








A histone-centric multi-omics study shows that increased H3K4 methylation sustains triple-negative breast cancer phenotypes

Received: 3 December 2024

Accepted: 27 August 2025

Published online: 30 September 2025

 Check for updates

Roberta Noberini ^{1,8} ✉, Giulia Robusti ^{1,8}, Alessandro Vai^{1,8}, Evelyn Oliva Savoia ¹, Maria Giovanna Jodice¹, Giovanni Bertalot ^{1,6,7}, Betül Çat ¹, Isabella Pallavicini¹, Giuseppina Bonizzi², Maria Capra², Claudia Anna Sangalli³, Federico Zambelli ⁴, Nicola Fusco^{2,5}, Salvatore Pece^{1,5}, Giulio Pavesi⁴, Saverio Minucci^{1,5} & Tiziana Bonaldi ^{1,5} ✉

Altered histone post-translational modifications are frequently associated with cancer. Here, we apply mass-spectrometry to study the epigenetic landscapes of breast cancer subtypes, with a particular focus on triple-negative breast cancers (TNBCs), a heterogeneous group lacking well-defined molecular targets and effective therapies. The analysis of over 200 tumors reveals epigenetic signatures that discriminate TNBCs from the other BC subtypes, and that distinguish TNBC patients with different prognoses. Employing a multi-OMICS approach integrating epigenomics, transcriptomics, and proteomics data, we investigate the mechanistic role of increased H3K4 methylation in TNBCs, demonstrating that H3K4me2 sustains the expression of genes associated with the TNBC phenotype. Through CRISPR-mediated editing, we establish a causal relationship between H3K4me2 and gene expression for several targets. Furthermore, treatment with H3K4 methyltransferase inhibitors reduce TNBC cell growth *in vitro* and *in vivo*. Collectively, our results unravel a novel epigenetic pathway implicated in TNBC pathogenesis and suggest new opportunities for targeted therapy.

Breast cancer (BC) is the most common cancer in women worldwide. Despite a significant improvement in patient survival over the last decades, 40% of BC patients experience relapse, eventually succumbing to their disease. BC therefore remains the first cause of cancer-related death in women. BC subtypes are typically defined based on the immunoreactivity of hormone receptors (Estrogen Receptor-ER, or Progesteron Receptor - PGR), the HER2 receptor and the proliferation marker Ki-67. The main breast cancer molecular subtypes include: Luminal A, Luminal B, HER2-positive and triple negative. Among BC subtypes, triple negative breast cancers (TNBCs),

which represent 15–20% of all breast cancers, are more aggressive and have a worse clinical outcome compared to other subtypes of breast tumors. Treatment of TNBC patients has been challenging due to the heterogeneity of the disease and the absence of molecular targets, and unlike ER-positive and HER2-positive breast cancer, there is no approved targeted therapy for TNBC. Recently, a step towards this goal was achieved through gene expression analysis of TNBCs, which suggested the existence of six subtypes – which were then revised to four – displaying unique gene expressions and ontologies and distinct responses to different chemotherapeutic agents^{1,2}.

¹Department of Experimental Oncology, IEO European Institute of Oncology IRCCS, Milan, Italy. ²Biobank for Translational and Digital Medicine, IEO European Institute of Oncology IRCCS, Milan, Italy. ³Clinical Trial Office, IEO European Institute of Oncology IRCCS, Milan, Italy. ⁴Dipartimento di Bioscienze, Università degli Studi di Milano, Milan, Italy. ⁵Department of Oncology and Hematology-Oncology, University of Milan, Milan, Italy. ⁶Present address: Unità Operativa Multizonale di Anatomia Patologica, APSS, Trento, Italy. ⁷Present address: Centre for Medical Sciences – CISMed, University of Trento, Trento, Italy. ⁸These authors contributed equally: Roberta Noberini, Giulia Robusti, Alessandro Vai. ✉e-mail: roberta.noberini@ieo.it; tiziana.bonaldi@ieo.it

Cancer arises as the result of the accumulation of genetic defects such as mutations and copy number changes^{3,4}. However, striking evidence has now shown how epigenetic changes, including histone post-translational modifications (PTMs), also play a crucial role in cancer initiation and progression, also in the context of adaptation to drug treatment^{5,6}. Histone PTMs represent a vast catalog of combinatorial events that occur on the tail of histone proteins, the critical unit of chromatin, and that contribute to gene regulation and phenotypic identities. Histone PTM alterations are often linked with cancer. Indeed, after the landmark discoveries of the loss of acetylation on H4-lysine 16 (H4K16ac) and of H4-lysine 20 trimethylation (H4K20me3) in cancer⁷, and the prognostic value of histone PTMs in various types of cancers^{8,9}, many more histone marks have been recognized as possible cancer biomarkers¹⁰. In breast cancer, a significant correlation has been shown between aberrant histone PTM pattern, tumor biomarker phenotype and clinical outcome by immunohistochemistry¹¹. By using state-of-the-art mass spectrometry (MS)-based approaches, which allow an unbiased, comprehensive and quantitative analysis of histone PTMs¹², we also identified several marks distinguishing BC from normal tissue, as well as different molecular subtypes, in pilot studies^{13–16}.

Global alterations of histone marks observed in cancer are often caused by aberrations in the activity or expression of histone modifying enzymes¹⁷. Since epigenetic changes, unlike genetic ones, are intrinsically reversible and can be overturned, targeting epigenetic enzymes for therapeutic use has emerged as a promising avenue in translational research^{18,19}. Therefore, profiling histone modifications in disease can not only help to uncover possible epigenetic mechanisms underlying different pathologies but also suggest novel epigenetic pathways targetable for therapy.

In this study, we perform an epi-proteomics analysis of large cohorts ($n = 202$ total samples) of BC patient samples belonging to different subtypes. We find that TN samples are characterized by a specific epigenetic profile that distinguishes them from the other subtypes, and that changes in a few histone marks, particularly histone H3K4 methylation, are associated with TNBC prognosis. To investigate this further, we employ a multi-OMICs approach integrating epigenomics, transcriptomics, and proteomics data from BC clinical samples, and we observe that H3K4me2 drives the expression of multiple genes associated with the TNBC phenotype. Importantly, we mechanistically prove this causal link for a subset of genes, by modulating H3K4me2 levels through CRISPR-mediated epigenome editing. In addition, we show that pharmacologically reducing H3K4 methylation levels decreases TNBC cell line growth in vitro and tumor growth in a xenograft in vivo model, unraveling a potential novel therapeutic avenue for the treatment of this tumor.

Results

Mass spectrometry-based profiling of BC clinical samples identifies a histone PTM signature of triple negative breast cancer

Liquid chromatography coupled with MS has emerged as the most powerful tool for the analysis of histone PTMs and variants. During the last years we have developed various bottom-up MS-based methods that allow profiling up to 100 differentially modified peptides from formalin-fixed paraffin embedded (FFPE) and frozen patient tissues^{13–15,20,21}. We applied these methods to profile histone PTMs in a cohort of 120 breast cancer patient tissues comprising 22 Luminal-A like (LuA), 23 Luminal-B like, 19 HER-positive (HER+) and 56 triple negative (TN) samples (Supplementary Fig. 1 and Supplementary Dataset 1). We also included in the analysis 21 normal breast tissues adjacent to tumor tissue, some of which were matched to the tumors analyzed, from previous studies^{14,16}. Histone H3 was processed with an in-gel digestion protocol involving the derivatization of lysines with deuterated acetic anhydride¹⁵, while histone H4 was digested in solution with the Arg-C protease¹⁶ (Fig. 1a), which were the most efficient approaches to comprehensively profile histone acetylations and

methylations on histones H3 and H4 at the time these experiments were carried out. Because the amount of tissue necessary to complete both digestions was not available for all the samples, histone H4 modifications were profiled in 60 samples (9 LuA, 8 LuB, 6 HER+ and 37 TN). Heavy-isotope labeled histones were added as a spike-in standard to each of the samples prior to digestion and used as an internal standard, to improve quantitation accuracy^{22,23}.

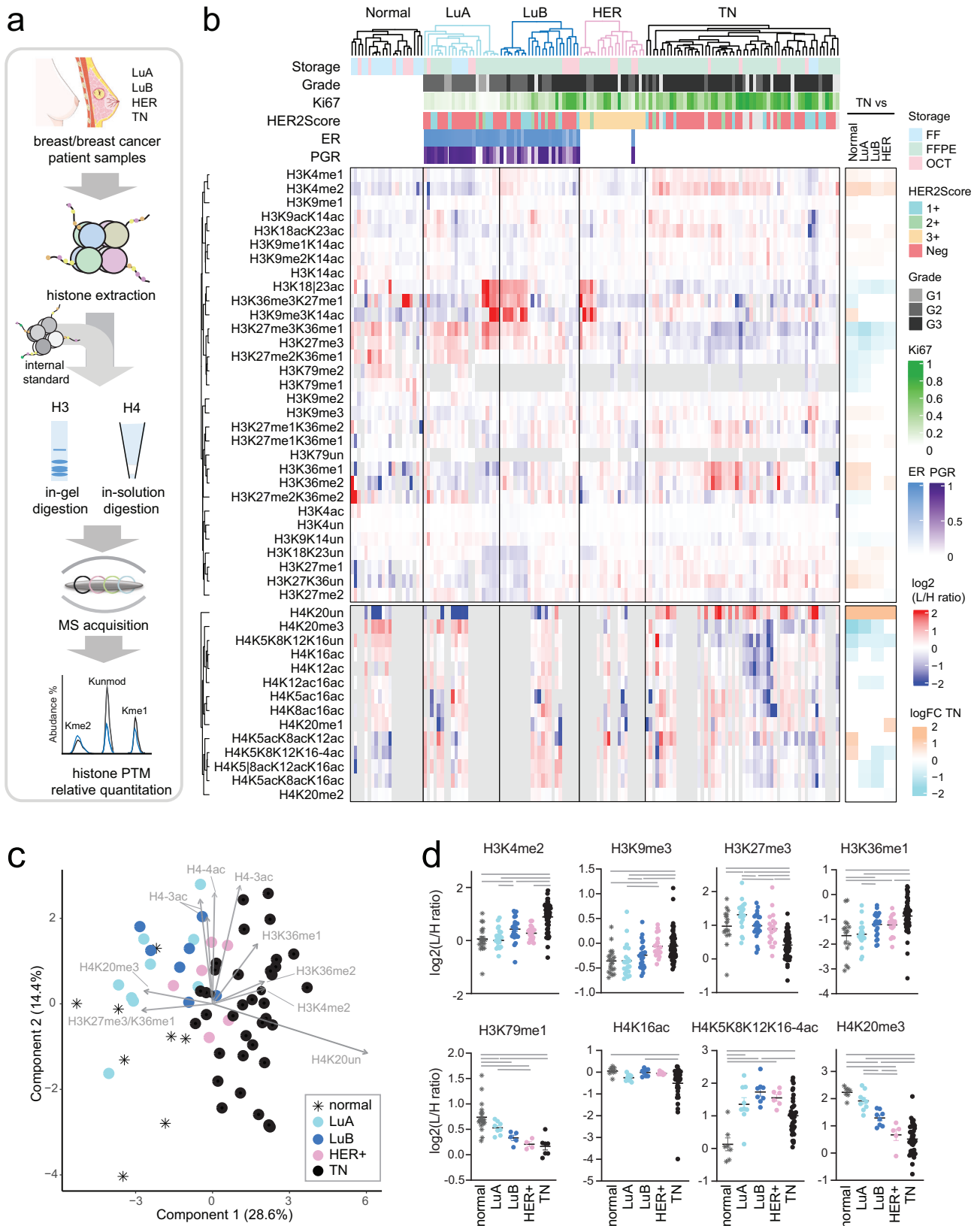
The heatmap in Fig. 1b summarizes the results from our epigenetic profiling and highlights several significant differences among the sample groups. In particular, 40 of the 45 total differentially modified peptides quantified in this cohort showed at least one significant difference in the group comparison (the right panel in Fig. 1b shows the results for the comparisons involving the TN subtype), highlighting substantial differences in the epigenetic landscape of different BC subtypes. The principal component analysis (PCA, Fig. 1c) shows that LuA samples are the most similar to normal tissue and overlap with LuB samples. HER+ tumors partially overlap with all the other subtypes. The TN group constitutes the most divergent group compared to normal tissue, and clusters separately from luminal samples. TN tumors are also the most epigenetically heterogeneous, in agreement with the known heterogeneity of this subtype²⁴. Changes in histone PTMs compared with normal breast tissues are generally few and mild in LuA samples, and are progressively more marked in LuB, HER+ and TN tumors (Fig. 1d, Supplementary Fig. 2). Interestingly, H4K20me3 and H4K16ac, whose decrease has already been reported as a general hallmark of cancer⁷, progressively decrease in BC tumors, with more marked fold changes in TN. The decrease of H4K20me3 was paralleled by a marked increase of unmodified H4K20, while the mild decrease of H4K16ac is paralleled by an increase of the hyperacetylated forms of the histone H4 tail (H4K5K8K12K16-4ac). Overall, the “epigenetic signature” of the TN subtype is characterized by an increase of H3K4 methylations (H3K4me1/me2), H3K9me3, H3K36 methylation (H3K36me1/me2) and of histone H4 hyper-acetylation. Conversely, H3K27me3-containing peptides, H3K79 methylation (H3K79me1/me2), H4K16ac and H4K20me3 decrease in TNBCs.

These results represent the first comprehensive and systematic profiling of histone PTMs in breast cancer clinical samples, and reveal specific epigenetic profiles associated with different subtypes, with TN samples showing the most divergent profiles.

Epigenetic profiling of a validation cohort comprising LuA and TN samples

To validate the findings from our epigenetic profiling, we gathered an independent cohort of 31 fresh-frozen samples, comprising 15 LuA and 16 TN cases (Supplementary Fig. 1 and Supplementary Dataset 1). The LuA subtype was chosen as the most divergent subtype - not only in terms of epigenetic profiles, but also clinical and survival features - compared to TN tumors. In this case, histones were proteolytically digested with an optimized in-gel digestion protocol involving a double derivatization of histone peptides²⁵. This protocol allows the concomitant profiling of histone H3 and H4 PTMs from lower starting amount of tissue, and improves the detection of short and hydrophilic peptides, such as H3K4me3, which could not be reliably measured in the previous sample set²⁵. This protocol also allows the analysis of several histone H2A PTMs. The results from our validation cohort are remarkably similar to those obtained from the comparison of LuA and TN samples from our first set of samples, as most of the differences observed in Cohort 1 were also present in Cohort 2 (Fig. 2a, Supplementary Fig. 3). In addition, in Cohort 2 we were able to detect an increase in the TN subtype of H3K4me3 and H2AXK5acK9ac, two modified peptides that were not detectable in Cohort 1.

Because histone PTM abundances have been shown to change during the cell cycle^{26,27}, we tested whether the differences observed in BC samples could be influenced by the different proliferation rates of the subtypes. We analyzed histone PTM levels in MDA-MB-231 TN



cells expressing the FUCCI system, which enables the isolation of cells in different cell cycle states based on the expression of specific fluorophores²⁸ (Fig. 2b and Supplementary Fig. 3c). The comparison of G2/M cells with unsorted cells showed that most of the histone PTMs changes identified in TNBCs, including H3K4me2/me3, H3K9me3, H3K79me1/me2, and H4K20me3 are independent from

the specific cell cycle phase (Fig. 2b), ruling out that the observed difference among subtypes could be explained merely by differences in proliferation rates. On the contrary, H3K27me3, H3K36me1/me2, and H4K16ac displayed in G2/M cells a change similar to that observed in TN, suggesting that increased proliferation may at least partially account for the difference observed in TN samples.

Fig. 1 | Mass spectrometry-based histone PTM profiling of breast cancer molecular subtypes. **a** Pipeline of the MS-based experiment showing histone extraction, histone derivatization, mixing with an internal standard, histone digestion (lysine acylation followed by in-gel trypsin digestion for histone H3, and in-solution Arg-C digestion for histone H4), and LC-MS/MS acquisition and quantification of histone PTMs levels in patient-derived tissues. Histone H3 PTMs were analyzed in 21 normal breast and 22 luminal-A-like (LuA), 23 luminal-B-like (LuB), 19 HER-positive (HER+) and 56 triple-negative samples. Histone H4 PTMs were analyzed in 9 normal breast and 9 LuA, 8 LuB, 6 HER+ and 37 TN samples. Breast and breast cancer cartoons were provided by Servier Medical Art (<https://smart.servier.com/>), licensed under CC BY 4.0 (<https://creativecommons.org/licenses/by/4.0/>). **b** Heatmap display of the $\log_2(L/H)$ ratios obtained for histone PTMs in normal breast or breast tumors belonging to different molecular subtypes. L/H (light/heavy) relative abundance ratios were obtained using a spike-in strategy (light channel: sample, heavy channel: spike-in standard), and were normalized to the

average ratios across samples. The storage method, grade, ER/PGR status, HER status and Ki67 index are indicated. The grey color indicates peptides that were not quantified. Pearson's correlation distance and average linkage were employed both for rows and columns clustering. The right panel highlights significant changes ($p < 0.05$ by two-tailed limma pairwise comparisons) for the comparison of the different groups to the TNBC subtype. Red indicates a significant increase, blue a significant decrease in TNBC, and white no significant changes. **c** Principal component analysis (PCA) based on histone PTM data obtained from the samples shown in **b**. Patients with a complete absence of values for histone H4 peptides were excluded. Missing values were replaced with 0. **d** $\log_2(L/H)$ ratios for selected histone PTMs from the patient samples shown in **b**. The mean is indicated for each group. Bar: $p < 0.05$ by two-tailed limma pairwise comparisons. The number of samples analyzed for each modification and the exact p -values are reported in Dataset S2. The symbol “|” indicates that the modification is present on only one of the indicated residues.

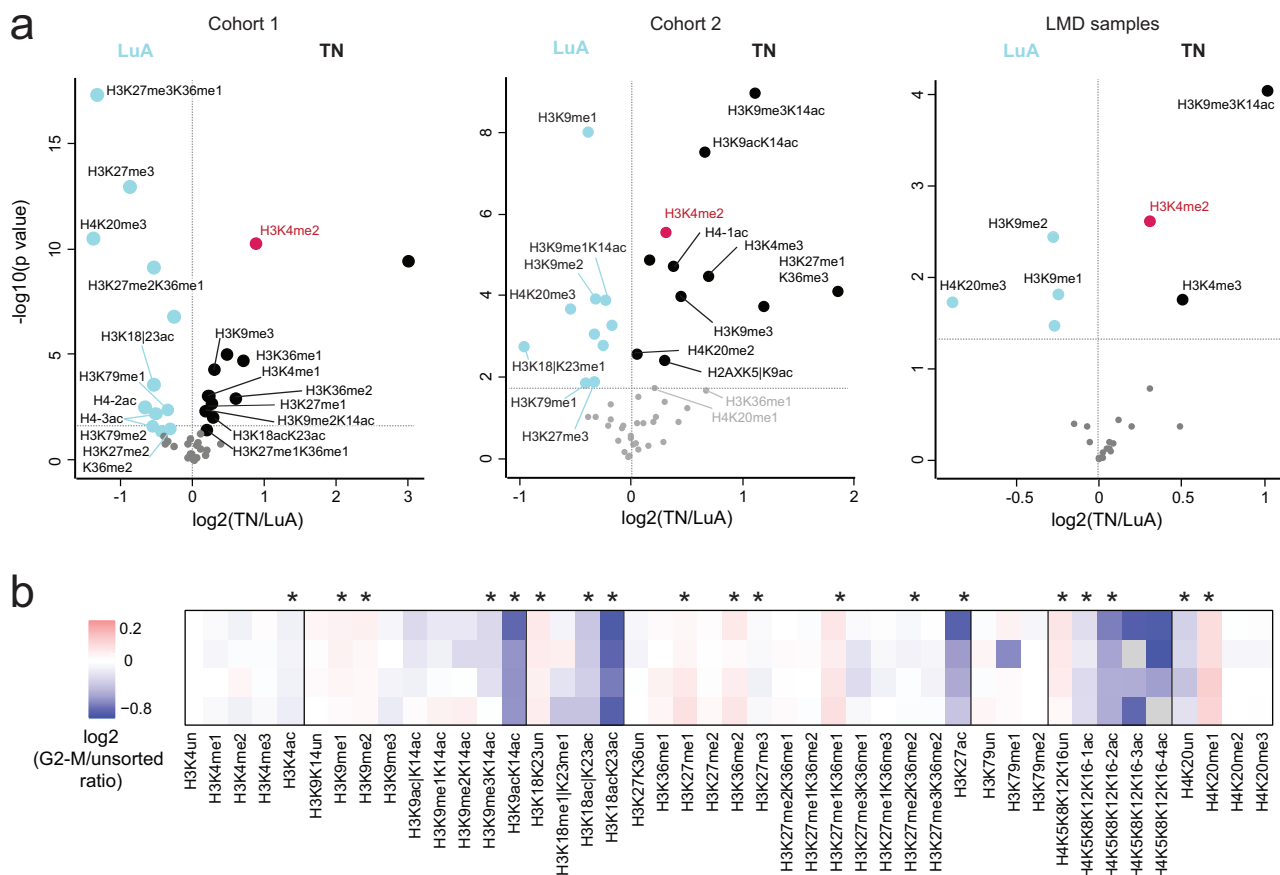


Fig. 2 | Validation of the TNBC epigenetic signature. **a** Volcano plots displaying histone PTM changes versus P values in TN compared to LuA tumors in Cohort 1 (22 LuA, 56 TN), Cohort 2 (15 LuA, 16 TN), and in a subset of laser microdissected (LMD) samples (4 LuA, 12 TN). The horizontal line represents the threshold of statistical significance ($p < 0.05$ by two-tailed limma moderated t-test). Unmodified peptides are not labeled. The symbol “|” indicates that the modification is present on only

one of the indicated residues. **b** Heatmap display of histone PTM levels in MDA-MB-231FUCCI cells, sorted in G2/M phase and normalized to unsorted cells. *: Significant changes in G2/M-sorted cells versus unsorted cells (p -value < 0.05 by two-tailed paired Student's t test). Exact p -values are reported in the Source Data file. Grey: not detected.

As an additional control, for a subset of samples (4 LuA and 12 TN tissues, Supplementary Fig. 1), we carried out an MS-based profiling of homogeneous tumor populations obtained by laser microdissection, to verify that the changes observed between subtypes are due to changes in tumor cells and not to the contribution of normal cells or of tumor microenvironment (Fig. 2a, right panel). Despite the limited number of samples profiled, the most pronounced and significant differences found in the previous experiments were confirmed also in

laser microdissected samples. These include the increase of H3K4me2/me3 and H3K9me3, and the decrease of H4K20me3. Due to the low amount of material, which limits the number of measurable modifications, PTMs on histone H3 K27, K36 and K79 could not be measured. Nevertheless, we have shown changes of H3K27 and H3K36 methylations in TN vs LuA samples in homogeneous laser microdissected population in a previous study¹³. Altogether, these results confirm the combinatorial epigenetic differences distinguishing TN

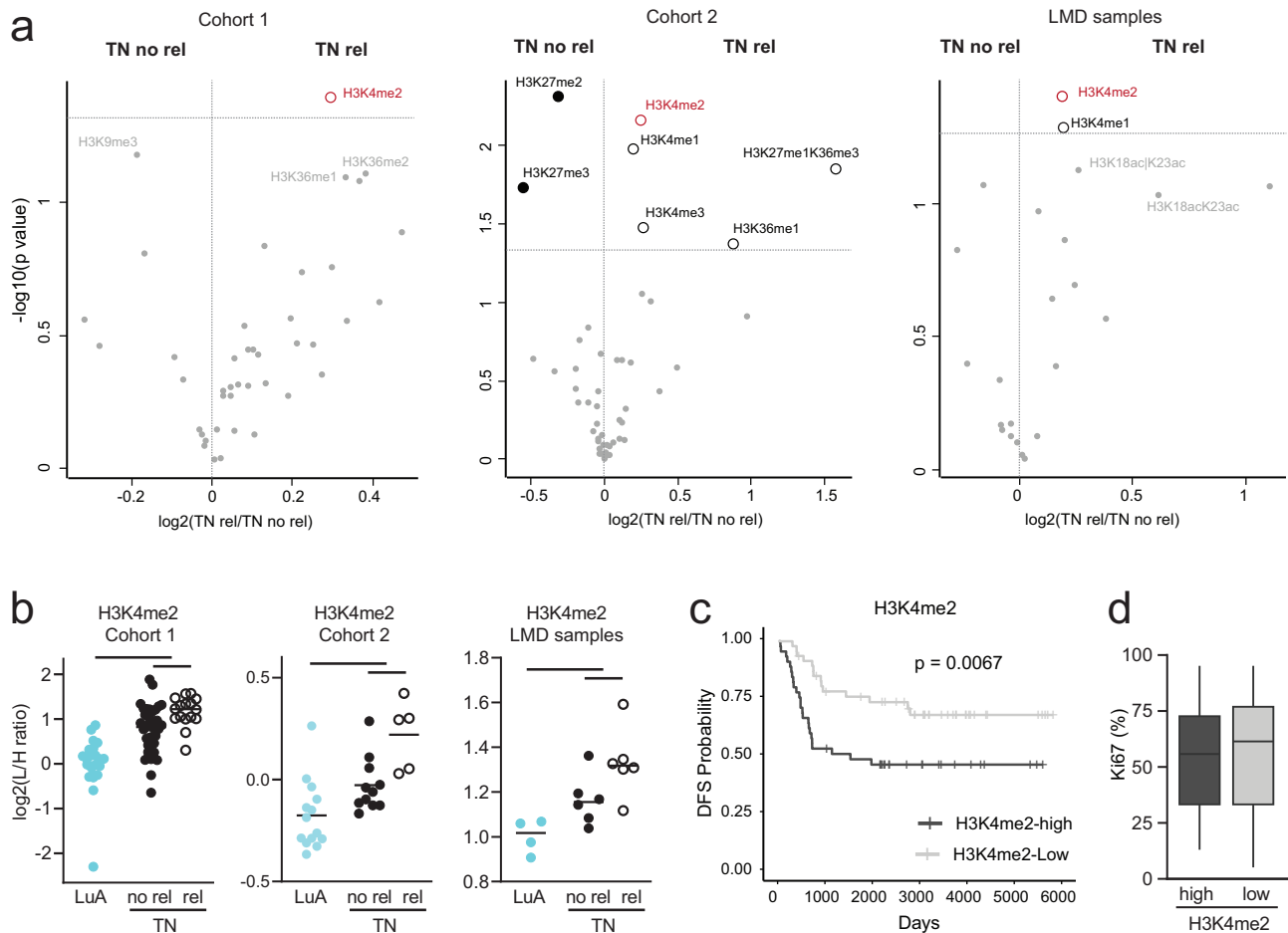


Fig. 3 | The H3K4me2 mark is associated with worse prognosis in TNBC.

a Volcano plots showing histone PTM changes versus P values in TNBCs with relapse (rel) versus TNBCs without relapse (no rel) three years after chemotherapy in Cohort 1 (29 no rel, 11 rel), Cohort 2 (11 no rel, 5 rel) and in a subset of laser microdissected (LMD) samples (6 no rel, 6 rel). The horizontal line represents the threshold of statistical significance ($p < 0.05$ by two-tailed limma moderated t -test). Unmodified peptides are not labeled. The symbol “|” indicates that the modification is present on only one of the indicated residues. **b** $\log_2(L/H)$ ratios for H3K4me2 in Cohorts 1 and 2, and in laser microdissected samples. Bar: $p < 0.05$ by two-tailed limma moderated t -test. The mean is shown for each group. Cohort 1: 22 LuA, 29 TN

no rel, 11 TN rel. Cohort 2: 15 LuA, 11 TN no rel, 5 TN rel. LMD samples: 4 LuA, 6 TN no rel, 6 TN rel. exact p values are reported in Dataset S2. **c** Kaplan-Meier curve displaying the estimated disease-free survival (DFS) probability for two groups of TNBC patients defined based on the level of H3K4me2. The analysis was performed in TNBC patients comprising Cohort 1 ($n = 56$), Cohort 2 ($n = 16$) and Cohort 3 ($n = 35$). The log-rank test indicates a significant difference between the survival curves. **d** Ki67 % in H3K4me2-high ($n = 49$) and -low ($n = 50$) samples. A box plot shows the median (center line), the 25th and 75th percentiles (box edges), and whiskers that extend to the most extreme values within 1.5 \times the interquartile range.

from LuA BC samples and indicate that they are due –for the major part– to changes that are independent from the proliferation rate and occur in the tumor cells and not in the tumor microenvironment.

Epigenetic marks show prognostic value in TNBC

We next focused on the TN subtype and asked whether histone PTM levels can stratify patients based on their survival. The comparison of the epigenetic profiles of TNBC primary tumors from patients with or without relapse 3 years after chemotherapy showed few significant/close to significant changes in histone PTMs (Fig. 3a and Supplementary Fig. 4a). Interestingly, the most marked and reproducible alteration in relapsing patients was an increase of H3K4me2, which was found significantly different in all the dataset tested, including Cohorts 1 and 2, and –remarkably– also in the laser microdissected sample group (Fig. 3a, b). In addition, H3K4me1 and me3 were increased in Cohort 2, and H3K4me1 also in the LMD set.

Additionally, we analyzed the association between histone PTMs and the probability of survival in a total of 107 TNBCs comprising 56 samples from Cohort 1, 16 from Cohort 2 and additional 35 samples (Cohort 3, Supplementary Fig. 1). H3K4me2 displayed once again the

most significant association with both overall survival and disease-free survival (Fig. 3c and Supplementary Fig. 5), thus emerging as an attractive potential prognostic marker in TNBC. Importantly, H3K4me2-high and H3K4me2-low tumors did not show differences in Ki67 levels (Fig. 3d and Supplementary Fig. 5), further suggesting that the changes observed in this PTM are not merely a reflection of proliferation changes. Interestingly, clustering of TNBC samples based on histone H3 PTMs (which was performed separately for Cohort 1 and Cohorts 2 + 3) defined two clusters that were characterized by similar histone PTM levels (Supplementary Fig. 4b), possibly suggesting that specific histone PTM patterns could characterize TNBC subgroups. One of the two clusters showed significantly higher H3K4me2, which was accompanied by higher levels of H3K4me1, H3K36 methylations, H3K18ackK23ac, and lower levels of H3K27 methylations (Supplementary Fig. 4c).

A multi-OMICs approach to dissect the downstream effects of increased H3K4me2 levels in TNBC

H3K4me2 is typically associated with positive regulation of transcription and has been shown to localize at cis-regulatory regions, such as promoters and enhancers²⁹. To study the downstream effects of

Fig. 4 | A multi-OMICs approach to dissect the downstream effects of increased H3K4me2 levels in TNBC. **a** Multi-OMICs analysis was performed in a small set of LuA and TN fresh frozen samples. ChIP-seq data from 7 LuA and 8 TN samples were integrated with RNA-seq from matched samples or TCGA data (shown) and proteomics data from matched samples/CPTAC data. **b** Multidimensional scaling (MDS) based on H3K4me2 read density at promoters and distal regions of LuA and TN samples. The distance between points indicates the leading fold-change (logFC) of the corresponding samples. The leading logFC is defined as the root-mean-square of the top 500 largest log₂ fold changes between each pair of samples. **c** Plots of the mean read density of TN (black) and LuA (light blue) in TN unique ($n = 5083$), common ($n = 21253$) and LuA unique ($n = 2936$) regions. Solid lines show the mean density. Shaded area depicts the 95% confidence interval of the mean. RPGC: Reads per Genomic Content. **d** Peak annotation of TN unique, common and LuA unique peaks, which was performed at base pair level. **e** Distribution of RNA quantification values for genes that harbor H3K4me2 peaks in promoters and enhancers only in TN, both in LuA and TN and only in LuA samples. Box plots show

the median (center line), the 25th and 75th percentiles (box edges), and whiskers that extend to the most extreme values within $1.5 \times$ the interquartile range. Outlier values are not displayed. TPM: transcripts per million. P values by two-tailed Wilcoxon rank-sum test: TN only promoters: 0.00068; TN only enhancers: 1.9×10^{-10} ; LuA only promoter: 0.002; LuA only enhancer: 3.53×10^{-13} ; Common promoter: 1.6×10^{-5} ; Common enhancer: 1.9×10^{-5} . **f** Volcano plot showing TCGA gene expression comparison of TN and LuA tumor samples. Differentially expressed genes harboring H3K4me2 peaks only in LuA or only in TN are colored in light blue and black, respectively. Dark gray colored points depict differentially expressed genes that do not harbor peaks unique to either subtype. The DESeq2 two-tailed Wald test was used to identify differentially expressed genes ($FDR < 0.05$, $\log_2\text{FoldChange} > |1|$). **g** Over-representation analysis of genes with TN-unique H3K4me2 peaks and overexpressed in the same subtype. A one-sided hypergeometric test, implemented in the R package clusterProfiler, was used for enrichment analysis. Dn: down. P.adjust: p -value adjusted for multiple comparisons using the Benjamini-Hochberg correction.

samples ($n = 5083$), and vice versa ($n = 2936$) (Fig. 4c). The majority of common peaks (approximately 41%) localized to promoter regions and transcription start sites, as previously described³⁰. Instead, H3K4me2 peaks unique to the TN or LuA subtypes were mostly localized in gene bodies and intergenic regions (Fig. 4d), suggesting that enhancers may play a more important role in defining BC subtypes, through the modulation of subtype-specific gene expression.

In order to verify whether the presence of a H3K4me2-unique peak correlates with increased transcript levels, we integrated the ChIP-seq data with LuA and TN transcriptomic data from The Cancer Genome Atlas Program (TCGA). This analysis unraveled a clear correlation between H3K4me2 presence at promoters and enhancers (retrieved from³¹) and the estimated transcript levels of the corresponding genes (Fig. 4e). Most of the common H3K4me2 peaks corresponded to genes with similar expression in the two subtypes, while unique peaks were more frequently associated to genes with higher expression in the respective subtype, both when we interrogated transcriptomics data from TCGA (Fig. 4e) and the RNA-seq data acquired from 9 samples matching those used for the ChIP-seq experiments (Supplementary Fig. 6b). Interestingly, we observed a more marked increase in expression when H3K4me2 localized at enhancers, particularly in the case of matched ChIP-seq/RNA-seq data, in agreement with the reported role of enhancers in boosting gene expression³². The intersection of genes with H3K4me2 TN specific gains with genes up-regulated at the RNA level in TN compared to LuA samples yielded 237 genes whose expression is putatively regulated by H3K4me2 (Fig. 4f). Notably, the functional categories associated with this set of genes that emerged from an over-representation analysis revealed terms related to the basal features and “core” phenotype of TNBCs (Fig. 4g). We also added to our analysis the layer of protein expression, by using the Clinical Proteomic Tumor Analysis Consortium (CPTAC) breast cancer dataset³³. The functional analysis of genes with a H3K4me2 peak unique to TN and upregulated both at the transcript and protein levels confirmed the enrichment of the same terms (Fig. 4g), which were also found by 2D annotation enrichment analysis³⁴ of matched samples for which we acquired transcriptomes and proteomes (Supplementary Fig. 6a, c).

Taken together, the intersection of ChIP-seq and transcriptomics/proteomics data revealed that H3K4me2 potentially regulates multiple genes related with the TNBC phenotype.

H3K4me2 regulates the expression of genes linked with the TNBC phenotype

To mechanistically prove the causal link between the presence of H3K4me2 and the expression of genes associated with the TNBC

phenotype that emerged from our multi-OMICs analysis, we took advantage of the CRISPR interference (CRISPRi) technology, which allows epigenetic editing and targeting of specific genomic regions of interest by using protein domains specific for histone PTMs removal or deposition³⁵. Here we employed a previously reported CRISPRi system that is particularly well suited for our needs³⁶, as it involves a deactivated d-Cas9 fused with LSD1, which is a demethylase specific for H3K4me2. We focused on six representative genes, three of which present TN-specific H3K4me2 peaks at promoters (*PLA2G4A*, *IGF2BP3*, *AIM2*) and three at enhancers (*GLS*, *GLIPR2*, *PLCG2*). The selection of these genes satisfied several criteria: (1) high expression fold change in TN compared with LuA samples at both the gene and protein levels (Figs. 5a), (2) the presence of a peak with a strong enrichment to be targeted by CRISPRi (Fig. 5b and Supplementary Fig. 7), and (3) a reported implication in the development or progression of cancer. *PLA2G4A* is a phospholipid metabolizing enzyme overexpressed in BC that releases free fatty acids and that contributes to the development of the tumor microenvironment, promoting immune evasion, angiogenesis, tumor growth, and invasiveness³⁷. *IGF2BP3* (Insulin-like Growth Factor 2 mRNA-Binding Protein 3) is an RNA-binding protein that has been described to contribute to the aggressive behavior of TNBCs and to BC chemo-resistance³⁸. *AIM2* promotes tumor growth by fostering a pro-inflammatory tumor microenvironment³⁹ and has been proposed as a negative prognostic marker for TNBC with brain metastases⁴⁰. *GLS* encodes glutaminase, a metabolic enzyme upregulated in highly proliferative tumors like TNBC⁴¹. *GLIPR2* (GLI pathogenesis-related 2) has been associated with positive regulation of epithelial cell migration and with epithelial to mesenchymal transition⁴². Finally, *PLCG2* is involved in inflammation and immune response, and its overexpression is linked to increased migration and invasion in vitro, and higher metastatic potential in vivo⁴³.

We engineered the MDA-MB-231 cell line to stably express d-Cas9, which was tagged with the mCherry fluorescent marker, and fused to the effector domain of LSD1 (dCas9LSD1) (Supplementary Fig. 8). We designed two guide RNAs (sgRNAs) for each target gene (Supplementary Table 1) in order to target the promoter or the enhancer region occupied by H3K4me2. A ChIP-PCR analysis showed decreased levels of H3K4me2 at all the targeted regions, with an effect that was more pronounced when the dCas9-LSD1 system was directed at enhancers (Fig. 5c). The decrease of H3K4me2 was paralleled by a significant reduction in gene expression upon the epigenetic modulation of all the genes tested, mechanistically proving that the presence of H3K4me2 at these regions directly regulated their expression (Fig. 5d).

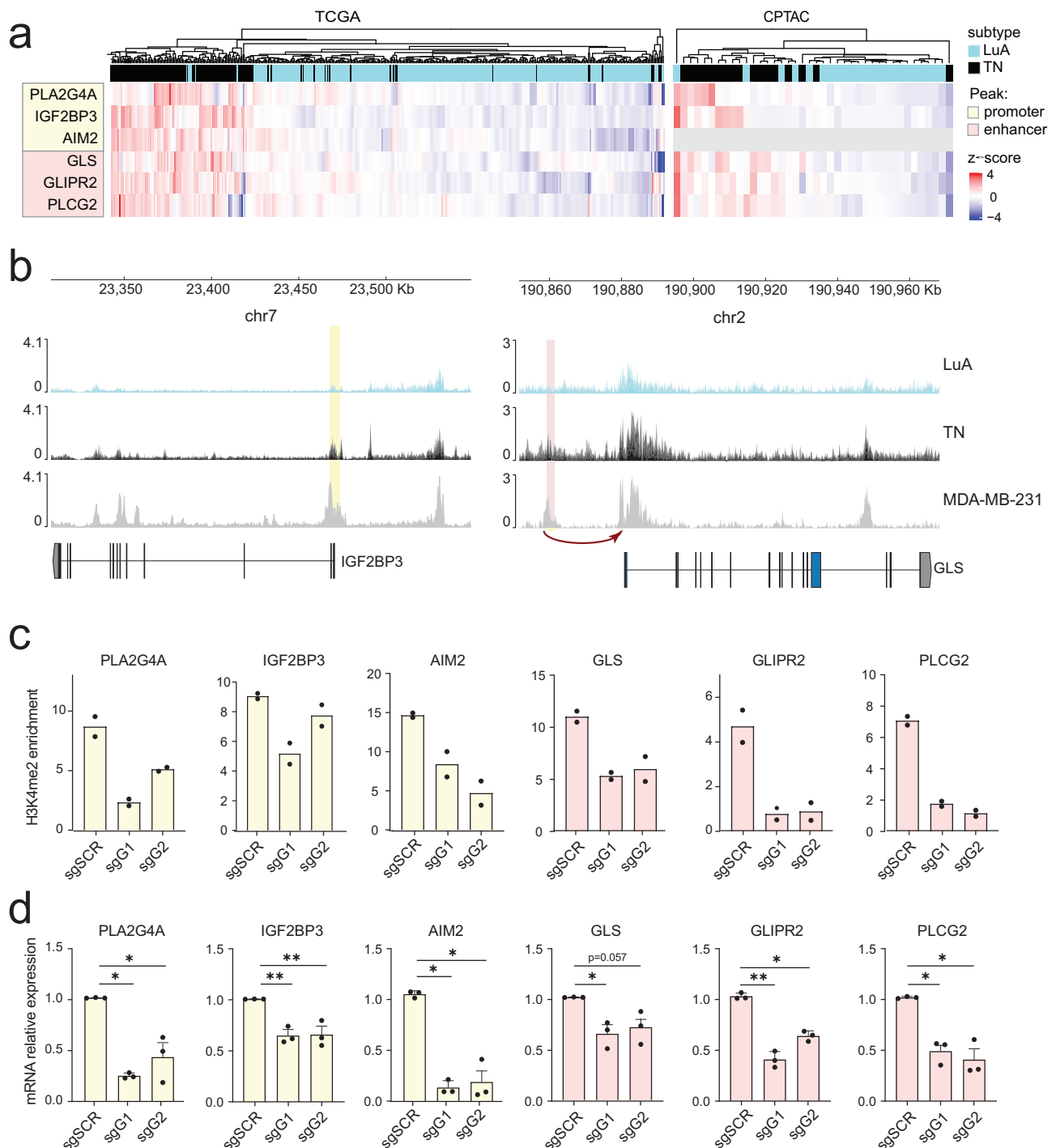


Fig. 5 | H3K4me2 regulates the expression of genes linked with the TNBC aggressive phenotype. a Heatmap showing the gene expression levels (from TCGA data, left panel) and protein expression levels (from CPTAC data, right panel) of six genes selected for CRISPRi. For column clustering, Pearson's correlation distance and average linkage were used. AIM2 was absent in the CPTAC dataset. **b** Representative genomic tracks of H3K4me2 ChIP-seq for two genes selected for CRISPRi. Left panel: IGF2BP3, peak at TSS. Right panel: GLS, peak at enhancer. The promoter region is highlighted in yellow. The region highlighted in red represents the distal genomic region found in TN samples but not in LuA samples and

confirmed in the MDA-MB-231 cell line. **c** H3K4me2 ChIP-qPCR profiling of promoters or enhancers in genes targeted by sgRNA ($n=2$ biological replicates). Data are displayed as percentage of enrichment relative to the input. **d** Quantitative PCR analysis of gene expression levels in MDA-MB-231 cells infected with sgRNAs for the indicated genes ($n=3$ biological replicates). Bar graphs represent mean \pm standard error (SEM) from 2-3 biological replicates. $**p < 0.01$, $***p < 0.001$ by one-way ANOVA followed by Dunnett's test. Exact p -values are reported in the Source Data file. The cells were induced for 72 h by adding 1 μ g/ml Doxycycline to the culture medium.

Inhibiting H3K4 methyltransferases reduces TNBC cell growth in vitro and in vivo

Our results showed that H3K4me2 is increased in TNBCs compared with normal tissues and tumors belonging to other BC molecular

subtypes, and that the increase in this epigenetic mark is associated with worse overall and disease-free survival in TNBCs. In addition, we found that the increase in H3K4me2 at specific genomic regions in TNBC favors the expression of a set of genes that may drive the TNBC

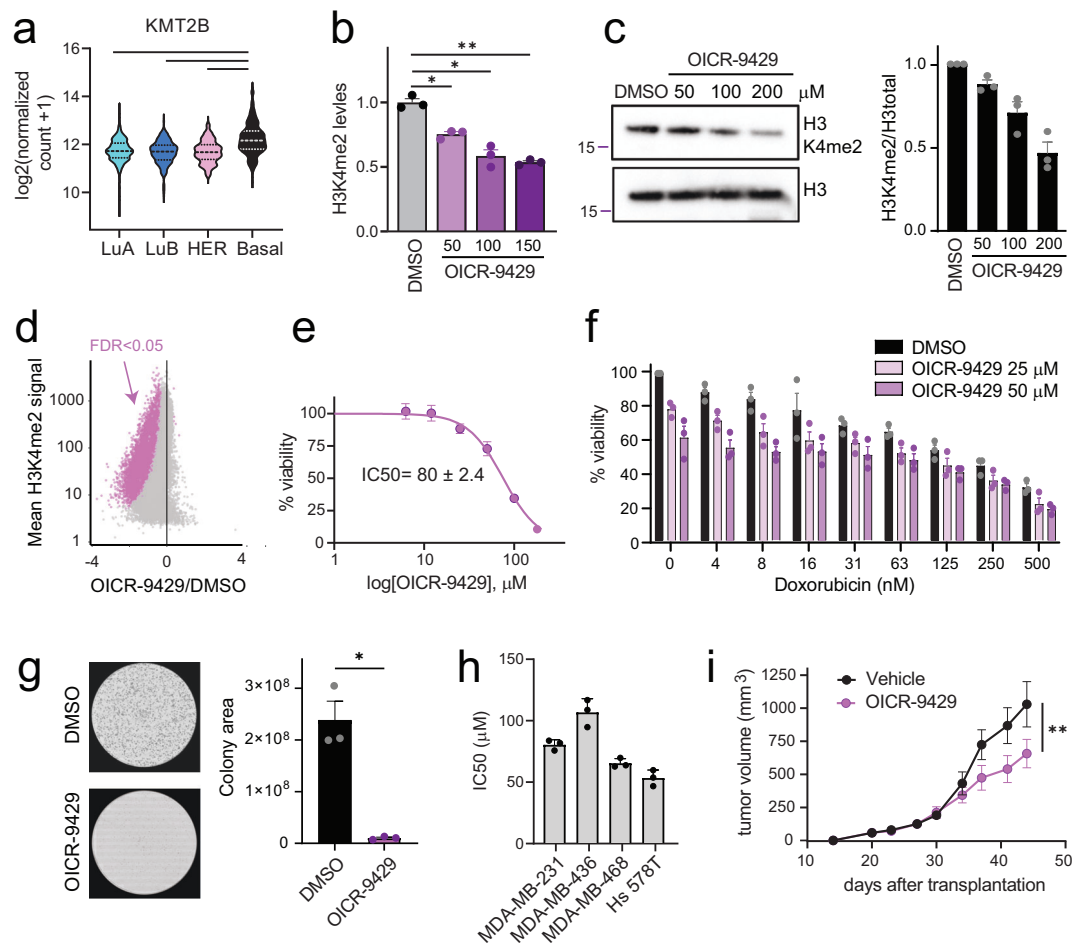


Fig. 6 | Inhibiting H3K4 methyltransferases reduces TNBC cell growth in vitro and in vivo. **a** KMT2B mRNA levels in different BC subtypes (571 LuA, 209 LuB, 82 HER+ and 197 TN). P_{adj} Basal vs LuA=1.22E-39, Basal vs LuB=1.38E-35, Basal vs HER2=8.51E-19 by DESeq2 two-tailed Wald test. **b** H3K4me2 levels measured by MS in MDA-MB-231 cells upon treatment with increasing concentrations of OICR-9429. The data were normalized to the DMSO condition. P values by one-way ANOVA, followed by Dunnett's test for DMSO vs OICR-9429 50, 100 and 150 μM were 0.0074, <0.00001 and 0.0001, respectively. **c** Western Blot showing H3K4me2 and total histone H3 levels in MDA-MB-231 cells upon treatment with OICR-9429. The bar plot on the right shows H3K4me2 normalized to total histone H3 levels and the DMSO condition. **d** MA plot illustrating H3K4me2 differentially bound regions upon OICR-9429 treatment versus DMSO control ($n=3$ biological replicates). The x-axis represents the \log_2 fold change in H3K4me2 signal, while the y-axis shows the mean average signal across both conditions. 17415 regions out of

51921 (34%) were significantly downregulated. Regions that are significantly ($\text{FDR} < 0.05$ by DESeq2 two-tailed Wald test) differentially bound are highlighted in pink. **e** Viability dose-response curve for MDA-MB-231 cells treated with increasing concentrations of OICR-9429. **f** Percentage of viable MDA-MB-231 cells after treatment with OICR-9429 and doxorubicin. The data were normalized to the untreated condition. **g** Colony formation assay in MDA-MB-231 cells treated with DMSO (control) or 50 μM OICR-9429 for 72 hours. The cells were stained with crystal violet. $*p = 0.018$ by two-tailed paired t-test ($n = 3$). **h** IC_{50} values from the viability assays performed in TN cell lines in the presence of OICR-9429. **i** Tumor growth curves for mice subcutaneously injected with 5×10^5 MDA-MB-231 cells and treated with vehicle ($n = 6$) or OICR-9429 30 mg/kg ($n = 7$). Data are shown as mean \pm s.e.m. $**p = 0.0064$ by two-way ANOVA. In b, c, e-h the data are presented as mean values \pm s.e.m. from $n = 3$ biological replicates (cells treated and collected in independent experiments).

aggressive phenotype. Interestingly, one of the methyltransferases responsible for the deposition of this PTM, *KMT2B* (also known as MLL2) is increased in TNBC compared with all the other BC subtypes (Fig. 6a and Supplementary Fig. 9). Therefore, we reasoned that decreasing the levels of H3K4me2 by pharmacological treatment could reduce the expression of multiple genes linked to the TNBC phenotype and therefore may represent a novel therapeutic strategy for the treatment of this tumor. Several inhibitors have been developed to target the methyltransferases responsible for the deposition of H3K4 methylation⁴⁴. Here, we tested the compounds OICR-9429, MI-136, MI-463, and MI-503. OICR-9429 functions as an antagonist of the WDR5 domain, competing with the interaction between WDR5 and K4-specific methyltransferases MLL (including *KMT2B*) and thus resulting in the inhibition of histone H3K4 methylation. OICR-9429 has been proposed for treatment in a range of cancer types, including tumors of hematological origin, such as non-MLL-rearranged leukemia, and solid

tumors, and has already been used in vivo^{44,45}. MI-136, MI-463 and MI-503 are instead inhibitors of the menin-MLL interaction^{46,47}.

In MDA-MB-231 cells, treatment for 72 hours with OICR-9429 reduced the level of H3K4me2 in a dose-dependent manner, as assessed by MS (Fig. 6b), immunoblot (Fig. 6c and Supplementary Fig. 10a) and ChIP-seq analysis (Fig. 6d). We then used H3K4me2 ChIP-seq data to assess the effect of OICR-9429 on a small set of genes associated with tumor aggressiveness, which were identified through our multi-OMICS analysis. They included the genes analyzed in Fig. 5 and *CHRM3*, *GABRE*, *TMEM71* and *TRVP4*⁴⁸⁻⁵¹. OICR-9429-treated cells showed a decreasing trend in H3K4me2 levels and a corresponding reduced gene expression in 8 of the 10 genes tested (Supplementary Fig. 10b, c). OICR-9429 reduced cell viability with an IC_{50} of 80 μM after 72 hours of treatment (Fig. 6e). While the menin inhibitors appeared to be more potent (IC_{50} values comprised between 3 and 6 μM), they showed a significant reduction of H3K4me2 levels only at doses that

exceeded the inhibitor IC₅₀ values determined in cell viability assays (Supplementary Fig. 11a, b).

We therefore focused on OICR-9429, which we further characterized at the phenotypic level. OICR-9429 inhibited the growth of MDA-MB-231 cells in combination with doxorubicin, showing an additive effect (Bliss synergy score = -2.55, Fig. 6f), and dramatically reduced the cell colony forming ability (Fig. 6g and Supplementary Fig. 10d). Furthermore, OICR-9429 reduced cell viability in three additional TN cell line models (MDA-MB-436, MDA-MB-468 and Hs 578 T, Fig. 6h and Supplementary Fig. 11c), as well as colony formation and cell viability in combination with chemotherapy in Hs 578T cells (Supplementary Fig. 12). Finally, treatment with OICR-9429 significantly reduced tumor growth in a MDA-MB-231 mouse xenograft model (Fig. 6i).

Discussion

Our epi-proteomics analysis of BC subtypes revealed a TNBC-specific histone PTM profile, which showed differences compared with other subtypes that were reproducible across different independent patient tissue cohorts. This first comprehensive and quantitative description of the epigenomics landscapes of BC subtypes was made possible thanks to MS-based approaches, which enabled us to profile all the best-characterized histone marks at once, in a highly quantitative manner. This represents a major advancement compared to antibody-based methods, such as immunohistochemistry, which are more commonly employed to study histone PTMs in clinical samples, but suffer from a number of limitations, including cross reactivity, antigen masking, poor quantitative accuracy and a limited number of simultaneously analyzable epigenetic marks. Our analysis primarily focused on methylations and acetylations on histone H3 and H4, which have been extensively characterized and can be associated with a functional outcome on gene expression. However, modifications localized on other histones or uncommon and less characterized PTMs may also contribute to the distinct epigenetic landscapes of breast cancer subtypes.

TNBCs are characterized by loss of H4K20me₃, which is paralleled by an increase of unmodified H4K20, compared with all the other subtypes except the HER⁺. H4K20me₃ is generally localized at repetitive regions and is associated with transcriptional repression⁵². It has been defined to be a hallmark of cancer⁷ and was found to associate with lower disease-free survival in breast cancer samples⁵³. Despite the widespread changes observed in cancer, its role has not yet been clearly characterized, possibly due to its peculiar localization in repetitive regions that are difficult to sequence with standard methods. TN samples also show a consistent increase of H3K9me₃ and H3K36 methylations and a decrease of H3K27me₃ and H3K79 methylations. These results confirm and expand previous findings obtained by our group in proof-of-concept studies involving histone H3 profiling of small patient cohorts^{13,15}. In addition, higher H3K9me₃ levels were observed in multiple tumors compared to their normal counterparts, including TNBCs¹⁴. Loss of H3K27me₃ has been previously reported to be a predictor of poor outcome in breast cancer and other cancer types⁵⁴, and has been recently associated with chemo-resistance in TNBC⁵⁵. However, a decrease in H3K27me₃ levels in TNBC might depend at least partially on the high proliferation rates that characterize TNBC tumors, as we found H3K27me₃ to be decreased in cells in the G2/M phase. This aspect has been previously investigated with discordant results, showing either a decrease²⁶ or a slight increase²⁷ of global acetylation and H3K27/K36 methylation during the G2/M phase.

Despite our follow-up work focused in particular on dissecting the downstream consequences of H3K4me₂ increase in TNBC, our epigenetic profiling highlighted several other changes worth of further investigation, both as potential biomarkers and as the starting point for mechanistic investigations aimed at identifying novel therapeutic avenues. Histone PTM changes in cancer are determined by multiple

factors, but -in the simplest scenario- they are the consequence of alterations in the levels of the histone modifying enzymes responsible for their deposition and removal, which can be targeted for therapeutic intervention with epigenetic drugs. When searching for possible associations between the gene expression of histone modifying enzymes and the changes observed in the levels of histone PTMs in the TNBC subtype (Supplementary Fig. 9), we found a remarkable correlation between the increase of H3K9me₃ and the levels of its methyltransferases and demethylases. Indeed, the H3K9me₃ methyltransferases *SUV39H1*, *SUV39H2*, and *SETDB1* were increased in TNBC, while the H3K9-specific demethylases *KDM3B* and *PHF2* decreased. The increase of H3K36 methylation in TNBC was also paralleled by an increase of the H3K36-specific methyltransferase *SMYD2*. On the contrary, H3K27-methyltransferase *EZH2* increased in TNBCs, oppositely from the decrease in H3K27me₃ levels. As previously mentioned, H3K27me₃ could be influenced by the proliferative status of the cells. Additionally, in other tumor types non-enzymatic reactions driven by the metabolic condition of the tumor cells have been shown to cause lower levels of H3K27me₃⁵⁶. An additional factor concurring to the level of histone PTMs includes the inter-dependence of histone PTM and DNA methylation levels⁵⁷⁻⁵⁹. This contribution may be particularly relevant in the case of H4K20me₃, whose loss in cancer has been associated with DNA hypomethylation in repetitive sequences⁷. Because the decrease of H4K20me₃ in TNBC is not paralleled by a concordant change of its specific histone modifying enzymes, it is possible that is influenced by DNA methylation, which would be interesting to study in this context as a future perspective.

Among the histone PTMs emerging from our BC profiling, H3K4me₂ stood out as the most interesting mark. Indeed, H3K4me₂ was increased in TNBC compared with all the other BC subtypes, was consistently identified in all the cohorts analyzed, including laser microdissected samples, and was associated with diminished disease-free survival in TNBC. Consistent with the higher levels of H3K4me₂ in TN tumors, the H3K4-specific *KMT2B* methyltransferase is also increased in this subtype. While increased *KMT2B* expression may contribute to the aberrant regulation of H3K4 methylation in TNBC, the specific mechanisms remain to be elucidated, and likely involve complex regulatory networks and multiple contributing factors. For instance, we recently found that a truncated and catalytically inactive isoform of the *KDM5B* demethylase accumulates in breast cancer cells and leads to increased bulk H3K4 methylation⁶⁰, which could also influence H3K4me₂ levels.

To investigate the downstream effects of the changes observed in H3K4me₂ levels, we exploited the complementarity of epi-proteomic data, which provides quantitative but bulk histone PTM levels, with more conventional epi-genomic and transcriptomic methods, in order to study H3K4me₂ genomic localization and its effects at the gene expression level. The multi-OMICS analysis of a small set of clinical samples that we carried out showed that genes displaying H3K4me₂ peaks uniquely in TN samples presented higher expression levels in this subtype. While this applies to both promoters or enhancers, it appears like the differences between the TN and LuA samples based on peak distribution (MDS in Fig. 4b) and gene expression (Fig. 4e) are more marked when considering enhancers. This is consistent with the notion that enhancers 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 characteristics⁶¹. In our study, we proved a direct, causative link between the expression of representative genes involved in TNBC phenotype and the presence of H3K4me₂ at their promoters or enhancers by using a CRISPRi system.

Because our data suggests that reducing the levels of H3K4me₂ could reduce the expression of several pro-disease genes, we tested several inhibitors targeting the interaction of MLL methyltransferases with WDR5 or menin. While menin inhibitors were more potent in

phenotypic assays, they did not elicit significant changes in H3K4me2 levels at IC_{50} concentrations. This may be explained by recent findings showing that the menin–MLL complex occupies fewer than 200 active promoters in leukemia cells, with only a subset of these being affected by menin inhibitors⁶², potentially accounting for the absence of a measurable global change in H3K4me2. Alternatively, the observed effects on cell viability may result from off-target mechanisms. On the contrary, while displaying substantially higher IC_{50} values, OICR-9429 induced a marked reduction in bulk H3K4me2 levels and decreased 34% of H3K4me2 peaks at concentrations comparable to those affecting cell viability, thus representing a better candidate to globally decrease this histone modification. OICR-9429 was able to reduce the viability of different TN cell lines in vitro, both as a single agent and in combination with chemotherapy, and inhibited tumor growth in vivo. Of note, OICR-9429 also inhibited cell growth in LuA and HER+ cell line models (Supplementary Fig. 11d, e), potentially suggesting that inhibiting H3K4 methyltransferases may be beneficial in other BC subtypes. We are aware that these inhibitors do not selectively target H3K4me2, but also H3K4me3. However, we found that H3K4me3 is also significantly increased in TN compared with LuA in both Cohort 2 and the laser-microdissected samples, and showed increased levels in TNBCs with worse prognosis in Cohort 2. Therefore, inhibiting the deposition of both methylation states may actually be beneficial, rather than detrimental. Because the levels of H3K4me2 are rather heterogeneous within the TNBC group, and higher levels associate with worse prognosis, we envision that stratifying tumors based on H3K4 methylation levels may aid the identification of patients that could particularly benefit from treatment with this type of inhibitors. It must be noted that OICR-9429 requires relatively high concentrations to elicit measurable effects, which limits its potential for therapeutic applications. Additional, more potent inhibitors with broad effects on H3K4 methylation may be tested, also in combination with standard therapies.

Overall, our MS-based profiling integrated with a multi-OMIC analysis of clinical samples identified H3K4me2 as a potential epigenetic driver of TNBC, suggesting strategies to reduce its levels as a therapeutic avenue for the treatment of this tumor. In particular, the inhibitor OICR-9429, which reduces H3K4 methylation levels and has been shown in our study to inhibit tumor growth in a xenograft model, may warrant further investigation in preclinical and clinical settings.

Methods

Ethical statement

The research described complies with all relevant ethical regulations. Breast cancer and normal breast samples were obtained from patients undergoing surgery at the European Institute of Oncology (Milan). The patients provided informed consent and this study was approved by the Ethical Committee of the European Institute of Oncology (Study UID 2550).

The animal experiments were approved by the Italian Ministry of Health (Project no. 679/20 and 485/2025), conducted in accordance with Italian law and under the control of institutional (European Institute of Oncology) local animal welfare (Cogentech Organismo Preposto al Benessere degli Animali (OPBA)) and ethical committees.

Tissue specimens

The levels of hormone receptors, HER-2, and Ki-67 of tumor samples were ascertained by immunohistochemistry, and breast cancer subtypes were defined in accordance to the ASCO/CAP guidelines. Samples with infiltrating carcinoma were selected to have a tumor cellularity of at least 40% for histone PTM analysis, as assessed by hematoxylin and eosin (H&E) staining. The samples subjected to the multi-OMICS analysis had a minimum tumor cellularity of 60%. Specimens with in situ carcinoma areas, large necrosis areas, and massive fogistic infiltrate were discarded. For storage, samples were collected

and snap frozen in liquid nitrogen, frozen in optimal cutting temperature compound (OCT) or fixed overnight in 4% formalin and embedded in paraffin. In some cases, homogenous tumor populations were isolated from fresh-frozen tissue sections by laser capture microdissection (25000–100000 cells were collected per sample)²⁵. The list of patient samples analyzed in this study, and the storage method used are summarized in Supplementary Dataset 1.

Sample preparation for MS analysis

Histones were enriched from tissue samples through the PAT-H-MS method²¹, and from cell lines through nuclei isolation²¹. Prior to digestion, histone samples were mixed with a heavy-labeled histone super-stable isotope labeling by amino acids in cell culture (SILAC) mix, which was used as an internal standard for quantification^{15,22}. The super-SILAC mix was generated from a mix of heavy isotope-labeled breast cancer cell lines²² or a mix of cancer cell lines²³ (for microdissected samples and treated cell lines). About 3 μ g of histones per run per sample were mixed with an approximately equal amount of super-SILAC mix and separated on a 17% SDS-PAGE gel. In case of laser microdissected tissues, the whole protein extract was loaded on the gel together with 1.5 μ g of super-SILAC mix. For in-gel digestions, a band corresponding to the histone octamer was excised, chemically acylated, and digested with trypsin. Chemical acylation was performed with deuterated acetic anhydride¹⁵ for Cohort 1, and with propionic anhydride for the other samples analyzed. After elution, the samples derivatized with propionic anhydride were also derivatized with phenyl isocyanate (PIC)²⁵. The samples were desalted on handmade StageTips¹⁵. A subset of samples (Supplementary Fig. 6A) was also digested in solution using the iST Sample Preparation Kit (PreOmics), for total proteome analysis. The digestion was performed following the manufacturer guidelines, starting from 100 μ g of protein extracts.

LC-MS/MS analysis

Peptide mixtures were separated by reversed-phase chromatography on an EASY-nLC 1200 high performance liquid chromatography (HPLC) system through an EASY-Spray column (Thermo Fisher Scientific), 25-cm long (inner diameter 75 μ m, PepMap C18, 2 μ m particles), which was connected online to a Q Exactive Plus or HF (Thermo Fisher Scientific) instrument through an EASY-Spray™ Ion Source (Thermo Fisher Scientific). Solvent A was 0.1% formic acid (FA) in ddH₂O and solvent B was 80% ACN plus 0.1% FA. Histone samples were injected in an aqueous 1% TFA solution at a flow rate of 500 nL/min and were separated with a 50-min linear gradient of 0–35% solvent B for samples derivatized with deuterated acetic anhydride, or a 50-min linear gradient of 10–45% for PRO-PIC digested samples. The Q Exactive instruments were operated in the data-dependent acquisition (DDA) mode to automatically switch between full scan MS and MS/MS acquisition. 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 states comprised between 2 and 4 were sequentially isolated to a target value for MS1 of 3×10^6 and fragmented by HCD 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 1×10^5 . The dynamic exclusion time was set to 10 sec, and the standard mass spectrometric conditions for all experiments were as follows: spray voltage of 1.8 kV, no sheath and auxiliary gas flow. The samples for whole proteome analysis were acquired as described⁶³.

MS data analysis

The acquired histone RAW data were analyzed using EpiProfile 2.0⁶⁴, selecting the SILAC option, or manually¹⁵. For the manual analysis, the RAW data were analyzed using the integrated MaxQuant software v.1.6.2.10, against the UniProt human proteome UP000005640, downloaded on 15/02, filtered to retain only histone sequences¹⁵.

Identifications, retention times and elution patterns were used to guide the manual quantification of each modified peptide using QualBrowser version 2.0.7 (Thermo Fisher Scientific). For each histone-modified peptide, the % relative abundance (%RA) was estimated by dividing the area under the curve (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. The AUC values for all the samples analyzed are reported in Supplementary Dataset 2. Data from our previously published pilot studies were also included in the analysis (Supplementary Dataset 3). Heatmap displays were generated using the R package ComplexHeatmap⁶⁵ from mean-centered L/H ratios. Outliers were identified using the boxplot method and subsequently removed. Statistical testing was carried out via the R package limma⁶⁶, specifying the parameter robust = TRUE to the eBayes function. Whole proteome data were analyzed using MaxQuant and the UniProt human proteome UP000005640 comprising canonical and isoform sequences, downloaded on 22/06. The TN and LuA proteome data were retrieved from Proteomics Data Common (<https://proteomic.datacommons.cancer.gov/pdc>) with PDC Study ID PDC000173. A Two-component normalization of iTRAQ ratios was carried out as described in the original publication³³. Differential expression analysis was performed via the R package DEqMS⁶⁷ using the normalized iTRAQ ratios and PSM count tables as input. Overexpressed proteins in TN samples were identified after applying a spectral count-adjusted p-value cut-off of 0.05 and log₂ fold change greater than 1.

Chromatin Immunoprecipitation (ChIP) and ChIP-PCR

The chromatin was prepared from 1.5×10^7 cells, or 25–30 mg of fresh frozen tissues. The tissues were homogenized in PBS with protease inhibitors (0.5 mM PMSF, 5 μ M Leupeptin, 5 μ M Aprotinin, 5 mM N-butyrate) using a Dounce homogenizer. The samples were fixed in 1% v/v formaldehyde (Sigma, F8775) at room temperature for 10 minutes, followed by the addition of 125 mM Glycine (Serva, SE39004) for 5 min. The samples were washed twice in PBS and then resuspended in ChIP SDS lysis buffer (100 mM NaCl, 50 mM Tris-Cl pH 8.1, 5 mM EDTA pH 8, 0.2% Na₃N, and 0.5% SDS). Nuclei were pelleted for 6 minutes at 16000 rcf rpm at 4 °C. The samples were resuspended in 3 ml (cell lines) or 1 ml (fresh frozen tissues) ice-cold IP buffer (100 mM Tris-Cl pH 8.6, 100 mM NaCl, 5 mM EDTA pH 8.0, 0.2% Na₃N, 5% (v/v) Triton 100X) and sonicated with a Branson Digital Sonifier 250, with a 3 mm microtip for fresh frozen samples or a large tip for cell lines (30% amplitude, 15 second on/15 second off), in order to obtain a DNA length of 200–300 bp. 200–400 μ g of proteins were incubated overnight at 4 °C with an anti-H3K4me2 antibody (ab7766, Abcam, 1:100) or an IgG control (normal rabbit IgG, Merck Life Science S.R.L.), followed by a 3 hour incubation with Dynabeads Protein G (Invitrogen, 10004D). When indicated, a 5% *Drosophila* spike-in (generated as described⁶⁸) was added before the incubation with antibodies. The samples were washed twice with ChIP wash buffer 150 mM (1% Triton 100, 0.1% SDS, 150 mM NaCl, 2 mM EDTA pH8.0, 20 mM Tris-Cl pH 8.0), once with wash buffer 500 mM (1% Triton 100, 0.1% SDS, 500 mM NaCl, 2 mM EDTA pH8.0, 20 mM Tris-Cl pH 8.0), and once in TE 1x. The DNA was eluted and de-crosslinked in SDS 2% in TE 1x and Proteinase K (Merck Life Science S.R.L.), overnight at 65 °C. DNA fragments were purified with the Qiagen PCR purification kit (Qiagen, 28106). ChIP-PCRs were performed in two biological replicates and quantitative PCR amplifications were carried out in the Vii7 PCR machine (BioRad). The IP enrichment compared to the input was calculated as % of input $\times 2^{-(\Delta Ct)}$ where $\Delta Ct = Ct_{input} - Ct_{IP}$. The primers used for ChIP-PCR are listed in Supplementary Table 2.

ChIP Sequencing (ChIP-seq) and ChIP-seq analysis

DNA libraries were prepared from 5–10 ng of input and IP DNA through an in-house protocol⁶⁹ and sequenced on a NovaSeq 6000 (Illumina)

instrument with a read length of 50–60 bp PE (Paired End) and a sequencing depth of 50 million reads per sample. Reads were mapped to the hg38 version of the reference human genome assembly using Bowtie⁷⁰, keeping only uniquely mapping reads and allowing for no more than two mismatches per read. The options “–best” and “–strata” were enabled. The maximum insert size for valid paired-end alignments was set to 700 bp. To identify H3K4me2 enriched regions, MACS2 was employed⁷¹, enabling the –BROAD option and setting both the q-value and the BROAD-CUTOFF parameter to 0.001. LuA-only peaks were defined as those peaks overlapping in four or more LuA samples and with no overlap with any of the TN samples. Analogously, TN-only peaks were peaks overlapping in four or more TN samples and with no overlap with any of the LuA samples. Common peaks were peaks found in at least four LuA samples and in at least four TN samples. These overlaps were performed using BEDOPS⁷². Genomic annotations of peaks were performed at base pair level using bedtools⁷³ and using the RefSeqCurated (hg38, release date: 2021-05-28) genome annotation. Peaks were associated to gene promoter regions if overlapping in a window of 3 kb centered around the TSS of a gene. Distal peaks were associated to genes according to the TCGA-BRCA ATAC-seq peak-to-gene links retrieved from³¹. To identify differences in the reads density distribution either at promoters or at distal regions (peaks not falling in the interval TSS \pm 3 kb), we counted mapped reads to the common peaks set using featureCounts⁷⁴, and counts were normalized as CPM (counts per million). Regions with low counts were removed with the edgeR⁷⁵ function filterByExpr with default parameters. Remaining regions were used as input for the multi-dimensional scaling (MDS) plot, produced with the plotMDS function implemented in edgeR.

For the analysis of ChIP-seq data from MDA-MB-231 cells treated with OICR-9429 with the *Drosophila* spike-in, the SpikeFlow pipeline was employed⁷⁶. Scaling factors were computed by dividing the ratio of reads mapped to the human genome in the IP and reads mapped to *Drosophila* in the IP by the ratio of reads mapped to human in the input and reads mapped to *Drosophila* in the input⁷⁷. The inverse of the scaling factors was supplied to DESeq2⁷⁸ as size factors for the identification of differentially bound regions. To this aim, featureCounts was used to count reads mapped to the list of peaks called from the untreated MDA-MB-231, and the obtained count matrix was supplied to DESeq2. Regions with an FDR < 0.05 were deemed as differentially bound regions.

Visualization of genomic tracks

Big Wig enrichment tracks were generated starting from BAM files with the bamCoverage function of deepTools⁷⁹ with the following parameters: –ignoreDuplicates –centerReads –bin-size 10 bp –normalizeUsing RPGC (reads per genomic content) –effectiveGenomeSize 2701495761. Density plots were generated starting from a data matrix obtained through computeMatrix with parameters –scale-regions -a 1000 -b 1000⁸⁰. Track displays were produced using pyGenomeTracks⁸⁰

RNA extraction, RT-qPCR analysis and RNA sequencing

Total RNA was extracted from $0.5\text{--}1 \times 10^6$ cells or 25–30 mg of frozen tissue samples (which were quickly homogenized by passing the suspension in RNA lysis buffer through a 18 gauge needle) using the Quick RNA miniprep kit (Zymo Research). 1 μ g of RNA was reverse-transcribed using the OneScript Plus cDNA synthesis kit (Applied Biological Materials Inc, Cat. No. G236). Quantitative real-time PCRs were carried out with the Fast SYBR Green PCR kit (Applied Biosystems, Thermo Fisher Scientific) in a Vii7 PCR machine (BioRad). Relative expression levels were calculated with the $2^{-[\Delta\Delta Ct]}$ method by normalization on the geomean of housekeeping genes (GAPDH and β ACTIN) or on the CT values of 18S ribosomal RNA and was expressed as a “fold change” relative to the control sample. Primers used for RT-qPCR are listed in Supplementary Table 2. Libraries for RNA

sequencing were prepared with the TruSeq RNA Sample Preparation v2 kit (RS-122-2002, Illumina), starting from 100 ng of total RNA for clinical samples. Sequencing was performed using a NovaSeq 6000 (Illumina) instrument with a read length of 150 bp PE and a sequencing depth of 80 million reads per sample for breast clinical samples.

RNA-seq analysis

For TN vs LuA comparison, the harmonized HTSeq count table was retrieved from TCGA on the Gencode 22 gene annotation. Short non-coding RNA and mitochondrial genes were removed from the count table prior to further processing. Samples were assigned to LuA and TN subtype according to Table S3 of ref. 2. Differentially expressed genes were identified using the R package DESeq2. ComBat-seq⁸¹ was employed to correct for batch effects retrieved from (<http://bioinformatics.mdanderson.org/tcgambatch/>). Genes were considered differentially expressed when presenting an absolute log₂ fold change equal to or greater than 1 and an adjusted p-value lower than 0.05. TPMs were calculated starting from FPKM values obtained through the DESeq2 with the option robust=TRUE. For the comparison of the four BC subtypes, data from the TCGA-BCRA project were retrieved with the help of the GDCquery function from the R package TCGAbiolinks⁸², with the following parameters: data.category="Transcriptome Profiling", data.type="Gene Expression Quantification", workflow.type="STAR - Counts", sample.type="Primary Tumor". The information stored in "paper_BRCA_Subtype_PAM50" was used to annotate patients to corresponding BC subtype. For internally acquired data, the BBduk program from the BBtools suite tools (sourceforge.net/projects/bbmap/) was used to perform adapter removal and read trimming with the following parameters: 'ktrim=r k=23 mink=11 hdist=1 tpe tbo'. STAR⁸³ was used to map and count reads against the hg38 assembly of the human genome, providing the RefSeqCurated annotation (release date: 2021-05-28) as GTF file. For TPM calculations, the top 50 genes with the highest difference between the sample with the maximum TPM value and the sample with the minimum TPM value were discarded, and TPM values were recomputed without considering these genes.

Over-representation and survival analyses

For the over-representation analysis the R package clusterProfiler⁸⁴ was used, using the curated gene sets (C2) from MSigDB⁸⁵ provided in the R package msigdb. The 2D annotation enrichment analysis was performed in Perseus v2.0.7.0. Survival analyses were carried out with the help of the R package survival and Kaplan-Meier plotted using the R package survminer. For each histone PTM, patients were stratified based on the median value, computed for each cohort separately. Significant difference in the survival of the two groups was assessed through the log-rank test. Time for overall survival was defined as the time elapsing between the date of surgery and the date of death or date of last follow-up. Time for disease-free survival was defined as the time elapsing between the date of surgery and the date of relapse, if the patient experienced the event, or the time of last follow-up.

Cell lines and cell culture

A list of the breast cancer cell lines used and their growth media is provided in Supplementary Table 3. The tFucci(CA)2/pCSII-EF vector was provided by the RIKEN BRC through the National BioResource Project of the MEXT, Japan (cat. RDB15446)²⁸ and used to generate a stable MDA-MB-231 FUCCI cell line. The media were supplemented with penicillin (100 U/ml) and streptomycin (100 mg/ml), and the cells were grown at 37 °C in a 5% CO₂ humidified atmosphere.

CRISPRi: generation of stable cell line expressing dCas9

Lentiviral transductions were performed in a Biosafety level 2 laboratory (BSL-2). Lentiviral constructs were transiently transfected in HEK-

293T cells by using the calcium phosphate transfection method together with the packaging plasmid pCMV-DR8.91 and the envelope plasmid pMD2G-VSVG. One 10 cm plate of HEK-293T was used to infect a well of a 6-well plate of MDA-MB-231. After 12 hours of transfection, the supernatant was removed, and 6 ml of fresh medium was added. After 48 and 72 hours from transfection, the supernatant containing the virus was harvested and used to infect MDA-MB-231 with the addition of Polybrene 1x. To generate a CRISPRi stable cell line, MDA-MB-231 cells were transduced with a lentivirus expressing Tet-On 3 G rtTA transactivator with Blasticidin resistance (pLVX-Tet3G Blasticidin, Addgene, #128061) and selected for 10 days with the addition of 10 µg/µl Blasticidin to the cell media. Next, the cells were transduced with dCas9-LSD1-mCherry (pHR_TRE3G-LSD1-dCas9-P2A-mCherry, Addgene #138462). The cells were grown in culture medium with the addition of tetracycline-free serum (Tetracline free Serum, Euroclone Spa, ECS0182L) and 1 µg/ml doxycycline (Thermo Fisher Scientific) was added following infection. Cells expressing mCherry were isolated by FACS sorting.

CRISPRi: sgRNA design and vector cloning

Target regions for sgRNAs were chosen by selecting two regions within the signal of H3K4me2 in the promoter or enhancer region (Supplementary Table 1). SgRNA sequences were designed with the CRISPOR Tool⁸⁶, taking into account the specificity and off-target scores. sgRNAs were cloned into the Addgene Vector pKLV2-U6gRNAs(BbsI)-PGKpuro2ABFP-W (Addgene, #67974), which carries the BFP marker. The list of oligos for sgRNA synthesis is provided in Supplementary Table 1. The negative control sgScramble was previously published⁸⁷. MDA-MB-231 dCas9-LSD1mCherry cells were plated at 0.5 x 10⁶ per well in 6 well plate. 24-48 hours after plating, each well was transduced with lentiviruses containing sgRNAs, which were packaged in HEK-293T cells as described above. 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 72 hours post-induction. The expression of the sgRNA and dCas9 system was assessed both by FACS analysis and by acquisition of images for the BFP and mCherry signal (Supplementary Figs. 13, 14).

Immunoblot analysis

MDA-MB-231 cells treated with methyltransferase inhibitors were lysed in nuclei isolation buffer as described above. 1 µg of protein was separated on a 4-12% PAA pre-cast gel, transferred onto a PVDF (Immobilon, Merck Life Science S.R.L.) membrane and probed by immunoblotting with primary antibodies (Anti-CRISPR-Cas9 antibody [7A9-3A3] (ab191468, Abcam, 1 µg/ml), H3K4me2 antibody (ab7766, Abcam, 1 µg/ml), anti-histone H3 Antibody (ab1791, Abcam, 1:5000)), followed by secondary peroxidase-conjugated antibodies (Thermo Scientific). Images were acquired on a ChemiDoc XRS instrument (Biorad) and quantified using the Image Lab software (Version 6.1).

Flow cytometry and cell sorting

Flow cytometry analyses were performed using a FACSCelesta and the FACSDiva Software v8.0L1, and were analyzed using the FlowJo 9.3 Software. For the analysis of CRISPRi cells, 0.5-1 x 10⁶ cells were induced with doxycycline for 72 hours, fixed with formaldehyde 1% in PBS and resuspended in 500 µl of PBS. Cell sorting was performed on sorterMelody and FACSARIA. Prior to sorting, cells were resuspended in sorting medium at 5 x 10⁶ cells/ml. Then the cells were filtered with a 100 µm cell strainer and dispensed in sterile FACS tubes. CRISPRi cells were collected in collection medium (Medium + 33% FBS + 3% P/S + 0.3% Gentamicin) and kept in culture. MDA-MB-231 FUCCI cells were sorted based on fluorescent markers (G1: mCherry-positive; G2/M: mCherry/mVenus-positive), collected in PBS, and pelleted prior to analysis.

Cell treatment and phenotypic assays

OICR-9429 was purchased from MedChemExpress and Target Mol. MI-136, MI-463, and MI-503 were purchased from MedChemExpress. All drugs were resuspended in DMSO for in vitro experiments. Doxorubicin was purchased from Pfizer. Cell viability was assessed with the Cell Titer Glo (Promega) assay following the manufacturer instructions and was performed in 96 well plate after seeding 2000 cells/well. Breast cancer cell lines were treated with either DMSO or OICR-9429 at 25–50 μ M, alone or in combination with different doses of Doxorubicin for 72 hours. The luminescence was measured with Glomax Explorer. IC₅₀ values were calculated on GraphPad Prism (9.3.1). The synergy between OICR-9429 and Doxorubicin was evaluated using Sinergy Finder⁸⁸ and Bliss score. Clonogenic assays were performed as described in⁸⁹. 4000 cells were seeded in p60, treated for 72 hours and grown for up to 8 days with the addition of fresh medium. Cells were fixed in methanol and stained with a crystal violet solution (V5265, Merck Life Science S.R.L.). Images were acquired with a Nikon Eclipse Ti2 at a 4x magnification and were analyzed using Fiji Image J (Version 2.14.0/1.54 f).

In vivo animal experiments

Mice were purchased from Charles River Italia and bred and housed under pathogen-free conditions in the animal facilities at the European Institute of Oncology–Italian Foundation for Cancer Research (FIRC) Institute of Molecular Oncology (IEO–IFOM, Milan, Italy) campus. Animals were housed at constant temperature (23 \pm 2 $^{\circ}$ C) and approximately 66% humidity in 12 hour light/dark cycles. Prior to injection, tumor cells MDA-MB-231 were trypsin detached, washed, and resuspended in medium to a final concentration of 500000 cells/10 μ L. The cell suspension was then mixed with 10- μ L growth factor-reduced Matrigel (Corning 356231) and maintained on ice until injection. Female NOD SCID IL2RGnull mice (NSG from Charles River, Italy, 8 weeks old) were anesthetized with isoflurane and injected with 20 μ L of cell suspension in Matrigel directly in the right fourth mammary fad pad with a Hamilton syringe. Tumor growth was monitored twice a week, by bi-dimensional measurements using a caliper, and the tumor volume was calculated according to the formula: $L \times W^2/2 = \text{mm}^3$, where “w” and “W” are “minor side” and “major side” (in mm), respectively. When the major side of the tumor reached 15 mm, the mice were sacrificed. After the tumors appeared as established palpable masses (~20 days after cell injection), the mice were randomized in different groups. OICR-9429, at 30 mg/kg dissolved in 10% DMSO + 40% PEG300 + 5% Tween 80 + 45% physiological solution, was administered three times a week via intraperitoneal injection, for 3 weeks

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The MS proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE⁹⁰ partner repository with the dataset identifiers [PXD057150](https://doi.org/10.26434/chemrxiv-2025-pxd05) and [PXD064421](https://doi.org/10.26434/chemrxiv-2025-pxd06). The ChIP-seq data relative to MDA-MB-231 treated with OICR-9429 have been deposited in NCBI's Gene Expression Omnibus⁹¹ and are accessible through GEO Series accession number [GSE283103](https://doi.org/10.5555/283103). The patient ChIP-seq data are not publicly available due to privacy concerns and restrictions related to the protection of participant confidentiality. The data will be shared on a reasonable request to the corresponding author. Source data are provided with this paper.

Code availability

The code used in this study has been deposited on GitHub (<https://github.com/alessandro-vai/BC-H3K4me2>)⁹².

References

1. Lehmann, B. D. et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J. Clin. Invest.* **121**, 2750–2767 (2011).
2. Lehmann, B. D. et al. Refinement of triple-negative breast cancer molecular subtypes: implications for neoadjuvant chemotherapy selection. *PLoS One* **11**, e0157368 (2016).
3. Rheinbay, E., Louis, D. N., Bernstein, B. E. & Suva, M. L. A tell-tail sign of chromatin: histone mutations drive pediatric glioblastoma. *Cancer Cell* **21**, 329–331 (2012).
4. Vogelstein, B. et al. Cancer genome landscapes. *Science* **339**, 1546–1558 (2013).
5. Nguyen, V. T. et al. Differential epigenetic reprogramming in response to specific endocrine therapies promotes cholesterol biosynthesis and cellular invasion. *Nat. Commun.* **6**, 10044 (2015).
6. Suva, M. L., Riggi, N. & Bernstein, B. E. Epigenetic reprogramming in cancer. *Science* **339**, 1567–1570 (2013).
7. Fraga, M. F. et al. Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. *Nat. Genet.* **37**, 391–400 (2005).
8. Seligson, D. B. et al. Global levels of histone modifications predict prognosis in different cancers. *Am. J. Pathol.* **174**, 1619–1628 (2009).
9. Seligson, D. B. et al. Global histone modification patterns predict risk of prostate cancer recurrence. *Nature* **435**, 1262–1266 (2005).
10. Khan, S. A., Reddy, D. & Gupta, S. Global histone post-translational modifications and cancer: Biomarkers for diagnosis, prognosis and treatment? *World J. Biol. Chem.* **6**, 333–345 (2015).
11. Elsheikh, S. E. et al. Global histone modifications in breast cancer correlate with tumor phenotypes, prognostic factors, and patient outcome. *Cancer Res.* **69**, 3802 (2009).
12. Soldi, M., Bremang, M. & Bonaldi, T. Biochemical systems approaches for the analysis of histone modification readout. *Biochim Biophys. Acta* **1839**, 657–668 (2014).
13. Noberini, R. et al. PAT-H-MS coupled with laser microdissection to study histone post-translational modifications in selected cell populations from pathology samples. *Clin. Epigenet.* **9**, 69 (2017).
14. Noberini, R. et al. Profiling of epigenetic features in clinical samples reveals novel widespread changes in cancer. *Cancers.* **11**, 723 (2019).
15. 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. Proteom.* **15**, 866–877 (2016).
16. Restellini, C. et al. Alternative digestion approaches improve histone modification mapping by mass spectrometry in clinical samples. *Proteom. Clin. Appl.* **13**, e1700166 (2019).
17. Chi, P., Allis, C. D. & Wang, G. G. Covalent histone modifications—miswritten, misinterpreted and mis-erased in human cancers. *Nat. Rev. Cancer* **10**, 457–469 (2010).
18. West, A. C. & Johnstone, R. W. New and emerging HDAC inhibitors for cancer treatment. *J. Clin. Invest.* **124**, 30–39 (2014).
19. Nebbioso, A., Carafa, V., Benedetti, R. & Altucci, L. Trials with ‘epigenetic’ drugs: an update. *Mol. Oncol.* **6**, 657–682 (2012).
20. Noberini, R. et al. Extensive and systematic rewiring of histone post-translational modifications in cancer model systems. *Nucleic Acids Res.* **46**, 3817–3832 (2018).
21. Noberini, R., Restellini, C., Savoia, E. O. & Bonaldi, T. Enrichment of histones from patient samples for mass spectrometry-based analysis of post-translational modifications. *Methods* **184**, 19–28 (2019).
22. Noberini, R. & Bonaldi, T. A Super-SILAC strategy for the accurate and multiplexed profiling of histone posttranslational modifications. *Methods Enzymol.* **586**, 311–332 (2017).

23. Noberini, R., Longhi, E. & Bonaldi, T. A Super-SILAC approach for profiling histone posttranslational modifications. *Methods Mol. Biol.* **2603**, 87–102 (2023).
24. Asleh, K., Riaz, N. & Nielsen, T. O. Heterogeneity of triple negative breast cancer: Current advances in subtyping and treatment implications. *J. Exp. Clin. Cancer Res* **41**, 265 (2022).
25. Noberini, R. et al. Spatial epi-proteomics enabled by histone post-translational modification analysis from low-abundance clinical samples. *Clin. Epigenet.* **13**, 145 (2021).
26. Bonenfant, D. et al. Analysis of dynamic changes in post-translational modifications of human histones during cell cycle by mass spectrometry. *Mol. Cell Proteom.* **6**, 1917–1932 (2007).
27. 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. *Mol. Biol. Cell* **28**, 2290–2302 (2017).
28. Sakaue-Sawano, A. et al. Genetically encoded tools for optical dissection of the mammalian cell cycle. *Mol. Cell* **68**, 626–640 e5 (2017).
29. Black, J. C., Van Rechem, C. & Whetstine, J. R. Histone lysine methylation dynamics: establishment, regulation, and biological impact. *Mol. Cell* **48**, 491–507 (2012).
30. 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* **450**, 309–313 (2007).
31. Corces, M. R. et al. The chromatin accessibility landscape of primary human cancers. *Science* **362**, eaav1898 (2018).
32. Yoshino, S. & Suzuki, H. I. The molecular understanding of super-enhancer dysregulation in cancer. *Nagoya J. Med. Sci.* **84**, 216–229 (2022).
33. Mertins, P. et al. Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature* **534**, 55–62 (2016).
34. Cox, J. & Mann, M. 1D and 2D annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data. *BMC Bioinforma.* **13**, S12 (2012).
35. Gilbert, L. A. et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442–451 (2013).
36. Li, K. et al. Interrogation of enhancer function by enhancer-targeting CRISPR epigenetic editing. *Nat. Commun.* **11**, 485 (2020).
37. Vecchi, L. et al. Phospholipase A(2) Drives Tumorigenesis and cancer aggressiveness through its interaction with Annexin A1. *Cells* **10**, 1472 (2021).
38. Zhang, X. et al. IGF2BP3 mediates the mRNA degradation of NF1 to promote triple-negative breast cancer progression via an m6A-dependent manner. *Clin. Transl. Med.* **13**, e1427 (2023).
39. Zhu, H. et al. The complex role of AIM2 in autoimmune diseases and cancers. *Immun. Inflamm. Dis.* **9**, 649–665 (2021).
40. Huang, Q. F., Fang, D. L., Nong, B. B. & Zeng, J. Focal pyroptosis-related genes AIM2 and ZBP1 are prognostic markers for triple-negative breast cancer with brain metastases. *Transl. Cancer Res.* **10**, 4845–4858 (2021).
41. Vidula, N., Yau, C. & Rugo, H. S. Glutaminase (GLS1) gene expression in primary breast cancer. *Breast Cancer* **30**, 1079–1084 (2023).
42. Huang, S. et al. GLIPR-2 overexpression in HK-2 cells promotes cell EMT and migration through ERK1/2 activation. *PLoS One* **8**, e58574 (2013).
43. Jackson, J. T., Mulazzani, E., Nutt, S. L. & Masters, S. L. The role of PLCgamma2 in immunological disorders, cancer, and neurodegeneration. *J. Biol. Chem.* **297**, 100905 (2021).
44. Yang, L., Jin, M. & Jeong, K. W. Histone H3K4 Methyltransferases as targets for drug-resistant cancers. *Biology* **10**, (2021).
45. Punzi, S. et al. WDR5 inhibition halts metastasis dissemination by repressing the mesenchymal phenotype of breast cancer cells. *Breast Cancer Res.* **21**, 123 (2019).
46. Borkin, D. et al. Pharmacologic inhibition of the Menin-MLL interaction blocks progression of MLL leukemia in vivo. *Cancer Cell* **27**, 589–602 (2015).
47. Malik, R. et al. Targeting the MLL complex in castration-resistant prostate cancer. *Nat. Med.* **21**, 344–352 (2015).
48. Azimi, I. et al. Activation of the ion channel TRPV4 induces epithelial to mesenchymal transition in breast cancer cells. *Int. J. Mol. Sci.* **21**, 9417 (2020).
49. Calaf, G. M., Crispin, L. A., Munoz, J. P., Aguayo, F. & Bleak, T. C. Muscarinic receptors associated with cancer. *Cancers* **14**, 2322 (2022).
50. Sizemore, G. M., Sizemore, S. T., Seachrist, D. D. & Keri, R. A. GABA(A) receptor pi (GABRP) stimulates basal-like breast cancer cell migration through activation of extracellular-regulated kinase 1/2 (ERK1/2). *J. Biol. Chem.* **289**, 24102–24113 (2014).
51. Xie, J. et al. TMEM71 is crucial for cell proliferation in lower-grade glioma and is linked to unfavorable prognosis. *Cancer Cell Int.* **25**, 109 (2025).
52. Schotta, G. et al. A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes Dev.* **18**, 1251–1262 (2004).
53. Yokoyama, Y. et al. Loss of histone H4K20 trimethylation predicts poor prognosis in breast cancer and is associated with invasive activity. *Breast Cancer Res.* **16**, R66 (2014).
54. Wei, Y. et al. Loss of trimethylation at lysine 27 of histone H3 is a predictor of poor outcome in breast, ovarian, and pancreatic cancers. *Mol. Carcinog.* **47**, 701–706 (2008).
55. Marsolier, J. et al. H3K27me3 conditions chemotolerance in triple-negative breast cancer. *Nat. Genet.* **54**, 459–468 (2022).
56. Michealraj, K. A. et al. Metabolic Regulation of the Epigenome Drives Lethal Infantile Ependymoma. *Cell* **181**, 1329–1345 e24 (2020).
57. Bachman, K. E. et al. Histone modifications and silencing prior to DNA methylation of a tumor suppressor gene. *Cancer Cell* **3**, 89–95 (2003).
58. Espada, J. et al. Human DNA methyltransferase 1 is required for maintenance of the histone H3 modification pattern. *J. Biol. Chem.* **279**, 37175–37184 (2004).
59. Tamaru, H. & Selker, E. U. A histone H3 methyltransferase controls DNA methylation in *Neurospora crassa*. *Nature* **414**, 277–283 (2001).
60. Di Nisio, E. et al. 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.* **30**, 822–832 (2023).
61. Huang, H. et al. Defining super-enhancer landscape in triple-negative breast cancer by multiomic profiling. *Nat. Commun.* **12**, 2242 (2021).
62. Freire, P. R. et al. Identification of a Novel Chromatin structure associated with the transcriptional response to menin inhibitors in AML. *Blood* **144**, 952 (2024).
63. Noberini, R. et al. Label-Free Mass Spectrometry-based quantification of Linker Histone H1 variants in clinical samples. *Int. J. Mol. Sci.* **21**, 7330 (2020).
64. Yuan, Z. F. et al. EpiProfile 2.0: A computational platform for processing epi-proteomics Mass Spectrometry Data. *J. Proteome Res.* **17**, 2533–2541 (2018).
65. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
66. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
67. Zhu, Y. et al. DEqMS: A method for accurate variance estimation in differential protein expression analysis. *Mol. Cell Proteom.* **19**, 1047–1057 (2020).

68. Greulich, F., Mechtidou, A., Horn, T. & Uhlenhaut, N. H. Protocol for using heterologous spike-ins to normalize for technical variation in chromatin immunoprecipitation. *STAR Protoc.* **2**, 100609 (2021).
69. Blecher-Gonen, R. et al. High-throughput chromatin immunoprecipitation for genome-wide mapping of in vivo protein-DNA interactions and epigenomic states. *Nat. Protoc.* **8**, 539–554 (2013).
70. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
71. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
72. Neph, S. et al. BEDOPS: high-performance genomic feature operations. *Bioinformatics* **28**, 1919–1920 (2012).
73. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
74. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
75. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
76. Bressan, D., Fernandez-Perez, D., Romanel, A. & Chiacchiera, F. SpikeFlow: automated and flexible analysis of ChIP-Seq data with spike-in control. *NAR Genom. Bioinform.* **6**, lqae118 (2024).
77. Krug, B. et al. Pervasive H3K27 acetylation leads to ERV Expression and a therapeutic vulnerability in H3K27M Gliomas. *Cancer Cell* **35**, 782–797 e8 (2019).
78. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
79. Ramirez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
80. Lopez-Delisle, L. et al. pyGenomeTracks: reproducible plots for multivariate genomic datasets. *Bioinformatics* **37**, 422–423 (2021).
81. Zhang, Y., Parmigiani, G. & Johnson, W. E. ComBat-seq: batch effect adjustment for RNA-seq count data. *NAR Genom. Bioinform.* **2**, lqaa078 (2020).
82. Colaprico, A. et al. TCGAAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* **44**, e71 (2016).
83. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
84. Wu, T. et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation.* **2**, 100141 (2021).
85. Liberzon, A. et al. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
86. Concordet, J.-P. & Haeussler, M. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res.* **46**, W242–W245 (2018).
87. Kelly, M. R. et al. A multi-omic dissection of super-enhancer driven oncogenic gene expression programs in ovarian cancer. *Nat. Commun.* **13**, 4247 (2022).
88. Zheng, S. et al. SynergyFinder plus: toward better interpretation and annotation of drug combination screening datasets. *Genomics Proteom. Bioinforma.* **20**, 587–596 (2022).
89. Franken, N. A., Rodermond, H. M., Stap, J., Haveman, J. & van Bree, C. Clonogenic assay of cells in vitro. *Nat. Protoc.* **1**, 2315–2319 (2006).
90. Perez-Riverol, Y. et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **47**, D442–D450 (2019).
91. Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207–210 (2002).
92. Vai, A. Code used for the data analysis of the manuscript: “A Histone-Centric Multi-Omics Study Shows that Increased H3K4 Methylation Sustains Triple Negative Breast Cancer Phenotypes. (Zenodo, 2025).

Acknowledgements

Italian Ministry of Health, grant number GR-2016-02361522 (RN). -Italian Association for Cancer Research (AIRC), grant numbers IG-2018-21834 (TB) and IG-2023-28767 (TB). Fondazione IEO-Monzino (FIEORDT-2023-BONALDI) (TB). EPIC-XS, project number 823839, funded by the Horizon 2020 program of the European Union. Alessandro Vai is a PhD student within the European School of Molecular Medicine (SEMM) and is supported by an AIRC fellowship for Italy. We thank Camilla Restellini and Alberto Dalmaso for technical help and Paola Scaffidi for insightful comments on the manuscript. The results shown here are in part based upon data generated by the TCGA Research Network: <https://www.cancer.gov/tcga>.

Author contributions

Conceptualization: T.B., R.N., S.M. Methodology: R.N., E.O.S., G.R. Formal Analysis: A.V. Investigation: G.R., R.N., A.V., E.O.S., I.P., M.G.J., G.Be, C.A.S., B.C., G.Bo, M.C., F.Z., S.P. Visualization: A.V., R.N. Funding acquisition: R.N., T.B. Project administration: T.B. Supervision: T.B., R.N., N.F., S.M., G.P. Writing – original draft: R.N., G.R. Writing – review & editing: T.B, A.V.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-63745-z>.

Correspondence and requests for materials should be addressed to Roberta Noberini or Tiziana Bonaldi.

Peer review information *Nature Communications* thanks Simone Sidoli and the other anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025