ALDH1 subtype-specific inhibitor targets key cancerous epithelial cell populations in aggressive subtypes of breast cancer

Raquel Pequerul[1,2], Andrada Constantinescu[2]*, Bassam Janji[3], Akinchan Kumar[3], Xavier Parés[1], Oscar Palacios[4], Delphine Colignon[5], Axelle Berrou[6], Guy Fournet[5], Paul Berchard[2], Guillaume Martin[2], Ismail Ceylan[2], Rocio Rebollido-Ríos[7]*, Jaume Farrés[1]*, Mileidys Perez-Alea[2]**

AFFILIATIONS

[1]Department of Biochemistry and Molecular Biology, Faculty of Biosciences, Universitat Autònoma de Barcelona. Cerdanyola del Vallès (Barcelona), Spain

[2]Advanced BioDesign, Saint Priest, France

[3]Tumor Immunotherapy and Microenvironment (TIME) Group, Department of Cancer Research, Luxembourg Institute of Health (LIH), Luxembourg City, Luxembourg

[4]Department of Chemistry, Faculty of Sciences, Universitat Autònoma de Barcelona. Cerdanyola del Vallès (Barcelona), Spain

[5]Institut de Chimie et Biochimie Moléculaires et Supramoléculaires, CNRS, Université de Lyon 1

[6]CEISAM laboratory, UMR CNRS 6230, Nantes University, 2 chemin de la Houssinière, 44322 Nantes

[7]University of Cologne, Faculty of Medicine and University Hospital Cologne, Department I of Internal Medicine, Center for Integrated Oncology Aachen Bonn Cologne Duesseldorf, Center for Molecular Medicine Cologne, CECAD Center of Excellence on Cellular Stress Responses in Aging-Associated Diseases, Cologne 50931, Germany

CORRESPONDING AUTHOR(S) E-MAIL ADDRESS(ES)

mileidys.perez@a-biodesign.com, jaume.farres@uab.cat, rocio.rebollido-rios@uk-koeln.de

ABSTRACT

Aggressive breast cancer subtypes, particularly triple-negative breast cancer (TNBC), lack effective targeted therapies, requiring novel approaches. This study focuses on the aldehyde dehydrogenase 1A (ALDH1A) subfamily, comprising ALDH1A1, ALDH1A2, and ALDH1A3, and their roles in tumor biology and the tumor microenvironment. Comprehensive transcriptomic and single-cell analyses revealed subtype- and cell-specific expression patterns of ALDH1A isoforms, with *ALDH1A3* predominantly expressed in epithelial cancer cells of basal-like tumors, while *ALDH1A1* and *ALDH1A2* were expressed in stromal and immune-associated subpopulations. Guided by these findings, we developed ABD0171, a selective ALDH1A3 inhibitor that demonstrated potent isoform-specific activity. ABD0171 effectively disrupted key pathways in TNBC cells *in vitro*, including IL6/JAK/STAT3, tPA and Src/FAK, and exhibited superior selectivity, robust antitumor and antimetastatic effects, and a favorable safety profile *in vivo*. These results establish ALDH1A3 as a promising therapeutic target and validate ABD0171 as a candidate to address current challenges in aggressive breast cancers.

INTRODUCTION

Breast cancer is the most prevalent malignancy among women worldwide, with metastatic progression remaining the leading cause of mortality[1]. This disease exhibits extensive clinical and molecular heterogeneity, which complicates treatment strategies. Over recent years, the classification of breast cancer has been expanded beyond the four major subtypes defined by the presence or absence of estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2)[2–5]. Triple-negative breast cancer (TNBC), defined by the lack of expression of ER, PR and HER2, represents a particularly challenging subset[6]. Advanced molecular profiling has further subdivided breast cancers into intrinsic subtypes, including Luminal A, Luminal B, HER2-enriched, Basal-like, and normal-like entities, providing greater insight into tumor biology and therapeutic opportunities[4,7,8]. Notably, TNBCs, which frequently align with the Basal-like subtype, account for ~15% of invasive cases. These tumors are characterized by high aggressiveness, a lack of targeted therapies, and poor clinical outcomes, making them a critical focus for research and therapeutic development[6,9].

The aldehyde dehydrogenase 1A (ALDH1A) subfamily of enzymes -comprising ALDH1A1, ALDH1A2, and ALDH1A3- has gained significant attention as a key player in breast cancer biology, particularly in TNBC. Historically, ALDH1 activity has served as a biomarker of cancer stem cells, and has been functionally implicated in tumor initiation, progression, and therapy resistance, primarily through its roles in oxidative stress mitigation and detoxification of reactive aldehydes derived from aberrant metabolism or cytotoxic agents. Recent findings, however, underscore isoform-specific contributions: ALDH1A1 expression is linked to chemoresistant breast cancer stem cells[10–12], ALDH1A2 is proposed as a diagnostic biomarker for early breast cancer[13], and ALDH1A3 is strongly associated with proliferation and poor prognosis in triple negative breast cancer (TNBC)[11,14–18]; likely due to their role in retinoic acid (RA) synthesis—a crucial regulator of differentiation and proliferation[19]. Despite mounting evidence supporting ALDH1A isoforms as important biomarkers and therapeutic targets in breast cancer, the precise contribution of each isoform to tumor pathogenesis, including their distribution within the tumor microenvironment, and their expression patterns across breast cancer molecular subtypes remain poorly understood. This gap in knowledge has hindered the development of isoform-selective therapies to effectively target hard-to-treat breast cancer subtypes like TNBC.

In this study, we investigated the distribution and abundance of *ALDH1A1*, *ALDH1A2*, and *ALDH1A3* across breast cancer molecular subtypes and at single-cell resolution within the tumor microenvironment. Guided by these insights, we developed a series of ALDH1 inhibitors based on the structure of DIMATE (4-dimethylamino-4-methyl-pent-2-ynthioic acid-S-methyl ester), an ALDH inhibitor currently in Phase I clinical trial ("ODYSSEY", NCT05601726) for acute myeloid leukemia, with demonstrated preclinical efficacy in leukemia, melanoma, and non-small cell lung cancer. The lead compound, ABD0171, demonstrated potent isoform-specific activity, disrupted key pathways in TNBC *cells in vitro*, and exhibited superior selectivity, robust antitumor and antimetastatic effects, and a favorable safety profile *in vivo*.

RESULTS

ALDH1A isoforms exhibit specific expression patterns across molecular subtypes and cell populations in breast tumors

We preprocessed and integrated three independent RNA-seq datasets from The Cancer Genome Atlas (TCGA)[20,21] to generate a comprehensive bulk RNA-seq dataset comprising 1103 primary human breast invasive carcinoma samples (Fig.1a). Unsupervised clustering of the integrated dataset, incorporating 59 genes- including the

PAM50 signature,[22,23] claudin-low associated-genes[24] and *ALDH1A1*, *ALDH1A2*, and *ALDH1A3* classified the tumors into four intrinsic molecular subtypes according to the PAM50 classifier: Basal-like (17.8%), HER2-enriched (36.4%), Luminal A (25.4%) and Luminal B (18.5%). Additionally, a small subset (1.9 %) exhibited claudin-low molecular features, consistent with the reported prevalence of claudin-low tumors, which account for 1.5% to 14% of all breast tumors (Fig. 1b, Supplementary Data 1A).

We also retrieved RNA-seq data for 50 human breast cancer cell lines from the Dependency Map (DepMap)[25] portal and performed the subsequent preprocessing steps. Unsupervised clustering of these cell lines revealed four distinct molecular subgroups, aligning with established classifications of breast cell lines [26]: Basal A (32%), Basal B (18%), HER2-enriched (26%), and Luminal (24%) (Fig. 1a, c, Supplementary Data 1B). We used these classifications to analyze the expression of *ALDH1A1*, *ALDH1A2* and *ALDH1A3* across breast cancer subtypes, aiming to clarify their expression profile and assess how well cell lines reflect the primary tumors. Among the cell lines, *ALDH1A3* was the most predominantly expressed isoform, with significantly higher expression in Basal A compared to Basal B, HER2-enriched and luminal subtypes (adjusted *p*-value < 0.001, Fig.1d and Supplementary Data 1C). This finding was consistent with ALDH1A3 protein expression profiles (Supplementary Fig. 1a) and supported by previous studies that associate high ALDH1A3 expression with TNBC[16,17], an entity identified by surrogate immunohistochemistry as negative for hormone receptors and HER2, reported to overlap significantly (75%) with the Basal-like subtype[27]. Across all breast cancer cell lines, *ALDH1A1* and *ALDH1A2* were expressed at low levels, regardless of molecular subtype. Basal A cell lines have been reported to closely resemble the Basal-like tumors in clinical samples, sharing key characteristics such as epithelial properties and highly proliferative capacity[28]. We observed a strong positive correlation ($\rho = 0.71$) between Basal A subtype in cell lines and the Basal-like subtype in patient samples, further confirming their shared molecular features (Fig.1e). Notably, this similarity extends to the significant expression of *ALDH1A3* in both Basal A cell lines and Basal-like breast cancer patients (adjusted *p*-value < 0.001, Fig. 1d and Supplementary Data 1D and 1E). *ALDH1A2* expression remained low across all intrinsic molecular subtypes in patient samples, reflecting a trend like that observed in the cell line data. In contrast, *ALDH1A1* exhibited higher expression in all patient subtypes, suggesting that its expression may originate from diverse cell types within the tumor microenvironment, which are not fully represented in traditional cell line models (Fig. 1d).

Interestingly, the specific subset of patients with claudin-low tumors exhibited significant expressions of both *ALDH1A1* and *ALDH1A2* compared to the other tumor types (all

adjusted *p*-values < 0.001, Fig. 1d and Supplementary Data 1E). Although claudin-low tumors are challenging to model *in vitro* due to their characteristic stromal compartment, three cell lines (HCC1143, MDAMB436 and DU4475) that had been previously identified as molecularly and phenotypically similar to claudin-low tumors[5] also displayed high ALDH1A2 protein levels and double positivity for ALDH1A1 protein (Supplementary Fig. 1a). Together, these observations suggest that claudin-low breast cancers possess a distinct ALDH1A expression profile.

To dissect the individual contributions of different cell types to the overall *ALDH1A* gene expression in both cell lines and tumor populations, we analyzed single-cell RNA sequencing (scRNA-seq) data collected from two publicly independent studies targeting breast cancer cell lines and patients, respectively. One study was based on 32 breast cancer cell lines[29] while another based on primary tumor biopsies from 26 patients with invasive breast carcinoma including a total of 100,064 cells[30]. The earlier study, which included the breast cancer cell lines, revealed that basal phenotype cell lines form a distinct cluster, separate from the luminal and HER2-enriched groups[29]. Focusing only on the 16 basal cell lines including a total of 20,485 cells (Fig. 2a and Supplementary Data 1F), we identified a distinct cluster of Basal B cells, with two Basal A cell lines (HCC1143 and JIMT1) mixed within a broader Basal A population (Fig. 2b). Differential gene expression analysis between these basal phenotypes confirmed that *ALDH1A3* is highly expressed in the Basal A subtype (Fig. 2b), consistent with the findings shown in Fig. 1d. The expression levels of *ALDH1A1* and *ALDH1A2* remain very low or nearly absent in both Basal A and Basal B cell lines subtypes, also aligning with Fig. 1d.

Next, we explored the heterogeneity of the ALDH1A subfamily across breast cancer subtypes by analyzing the patient samples[30]. Using canonical cell markers, we identified 10 major cell types along with a small population of myoepithelial cells (Fig. 2c). The major cell types were annotated as epithelial, stromal (including endothelial, mesenchymal stem cells (MSCs), cancer-associated fibroblasts (CAFs), myofibroblastic CAFs-like (my-CAFs-like) and inflammatory CAF-like (iCAFs-like) and immune cells (T-cells, myeloid cells, B cells and plasmablasts). These cell populations were found across all five intrinsic molecular subtypes, except for CAFs and myoepithelial cells, which were exclusively detected in the Basal-like subtype (Fig. 2c and Supplementary Data 1G).

Our analyses revealed unique expression patterns of the *ALDH1A* isoforms within specific cell populations (Fig. 2d, e). Among the three *ALDH1A* isoforms, only *ALDH1A1* was differentially expressed in the myCAFs-like population. *ALDH1A1* was expressed in 2,853 cells, *ALDH1A3* in 1,956 cells, and *ALDH1A2* in only 502 cells. Comparative

analysis of isoform expression patterns across cell populations and molecular subtypes is visualized in two sets of pie charts in Fig. 2e. The data reveal unique expression profiles for *ALDH1A1, ALDH1A2* and *ALDH1A3* across different cell types within breast tumors. Notably, *ALDH1A1* and *ALDH1A3* are uniquely expressed in myeloid cells (33% and 1%, respectively) and epithelial cells (4% and 53%, respectively). *ALDH1A1* and *ALDH1A2* display similar positivity within the myCAFs-like population (50% and 46%, respectively) while *ALDH1A*3 is 27%. Approximately, 70-80% of the cells expressing *ALDH1A1, ALDH1A2*, or *ALDH1A3* originate from Basal-like and Luminal B subtypes. Interestingly, the HER2-enriched subtype contributes to 19% of the ALDH1A3-expressing cells, a higher proportion compared to its representation in ALDH1A1 and ALDH1A2 positive expression profiles.

Overall, these insights emphasize the complex, cell population-specific expression patterns of ALDH1A isoforms in breast tumors, underscoring their unique roles in different breast cancer subtypes and their potential impact on targeted therapies.

Novel compounds with selective inhibitory activity towards ALDH1A3 show enhanced cytotoxicity in basal phenotype breast cancer

Building upon our characterization of the ALDH1A landscape in breast tumors, we next sought to explore the therapeutic potential of targeting these enzymes. To this end, we designed and synthesized a series of novel compounds, categorized into different families by their chemical substituents (Fig. 3a). DIMATE and MATE represent prior α,β-acetylenic aminothiolester compounds featuring a pentynethioic acid backbone[31]. In previous efforts to generate potent ALDH inhibitors, we explored pharmacological modulation of DIMATE at the N atom with alkyl, aryl or heterocyclic substituents (series I and II) and/or the substitution of the sulphur atom of the thiolester functional group with oxygen (series III). However, these approaches did not yield increased potency or selectivity for this family of compounds[31]. Here, we performed novel targeted structural modifications mainly at the nitrogen atom of the original aminothiolester core structure. The new compounds were designed with a -N-C-C-O- pattern which corresponds to the morpholine ring opening in MATE (IV, V, VI, and VII series) (Fig. 3a)[32,33]. We next evaluated their effects on the enzymatic activity of recombinant human ALDH1A1, ALDH1A2, and ALDH1A3, as well as ALDH2 (the isoform most closely related to ALDH1A and known for its cardioprotective effects in the heart) and the critical corneal ALDH3A1, through a single-point enzymatic inhibition analysis (Fig. 3b and Supplementary Data 2A). The most effective compounds against ALDH1A1, ALDH1A2

and ALDH1A3 activity belonged to series V of analogs, characterized by a (arylmethoxy)ethylamino substituents compared to series IV (alkylmethoxy)ethylamino substituents (e.g. ABD0118). Notably, three compounds within the same family V, containing an (heteroarylmethoxy)ethylamino group substitution instead (i.e. ABD0257, ABD0258 and ABD0260), were as ineffective as, for example, ABD0118 in inhibiting ALDH1A activity. This suggests that the aryl moiety is a crucial modification for achieving enhanced and selective inhibitory activity against all three human ALDH1A isoforms, while preserving the functions of ALDH2 and ALDH3A1.

To investigate the anticancer potential of synthetized compounds, we conducted cytotoxic drug-response assays using a panel of 18 human breast cancer cell lines representing the four major molecular subtypes (Supplementary Data 3A). We reported the results as half-growth inhibition (GI_{50}) values calculated from the corresponding dose-response curves (Supplementary Fig. 1b, Supplementary Data 3B) and represented them as a heatmap to highlight compounds with enhanced activity against different subtypes alongside their ALDH inhibitory profiles (Fig. 3c). Compounds in series V, characterized by aryl substitutions- except for compound ABD0369, which features a bicyclic aromatic hydrocarbon (i.e. naphthalene)- demonstrated the most potent cytotoxic effects. ABD0171 exhibited exceptional potency, with GI_{50} values consistently below 10 μM in 95% of tested cell lines (Supplementary Data 3B).

Additionally, compounds ABD0080, featuring a benzyl substitution on nitrogen, and ABD0079, forming a heterocycle with nitrogen, also showed notable efficacy. Importantly, these two compounds, along with most aryl-substituted compounds in series V (except ABD0405 and ABD0158) exhibited selective cytotoxic activity toward Basal A and Basal B cell lines, with ABD0171 being the most effective.

In contrast, compounds from series I, II, III and IV, particularly those with minimal inhibitory effects on ALDH1A1 and ALDH1A3, showed little to no cytotoxic activity against the majority of basal phenotype cell lines. These compounds were notably ineffective against BT20, HCC1143, HCC1937, and HCC1954 Basal A cell lines, as well as the HER2-enriched HCC1419 cells, which express high ALDH1A3 protein levels. This suggests a specific association between the compounds' efficacy and the ALDH1A3 inhibition. Intriguingly, HER2-enriched cells, which co-expressed ALDH1A1 and ALDH1A3 in 4 out of 5 cell lines, showed high sensitivity to DIMATE and compounds within family IV, despite their weaker ALDH1A inhibitory effects. This finding raises the possibility that the single-point enzymatic analysis may not fully capture their inhibitory effects on ALDH1A1 and/or ALDH1A3, or that their cytotoxicity may result from inhibition

of another ALDH isoform or an off-target mechanism critical for HER2-enriched cell survival.

Next, as shown in Fig. 3d, we evaluated the impact of ABD0171 and DIMATE on the self-renewal and sustain proliferation of breast cancer patient-derived xenograft (PDX) cells. These PDX models were characterized by immunochemistry as HER2-enriched subtype (n=4) and triple-negative breast cancer (TNBC, n=7). Both compounds displayed similar inhibitory effects on long-term proliferation in colony formation assays across both PDX subtypes, suggesting minimal differences in their tumorigenic potential in this context (Supplementary Data 3C).

To further investigate the relationship between predicted and observed drug response in both PDX models and cancer cell lines, we employed a linear mixed-effects model (LMM). This analysis incorporated GI_{50} values from TNBC and HER2-enriched PDX as well as cell lines. The average GI_{50} for ABD0171 in Basal A cell lines was used as the baseline for the LMM analysis, given their demonstrated sensitivity to the compound. As shown in Fig. 3e, the LMM revealed a robust relationship between predicted and observed GI_{50} values for both TNBC and HER2-enriched models treated with ABD0171 and DIMATE. In TNBC models, ABD0171 showed a strong predictive performance, as evidenced by the close alignment of observed and predicted GI_{50} values for most cell lines and PDX models (Fig. 3e, Supplementary Data 3D). In contrast, DIMATE-treated TNBC models showed greater variability, indicating that while potent, its effects might be modulated by additional factors not fully captured by the model. In HER2-enriched models, both compounds performed similarly. DIMATE exhibited the strongest agreement between predicted and observed GI_{50} values, whereas ABD0171, though effective, showed a few outliers in PDX models, suggesting the presence resistant mechanisms or differential drug activity in HER-2-enriched cells that warrants further mechanistic investigation.

ABD0171 is a potent irreversible inhibitor and specific affinity-label against ALDH1A3

Given the strong ALDH1A inhibitory effect and potent cytotoxicity of ABD0171 against Basal-like cancer cells, we conducted a detailed enzymatic and biochemical characterization of ABD0171 alongside DIMATE. We determined the IC_{50} values and the type of inhibition for both compounds against ALDH1A1, ALDH1A2, ALDH1A3, ALDH2 and ALDH3A1 isoforms, using both routine (hexanal) and physiological (all-*trans*-retinaldehyde, atRAL) substrates and for ALDH3A1 its preferred 4-nitrobenzaldehyde substrate (Table 1 and Supplementary Table S1). These assays revealed that inhibition

was contingent on the order of reagent addition and incubation times. No inhibition could be observed when the cofactor was added prior to the inhibitors, suggesting the existence of steric or conformational impediments. Both compounds exhibited time-dependent inhibition of all ALDH isoforms (Fig. 4a and Supplementary Fig. 2a), except for ALDH3A1 (Supplementary Fig. 2a), suggesting that DIMATE and ABD0171 act as irreversible inhibitors for all the other isoforms. The ALDH1A IC_{50} values with the physiological atRAL substrate ranged from 0.34 to 6 μM for ABD0171 and from 5.7 to 25 for DIMATE, after a 20-min preincubation (Table 1). In contrast, IC_{50} values for ALDH2 and ALDH3A1 remained substantially higher for both inhibitors, highlighting their selectivity against ALDH1A isoforms (Supplementary Table S1). Additionally, kinetic analysis confirmed non-competitive inhibition of ALDH3A1 by both DIMATE and ABD0171 with K_i values of 253 ± 48 μM and 233 ± 16 μM, respectively (Supplementary Fig. 2a and Supplementary Table S1).

To assess label specificity, we plotted k_{obs} *versus* [I] and calculated the k_{inact}/K_i value, representing the apparent second-order bimolecular rate constant (Table 1 and Supplementary Table S1). ABD0171 displayed the highest k_{inact}/K_i values against ALDH1A3, showing specific affinity-labeling for this isoform, as indicated by the hyperbolic k_{obs} *versus* [I] plot (Fig. 4a). In contrast, DIMATE showed specific affinity-labelling for ALDH1A1 (Fig 4a) but non-specific labeling for ALDH1A2 and ALDH2 (Fig. 2a and Supplementary Fig. 2a). At low concentrations (0-50 μM), DIMATE interacted specifically with ALDH1A3 (Supplementary Fig. 2b), while at higher concentrations, it acted as a non-specific affinity-label (Fig. 4a).

We confirmed the irreversible binding of DIMATE and ABD0171 to ALDH1A3 using electrospray ionization time-of-flight mass spectrometry (ESI-TOF-MS). Under varying voltage and temperature conditions, treatment with 167 μM DIMATE (1:10, ALDH1A3: inhibitor ratio) induced a complete shift in ALDH1A3 subunit mass resulting in species approximately 185 Da larger than the native protein subunit (Fig. 4b). We observed similar results with ABD0171, which generated species with a molecular mass increase of 333 Da compared to the untreated sample. At 50 μM (1:3, ALDH1A3: inhibitor ratio), DIMATE specifically modified ALDH1A3, leading to partial alteration of the native protein (Supplementary Fig. 2c). The mass spectra suggest irreversible modification with one DIMATE or ABD0171 molecule per protein subunit, corroborating our inhibition kinetics findings.

ABD0171 induces death receptor-mediated extrinsic apoptosis and sensitizes basal subtype cells to anoikis by inhibiting key signaling pathways

Previous reports showed that DIMATE was able to induce cell death in p53-mutated cells[34,35]. To investigate thoroughly the mechanism of cell death of ABD0171 and DIMATE in TNBC cell lines, we used the MDAMB231 (Basal B, p53-mutated), MDAMB468 and HCC70 cells (Basal A, defective and wild-type p53, respectively). Annexin V/PI staining indicated that both compounds induce apoptosis at equimolar concentrations (Fig. 5a). Apoptosis triggered by ABD0171 and DIMATE was characterized by reduced expression of FLICE-inhibitory protein (c-FLIP), decreased of cellular X-linked inhibitor of apoptosis proteins (cIAP2 and XIAP), and activation of initiator caspase-8. This sequence of events is indicative of the extrinsic apoptosis pathway activation. Consequently, activated forms of effectors caspase-7, caspase-3 and its cleavage product cPARP were prominently detected in all cell lines treated with ABD0171, and to a lower extent in cells treated with DIMATE (Fig. 5b). Activation of death receptors was accompanied by reactive oxygen species (ROS) generation (Supplementary Fig. 3a) and sustained activation of the oxidative stress-sensing JNK pathway and its downstream effector c-Jun, for both compounds (Fig. 5b). Stress activation of p38 was transient, with DIMATE-treated cells showing greater prominence. Both extrinsic apoptosis and JNK pathway activation can influence inflammation and the metastatic potential of TNBC through several mechanisms. Therefore, we further investigate the effects of ABD0171 and DIMATE on cell migration, invasion and anoikis, which are critical processes in the metastatic cascade. At sub-lethal concentrations, ABD0171 and DIMATE attenuated migration and significantly reduced the invasion capacity of MDAMB231, MDAMB468, and HCC70 cells (Fig. 5c and Supplementary Fig. 3b). Similarly, the compounds significantly diminished the ability of these cell lines to resist anoikis in a dose-dependent manner after 24 hours (Fig. 5d and Supplementary Fig. 3b). Intriguingly, Basal A MDAMB468 and HCC70 cell lines displayed a shift from ALDH1A3 to ALDH1A1 under non-anchored conditions (Fig. 5e), becoming more evident by day 10, when cells formed spheroids with distinct morphological organization. This shift may reflect the acquisition of EMT-like features, as ALDH1A1 has been previously associated with mesenchymal traits and stemness, which are often enriched in spheroids formed by TNBC cell lines[36].

As early as 24 hours, we observed robust activation of STAT3 in all cell lines under non-anchored conditions (Fig. 5f). The observation that ABD0171 and DIMATE reduce anoikis resistance in these cells suggested that both inhibitors could disrupt oncogenic reprogramming mediated by STAT3, triggered by detachment from the extracellular cell matrix. Consistently, ABD0171 and DIMATE simultaneously suppressed tPA, Src/FAK and IL6/JAK/STAT3 signaling pathways in anchored and non-anchored conditions, with

ABD0171 demonstrating a markedly stronger effect (Fig. 5f, g). In contrast, DIMATE elicited pro-survival activation of AKT in MDAMB468 cells and ERK in HCC70 cells, suggesting potential compensatory mechanisms that were not observed with ABD0171 (Supplementary Fig. 3c).

ABD0171 reduces TNBC cell proliferation and metastasis in mouse models, with demonstrated *in vivo* safety

Before evaluating *in vivo* efficacy, we assessed potential systemic, hepatic, and/or renal toxicity following repeated administration of ABD0171 and DIMATE in healthy animals. We administered equimolar doses of both compounds (30 and 40 mg/kg for ABD0171; 15 and 20 mg/kg for DIMATE) three times per week over 29 days. Neither 30 nor 40mg/kg of ABD0171, nor 15 mg/kg of DIMATE caused body weight loss in healthy animals (Supplementary Fig. 4a and Supplementary Data 3E) or produced visible signs of adverse events. In these groups we did not observe any significant alterations in serum biochemical parameters, including liver enzymes (alkaline phosphatase, ALP; alanine aminotransferase, ALT; gamma-glutamyl transferase, GGT), bile acids (BA), total bilirubin (TBIL), albumin (ALB), renal markers (blood urea nitrogen, BUN; creatinine, CRE), or electrolytes (Na^+, K^+, Cl^-). Other factors such as cholesterol (CHOL), glucose (GLU), calcium (Ca^{2+}), phosphate (PHOS), and total CO_2 (tCO_2) also remained unaffected, suggesting no impairment of hepatic or renal function. Hematological analysis showed no significant changes in red blood cell count (RBC), hemoglobin (Hb) levels, leukocyte numbers, or other key hematological parameters (Supplementary Data 3F). However, in the group treated with 20 mg/kg of DIMATE, we observed a transient decrease in body weight (Supplementary Fig. 4a and Supplementary Data 3E), accompanied by a moderate increase in tCO_2 levels ($p=0.003$) compared to controls, remaining within a non-critical range (Supplementary Data 3F).

We next evaluated the anticancer efficacy of ABD0171 and DIMATE *in vivo*, using two different TNBC orthotopic mouse models: the MDAMB231-GFP xenograft and the immunocompetent 4T1-Luc_M3 model. In both models, treatment with ABD0171 was well tolerated without observed weight loss (Supplementary Fig. 4b, c). Contrarily, DIMATE administration caused noticeable loss of weight in immunocompetent mice and lethal acute toxicity in 6 animals in the immunosuppressed MDAMB231-GFP model towards the end of the study (Fig. 6a, Supplementary Fig. 4b, c and Supplementary Data 3G). Both compounds resulted in significant inhibition of tumor growth (Fig. 6a, b). On day 30 of the experiment, treated mice bearing MDAMB231-GFP tumors showed tumor

growth reductions of 54% and 32% with 40 mg/kg ABD0171 and 20 mg/kg DIMATE, respectively, compared to controls ($p < 0.0001$, $p = 01710$) (Fig. 6a). Similarly, both compounds reduced 4T1-Luc_M3 tumor growth by 14% and 20% on day 24 of the experiment ($p = 0.0430$, $p = 0.0004$, respectively) (Fig. 6b and Supplementary Data 3H).

Histopathological analysis of MDAMB231 tumor tissues revealed neoplastic proliferation with pleomorphic cells, absence of glandular or ductal structures, poor vascularization, and extensive necrosis (Fig. 6c). Treatment with DIMATE and ABD0171 resulted in a significant increase in cleaved caspase-3 and cleaved PARP-positive cells compared to controls, indicating enhanced apoptosis (Fig. 6c and Supplementary Data 3H).

Similarly, 4T1-M3 tumor tissues displayed disorganized architecture, dense cellularity, pleomorphic cells with scant cytoplasm, and high Ki67 positivity, consistent with elevated proliferation (Fig. 6d). While DIMATE-treated tumors showed increased cleaved caspase-3 positivity, this effect was not observed with ABD0171 treatment. However, ABD0171-treated tumors demonstrated heightened T-cell infiltration, predominantly CD8+ and, to a lesser extent, CD4+ T cells, suggesting a potential immunomodulatory effect (Fig. 6d, e and Supplementary Data 3H). Complementing these primary tumor effects, treatment with ABD0171 and DIMATE significantly reduced metastasis to lymph nodes, brain, lumbar spine, and kidneys in this model (Fig. 6f and Supplementary Data 3I), supporting the antimetastatic activity observed in the functional *in vitro* studies and highlighting the therapeutic potential of these compounds in targeting both primary tumor growth and metastatic spread.

DISCUSSION

Drug resistance and the persistence of breast cancer cells with stem/progenitor properties remain significant challenges in treating aggressive breast cancer[37,38]. ALDH1 activity has been closely associated with stemness features and plays crucial roles in driving this therapeutic resistance[10,11,37–39]. In this study, we thoroughly characterized the expression diversity of the ALDH1A subfamily across molecular subtypes and cellular populations in breast cancer, uncovering distinct profiles of *ALDH1A1*, *ALDH1A2*, and *ALDH1A3* in both patient-derived tumors and breast cancer cell lines. Our analysis highlights the varied roles that these isoforms may play within different pathological entities and the tumor microenvironment, emphasizing their potential as therapeutic targets.

One of the key findings of this study is the predominant expression of *ALDH1A3* in the Basal-like molecular subtype of breast cancer, where it clusters with Basal-like gene

markers in the PAM50 signature[22,23]. The elevated expression of *ALDH1A3*, primarily attributed here to the epithelial cancer cell compartment within tumors, is consistent with the significantly higher levels of ALDH1A3 observed in the Basal A subtype of TNBC cell lines, which are also characterized by epithelial-like features[26]. These results complement and expand prior studies linking ALDH1A3 expression to aggressive TNBC phenotypes and poor clinical outcomes[15,17,18,40]. TNBC and Basal-like breast cancers often exhibit overlapping morphological characteristics and biological aggressiveness; however, they are not entirely identical entities[28,41–43]. Basal-like TNBC tumors exhibit distinct molecular alterations (e.g. TP53 mutations and increased mitotic activity), unique patterns of organ-specific metastases as well as worse survival outcomes compared to TNBC tumors that are not Basal-like[42,43]. Additionally, other breast cancer subtypes, including specific HER2-enriched tumors, certain luminal B tumors, and the majority of claudin-low phenotype tumors, can exhibit molecular and histological characteristics similar to Basal-like cancers[24,28].

Interestingly, recent findings from a comprehensive multi-dimensional atlas of breast tissues identified a distinct subset of cells termed basal luminal alveolar cells[44]. These epithelial cells accumulate with age, exhibit low lineage fidelity, and express high levels of stem/progenitor markers ALDH1A3 and CD73. Notably, their gene expression signatures strongly associate with Basal-like breast cancers, aligning with our observations and suggesting that they may represent a critical ALDH1A3-enriched population implicated in tumor initiation. These findings highlight the potential of ALDH1A3 as both a Basal-like biomarker surrogate and a therapeutic target. While further investigation is necessary, incorporating ALDH1A3 into immunohistochemistry strategies could enhance the sensitivity of existing biomarker panels for Basal-like TNBC, which currently rely on basal marker staining (cytokeratin 5/6, -14, or -17, and/or EGFR) in combination with hormone receptor and HER2 negativity, achieving a reported sensitivity of 76%[45].

Our single-cell analysis also provides valuable insights into the expression of *ALDH1A1*, a marker associated with breast cancer stem cells and poor prognosis[12], and *ALDH1A2* isoforms within breast tumors. Unlike *ALDH1A3*, *ALDH1A1* and *ALDH1A2* were less prominently expressed in cancer epithelial cells, displaying distinct cellular distribution profiles. ALDH1A1 was predominantly enriched in myeloid cells, while ALDH1A2 was primarily expressed in endothelial cells. Although the three isoforms were detected in myCAF-like cells, only ALDH1A1 was significantly expressed in this relevant subtype of cancer-associated fibroblasts known for its contribution to mechanical support,

extracellular matrix remodeling, and stromal stiffness, facilitating tumor invasion. An intriguing aspect of our findings is the high expression of *ALDH1A1* and *ALDH1A2* in claudin-low breast cancers and in cell lines resembling these tumors. This observation aligns with a recent study that identified *ALDH1A1* enrichment in a specific claudin-low subgroup of breast tumors, characterized by high stemness scores, immune cell infiltration, and strong epithelial-to-mesenchymal transition (EMT) marker expression[2]. The authors suggested that the observed *ALDH1A1* expression could partly result from infiltrating immune cells. Furthermore, our *in vitro* experiments demonstrated that breast cancer cells can transition from ALDH1A3 to ALDH1A1 protein expression under EMT-inducing conditions, such as anchorage-independent growth, highlighting a plasticity in ALDH1 isoforms expression in response to tumor microenvironment cues. The prominent expression of ALDH1A1 and ALDH1A2 in different stromal populations and in claudin-low tumors, known for their stromal richness, suggests distinct functional roles within these tumors. ALDH1A1 likely contributes to myeloid-associated tumor immunity by modulating the metabolic adaptation of myeloid cells within the tumor microenvironment[30,46] and promoting the mesenchymal traits characteristic of these primitive tumors. These findings highlight the need for further investigation into the specific roles of ALDH1 isoforms in aggressive breast cancer phenotypes and emphasize the importance of further exploring the role of ALDH1A1 in immunosuppression, ALDH1A2 in angiogenesis, and the broader contribution of ALDH1 isoforms on matrix remodeling within the tumor microenvironment.

The profiling of ALDH1A isoforms and the therapeutic hypotheses that they prompt guided our efforts to develop novel ALDH1A inhibitors, designed as structural analogues of DIMATE, a clinical-stage ALDH inhibitor[47]. Among these, ABD0171 emerged as a potent irreversible inhibitor with specific affinity labeling properties for ALDH1A3. Belonging to the series V of analogs, ABD0171 demonstrated enhanced efficacy against ALDH1A3-expressing Basal-like cancer cells compared to DIMATE. Structure-activity relationship analyses revealed the critical role of aryl substitutions in series V analogs for enhancing inhibitory activity, enabling ABD0171 to achieve GI50 values below 10 μM in most Basal A and Basal B cell lines.

A key challenge in translating novel ALDH1A3 inhibitors into the clinical setting lies in achieving selectivity while sparing ALDH2, the closest related enzyme within the ALDH1A subfamily based on evolutionary divergence[48]. Unlike the cytoplasmic ALDH1A isoforms, ALDH2 is localized to mitochondria and plays an essential role in hematopoiesis, the prevention of mutagenesis in somatic tissues and cardioprotection through its enzymatic activity, which mitigates oxidative stress and clears toxic

endogenous acetaldehyde and formaldehyde[49]. ABD0171 exerts only marginal inhibitory effects on ALDH2 and demonstrates low potency against ALDH3A1, an isoform that constitutes 10–40% of the water-soluble protein in the mammalian cornea and plays a crucial role in protecting this tissue from oxidative stress[50]. This selective inhibition of ALDH1A3, combined with favorable cytotoxicity profiles, underscores the potential of ABD0171 to overcome the safety challenges observed with broader ALDH inhibitors, making it a highly promising therapeutic candidate for further clinical development.

In comparative functional studies, ABD0171 exhibited higher predictive cytotoxicity in TNBC models, as determined through linear mixed modeling. It also demonstrated a stronger induction of death receptor-mediated apoptosis in TNBC cell lines compared to DIMATE. These apoptotic events were associated with elevated ROS levels and sustained activation of the oxidative stress-sensing JNK pathway (JNK/cJun). The differences in apoptotic effector activation between ABD0171 and DIMATE may be attributed to the concurrent activation of prosurvival pathways, such as AKT and ERK, and the transient activation of p38 observed prominently in DIMATE-treated cells. While AKT and ERK signaling are well known to promote cell survival and counteract apoptosis, the transient activation of p38 may further induce a prosurvival stress response that mitigates caspase activation. This mechanism has been previously reported under chemotherapy-induced stress conditions, such as those induced by Paclitaxel and doxorubicin in TNBC cells[51,52]. This further underscores the differential capacities of ABD0171 and DIMATE to modulate stress-response pathways in TNBC models.

In vivo, both ABD0171 and DIMATE significantly inhibited tumor growth in MDA-MB-231 xenografts and syngeneic 4T1-Luc_M3 TNBC models. ABD0171 demonstrated superior tolerability, reflected in its favorable safety profile, with minimal systemic, hepatic, or renal toxicity. Notably, ABD0171 uniquely enhanced cytotoxic CD8+ T cell infiltration into the tumor microenvironment, an observation that warrants further investigation, particularly in TNBC, where the tumor microenvironment is predominantly immunosuppressive, often compromising responses to immunotherapies. . This finding raises the possibility of ABD0171 not only as a direct antitumor agent but also as a potential modulator of tumor immunity, offering potential benefits in combination with immune checkpoint inhibitors. Moreover, ABD0171 was more effective than DIMATE in reducing tumor burden and metastasis to distant organs, including lymph nodes and the lumbar spine, while both compounds exhibited comparable antimetastatic activity in the brain and kidney.

These findings align with our *in vitro* studies, where sub-lethal concentrations of ABD0171 and DIMATE suppressed key metastatic traits such as migration, invasion, and anoikis resistance in TNBC cells by disrupting the tPA, Src/FAK, and IL6/JAK/STAT3 signaling pathways. The modulation of these pathways through ALDH1A3 inhibition underscores its pivotal role in TNBC progression. tPA, a key mediator of aggressive breast cancer phenotypes, has been shown to promote invasion and lymph node metastasis through its regulation by ALDH1A3-derived RA, the principal product of ALDH1A enzymatic activity and a major regulator of gene expression in tumor biology[16]. RA is reported also to modulate the expression of Src activators such as integrins and EGFR, which are frequently upregulated in TNBC[53]. These activators, together with elevated tPA levels, converge to activate Src kinase and FAK, driving cytoskeletal remodeling and cell motility, hallmarks of metastatic dissemination54,55.

RA is also reported to influence cytokine production, notably IL6, a critical activator of the JAK/STAT3 pathway56. This central axis governs multiple oncogenic processes, including epithelial-to-mesenchymal transition (EMT), tumor proliferation, chemoresistance, immune evasion, and metastasis57,58. Aberrant activation of STAT3, observed in approximately 50–70% of TNBC cases, amplifies a plethora of oncogenic signaling, making it a key driver of TNBC pathogenesis and a significant therapeutic target in breast cancer research59–61. Our findings underscore the potential of ALDH1A3 inhibition to disrupt a broad spectrum of signaling networks integral to Basal-like or TNBC subtypes in general and suggest additional therapeutic potential in combination strategies to overcome immune evasion in aggressive breast cancers.

In conclusion, the distinct expression patterns of ALDH1A3 across breast cancer subtypes and tumor cell populations points to its association with a specific subset of epithelial cells, characterized by unique properties that strongly contribute to disease initiation and progression. Furthermore, the predominance of ALDH1A1 and ALDH1A2 in specific stromal cells provides valuable insights into the complex biology of breast cancer, highlighting the need to address both tumor genetic heterogeneity and the intricate tumor-stroma interactions when designing ALDH-targeted therapies.

The development of selective ALDH1A3 inhibitors, such as ABD0171, represents a promising approach to tackle the therapeutic challenges linked to aggressive breast cancer subtypes, including TNBC and Basal-like cancers. By targeting key signaling pathways, ABD0171 has the potential to significantly enhance patient outcomes in these subtypes, where effective treatment options remain limited. Future research should focus on evaluating combinatorial therapies and deepening our understanding of the roles of

ALDH1A in tumor immunity and resistance mechanisms. These steps will be critical to unlock the full clinical potential of ALDH1A inhibition as a transformative strategy in breast cancer therapy.

METHODS

Public data-based transcriptomic profiling of breast cancer cell lines and tumors

Breast cancer cell lines: CCLE 2019 bulk RNA-seq data (raw read counts) for 50 human breast cancer cell lines were retrieved from the Dependency Map portal (DepMap)[25], where reads had previously been aligned to the reference genome (GRCh37.p13, GENCODE release 19). Genes with low expression were filtered out, retaining only those with sufficiently high counts for further statistical analysis. Using the edgeR package[62], normalization factors were calculated to scale the raw library sizes via the trimmed mean of M-values (TMM) method. Log_2 counts per million (CPM) values were then computed as a measure of gene expression levels. The molecular subtypes of the cell lines were kept as annotated in DepMap, ensuring consistency with existing classification standards. For unsupervised clustering, the log_2 expression of 59 genes, which included the well-known PAM50 signature[22,23], *CHD1, CLDN3, CLDN4, CLDN5, CLDN7, OCLN*, and *ALDH1A* isoforms (*ALDH1A1, ALDH1A2, ALDH1A3*), was used as input. This analysis was conducted using the partitioning around medoids (PAM) algorithm set to k=4 (Supplementary Data 1B and Fig. 1c). Subsequently, the median expression of these 59 genes was calculated for each of the four groups, based on the standard molecular classification of breast cancer cell lines: Basal A, Basal B, HER2-enriched, and Luminal.

Breast tumor patient samples: mRNA (bulk) gene expression data from three breast cancer studies were retrieved from cBioPortal (https://www.cbioportal.org): Breast Invasive Carcinoma (TCGA, Cell 2015)[21], Breast Invasive Carcinoma (TCGA, PanCancer Atlas)[20], and Breast Invasive Carcinoma (TCGA, Firehose Legacy). After merging the individual patient samples, a total of 1103 unique samples were selected across the three studies. As described above, using the edgeR package, normalization factors were calculated to scale the raw library sizes via the trimmed mean of M-values (TMM) method. Log_2 counts per million (CPM) values were then computed as a measure of gene expression levels. Next, the log_2 expression of 59 genes was used as input for unsupervised clustering using the PAM algorithm with k=5 (Supplementary Data 1A and Fig. 1b). Following this, the median expression of these 59 genes was calculated for

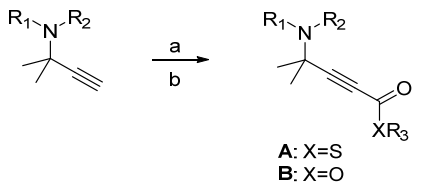
each of the five groups, based on the molecular classification: Basal-like, Claudin-low, HER2-enriched, Luminal A, and Luminal B.

Single-cell of breast cancer cell lines: Pre-processed scRNA-seq data from 32 cell lines, initially comprising 35,276 cells, was retrieved from: https://figshare.com/articles/dataset/Single_Cell_Breast_Cancer_cell-line_Atlas/15022698[29]. Subsequently, genes with high relative expression per cell were filtered out, and the analysis focused on a subset consisting only of cell lines classified as Basal A or Basal B, which represents 16 cell lines encompassing a total of 20,485 cells after filtering. Further data processing including normalization, dimensionality reduction, and clustering was conducted with the Seurat R package (v4.4)[63]. Normalization was performed using the sctransform v2 method with the glmGamPoi package[64]. The FeaturePlot and VlnPlot functions from Seurat were used to visualize specific gene expressions across identified clusters.

Single-cell of breast tumor patients: Pre-processed scRNA-seq data from breast cancer patients, comprising 26 individual patient samples and a total of 100,064 cells, was retrieved from https://singlecell.broadinstitute.org/single_cell/study/SCP1039[30]. Further data processing including normalization, dimensionality reduction, and clustering was conducted with the Seurat R package (v4.4)[63]. Prior to data integration, individual patient samples were grouped according to their molecular classification: Basal-like, HER2-enriched, Luminal B/HER2+, Luminal A, and Luminal B. Normalization was performed using the sctransform v2 method with the glmGamPoi package. Additionally, to remove potential confounding sources of variation, the mitochondrial mapping percentage was regressed during normalization. Data integration was performed using the SCT method and 3000 features for anchoring. The number of principal components was set to 30. The FeaturePlot function from Seurat was used to visualize specific gene expressions across identified clusters. Visualization was conducted using ggplot2[65], GGally[66], complexHeatmap[67] and Seurat[63] R packages.

Chemical synthesis of compounds.

Various structural derivatives of DIMATE were synthesized according to the following scheme:



Aminothiolesters synthesis: **a:** nBuLi, THF, -70°C to 0°C – b: COS, -70°C to 0°C / R3I, 0°C [method 1] or CO2, 0°C to 20°C / isobutyl chloroformate, 0°C to 20°C/ NaSMe, 20°C [method 2]

The process and experimental details for the preparation of compounds **A** or **B** have been described previously[32,33,68,69]. Briefly, aminothiolesters **A** were obtained by deprotonation of corresponding acetylenic amines with n-Butyllithium. The resulting acetylides were treated with carbonyl sulfide (COS) then with alkyl iodide [method 1]. An alternative process involved the reaction of lithium acetylide intermediate with carbon dioxide followed by treatment with isobutyl chloroformate and finally with sodium methanethiolate (NaSMe) [Method 2]. CEY1410 (R1, R2=Me, X=O, R3=H) was performed by treatment of lithium acetylide intermediate by carbon dioxide followed by an acidic work-up. Corresponding alkylesters **B** (R3=Me, Et) were obtained by Fisher esterification of CEY1410.

Protein expression and purification

Escherichia coli BL21 (DE3) pLysS strain was transformed with the plasmid pET-30 Xa/LIC that encoded ALDH1A1, ALDH1A2, ALDH1A3, ALDH3A1 or ALDH2 fused to the N-terminal $His_{(6)}$ tag. The ALDH isoforms were expressed and purified as previously described[70] using a nickel-charged chelating Sepharose™ Fast Flow 5 mL column (HisTrap) and an ÄKTA™ FPLC purification system.

Enzymatic activity assay

Standard activities were measured prior to each inhibition experiment by using hexanal (for ALDH1A and ALDH2) or 4-nitrobenzaldehyde (for ALDH3A1) as a substrate as described[70]. DTT was removed from the reaction buffer and the pH was adjusted at 7.2. Thus, 50 mM HEPES, 0.5 mM EDTA, pH 7.2 was used for ALDH1A1, ALDH1A2 and ALDH2; 50 mM HEPES, 30 mM $MgCl_2$, pH 7.2 was used for ALDH1A3; and 50 mM Tris/HCl, pH 7.2 was used for ALDH3A1. Enzymatic activity assays were performed at 37°C using a Cary Eclipse fluorescence spectrometer.

Inhibitor screening for human ALDH1A, ALDH3A1 and ALDH2 isoforms and determination of the IC_{50} values

Single-point analyses of enzymatic activity at 10 μM of inhibitor were performed by the fluorimeter methodology described previously[70]. Enzymatic inhibition assays were performed at 37°C, in a final volume of 1 mL and in an appropriate reaction buffer. All compounds tested as inhibitors (including DIMATE and ABD0171, Fig. 3b) were dissolved in ethanol and assayed in a final concentration of 1% ethanol (v/v) using hexanal (for ALDH1A and ALDH2) and 4-nitrobenzaldehyde (for ALDH3A1) as a substrate. The inhibitor effect was checked at both saturating and K_m concentrations of substrate[71]. Usually, we performed the inhibition experiments at saturating concentrations of substrate for irreversible and non-competitive inhibitors but substrate concentrations at K_m values were also used to identify possible competitive inhibition. For this reason, the inhibitory effect of compounds was tested at two substrate concentrations: 10 and 0.25 μM hexanal for ALDH1A1, at 250 and 10 μM hexanal for the remaining ALDH1A and ALDH2 enzymes and at 250 and 5 μM 4-nitrobenzaldehyde for ALDH3A1. The concentration of enzyme was kept from 50- to 100-fold lower than that of the substrate. The reaction mixture containing enzyme and inhibitor in the appropriate reaction buffer was incubated for 20 min at 37°C, before adding the corresponding NAD^+/$NADP^+$ cofactor, NADH/NADPH internal standard and substrate (hexanal/4-nitrobenzaldehyde). The percentage of remaining activity was calculated from the ratio of activity at a given inhibitor concentration to the control activity with 1% ethanol (v/v) without inhibitor. The IC_{50} values were calculated at saturating concentrations of substrate by fitting the initial rates to the appropriate equation using GraFit 5.0 (Erithacus Software) and the values were given as the mean ± standard error.

Time-dependent inhibition assay and determination of k_{obs} and k_{inact}/K_i parameters

Different concentrations of DIMATE and ABD0171 were incubated in a range of time including 0, 2, 5, 10, 15, 20, 30, 45, 60, 90, 120, 180 min at 37°C with ALDH enzymes (ALDH1A and ALDH2) in the corresponding reaction buffer, before to the addition of NAD^+ cofactor, NADH internal standard and hexanal at saturating concentrations and at concentrations near K_m values. The corresponding inhibitor solutions were prepared to be assayed in a final concentration of 1% ethanol (v/v). The concentration of enzyme was kept from 50- to 500-fold lower than that of the substrate for all enzymatic assays and 500 μM NAD^+ was added after preincubation time. The inhibitory effect of compounds was tested using the fluorometric assay, at 10 and 0.25 μM hexanal for ALDH1A1, at 250 and 10 μM hexanal for ALDH1A2, ALDH1A3 and ALDH2 enzymes, to determine the effect of substrate concentrations on the inhibition. In all cases, k_{obs} was obtained at saturating concentrations of substrate by fitting the data using GraFit 5.0 to

the following equation: $v_t = v_i \cdot exp\ (-k_{obs} \cdot t)$, where v_t is the measured steady state velocity and v_i is the initial velocity at short preincubation times (t). Finally, the dependence of k_{obs} on inhibitor concentration was represented to obtain the k_{inact}/K_i rate constant, which is the most reported parameter to describe time-dependent inhibitors[72].

Non-competitive inhibition of DIMATE and ABD0171 against ALDH3A1 and determination of K_i value

The inhibitory effect of DIMATE and ABD0171 against ALDH3A1 was analyzed fluorometrically using 4-nitrobenzaldehyde as a substrate, 1 mM $NADP^+$ as a cofactor and 5 μM NADPH as an internal standard in the corresponding reaction buffer. Activity assays to determine the type of inhibition and K_i values were performed using various substrate concentrations at fixed inhibitor concentrations (0, 50, 250 and 500 μM DIMATE and 0, 100, 250 and 500 μM ABD0171). The data of enzymatic activities were fitted to the Michaelis-Menten equation using GraFit 5.0, to determine the values of K_m and V_{max}. Then, results were fitted to the equations of the different types of enzymatic inhibition and K_i values were determined choosing the type of inhibition yielding the best fit. Values were expressed as the mean ± standard error.

HPLC-based assay for the inhibition characterization using all-*trans*-retinaldehyde as a substrate

To evaluate the retinaldehyde dehydrogenase activity of human ALDH1A isoforms, an HPLC method was used in presence of all-*trans*-retinaldehyde as a substrate and DIMATE and ABD0171 as inhibitors. Stock solutions of retinoid substrate were stored in ethanol. Working solutions of retinoids were prepared by 10-min sonication in the presence of equimolar delipidated BSA[73]. The actual concentration of solubilized retinoid was determined based on the corresponding molar absorption coefficient (ε_{370} = 29,500 $M^{-1} \cdot cm^{-1}$). The enzymatic reaction was carried out for 15 min at 37°C in a final volume of 500 μL, after 20-min preincubation with different concentrations of DIMATE and ABD0171 and using a saturating concentration of cofactor (500 μM NAD^+) in the corresponding reaction buffer. With the aim to measure the steady-state enzymatic activity, the concentration of enzymes was kept from 50- to 100-fold lower than that of the substrate for all enzymatic assays. The reaction was stopped with 1 mL cold hexane:dioxane:isopropanol (50:5:1, v/v/v), and retinoids were extracted by a two-step procedure with the same solvent mixture[74,75]. The aqueous phase was removed, and the organic phase was evaporated under a N_2 stream. Retinoids were dissolved in 200 μL of hexane and analyzed by a modification of a published method[76], as follows. Retinoids were separated by HPLC on a Novapak® Silica 4 μm, 3.9 x 150 mm in hexane/methyl-

tert-butyl ether (96:4, v/v) mobile phase, at a flow rate of 2 mL/min, using a Waters Alliance 2695 HPLC. Elution was monitored at 370 nm and 350 nm for all-*trans*-retinaldehyde and all-*trans*-retinoic acid, respectively, using a Waters 2996 photodiode array detector. Quantification of retinoids was performed by interpolation of HPLC peak areas into a calibration curve of known retinoid concentration (data not shown). IC_{50} and k_{obs} parameters were calculated by fitting the initial rates to the appropriate equation using GraFit 5.0 and values were given as the mean ± standard error. All compound manipulations were performed under dim red light to prevent retinoid photoisomerization.

Electrospray Ionization Time-of-Flight Mass Spectrometry (ESI-TOF-MS) assays

ALDH1A3 wild type was stored at −20°C in 100 mM ammonium bicarbonate, pH 7.5, at 3.75 mg/mL (16.74 μM). The concentration of DIMATE and ABD0171 stock was prepared 100-fold higher in order to work at 1% EtOH (v/v). Fresh protein samples were incubated with the corresponding compound for 45 min at 37°C. The molecular weight of 16.7 μM ALDH1A3 formed after 45 min-incubation with 167 μM DIMATE, 50 μM DIMATE and 167 μM ABD0171 was determined by Electrospray Ionization Time-of-Flight Mass Spectrometry (ESI-TOF-MS) on a Micro TOF-Q instrument (Bruker Daltonics, Bremen, Germany), connected to a Series 1200 HPLC Agilent pump and controlled by the Compass Software. Samples were analyzed under neutral conditions (pH 7.0), using as running buffer a mixture of acetonitrile:ammonium acetate (15 mM) solution (20:80, v/v). For each analysis, 20 μL of protein solution were injected at 30-40 $\mu L \cdot min^{-1}$ and analyzed under the following conditions: capillary counter-electrode voltage, 3500-5000 V; dry temperature, 90-110°C; dry gas, 6 $L \cdot min^{-1}$; m/z range 800-3000.

Cell lines and culturing

The human breast cancer cell lines (Supplementary Data 3A) used in this study included DU4475, HCC1143, HCC1937, MDAMB231, MDAMB468, MDAMB453, SKBR3, MCF7, and BT474, obtained from the German Collection of Microorganisms and Cell Cultures GmbH (DSMZ), as well as HCC70, BT20, and MDAMB436, purchased from the American Type Culture Collection (ATCC). Cell lines HCC1419, HCC1569, HCC1954, MDAMB361, and T47D were generously provided by Dr. Ana Ruiz-Saenz (Dept. of Cell Biology, Erasmus Medical Center, Rotterdam, Netherlands), and Hs578T was kindly provided by Dr. Aida Zulueta-Morales (Dept. of Health Sciences, University of Milan, Italy).

The authenticity of cell lines was verified by polymorphic short tandem repeat (STR) profiling (Microsynth, Switzerland). All cell lines were cultured following the supplier's

recommended media and supplements and were maintained at 37°C in a humidified incubator with 5% CO_2.

Cell cytotoxicity assay

Cells (0.5 x10^4 cells/well) were seeded into 96-well cell culture plates and incubated with various concentrations of drugs (1–50μM) or control medium for 48 hours. The growth-inhibitory effect of the drugs was analyzed using an *in vitro*, resazurin-based toxicity assay, following the manufacturer's instructions (Sigma-Aldrich, Darmstadt). The drug response was quantified by the half-maximal inhibitory growth concentration (GI_{50}) and determined by non-linear regression analysis of log-dose/response curves. The cut-off value to define cell resistance was determined statistically (above GI50 geometric mean+ S.D.).

3D clonogenic assay

Patient-derived tumor xenografts (PDX) cells were plated in ultra-low attachment 96-well plates containing soft-agar medium and incubated with various concentrations of two drugs, ABD0171 and DIMATE (1-30μM) or control medium for 8 to 13 days, at 37°C and humidified air with 7.5% CO2. Colonies were stained with 2-(4-iodophenyl)-3-(4-nitrophenyl)-5-phenyltetrazolium chloride and counted with an automatic image analysis system (CellInsight NXT, Thermo Scientific or Bioreader 5000 V, BIO-SYS).

Data was collected from 11 models (7 TNBC, and 4 HER2-enriched), each treated with 9 different drug concentrations, along with control samples (untreated). Control values were calculated using the median across replicates and were used to normalize the response for each drug concentration. After normalization, the mean was calculated for two or three replicates per concentration, depending on the PDX model, to generate the final response values for each drug concentration. Dose-response curves were fitted for each model and drug using the dr4pl R package[77], and the GI_{50} values were extracted from the fitted models.

Linear mixed-effects model analysis

GI_{50} values, representing drug efficacy in both cell lines and PDX models, were selected for analysis. In the case of cell lines, molecular subtypes "Basal A" and "Basal B" were reclassified as triple-negative breast cancer (TNBC) to match the classification used for the PDX models, where only HER2-enriched and TNBC subtypes were available. A combined dataset of cell lines and PDX models was used to fit a linear mixed-effects model (LMM) using the lmerTest R package[78]. This model assessed the effects of drug (DIMATE or ABD0171), molecular subtype (TNBC or HER2-enriched), and their

interaction on the observed GI_{50} values. The model included a random intercept to account for variability between individual models (both cell lines and PDXs). Satterthwaite's approximation, provided by the lmerTest package, was applied to calculate degrees of freedom and p-values for the fixed effects. Both observed and predicted GI_{50} values were extracted for further analysis and interpretation.

Apoptosis assays

Apoptotic cells were quantified using eBioscience Annexin V Apoptosis Detection Kit FITC (Thermo Scientific, Strasbourg, FR). Cells were incubated with Annexin V-FITC and propidium iodide (PI) according to the manufacturer's instructions. Staining was performed for 15 minutes at room temperature, followed by analysis via flow cytometry. Annexin V-positive/PI-negative cells were classified as early apoptotic, while Annexin V-positive/PI-positive cells were considered late apoptotic or necrotic.

Capillary electrophoresis immunoassay

For protein expression analysis, cells were lysed in RIPA buffer containing phosphatase and protease inhibitors (Calbiochem, San Diego, CA, USA). Total protein extracts were quantified and separated via capillary electrophoresis, using the WES Simple Western system and following manufacture's specifications (Bio-techne, Minneapolis, USA). Immunodetection was performed using WES standard protocols[79]. Quantification of chemiluminescence signals and imaging were performed using Compass software for WES Simple western instruments (Bio-techne). Antibodies used in these studies are listed in Supplementary Table S2.

Cell migration and invasion assays

Migration assays were performed using the Oris Cell Migration Assembly Kit-FLEX (PlatypusTechnologies, Madison, USA). Cells (5×10^4 per well) were seeded into a 96-well plate containing an insert that created a defined, cell-free gap. After 24h of incubation to reach 100% confluence, ALDH inhibitors were added to the culture media, and the inserts were removed to allow migration. The cell-free area was monitored over 24h and quantified using the SpectraMax I3-Minimax station (Molecular Devices, USA) and ImageJ software[80]. Calcein AM staining (Fisher Scientific) was used for imaging at 0 and 24h. Data were obtained in triplicate for each condition.

For invasion assays[81], cells ($5–10 \times 10^4$) were seeded onto ECM-coated filter membranes within transwell inserts (Corning, Lyon, FR). The lower chamber of a 24-well plate was filled with 600 µL of complete media, with or without ALDH inhibitors. After 24–48h of incubation, non-migrated cells on the upper membrane surface were removed

using cotton swabs. Migrated cells on the underside of the membrane were fixed in 70% ethanol, stained with 0.2% crystal violet, and quantified by manual counting under a microscope.

Anoikis assay

Cells were cultured in 96-well plates coated with poly-HEMA (non-anchored conditions) or uncoated plates (anchored conditions) and treated with DIMATE or ABD0171 for 24h. Calcein AM and propidium iodide (PI) were sequentially added to each well, incubated for 20 min at 37°C, and fluorescence signals were detected using the SpectraMax MiniMax 300. Green fluorescence (Ex 495 nm; Em 530 nm) indicated live cells stained with Calcein AM, while red fluorescence (Ex 530 nm; Em 645 nm) indicated dead cells stained with PI. For further investigation, the assay was scaled up using 100-mm dishes under similar conditions. After 14 h of treatment with DIMATE or ABD0171, capillary electrophoresis immunoassay was performed as described above.

Quantification of soluble IL-6 and tPA proteins

Cells were cultured in 100-mm dishes under anchored (uncoated) or non-anchored (poly-HEMA-coated) conditions. After adherence, cells were treated for 14h with DIMATE, ABD0171, or vehicle control. Supernatants were collected, centrifuged, and assayed for IL-6 and tPA protein levels using the Human IL-6 and Human t-Plasminogen Activator DuoSet ELISA Development Kit (R&D Systems, Minneapolis, MN, USA), following the manufacturer's instructions.

Animal experiments

All mice were obtained from Janvier Labs (Le Genest-Saint-Isle, France) or Charles River GmbH (Sulzfeld, Germany) and housed at the Commune Scar Rockefeller animal facility (Lyon, France) or Reaction Biology Europe GmbH (Freiburg, Germany), following institutional and governmental animal welfare regulations.

For toxicity studies, four female SCID-CB17 immunocompromised mice per group were treated with intra-peritoneal injections of DIMATE (15 or 20 mg/kg) or ABD0171 (30 or 40 mg/kg), t.i.w. for four weeks. The control group received vehicle (PBS containing 10% Kolliphor® HS15). Mice were monitored for body weight, physical condition, and blood sampling was conducted (days 5, 19, and 26) for haematological and enzymatic analyses.

For xenograft models, 5×10^6 MDAMB231-GFP cells were orthotopically implanted into 50 SCID-CB17 immunocompromised mice. For syngeneic models, 0.5×10^6 4T1_Luc-M3 cells were orthotopically implanted into 36 BALB/cAnNCrl immunocompetent mice.

Once tumors reached ~50 mm^3, mice were randomized into treatment cohorts: DIMATE (20 mg/kg), ABD0171 (40 mg/kg), or vehicle control (n = 10 per group for MDAMB231-GFP, n = 12 per group for 4T1_Luc-M3). Treatments were administered intra-peritoneally, t.i.w. for three weeks. Tumor growth, body weight, and overall health were monitored daily. Sacrifice occurred at the study endpoint (week 3) or earlier if tumor volumes reached 700 mm^3 (MDAMB231-GFP) or 1600 mm^3 (4T1_Luc-M3). For the 4T1_Luc-M3 model, various organs were harvested for *ex vivo* luciferase assays to assess metastasis.

Immunohistochemical stainings

Formalin-fixed, paraffin-embedded (FFPE) tumor samples were analyzed by immunohistochemistry (IHC) at CIQLE, Lyon (UAR3453 CNRS, US7 Inserm, UCBL). Hematoxylin and eosin (H&E) staining was performed on 3 µm sections using standard protocols. For IHC, sections were incubated with the following primary antibodies: anti-Ki67 (Abcam, Paris), anti-cleaved caspase-3 (Cell Signaling, Danvers, MA), anti-cleaved PARP (Cell Signaling), anti-CD4 (Abcam), and anti-CD8α (Cell Signaling). Secondary antibodies conjugated to biotin were applied, and staining was performed using the Roche Ventana Discovery XT immunostainer (Roche Diagnostics, Meylan). Antibody detection was achieved with the DAB Map Detection Kit (Roche Diagnostics). Image analysis and quantification were carried out using the open-source software QuPath.

Ex vivo bioluminescence imaging (BLI) analysis

Organs from experimental animals were analyzed for bioluminescent signals. The control group data were separated to facilitate outlier detection and removal using the interquartile range (IQR) method. Median values from the control samples were calculated for each organ and plate, and these were used to normalize the replicate measurements of the experimental samples. Negative normalized values were adjusted to zero to ensure all values represented positive biological signals. To identify potential outliers, group-specific criteria were applied. For Group 1 (PBS), both lower and upper bounds for outlier detection were defined based on the IQR, while for Groups 2 and 3 (DIMATE and ABD0171), only upper-bound outliers were flagged. Subsequently, outliers were removed from each group. Ex vivo BLI values were calculated for each sample by adjusting the median of the normalized replicates according to sample volume (10 µL) and animal body weight (g), yielding values expressed as BLI per gram of tissue. Statistical comparisons between treatment groups and controls were performed using the Wilcoxon rank-sum test, with p-values adjusted for multiple comparisons using the Benjamini-Hochberg method. A threshold of $p \leq 0.05$ was used to determine significance.

Statistical analyses of experimental data

GraphPad Prism (GraphPad Software, Inc., La Jolla, CA, USA) was used for the graphic representations and statistical analysis of the data. Mann-Whitney test was used for the comparison between two groups and Kruskal-Wallis test for multiple comparisons. The differences between different groups with the control were considered as significant if the p-values were $\leqslant 0.05$.

ACKNOWLEDGEMENTS

This research was partially funded by the Spanish Ministerio de Ciencia e Innovación (Agencia Estatal de Investigación, grant number PID2020-119424RB-I00 / AEI / 10.13039/501100011033), Plan France Relance by the French Ministère de l'Économie et des Finances, and received funding from Xerys Invest. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The results published here are in part based upon data collected from the TCGA Research Network: https://www.cancer.gov/tcga.

AUTHOR CONTRIBUTIONS

RRR carried out bioinformatics and statistical analyses. RP, XP, and JF designed, conducted, and interpreted enzymatic and biochemical assays. OP performed mass spectroscopy assays. AC conducted and analyzed the *in vitro* cytotoxicity screenings and functional assays. AC, PB, BJ and AK performed the *in vivo* studies and immunohistochemical (IHC) analyses. DC, AP, GF, and GM were responsible for the chemical synthesis and corresponding analytical studies. MP, RRR, RP, and AC interpreted data and prepared the figures. MP, JF, and RRR conceptualized, interpreted and supervised the study, with major contributions to the writing and editing of the manuscript.

DECLARATION OF INTERESTS

MP, GM, and IC are shareholders in Advanced BioDesign.

FIGURE LEGENDS

Figure 1. Data-based transcriptomic profiling of breast cancer cell lines and patient tumors. **(a)** Overview of data collection, detailing the number of human cell lines and patient samples analyzed. **(b-c)** Heatmaps showing the differential expression of

PAM50 genes and ALDH1 isoforms (*ALDH1A1*, *ALDH1A2*, and *ALDH1A3*) in cell lines and patient samples, respectively, with molecular subtypes represented as annotated rows to the left of the heatmaps. **(d)** Median bulk expression of 59 genes in breast cancer cell lines and patient samples, grouped by molecular classification. Dot size represents median absolute deviation (MAD), with color intensity reflecting $\log_2$ expression levels. **(e)** Spearman correlation of gene expression between groups in **(d)**, with significant correlation (p-value ≤ 0.01) marked by stars. A gray square highlights a strong positive correlation ($\rho = 0.71$) between Basal A (cell lines) and Basal-like (patient samples) subtypes.

Figure 2. Single-cell RNA sequencing reveals distinct ALDH1A isoform expression across tumor populations in breast cancer. (a,b) Expression patterns of *ALDH1A1, ALDH1A2*, and *ALDH1A3* across 20,485 cells from 16 Basal-like classified breast cancer cell lines, visualized using feature and violin plots. Classification into Basal A and B subgroups highlights intra-group heterogeneity and variability. **(c)** Single-cell analysis of 26 patient samples encompassing 100,064 cells, detailing the *ALDH1A* expression landscape. **(d)** Distribution of cell populations by molecular classification, illustrating cellular heterogeneity within tumors. **(e)** Feature plots display gene expression distribution and intensity for *ALDH1A1*, *ALDH1A2*, and *ALDH1A3* among individual cells (left). Corresponding pie charts depict comparative gene expression of each ALDH1A isoform, indicating the percentage of cells grouped by population and molecular subtype (right). Cell types analyzed include cancer-associated fibroblasts (CAFs), myofibroblasts (myCAFs), inflammatory CAFs (iCAFs), and mesenchymal stem cells (MSCs).

Figure 3. Novel chemical compounds targeting ALDH1A isoforms and their cytotoxic effects on breast cancer cell lines and patient-derived xenograft (PDX) models. (a) Schematic representation of the structural analogs of DIMATE and MATE developed for selective ALDH1A inhibition. **(b)** Circular stacked barplots showing the remaining activity (%) of compounds against five aldehyde dehydrogenase isoforms. ALDH1A1, ALDH1A2, and ALDH1A3 are represented in shades of green (dark to light), while ALDH2 and ALDH3A1 are depicted in shades of gray. The segment lengths indicate the proportion of remaining activity, providing a comparative view of compound efficacy. **(c)** Heatmap of cytotoxic activity (GI50 values) of the compounds, grouped by series number and their inhibitory effects on ALDH1 isoforms, evaluated across 18 breast cancer cell lines stratified by molecular subtypes. **(d)** Experimental design for 3D clonogenic assays using patient-derived cells from HER2-enriched and TNBC xenograft models (top). Violin plots display the GI50 distribution for DIMATE and ABD0171 in these models (bottom). **(e)** Linear mixed-effects model analysis showing predicted versus

observed cytotoxic effects (GI50 values) for DIMATE and ABD0171 in TNBC and HER2-enriched cell lines and PDX models.

Figure 4. Kinetic analyses and mass spectrometry validation of DIMATE and ABD0171 as ALDH1A inhibitors. (a) k_{inact}/K_i values calculated from the plots of k_{obs} *versus* concentrations of DIMATE and ABD0171 for ALDH1A1, ALDH1A2 and ALDH1A3 (inner plots). For DIMATE, tested concentration ranges were 0–500 μM for ALDH1A1, 0–200 μM for ALDH1A2, and 0–300 μM for ALDH1A3. For ABD0171, concentration ranges were 0–300 μM for ALDH1A1, 0–3 μM for ALDH1A2, and 0–200 μM for ALDH1A3. Enzymatic activities were measured at saturating substrate concentrations (250 μM hexanal) in the corresponding reaction buffer. Data are presented as mean ± standard deviation. **(b)** ESI-MS spectra of ALDH1A3 treated with DIMATE (left) and ABD0171 (right). Single-charged peaks (1+) were observed at *m/z* 61075–61119. Red signals represent ALDH1A3 treated with 167 μM DIMATE or ABD0171 (at an enzyme concentration of 16.7 μM), while black signals correspond to the untreated ALDH1A3 control. The spectra confirm irreversible binding of ALDH1A3 to DIMATE and ABD0171. Samples were prepared in 100 mM ammonium bicarbonate (pH 7.5) and pre-incubated for 45 min at 37°C.

Figure 5. **Effects of ALDH inhibitors ABD0171 and DIMATE on cell viability, invasion, migration and anoikis in triple-negative breast cancer (TNBC) cell lines. (a)** Flow cytometry analysis of early and late apoptosis in TNBC cells treated with ABD0171 or DIMATE for 12h, using Annexin V/PI staining. The lower-left quadrant indicates healthy cells, while the lower-right and upper-right quadrants show early and late apoptotic cells, respectively. **(b)** Protein analysis using the WES (Western enhanced sensitivity) capillary electrophoresis system, illustrating the activation of the extrinsic apoptotic pathway and the JNK (c-Jun N-terminal kinase)/ cJun signaling pathway in response to equimolar concentrations of ABD0171 and DIMATE. **(c)** Representative images from invasion and migration assays in MDAMB231 cells treated with ABD0171 or DIMATE for 24- 48h. Scatter plots (bottom) display quantification from two independent experiments (median ± standard deviation). Statistical analysis was performed using Kruskal-Wallis test. **(d)** Evaluation of anoikis resistance in MDAMB231 cells cultured on anchored and non-anchored (poly-HEMA-coated) surfaces and treated with either vehicle control, DIMATE or ABD0171 for 24h. Representative images of cells stained with calcein-AM and propidium iodide are shown, illustrating cell viability and death. The resazurin-based toxicity assay was used to quantify cell death, with results presented as median values ± standard deviation from triplicate experiments (bottom). Statistical significance was determined using an unpaired t-test with Welch's correction.

(e) WES capillary immunoelectrophoresis analysis of ALDH1A1, ALDH1A2 and ALDH1A3 protein expression levels in cells subjected non-anchored culture conditions for 10 days**. (f)** WES capillary immunoelectrophoresis analysis showing the activation STAT3 and associated proteins involved in its signalling pathway following 24h of treatment with ABD0171 or DIMATE. **(g)** Quantification of IL-6 and tPA protein levels by ELISA in TNBC cell lines, either untreated and treated with DIMATE and ABD0171 for 14h, under both anchored and non-anchored conditions.

Figure 6. ALDH inhibitors reduce tumor growth in immunocompromised and immunocompetent TNBC cancer models. (a, b) Assessment of survival and antitumor efficacy of DIMATE and ABD0171 in two murine TNBC models. SCID-CB17 mice bearing MDAMB231-GFP tumors and BALB/c mice with 4T1-Luc_M3 tumors were treated intraperitoneally with equimolar concentrations of DIMATE (20 mg/kg) or ABD0171 (40 mg/kg), administered three times per week for 3 weeks. Control groups were treated with compound's vehicle. Tumor growth inhibition is shown in the right panel, and overall survival rates are presented in the left panel. **(c, d, e)** Representative immunohistochemistry images for the indicated antibodies are provided. Quantification of positive cells was performed using QuPath. **(f)** Quantification of metastatic spread in the indicated organs, determined by luminescence intensity through bioluminescence imaging.

REFERENCES

1. Arnold, M. *et al.* Current and future burden of breast cancer: Global statistics for 2020 and 2040. *Breast* **66**, 15–23 (2022).
2. Pommier, R. M. *et al.* Comprehensive characterization of claudin-low breast tumors reflects the impact of the cell-of-origin on cancer evolution. *Nat Commun* **11**, (2020).
3. Perou, C. M. Molecular Stratification of Triple-Negative Breast Cancers. doi:10.1634/theoncologist.
4. Charles M. Perou *et al.* Molecular portraits of human breast tumours. *letters to nature* **406**, (2000).
5. Prat, A. *et al.* Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Research* **12**, (2010).
6. Obidiro, O., Battogtokh, G. & Akala, E. O. Triple Negative Breast Cancer Treatment Options and Limitations: Future Outlook. *Pharmaceutics* vol. 15 Preprint at (2023).
7. Prat, A. *et al.* Clinical implications of the intrinsic molecular subtypes of breast cancer. *Breast* **24**, S26–S35 (2015).
8. Koboldt, D. C. *et al.* Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).

9. Prat, A. *et al.* Molecular Characterization of Basal-Like and Non-Basal-Like Triple-Negative Breast Cancer. *Oncologist* **18**, 123–133 (2013).

10. Liu, Y. *et al.* ALDH1A1 expression correlates with clinicopathologic features and poor prognosis of breast cancer patients: A systematic review and meta-analysis. *BMC Cancer* **14**, (2014).

11. Croker, A. K. *et al.* Differential functional roles of ALDH1A1 and ALDH1A3 in mediating metastatic behavior and therapy resistance of human breast cancer cells. *Int J Mol Sci* **18**, (2017).

12. Charafe-Jauffret, E. *et al.* Aldehyde dehydrogenase 1-positive cancer stem cells mediate metastasis and poor clinical outcome in inflammatory breast cancer. *Clinical Cancer Research* **16**, 45–55 (2010).

13. Zhang, W., Wang, H., Qi, Y., Li, S. & Geng, C. Epigenetic study of early breast cancer (EBC) based on DNA methylation and gene integration analysis. *Sci Rep* **12**, (2022).

14. Pan, M. *et al.* Knockdown of ALDH1A3 reduces breast cancer stem cell marker CD44 via the miR-7-TGFBR2-Smad3-CD44 regulatory axis. *Exp Ther Med* **22**, (2021).

15. Yamashita, D. *et al.* Identification of ALDH1A3 as a viable therapeutic target in breast cancer metastasis-initiating cells. *Mol Cancer Ther* **19**, 1134–1147 (2020).

16. Bharadwaj, A. G. *et al.* ALDH1A3 promotes invasion and metastasis in triple-negative breast cancer by regulating the plasminogen activation pathway. *Mol Oncol* **18**, 91–112 (2024).

17. Marcato, P. *et al.* Aldehyde dehydrogenase 1A3 influences breast cancer progression via differential retinoic acid signaling. *Mol Oncol* **9**, 17–31 (2015).

18. Marcato, P. *et al.* Aldehyde dehydrogenase activity of breast cancer stem cells is primarily due to isoform ALDH1A3 and its expression is predictive of metastasis. *Stem Cells* **29**, 32–45 (2011).

19. Ginestier, C. *et al.* Retinoid signaling regulates breast cancer stem cell differentiation. *Cell Cycle* vol. 8 3297–3302 Preprint at (2009).

20. Liu, J. *et al.* An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell* **173**, 400-416.e11 (2018).

21. Ciriello, G. *et al.* Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. *Cell* **163**, 506–519 (2015).

22. Parker, J. S. *et al.* Supervised risk predictor of breast cancer based on intrinsic subtypes. *Journal of Clinical Oncology* **27**, 1160–1167 (2009).

23. Picornell, A. C. *et al.* Breast cancer PAM50 signature: Correlation and concordance between RNA-Seq and digital multiplexed gene expression technologies in a triple negative breast cancer series. *BMC Genomics* **20**, (2019).

24. Fougner, C., Bergholtz, H., Norum, J. H. & Sørlie, T. Re-definition of claudin-low as a breast cancer phenotype. *Nat Commun* **11**, (2020).

25. Tsherniak, A. *et al.* Defining a Cancer Dependency Map. *Cell* **170**, 564-576.e16 (2017).

26. Neve, R. M. *et al.* A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell* **10**, 515–527 (2006).

27. Yin, L., Duan, J. J., Bian, X. W. & Yu, S. C. Triple-negative breast cancer molecular subtyping and treatment progress. *Breast Cancer Research* vol. 22 Preprint at (2020).

28. Prat, A. & Perou, C. M. Deconstructing the molecular portraits of breast cancer. *Molecular Oncology* vol. 5 5–23 Preprint at (2011).

29. Gambardella, G. *et al.* A single-cell analysis of breast cancer cell lines to study tumour heterogeneity and drug response. *Nat Commun* **13**, (2022).

30. Wu, S. Z. *et al.* A single-cell and spatially resolved atlas of human breast cancers. *Nat Genet* **53**, 1334–1347 (2021).

31. Quash, G. *et al.* Novel competitive irreversible inhibitors of aldehyde dehydrogenase (ALDH1): restoration of chemosensitivity of L1210 cells overexpressing ALDH1 and induction of apoptosis in BAF 3 cells overexpressing bcl 2. *Biochem Pharmacol* **64**, 1279–1292 (2002).

32. Fournet, G., Quash, G. A., & Gore, J. Aminothiol ester compounds, pharmaceutical and cosmetics compositions containing same and uses thereof. U.S. Patent 7,078,402 B2. (2001).

33. Ceylan, I., Quash, G., Pérez-Alea, M., & Martin, G. Aminothiolester compounds or pharmaceutically acceptable salts thereof, for use in the treatment of cancer. European Patent EP3362056B1. (2016).

34. Pérez-Alea, M. *et al.* ALDH1A3 is epigenetically regulated during melanocyte transformation and is a target for melanoma treatment. *Oncogene* **36**, 5695–5708 (2017).

35. Rebollido-Rios, R. *et al.* Dual disruption of aldehyde dehydrogenases 1 and 3 promotes functional changes in the glutathione redox system and enhances chemosensitivity in non-small cell lung cancer. *Oncogene* **39**, 2756–2771 (2020).

36. Tomita, H., Tanaka, K., Tanaka, T. & Hara, A. Aldehyde dehydrogenase 1A1 in stem cells and cancer. *Oncotarget* **7**, (2016).

37. Butti, R., Gunasekaran, V. P., Kumar, T. V. S., Banerjee, P. & Kundu, G. C. Breast cancer stem cells: Biology and therapeutic implications. *Int J Biochem Cell Biol* **107**, 38–52 (2019).

38. Zhang, L., Chen, W., Liu, S. & Chen, C. Targeting Breast Cancer Stem Cells. *International Journal of Biological Sciences* vol. 19 552–570 Preprint at (2023).

39. Liu, S. *et al.* Breast cancer stem cells transition between epithelial and mesenchymal states reflective of their normal counterparts. *Stem Cell Reports* **2**, 78–91 (2014).

40. McLean, M. E. *et al.* The Expanding Role of Cancer Stem Cell Marker ALDH1A3 in Cancer and Beyond. *Cancers (Basel)* **15**, (2023).

41. Badve, S. *et al.* Basal-like and triple-negative breast cancers: A critical review with an emphasis on the implications for pathologists and oncologists. *Modern Pathology* **24**, 157–167 (2011).

42. Cheang, M. C. U. *et al.* Basal-like breast cancer defined by five biomarkers has superior prognostic value than triple-negative phenotype. *Clinical Cancer Research* **14**, 1368–1376 (2008).

43. Rakha, E. A. *et al.* Triple-negative breast cancer: Distinguishing between basal and nonbasal subtypes. *Clinical Cancer Research* **15**, 2302–2310 (2009).

44. Gray, G. K. *et al.* A human breast atlas integrating single-cell proteomics and transcriptomics. *Dev Cell* **57**, 1400-1420.e7 (2022).

45. Nielsen, T. O. *et al.* Immunohistochemical and Clinical Characterization of the Basal-Like Subtype of Invasive Breast Carcinoma. *Clinical Cancer Research* **10**, 5367–5374 (2004).

46. Liu, C. *et al.* ALDH1A1 Activity in Tumor-Initiating Cells Remodels Myeloid-Derived Suppressor Cells to Promote Breast Cancer Progression. *Cancer Res* **81**, 5919–5934 (2021).

47. US National Library of Medicine. ClinicalTrials.gov https://clinicaltrials.gov/study/NCT05601726.

48. Vasiliou, V. & Nebert, D. W. Analysis and update of the human aldehyde dehydrogenase (ALDH) gene family. *Hum Genomics* **2**, 138 (2005).

49. Dingler, F. A. *et al.* Two Aldehyde Clearance Systems Are Essential to Prevent Lethal Formaldehyde Accumulation in Mice and Humans. *Mol Cell* **80**, 996-1012.e9 (2020).

50. Koppaka, V. *et al.* ALDH3A1 plays a functional role in maintenance of corneal epithelial homeostasis. *PLoS One* **11**, (2016).

51. Wee, Z. N. *et al.* IRAK1 is a therapeutic target that drives breast cancer metastasis and resistance to paclitaxel. *Nat Commun* **6**, (2015).

52. Lin, Y. F. *et al.* High-level expression of ARID1A predicts a favourable outcome in triple-negative breast cancer patients receiving paclitaxel-based chemotherapy. *J Cell Mol Med* **22**, 2458–2468 (2018).

53. Balmer, J. E. & Blomhoff, R. Gene expression regulation by retinoic acid. *J Lipid Res* **43**, 1773–1808 (2002).

54. Tai, Y. L. *et al.* An EGFR/Src-dependent β4 integrin/FAK complex contributes to malignancy of breast cancer. *Sci Rep* **5**, (2015).

55. Ortiz, M. A. *et al.* Src family kinases, adaptor proteins and the actin cytoskeleton in epithelial-to-mesenchymal transition. *Cell Communication and Signaling* **19**, (2021).

56. Mezquita, B. *et al.* All-trans-retinoic acid activates the pro-invasive Src-YAP-Interleukin 6 axis in triple-negative MDA-MB-231 breast cancer cells while cerivastatin reverses this action. *Sci Rep* **8**, (2018).

57. Marotta, L. L. C. *et al.* The JAK2/STAT3 signaling pathway is required for growth of CD44 +CD24- stem cell-like breast cancer cells in human tumors. *Journal of Clinical Investigation* **121**, 2723–2735 (2011).

58. Jin, W. Role of JAK/STAT3 signaling in the regulation of metastasis, the transition of cancer stem cells, and chemoresistance of cancer by epithelial– mesenchymal transition. *Cells* **9**, (2020).

59. Ma, J. H., Qin, L. & Li, X. Role of STAT3 signaling pathway in breast cancer. *Cell Communication and Signaling* vol. 18 Preprint at https://doi.org/10.1186/s12964-020-0527-z (2020).

60. Qin, J. J., Yan, L., Zhang, J. & Zhang, W. D. STAT3 as a potential therapeutic target in triple negative breast cancer: A systematic review. *Journal of Experimental and Clinical Cancer Research* vol. 38 Preprint at (2019).

61. Zhang, G., Hou, S., Li, S., Wang, Y. & Cui, W. Role of STAT3 in cancer cell epithelial-mesenchymal transition (Review). *International Journal of Oncology* vol. 64 Preprint at (2024).

62. 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 (2009).

63. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).

64. Ahlmann-Eltze, C. & Huber, W. glmGamPoi: Fitting Gamma-Poisson generalized linear models on single cell count data. *Bioinformatics* **36**, 5701–5702 (2020).

65. Hadley Wickham. *Ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York, 2016).

66. Schloerke, B. *et al. GGally: Extension to 'Ggplot2'*. (2024).

67. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).

68. Quash, G. et al., European Journal of Medicinal Chemistry 43, 906–916. (2008).

69. Ceylan, I., Martin, G., Pérez, M., & Berrou, A. Aminothiolester compounds and uses thereof. European Patent EP3833657B1. (2020).

70. Pequerul, R., Porté, S., Parés, X., Pérez-Alea, M. & Farrés, J. Production, Purification, and Fluorometric Activity Assay of Human Aldehyde Dehydrogenases. *Bio Protoc* **12**, (2022).

71. Pequerul, R. *et al.* Structural and kinetic features of aldehyde dehydrogenase 1A (ALDH1A) subfamily members, cancer stem cell markers active in retinoic acid biosynthesis. *Arch Biochem Biophys* **681**, (2020).

72. Copeland, R. A. *EVALUATION OF ENZYME INHIBITORS IN DRUG DISCOVERY A Guide for Medicinal Chemists and Pharmacologists Second Edition*. (A John Wiley & Sons, INC., PUBLICATION, 2013).

73. Gallego, O. *et al.* Comparative functional analysis of human medium-chain dehydrogenases, short-chain dehydrogenases/reductases and aldo-keto reductases with retinoids. *Biochemical Journal* **399**, 101–109 (2006).

74. McCaffery, P. *et al.* Retinoid quantification by HPLC/MSn. *J Lipid Res* **43**, 1143–1149 (2002).

75. Evans James E. and McCaffery, P. HPLC/MSN Analysis of Retinoids. in *Retinoids: Methods and Protocols* (ed. Sun Hui and Travis, G. H.) 149–162 (Humana Press, Totowa, NJ, 2010).

76. Kane, M. A., Chen, N., Sparks, S. & Napoli, J. L. *Quantification of Endogenous Retinoic Acid in Limited Biological Samples by LC/MS/MS*. *Biochem. J* vol. 388 (2005).

77. Landis, J., An, H., Bailey, A., Dittmer, D. & Marron, J. *Dr4pl: Dose Response Data Analysis Using the 4 Parameter Logistic (4pl) Model*. vol. R package version 2 (2021).

78. Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. lmerTest Package: Tests in Linear Mixed Effects Models. *J Stat Softw* **82**, 1–26 (2017).

79. Nguyen, U., Squaglia, N., Boge, A. & Fung, P. A. The Simple Western™: a gel-free, blot-free, hands-free Western blotting reinvention. *Nat Methods* **8**, v–vi (2011).

80. Hulkower, K. I. & Gehler, S. R. How to Measure Area Closure Using the Oris™ Pro Cell Migration Assay. in *Platypus Technologies, LLC* (2016).

81. Justus, C. R., Leffler, N., Ruiz-Echevarria, M. & Yang, L. V. In vitro cell migration and invasion assays. *Journal of Visualized Experiments* (2014).

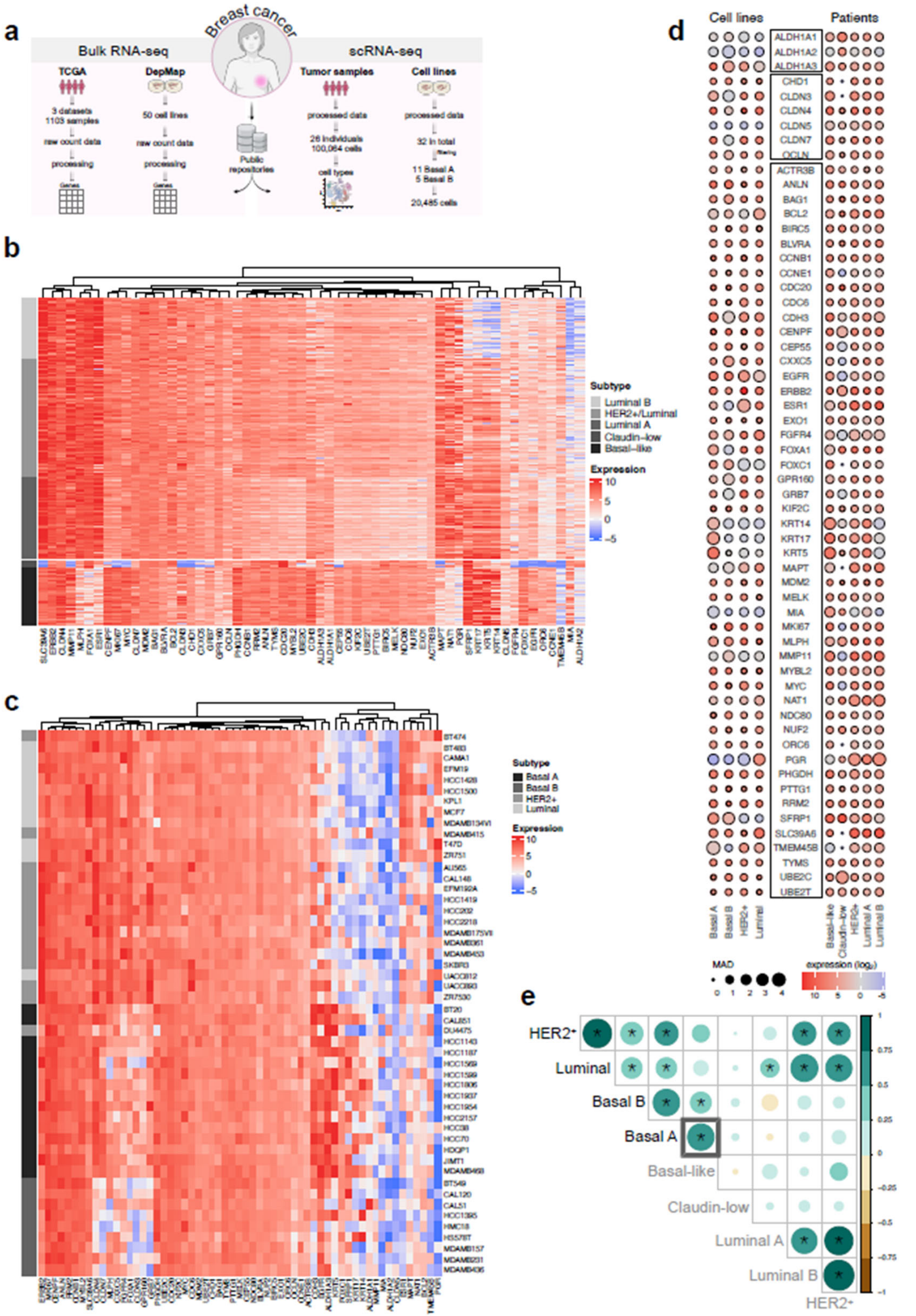

a
Breast cancer
Bulk RNA-seq
scRNA-seq
TCGA
DepMap
Tumor samples
Cell lines
3 datasets
1103 samples
raw count data
processing
Genes
50 cell lines
raw count data
processing
Genes
Public repositories
processed data
26 individuals
100,064 cells
cell types
processed data
32 in total
11 Basal A
5 Basal B
20,485 cells
b
Subtype
Luminal B
HER2+/Luminal
Luminal A
Claudin-low
Basal-like
Expression
10
5
0
−5
c
Subtype
Basal A
Basal B
HER2+
Luminal
Expression
10
5
0
−5
d
Cell lines
Patients
ALDH1A1
ALDH1A2
ALDH1A3
CHD1
CLDN3
CLDN4
CLDN5
CLDN7
OCLN
ACTR3B
ANLN
BAG1
BCL2
BIRC5
BLVRA
CCNB1
CCNE1
CDC20
CDC6
CDH3
CENPF
CEP55
CXXC5
EGFR
ERBB2
ESR1
EXO1
FGFR4
FOXA1
FOXC1
GPR160
GRB7
KIF2C
KRT14
KRT17
KRT5
MAPT
MDM2
MELK
MIA
MKI67
MLPH
MMP11
MYBL2
MYC
NAT1
NDC80
NUF2
ORC6
PGR
PHGDH
PTTG1
RRM2
SFRP1
SLC39A6
TMEM45B
TYMS
UBE2C
UBE2T
Basal A
Basal B
HER2+
Luminal
Basal-like
Claudin-low
HER2+
Luminal A
Luminal B
MAD
0
1
2
3
4
expression (log2)
10
5
0
−5
e
HER2+
Luminal
Basal B
Basal A
Basal-like
Claudin-low
Luminal A
Luminal B
HER2+
1
0.75
0.5
0.25
0
−0.25
−0.5
−0.75
−1

Figure 1

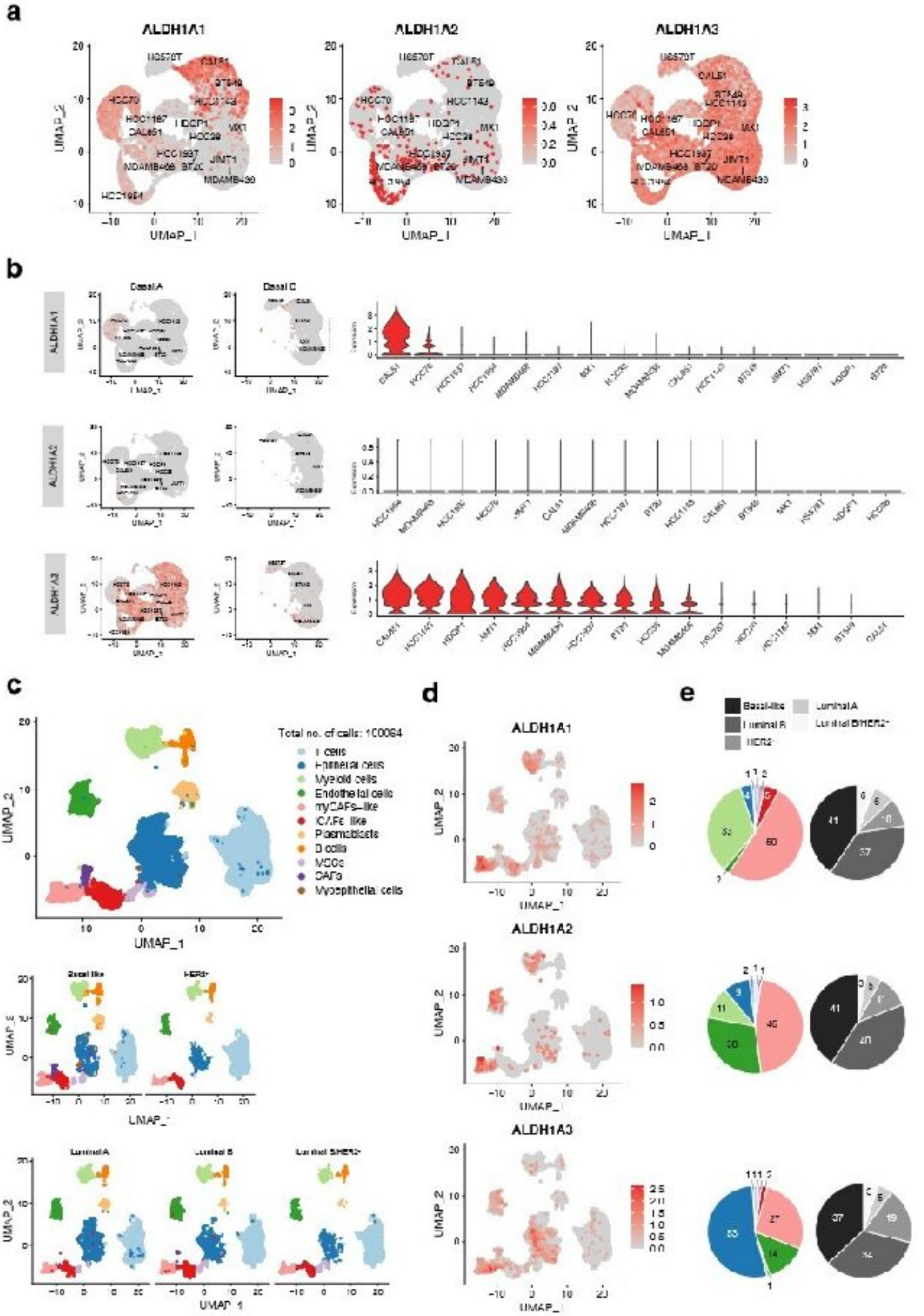

a
ALDH1A1
ALDH1A2
ALDH1A3
UMAP_1
UMAP_2
b
Basal A
Basal B
ALDH1A1
ALDH1A2
ALDH1A3
c
Total no. of cells: 100064
T cells
Epithelial cells
Myeloid cells
Endothelial cells
myCAFs-like
iCAFs-like
Plasmablasts
B cells
MSCs
CAFs
Myoepithelial cells
Basal like
HER2+
Luminal A
Luminal B
Luminal SHER2+
d
ALDH1A1
ALDH1A2
ALDH1A3
e
Basal-like
Luminal A
Luminal B
Luminal SHER2+
HER2

Figure 2

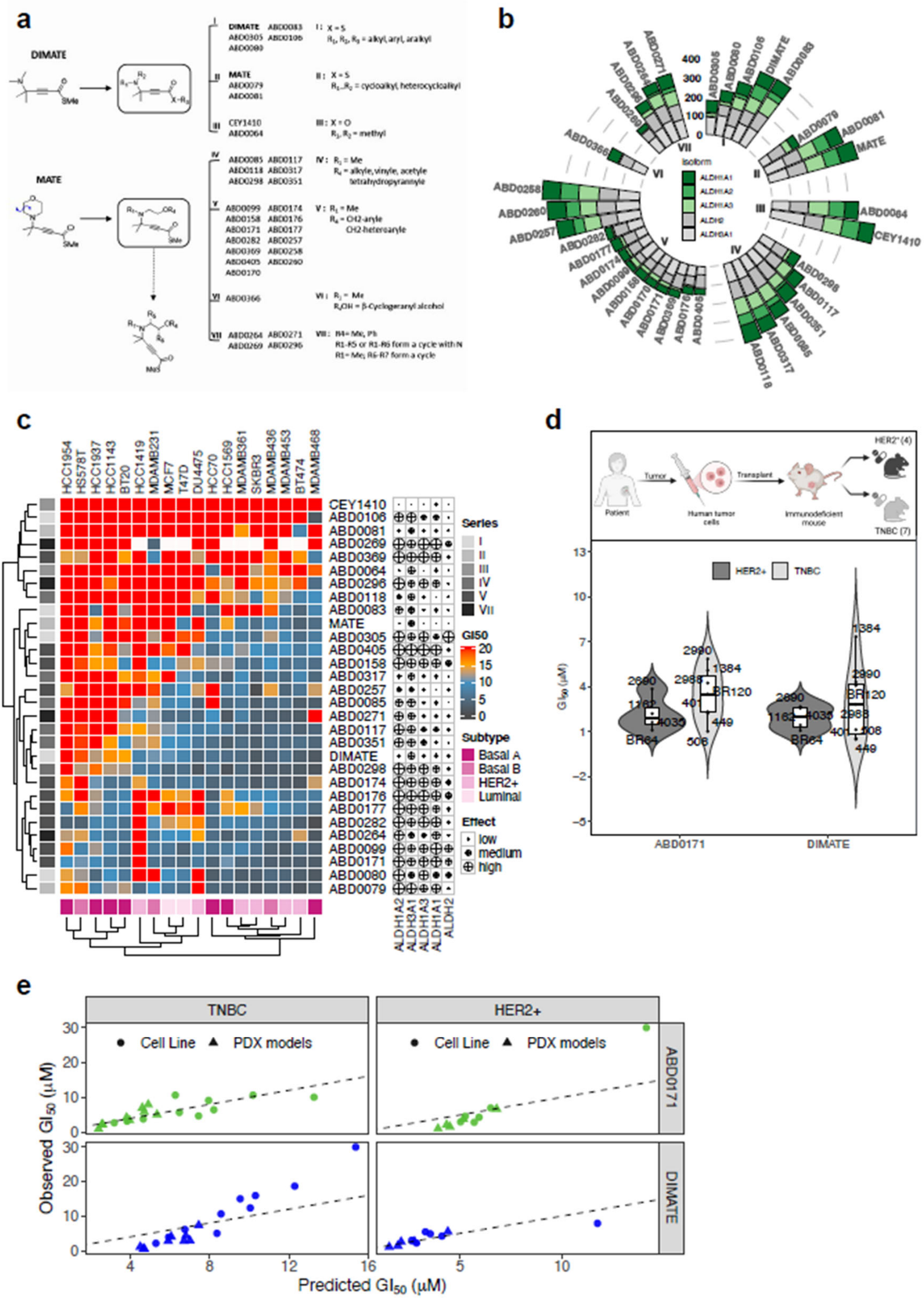

a
DIMATE
MATE
b
Isoform
ALDH1A1
ALDH1A2
ALDH1A3
ALDH2
ALDH3A1
c
Series
Subtype
Basal A
Basal B
HER2+
Luminal
Effect
low
medium
high
GI50
d
Patient
Tumor
Human tumor cells
Transplant
Immunodeficient mouse
HER2+ (4)
TNBC (7)
HER2+
TNBC
ABD0171
DIMATE
e
TNBC
HER2+
Cell Line
PDX models
ABD0171
DIMATE
Observed GI50 (μM)
Predicted GI50 (μM)

Figure 3

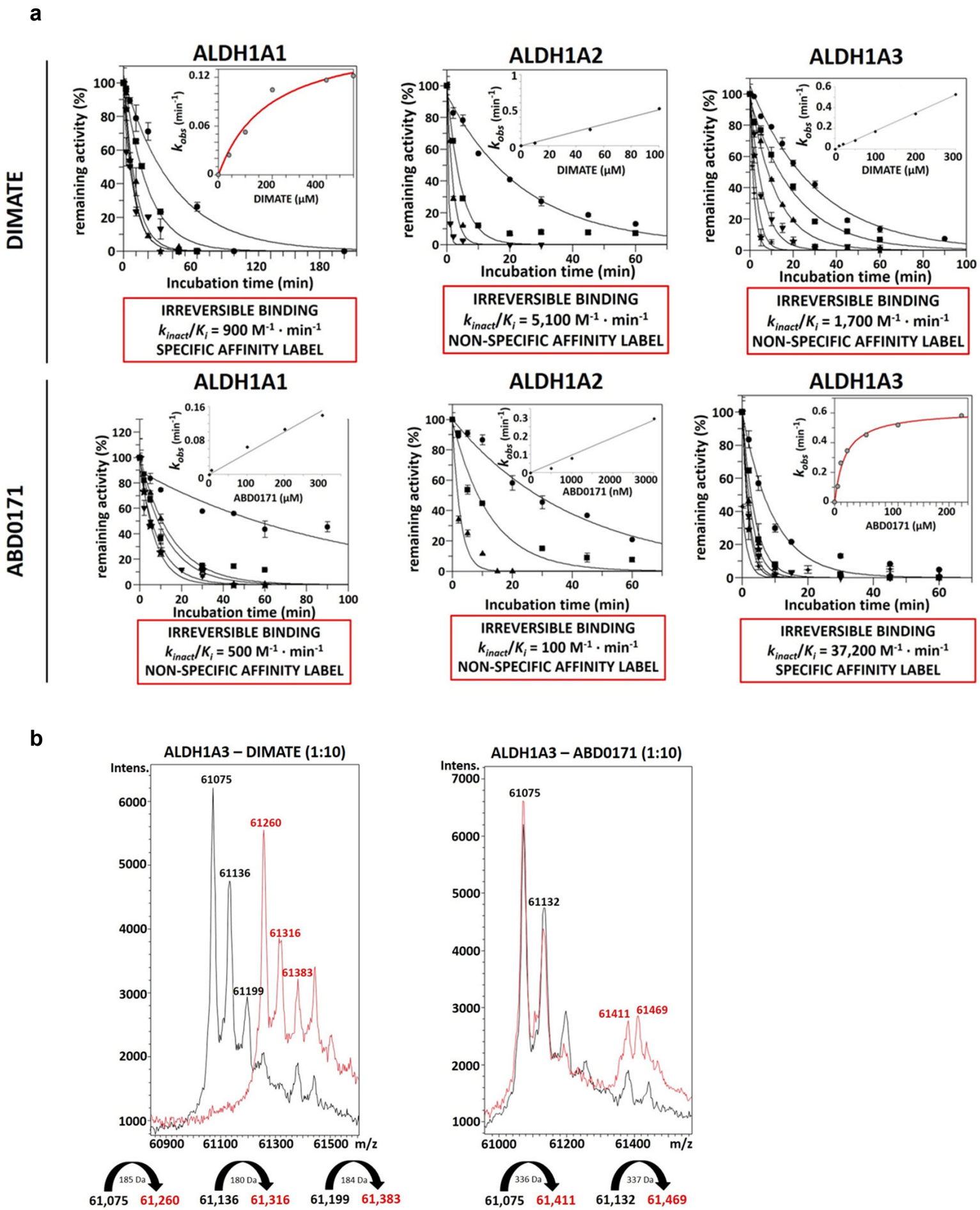

a
DIMATE
ABD0171
ALDH1A1
ALDH1A2
ALDH1A3
remaining activity (%)
Incubation time (min)
k_{obs} (min^{-1})
DIMATE (μM)
ABD0171 (μM)
ABD0171 (nM)
IRREVERSIBLE BINDING
k_{inact}/K_i = 900 M^{-1} · min^{-1}
SPECIFIC AFFINITY LABEL
IRREVERSIBLE BINDING
k_{inact}/K_i = 5,100 M^{-1} · min^{-1}
NON-SPECIFIC AFFINITY LABEL
IRREVERSIBLE BINDING
k_{inact}/K_i = 1,700 M^{-1} · min^{-1}
NON-SPECIFIC AFFINITY LABEL
IRREVERSIBLE BINDING
k_{inact}/K_i = 500 M^{-1} · min^{-1}
NON-SPECIFIC AFFINITY LABEL
IRREVERSIBLE BINDING
k_{inact}/K_i = 100 M^{-1} · min^{-1}
NON-SPECIFIC AFFINITY LABEL
IRREVERSIBLE BINDING
k_{inact}/K_i = 37,200 M^{-1} · min^{-1}
SPECIFIC AFFINITY LABEL
b
ALDH1A3 – DIMATE (1:10)
ALDH1A3 – ABD0171 (1:10)
Intens.
61075
61136
61199
61260
61316
61383
61132
61411
61469
m/z
185 Da
180 Da
184 Da
336 Da
337 Da
61,075 61,260
61,136 61,316
61,199 61,383
61,075 61,411
61,132 61,469

Figure 4

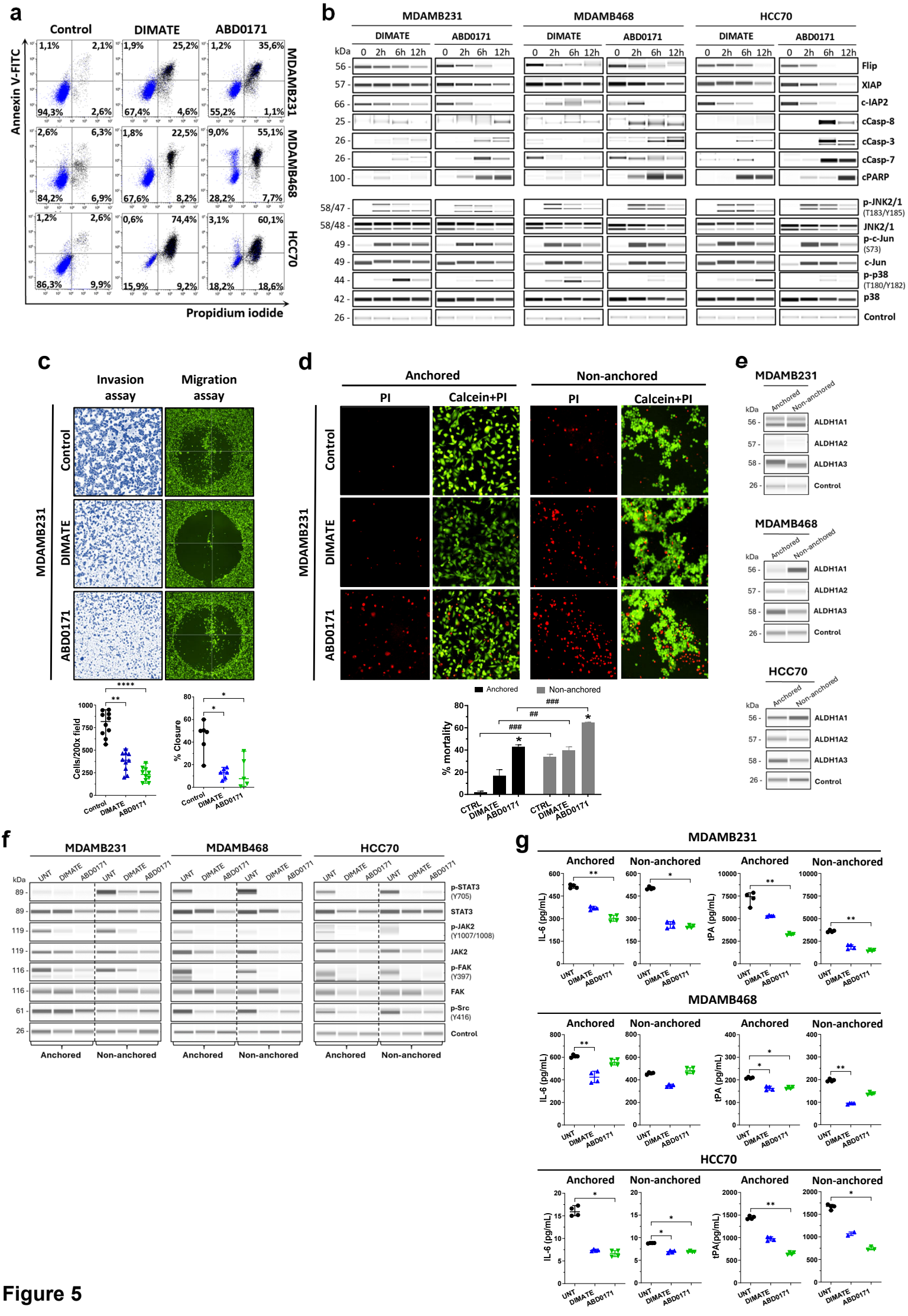

a
Control
DIMATE
ABD0171
Annexin V-FITC
Propidium iodide
MDAMB231
MDAMB468
HCC70
b
MDAMB231
MDAMB468
HCC70
DIMATE
ABD0171
kDa
0 2h 6h 12h
Flip
XIAP
c-IAP2
cCasp-8
cCasp-3
cCasp-7
cPARP
p-JNK2/1 (T183/Y185)
JNK2/1
p-c-Jun (S73)
c-Jun
p-p38 (T180/Y182)
p38
Control
c
Invasion assay
Migration assay
Control
DIMATE
ABD0171
Cells/200x field
% Closure
d
Anchored
Non-anchored
PI
Calcein+PI
% mortality
CTRL
e
ALDH1A1
ALDH1A2
ALDH1A3
f
p-STAT3 (Y705)
STAT3
p-JAK2 (Y1007/1008)
JAK2
p-FAK (Y397)
FAK
p-Src (Y416)
UNT
g
IL-6 (pg/mL)
tPA (pg/mL)

Figure 5

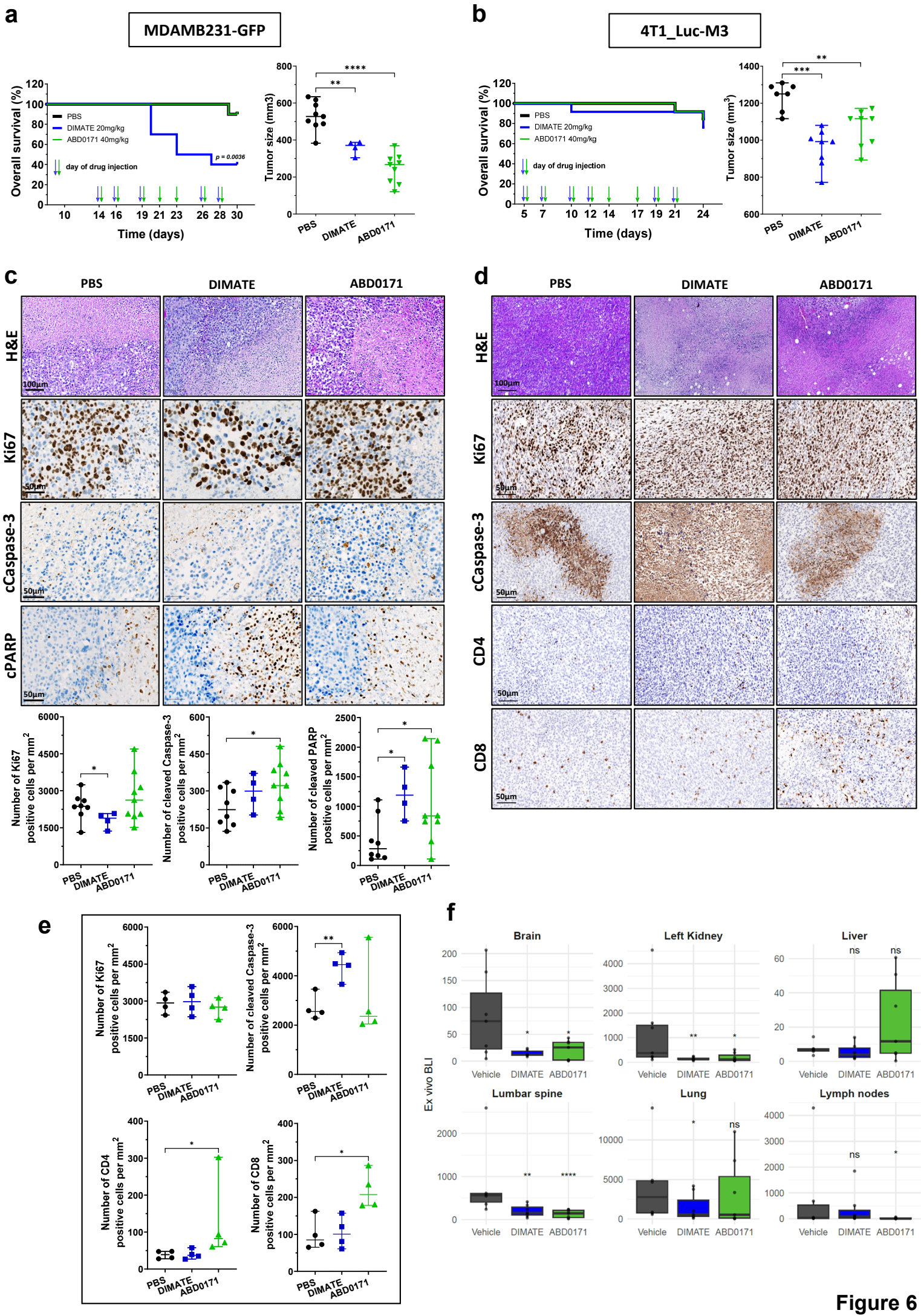

a
MDAMB231-GFP
Overall survival (%)
PBS
DIMATE 20mg/kg
ABD0171 40mg/kg
day of drug injection
p = 0.0036
Time (days)
Tumor size (mm3)
PBS
DIMATE
ABD0171
b
4T1_Luc-M3
Overall survival (%)
PBS
DIMATE 20mg/kg
ABD0171 40mg/kg
day of drug injection
Time (days)
Tumor size (mm3)
PBS
DIMATE
ABD0171
c
PBS
DIMATE
ABD0171
H&E
Ki67
cCaspase-3
cPARP
100µm
50µm
Number of Ki67 positive cells per mm2
Number of cleaved Caspase-3 positive cells per mm2
Number of cleaved PARP positive cells per mm2
d
PBS
DIMATE
ABD0171
H&E
Ki67
cCaspase-3
CD4
CD8
100µm
50µm
e
Number of Ki67 positive cells per mm2
Number of cleaved Caspase-3 positive cells per mm2
Number of CD4 positive cells per mm2
Number of CD8 positive cells per mm2
PBS
DIMATE
ABD0171
f
Brain
Left Kidney
Liver
Lumbar spine
Lung
Lymph nodes
Ex vivo BLI
Vehicle
DIMATE
ABD0171
ns

Figure 6

TABLE 1. IC_{50} (μM) and k_{inact}/K_i ($M^{-1} \cdot min^{-1}$) values of DIMATE and ABD0171 for ALDH1A1, ALDH1A2 and ALDH1A3 using hexanal and all-*trans*-retinaldehyde as substrates.

	ALDH1A1			ALDH1A2			ALDH1A3			
	atRAL	HEXANAL		atRAL	HEXANAL		atRAL		HEXANAL	
COMPOUND	**IC_{50} (μM)**	**IC_{50} (μM)**	**k_{inact}/K_i ($M^{-1}\cdot min^{-1}$)**	**IC_{50} (μM)**	**IC_{50} (μM)**	**k_{inact}/K_i ($M^{-1}\cdot min^{-1}$)**	**IC_{50} (μM)**	**k_{inact}/K_i ($M^{-1}\cdot min^{-1}$)**	**IC_{50} (μM)**	**k_{inact}/K_i ($M^{-1}\cdot min^{-1}$)**
DIMATE	25 ± 2	37 ± 5	900	6.4 ± 0.9	18 ± 4	5,100	5.7 ± 0.5	80,100	20 ± 2	1,700
ABD0171	6.6 ± 1	3.8 ± 1.1	500	0.34 ± 0.03	0.57 ± 0.1	100	0.79 ± 0.2	115,400	3.4 ± 0.1	37,200

IC_{50} values were obtained fluorometrically and by an HPLC-based method when hexanal and all-*trans*-retinaldehyde (atRAL) were used as substrates, respectively. The reaction mixture was incubated in the corresponding reaction buffer[71] for 20 min at 37ºC, before adding NAD^+ as cofactor, NADH as internal standard and the corresponding substrate. k_{inact}/K_i values and the affinity labelling type were obtained by k_{obs} *vs* [I] plot. Experimental values are the mean ± standard error.