Distinct prognostic values and potential drug targets of ALDH1 isoenzymes in non-small-cell lung cancer

Qinghua You1
Huanchen Guo2
Dongxiang Xu3

1Department of Pathology, Shanghai Pudong Medical Center, Fudan University, Shanghai, 2Department of Respiratory Medicine, Shouguang Hospital of Traditional Chinese Medicine, Shouguang, 3Department of Endocrinology, Shanghai Pudong Hospital, Fudan University, Pudong Medical Center, Shanghai, People’s Republic of China

Abstract: Increased aldehyde dehydrogenase 1 (ALDH1) activity has been found in the stem cell populations of leukemia and some solid tumors including non-small-cell lung cancer (NSCLC). However, which ALDH1 isoenzymes are contributing to ALDH1 activity remains elusive. In addition, the prognostic value of individual ALDH1 isoenzyme is not clear. In the current study, we investigated the prognostic value of ALDH1 isoenzymes in NSCLC patients through the Kaplan–Meier plotter database, which contains updated gene expression data and survival information from a total of 1,926 NSCLC patients. High expression of ALDH1A1 mRNA was found to be correlated to a better overall survival (OS) in all NSCLC patients followed for 20 years (hazard ratio [HR] 0.88 [0.77–0.99], P=0.039). In addition, high expression of ALDH1A1 mRNA was also found to be correlated to better OS in adenocarcinoma (Ade) patients (HR 0.71 [0.57–0.9], P=0.0044) but not in squamous cell carcinoma (SCC) patients (HR 0.92 [0.72–1.16], P=0.48). High expression of ALDH1A2 and ALDH1B1 mRNA was found to be correlated to worser OS in all NSCLC patients, as well as in Ade, but not in SCC patients. High expression of both ALDH1A3 and ALDH1L1 mRNA was not found to be correlated to OS in all NSCLC patients. These results strongly support that ALDH1A1 mRNA in NSCLC is associated with better prognosis. In addition, our current study also supports that ALDH1A2 and ALDH1B1 might be major contributors to the ALDH1 activity in NSCLC, since high expression of ALDH1A2 and ALDH1B1 mRNA was found to be significantly correlated to worser OS in all NSCLC patients. Based on our study, ALDH1A2 and ALDH1B1 might be excellent potential drug targets for NSCLC patients.

Keywords: NSCLC, ALDH1, cancer stem cell, prognosis, drug target, KM plotter, hazard ratio

Introduction
Non-small-cell lung cancer (NSCLC) includes the following major histological types: adenocarcinoma (Ade) and squamous cell carcinoma (SCC). NSCLC accounts for approximately 85% of all lung cancers, which makes NSCLC the leading cause of cancer-related deaths worldwide.1,2 Despite the advances in early detection, radical cure operation, and multimodal therapeutic modalities, at diagnosis, there are approximately 80% of NSCLC cases in advanced stage. The incidence of NSCLC is still increasing, and systemic chemotherapy that remains in the standard care only provides marginal improvement in survival.3 Therefore, further investigation on the mechanism of initiation, progression, and identification of prognostic markers and potential drug targets is still needed and will help select the patients with higher chance of lung cancer recurrence in order to provide better prognosis and individualized treatment.
Aldehyde dehydrogenase 1 (ALDH1) family is composed of enzymes that are expressed at high levels in stem cells (SCs) and contributes to the oxidation of retinol to retinoic acid in SC differentiation.4–6 Elevated ALDH1 activity has been detected in the SC populations of human acute myeloid leukemia, multiple myeloma, and a number of solid tumors.7–11 Therefore, examination of ALDH1 activity might be used as a common marker for both normal and malignant SC populations. Increased ALDH1 expression has been reported in some lung cancer cell lines, malignant transformation of lung cells,12,13 and lung tissues.14 However, which ALDH1’s isoenzymes are contributing to ALDH1 activity in NSCLC has not been determined. In addition, the prognostic value of individual ALDH1 isoenzyme in NSCLC remains elusive. The “Kaplan–Meier plotter” (KM plotter) generated data from Gene Expression Omnibus (GEO; www.ncbi.nlm.nih.gov/geo) database. KM plotter database was initially established using data on expression manifested by 22,277 genes in a group of 1,809 breast cancer patients.15,16 Later, this database also included gene expression data and survival information from a total of 1,715 NSCLC patients.17 Currently, they increased the patient number to a total of 1,926. Thus, KM plotter can be utilized for the analysis of individual genes with clinical results to relapse-free survival and total survival of the patients. So far, a number of genes have been identified and/or validated by KM plotter in breast cancer18–26 as well as in NSCLC.27 In this study, we determined the prognostic value of individual ALDH1 isoenzymes in human NSCLC patients using KM plotter database.

Materials and methods

We used an online database15 to determine the relevance of individual ALDH1 members’ mRNA expression to relapse-free survival. Currently, they have breast cancer,15 lung cancer,17 ovarian cancer,28 and gastric cancer database. NSCLC patients in the database were identified from Cancer Biomedical Informatics Grid (http://cabig.cancer.gov/), microarray samples are published in the caBIG cancer project, the GEO (www.ncbi.nlm.nih.gov/geo/), and The Cancer Genome Atlas (TCGA; http://cancergenome.nih.gov/) lung cancer datasets.17 They collected clinical data including only age, sex, smoking history, histology, stage, grade, success of surgery, radiotherapy, and applied chemotherapy for all patients in WinStat 2013. The database was established using gene expression data and survival information of 1,926 NSCLC patients downloaded from GEO. Briefly, five ALDH1 sub-members (ALDH1A1, ALDH1A2, ALDH1A3, ALDH1B1, and ALDH1L1) were entered into the database (http://kmplot.com/analysis/index.php?p=service&cancer=breast) to obtain Kaplan–Meier survival plots in which the number at risk is indicated below the main plot. Hazard ratio (HR; and 95% confidence intervals) and log rank P were calculated and displayed on the webpage.

Results

The ALDH1 family is composed of six sub-members, and their characteristics are shown in Table 1. As previously indicated,26 among all the six ALDH1 isoenzymes, only ALDH1L2 was not found in www.kmplot.com, probably due to its low expression.

We first examined the prognostic value of ALDH1A1 mRNA expression in www.kmplot.com. The desired Affymetrix ID is valid: 212224_at (ALDH1A1). Survival curves are plotted for all patients (n=1,926) (Figure 1A), for Ade (n=719) (Figure 1B), and for SCC (n=525) (Figure 1C). High expression of ALDH1A1 mRNA was found to be correlated to better overall survival (OS) in all NSCLC patients followed for 20 years (HR 0.88 [0.77–0.99], P=0.039). In addition, high expression of ALDH1A1 mRNA was also found to be correlated to better OS in Ade patients (HR 0.71 [0.57–0.9], P=0.0044) but not in SCC patients (HR 0.92 [0.72–1.16], P=0.48).

Table 1 Alternatively spliced variants and characterization of ALDH1 isoenzymes

<table>
<thead>
<tr>
<th>Isoenzymes</th>
<th>Alternatively spliced variants</th>
<th>Cellular localization</th>
<th>Tissue distribution</th>
<th>Associated diseases</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALDH1A1</td>
<td>ALDH1A1_v2</td>
<td>Cytosol</td>
<td>Lung, breast, brain, pancreas, liver, kidney, etc</td>
<td>Alcoholism</td>
</tr>
<tr>
<td>ALDH1A2</td>
<td>ALDH1A2_v2, ALDH1A2_v3, ALDH1A2_v4</td>
<td>Cytosol</td>
<td>Kidney, testis, liver</td>
<td>Schizophrenia, spina bifida</td>
</tr>
<tr>
<td>ALDH1A3</td>
<td>ALDH1A3_v2</td>
<td>Cytosol</td>
<td>Skeletal muscle, lung, breast, kidney, etc</td>
<td>Autosomal recessive anophthalmia/microphthalmia</td>
</tr>
<tr>
<td>ALDH1B1</td>
<td>N/A</td>
<td>Mitochondria</td>
<td>Liver, heart, kidney, brain, prostate</td>
<td>N/A</td>
</tr>
<tr>
<td>ALDH1L1</td>
<td>N/A</td>
<td>Cytosol</td>
<td>Kidney, liver, skeletal muscle, brain, prostate</td>
<td>Ischemic stroke</td>
</tr>
<tr>
<td>ALDH1L2</td>
<td>ALDH1L2_v2, ALDH1L2_v2</td>
<td>Mitochondria</td>
<td>Pancreas, heart, and brain</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Abbreviations: ALDH1, aldehyde dehydrogenase 1; N/A, not available.
Figure 1. The prognostic value of ALDH1A1 expression.

Notes: The desired Affymetrix ID is valid: 212224_at (ALDH1A1). (A) Survival curves are plotted for all patients (n=1,926). (B) Survival curves are plotted for adenocarcinoma (n=719). (C) Survival curves are plotted for squamous cell carcinoma (n=525). Data was analyzed using Kaplan Meier Plotter (www.kmplot.com).

Abbreviation: HR, hazard ratio.
We then examined the prognostic value of ALDH1A2 mRNA expression in www.kmplot.com. The desired Affymetrix ID is valid: 207015_s_at (ALDH1A2). High expression of ALDH1A2 mRNA was found to be correlated to worse OS in all NSCLC patients (HR 1.24 [1.09–1.4], \( P=0.0093 \)) (Figure 2A). In addition, high expression of ALDH1A2 mRNA was also found to be correlated to worse OS in Ade patients (HR 1.57 [1.24–1.98], \( P=0.00015 \)) (Figure 2B) but not in SCC patients (HR 1.14 [0.9–1.44], \( P=0.29 \)) (Figure 2C).

Figure 3 shows the prognostic value of ALDH1A3 mRNA expression in www.kmplot.com. The desired Affymetrix ID is valid: 203180_at (ALDH1A3). The curves show that ALDH1A3 expression above or below the median does not separate the cases into significantly different prognostic groups in NSCLC patients (HR 0.99 [0.88–1.13], \( P=0.94 \)) (Figure 3A) or Ade patients (HR 0.97 [0.77–1.22], \( P=0.79 \)) (Figure 3B), or SCC patients (HR 1.04 [0.82–1.31], \( P=0.77 \)) (Figure 3C).

Figure 4 shows the prognostic value of ALDH1B1 mRNA expression in www.kmplot.com. The desired Affymetrix ID is valid: 20964_x_at (ALDH1B1). High expression of ALDH1B1 mRNA was found to be correlated to worse OS in all NSCLC patients (HR 1.53 [1.35–1.74], \( P=4.5e–11 \)) (Figure 4A). In addition, high expression of ALDH1B1 mRNA was also found to be correlated to worse OS in Ade patients (HR 2.03 [1.6–2.59], \( P=5.2e–09 \)) (Figure 4B) but not in SCC patients (HR 1.07 [0.85–1.36], \( P=0.57 \)) (Figure 4C).

Finally, we examined the prognostic value of ALDH1L1 mRNA expression in www.kmplot.com. The desired Affymetrix ID is valid: 205208_at (ALDH1L1). The curves show that ALDH1L1 expression above or below the median does not separate the cases into significantly different prognostic groups in NSCLC patients (HR 1.1 [0.97–1.25], \( P=0.13 \)) (Figure 5A) or Ade patients (HR 1.19 [0.94–1.5], \( P=0.14 \)) (Figure 5B), or SCC patients (HR 0.87 [0.69–1.1], \( P=0.26 \)) (Figure 5C).

To further determine the correlation of individual ALDH1 isoenzymes with other clinicopathological features, we examined the correlation with smoking status (Table 2), clinical stages (Table 3), and chemotherapy (Table 4) of NSCLC patients. As shown in Table 2, all the individual ALDH1s except ALDH1A3 are significantly associated with smoking status of NSCLC patients. As shown in Table 3, all the individual ALDH1s except ALDH1A1 are significantly associated with clinical stages of NSCLC patients. As shown in Table 4, only ALDH1L1 is significantly associated with chemotherapy of NSCLC patients, probably due to relatively limited number of patients.

Discussion

Normal SCs are considered to have ability to undergo self-renewal and multilineage differentiation.\(^{29}\) Accumulating evidence has proposed a model in which tumorigenesis is driven by cancer stem cells (CSCs) that are derived from mutated adult SCs.\(^{30}\) CSCs initiate and drive carcinogenesis and differentiation through the deregulation of the self-renewal process, contribute to tumor cellular heterogeneity, develop into phenotypically diverse cancer cell populations, proliferate extensively, and drive both continuous expansion of malignant cells and resistance to chemotherapy; thus, CSCs represent the apex in the hierarchical model of tumor genesis, heterogeneity, and metastasis.\(^{31–33}\) Many human cancers including NSCLC possess an enhanced tumor-initiating capacity and partially recreate the cellular heterogeneity of the parental tumor. Previous studies have identified CSCs as responsible risk factors for tumor initiation, growth, and metastasis in solid tumors.\(^{34,35}\)

ALDH1 belongs to a family of detoxifying enzymes that convert aldehydes to their corresponding carboxylic acids, and members of this family are present in many types of normal tissues.\(^{36,37}\) Currently, the “gold standard” of the measurement of the activity of ALDH1 in viable cells has been the use of flow cytometry and fluorescent substrates for ALDH1.\(^{10,38,39}\) Jiang et al first observed that the ALDH1-positive lung cancer cells exhibited the important CSC properties: in vitro self-renewal, differentiation, and multidrug resistance capacities, expression of SC marker, in vivo tumor initiation, and occurrence of a heterogeneous population of cancer cells.\(^{14}\) They also found that relatively high ALDH1 protein levels were positively associated with stage and grade of the tumors and inversely related to the patients’ survival, and their data suggested that ALDH1 might be a lung tumor SC marker and a potential prognostic factor and a therapeutic target for efficient treatment of lung cancer.\(^{14}\) A multivariate analysis also identified ALDH1 expression in NSCLC as significantly independent prognostic factors for disease-free survival (\( P=0.045 \)) and indicated that the immunophenotypes of ALDH1 in cancer cells could have prognostic value in patients with NSCLC who are treated with neoadjuvant therapy.\(^{40}\) However, an exploratory and retrospective study in NSCLC indicates that, unlike breast cancer in which ALDH1 expression predicts poor outcome, ALDH1 expression is associated with favorable outcome.\(^{31}\) The prognostic value of ALDH1 isoenzyme mRNA in NSCLC patients was not reported. Using KM plotter, we found that high expression of ALDH1A1 mRNA was correlated to better OS in all NSCLC patients. In addition, high expression of ALDH1A1 mRNA...
Figure 2 The prognostic value of ALDH1A2 expression.

Notes: The desired Affymetrix ID is valid: 207015_s_at (ALDH1A2). (A) Survival curves are plotted for all patients (n=1,926). (B) Survival curves are plotted for adenocarcinoma (n=719). (C) Survival curves are plotted for squamous cell carcinoma (n=525). Data was analyzed using Kaplan Meier Plotter (www.kmplot.com).

Abbreviation: HR, hazard ratio.
Figure 3 The prognostic value of ALDH1A3 expression.

Notes: The desired Affymetrix ID is valid: 203180_at (ALDH1A3). (A) Survival curves are plotted for all patients (n=1,926). (B) Survival curves are plotted for adenocarcinoma (n=719). (C) Survival curves are plotted for squamous cell carcinoma (n=525). Data was analyzed using Kaplan Meier Plotter (www.kmplot.com).

Abbreviation: HR, hazard ratio.
Figure 4 The prognostic value of ALDH1B1 expression.

Notes: The desired Affymetrix ID is valid: 209646_x_at (ALDH1B1). (A) Survival curves are plotted for all patients (n=1,926). (B) Survival curves are plotted for adenocarcinoma (n=719). (C) Survival curves are plotted for squamous cell carcinoma (n=525). Data was analyzed using Kaplan Meier Plotter (www.kmplot.com).

Abbreviation: HR, hazard ratio.
Figure 5 The prognostic value of ALDH1L1 expression.

Notes: The desired Affymetrix iD is valid: 205208_at (ALDH1L1). (A) Survival curves are plotted for all patients (n=1,926). (B) Survival curves are plotted for adenocarcinoma (n=719). (C) Survival curves are plotted for squamous cell carcinoma (n=525). Data was analyzed using Kaplan Meier Plotter (www.kmplot.com).

Abbreviation: HR, hazard ratio.
was also found to be correlated to better OS in Ade patients (HR 0.71 [0.57–0.9], P=0.0044) but not in SCC patients (HR 0.92 [0.72–1.16], P=0.48). High expression of ALDH1A2 and ALDH1B1 mRNA was found to be correlated to worse OS in all NSCLC patients, as well as in Ade, but not in SCC patients. High expression of both ALDH1A3 and ALDH1L1 mRNA was not found to be correlated to OS in all NSCLC patients. Real-time polymerase chain reaction performed on an array of human tissues has shown that ALDH1L2 is expressed in the liver, kidney, pancreas, heart, and brain; no information is available on its expression in lung tissue. No survival information on ALDH1L2 in NSCLC patients is available, probably due to its low expression in lung tissue and NSCLC.

Previous studies have focused on the relationship between the expression of ALDH1A1 protein and the clinicopathologic parameters, including prognosis of tumor patients. In most types of tumors, such as breast cancer, colorectal carcinoma, esophageal SCC, clear cell renal cell carcinoma, gastric cancer, SCC of the head and neck, and urothelial carcinomas of urinary bladder, high expression of ALDH1A1 protein was associated with tumor metastasis and poor prognosis. In contrast to earlier studies, there was also evidence of ALDH1A1 as a marker of astrocytic differentiation during brain development and of better prognosis in patients suffering from primary glioblastoma. Jiang et al reported that isolated lung cancer cells with relatively high ALDH1 activity display in vitro features of CSCs, including capacities for proliferation, self-renewal, and differentiation, resistance to chemotherapy, and expression of CSC surface marker CD133. The ALDH1-positive cells could generate tumors, and expression of ALDH1 is positively correlated to the stage and grade of lung tumors and related to a poor prognosis in the patients with early-stage lung cancer. In another report, the enzymatic activity of human ALDH1A2 and ALDH2 in lung cancer cells is detected by Aldefluor and inhibited by diethylaminobenzaldehyde and has significant effects on cell proliferation and drug resistance. However, Shao et al showed that ALDH1A3 is the predominant ALDH isoenzyme responsible for ALDH1 activity and tumorigenicity in most NSCLCs, and that inhibiting either ALDH1A3 or the signal transducer and activator of transcription (STAT)3 pathway is the potential therapeutic strategy to eliminate the ALDH(+) subpopulation in NSCLCs. Our current study found that unlike breast cancer, ALDH1A1 mRNA in NSCLC is associated with better prognosis. In addition, our current study also supports that ALDH1A2 and ALDH1B1 might be major contributors to the ALDH1 activity in NSCLC, since high expression of ALDH1A2 and ALDH1B1

### Table 2 Correlation of ALDH1 isoenzymes with smoking status of NSCLC patients

<table>
<thead>
<tr>
<th>Isoenzymes</th>
<th>Smoking status</th>
<th>Cases</th>
<th>HR</th>
<th>95% CI</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALDH1A1</td>
<td>Never smoked</td>
<td>205</td>
<td>0.57</td>
<td>0.32–1</td>
<td>0.049</td>
</tr>
<tr>
<td>ALDH1A2</td>
<td>Never smoked</td>
<td>205</td>
<td>0.85</td>
<td>0.69–1.05</td>
<td>0.13</td>
</tr>
<tr>
<td>ALDH1A3</td>
<td>Never smoked</td>
<td>205</td>
<td>0.76</td>
<td>0.43–1.33</td>
<td>0.33</td>
</tr>
<tr>
<td>ALDH1B1</td>
<td>Never smoked</td>
<td>205</td>
<td>0.99</td>
<td>0.81–1.22</td>
<td>0.94</td>
</tr>
<tr>
<td>ALDH1L1</td>
<td>Never smoked</td>
<td>205</td>
<td>2.19</td>
<td>1.22–3.34</td>
<td>0.0073</td>
</tr>
<tr>
<td></td>
<td>Smoked</td>
<td>821</td>
<td>1.16</td>
<td>0.94–1.43</td>
<td>0.16</td>
</tr>
</tbody>
</table>

**Abbreviations:** ALDH1, aldehyde dehydrogenase 1; NSCLC, non-small-cell lung cancer; HR, hazard ratio; CI, confidence interval.

### Table 3 Correlation of ALDH1 isoenzymes with clinical stages of NSCLC patients

<table>
<thead>
<tr>
<th>Isoenzymes</th>
<th>Clinical stages</th>
<th>Cases</th>
<th>HR</th>
<th>95% CI</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALDH1A1</td>
<td>I</td>
<td>578</td>
<td>0.9</td>
<td>0.69–1.18</td>
<td>0.44</td>
</tr>
<tr>
<td>ALDH1A2</td>
<td>II</td>
<td>244</td>
<td>1</td>
<td>0.7–1.45</td>
<td>0.98</td>
</tr>
<tr>
<td>ALDH1A3</td>
<td>III</td>
<td>70</td>
<td>0.99</td>
<td>0.57–1.72</td>
<td>0.97</td>
</tr>
<tr>
<td>ALDH1B1</td>
<td>II</td>
<td>244</td>
<td>1.58</td>
<td>1.1–2.28</td>
<td>0.013</td>
</tr>
<tr>
<td>ALDH1L1</td>
<td>III</td>
<td>70</td>
<td>0.78</td>
<td>0.45–1.35</td>
<td>0.38</td>
</tr>
</tbody>
</table>

**Abbreviations:** ALDH1, aldehyde dehydrogenase 1; NSCLC, non-small-cell lung cancer; HR, hazard ratio; CI, confidence interval.

### Table 4 Correlation of ALDH1 isoenzymes with chemotherapy of NSCLC patients

<table>
<thead>
<tr>
<th>Isoenzymes</th>
<th>Chemotherapy</th>
<th>Cases</th>
<th>HR</th>
<th>95% CI</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALDH1A1</td>
<td>No</td>
<td>310</td>
<td>0.9</td>
<td>0.69–1.18</td>
<td>0.44</td>
</tr>
<tr>
<td>ALDH1A2</td>
<td>Yes</td>
<td>176</td>
<td>0.75</td>
<td>0.53–1.04</td>
<td>0.086</td>
</tr>
<tr>
<td>ALDH1A3</td>
<td>No</td>
<td>310</td>
<td>1.08</td>
<td>0.77–1.5</td>
<td>0.67</td>
</tr>
<tr>
<td>ALDH1B1</td>
<td>No</td>
<td>310</td>
<td>1.18</td>
<td>0.85–1.65</td>
<td>0.32</td>
</tr>
<tr>
<td>ALDH1L1</td>
<td>No</td>
<td>310</td>
<td>1.13</td>
<td>0.75–1.7</td>
<td>0.55</td>
</tr>
</tbody>
</table>

**Abbreviations:** ALDH1, aldehyde dehydrogenase 1; NSCLC, non-small-cell lung cancer; HR, hazard ratio; CI, confidence interval.
mRNA was found to be significantly correlated to worsen OS in all NSCLC patients. Since high-quality commercial antibodies against ALDH1A1 are already available, the analysis of ALDH1A1 expression by immunohistochemistry will also be important for the design of treatment and the assessment of the prognosis of NSCLC patients. Based on our study, ALDH1A2 and ALDH1B1 might be excellent potential drug targets for NSCLC patients.

**Disclosure**
The authors have no conflicts of interest with this work.

**References**
Prognostic values and potential drug targets of ALDH1 isoenzymes


52. Moreb JS, Ucar D, Han S, et al. The enzymatic activity of human aldehyde dehydrogenases 1A2 and 2 (ALDH1A2 and ALDH2) is detected by Aldefluor, inhibited by diethylaminobenzaldehyde and has significant effects on cell proliferation and drug resistance. *Chem Biol Interact*. 2012;195:52–60.