 cells at basal conditions. Overexpression of KDM5B-NTT had instead no effect on MDA-MB-231 cells, probably because this cell line is already characterized by higher constitutive levels of this isoform. This result is very interesting also in the context of clinical samples since differential expression of KDM5B-NTT, which works as a dominant-negative isoform, may explain the observed changes in H3K4 methylation levels between TNBC and Luminal A clinical samples, thus suggesting that additional, more complex mechanisms, may be involved in deposition of H3K4 methylation in composite systems, *in vivo*. Since this isoform might be subtype-dependent,

current work is devoted to measuring the levels of KDM5B-NTT and PLU-1 also in TNBC and Luminal A primary samples.

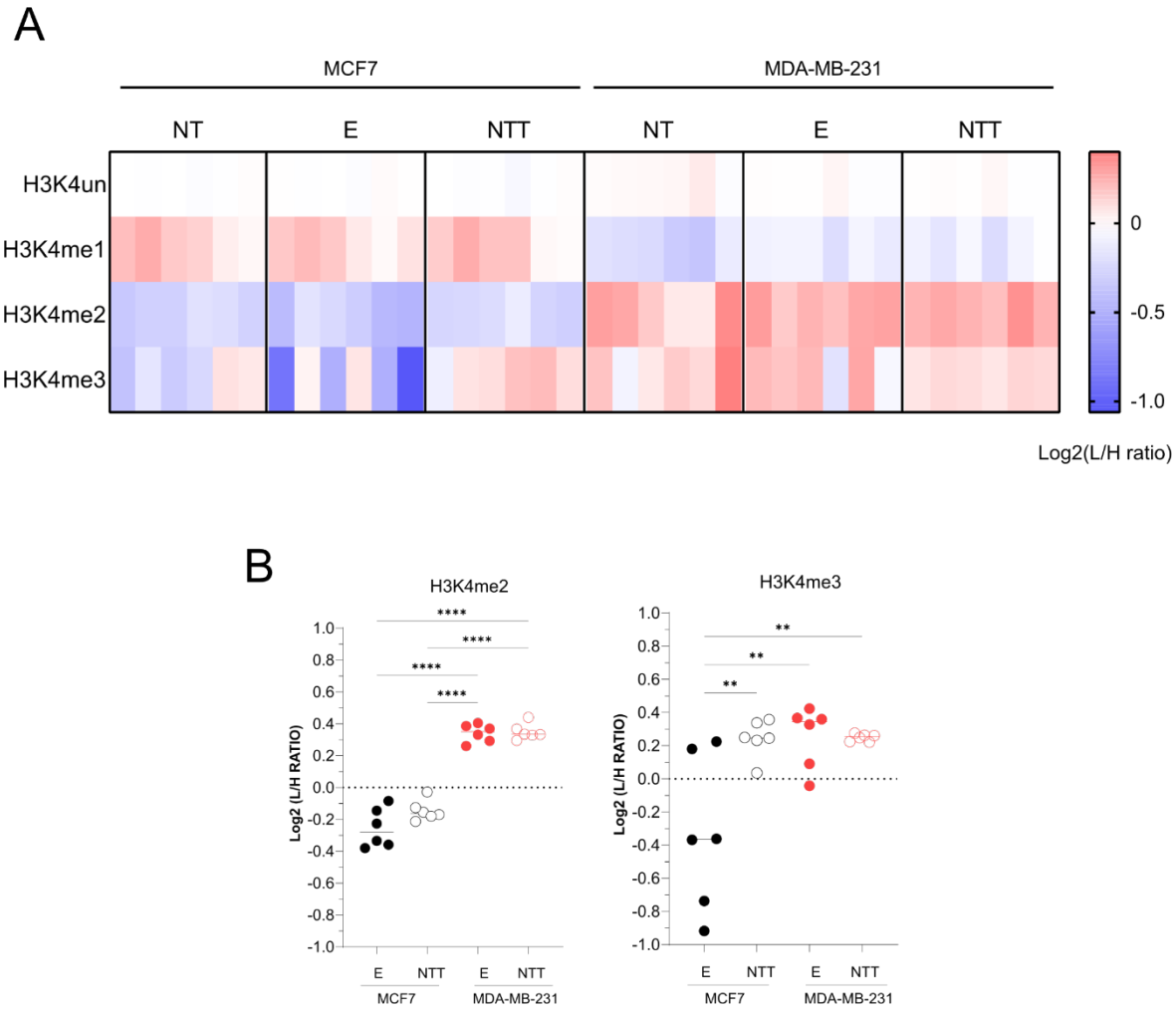


Figure 38: MS analysis of H3K4 methylation in MCF7 and MDA-MB-231 cell lines upon overexpression of KDM5B-NTT

*(A) Heatmap display of H3K4 methylation levels shown as log2 L/H ratios normalized on average value across all samples. NT (Not transfected cells), E (empty vector), NTT (cells with overexpression of KDM5B-NTT isoform). (B) H3K4me2 and H3K4me3 log2 transformed L/H ratios in MCF7 and MDA-MB-231 cell lines overexpressing KDM5B-NTT compared to the control E, empty vector. Statistical analysis was performed by ANOVA, followed by Tukey multiple comparison test. **p-value<0.01, ****p-value<0.0001. Adapted from (88)*

5.11 Modulation of H3K9me3-specific methyltransferases in breast cancer cell lines

Among the epigenetic changes observed in the different breast cancer subtypes, we found significant higher levels of H3K9me3 in TN compared to Luminal A samples. Interestingly, in line with this piece of data, by the analysis of histone modifying enzymes expression through the internal transcriptomic data and TCGA data, in TNBC samples, we identified a significant increase in the expression of methyltransferases specific to this modification, SUV39H1, SUV39H2, and SETDB1, which could be responsible for the increased H3K9me3 levels in the TNBC samples.

As done for H3K4me2 studies, we first analyzed several TN and Luminal A cell lines to identify the best cellular models to investigate the correlation between H3K9me3 levels and H3K9-specific methyltransferases by pharmacological and genetic manipulation. Among the Triple Negative cells, MDA-MB-231 and MDA-MB-436 displayed higher levels of H3K9me3 alone and in combination with K14ac, while we observed lower levels in ZR-751, MCF7 and, to a lesser extent, T47D. Therefore, the previously used MDA-MB-231 and MCF7 cell lines appear to be the best model cell lines to also study H3K9me3 modulation (**Figure 39**).

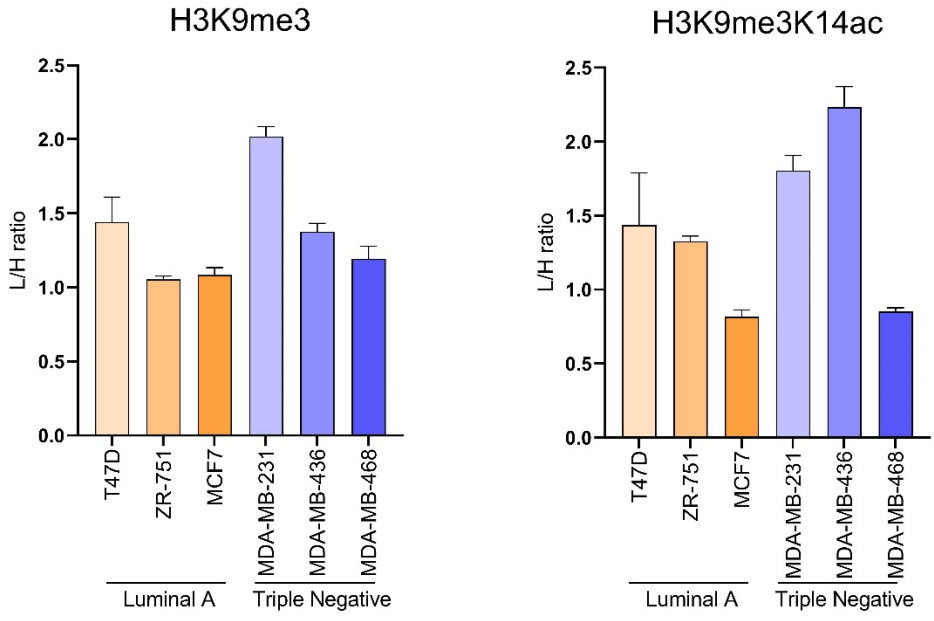


Figure 39: MS-based profiling of H3K9me3 in Luminal A and TN breast cancer cell lines
H3K9me3 and H3K9me3K14ac levels are shown as L/H ratios of relative abundances in different breast cancer cell lines. Luminal A: T47D, ZR-751, MCF7; Triple negative: MDA-MB-231, MDA-MB-436, MDA-MB-468.

5.11.1 Treatment with SUV39H2 inhibitor results in high phenotypic effect in MDA-MB-231 cells

We tested a commercially available selective inhibitor of SUV39H2, OTS186935 by assessing its effect in proliferation assay by Cell Titer Glo, at increasing concentrations (1.2 µM, 2.4µM, and 4.8 µM) for time intervals up to 96 hours and demonstrated that it inhibits cell proliferation more in TN MDA-MB-231 cells than in MCF7 at higher concentrations; a result consistent with the higher levels of SUV39H2 in cell lines, suggesting the impact of this enzyme on the cellular phenotypic features (**Figure 40**). Because histone methyltransferase inhibitors might have effect on multiple methyltransferases and no selective inhibitors targeting individually SUV39H1 and SETDB1 are commercially available, we decided to proceed by genetic depletion of these H3K9 specific methyltransferases by shRNA knockdown experiments in MDA-MB-231 cell line.

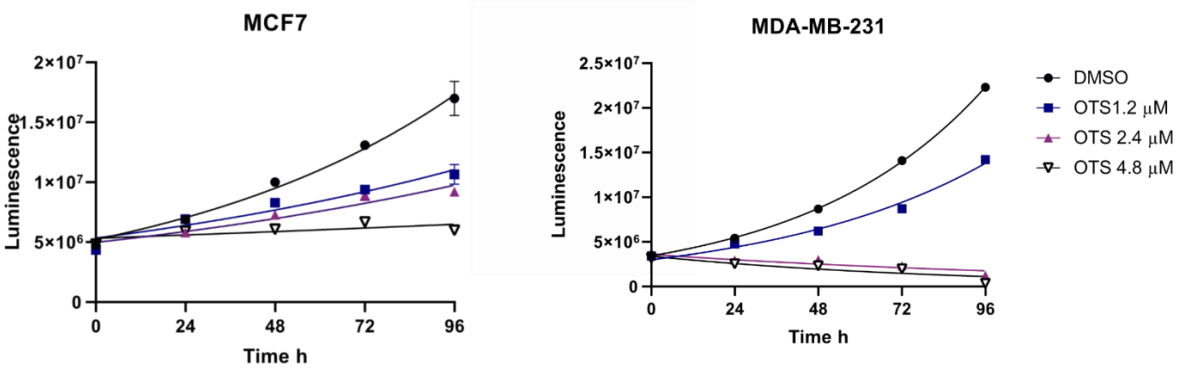


Figure 40: Effect of SUV39H2 inhibitor OTS186935 on BC cell proliferation

Cell proliferation was calculated by Cell Titer Glo and indicated as Luminescence values. MCF7 and MDA-MB-231 were treated for 24, 48, 72, and 96 hours with different concentrations of OTS186935 inhibitor.

5.11.2 Knockdown of H3K9 methyltransferases SETDB1, SUV39H1, and SUV39H2 results in phenotypic effects

MDA-MB-231 cells were infected with lentiviral vectors containing different shRNAs to target each H3K9-specific methyltransferases (SETDB1, SUV39H1, and SUV39H2). Sh1 and sh3 were used for knockdown of SETDB1, sh5 and sh6 were used for knockdown of SUV39H1, and sh9 and sh12 were used for knockdown of SUV39H2. In addition, a pLKO scramble vector was used as a control and to infect both MDA-MB-231 and MCF7 cells (as additional control for less aggressive phenotype). Forty-eight hours after puromycin selection, cells were pelleted for RT-qPCR to assess efficacy of knockdown and for MS

analysis of histone PTMs to assess effect on H3K9me3. Gene expression analysis by RT-qPCR confirmed the reduction of expression upon knockdown (**Figure 41A**). In **Figure 41B** mass spectrometry data are shown for total H3K9me3 levels calculated as the sum of H3K9me3 and H3K9me3K14ac. Upon enzyme knockdown, we found a reduction of H3K9me3 total levels when SUV39H1 and SUV39H2 were targeted, but we observed an increase of H3K9me3 levels upon SETDB1 knockdown. The increase observed upon SETDB1 knockdown might depend on compensation effect from the other H3K9 methyltrasferases or other mechanisms (*see Discussion section*).

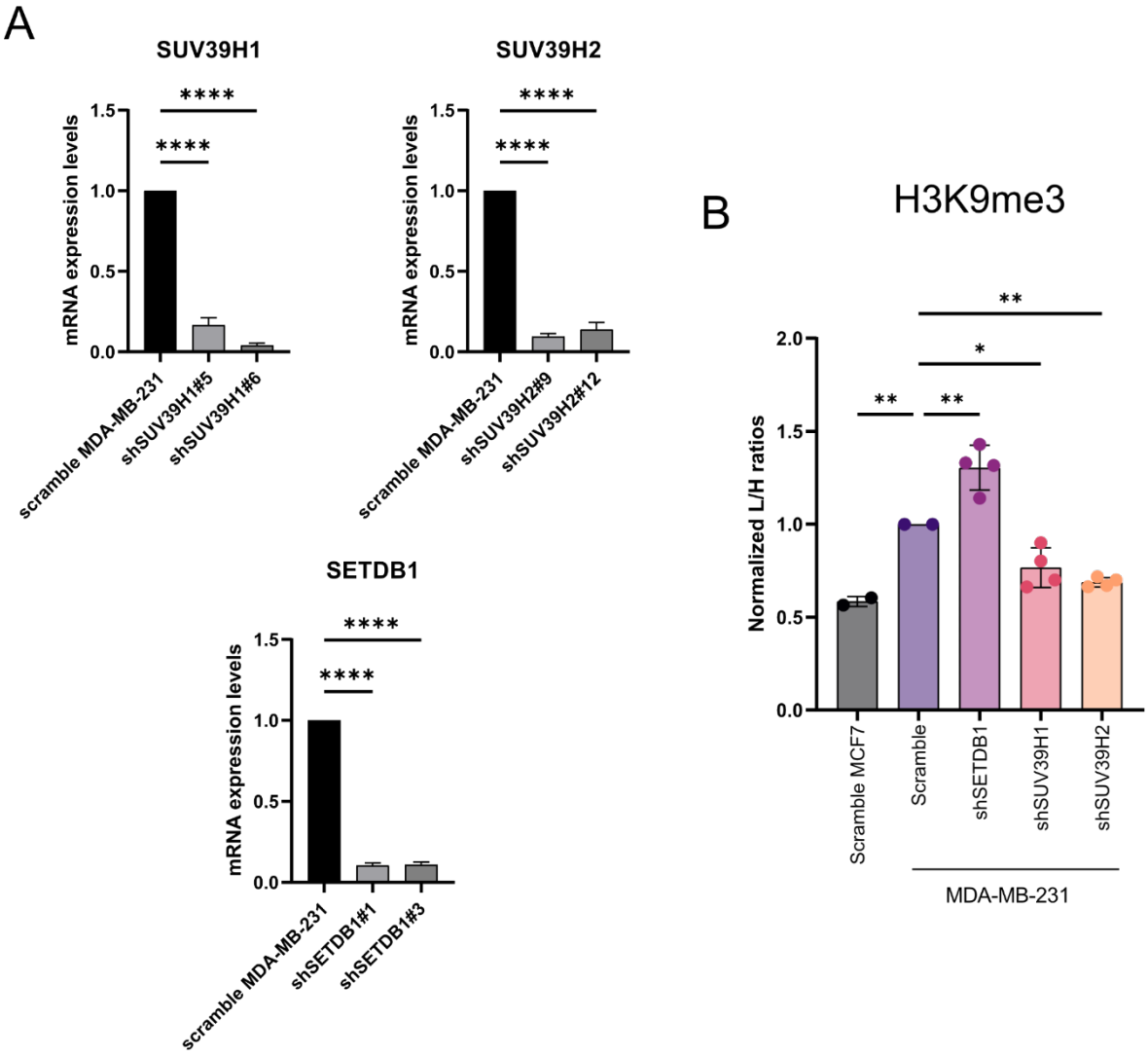


Figure 41: Modulation of H3K9me3 methyltransferases

*(A)RT-qPCR data normalized over GAPDH show mRNA levels compared to scramble vector control in MDA-MB-231 cell line for each shRNA. ****p-value< 0.0001 by t-test. (B)MS analysis of total H3K9me3 in MDA-MB-231 cell line upon knockdown by shRNA of SUV39H1, SUV39H2, and SETDB1. A Scramble vector was used as control in both MCF7 and MDA-MB-231 cell line. Values are displayed as average of H3K9me3 L/H values for shRNA targeting the same enzyme. Data are showed as L/H ratios normalized on MDA-MB-231 scramble control values. p-value was defined by one-way ANOVA *p-value < 0.05, **p-value< 0.01 by comparison with scramble MDA-MB-231.*

We performed phenotypic assays measuring cell proliferation, colony formation, and cell migration upon knockdown of SETDB1, SUV39H1, and SUV39H2. Cell proliferation was assessed by Cell Titer Glo assay at different time points (0, 24, 48,72, 96, and 120h). The shRNA for knockdown of SETDB1 produced a strong effect on cell proliferation already after 24h, while no marked effect was observed upon knockdown of SUV39H1 and SUV39H2 (**Figure 42**).

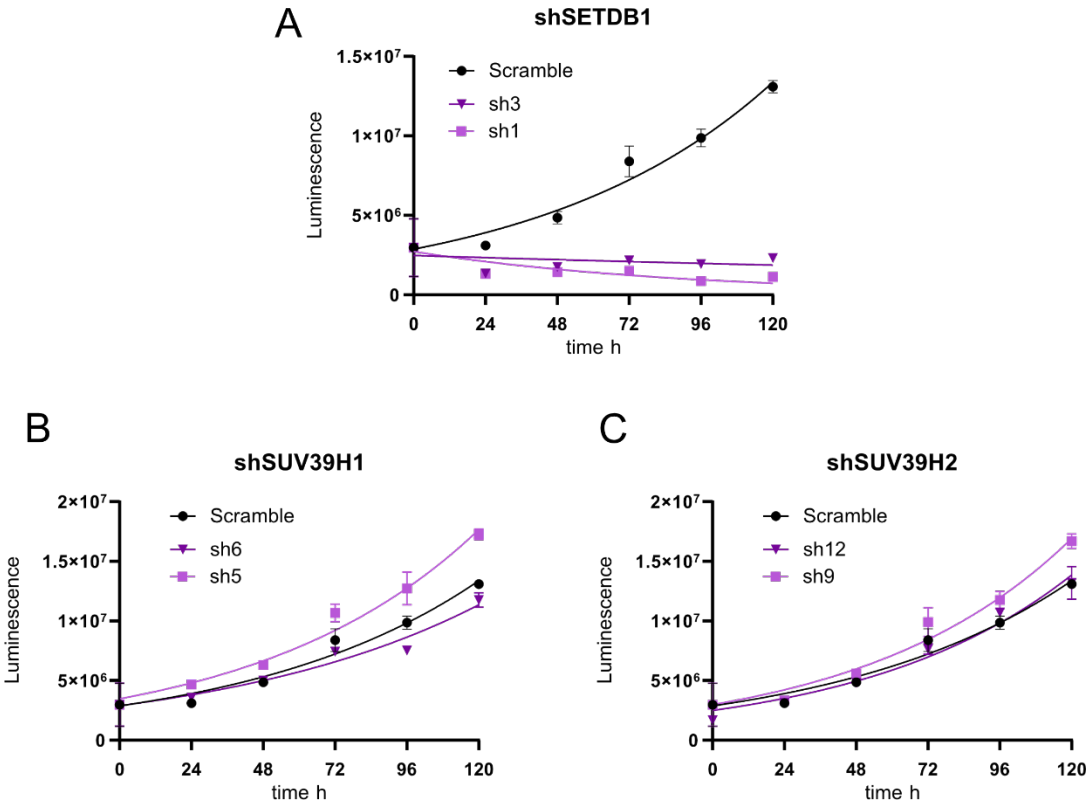


Figure 42: Proliferation assay upon knockdown of SETDB1, SUV39H1, and SUV39H2 in MDA-MB-231 cells

Proliferation assay upon knockdown of SETDB1 (A), SUV39H1 (B), and SUV39H2(C). Proliferation was calculated by Cell Titer Glo assay based on luminescence values at 0, 24, 48, 72, 96, and 120 hours. n=3.

In addition, we evaluated cell migration by wound healing assay, at 12h and 24h post wound creation through Ibidi insert (**Figure 43**). Similar to the results obtained for proliferation, we observed a strong effect on migration of cells with knockdown of SETDB1, while the effects were more variable upon SUV39H1 and SUV39H2 knockdown.

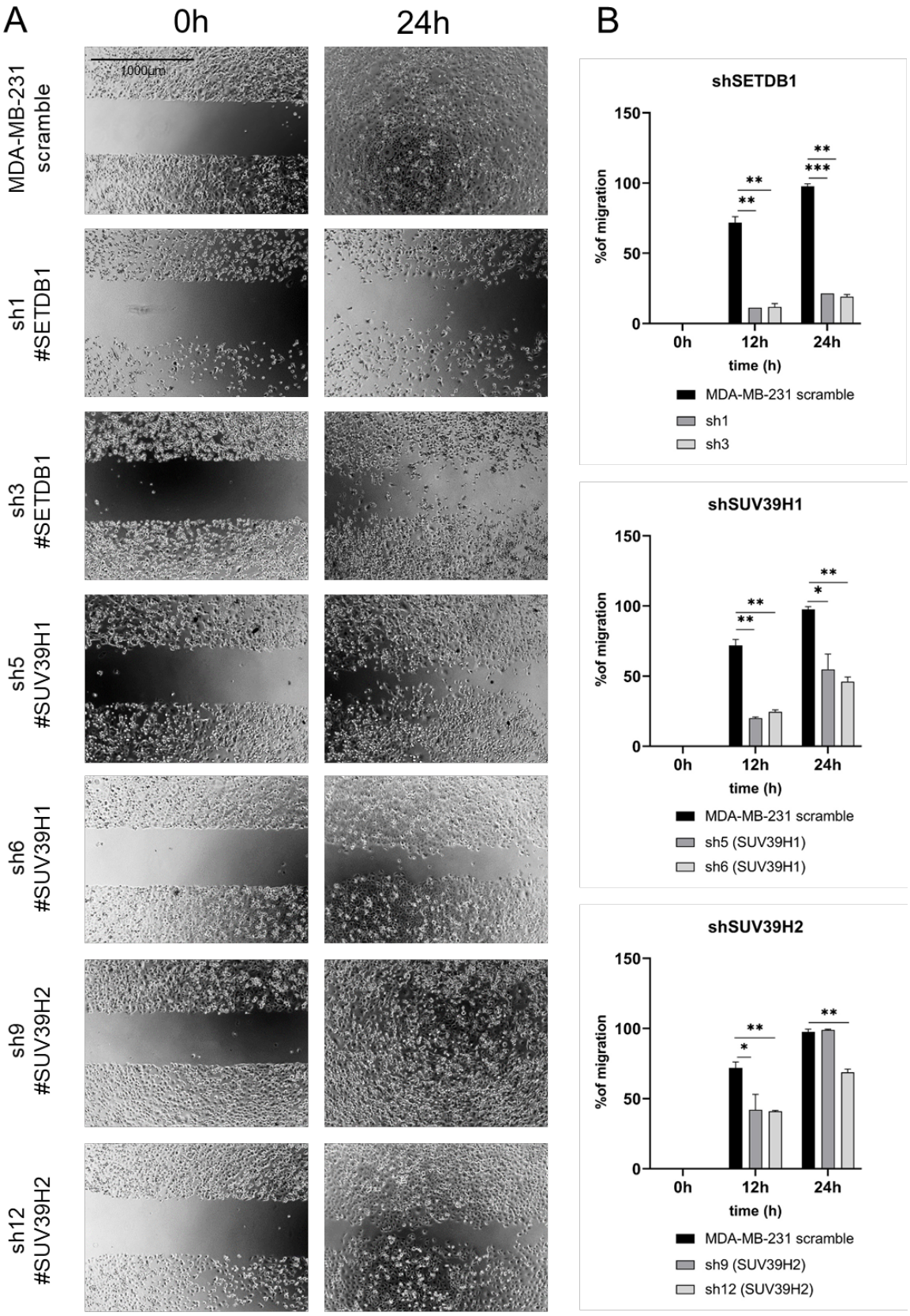


Figure 43:Migration assay upon knockdown of SETDB1, SUV39H1, and SUV39H2

*(A) Pictures taken at Evos 4x at 0h and 24h for scramble as control and upon enzyme knockdown. Scale bar=1000µm (B) Display of percentage of migration calculated at 0, 12 and 24 h. Wound was created by Ibidi insert. n=3. p-value was calculated by t-test at each time point. *p-value<0.05, **p-value<0.01, ***p-value<0.001.*

Interestingly, the most significant changes were observed in clonogenic assay, where a strong effect on colony formation upon knockdown of SETDB1, SUV39H1, and SUV39H2 was observed (**Figure 44**).

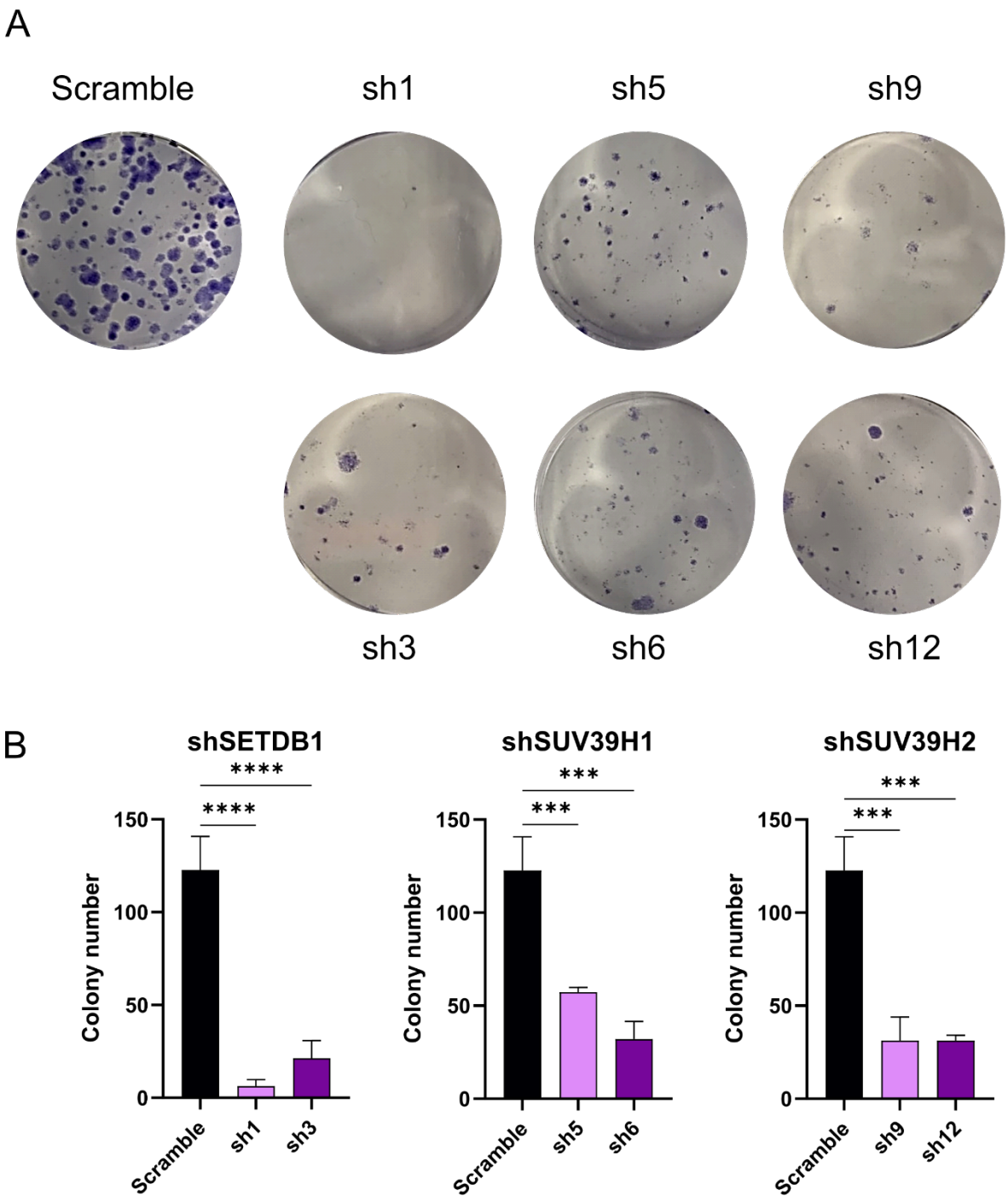


Figure 44:Colony formation assay upon knockdown of SETDB1, SUV39H1, and SUV39H2

*(A) Colony staining by crystal violet. 500 cells were seeded per well in a 6-well plate and colonies were colored after 14 days in culture. (B) Quantification of the number of colonies by ImageJ Fiji tool for colony count. n=3. p-value calculated by one-way ANOVA ***p-value<0.001, ****p-value<0.0001.*

Phenotypic analyses identified a more pronounced effect upon SETDB1 knockdown compared to SUV39H1 and SUV39H2 knockdown. Overall we identified an effect on colony formation and, to a much lesser extent, migration and proliferation. These findings confirm already published results (198,199) and suggest that the phenotypic effects of SETDB1 depletion are likely not due to changes in the level of its target modification H3K9me3, which in fact is not decreased, but may involve effect on other targets on both histone or non-histones. On the contrary, the inhibition of colony formation observed in SUV39H1 and SUV39H2 knockdowns might indeed be associated with the reduction of H3K9me3. Overall, these results indicate that H3K9 methyltransferases are potential targets to decrease cell proliferation and/or migration in TN tumor cells.

5.11.3 Documentation of a potential cross-talk between H3K9me3 and H4K20me3

Because MS is an unbiased method allowing the comprehensive profiling of histone PTMs, we were able to profile the levels of other histone PTMs upon modulation of H3K9me3 methyltransferases. Upon this comprehensive analysis we interestingly found that the reduction of H3K9me3 in MDA-MB-231 depleted of SUV39H1 and SUV39H2 was paralleled by the increase of H4K20me3 levels (**Figure 45**), suggesting a potential interplay between H3K9me3, H3K9 methyltransferases and H4K20me3. This increase was comparable to the levels quantified in the MCF7 Luminal-A cell line, which showed higher levels of H4K20me3 compared with the MDA-MB-231 cells in basal conditions.

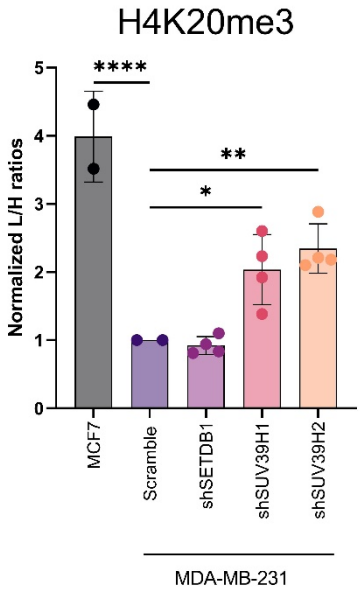


Figure 45(previous page): H4K20me3 levels upon H3K9 methyltransferases knockdown

*(A) MS analysis of H4K20me3 in MDA-MB-231 cell line upon knockdown by shRNA for SUV39H1, SUV39H2, and SETDB1. A Scramble vector was used as control in both MCF7 and MDA-MB-231 cell line. Values are displayed as average of H4K20me3 L/H values for shRNA targeting the same enzyme. Data are showed as L/H ratios normalized on MDA-MB-231 scramble control values. p-value was defined by one-way ANOVA *p-value < 0.05, **p-value< 0.01, ****p< 0.0001 by comparison with scramble MDA-MB-231.*

Moreover, the differential level of H4K20me3 between Luminal A and TN was then confirmed in an expanded panel of cells lines **(Figure 46)**, with sensible lower levels observed in all the TN cell lines investigated. H4K20me3 is a histone mark known to be present in constitutive heterochromatin, where also H3K9me3 is generally deposited. A mechanism of interaction between H3K9me3 and H4K20me3 has been previously reported (200). This mechanism involves the deposition of H3K9me3 via SUV39H1 and SUV39H2, the binding of the heterochromatin protein 1 (HP1) to H3K9me3 mark, and the subsequent recruitment of SUV4-20H enzymes (the enzymes responsible for the deposition of H4K20me3) to establish H4K20me3 in the same chromosomal regions. According to this mechanism, the levels of H3K9me3 and H4K20me3 should move in the same direction. Although this mechanism would not explain the opposite levels that we identified for H3K9me3 and H4K20me3 in both basal conditions and upon modulation of H3K9 enzymes, our hypothesis is that it might account for a redistribution of these histone PTMs on the chromatin.

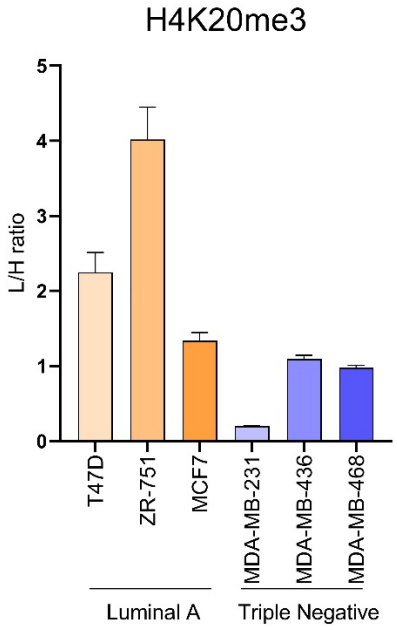


Figure 46 (previous page): H4K20me3 profiling in Luminal A and TN breast cancer cell lines
MS analysis of H4K20me3 in Luminal A (T47D, ZR-751, MCF7) and Triple Negative (MDA-MB-231, MDA-MB-436, MDA-MB-468) breast cancer cell lines displayed as L/H ratios of relative abundances.

5.12 Analysis of the genomic distribution of H3K9me3 and H4K20me3 by chromosome staining and ChIP-seq

5.12.1 Analysis of H3K9me3 and H4K20me3 in chromosome spreads in breast cancer cell lines

To test the hypothesis of a possible re-localization of H3K9me3 and H4K20me3 upon knockdown of H3K9-methyltransferases, we first investigated the distribution of these histone marks in basal conditions in a set of breast cancer cell lines including MCF7 and ZR-751 as Luminal A subtype and MDA-MB-231, MDA-MB-436, and MDA-MB-468 as TN subtype. First, we set up an immunofluorescence experiment on chromosome spreads to assess and visualize H3K9me3 and H4K20me3 localization to investigate their distribution over the chromosome and to verify the reported (201) presence at centromeric and peri-centromeric regions. After metaphase block, we performed the immuno-labeling of chromosome spreads, by a co-staining with H3K9me3 or H4K20me3 antibody and an antibody for CENP-A, as centromere marker. The analysis of H3K9me3 distribution in chromosome spreads revealed a localization at centromeric regions in both Luminal A and TN cells. The quantification of the co-localization with the centromeric marker CENP-A identified a higher presence at centromere in Luminal A samples (around 30%), whereas lower levels were identified in TN subtypes (around 20%) with some cell-dependent variability (**Figure 47**).

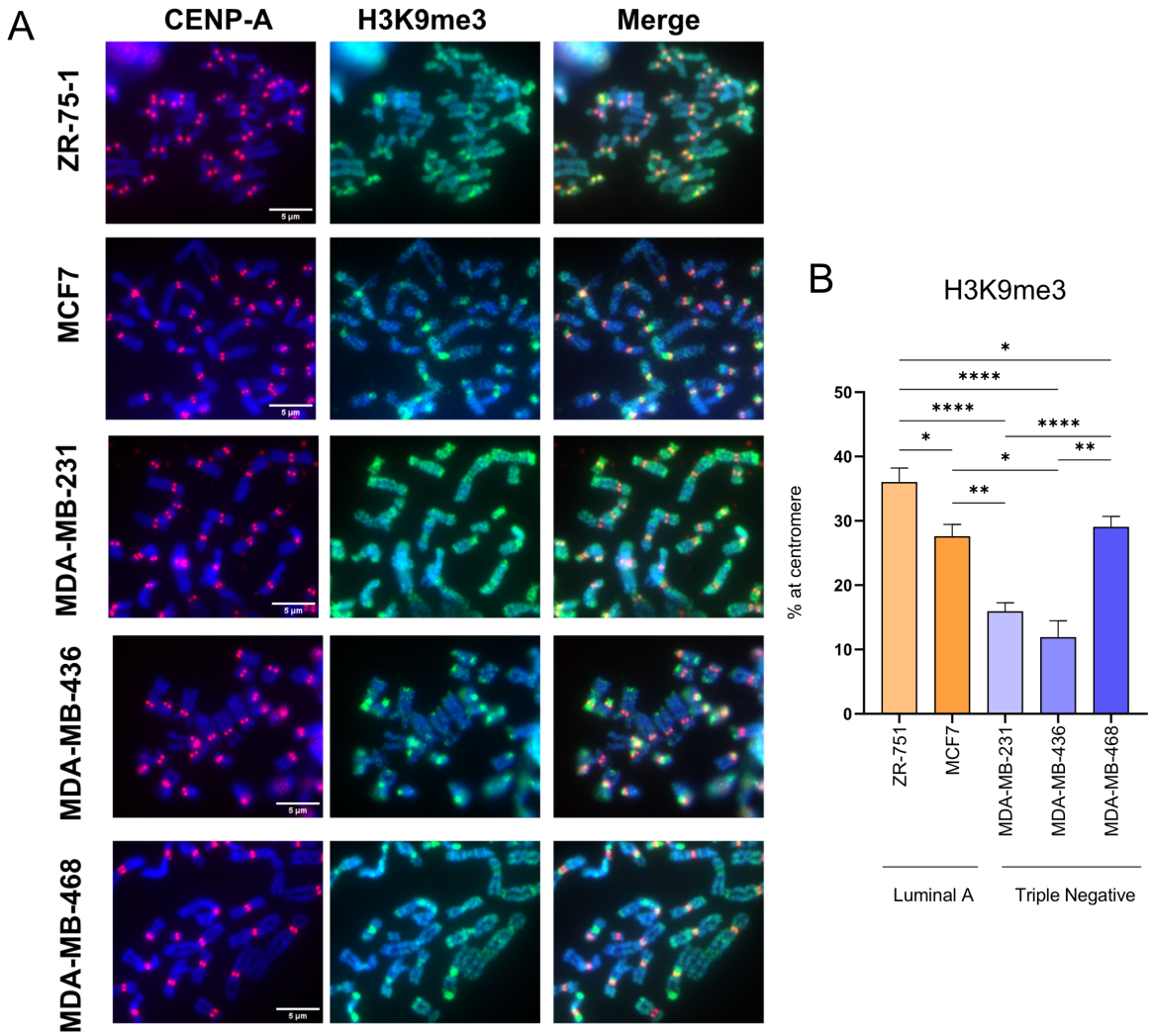


Figure 47: Immunofluorescence of chromosome spreads for H3K9me3

*(A) Images of chromosomes stained with H3K9me3 antibody and CENP-A antibody as centromere marker. Images were taken at 100x magnification. Scale bar= 5µM. (B)Quantification of H3K9me3 localization at centromere by co-localization with CENP-A represented for each cell line. p-value was calculated by ANOVA, *p-value<0.05, **p-value<0.01, ****p-value<0.0001. Average number of metaphases per sample analyzed = 30.*

However, in both cell lines H3K9me3 seemed quite spread along the chromosome suggesting that it might have a role not only in maintaining heterochromatin at centromeric regions but also in regulation of other genomic regions. A more marked difference in distribution at chromosome level was instead observed in the analysis of H4K20me3 localization (**Figure 48**). The immunostaining of H4K20me3 revealed a more pronounced localization in peri-centromeric regions in the two Luminal A cells, around 60%, in comparison with the TN cell lines which displayed around 20% of H4K20me3 at these regions. This analysis suggests that changes in the localization of H4K20me3 are subtype-specific, with a much more diffused distribution in TN samples.

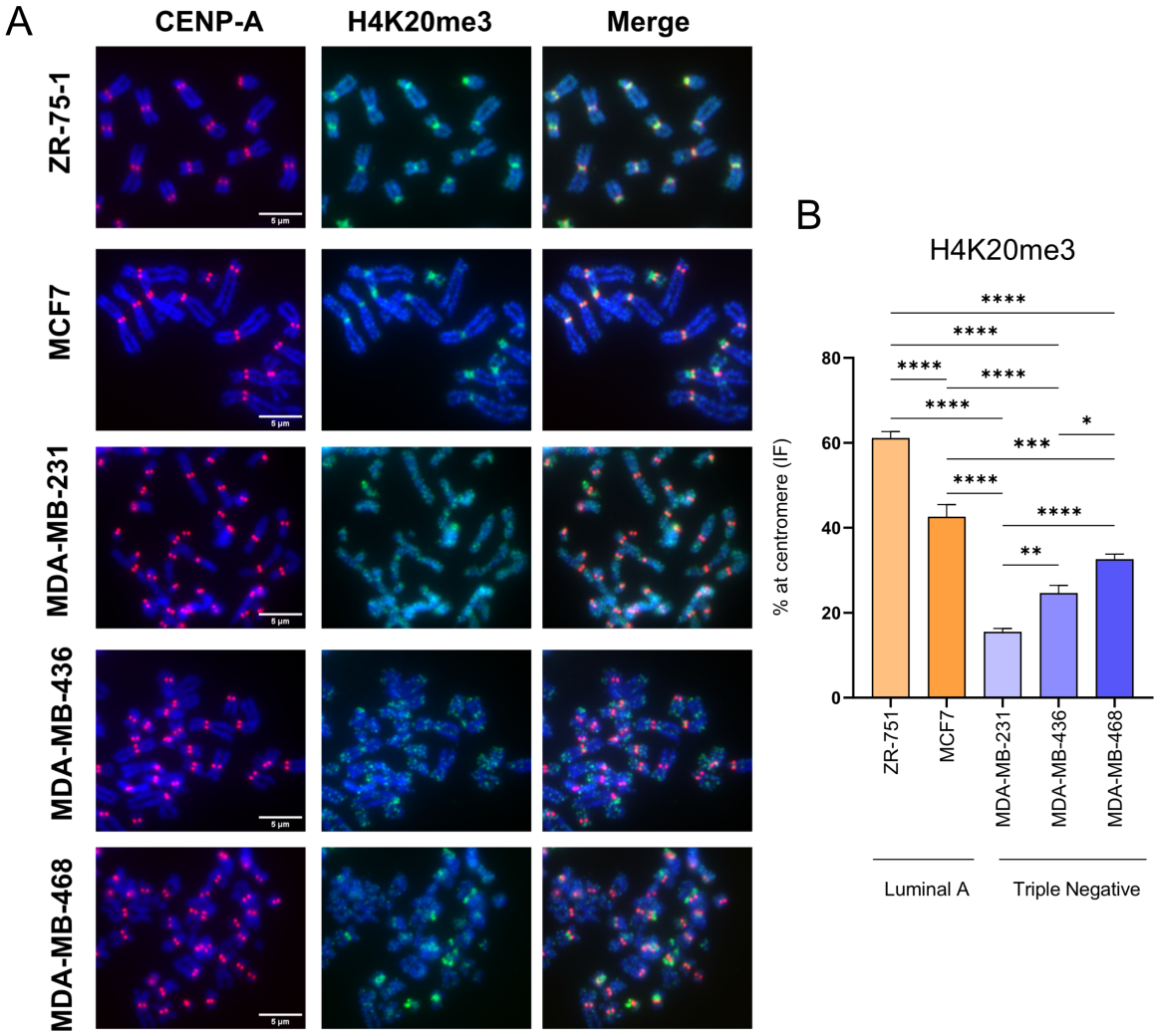


Figure 48**: **Immunofluorescence of chromosome spreads for H4K20me3

*(A) Images of chromosomes stained with H4K20me3 antibody and CENP-A antibody as centromere marker. Images were taken at 100x magnification. Scale bar= 5µM. (B)Quantification of H4K20me3 localization at centromere by co-localization with CENP-A represented for each cell line. p-value was calculated by ANOVA, *p-value<0.05, **p-value<0.01, ***p-value>0.001, ****p-value<0.0001. Average number of metaphases per sample analyzed = 60.*

5.12.2 Investigation of H3K9me3 and H4K20me3 genomic distribution by ChIP-seq analysis in breast cancer cells

Next, for the analysis of distribution at genome wide level we performed a ChIP-seq analysis. H3K9me3 and H4K20me3 are histone marks which are challenging to analyze by ChIP-seq for several reasons. First, these modifications are generally present with a broad distribution that makes it more difficult to associate them to specific genes and therefore to associate them with regulation of gene expression. In addition, they are often associated with repetitive regions and heterochromatin, meaning that most of the reads are lost in the analysis since reads aligned to repetitive regions are generally discarded in a common ChIP-seq analysis workflow. However, thanks to specific analysis tools for these histone PTMs and a higher number of reads sequenced (80 millions), we were able to analyze both H3K9me3 and H4K20me3 distributions in the same cell lines analyzed by immunofluorescence (**Figure 49**).

We found similar H3K9me3 peak distribution in Luminal A and TN cells, with most of the peaks present at gene body, mainly localizing at intronic and intergenic regions (**Figure 49A**). Interestingly, the Triple Negative cells showed higher H3K9me3 distribution at 5'UTR and exons compared to Luminal A cells. In addition, exploiting our ChIP-seq data, we also analyzed the localization of H3K9me3 at centromeric regions by directly mapping the sequenced reads at these specific genomic regions and by comparison of reads from the IP and from the INPUT chromatin (**Figure 49D**). This result was in line with the results obtained by the immunofluorescence experiment, with a higher number of reads at centromeric regions for Luminal A cells compared with TN cell lines. The same analysis was performed for H4K20me3 distribution. (**Figure 49C**). As in the case of H3K9me3 most of H4K20me3 peaks were present at intronic and intergenic regions. Interestingly, MDA-MB-231 cells were the most divergent in terms of distribution, with a large portion of peaks at exons and introns, and a reduced presence at intergenic regions. As in the case of H3K9me3, we aligned the reads with centromeric regions and obtained similar results to those achieved by immunofluorescence, where Luminal A cells displayed higher presence of H4K20me3 at centromeres (**Figure 49E**). These results confirm not only a different distribution of H3K9me3 and H4K20me3 in TN cells, but also suggest that the localization at centromeric regions is present not only in cells blocked in metaphase (as displayed by immunofluorescence), but also in the asynchronous cells analyzed through ChIP-seq approach, suggesting that the localization of these modifications at heterochromatic regions is not cell cycle dependent.

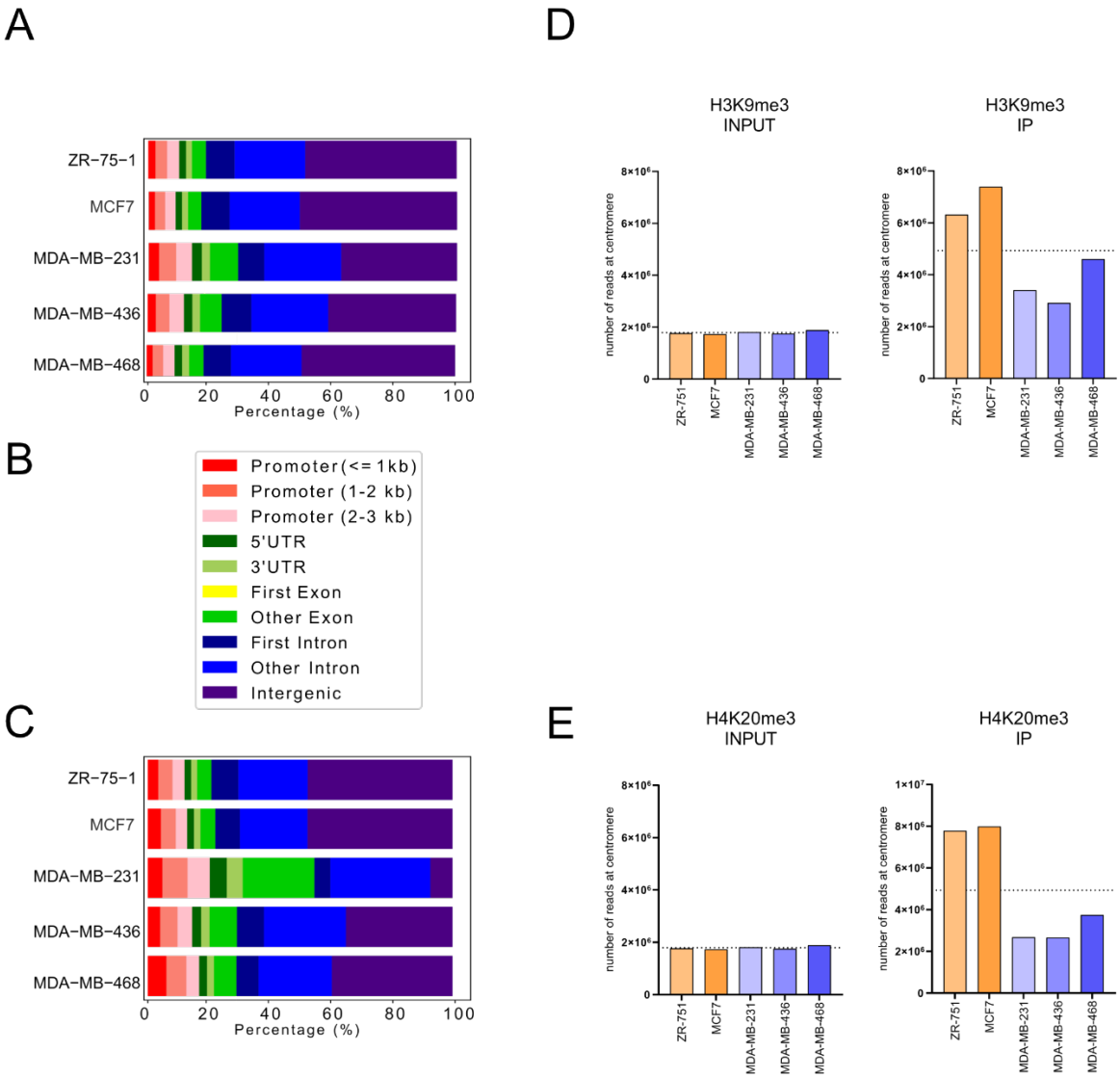


Figure 49: Investigation of H4K20me3 distribution in breast cancer cell lines by ChIP-seq

(A) H3K9me3 peak distribution in two Luminal A (ZR-751 and MCF7) and three TN (MDA-MB-231, MDA-MB-436, and MDA-MB-468) cell lines. Colors corresponding to specific genomic regions are described in the figure legend in B. (C) H4K20me3 peak distribution in two Luminal A (ZR-751 and MCF7) and three TN (MDA-MB-231, MDA-MB-436, and MDA-MB-468) cells as in A. (D)Analysis of H3K9me3 mapped reads at centromere for each breast cancer cell line in INPUT and IP samples. (E) Analysis of H4K20me3 mapped reads at centromere for each breast cancer cell line in INPUT and IP samples.

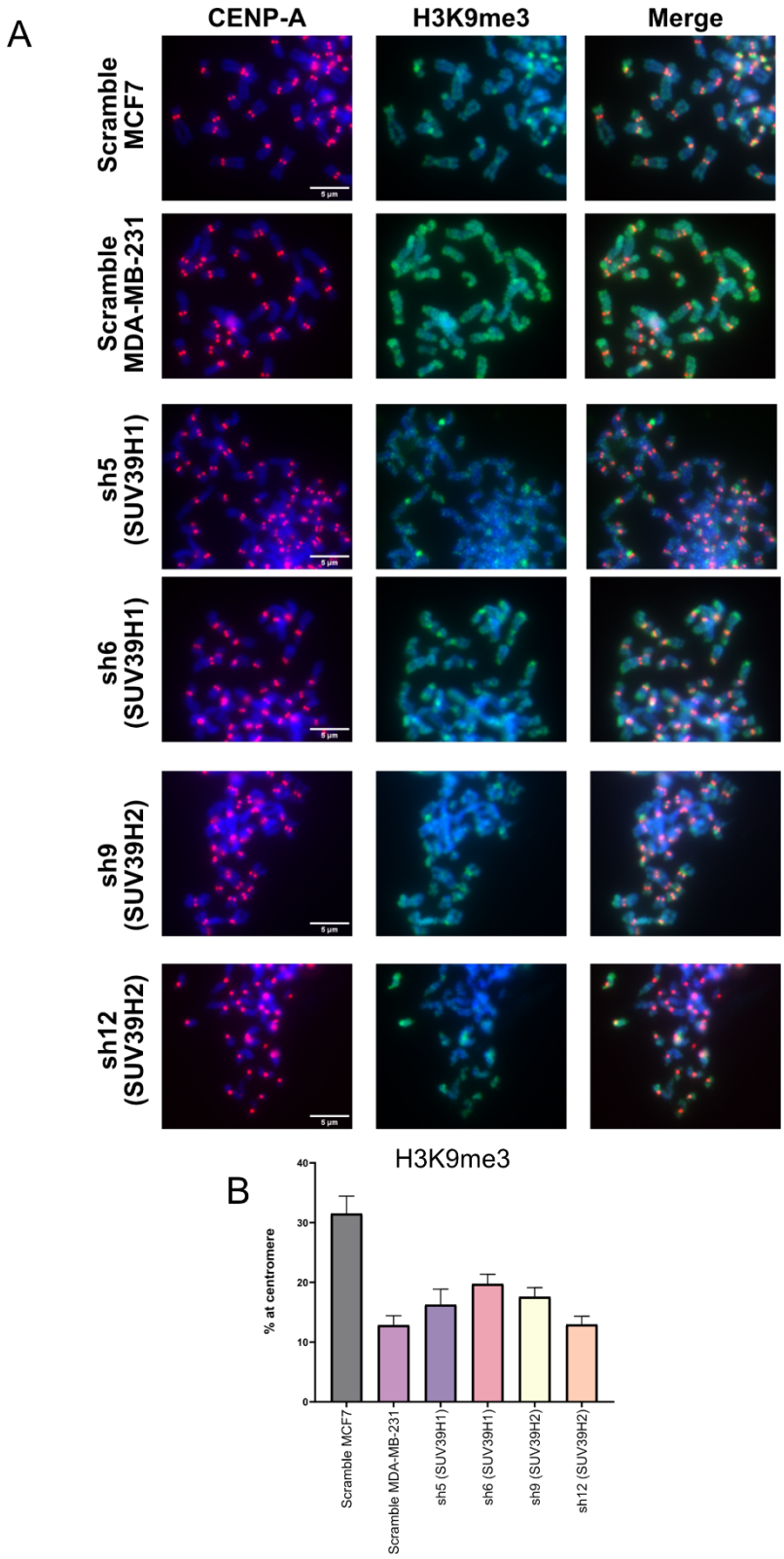
5.12.3 Analysis of H3K9me3 and H4K20me3 localization upon knockdown of the H3K9-specific methyltransferases SUV39H1 and SUV39H2

After observing a different localization of H3K9me3 and H4K20me3 in TN and Luminal A cell lines in basal conditions, we investigated whether the localization of these marks could also change as a consequence of SUV39H1 and SUV39H2 depletion in MDA-MB-231 and MCF7 cells, on the basis of our hypothesis of a possible re-localization of H3K9me3 and H4K20me3 upon modulation of SUV39H1 and SUV39H2.

Upon immunofluorescence of chromosome spreads, we identified an increased trend in the localization of H3K9me3 at centromeres in SUV39H1- and SUV39H2-depleted cells compared to cells infected with the scramble control (**Figure 50**), suggesting that this modification is probably lost in other genomic regions but maintained at the centromere, with some variability. However, to further corroborate this hypothesis, we will need to acquired additional data on H3K9me3 localization for which the analysis is ongoing.

Figure 50 (next page): Investigation of H3K9me3 distribution by immunofluorescence upon SUV39H1 and SUV39H2 knockdown

(A) Immunofluorescence of chromosome spreads at 100x magnification taken by Multifluo Leica microscope. CENP-A was used as centromere marker. Scale bar= 5μM. (B)Quantification of H3K9me3 localization at centromere by co-localization with CENP-A staining represented for controls (scramble MCF7 and scramble MDA-MB-231) and cells upon knockdown of SUV39H1 and SUV39H2.



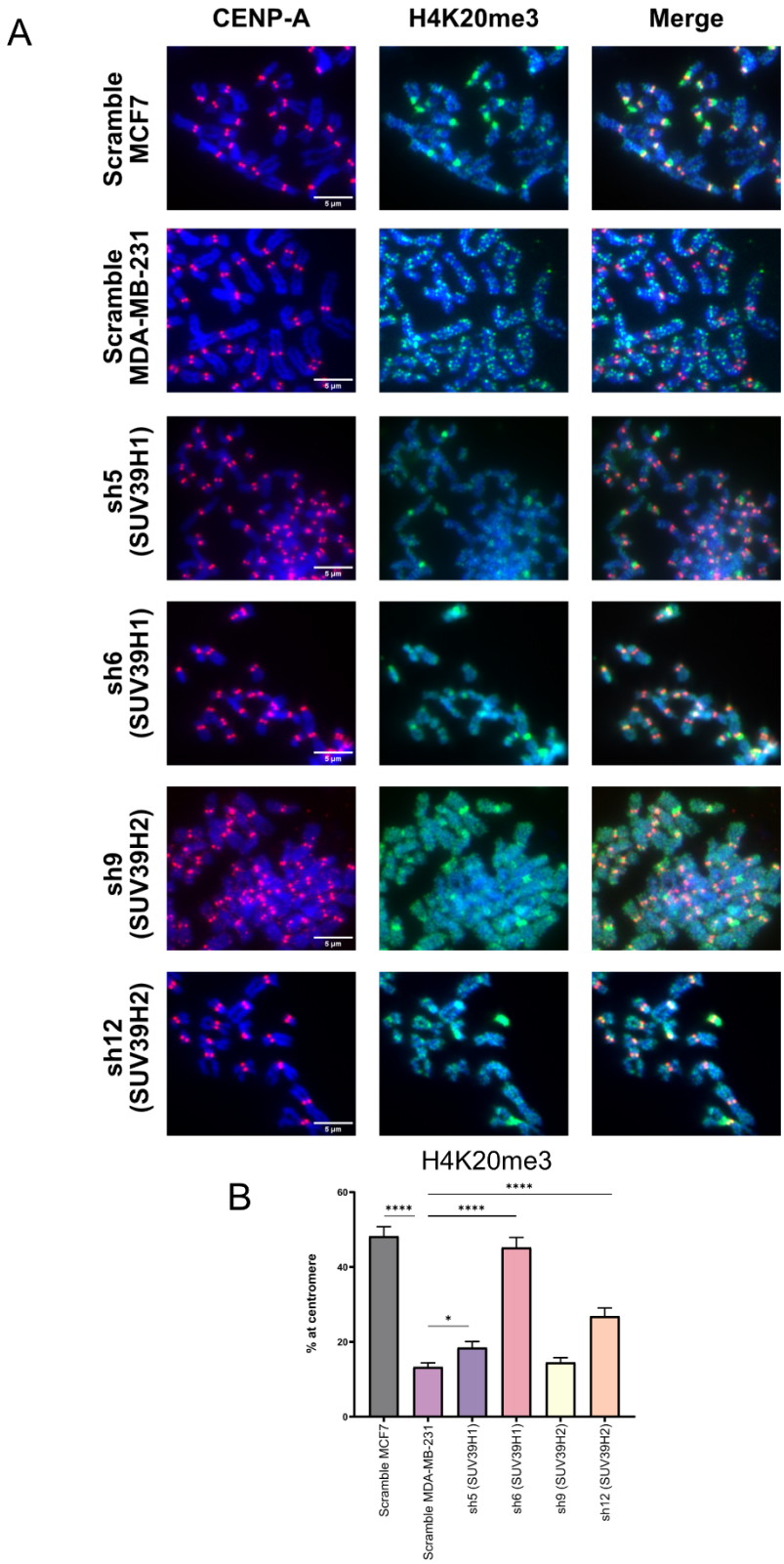
Interestingly, we obtained more consistent and significant results for H4K20me3 analysis (**Figure 51**). The depletion of SUV39H1 and SUV39H2 in MDA-MB-231 cells resulted in re-localization to peri-centromeric regions, exhibiting a pattern more similar to the MCF7 cell line, especially in the case of SUV39H1 knockdown. Of note, the shRNA leading to higher proliferation rates by phenotypic assay was also the one with less H4K20me3 at

centromeres, suggesting that H4K20me3 localization might be involved in cellular proliferation in the TNBC subtype.

Overall, the results achieved so far are all indicative of a potential cross-talk between H3K9me3, H4K20me3 and H3K9-methyltransferases which we still have to clarify. Upon modulation of SUV39H1 and SUV39H2, these two modifications show a similar behavior in terms of localization at centromere, suggesting that the knockdown of SUV39H1 and SUV39H2 might contribute to a re-localization of these histone PTMs. In addition, the alteration of these methyltransferases in TN tumors represents a potential cause of the altered levels of the H4K20me3 marker, whose decrease in this subtype was otherwise unexplained. This also suggests that targeting the activity of SUV39H1 and SUV39H2 methyltransferases might not only influence H3K9me3 levels but also H4K20me3 levels, potentially producing a more pronounced effect, especially at phenotypic level.

Figure 51 (next page): Investigation of H4K20me3 distribution by immunofluorescence upon SUV39H1 and SUV39H2 knockdown

*(A) Immunofluorescence of chromosome spreads at 100x magnification taken by Multifluo Leica microscope. CENP-A as centromere marker. Scale bar= 5µM. (B)Quantification of H4K20me3 localization at centromere by co-localization with CENP-A staining represented for controls (scramble MCF7 and scramble MDA-MB-231) and cells upon knockdown of SUV39H1 and SUV39H2. p-value was calculated by ANOVA *p<0.05, ****p<0.0001.*


A
CENP-A
H4K20me3
Merge
Scramble MCF7
Scramble MDA-MB-231
sh5 (SUV39H1)
sh6 (SUV39H1)
sh9 (SUV39H2)
sh12 (SUV39H2)
5 µm
B
H4K20me3
% at centromere
60
40
20
0
****
****
****
*
Scramble MCF7
Scramble MDA-MB-231
sh5 (SUV39H1)
sh6 (SUV39H1)
sh9 (SUV39H2)
sh12 (SUV39H2)

6. DISCUSSION

The MS-based profiling of histone PTMs previously performed in our laboratory identified an epigenetic signature distinguishing the breast cancer Triple Negative subtype from the other subtypes and in particular from the Luminal A subtype, which was the most divergent, not only in terms of prognosis but also from the overall epigenetic landscape. Based on these preliminary results, which represented the starting point of my Ph.D. project, the aim of this study was to dissect the molecular mechanisms involved in the epigenetic changes observed in the TNBC subtype using different strategies.

One of the most interesting histone mark displaying a significantly increased level in TNBC subtype compared to others was H3K4me2. This epigenetic mark is known to be associated with active transcription and generally localizes at active cis-regulatory regions, including both promoters and enhancers. Even though MS allows the unbiased and comprehensive quantification of bulk histone PTMs, it only provides information on global level of such modifications on chromatin, lacking any information about their locus/region-specific genomic localization.

To investigate the downstream effects of the changes observed in H3K4me2 levels, we exploited the complementarity of epi-proteomic data with more conventional epi-genomic and transcriptomic methods (ChIP-seq and RNA-seq), in order to understand and investigate H3K4me2 genomic localization and its effects at gene expression level. It must be noted that in the epi-proteomic profiling of histone PTMs we also observed an increase in the levels of H3K4me3. This histone mark is localized around the TSS in promoter regions and is strongly associated with active transcription. Although H3K4me3 is more thoroughly characterized compared to H3K4me2, its ChIP-seq profiles did not seem to discriminate TNBC and Luminal A samples, as in the case of H3K4me2. This somewhat surprising result could be explained by the fact that the qualitative analysis, which we performed, i.e. presence/absence of the peaks, did not show any difference. Therefore, we hypothesized that the actual difference is not in the presence/absence of peaks, but in the quantitative H3K4me3 levels at promoters. However, a quantitative analysis, i.e employing a *Drosphila* chromatin spike-in, was not possible, since the *Drosophila* chromatin is immunoprecipitated much more efficiently compared to the chromatin from clinical samples, thus masking the signal of our samples.

The multi-OMICs analysis of a small set of clinical samples that we carried out showed that genes displaying H3K4me2 peaks at either promoters or super-enhancers uniquely in TN samples presented higher expression levels in this subtype. Remarkably, this correlation not

only applied when H3K4me2 was present at promoters, but also when it localized at gene bodies and intergenic regions, and particularly at super-enhancers.

Recently, super-enhancers have been shown to play a critical role in driving the upregulation of cancer driver genes across various cancer types, including breast cancer, and contribute to the definition of subtype-specific characteristics (125). In our study, we assessed a direct link between the expression of potential genes involved in TNBC phenotype and the presence of H3K4me2 at their promoters by the use of CRISPRi. Candidate genes to follow up for validation of our model and functional characterization of their role in TN phenotype were selected based on increased expression in clinical samples, as well as based on their potential role in the context of cancer. Therefore, we first started, as a more explorative analysis, with modulation of H3K4me2 at promoter regions for the genes PLA2G4A, IGF2BP3, and SLAMF7 in MDA-MB-231 cells, which resulted in a direct association of H3K4me2 and gene expression. Interestingly, by the multi-OMICs analysis performed in clinical samples, targets carrying H3K4me2 at enhancer regions were also identified. Among these genes, we found CTL4A, which is a cytotoxic antigen for lymphocyte T cells and which has been previously described to be expressed in TNBC suggesting also the potential use of immune-checkpoint inhibitors in this type of tumor (202). We also identified ANLN, an actin-binding protein that plays a role in cell growth and migration, and that has been described to be associated with poor survival and recurrence. This gene was previously described as a TNBC-specific gene, regulated by super-enhancers (203). However, the modulation of H3K4me2 at these regions is more challenging, because it is more difficult to associate a super-enhancer region to the regulation of a specific gene. A selection of the best candidates is ongoing in the model breast cancer cells MDA-MB-231.

Because our data suggests that reducing the levels of H3K4me2 could reduce the expression of several targets linked with the TN aggressive phenotype, we are also in the process of testing several inhibitors targeting enzymes responsible for deposing H3K4me2, such as MLL methyltransferases. While these inhibitors have often been used in the context of acute myeloid leukemia, ongoing research is expanding their potential applications to solid tumors. It is worth noting that these enzymes often have more than one target and may not be entirely specific for H3K4me2. In addition, the enzymatic activity of H3K4 methyltransferases on its own is known to be very weak, while it is augmented by the binding of several cofactors. Therefore, the majority of MLL inhibitors have been developed to target directly the binding of these co-factors, such as WDR5 and Menin, with the MLL methyltransferases (204). These H3K4-specific methyltransferases have an effect on both H3K4me2 and H3K4me3 levels. Therefore, the phenotypic effect that we observed might derive from both H3K4me2

and H3K4me3 depletion. However, in our case it is not an issue, since increased levels of H3K4me3 were also found in TN compared with other BC subtypes.

One potential limitation of our study is that we are analyzing bulk tissues, and clinical samples can be quite heterogeneous and contain various cell types, which can make more complex the interpretation of results. This is particularly evident in cases like TNBC, where immune infiltration, such as Tumor-Infiltrating Lymphocytes (TILs), can play a significant role in shaping the disease phenotype. In our datasets, by RNA-seq analysis we identified a large number of genes involved in immune-response, which may be suggestive of an intricate interplay between TNBC cancer cells and infiltrating lymphocytes (188). To gain a more precise understanding of the role of H3K4me2 identified in TNBC, it would be beneficial to conduct spatial analyses through laser microdissection -to isolate specific cell populations- for the subsequent extraction and MS-analysis of both histones and total proteomes. However, it is essential to consider that working with low-abundance samples presents a number of challenges, including the need for protocols tailored for low sample amounts and highly sensitive analytical methods. While for MS analysis we have already demonstrated that it is possible to quantify histone PTMs from 1000 cells (156), and also RNA-seq is possible from low abundance samples, the analysis of low abundance samples by ChIP-seq is far more challenging. Recent methods such as Cut & Tag have been developed and allow analysis of histone PTMs from as low as 5000 cells; since this method has already been developed for the common histone PTMs (such as H3K27me3 and H3K4me3), we are planning to test it on the modifications of interest (H3K4me2, H3K9me3 and H4K20me3) in our laboratory.

As a second strategy, to find potential therapeutic targets and better understand the mechanisms at the base of deposition of aberrant histone PTMs in TNBC, we studied upstream mechanisms by correlating histone PTM and histone modifying enzyme levels. In most of the cases, we could not find any correlation between histone modifying enzyme levels and histone PTM levels in the TNBC subtype. This has been observed in different tumors, including breast, pancreatic, and ovarian cancer where levels of EZH2 were opposite to decreased levels of H3K27me3, as well as in a study performed by our group by comparison of tumor samples with normal tissues (55). Even though transcript levels of histone modifying enzymes are frequently used as indicators for histone PTM abundances, specific PTMs might be regulated independently of the expression levels of the enzymes responsible for their deposition and removal (205). For instance, histone modifying enzymes

could be characterized by deletions, amplifications, point mutations, and translocations, and in addition these enzymes are known to be part of multi-subunit complexes, where more than one enzyme is often present and which could as well influence their activities.

We, however, found an interesting correlation between the increase of H3K9me3 levels and the increased levels of the H3K9-specific methyltransferases SETDB1, SUV39H1 and SUV39H2 in TNBC samples. These enzymes have been previously investigated in various studies, in multiple tumor types, showing relevant effects at phenotypic level, both in culture models and on tumor growth, *in vivo*. In breast cancer, knockdown of SETDB1 resulted in suppression of cell proliferation, cell cycle progression and migration (198), while overexpression of SUV39H1 in breast and colorectal cancer cells resulted in increased cell migration (206). In a different study performed on hepatocellular carcinoma cells, depletion of SUV39H1 and SUV39H2 resulted in effect on cell growth (199) . However, in none of these studies the observed phenotype was correlated with an effective molecular characterization, for instance with the measurement of the levels of H3K9 methylation upon enzyme depletion, which we instead performed. While our knockdown experiments confirmed overall these phenotypic effects, the MS-based profiling of histone PTMs showed that SUV39H1 and SUV39H2 depletion leads to the expected decrease of the known target H3K9me3, this is not the case for SETDB1, which, despite causing the strongest phenotypic effects in terms of proliferation, migration, and colony formation, was not associated with H3K9me3 decrease. This unexpected result was not reported before, because site-specific measurement of this mark upon SETDB1 genetic depletion was never carried out in combination with cellular assays. This suggests that the strong phenotypic effects of this enzyme might not depend from the levels of H3K9me3 (which is probably compensated by the increased activity of endogenous SUV39H1 and SUV39H2), but rather from other mechanisms such as the modulation of other modification on both other histone or non-histone targets. To identify these potential targets, a methylproteome-based study would be highly informative and help to uncover which non-histone proteins undergo methylation changes upon SETDB1 depletion and may mediate its phenotypic consequences. Given the strong effect of SETDB1 on cell viability, using an inducible system for SETDB1 knockdown would be a better choice as this would allow for precise control of SETDB1 levels, minimizing the impact on cell viability. In addition, other mechanisms might contribute to the phenotypic effect observed and could involve the presence of other complexes or interactors with a role independent from the catalytic activity of this enzyme. For instance, SETDB1 has been demonstrated to interact and form complexes with other proteins including HDACs (207), HP1 (208), and DNMTs (209). In addition, experimental

evidence has shown that SETDB1 seems involved in methylation of free histones and it mainly methylates histone H3 in euchromatic regions while SUV39H1 and SUV39H2 bind mainly at pericentric heterochromatin (210). Therefore, the changes observed in H3K9me3 levels might also be linked to a different localization of these enzymes at chromatin level.

Interestingly, the knockdown of SUV39H1 and SUV39H2, but not SETDB1, was associated with a reproducible increase of H4K20me3 levels, suggesting a potential crosstalk between H3K9me3 and H4K20me3. A crosstalk between these two modifications has been already reported in the literature (19,200). According to the proposed model, the H3K9me3 mark deposed by SUV39H1 and SUV39H2 acts as a binding platform for HP1 proteins, which in turn recruit SUV4-20H enzymes to depose H4K20me3. This model would imply that the levels of these two histone PTMs would go in the same direction, which however is not what we observe in our case.

One possible explanation is that the modulation of K9 methyltransferases could result in the re-localization of H3K9me3, which could have an effect on the recruitment of HP1 and as a consequence affect the presence of H4K20me3. Immunostaining and ChIP-seq experiments for H3K9me3 and H4K20me3 in Luminal A and TN cells in basal condition revealed a much higher presence of these histone marks at centromeric regions in the Luminal A subtype. A marked re-localization of H4K20me3 was also observed upon knockdown of SUV39H1 and SUV39H2. We are currently investigating the role of HP1 protein in this potential cross-talk mechanism, by immunofluorescence of cells in basal conditions and in cells modulated for SUV39H1 and SUV39H2. Indeed, HP1 was shown to play a role in the maintenance of heterochromatin and described as a regulator of transcription (204) . Therefore, exploring the localization and interplay between SUV39H1, SUV39H2, HP1, H3K9me3 and H4K20me3 could provide valuable insights into the regulatory dynamics of heterochromatin for which no clear mechanisms have been described. Other future experiments could focus on modulating K9 specific-demethylases (KDM3B and PHF2), whose levels were also found to correlate with the altered H3K9me3 levels observed in TNBC.

The different breast cancer subtypes are characterized by different proliferation rates, in addition we observed changes in proliferation upon modulation of K9-specific methyltransferases. Therefore, to evaluate whether the changes observed in histone PTMs levels are dependent on different proliferation, we expanded previous acquired data with new experiments based on the analysis of histone PTM changes at the G2/M phase of the cell cycle. From the analysis performed in synchronized cells and MDA-MB-231FUCCI,

the levels of histone PTMs identified in the TNBC signature seemed not associated with G2/M phase of cell cycle except in the case of H3K27me3. Overall, this analysis ruled out a correlation between the histone PTMs we focused on in this study and the mitotic state (G2/M) of the cells.

The use of model breast cancer cell lines was a necessary step in our research plan. We have selected the most representative cell lines, even though it is well known that immortalized cells in culture are not representative of patient derived tissues and other models such as primary cells and organotypic cultures might be more representative of the primary tumor. In our case, the cell lines that were more similar to clinical samples were MCF7 as Luminal A and MDA-MB-231 as TN both in terms of histone PTM levels and distribution at genome wide level of selected histone modifications. This was also the limiting part for the selection of potential target genes (161). However, the use of mouse xenografts would be a better option, and as previously reported the use of CRISPRi in *in vivo* models would be advantageous for investigation of potential targets (194).

Taken together, our results show how starting from the descriptive and quantitative profiling of histone PTM changes by MS, it was possible to investigate and describe molecular mechanisms linked to these changes. We discovered that H3K4me2 regulates specific targets and this could open the possibility of testing novel therapeutic strategies to tackle H3K4me2 levels. Then we also identified an interesting potential cross-talk between the methyltransferases specific to H3K9me3 and H4K20me3. The alteration of these methyltransferases in TN tumors represents a potential cause of the abnormal levels of the H4K20me3 marker, the decrease of which was otherwise unexplainable in TN tumors. This also suggests that targeting the activity of the methyltransferases SUV39H1 and SUV39H2 may not only restore the levels of the specific target H3K9me3 but also those of H4K20me3, potentially leading to a more pronounced phenotypic effect.

7. References

1. Waddington CH. The epigenotype. 1942. Int J Epidemiol **2012**;41:10-3
2. Zhang L, Lu Q, Chang C. Epigenetics in Health and Disease. Adv Exp Med Biol **2020**;1253:3-55
3. Moore LD, Le T, Fan G. DNA methylation and its basic function. Neuropsychopharmacology **2013**;38:23-38.PMC3521964
4. Clapier CR, Iwasa J, Cairns BR, Peterson CL. Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. Nat Rev Mol Cell Biol **2017**;18:407-22.PMC8127953
5. Meller VH, Joshi SS, Deshpande N. Modulation of Chromatin by Noncoding RNA. Annu Rev Genet **2015**;49:673-95
6. Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature **1997**;389:251-60
7. Cutter AR, Hayes JJ. A brief review of nucleosome structure. FEBS Lett **2015**;589:2914-22.PMC4598263
8. Morrison O, Thakur J. Molecular Complexes at Euchromatin, Heterochromatin and Centromeric Chromatin. Int J Mol Sci **2021**;22.PMC8268097
9. Huang H, Lin S, Garcia BA, Zhao Y. Quantitative proteomic analysis of histone modifications. Chem Rev **2015**;115:2376-418.PMC4502928
10. Audia JE, Campbell RM. Histone Modifications and Cancer. Cold Spring Harb Perspect Biol **2016**;8:a019521.PMC4817802
11. Chervona Y, Costa M. Histone modifications and cancer: biomarkers of prognosis? Am J Cancer Res **2012**;2:589-97.PMC3433108
12. Khan SA, Reddy D, Gupta S. Global histone post-translational modifications and cancer: Biomarkers for diagnosis, prognosis and treatment? World J Biol Chem **2015**;6:333-45.PMC4657128
13. Kouzarides T. Chromatin modifications and their function. Cell **2007**;128:693-705
14. Di Cerbo V, Schneider R. Cancers with wrong HATs: the impact of acetylation. Brief Funct Genomics **2013**;12:231-43
15. Millan-Zambrano G, Burton A, Bannister AJ, Schneider R. Histone post-translational modifications - cause and consequence of genome function. Nat Rev Genet **2022**;23:563-80
16. Santos-Rosa H, Schneider R, Bannister AJ, Sherriff J, Bernstein BE, Emre NT, Schreiber SL, Mellor J, Kouzarides T. Active genes are tri-methylated at K4 of histone H3. Nature **2002**;419:407-11
17. Black JC, Van Rechem C, Whetstine JR. Histone lysine methylation dynamics: establishment, regulation, and biological impact. Mol Cell **2012**;48:491-507.PMC3861058
18. Wiles ET, Selker EU. H3K27 methylation: a promiscuous repressive chromatin mark. Curr Opin Genet Dev **2017**;43:31-7.PMC5447479
19. Schotta G, Lachner M, Sarma K, Ebert A, Sengupta R, Reuter G, Reinberg D, Jenuwein T. A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. Genes Dev **2004**;18:1251-62.PMC420351
20. Vardabasso C, Hasson D, Ratnakumar K, Chung CY, Duarte LF, Bernstein E. Histone variants: emerging players in cancer biology. Cell Mol Life Sci **2014**;71:379-404.PMC4025945
21. Corujo D, Buschbeck M. Post-Translational Modifications of H2A Histone Variants and Their Role in Cancer. Cancers (Basel) **2018**;10.PMC5876634
22. Long M, Sun X, Shi W, Yanru A, Leung STC, Ding D, Cheema MS, MacPherson N, Nelson CJ, Ausio J, Yan Y, Ishibashi T. A novel histone H4 variant H4G regulates rDNA transcription in breast cancer. Nucleic Acids Res **2019**;47:8399-409.PMC6895281

23. Zlatanova J, Doenecke D. Histone H1 zero: a major player in cell differentiation? FASEB J **1994**;8:1260-8
24. Zhang K, Dent SY. Histone modifying enzymes and cancer: going beyond histones. J Cell Biochem **2005**;96:1137-48
25. Biswas S, Rao CM. Epigenetic tools (The Writers, The Readers and The Erasers) and their implications in cancer therapy. Eur J Pharmacol **2018**;837:8-24
26. Marmorstein R, Zhou MM. Writers and readers of histone acetylation: structure, mechanism, and inhibition. Cold Spring Harb Perspect Biol **2014**;6:a018762.PMC4067988
27. Roth SY, Denu JM, Allis CD. Histone Acetyltransferases. Annual Review of Biochemistry **2001**;70:81-120
28. Dillon SC, Zhang X, Trievel RC, Cheng X. The SET-domain protein superfamily: protein lysine methyltransferases. Genome Biol **2005**;6:227.PMC1273623
29. Kouzarides T. Histone methylation in transcriptional control. Curr Opin Genet Dev **2002**;12:198-209
30. Hwang JW, Cho Y, Bae GU, Kim SN, Kim YK. Protein arginine methyltransferases: promising targets for cancer therapy. Exp Mol Med **2021**;53:788-808.PMC8178397
31. Rea S, Eisenhaber F, O'Carroll D, Strahl BD, Sun ZW, Schmid M, Opravil S, Mechtler K, Ponting CP, Allis CD, Jenuwein T. Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature **2000**;406:593-9
32. Hyun K, Jeon J, Park K, Kim J. Writing, erasing and reading histone lysine methylations. Exp Mol Med **2017**;49:e324.PMC6130214
33. van Nuland R, Gozani O. Histone H4 Lysine 20 (H4K20) Methylation, Expanding the Signaling Potential of the Proteome One Methyl Moiety at a Time. Mol Cell Proteomics **2016**;15:755-64.PMC4813698
34. Feng Q, Wang H, Ng HH, Erdjument-Bromage H, Tempst P, Struhl K, Zhang Y. Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. Curr Biol **2002**;12:1052-8
35. Ramaiah MJ, Tangutur AD, Manyam RR. Epigenetic modulation and understanding of HDAC inhibitors in cancer therapy. Life Sci **2021**;277:119504
36. Ropero S, Esteller M. The role of histone deacetylases (HDACs) in human cancer. Mol Oncol **2007**;1:19-25.PMC5543853
37. Seto E, Yoshida M. Erasers of histone acetylation: the histone deacetylase enzymes. Cold Spring Harb Perspect Biol **2014**;6:a018713.PMC3970420
38. D'Oto A, Tian QW, Davidoff AM, Yang J. Histone demethylases and their roles in cancer epigenetics. J Med Oncol Ther **2016**;1:34-40.PMC5279889
39. Poulard C, Corbo L, Le Romancer M. Protein arginine methylation/demethylation and cancer. Oncotarget **2016**;7:67532-50.PMC5341895
40. Chang B, Chen Y, Zhao Y, Bruick RK. JMJD6 is a histone arginine demethylase. Science **2007**;318:444-7
41. Perillo B, Tramontano A, Pezone A, Migliaccio A. LSD1: more than demethylation of histone lysine residues. Exp Mol Med **2020**;52:1936-47.PMC8080763
42. Shi Y, Lan F, Matson C, Mulligan P, Whetstine JR, Cole PA, Casero RA, Shi Y. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell **2004**;119:941-53
43. Fang R, Barbera AJ, Xu Y, Rutenberg M, Leonor T, Bi Q, Lan F, Mei P, Yuan GC, Lian C, Peng J, Cheng D, Sui G, Kaiser UB, Shi Y, Shi YG. Human LSD2/KDM1b/AOF1 regulates gene transcription by modulating intragenic H3K4me2 methylation. Mol Cell **2010**;39:222-33.PMC3518444
44. Hojfeldt JW, Agger K, Helin K. Histone lysine demethylases as targets for anticancer therapy. Nat Rev Drug Discov **2013**;12:917-30

45. Cloos PA, Christensen J, Agger K, Helin K. Erasing the methyl mark: histone demethylases at the center of cellular differentiation and disease. Genes Dev **2008**;22:1115-40.PMC2732404
46. He J, Kallin EM, Tsukada Y, Zhang Y. The H3K36 demethylase Jhdm1b/Kdm2b regulates cell proliferation and senescence through p15(Ink4b). Nat Struct Mol Biol **2008**;15:1169-75.PMC2612995
47. Frescas D, Guardavaccaro D, Bassermann F, Koyama-Nasu R, Pagano M. JHDM1B/FBXL10 is a nucleolar protein that represses transcription of ribosomal RNA genes. Nature **2007**;450:309-13
48. Dhalluin C, Carlson JE, Zeng L, He C, Aggarwal AK, Zhou MM. Structure and ligand of a histone acetyltransferase bromodomain. Nature **1999**;399:491-6
49. Ali HA, Li Y, Bilal AHM, Qin T, Yuan Z, Zhao W. A Comprehensive Review of BET Protein Biochemistry, Physiology, and Pathological Roles. Front Pharmacol **2022**;13:818891.PMC8990909
50. Yun M, Wu J, Workman JL, Li B. Readers of histone modifications. Cell Res **2011**;21:564-78.PMC3131977
51. Maeda R, Tachibana M. HP1 maintains protein stability of H3K9 methyltransferases and demethylases. EMBO Rep **2022**;23:e53581.PMC8982598
52. Jain K, Fraser CS, Marunde MR, Parker MM, Sagum C, Burg JM, Hall N, Popova IK, Rodriguez KL, Vaidya A, Krajewski K, Keogh MC, Bedford MT, Strahl BD. Characterization of the plant homeodomain (PHD) reader family for their histone tail interactions. Epigenetics Chromatin **2020**;13:3.PMC6979384
53. Sharma S, Kelly TK, Jones PA. Epigenetics in cancer. Carcinogenesis **2010**;31:27-36.PMC2802667
54. Fraga MF, Ballestar E, Villar-Garea A, Boix-Chornet M, Espada J, Schotta G, Bonaldi T, Haydon C, Ropero S, Petrie K, Iyer NG, Perez-Rosado A, Calvo E, Lopez JA, Cano A, Calasanz MJ, Colomer D, Piris MA, Ahn N, Imhof A, Caldas C, Jenuwein T, Esteller M. Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. Nat Genet **2005**;37:391-400
55. Noberini R, Restellini C, Savoia EO, Raimondi F, Ghiani L, Jodice MG, Bertalot G, Bonizzi G, Capra M, Maffini FA, Tagliabue M, Ansarin M, Lupia M, Giordano M, Osti D, Pelicci G, Chiocca S, Bonaldi T. Profiling of Epigenetic Features in Clinical Samples Reveals Novel Widespread Changes in Cancer. Cancers (Basel) **2019**;11.PMC6562406
56. Park YS, Jin MY, Kim YJ, Yook JH, Kim BS, Jang SJ. The global histone modification pattern correlates with cancer recurrence and overall survival in gastric adenocarcinoma. Ann Surg Oncol **2008**;15:1968-76
57. Wei Y, Xia W, Zhang Z, Liu J, Wang H, Adsay NV, Albarracin C, Yu D, Abbruzzese JL, Mills GB, Bast RC, Jr., Hortobagyi GN, Hung MC. Loss of trimethylation at lysine 27 of histone H3 is a predictor of poor outcome in breast, ovarian, and pancreatic cancers. Mol Carcinog **2008**;47:701-6.PMC2580832
58. Reddy D, Khade B, Pandya R, Gupta S. A novel method for isolation of histones from serum and its implications in therapeutics and prognosis of solid tumours. Clin Epigenetics **2017**;9:30.PMC5372264
59. Holdenrieder S, Stieber P, Bodenmuller H, Busch M, Fertig G, Furst H, Schalhorn A, Schmeller N, Untch M, Seidel D. Nucleosomes in serum of patients with benign and malignant diseases. Int J Cancer **2001**;95:114-20
60. Duforestel M, Briand J, Bougras-Cartron G, Heymann D, Frenel JS, Vallette FM, Cartron PF. Cell-free circulating epimarks in cancer monitoring. Epigenomics **2020**;12:145-55
61. Gezer U, Yoruker EE, Keskin M, Kulle CB, Dharuman Y, Holdenrieder S. Histone Methylation Marks on Circulating Nucleosomes as Novel Blood-Based Biomarker in Colorectal Cancer. Int J Mol Sci **2015**;16:29654-62.PMC4691123

62. Leszinski G, Gezer U, Siegele B, Stoetzer O, Holdenrieder S. Relevance of histone marks H3K9me3 and H4K20me3 in cancer. Anticancer Res **2012**;32:2199-205
63. Seligson DB, Horvath S, Shi T, Yu H, Tze S, Grunstein M, Kurdistani SK. Global histone modification patterns predict risk of prostate cancer recurrence. Nature **2005**;435:1262-6
64. Halasa M, Wawruszak A, Przybyszewska A, Jaruga A, Guz M, Kalafut J, Stepulak A, Cybulski M. H3K18Ac as a Marker of Cancer Progression and Potential Target of Anti-Cancer Therapy. Cells-Basel **2019**;8
65. Seligson DB, Horvath S, McBrian MA, Mah V, Yu H, Tze S, Wang Q, Chia D, Goodglick L, Kurdistani SK. Global levels of histone modifications predict prognosis in different cancers. Am J Pathol **2009**;174:1619-28.PMC2671251
66. Van Den Broeck A, Brambilla E, Moro-Sibilot D, Lantuejoul S, Brambilla C, Eymin B, Gazzeri S. Loss of histone H4K20 trimethylation occurs in preneoplasia and influences prognosis of non-small cell lung cancer. Clin Cancer Res **2008**;14:7237-45
67. Ashktorab H, Belgrave K, Hosseinkhah F, Brim H, Nouraie M, Takkikto M, Hewitt S, Lee EL, Dashwood RH, Smoot D. Global histone H4 acetylation and HDAC2 expression in colon adenoma and carcinoma. Dig Dis Sci **2009**;54:2109-17.PMC2737733
68. Elsheikh SE, Green AR, Rakha EA, Powe DG, Ahmed RA, Collins HM, Soria D, Garibaldi JM, Paish CE, Ammar AA, Grainge MJ, Ball GR, Abdelghany MK, Martinez-Pomares L, Heery DM, Ellis IO. Global histone modifications in breast cancer correlate with tumor phenotypes, prognostic factors, and patient outcome. Cancer Res **2009**;69:3802-9
69. Song JS, Kim YS, Kim DK, Park SI, Jang SJ. Global histone modification pattern associated with recurrence and disease-free survival in non-small cell lung cancer patients. Pathol Int **2012**;62:182-90
70. Ellinger J, Kahl P, von der Gathen J, Rogenhofer S, Heukamp LC, Gutgemann I, Walter B, Hofstadter F, Buttner R, Muller SC, Bastian PJ, von Ruecker A. Global levels of histone modifications predict prostate cancer recurrence. Prostate **2010**;70:61-9
71. McAnena P, Brown JA, Kerin MJ. Circulating Nucleosomes and Nucleosome Modifications as Biomarkers in Cancer. Cancers (Basel) **2017**;9.PMC5295776
72. Liu BL, Cheng JX, Zhang XA, Wang R, Zhang W, Lin H, Xiao XA, Cai S, Chen XY, Cheng H. Global Histone Modification Patterns as Prognostic Markers to Classify Glioma Patients. Cancer Epidem Biomar **2010**;19:2888-96
73. Duan JL, Nie RC, Xiang ZC, Chen JW, Deng MH, Liang H, Wang FW, Luo RZ, Xie D, Cai MY. Prognostic Model for the Risk Stratification of Early and Late Recurrence in Hepatitis B Virus-Related Small Hepatocellular Carcinoma Patients with Global Histone Modifications. J Hepatocell Carcinoma **2021**;8:493-505.PMC8170593
74. Shanmugam MK, Arfuso F, Arumugam S, Chinnathambi A, Jinsong B, Warrier S, Wang LZ, Kumar AP, Ahn KS, Sethi G, Lakshmanan M. Role of novel histone modifications in cancer. Oncotarget **2018**;9:11414-26.PMC5834259
75. Morgan MA, Shilatifard A. Chromatin signatures of cancer. Genes Dev **2015**;29:238-49.PMC4318141
76. Sheikh BN, Akhtar A. The many lives of KATs - detectors, integrators and modulators of the cellular environment. Nat Rev Genet **2019**;20:7-23
77. Iyer NG, Ozdag H, Caldas C. p300/CBP and cancer. Oncogene **2004**;23:4225-31
78. Gao Y, Geng J, Hong X, Qi J, Teng Y, Yang Y, Qu D, Chen G. Expression of p300 and CBP is associated with poor prognosis in small cell lung cancer. Int J Clin Exp Pathol **2014**;7:760-7.PMC3925924

79. Peters AH, O'Carroll D, Scherthan H, Mechtler K, Sauer S, Schofer C, Weipoltshammer K, Pagani M, Lachner M, Kohlmaier A, Opravil S, Doyle M, Sibilia M, Jenuwein T. Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. Cell **2001**;107:323-37
80. Ford DJ, Dingwall AK. The cancer COMPASS: navigating the functions of MLL complexes in cancer. Cancer Genet **2015**;208:178-91
81. Rabello Ddo A, de Moura CA, de Andrade RV, Motoyama AB, Silva FP. Altered expression of MLL methyltransferase family genes in breast cancer. Int J Oncol **2013**;43:653-60
82. Wu S, Yang Z, Ye R, An D, Li C, Wang Y, Wang Y, Huang Y, Liu H, Li F, He L, Sun D, Yu Y, Li Q, Huang P, Zhang M, Zhao X, Bi T, Zhuang X, Zhang L, Lu J, Sun X, Zhou F, Liu C, Yang G, Hou Y, Fan Z, Cai Z. Novel variants in MLL confer to bladder cancer recurrence identified by whole-exome sequencing. Oncotarget **2016**;7:2629-45.PMC4823060
83. Gan L, Yang Y, Li Q, Feng Y, Liu T, Guo W. Epigenetic regulation of cancer progression by EZH2: from biological insights to therapeutic potential. Biomark Res **2018**;6:10.PMC5845366
84. Tzao C, Tung HJ, Jin JS, Sun GH, Hsu HS, Chen BH, Yu CP, Lee SC. Prognostic significance of global histone modifications in resected squamous cell carcinoma of the esophagus. Mod Pathol **2009**;22:252-60
85. Rueda-Robles A, Audano M, Alvarez-Mercado AI, Rubio-Tomas T. Functions of SMYD proteins in biological processes: What do we know? An updated review. Arch Biochem Biophys **2021**;712:109040
86. Hayami S, Kelly JD, Cho HS, Yoshimatsu M, Unoki M, Tsunoda T, Field HI, Neal DE, Yamaue H, Ponder BA, Nakamura Y, Hamamoto R. Overexpression of LSD1 contributes to human carcinogenesis through chromatin regulation in various cancers. Int J Cancer **2011**;128:574-86
87. Berry WL, Janknecht R. KDM4/JMJD2 histone demethylases: epigenetic regulators in cancer cells. Cancer Res **2013**;73:2936-42.PMC3655154
88. Di Nisio E, Licursi V, Mannironi C, Buglioni V, Paiardini A, Robusti G, Noberini R, Bonaldi T, Negri R. A truncated and catalytically inactive isoform of KDM5B histone demethylase accumulates in breast cancer cells and regulates H3K4 tri-methylation and gene expression. Cancer Gene Ther **2023**;30:822-32.PMC10281864
89. Cheng Y, He C, Wang M, Ma X, Mo F, Yang S, Han J, Wei X. Targeting epigenetic regulators for cancer therapy: mechanisms and advances in clinical trials. Signal Transduct Target Ther **2019**;4:62.PMC6915746
90. Prachayasittikul V, Prathipati P, Pratiwi R, Phanus-Umporn C, Malik AA, Schaduangrat N, Seenprachawong K, Wongchitrat P, Supokawej A, Prachayasittikul V, Wikberg JE, Nantasenamat C. Exploring the epigenetic drug discovery landscape. Expert Opin Drug Discov **2017**;12:345-62
91. Ahuja N, Sharma AR, Baylin SB. Epigenetic Therapeutics: A New Weapon in the War Against Cancer. Annu Rev Med **2016**;67:73-89.PMC4937439
92. Issa JP, Kantarjian HM, Kirkpatrick P. Azacitidine. Nat Rev Drug Discov **2005**;4:275-6
93. Kim N, Norsworthy KJ, Subramaniam S, Chen H, Manning ML, Kitabi E, Earp J, Ehrlich LA, Okusanya OO, Vallejo J, Gehrke BJ, de Claro RA, Pazdur R. FDA Approval Summary: Decitabine and Cedazuridine Tablets for Myelodysplastic Syndromes. Clin Cancer Res **2022**;28:3411-6.PMC9378483
94. Schuh AC, Dohner H, Pleyer L, Seymour JF, Fenaux P, Dombret H. Azacitidine in adult patients with acute myeloid leukemia. Crit Rev Oncol Hematol **2017**;116:159-77
95. Kim A, Mo K, Kwon H, Choe S, Park M, Kwak W, Yoon H. Epigenetic Regulation in Breast Cancer: Insights on Epidrugs. Epigenomes **2023**;7.PMC9953240

96. Montalvo-Casimiro M, Gonzalez-Barrios R, Meraz-Rodriguez MA, Juarez-Gonzalez VT, Arriaga-Canon C, Herrera LA. Epidrug Repurposing: Discovering New Faces of Old Acquaintances in Cancer Therapy. Front Oncol **2020**;10:605386.PMC7708379
97. Hoy SM. Tazemetostat: First Approval. Drugs **2020**;80:513-21
98. Straining R, Eighmy W. Tazemetostat: EZH2 Inhibitor. J Adv Pract Oncol **2022**;13:158-63.PMC8955562
99. Yang GJ, Lei PM, Wong SY, Ma DL, Leung CH. Pharmacological Inhibition of LSD1 for Cancer Treatment. Molecules **2018**;23.PMC6320820
100. Nicosia L, Boffo FL, Ceccacci E, Conforti F, Pallavicini I, Bedin F, Ravasio R, Massignani E, Somervaille TCP, Minucci S, Bonaldi T. Pharmacological inhibition of LSD1 triggers myeloid differentiation by targeting GSE1 oncogenic functions in AML. Oncogene **2022**;41:878-94.PMC8830420
101. Sander JD, Joung JK. CRISPR-Cas systems for editing, regulating and targeting genomes. Nat Biotechnol **2014**;32:347-55.PMC4022601
102. Portela A, Esteller M. Epigenetic modifications and human disease. Nat Biotechnol **2010**;28:1057-68
103. Larsen MR, Trelle MB, Thingholm TE, Jensen ON. Analysis of posttranslational modifications of proteins by tandem mass spectrometry. Biotechniques **2006**;40:790-8
104. Walther TC, Mann M. Mass spectrometry-based proteomics in cell biology. J Cell Biol **2010**;190:491-500.PMC2928005
105. Kaufmann R. Matrix-assisted laser desorption ionization (MALDI) mass spectrometry: a novel analytical tool in molecular biology and biotechnology. J Biotechnol **1995**;41:155-75
106. Konermann L, Ahadi E, Rodriguez AD, Vahidi S. Unraveling the mechanism of electrospray ionization. Anal Chem **2013**;85:2-9
107. Rappsilber J, Ryder U, Lamond AI, Mann M. Large-scale proteomic analysis of the human spliceosome. Genome Res **2002**;12:1231-45.PMC186633
108. Ong SE, Mittler G, Mann M. Identifying and quantifying in vivo methylation sites by heavy methyl SILAC. Nat Methods **2004**;1:119-26
109. Soldi M, Cuomo A, Bonaldi T. Improved bottom-up strategy to efficiently separate hypermodified histone peptides through ultra-HPLC separation on a bench top Orbitrap instrument. Proteomics **2014**;14:2212-25
110. Noberini R, Robusti G, Bonaldi T. Mass spectrometry-based characterization of histones in clinical samples: applications, progress, and challenges. FEBS J **2022**;289:1191-213.PMC9291046
111. Yuan ZF, Sidoli S, Marchione DM, Simithy J, Janssen KA, Szurgot MR, Garcia BA. EpiProfile 2.0: A Computational Platform for Processing Epi-Proteomics Mass Spectrometry Data. J Proteome Res **2018**;17:2533-41.PMC6387837
112. Noberini R, Bonaldi T. A Super-SILAC Strategy for the Accurate and Multiplexed Profiling of Histone Posttranslational Modifications. Methods Enzymol **2017**;586:311-32
113. Noberini R, Longhi E, Bonaldi T. A Super-SILAC Approach for Profiling Histone Posttranslational Modifications. Methods Mol Biol **2023**;2603:87-102
114. Lin S, Wein S, Gonzales-Cope M, Otte GL, Yuan ZF, Afjehi-Sadat L, Maile T, Berger SL, Rush J, Lill JR, Arnott D, Garcia BA. Stable-isotope-labeled histone peptide library for histone post-translational modification and variant quantification by mass spectrometry. Mol Cell Proteomics **2014**;13:2450-66.PMC4159661
115. Noberini R, Uggetti A, Pruneri G, Minucci S, Bonaldi T. Pathology Tissue-quantitative Mass Spectrometry Analysis to Profile Histone Post-translational Modification Patterns in Patient Samples. Mol Cell Proteomics **2016**;15:866-77.PMC4813706

116. Robusti G, Vai A, Bonaldi T, Noberini R. Investigating pathological epigenetic aberrations by epi-proteomics. Clin Epigenetics **2022**;14:145.PMC9652867
117. Nakato R, Sakata T. Methods for ChIP-seq analysis: A practical workflow and advanced applications. Methods **2021**;187:44-53
118. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet **2009**;10:57-63.PMC2949280
119. Wilhelm BT, Marguerat S, Watt S, Schubert F, Wood V, Goodhead I, Penkett CJ, Rogers J, Bahler J. Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. Nature **2008**;453:1239-U39
120. Network TR. <https://www.cancer.gov/tcga>.
121. CPTAC. <https://proteomics.cancer.gov/programs/cptac >.
122. Babu M, Snyder M. Multi-Omics Profiling for Health. Mol Cell Proteomics **2023**;22:100561.PMC10220275
123. Nativio R, Lan Y, Donahue G, Sidoli S, Berson A, Srinivasan AR, Shcherbakova O, Amlie-Wolf A, Nie J, Cui X, He C, Wang LS, Garcia BA, Trojanowski JQ, Bonini NM, Berger SL. An integrated multi-omics approach identifies epigenetic alterations associated with Alzheimer's disease. Nat Genet **2020**;52:1024-35.PMC8098004
124. Kelly MR, Wisniewska K, Regner MJ, Lewis MW, Perreault AA, Davis ES, Phanstiel DH, Parker JS, Franco HL. A multi-omic dissection of super-enhancer driven oncogenic gene expression programs in ovarian cancer. Nat Commun **2022**;13:4247.PMC9307778
125. Huang H, Hu J, Maryam A, Huang Q, Zhang Y, Ramakrishnan S, Li J, Ma H, Ma VWS, Cheuk W, So GYK, Wang W, Cho WCS, Zhang L, Chan KM, Wang X, Chin YR. Defining super-enhancer landscape in triple-negative breast cancer by multiomic profiling. Nat Commun **2021**;12:2242.PMC8046763
126. Zheng Q, Maksimovic I, Upad A, David Y. Non-enzymatic covalent modifications: a new link between metabolism and epigenetics. Protein Cell **2020**;11:401-16.PMC7251012
127. Hammond CM, Stromme CB, Huang H, Patel DJ, Groth A. Histone chaperone networks shaping chromatin function. Nat Rev Mol Cell Biol **2017**;18:141-58.PMC5319910
128. Paddison PJ, Caudy AA, Bernstein E, Hannon GJ, Conklin DS. Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. Genes Dev **2002**;16:948-58.PMC152352
129. Ghavami S, Pandi A. CRISPR interference and its applications. Prog Mol Biol Transl Sci **2021**;180:123-40
130. Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, Stern-Ginossar N, Brandman O, Whitehead EH, Doudna JA, Lim WA, Weissman JS, Qi LS. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell **2013**;154:442-51.PMC3770145
131. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin **2021**;71:209-49
132. Harbeck N, Penault-Llorca F, Cortes J, Gnant M, Houssami N, Poortmans P, Ruddy K, Tsang J, Cardoso F. Breast cancer. Nat Rev Dis Primers **2019**;5:66
133. Zhang MH, Man HT, Zhao XD, Dong N, Ma SL. Estrogen receptor-positive breast cancer molecular signatures and therapeutic potentials (Review). Biomed Rep **2014**;2:41-52.PMC3916982
134. Lukasiewicz S, Czeczelewski M, Forma A, Baj J, Sitarz R, Stanislawek A. Breast Cancer-Epidemiology, Risk Factors, Classification, Prognostic Markers, and Current Treatment Strategies-An Updated Review. Cancers (Basel) **2021**;13.PMC8428369

135. Orrantia-Borunda E, Anchondo-Nunez P, Acuna-Aguilar LE, Gomez-Valles FO, Ramirez-Valdespino CA. Subtypes of Breast Cancer. In: Mayrovitz HN, editor. Breast Cancer. Brisbane (AU)2022.
136. Purdie CA, Quinlan P, Jordan LB, Ashfield A, Ogston S, Dewar JA, Thompson AM. Progesterone receptor expression is an independent prognostic variable in early breast cancer: a population-based study. Br J Cancer **2014**;110:565-72.PMC3915123
137. Nicolini A, Ferrari P, Duffy MJ. Prognostic and predictive biomarkers in breast cancer: Past, present and future. Semin Cancer Biol **2018**;52:56-73
138. Iqbal N, Iqbal N. Human Epidermal Growth Factor Receptor 2 (HER2) in Cancers: Overexpression and Therapeutic Implications. Mol Biol Int **2014**;2014:852748.PMC4170925
139. Nishimura R, Osako T, Okumura Y, Hayashi M, Toyozumi Y, Arima N. Ki-67 as a prognostic marker according to breast cancer subtype and a predictor of recurrence time in primary breast cancer. Exp Ther Med **2010**;1:747-54.PMC3445951
140. de Azambuja E, Cardoso F, de Castro G, Jr., Colozza M, Mano MS, Durbecq V, Sotiriou C, Larsimont D, Piccart-Gebhart MJ, Paesmans M. Ki-67 as prognostic marker in early breast cancer: a meta-analysis of published studies involving 12,155 patients. Br J Cancer **2007**;96:1504-13.PMC2359936
141. Prat A, Perou CM. Deconstructing the molecular portraits of breast cancer. Mol Oncol **2011**;5:5-23.PMC5528267
142. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO, Botstein D. Molecular portraits of human breast tumours. Nature **2000**;406:747-52
143. Wirapati P, Sotiriou C, Kunkel S, Farmer P, Pradervand S, Haibe-Kains B, Desmedt C, Ignatiadis M, Sengstag T, Schutz F, Goldstein DR, Piccart M, Delorenzi M. Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. Breast Cancer Res **2008**;10:R65.PMC2575538
144. Dias K, Dvorkin-Gheva A, Hallett RM, Wu Y, Hassell J, Pond GR, Levine M, Whelan T, Bane AL. Claudin-Low Breast Cancer; Clinical & Pathological Characteristics. PLoS One **2017**;12:e0168669.PMC5207440
145. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, Davies S, Fauron C, He X, Hu Z, Quackenbush JF, Stijleman IJ, Palazzo J, Marron JS, Nobel AB, Mardis E, Nielsen TO, Ellis MJ, Perou CM, Bernard PS. Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol **2009**;27:1160-7.PMC2667820
146. Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, Pietenpol JA. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J Clin Invest **2011**;121:2750-67.PMC3127435
147. Lehmann BD, Pietenpol JA, Tan AR. Triple-negative breast cancer: molecular subtypes and new targets for therapy. Am Soc Clin Oncol Educ Book **2015**:e31-9
148. Sher G, Salman NA, Khan AQ, Prabhu KS, Raza A, Kulinski M, Dermime S, Haris M, Junejo K, Uddin S. Epigenetic and breast cancer therapy: Promising diagnostic and therapeutic applications. Semin Cancer Biol **2022**;83:152-65
149. Lustberg MB, Ramaswamy B. Epigenetic Therapy in Breast Cancer. Curr Breast Cancer Rep **2011**;3:34-43.PMC3477864
150. Stefansson OA, Moran S, Gomez A, Sayols S, Arribas-Jorba C, Sandoval J, Hilmarsdottir H, Olafsdottir E, Tryggvadottir L, Jonasson JG, Eyfjord J, Esteller M. A DNA methylation-based definition of biologically distinct breast cancer subtypes. Mol Oncol **2015**;9:555-68.PMC5528700

151. Gomez-Miragaya J, Moran S, Calleja-Cervantes ME, Collado-Sole A, Pare L, Gomez A, Serra V, Dobrolecki LE, Lewis MT, Diaz-Lagares A, Eroles P, Prat A, Esteller M, Gonzalez-Suarez E. The Altered Transcriptome and DNA Methylation Profiles of Docetaxel Resistance in Breast Cancer PDX Models. Mol Cancer Res **2019**;17:2063-76
152. Marsolier J, Prompsy P, Durand A, Lyne AM, Landragin C, Trouchet A, Bento ST, Eisele A, Foulon S, Baudre L, Grosselin K, Bohec M, Baulande S, Dahmani A, Sourd L, Letouze E, Salomon AV, Marangoni E, Perie L, Vallot C. H3K27me3 conditions chemotolerance in triple-negative breast cancer. Nat Genet **2022**;54:459-68.PMC7612638
153. Restellini C, Cuomo A, Lupia M, Giordano M, Bonaldi T, Noberini R. Alternative digestion approaches improve histone modification mapping by mass spectrometry in clinical samples. Proteomics Clin Appl **2019**;13:e1700166
154. Sanchini V, Bonizzi G, Disalvatore D, Monturano M, Pece S, Viale G, Di Fiore PP, Boniolo G. A Trust-Based Pact in Research Biobanks. From Theory to Practice. Bioethics **2016**;30:260-71
155. Noberini R, Restellini C, Savoia EO, Bonaldi T. Enrichment of histones from patient samples for mass spectrometry-based analysis of post-translational modifications. Methods **2020**;184:19-28
156. Noberini R, Savoia EO, Brandini S, Greco F, Marra F, Bertalot G, Pruneri G, McDonnell LA, Bonaldi T. Spatial epi-proteomics enabled by histone post-translational modification analysis from low-abundance clinical samples. Clin Epigenetics **2021**;13:145.PMC8317427
157. Rappsilber J, Mann M, Ishihama Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat Protoc **2007**;2:1896-906
158. Noberini R, Morales Torres C, Savoia EO, Brandini S, Jodice MG, Bertalot G, Bonizzi G, Capra M, Diaferia G, Scaffidi P, Bonaldi T. Label-Free Mass Spectrometry-Based Quantification of Linker Histone H1 Variants in Clinical Samples. Int J Mol Sci **2020**;21.PMC7582528
159. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol **2008**;26:1367-72
160. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M. Andromeda: a peptide search engine integrated into the MaxQuant environment. J Proteome Res **2011**;10:1794-805
161. Noberini R, Osti D, Miccolo C, Richichi C, Lupia M, Corleone G, Hong SP, Colombo P, Pollo B, Fornasari L, Pruneri G, Magnani L, Cavallaro U, Chiocca S, Minucci S, Pelicci G, Bonaldi T. Extensive and systematic rewiring of histone post-translational modifications in cancer model systems. Nucleic Acids Res **2018**;46:3817-32.PMC5934616
162. Frank SR, Schroeder M, Fernandez P, Taubert S, Amati B. Binding of c-Myc to chromatin mediates mitogen-induced acetylation of histone H4 and gene activation. Genes Dev **2001**;15:2069-82.PMC312758
163. Abcam X-ChIP protocol <https://docs.abcam.com/pdf/chromatin/A-beginners-guide-to-ChIP.pdf>.
164. Blecher-Gonen R, Barnett-Itzhaki Z, Jaitin D, Amann-Zalcenstein D, Lara-Astiaso D, Amit I. High-throughput chromatin immunoprecipitation for genome-wide mapping of in vivo protein-DNA interactions and epigenomic states. Nat Protoc **2013**;8:539-54
165. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol **2009**;10:R25.PMC2690996

166. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. Model-based analysis of ChIP-Seq (MACS). Genome Biol **2008**;9:R137.PMC2592715
167. Neph S, Kuehn MS, Reynolds AP, Haugen E, Thurman RE, Johnson AK, Rynes E, Maurano MT, Vierstra J, Thomas S, Sandstrom R, Humbert R, Stamatoyannopoulos JA. BEDOPS: high-performance genomic feature operations. Bioinformatics **2012**;28:1919-20.PMC3389768
168. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics **2010**;26:841-2.PMC2832824
169. Stovner EB, Saetrom P. epic2 efficiently finds diffuse domains in ChIP-seq data. Bioinformatics **2019**;35:4392-3
170. Amemiya HM, Kundaje A, Boyle AP. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. Sci Rep **2019**;9:9354.PMC6597582
171. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-Andre V, Sigova AA, Hoke HA, Young RA. Super-enhancers in the control of cell identity and disease. Cell **2013**;155:934-47.PMC3841062
172. Jiang Y, Qian F, Bai X, Liu Y, Wang Q, Ai B, Han X, Shi S, Zhang J, Li X, Tang Z, Pan Q, Wang Y, Wang F, Li C. SEdb: a comprehensive human super-enhancer database. Nucleic Acids Res **2019**;47:D235-D43.PMC6323980
173. Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze AV, Mikheenko A, Vollger MR, Altemose N, Uralsky L, Gershman A, Aganezov S, Hoyt SJ, Diekhans M, Logsdon GA, Alonge M, Antonarakis SE, Borchers M, Bouffard GG, Brooks SY, Caldas GV, Chen NC, Cheng H, Chin CS, Chow W, de Lima LG, Dishuck PC, Durbin R, Dvorkina T, Fiddes IT, Formenti G, Fulton RS, Fungtammasan A, Garrison E, Grady PGS, Graves-Lindsay TA, Hall IM, Hansen NF, Hartley GA, Haukness M, Howe K, Hunkapiller MW, Jain C, Jain M, Jarvis ED, Kerpedjiev P, Kirsche M, Kolmogorov M, Korlach J, Kremitzki M, Li H, Maduro VV, Marschall T, McCartney AM, McDaniel J, Miller DE, Mullikin JC, Myers EW, Olson ND, Paten B, Peluso P, Pevzner PA, Porubsky D, Potapova T, Rogaev EI, Rosenfeld JA, Salzberg SL, Schneider VA, Sedlazeck FJ, Shafin K, Shew CJ, Shumate A, Sims Y, Smit AFA, Soto DC, Sovic I, Storer JM, Streets A, Sullivan BA, Thibaud-Nissen F, Torrance J, Wagner J, Walenz BP, Wenger A, Wood JMD, Xiao C, Yan SM, Young AC, Zarate S, Surti U, McCoy RC, Dennis MY, Alexandrov IA, Gerton JL, O'Neill RJ, Timp W, Zook JM, Schatz MC, Eichler EE, Miga KH, Phillippy AM. The complete sequence of a human genome. Science **2022**;376:44-53.PMC9186530
174. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics **2014**;30:923-30
175. Rao X, Huang X, Zhou Z, Lin X. An improvement of the 2^(-delta delta CT) method for quantitative real-time polymerase chain reaction data analysis. Biostat Bioinforma Biomath **2013**;3:71-85.PMC4280562
176. Zambelli F, Mastropasqua F, Picardi E, D'Erchia AM, Pesole G, Pavesi G. RNentropy: an entropy-based tool for the detection of significant variation of gene expression across multiple RNA-Seq experiments. Nucleic Acids Res **2018**;46:e46.PMC5934672
177. Concordet J-P, Haeussler M. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. Nucleic Acids Research **2018**;46:W242-W5
178. Terrenoire E, McRonald F, Halsall JA, Page P, Illingworth RS, Taylor AM, Davison V, O'Neill LP, Turner BM. Immunostaining of modified histones defines high-level features of the human metaphase epigenome. Genome Biol **2010**;11:R110.PMC3156949
179. Franken NA, Rodermond HM, Stap J, Haveman J, van Bree C. Clonogenic assay of cells in vitro. Nat Protoc **2006**;1:2315-9

180. Yokoyama Y, Matsumoto A, Hieda M, Shinchi Y, Ogihara E, Hamada M, Nishioka Y, Kimura H, Yoshidome K, Tsujimoto M, Matsuura N. Loss of histone H4K20 trimethylation predicts poor prognosis in breast cancer and is associated with invasive activity. Breast Cancer Res **2014**;16:R66.PMC4229880
181. Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, Mann M. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Mol Cell Proteomics **2014**;13:2513-26.PMC4159666
182. Izzo A, Kamieniarz-Gdula K, Ramirez F, Noureen N, Kind J, Manke T, van Steensel B, Schneider R. The genomic landscape of the somatic linker histone subtypes H1.1 to H1.5 in human cells. Cell Rep **2013**;3:2142-54
183. Bonenfant D, Towbin H, Coulot M, Schindler P, Mueller DR, van Oostrum J. Analysis of dynamic changes in post-translational modifications of human histones during cell cycle by mass spectrometry. Mol Cell Proteomics **2007**;6:1917-32
184. Zane L, Chapus F, Pegoraro G, Misteli T. HiHiMap: single-cell quantitation of histones and histone posttranslational modifications across the cell cycle by high-throughput imaging. Molecular Biology of the Cell **2017**;28:2290-302
185. Sakaue-Sawano A, Yo M, Komatsu N, Hiratsuka T, Kogure T, Hoshida T, Goshima N, Matsuda M, Miyoshi H, Miyawaki A. Genetically Encoded Tools for Optical Dissection of the Mammalian Cell Cycle. Mol Cell **2017**;68:626-40 e5
186. Yoshino S, Suzuki HI. The molecular understanding of super-enhancer dysregulation in cancer. Nagoya J Med Sci **2022**;84:216-29.PMC9350580
187. Garcia-Teijido P, Cabal ML, Fernandez IP, Perez YF. Tumor-Infiltrating Lymphocytes in Triple Negative Breast Cancer: The Future of Immune Targeting. Clin Med Insights Oncol **2016**;10:31-9.PMC4822722
188. Valenza C, Taurelli Salimbeni B, Santoro C, Trapani D, Antonarelli G, Curigliano G. Tumor Infiltrating Lymphocytes across Breast Cancer Subtypes: Current Issues for Biomarker Assessment. Cancers (Basel) **2023**;15.PMC9913599
189. Massihnia D, Galvano A, Fanale D, Perez A, Castiglia M, Incorvaia L, Listi A, Rizzo S, Cicero G, Bazan V, Castorina S, Russo A. Triple negative breast cancer: shedding light onto the role of pi3k/akt/mtor pathway. Oncotarget **2016**;7:60712-22.PMC5312414
190. Abdulkareem NM, Bhat R, Qin L, Vasaikar S, Gopinathan A, Mitchell T, Shea MJ, Nanda S, Thangavel H, Zhang B, De Angelis C, Schiff R, Trivedi MV. A novel role of ADGRF1 (GPR110) in promoting cellular quiescence and chemoresistance in human epidermal growth factor receptor 2-positive breast cancer. FASEB J **2021**;35:e21719.PMC8218746
191. Farhangnia P, Ghomi SM, Mollazadehghomi S, Nickho H, Akbarpour M, Delbandi AA. SLAM-family receptors come of age as a potential molecular target in cancer immunotherapy. Front Immunol **2023**;14:1174138.PMC10213746
192. Vecchi L, Araujo TG, Azevedo F, Mota STS, Avila VMR, Ribeiro MA, Goulart LR. Phospholipase A(2) Drives Tumorigenesis and Cancer Aggressiveness through Its Interaction with Annexin A1. Cells-Basel **2021**;10.PMC8231270
193. Huang W, Li Y, Zhang C, Zha H, Zhou X, Fu B, Guo J, Wang G. IGF2BP3 facilitates cell proliferation and tumorigenesis via modulation of JAK/STAT signalling pathway in human bladder cancer. J Cell Mol Med **2020**;24:13949-60.PMC7753985
194. Li K, Liu Y, Cao H, Zhang Y, Gu Z, Liu X, Yu A, Kaphle P, Dickerson KE, Ni M, Xu J. Interrogation of enhancer function by enhancer-targeting CRISPR epigenetic editing. Nat Commun **2020**;11:485.PMC6981169
195. Yang L, Jin M, Jeong KW. Histone H3K4 Methyltransferases as Targets for Drug-Resistant Cancers. Biology (Basel) **2021**;10.PMC8301125
196. Punzi S, Balestrieri C, D'Alesio C, Bossi D, Dellino GI, Gatti E, Pruneri G, Criscitiello C, Lovati G, Meliksetyan M, Carugo A, Curigliano G, Natoli G, Pelicci PG, Lanfrancone L. WDR5 inhibition halts metastasis dissemination by repressing

the mesenchymal phenotype of breast cancer cells. Breast Cancer Res **2019**;21:123.PMC6873410
197. Zhang J, Zhou Q, Xie K, Cheng L, Peng S, Xie R, Liu L, Zhang Y, Dong W, Han J, Huang M, Chen Y, Lin T, Huang J, Chen X. Targeting WD repeat domain 5 enhances chemosensitivity and inhibits proliferation and programmed death-ligand 1 expression in bladder cancer. J Exp Clin Cancer Res **2021**;40:203.PMC8215817
198. Wu M, Fan B, Guo Q, Li Y, Chen R, Lv N, Diao Y, Luo Y. Knockdown of SETDB1 inhibits breast cancer progression by miR-381-3p-related regulation. Biol Res **2018**;51:39.PMC6180515
199. Chiba T, Saito T, Yuki K, Zen Y, Koide S, Kanogawa N, Motoyama T, Ogasawara S, Suzuki E, Ooka Y, Tawada A, Otsuka M, Miyazaki M, Iwama A, Yokosuka O. Histone lysine methyltransferase SUV39H1 is a potent target for epigenetic therapy of hepatocellular carcinoma. Int J Cancer **2015**;136:289-98
200. Dambacher S, Hahn M, Schotta G. Epigenetic regulation of development by histone lysine methylation. Heredity (Edinb) **2010**;105:24-37
201. Fioriniello S, Marano D, Fiorillo F, D'Esposito M, Della Ragione F. Epigenetic Factors That Control Pericentric Heterochromatin Organization in Mammals. Genes (Basel) **2020**;11.PMC7349813
202. Kern R, Panis C. CTLA-4 Expression and Its Clinical Significance in Breast Cancer. Arch Immunol Ther Exp (Warsz) **2021**;69:16
203. Maryam A, Chin YR. ANLN Enhances Triple-Negative Breast Cancer Stemness Through TWIST1 and BMP2 and Promotes its Spheroid Growth. Front Mol Biosci **2021**;8:700973.PMC8280772
204. Jeon YH, Kim GW, Kim SY, Yi SA, Yoo J, Kim JY, Lee SW, Kwon SH. Heterochromatin Protein 1: A Multiplayer in Cancer Progression. Cancers (Basel) **2022**;14.PMC8833910
205. Leroy G, Dimaggio PA, Chan EY, Zee BM, Blanco MA, Bryant B, Flaniken IZ, Liu S, Kang Y, Trojer P, Garcia BA. A quantitative atlas of histone modification signatures from human cancer cells. Epigenetics Chromatin **2013**;6:20.PMC3710262
206. Yokoyama Y, Hieda M, Nishioka Y, Matsumoto A, Higashi S, Kimura H, Yamamoto H, Mori M, Matsuura S, Matsuura N. Cancer-associated upregulation of histone H3 lysine 9 trimethylation promotes cell motility in vitro and drives tumor formation in vivo. Cancer Sci **2013**;104:889-95.PMC7657232
207. Yang L, Mei Q, Zielinska-Kwiatkowska A, Matsui Y, Blackburn ML, Benedetti D, Krumm AA, Taborsky GJ, Jr., Chansky HA. An ERG (ets-related gene)-associated histone methyltransferase interacts with histone deacetylases 1/2 and transcription co-repressors mSin3A/B. Biochem J **2003**;369:651-7.PMC1223118
208. Kourmouli N, Sun YM, van der Sar S, Singh PB, Brown JP. Epigenetic regulation of mammalian pericentric heterochromatin in vivo by HP1. Biochem Biophys Res Commun **2005**;337:901-7
209. Li H, Rauch T, Chen ZX, Szabo PE, Riggs AD, Pfeifer GP. The histone methyltransferase SETDB1 and the DNA methyltransferase DNMT3A interact directly and localize to promoters silenced in cancer cells. J Biol Chem **2006**;281:19489-500
210. Wang H, An W, Cao R, Xia L, Erdjument-Bromage H, Chatton B, Tempst P, Roeder RG, Zhang Y. mAM facilitates conversion by ESET of dimethyl to trimethyl lysine 9 of histone H3 to cause transcriptional repression. Mol Cell **2003**;12:475-87

8. List of publications achieved during the Ph.D.

Noberini R, **Robusti G**, Bonaldi T. Mass spectrometry-based characterization of histones in clinical samples: applications, progress, and challenges. FEBS J. 2022 Mar;289(5):1191-1213. doi: 10.1111/febs.15707. Epub 2021 Jan 23. PMID: 33415821; PMCID: PMC9291046.

Robusti G, Vai A, Bonaldi T, Noberini R. Investigating pathological epigenetic aberrations by epi-proteomics. Clin Epigenetics. 2022 Nov 12;14(1):145. doi: 10.1186/s13148-022-01371-y. PMID: 36371348; PMCID: PMC9652867.

Di Nisio E, Licursi V, Mannironi C, Buglioni V, Paiardini A, **Robusti G**, Noberini R, Bonaldi T, Negri R. A truncated and catalytically inactive isoform of KDM5B histone demethylase accumulates in breast cancer cells and regulates H3K4 tri-methylation and gene expression. Cancer Gene Ther. 2023 Jun;30(6):822-832. doi: 10.1038/s41417-022-00584-w. Epub 2023 Jan 26. PMID: 36697763; PMCID: PMC10281864.

9. Acknowledgments

I would like to thank my Ph.D. supervisor, Dr. Tiziana Bonaldi, for giving me the opportunity to work in her laboratory and for her guidance in the research project during the years of my Ph.D. study. I also thank Roberta Noberini for her precious advices, suggestions and support in the research project.

I would like to thank my internal Advisor Prof. Saverio Minucci and my external Advisor Luca Magnani for their kind suggestions and comments on the project.

I would like to thank all the members of the TB group, in particular Alessandro for his important contribution in the bioinformatic analysis for epigenomic and transcriptomic data and Chiara for technical and practical support.

I would like to thank all friends and colleagues at IEO, with whom I shared this journey, especially Stefania, who shared with me positive and negative moments of this study.

Last but not least, I would like to thank my family for always supporting me.