ALDH1 subtype-specific inhibitor targets key cell populations in triple negative 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], 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, Cologne, Germany

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

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

ABSTRACT

Breast cancer exhibits significant heterogeneity, making the identification of molecular targets critical for the development of effective therapies. In this study, we investigate the expression patterns of the ALDH1A enzyme subfamily- ALDH1A1, ALDH1A2, and ALDH1A3- across diverse molecular subtypes and the tumor microenvironment, with a particular focus on triple-negative breast cancer (TNBC). Notably, ALDH1A3 is predominantly expressed in basal-like tumors and is significantly enriched in epithelial cells. In contrast, ALDH1A1 is uniquely expressed in myeloid cells and cancer-associated fibroblasts (CAFs). This dual expression pattern underscores the critical roles of ALDH1A3 and ALDH1A1 within the tumor microenvironment. We also evaluated the activity of novel ALDH1A inhibitors, specifically ABD0171, which demonstrated selective cytotoxicity in ALDH1A3-expressing basal-like cells and induced apoptosis through activation of the death receptor pathway. Both ABD0171 and its precursor, DIMATE, exhibited potent anti-metastatic effects, likely via inhibition of IL6/JAK/STAT3 signalling, and significantly inhibited tumor growth in preclinical TNBC models. These findings underscore the therapeutic potential of targeting ALDH1A isoforms, particularly ALDH1A3, in aggressive breast cancer subtypes and highlight the importance of considering tumor heterogeneity in the development of targeted therapies.

INTRODUCTION

Breast cancer remains the most common malignancy among women worldwide, with an estimated 2.3 million new cases and 685,000 deaths reported in 2020. This disease is characterized by considerable heterogeneity, both molecularly and clinically, which complicates treatment strategies. Based on the expression of estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2), breast cancer is classified into four major subtypes: luminal A (ER+, PR+, HER2−), luminal B (ER+, PR−/HER2+), HER2-positive, and triple-negative breast cancer (TNBC), which lacks expression of all three receptors. Among these, TNBC is particularly aggressive and lacks effective targeted therapies, making it a critical focus for research and therapeutic development.

The aldehyde dehydrogenase 1 (ALDH1) family of enzymes has gained attention as a key player in cancer biology, particularly in tumor initiation, progression, and resistance to therapy. Within this family, the ALDH1A subfamily- comprising ALDH1A1, ALDH1A2, and ALDH1A3- has been implicated in maintaining cancer stem cell properties and

contributing to chemoresistance. Among these isoforms, ALDH1A3 has been strongly associated with aggressive cancer phenotypes and poor clinical outcomes, particularly in TNBC [1–6]. Despite growing evidence supporting the role of ALDH1A isoforms as important biomarkers and therapeutic targets, their specific expression patterns and functional roles across different molecular subtypes and within the tumor microenvironment remain poorly understood. This gap has limited the potential for developing selective inhibitors to improve therapeutic efficacy, particularly in difficult-to-treat subtypes like TNBC.

In this study, we sought to elucidate the expression patterns of ALDH1A isoforms across diverse breast cancer subtypes and within the tumor microenvironment, while also evaluating the therapeutic potential of novel ALDH1A inhibitors, specifically ABD0171 and its precursor, DIMATE. DIMATE (4-dimethylamino-4-methyl-pent-2-ynthioic acid-S-methylester) is the lead compound of a series of α,β-acetylenic amino thiolesters and is considered an irreversible, competitive inhibitor of ALDH1 and ALDH3 subfamilies. It has demonstrated tumor growth inhibition *in vitro* and *in vivo* in leukemia and melanoma models with low toxicity to healthy cells. DIMATE has also shown efficacy in non-small cell lung cancer (NSCLC) models, either alone or in combination with platinum-based chemotherapies, by inducing oxidative stress through the accumulation of aldehyde-protein adducts, without causing significant toxicity in animal models. A multicenter Phase I clinical trial (NCT05601726) is currently investigating a lipidic nanoformulation of DIMATE in acute myeloid leukemia (AML) patients.

In the present study, we developed and enzymatically characterized novel analogues of DIMATE, including ABD0171, and evaluated their inhibition profiles against multiple ALDH isoforms. We explored their cytotoxic effects in 18 breast cancer cell lines representing different molecular subtypes. Furthermore, we performed extensive data mining on 59 genes, including ALDH1A1, ALDH1A2, and ALDH1A3, using the Cancer Cell Line Encyclopedia (CCLE) and The Cancer Genome Atlas (TCGA) to gain deeper insights into the ALDH1A subfamily's role in breast cancer. The study also focused specifically on TNBC cell lines (MDAMB231, MDAMB468, and HCC70) to assess the impact of DIMATE and ABD0171 on key molecular pathways associated with stress response, proliferation, and metastatic activity. Finally, we demonstrated the antitumor and anti-metastatic effects of both compounds *in vivo* using TNBC xenograft and syngeneic mouse models.

RESULTS

ALDH1A subfamily exhibits specific expression patterns across molecular subtypes and cell subpopulations in breast tumors

We preprocessed and integrated three independent RNA-seq datasets from The Cancer Genome Atlas (TCGA)[7,8] 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[9,10] claudin-low associated-genes[11] 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)[12] portal and performed the subsequent preprocessing steps. Unsupervised clustering of these cell lines grouped them into four recognizable molecular subgroups consistent with known breast cell line classifications[13]: 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 differential 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 (all adjusted *p*-values < 0.001, Fig.1d and Supplementary Data 1C). This result was consistent with ALDH1A3 protein expression profiles (Supplementary Fig. 1a) and corroborated previous studies linking high ALDH1A3 expression to TNBC[4,5], which significantly overlaps with the basal-like subtype[14]. 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[15]. We observed a strong positive correlation (r= 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 (all adjusted

p-values < 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.

Interestingly, the specific subset of patients with claudin-low tumors exhibited significant expression 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[16] 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 based on 32 breast cancer cell lines[17] while another based on primary tumor biopsies from 26 patients including a total of 100,064 cells[18].

The earlier study, which included a total of 32 breast cancer cell lines, revealed that basal phenotype cell lines form a distinct cluster, separate from the luminal and HER2-enriched groups[17]. 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.

We further explored the heterogeneity of the ALDH1A subfamily across breast cancer subtypes by analyzing the patient samples[18]. 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 and inflammatory CAF-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+ 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 enhanced inhibitory activity towards ALDH1A show selective 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[19]. 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[19]. 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) (WO2001092220 and WO2017064247 patents). 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 2). The most effective compounds against ALDH1A1, ALDH1A2 and ALDH1A3 activity belonged to series V of analogs, characterized by an (arylmethoxy)ethylamino substituent 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, in particular, exhibited exceptional potency, with GI_{50} values consistently below 10 μM in 95% of tested cell lines. 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 tend to express higher 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 have 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 that warrants further mechanistic investigation.

ABD0171 is a potent covalent 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, ALDH3A1 and ALDH2 isoforms, using both routine (hexanal) and physiological (all-*trans*-retinaldehyde) substrates (Table 1 and Supplementary Table S1). Our 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. Both compounds exhibited time-dependent inhibition of all ALDH isoforms (Fig. 4a), except for ALDH3A1 (Supplementary Fig. 2a), suggesting that DIMATE and ABD0171 act as covalent irreversible inhibitors for all the other isoforms. The IC_{50} values ranged from 0.34 to 303 μM after a 20-min preincubation with the inhibitors (Table 1 and Supplementary Table S1). Our 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 (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 covalent 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 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 spectrum confirmed covalent 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 the IL6/JAK/STAT3 signaling pathway.

Previous reports showed that DIMATE was able to induce cell death in p53-mutated cells[20,21]. 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). Similarly, 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. Stress activation of p38 was transient, with DIMATE-treated cells showing greater prominence.

Both, extrinsic apoptosis and JNK pathway activation can influence 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, c). 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, c). At this time point, we noted robust activation of the Src- STAT3 signaling pathway in all cell lines under non-anchored conditions, with pro-survival AKT and ERK remaining unaffected (Fig. 5e). The observation that ABD0171 and DIMATE reduce anoikis resistance in these cells suggested that pharmacological inhibition of ALDH1A could disrupt oncogenic reprogramming mediated by STAT3, triggered by detachment from the extracellular cell matrix. Our analysis of the molecular response to ABD0171 and DIMATE revealed simultaneous inhibition of FAK and JAK/STAT3 signaling pathways in both anchored and non-anchored cell conditions, with ABD0171 showing a markedly stronger effect (Fig. 5e).

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. 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 or 15 mg/kg of DIMATE caused body weight loss (Supplementary Data 3E). or produced any 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. In addition, 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, accompanied by a moderate increase in tCO_2 levels ($p=0.003$) compared to controls, remaining within a non-critical range.

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. Contrarily, DIMATE administration caused acute toxicity in 6 animals from the immunosuppressed MDAMB231-GFP animals towards the end of the study (Supplementary Fig. 4b, c and Supplementary Data 3G). Both compounds resulted in significant inhibition of tumor growth. At day 30 of 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% after 24 days of treatment ($p = 0.0430$, $p = 0.0004$), despite the higher aggressiveness of these TNBC cells (Fig. 6a, b and Supplementary Data 3H), which are also known for multiorgan metastasis. Notably, treatment with ABD0171 and DIMATE significantly reduced metastasis to lymph nodes, brain, lumbar spine, and kidneys in this model (Fig. 6c and Supplementary Data 3I), reinforcing the antimetastatic activity observed in the functional *in vitro* studies.

DISCUSSION

The results of our study provide new insights into the expression patterns of the ALDH1A subfamily across molecular subtypes and cellular subpopulations of breast cancer, with a particular focus on the potential of ALDH1A isoforms as therapeutic targets. The identification of distinct expression profiles of ALDH1A1, ALDH1A2, and ALDH1A3 across both patient-derived tumors and breast cancer cell lines highlights the complexity of the tumor microenvironment and the diverse roles these enzymes play in different cancer contexts.

One of the key findings of this study is the predominant expression of ALDH1A3 in basal-like breast cancer, as well as its correlation with the Basal A cell line subtype. This observation is consistent with prior studies linking ALDH1A3 to aggressive cancer phenotypes and poor prognosis, particularly in TNBC, a classification with high overlap with basal-like molecular subtypes, for which effective targeted therapies are limited. The high expression of ALDH1A3 in both basal-like tumors and Basal A cell lines further supports the relevance of this enzyme as a potential biomarker and therapeutic target in TNBC. Additionally, our study reveals that while ALDH1A1 is expressed across all breast cancer subtypes, its expression in patient samples may originate from diverse cell types within the tumor microenvironment, such as stromal and immune cells, which are not fully represented in traditional cell line models. This finding emphasizes the importance of considering the tumor microenvironment when assessing ALDH1A expression and function.

The claudin-low breast cancer subtype presented a unique expression profile, with significant expression of both ALDH1A1 and ALDH1A2. These tumors, known for their stromal richness and mesenchymal characteristics, are difficult to model *in vitro*, yet our identification of three cell lines that resemble claudin-low tumors provides a valuable resource for further study. The high expression of ALDH1A1 and ALDH1A2 in these tumors suggests that these isoforms may play distinct roles in this challenging subtype, possibly contributing to the mesenchymal phenotype and drug resistance observed in claudin-low tumors.

In our exploration of novel ALDH1A inhibitors, we demonstrated the potent and selective activity of ABD0171, particularly against ALDH1A3-expressing basal-like cells. The structure-activity relationship analysis of our compound library revealed that the aryl substitution in series V analogs is crucial for achieving enhanced inhibitory activity. ABD0171 exhibited exceptional potency, with GI50 values below 10 μM in most basal-A and basal-B cell lines, underscoring its potential as a therapeutic agent for TNBC and other aggressive breast cancer subtypes. The observed cytotoxicity of ABD0171 is consistent with the known role of ALDH1A3 in maintaining the self-renewal capacity of cancer stem cells. These findings align with previous reports that have implicated ALDH1A3 in cancer stem cell biology and drug resistance, making it an attractive target for therapeutic intervention.

Interestingly, while ABD0171 showed significant cytotoxic activity in basal-like and HER2-enriched subtypes, its impact on HER2-enriched PDX models was somewhat variable. This suggests that the efficacy of ALDH1A-targeting compounds may be modulated by factors beyond ALDH1A inhibition alone, such as other molecular mechanisms or differential expression of other ALDH isoforms. The possibility of additional resistance mechanisms in HER2-enriched models, particularly those co-expressing ALDH1A1 and ALDH1A3, warrants further investigation to elucidate the precise role of these enzymes in HER2-enriched breast cancer and to optimize the therapeutic potential of ALDH inhibitors in this context. The differential performance of DIMATE and ABD0171 in TNBC and HER2-enriched models also points to the need for further mechanistic studies to understand the factors driving variability in drug response. While both compounds demonstrated strong predictive performance in TNBC models, the variability observed in DIMATE-treated HER2-enriched models suggests the involvement of additional regulatory pathways that modulate ALDH1A activity or drug sensitivity in this subtype.

In addition to this molecular characterization, ABD0171 and DIMATE were shown to induce apoptosis and impair the metastatic potential of TNBC cells, particularly in basal-like subtypes. ABD0171 induced death receptor-mediated extrinsic apoptosis and sensitized basal subtype cells to anoikis by inhibiting the IL6/JAK/STAT3 signaling pathway. Both compounds induced apoptosis across TNBC cell lines, as evidenced by the reduction in c-FLIP and cellular IAPs (cIAP2, XIAP), and activation of caspase-8, caspase-3, and cPARP. ABD0171's effect was more robust, accompanied by sustained activation of the oxidative stress-sensing JNK pathway, while DIMATE triggered more transient p38 activation. These findings suggest ABD0171 may provide a stronger and

more sustained apoptotic response compared to DIMATE, particularly in basal-like TNBC cells.

Both inhibitors also exhibited potent anti-metastatic effects. At sub-lethal concentrations, they attenuated migration and invasion in TNBC cells and diminished anoikis resistance by disrupting FAK and JAK/STAT3 signaling pathways. ABD0171 exerted a more pronounced effect on these survival pathways, reinforcing its superior anti-metastatic activity compared to DIMATE.

In vivo, both ABD0171 and DIMATE demonstrated significant tumor growth inhibition in MDA-MB-231 xenografts and 4T1-Luc_M3 models of TNBC. ABD0171 showed greater efficacy in reducing tumor burden and metastasis to distant organs such as lymph nodes, brain, lumbar spine, and kidneys. These results corroborated the antimetastatic activity observed in vitro. Both compounds were well tolerated in preclinical models, exhibiting favorable safety profiles, with minimal systemic, hepatic, or renal toxicity.

Overall, our study highlights the therapeutic promise of targeting ALDH1A3, particularly in basal-like or TNBC subtypes in general. The selective cytotoxicity of ABD0171 and its potent inhibitory activity against ALDH1A isoforms provide a compelling rationale for further preclinical development. Future studies should focus on exploring combination strategies with existing therapies to overcome resistance mechanisms. Additionally, our findings underscore the importance of considering tumor heterogeneity and the tumor microenvironment when developing targeted therapies for breast cancer, as ALDH1A isoforms play distinct roles in different cellular contexts.

In conclusion, the specific expression patterns of ALDH1A isoforms across breast cancer subtypes and their distinct roles in the tumor microenvironment offer valuable insights into the biology of breast cancer and open new avenues for targeted therapy. The development of selective ALDH1A inhibitors, such as ABD0171, holds promise for improving outcomes in patients with aggressive breast cancer subtypes, particularly TNBC and basal-like cancers.

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)[12],

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[22], 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. 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[9,10], *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, $HER2^+$, and Luminal.

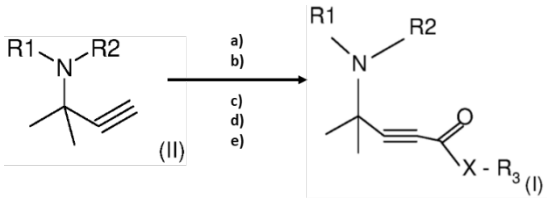
Breast tumor patient samples: mRNA (bulk) gene expression data from three breast cancer studies were retrieved from cBioPortal (https://www.cbioportal.org): Breast Invasive Carcinoma (TCGA, Cell 2015)[8], Breast Invasive Carcinoma (TCGA, PanCancer Atlas)[7], and Breast Invasive Carcinoma (TCGA, Firehose 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, $HER2^+$, 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[17]. 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)[23].

Normalization was performed using the sctransform v2 method with the glmGamPoi package[24]. 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[18]. Further data processing including normalization, dimensionality reduction, and clustering was conducted with the Seurat R package (v4.4)[23]. Prior to data integration, individual patient samples were grouped according to their molecular classification: Basal-like, HER2+, 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 out 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[25], GGally[26], complexHeatmap[27] and Seurat[23] R packages.

Chemical synthesis of compounds. The process was described in patents WO2001092220 and WO2017064247. Briefly, a compound of formula (I) is obtained by



a) reacting a compound of formula (II) with an organic or inorganic acid.
b) reacting the compound obtained in step a) with a base.
c) reacting the compound obtained in step b) with CO_2.
d) reacting the compound obtained in step c) with alkyl chloroformate, a reagent able of forming, with the compound obtained in step c), an acid halide or a reagent able of forming, with the compound obtained in step c), a mixed anhydride.
e) reacting the compound obtained in step d) with an anion precursor compound SMe-.

The structural analog compounds are related to the formula (I), in which:
- X is an atom chosen from O or S.

- R1 and R2 identical or different are independently chosen from: linear or branched (C_1-C_7)alkyl, linear or branched (C_2-C_7)alkenyl, aryl, heteroaryl, $CHR_5CHR_6OR_4$ and $(CHR_5)_vOR_4$, said aryl and heteroaryl being optionally substituted by one or more substituents chosen from: linear or branched (C_1-C_7)alkyl, halogen, NO_2 and $CONH_2$.
- v is chosen from 2 to 4.
- R3 is chosen from linear or branched (C_1-C_7)alkyl and linear or branched (C_1-C_7)alkyl-NY_1Y_2; said linear or branched (C_1-C_7)alkyl-NY_1Y_2 being optionally substituted by (C_1-C_7)alkyl -CO_2Z.
- R4 is chosen from: H, linear or branched (C_2-C_7)alkyl, linear or branched (C_2-C_7)alkenyl, aryl, heteroaryl, linear or branched -(C_1-C_7)alkyl-aryl and linear or branched -(C_1-C_7)alkyl-heteroaryl; said aryl and heteroaryl being optionally substituted by one or more substituents chosen from: linear or branched (C_1-C_7)alkyl optionally substituted by one or more halogen atom, linear or branched (C_1-C_7)alkoxy optionally substituted by one or more halogen atom, -COOH, aryl, -NRR', -NO_2 or said aryl and heteroaryl being optionally fused to form an heterocycloalkyl.
- R5 and R6 identical or different are independently chosen from:
    - H and linear or branched (C_1-C_7)alkyl, or
    - R5 and R6 are linked together to form with the carbon atoms to which they are attached a cycloalkyl, aryl or
      heteroaryl, or
    - R5 is H and R1 and R6 are linked together to form with the nitrogen atom linked to R1 an heterocycloalkyl or
      heteroaryl, or
    - R6 is H and R1 and R5 are linked together to R1 to form with the nitrogen atom linked to R1 an heterocycloalkyl.
- R and R' identical or different, are independently chosen from H and linear or branched (C_1-C_7)alkyl.
- Y_1 and Y_2 identical or different are independently chosen from H and -CO-(C_i-C_7)alkyl.
- Z is chosen from H and linear or branched (C_1-C_7)alkyl.
- and in which, at least one of R1 and R2 is $CHR_5CHR_6OR_4$ or $(CHR_5)_vOR_4$ when X is S and R3 is linear or branched (C_1-C_7)alkyl; or its pharmaceutically acceptable salts or optical isomers, racemates, diastereoisomers, enantiomers or tautomers.

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[28] 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[28]. 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[28]. 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[29]. 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 commonly reported parameter to describe time-dependent inhibitors[30].

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[31]. 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 37ºC 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 enzyme 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[32,33]. 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[34], 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 −20ºC 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 37ºC. 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· min^{-1}; m/z range 800-3000.

Cell lines and culturing

The human breast cancer cell lines (Supplementary Data 3A) used in this study include 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). Additional cell lines such as 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 analysed 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[35], and the GI_{50} values were extracted from the fitted models.

Linear mixed-effects model analysis

GI50 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+ 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[36]. This model assessed the effects of drug (DIMATE or ABD0171), molecular subtype (TNBC or HER2+), and their interaction on the observed GI50 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 assay

Cells were stained with annexin V–FITC together with propidium iodide (PI) and evaluated for apoptosis by flow cytometry according to the manufacturer's protocol (Thermo Scientific, Strasbourg, FR). For acquisition and analysis an Accuri C6 flow cytometer with BD Accuri C6 software (Beckton Dickinson) was used.

Detection of ROS

Cellular ROS production was measured using 2’,7’-dichlorodihydrofluorescein-diacetate according to the manufacturer’s protocol (VWR) and evaluated by flow cytometry.

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 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. 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

For the migration assay, cells were seeded at 5 × 10^4 cells per well in a 96-well plate from the Oris Cell Migration Assembly Kit-FLEX containing a well insert with a defined round, cell-free gap (PlatypusTechnologies, Madison, USA). Cells were incubated and allowed to growth for 24 h. Once the cells reached 100% confluence, ALDH inhibitors were added into the media, the culture insert was removed and the area that remained clear of cells was quantified for 24 h using the SpectraMaxI3-Minimax station (Molecular Devices, USA) and the ImageJ analysis software[37]. For digital images at 0- and 24-hours cells were stained with Calcein AM (Fisher Scientific). Data consisted of triplicates for each experimental condition.

For invasion assay[38], cells were seeded at 5 - 10 × 10^4 cells on top of ECM-coated filter membrane in transwell inserts (Corning, Lyon, FR) and 600µL of complemented media, with or without ALDH inhibitors, were carefully added in the bottom of a lower chamber in a 24-well plate. Following incubation for 24 - 48 hours, cells from the upper chamber were removed with cotton swabs and migrated cells were fixed in 70% ethanol, stained with 0.2% crystal violet and quantified by counting underneath a microscope.

Anoikis assay

Cells were cultured in 96-well plates coated with (non-anchored plate) or without poly-HEMA (anchored plate), and then treated with DIMATE or ABD0171 for 24h. Calcein was added to each well and incubated 20 min at 37°C, and then propidium iodide (PI). SpectraMax MiniMax 300 was used to detect the green calcein AM fluorescence (Ex 495nm; Em 530nm) staining live cells and red PI fluorescence (Ex 530nm; Em 645nm) staining dead cells Then, the experiment was translated in 100mm dishes and cells were treated with DIMATE or ABD0171 for 14h to perform capillary electrophoresis immunoassay as described previously.

Animal experiments

All the mice were provided by Janvier Labs (Le Genest-Saint-Isle, France) or Charles River GmbH (Sulzfeld, Germany) and cared in the animal facility Commune Scar Rockefeller (Lyon, France) or Reaction Biology Europe GmbH (Freiburg, Germany) in accordance with animal welfare regulations.

For toxicity studies, groups of 4 female SCID-CB17 immunocompromised mice were treated receiving an intra-peritoneal injection of DIMATE (15 or 20mg/kg) or ABD0171 (30 or 40mg/kg) once every third day for 4 weeks. The control group was treated with vehicle (PBS containing 10% of Kolliphor® HS15). Mice were monitored for body weight and physical aspect. Weekly blood sampling was performed after every week of treatment (day 5, 12, 19 and 26) for hematological and enzymatic tests.
For xenograft animal model, 5 x 10^6 MDAMB231-GFP cells were orthotopically implanted in 50 SCID-CB17 immunocompromised mice. For syngeneic animal model, 0.5 x 10^6 4T_Luc-M3 cells were orthotopically implanted in 36 BALBL/cAnNCrl immunocompetent mice. When the tumors reached a volume approximative of 50mm3, mice with similarly sized tumors were randomized into treatment cohorts (n = 10 per group for MDAMB231-GFP, n = 12 for 4T1_Luc-M3). Treated groups received an intra-peritoneal injection of DIMATE (20mg/kg) or ABD0171 (40mg/kg) once every third day for 3 weeks. The control group was treated with vehicle. Mice were monitored for body weight and tumor growth for 3 weeks, and daily for behavior and survival. Sacrifice was performed at the end of week 3 or when the tumor reached a maximum volume of 700mm3 for MDAMB231 and 1600 mm3 for 4T1_Luc-M3. Tumors were cut and one part was frozen in liquid nitrogen and kept at -80°C, and the other part paraffine-embedded for immunohistochemical assays. For MDAMB231-GFP model, lung was collected and processed for ex vivo metastasis GFP monitoring by flow cytometry; for 4T1_Luc-M3 model, several organs were collected and processed for ex vivo metastasis luciferase assay.

Ex vivo bioluminescence imaging (BLI) analysis

Data for ex vivo BLI analysis consisted of organ-specific raw measurements across multiple animal groups. 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 (in grams), resulting in values expressed as BLI per gram of tissue. Statistical analysis was performed using Wilcoxon rank-sum tests, comparing experimental groups to the control (PBS). P-values

were adjusted for multiple testing using the Benjamini-Hochberg method, and significance was determined at a threshold of $p \leq 0.05$.

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. Comparisons between groups were performed with Kruskal-Wallis test. 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 also 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. Additionally, RP received financial support from Advanced BioDesign through a research contract agreement with the Universitat Autònoma de Barcelona. 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 the bioinformatics and statistical analyses. RP, XP, and JF designed, conducted, and interpreted the enzymatic and biochemical assays. AC conducted and analyzed the *in vitro* cytotoxicity screenings and functional assays. AC, 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. RRR, MP, 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)** Schematic overview of data collection, showing the number of human cell lines and patient samples sourced and analyzed. **(b-c)** Heatmap with the differential expression of PAM50 genes *and ALDH1A1, ALDH1A2* and *ALDH1A3* in cell lines and patient samples, respectively **(d)** Median bulk expression of 59 genes across cell lines and patient samples in breast cancer, categorized by molecular classification. Dot size depicts median absolute deviation (MAD), with color intensity indicating log2 expression levels. **(e)** Spearman correlation of gene expression across groups from **(d)**, highlighting significant correlations (p-value ≤ 0.01) with stars. A gray square emphasizes a positive correlation of 0.71 between Basal A and Basal-like subtypes from cell lines and patient samples, respectively.

Figure 2. Single-cell RNA sequencing uncovering distinct ALDH1A isoform expression across tumor subpopulations in breast Cancer. (a,b) Expression of *ALDH1A1, ALDH1A2*, and *ALDH1A3* across 20,485 cells from 16 basal-like classified breast cancer cell lines depicted through feature and violin plots. Classification into Basal A and B subgroups capture intra-group heterogeneity and variability. **(c)** single-cell analysis of 26 patient samples comprising 100.064 cells, to elucidate ALDH1A expression landscape. **(d)** Cell population distribution by molecular classification highlighting cellular heterogeneity **(e)** Feature plots for *ALDH1A1, ALDH1A2*, and *ALDH1A3* show gene expression distribution and intensity among individual cells (left) and Pie charts for comparative gene expression of each ALDH1A isoform, showing the percentage of cells grouped by population and molecular subtype (right). Cell types: cancer-associated fibroblasts (CAFs), myofibroblasts (myCAFs), inflammatory (iCAFs), MSCs: mesenchymal stem cells.

Figure 3. Novel chemical compounds with inhibitory activity against ALDH1A and cytotoxicity towards breast cancer cell lines (a) Schematic representation of novel structural analogs of DIMATE and MATE **(b)** Circular stacked barplots illustrating the remaining activity of drugs against five aldehyde dehydrogenase isoforms. Each segment represents a distinct isoform. ALDH1A1, ALDH1A2 and ALDH1A3 are represented in shades of green, with darker to lighter shades, respectively. ALDH2 and ALDH3A1 isoforms are depicted in shades of grey, following the same pattern. The length of each stacked segment indicates the proportion of remaining drug activity against the respective isoform, providing a comparative overview of drug efficacy across different targets **(c)** Heatmap showing the cytotoxic activity (GI50 values) of the novel compounds, grouped by their series number, and their inhibitory effects on ALDH1A1,

ALDH1A2, and ALDH1A3. The analysis was performed across a panel of 18 breast cancer cell lines, stratified by their respective molecular subtypes**. (d)** Schematic representation of the experimental design for 3D clonogenic assays using PDX cells derived from advanced HER2-positive and TNBC breast cancer models (top). Violin plots illustrate the cytotoxic activity (GI50 values) of the compounds DIMATE and ABD0171 on these cells (bottom). **(e**) Plots from a linear mixed-effects model showing the predicted versus observed effects of ABD0171 and DIMATE in TNBC and HER2-positive breast cancer subtypes.

Figure 4. (a) Calculation of k_{inact}/K_i values from the plots of k_{obs} *versus* concentrations of DIMATE and ABD0171 for ALDH1A1, ALDH1A2 and ALDH1A3. For DIMATE, concentrations ranging from 0 to 500 μM for ALDH1A1, 0 to 200 μM for ALDH1A2, and 0 to 300 μM for ALDH1A3 were tested. For ABD0171, concentrations ranged from 0 to 300 μM for ALDH1A1, 0 to 3 μM for ALDH1A2, and 0 to 200 μM for ALDH1A3. Activities were measured at saturating substrate concentrations (250 μM hexanal) in the corresponding reaction buffer. Experimental data are presented as mean ± standard deviation. **(b)** ESI-MS spectra of ALDH1A3 treated with DIMATE (left) and ABD0171 (right). The spectra show single-charged peaks (1+) at m/z between 61075 and 61119. The resulting peaks in the ESI-MS spectra correspond to 16.7 μM ALDH1A3 treated with 167 μM DIMATE (left) or 167 μM ABD0171 (right) (red signal), while the black signal represents the untreated ALDH1A3 control. The analysis of these spectra confirmed covalent binding of ALDH1A3 to DIMATE and its analogue ABD0171. Samples were prepared in 100 mM ammonium bicarbonate (pH 7.5) and pre-incubated for 45 minutes at 37°C.

Figure 5. **Effects of ALDH inhibitors ABD0171 and DIMATE on cell viability, invasion, and apoptosis in triple-negative breast cancer (TNBC) cell lines.** **(a)** Apoptosis induction. Flow cytometry analysis of early and late apoptosis in TNBC cells treated with ABD0171 or DIMATE for 12 hours, 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)** Activation of apoptotic pathways. WES immunoblots illustrating the activation of the extrinsic apoptotic pathway and the JNK (c-Jun N-terminal kinase) signaling pathway in response to oxidative stress triggered by ABD0171 and DIMATE. Both compounds modulate the expression of pro- and anti-apoptotic proteins after 12 hours of treatment. **(c)** Inhibition of cell migration and invasion. Representative images from migration and invasion assays in MDAMB231 cells treated with ABD0171 or DIMATE for 24 hours. Scatter plots display quantification from three independent experiments (median with

range). Statistical analysis was performed using Welch's t-test. Images were acquired with the SpectraMax® MiniMax™ 300 Imaging Cytometer (Molecular Devices, USA). **(d)** Anoikis resistance assay. Evaluation of anoikis resistance in MDAMB231 cells cultured on adherent (control) and non-adherent (poly-HEMA-coated) plates. Cells are stained with calcein-AM and propidium iodide, with representative images captured using the SpectraMax® MiniMax™ 300 Imaging Cytometer. The resazurin-based toxicity assay was used to quantify cell death under different conditions, with results presented as median values with the range of triplicates. Statistical analysis was performed using an unpaired t-test with Welch's correction. Anoikis resistance was measured after 24 hours of treatment with ABD0171 or DIMATE, showing reduced resistance in MDAMB231 cells. **(e)** WES immunoblots showing the activation of the STAT3 pathway, a key regulator in anoikis, after 24 hours of treatment with ABD0171 or DIMATE.

Figure 6. ALDH inhibitors reduce tumor growth in immunocompromised and immunocompetent triple negative breast cancer models. (a, b) Evaluation of survival and antitumor efficcacy of DIMATE and ABD0171 in two murine models. SCID-CB17 mice bearing MDAMB231-GFP tumors and BALB/c mice with 4T1-Luc_M3 tumors were treated intraperitoneally with DIMATE (20 mg/kg) or ABD0171 (40 mg/kg), administered three times per week for 3 or 4 weeks. Control groups received vehicle treatment. Tumor growth inhibition was assessed, and overall survival rates were recorded. **(c)** Quantification of metastatic spread in the indicated organs, determined by luminescence intensity using bioluminescence imaging.

REFERENCES

1. 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).

2. 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).

3. 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).

4. 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).

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

6. 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).

7. 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).

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

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

10. 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).

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

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

13. 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).

14. 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).

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

16. Prat, A. *et al.* Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Research* **12**, (2010).

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

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

19. 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).

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

21. 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).

22. 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).

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

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

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

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

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

28. 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).

29. 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).

30. 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).

31. 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).

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

33. 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). doi:10.1007/978-1-60327-325-1_8.

34. 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).

35. 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).

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

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

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

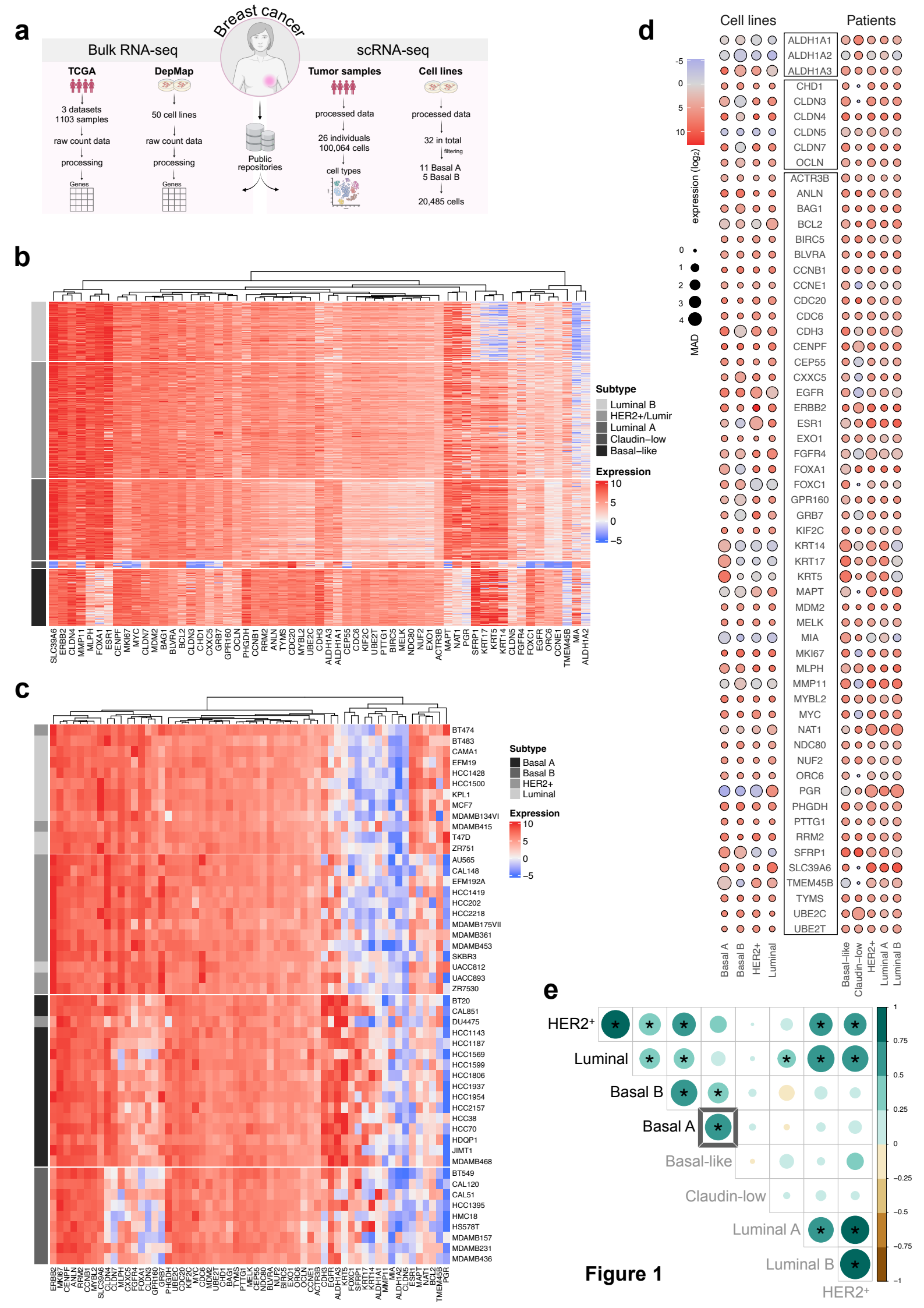

a
Breast cancer
Bulk RNA-seq
scRNA-seq
TCGA
DepMap
Tumor samples
Cell lines
3 datasets
1103 samples
raw count data
processing
Genes
50 cell lines
raw count data
processing
Genes
Public repositories
processed data
26 individuals
100,064 cells
cell types
processed data
32 in total
filtering
11 Basal A
5 Basal B
20,485 cells
b
Subtype
Luminal B
HER2+/Lumir
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
expression (log2)
MAD
Basal A
Basal B
HER2+
Luminal
Basal–like
Claudin–low
HER2+
Luminal A
Luminal B
e
HER2+
Luminal
Basal B
Basal A
Basal-like
Claudin-low
Luminal A
Luminal B
HER2+

Figure 1

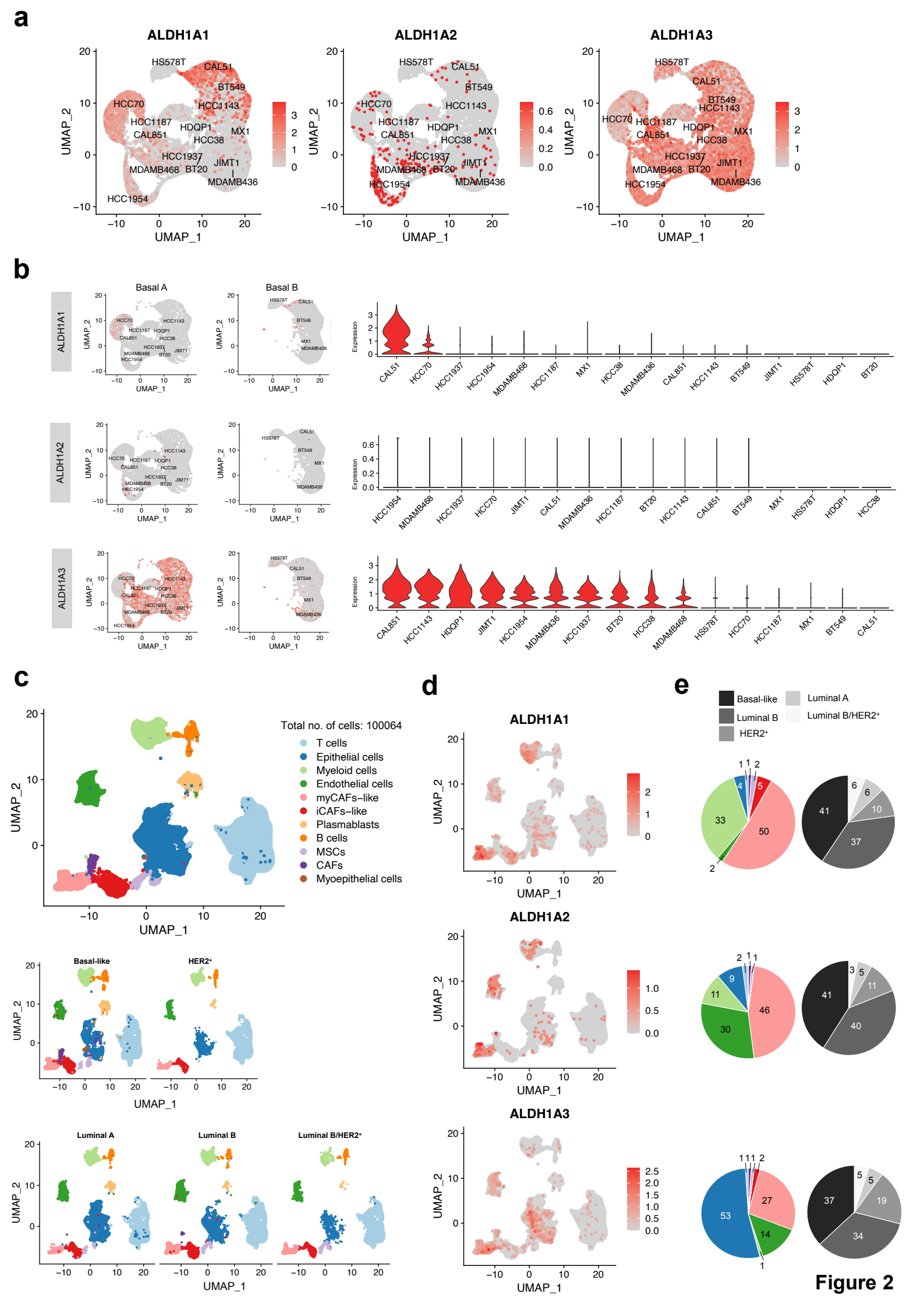

a
ALDH1A1
ALDH1A2
ALDH1A3
UMAP_1
UMAP_2
HS578T
CAL51
BT549
HCC70
HCC1143
HCC1187
HDQP1
CAL851
MX1
HCC38
HCC1937
JIMT1
MDAMB468
BT20
MDAMB436
HCC1954
b
Basal A
Basal B
Expression
c
Total no. of cells: 100064
T cells
Epithelial cells
Myeloid cells
Endothelial cells
myCAFs−like
iCAFs−like
Plasmablasts
B cells
MSCs
CAFs
Myoepithelial cells
Basal-like
HER2⁺
Luminal A
Luminal B
Luminal B/HER2⁺
d
e
Basal-like
Luminal A
Luminal B
Luminal B/HER2⁺
HER2⁺

Figure 2

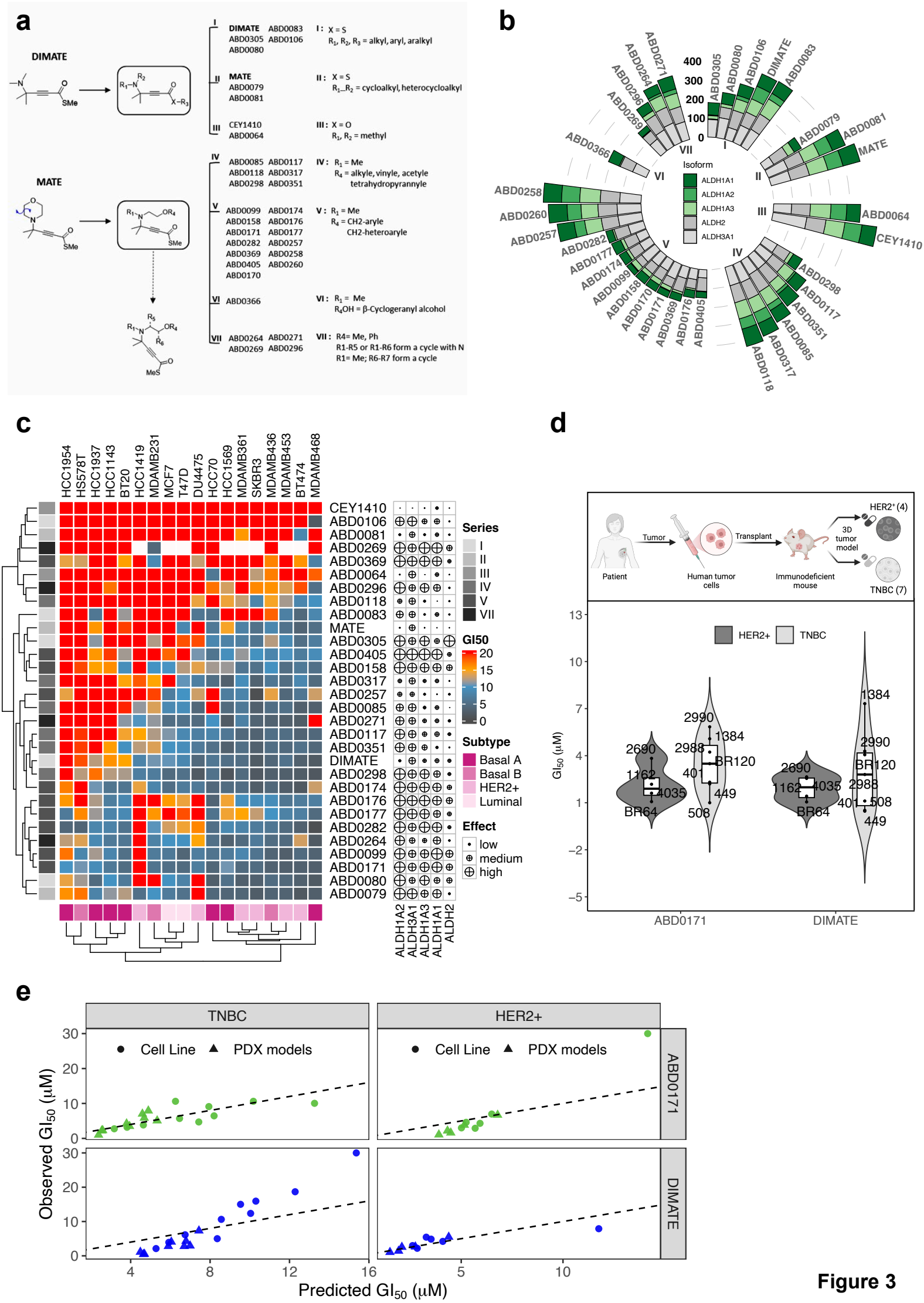
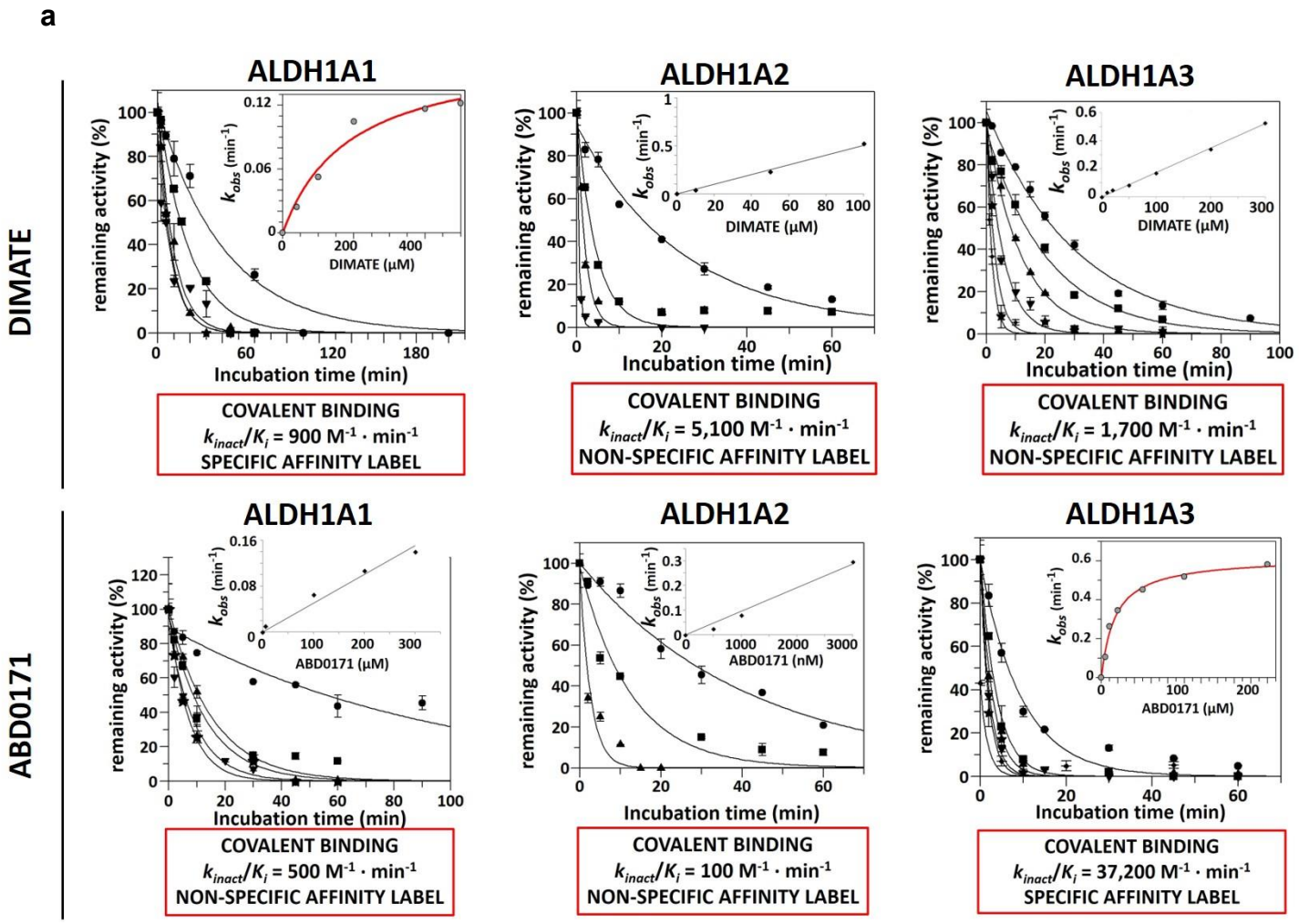

a
DIMATE
MATE
I DIMATE ABD0083 ABD0305 ABD0106 ABD0080
I : X = S
R1, R2, R3 = alkyl, aryl, aralkyl
II MATE ABD0079 ABD0081
II : X = S
R1_R2 = cycloalkyl, heterocycloalkyl
III CEY1410 ABD0064
III : X = O
R1, R2 = methyl
IV ABD0085 ABD0117 ABD0118 ABD0317 ABD0298 ABD0351
IV : R1 = Me
R4 = alkyle, vinyle, acetyle
tetrahydropyrannyle
V ABD0099 ABD0174 ABD0158 ABD0176 ABD0171 ABD0177 ABD0282 ABD0257 ABD0369 ABD0258 ABD0405 ABD0260 ABD0170
V : R1 = Me
R4 = CH2-aryle
CH2-heteroaryle
VI ABD0366
VI : R1 = Me
R4OH = β-Cyclogeranyl alcohol
VII ABD0264 ABD0271 ABD0269 ABD0296
VII : R4= Me, Ph
R1-R5 or R1-R6 form a cycle with N
R1= Me; R6-R7 form a cycle
b
Isoform
ALDH1A1
ALDH1A2
ALDH1A3
ALDH2
ALDH3A1
c
Series
Gl50
Subtype
Basal A
Basal B
HER2+
Luminal
Effect
low
medium
high
d
HER2+ (4)
TNBC (7)
Patient
Tumor
Human tumor cells
Transplant
Immunodeficient mouse
3D tumor model
HER2+
TNBC
ABD0171
DIMATE
e
TNBC
HER2+
Cell Line
PDX models
ABD0171
DIMATE
Observed Gl50 (μM)
Predicted Gl50 (μM)

Figure 3


a
DIMATE
ALDH1A1
remaining activity (%)
Incubation time (min)
k_{obs} (min^{-1})
DIMATE (μM)
COVALENT BINDING
k_{inact}/K_i = 900 M^{-1} · min^{-1}
SPECIFIC AFFINITY LABEL
ALDH1A2
COVALENT BINDING
k_{inact}/K_i = 5,100 M^{-1} · min^{-1}
NON-SPECIFIC AFFINITY LABEL
ALDH1A3
COVALENT BINDING
k_{inact}/K_i = 1,700 M^{-1} · min^{-1}
NON-SPECIFIC AFFINITY LABEL
ABD0171
ALDH1A1
ABD0171 (μM)
COVALENT BINDING
k_{inact}/K_i = 500 M^{-1} · min^{-1}
NON-SPECIFIC AFFINITY LABEL
ALDH1A2
ABD0171 (nM)
COVALENT BINDING
k_{inact}/K_i = 100 M^{-1} · min^{-1}
NON-SPECIFIC AFFINITY LABEL
ALDH1A3
ABD0171 (μM)
COVALENT BINDING
k_{inact}/K_i = 37,200 M^{-1} · min^{-1}
SPECIFIC AFFINITY LABEL

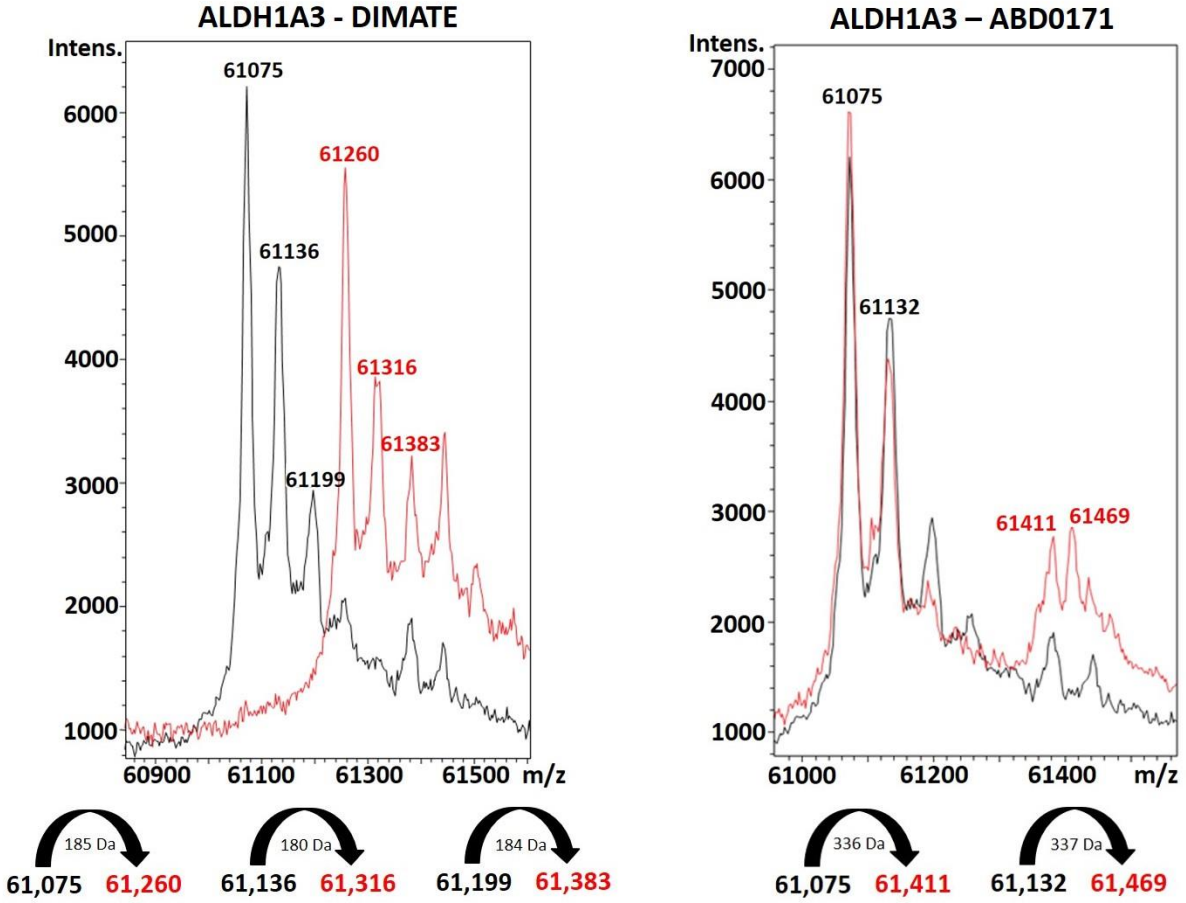

b
ALDH1A3 - DIMATE
Intens.
61075
61260
61136
61316
61383
61199
m/z
185 Da
180 Da
184 Da
61,075 61,260
61,136 61,316
61,199 61,383
ALDH1A3 – ABD0171
Intens.
61075
61132
61411
61469
m/z
336 Da
337 Da
61,075 61,411
61,132 61,469

Figure 4

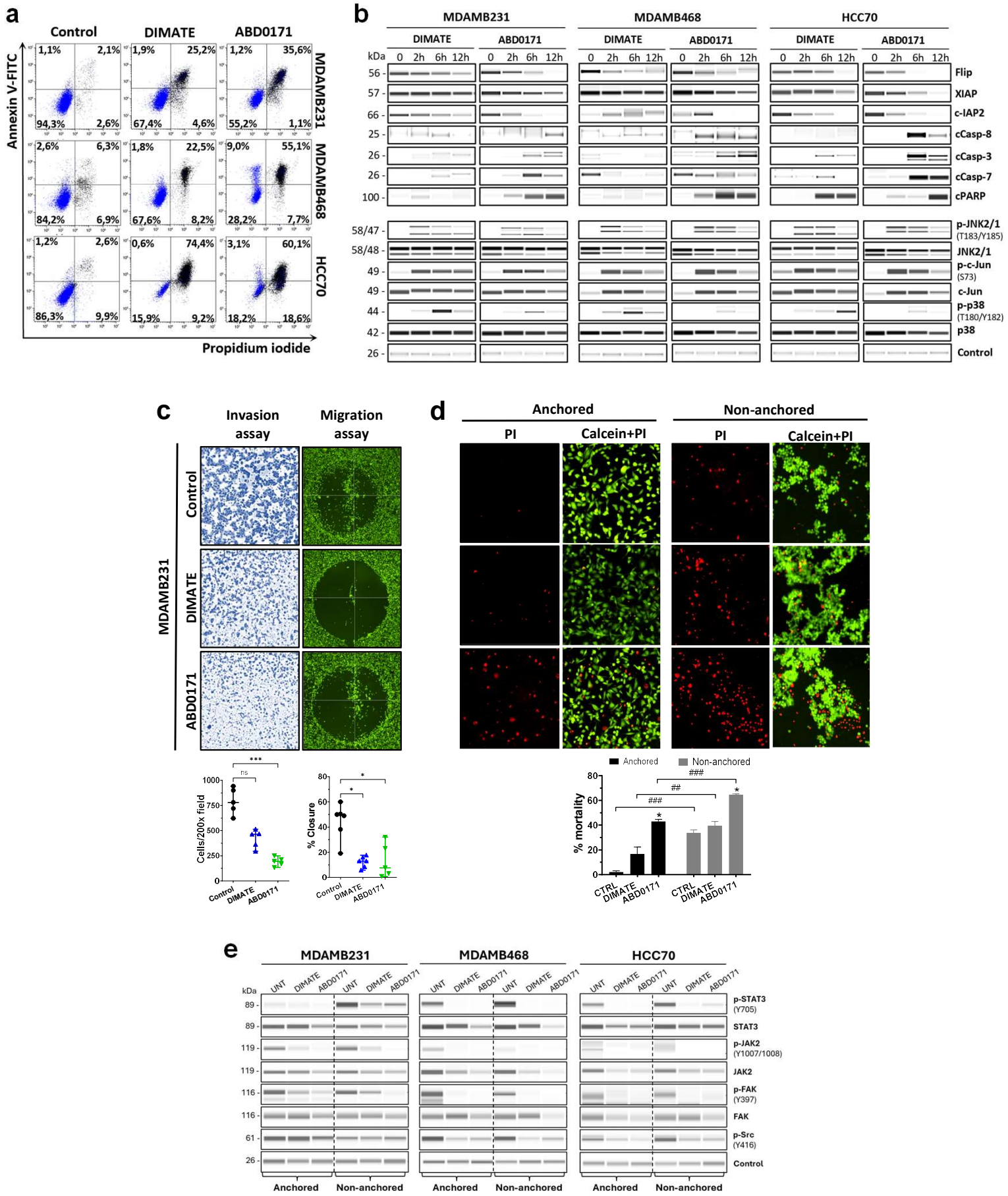
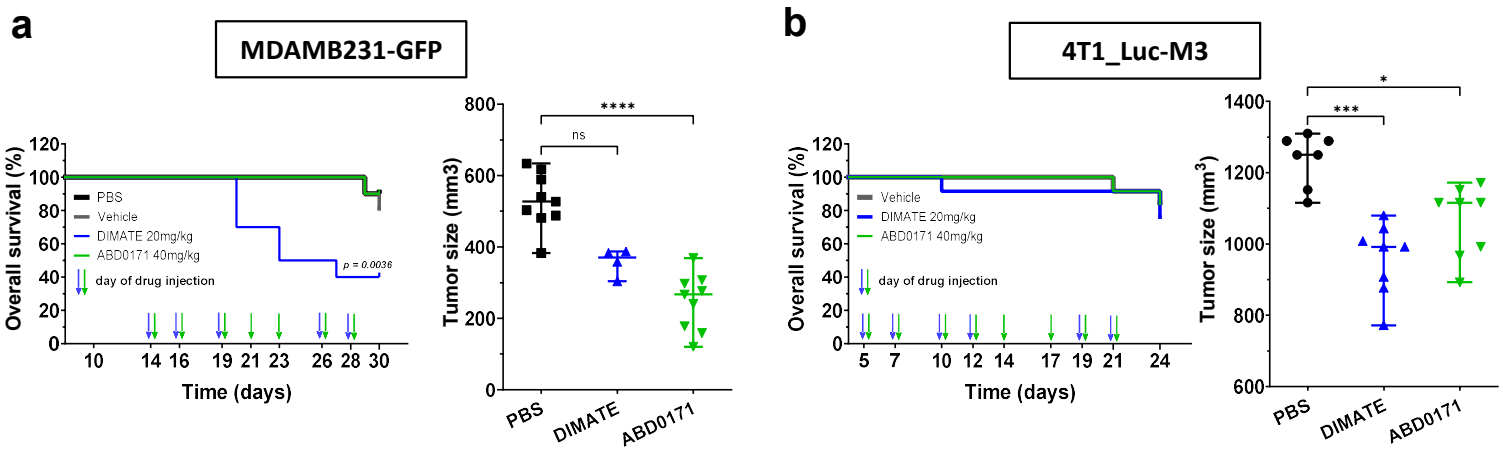

a
Control
DIMATE
ABD0171
Annexin V-FITC
Propidium iodide
MDAMB231
MDAMB468
HCC70
b
MDAMB231
MDAMB468
HCC70
DIMATE
ABD0171
kDa
0 2h 6h 12h
Flip
XIAP
c-IAP2
cCasp-8
cCasp-3
cCasp-7
cPARP
p-JNK2/1 (T183/Y185)
JNK2/1
p-c-Jun (S73)
c-Jun
p-p38 (T180/Y182)
p38
Control
c
Invasion assay
Migration assay
Control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

Figure 5


a
MDAMB231-GFP
Overall survival (%)
PBS
Vehicle
DIMATE 20mg/kg
ABD0171 40mg/kg
day of drug injection
p = 0.0036
Time (days)
Tumor size (mm3)
****
ns
PBS
DIMATE
ABD0171
b
4T1_Luc-M3
Overall survival (%)
Vehicle
DIMATE 20mg/kg
ABD0171 40mg/kg
day of drug injection
Time (days)
Tumor size (mm3)
***
*
PBS
DIMATE
ABD0171

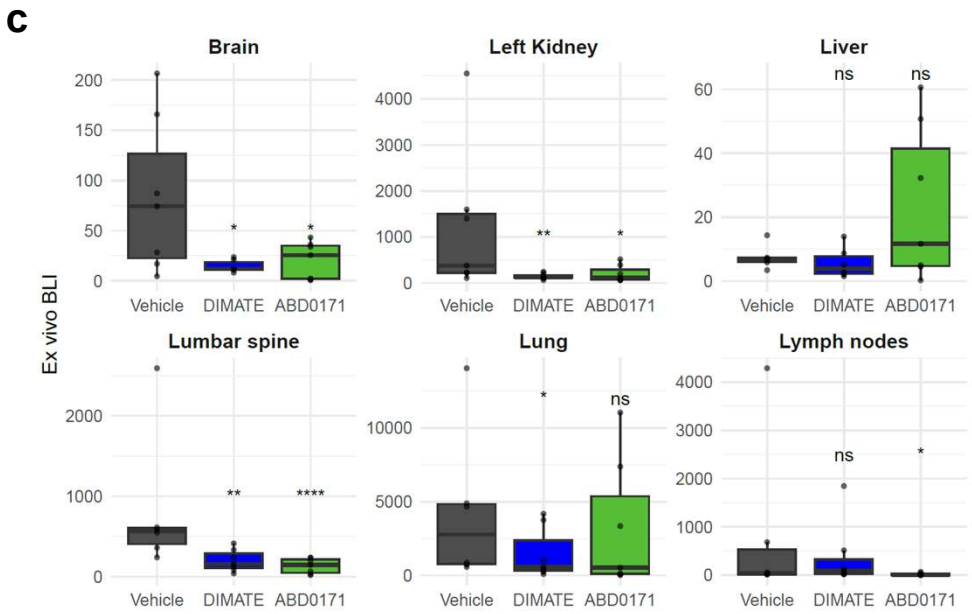

c
Brain
Left Kidney
Liver
Lumbar spine
Lung
Lymph nodes
Ex vivo BLI
Vehicle
DIMATE
ABD0171
ns

Figure 6

Table S1. IC_{50} (μM) and k_{inact}/K_i ($M^{-1} \cdot min^{-1}$) or K_i values of DIMATE and ABD0171 for ALDH3A1 and ALDH2 using 4-nitrobenzaldehyde and hexanal as substrate, respectively.

	ALDH3A1		ALDH2	
COMPOUND	**IC_{50} (μM)**	**K_i (μM)**	**IC_{50} (μM)**	**k_{inact}/K_i ($M^{-1} \cdot min^{-1}$)**
DIMATE	303 ± 46	253 ± 48	72 ± 9	425
ABD0171	242 ± 11	233 ± 16	12 ± 3	540

IC_{50} values were obtained fluorometrically using 4-nitrobenzaldehyde for ALDH3A1 and hexanal for ALDH2 as substrates. The reaction mixture was incubated in the corresponding reaction buffer[28] for 20 min at 37ºC, before adding $NAD^+/NADP^+$ cofactor, NADH/NADPH internal standard and the corresponding substrate. k_{inact}/K_i values were obtained by k_{obs} *vs* [I] plot. K_i values were obtained using a non-competitive fit of DIMATE and ABD0171 inhibition against ALDH3A1 at various concentrations of inhibitor compound. Experimental values are the mean ± standard error.

TABLE S2. Antibodies used for capillary electrophoresis immunoassays.

Antibody	Provider	Reference
ALDH1A1	Proteintech	15910-1-AP
ALDH1A2	Proteintech	13951-1-AP
ALDH1A3	Novus	NBP2-15339
JNK	Cell Signaling	9252
Phopsho-JNK (T183/Y185)	Cell Signaling	4668
p38	Cell Signaling	8690
Phospho-p38 (T180/Y182)	Cell Signaling	4511
c-Jun	Cell Signaling	9165
Phospho-c-Jun Ser73 (S73)	Cell Signaling	3270
Flip	Cell Signaling	56343
XIAP	Cell Signaling	14334
C-IAP2	Cell Signaling	3130
Cleaved Caspase 8	Cell Signaling	9496
Caspase 3	Cell Signaling	9662
Caspase 7	Cell Signaling	9492
Cleaved PARP	Cell Signaling	5625
STAT3	Cell Signaling	9139
Phospho-STAT3 (Y705)	Cell Signaling	9145
JAK2	Cell Signaling	3230
Phospho-JAK2 (Y1007/1008)	Cell Signaling	3776
FAK	Cell Signaling	3285
Phospho-FAK (Y397)	Cell Signaling	8556
Phospho-Src (Y416)	Cell Signaling	20101
Loading Control	Protein Simple	042-196