January 2016

Aldehyde Dehydrogenase 1 (aldh1) Genes In Cancer Clinical Prognosis Outcomes

Mengyi Ding
Yale University, mengyi.ding@yale.edu

Aldehyde Dehydrogenase 1 (*ALDH1*) Genes in Cancer Clinical Prognosis Outcomes

Mengyi Ding

M.P.H. Thesis

Class of 2016

Department of Environmental Health Science

Yale University School of Public Health

First Reader: Vasilis Vasiliou, PhD

Second Reader: Feng Dai, PhD

ABSTRACT

Background & Hypothesis

It has become clear that *ALDH1* genes are involved in the pathobiology of various human cancers. Several lines of evidence indicated that *ALDH1* gene expression in tumors may be associated with clinical prognosis outcomes. This hypothesis of our study is that *ALDH1* genes may be used to predict human cancer prognosis. In order to test this hypothesis, a systematic review of published articles, a meta-analysis using Random-Effects model was conducted to assess the association between *ALDH1* genes and clinicopathological features as well as survival outcomes.

Methods

Using PubMed, EMBASE and Web of Science, this study identified original English researches targeted for association between *ALDH1* genes and cancer prognosis for more than 20 patients during survival analysis follow-up. This meta-analysis included original studies that evaluated a major clinical outcomes (overall survival, disease progression, recurrence, and metastasis) in agnostic format for a variety of cancer types and *ALDH1* genes. Association of *ALDH1* expression and clinicopathological outcomes were evaluated using the Review Manager 5.3 software.

Results

One hundred and twenty one original researches were eligible for inclusion in this meta-analysis. *ALDH1* expression was significantly associated with poor overall survival of breast cancer, colon cancer, non-small cell lung cancer, and ovarian cancer. *ALDH1* expression was also associated with poor prognosis of disease-free survival of breast cancer, non-small cell lung cancer, and rectal cancer. This meta-analysis showed no association of *ALDH1* expression with prognosis of esophageal squamous carcinoma and Head and Neck Squamous Cell Carcinoma (HNSCC).

Conclusion

Expression of *ALDH1* genes is associated with poor prognosis of breast cancer, colon cancer, non-small cell lung cancer, ovarian cancer and rectal cancer.

Background and Rationale

The human aldehyde dehydrogenase (ALDH) gene superfamily comprises nineteen genes that are classified into eleven families and four subfamilies family. The ALDH genes encode proteins (isozymes) that are catalytically-active, although some ALDH proteins appear to be catalytically-inactive. The ALDH gene family also play an important physiological role in encoding proteins, leading to the formation of either catalytically-active or catalytically-inactive proteins [1]. ALDH isoenzymes participate in numerous biological processes mainly by catalyzing the oxidation of a wide spectrum of aldehyde to carboxylic acids [2]. Aldehydes are highly reactive molecules that are generated during the metabolism of either endogenous (e.g., amino acids, neurotransmitters, and carbohydrates) or exogenous (cigarette smoke, food) agents [3]. The ALDH proteins ubiquitously exist in nearly all subcellular tissues [4], with the majority of the ALDHs broadly distributed in tissues and hence display distinct substrate specificity [5].

Historically the *ALDH1* and *ALDH2* genes are the most commonly studied enzymes that are responsible for aldehyde oxidation and the enzymes have the highest concentration in the liver [6]. The significance of ALDHs in physiological processes is based on compelling evidence that mutations and genetic polymorphisms in ALDH genes are associated with phenotypes that extensively exist in humans and rodents [7], such as alcohol-related diseases [8], cancer [9] and other neuro and endocrine diseases [10, 11]. In addition to the association between mutations in ALDH genes and clinical phenotypes, studies with transgenic knockout mice further support the notion that ALDHs are crucial in less-studied physiological functions such as embryogenesis and development [12].

The *ALDH1* family is consists of six enzymes including *ALDH1A1, ALDH1A2, ALDH1A3, ALDH1B1, ALDH1L1 and ALDH1L2*. *ALDH1* genes catalyzes the detoxification of endogenous as well as exogenous aldehydes, oxidization of retinol to synthesize retinoic acid [13]. *ALDH1* is responsible for the oxidation of intracellular aldehydes [14], contributing to the oxidation of retinol to retinoic acid in stem cell differentiation process and has been widely identified as a novel tumor stem-like cell marker in malignancies. *ALDH1* is highly expressed in many stem and progenitor cells in several tissue types,

which is believed to play a role in cancer [15]. As stem cells can asymmetrically divide into stem cell or progenitor cell, and cancer is often regarded as uncontrolled proliferation of such stem cell [16]. Hence *ALDH1* may be a biomarker of stemness. To date, *ALDH1* activity has been used as a stem cell marker for head and neck squamous cell carcinoma [17], lung cancer [18], prostate cancer [19], pancreas cancer [20] and breast cancer [21]. *ALDH1*-positive cells can generate tumors from xenotransplantation situation, thus *ALDH1* might be used to describe the long-time-sought cancer stem cells and even cancer prognosis [22]. According to Ginestier [23], the expression of *ALDH1* in tumor cells can disclose an earlier phase of progenitor cells. In addition, *ALDH1*-positive tumor cells may have inherited aggressive properties including ability to self-renew, high proliferation potential, and resistance to damaging agents. *ALDH1* expression is associated with self-renewal of normal cells and can be a predictor of poor prognosis among cancer patients [24]. The molecular level of *ALDH1* researches can be transferred to practical utility of *ALDH1* in clinical diagnosis and prognosis.

The primary function of *ALDH1A1* concerns with encoding a homotetramer, which are ubiquitously distributed in epithelium of various organs such as testis, brain, eye lens, liver, kidney, lung and retina [25]. Recent studies suggest that *ALDH1A1* may play an important role in cancer therapeutics prognosis effect and mechanism [26], which may result from a decrease in effectiveness of anticancer drugs. Because *ALDH1A1* can detoxify major active aldehyde metabolites especially in breast cancer, *ALDH1A1* expression in the breast is associated with unfavorable clinical outcomes [21]. So breast cancer patients who expressed lower *ALDH1A1* expression status were likely to respond to CP-based treatment significantly more compared to those who have higher *ALDH1A1* expression level. Previous studies also indicated that *ALDH1A1* may be a predictor of the drug's therapeutic effectiveness among non-small cell lung cancer patients [27]. *ALDH1A1* is not only a potential marker of cancer stem cells, but also involves in the formation of tumor-initiating cells in ovarian tissues [28]. Besides this, a variety of non-cancerous cells including hematopoietic progenitor cells can express higher *ALDH1A1* levels [4]. Original researches also indicated *ALDH1* positive status is associated with epithelial-mesenchymal transition, the process of which is considered to be prerogative in tumor metastases [29]. *ALDH1A1* positive cells are enriched in CSCs and are associated with progression of bladder cancer [30]. Also the

ALDH1A1 positive cells define invasive CSCs and *ALDH1A1* predicts poor prognosis in esophageal squamous cell carcinoma [31]. *ALDH1A1* can downregulate certain cancers and has also been shown to interact with certain anticancer drugs including daunorubicin and flavopiridol [32].

Like *ALDH1A1*, the isoenzyme of *ALDH1A2* also has the function of encoding a cytosolic homotetramer, which are expressed and exist in various embryonic tissues [33]. *ALDH1A2* plays a crucial role in regulating RA synthesis, therefore *ALDH1A2* may affect cell growth and differentiation as well as apoptosis, leading to an anticancer effect [4]. Previous studies indicated *ALDH1A2* is a candidate tumor suppressor for prostate cancer [34]. *ALDH1A2* may be an excellent potential target for individualized treatment for gastric cancer patients because *ALDH1A2* demonstrates the association with prognosis [35]. Other studies also suggest the implication of *ALDH1A2* in for non-small cell lung cancer [36]. Low *ALDH1A2* expression is associated with unfavorable recurrence-free survival in non-small cell lung cancer patients [37].

Expression of *ALDH1A3* has been found in a variety of organs such as salivary gland [38], stomach [39], breast [40], kidney [41] and fetal nasal mucosa [42]. *ALDH1A3* has been shown to play a critical role in development of human tissues. Several studies have demonstrated that *ALDH1A3* deficiency may be correlated with prognosis of certain cancer types. Like *ALDH1A1* and *ALDH1A2*, *ALDH1A3* is also a cytosolic homodimer and participates in the synthesis of RA and even embryonic development [43]. *ALDH1A3* can be expressed in various late-stage embryonic and adult rodent tissues. Negative *ALDH1A3* expression in mouse embryos is leading such mice more likely to die from defects in nasal development [44]. Previous studied reported that low *ALDH1A3* expression status may play a critical role in a variety of cancers [4]. *ALDH1A3* expression has been found to be downregulated in human breast cancer MCF-7 cells [45] and upregulated by induction of wild type p53 in cultured human colon cancer cells [46]. *ALDH1A3* has been proposed as a prognostic marker for nonmuscle invasive bladder cancer [47]. *ALDH1A3* expression is methylation-silenced in gastric cancer cells [48] and can be induced by the antitumor agent IL-13 cytotoxin in glioblastoma cells [49], which results in different prognosis outcomes in those tumor cells. *ALDH1A3* belongs to the five candidate genes(*Aldh1a3, Chd2,*

Nipa2, Pcsk6, and Tubgcp5) within a region related to mammary tumorigenesis [50]. In humans, *ALDH1A3* expression may be associated with enhancing malignant behavior of certain cancer types, and *ALDH1A3* might be a new therapeutic target for cancer treatment [51].

ALDH1B1 is a mitochondrial protein expressed and exist extensively in various human tissues including liver, testis, kidney, skeletal muscle, heart, placenta, brain and lung [52]. Recent studies have shown that *ALDH1B1* is involved in the metabolism of the ethanol-derived acetaldehyde and may represent a link between alcohol consumption and diabetes [53]. *ALDH1B1* might be a crucial isozyme for colon cancer tumorigenesis, because *ALDH1B1* can modulate related signal pathways [53]. *ALDH1B1* displays relatively high affinity for acetaldehyde and is believed to play a major role in acetaldehyde oxidation in vivo [54].

The primary function of *ALDH1L1* concerns with catalyzing the formation of tetrahydrofolate from 10-formyltetrahydrofolate [55]. *ALDH1L1* also has the function of cellular proliferation, so *ALDH1L1* might be closely associated with cancer formation and progression. The positive expression of *ALDH1L1* in different cancer cell can result in suppressed cellular proliferation and increased cytotoxicity, which might be attributed to its catalytic function[56]. *ALDH1L1* is significantly downregulated in human liver, lung, prostate, pancreas and ovarian cancers, which may enhance tumor proliferation [57]. There consists with two intronic SNPs in *ALDH1L1* and they are associated with one increasing and one decreasing risk respectively for breast cancer patients, indicating potential influence on breast cancer [58]. What's more, *ALDH1L1* is reported to have protective role for retinal cells [59]. Lower retinal tetrahydrofolate levels can affect *ALDH1L1* in formate oxidation because of its additional role of methanol toxicity [4]. *ALDH1L1* may work as a good target for personal treatment among gastric cancer patients, and *ALDH1L1* may be associated with better overall survival in breast cancer patients [38]. *ALDH1L2* is one of the most recently found isoenzyme in ALDH superfamily and is mainly expressed in spleen and corpus callosum tissue [60]. According to limited researches found, breast cancer treatment that uses anti-inflammatory agent can upregulate *ALDH1L2* expression [61].

The total studies above indicate that enhanced *ALDH1* expression might be a hallmark for cancer stem cells(CSC) [62] and cancer stem cells are believed to possess characteristics of tumorigenesis in particular cancer types [16]. The objective of this study was to systematically investigate the significance of *ALDH1* genes for prognosis and clinical outcomes in cancer patients. Hazard Ratio can represent instantaneous risk over the follow-up time period and can indicate risks for the cumulative follow-up period [63]. The survival analysis involves a series of follow-up time intervals between a fixed starting point and the terminating event and in this study it's the death of cancer patient [64]. The calculation of Hazard Ratio differs from Relative Risk or Odds Ratio in case-control studies is that the time contribution of individual cancer patients vary by the time of termination and their full survival times remain unknown. The Kaplan-Meier survival curve and log-rank test to investigate difference between two groups are examples of univariate analysis method, in which survival is described with respect to the factor while ignoring other variables' influence [65] . It's more common in clinical researches that more than one covariates or variables will exert influence on cancer patients' prognosis, including different genotypes, drug treatment, age, race or combination of these covariates. Therefore, it's more desirable to adjust these covariates while investigating the cancer patients' survival in relation to *ALDH1* status. In previous survival analysis of cancer patients with ALDH, Hazard Ratio is the comparison of death or recurrence corresponding to survival in patients between ALDH-positive and ALDH-negative groups. The Cox Regression model [66] seeks to describe association between the event incidence by hazard function and a set of covariates and the hazard is the instantaneous event probability at a given time or the probability that an individual cancer patient under follow-up time period the event in a time interval centered around that point.

In meta-analysis, the clinical outcomes for time-to-event survival analysis is Hazard Ratio for overall survival or disease-free survival. However, not all studies include individual patient data and carry out the Cox regression analysis for Hazard Ratio. Methods are still available to obtain HRs associated statistics by carefully manipulating the published data that only include Hazard Ratio or Kaplan-Meier curve[67]. It may be possible to extract data from published Kaplan-Meier curves by digitizing data from a number of time intervals on the curves and then pool across these time intervals within a trial to

estimate HR that can represent the entire curve. The practical method proposed by Tierney[67] for incorporating summary hazard ratio into meta-analysis can provide stronger analysis because excluding researches that didn't calculate HR may introduce a bias and may not report the necessary statistical information to allow estimation of entire HRs. More often, the researches can present the outcomes in different ways and by different cut-off point standard.

This study helps fill the gap of the efficacy and time-to-event association between candidate ALDH genes with various cancer types. This study provides a more comprehensive review for *ALDH1* genes and cancer types, the result of which shed interesting light on whether *ALDH1* is a good biomarker for cancer patient prognosis. The association of *ALDH1* and clinicopathological features of cancer patients with corresponding prognostic outcomes remain controversial. The clinical significance of this meta-analysis study is the implementation *ALDH1* in clinical prognosis prediction broadens the research area of the already studied enzyme and pushes forward clinical utility in improving quality of cancer patients.

Methods

The articles for this meta-analysis study is identified by searching the PubMed, EMBASE and Web of Science databases. We searched English language studies that analyzed the associations between *ALDH1* genes expression and prognosis in cancer patients. The search strategy used the clinical queries prognosis filter in databases mentioned above. And the key words for searching are as following: (Prognosis/Broad[filter])AND(*ALDH1* OR *ALDH1A1* OR *ALDH1A2* OR *ALDH1A3* OR *ALDH1B1* OR *ALDH1L1* OR *ALDH1L2*)AND(cancer OR tumor OR neoplas* OR malignan* OR metastat* OR recurrence) AND(Humans[Mesh] AND English[lang]).

The search results were then screened according to the following inclusion criteria:

1) evaluation of the association between *ALDH1* genes expression and overall survival(OS) or disease-free survival(DFS) or other prognostic factors among all types of cancer patients;

2) inclusion of validated data to calculate hazard ratio(HR) with a 95% confidence interval(95%CI) for Overall Survival, Disease-Free Survival, or other prognostic outcomes among cancer patients;

3) inclusion of Kaplan-Meier survival curve to carry out data extraction and calculate unadjusted hazard ratio(HR) based on follow-up information;

4) English language original researches;

5) inclusion of original researches with sufficient sample size of more than twenty patients;

6) articles published as original researches. Reviews were excluded.

The following five criteria were implemented to assess the quality of the original researches:

1) appropriate research design for survival analysis in cancer patients;

2) meeting the inclusion criteria stated in the previous paragraph;

3) clear research objectives for *ALDH1* genes prognosis for different types of cancers;

4) appropriate statistical analysis for Hazard Ratio of clinical outcomes in prognosis prediction,

5) consideration of research bias and standardization.

The following information was extracted from each published researches: title, first author, publication year, key words, number of patients, histopathological cancer type, analysis method applied, cutoff value of *ALDH1* expression, Hazard Ratio, 95%CI for HR, p-value, Kaplan-Meier curve. The original researches for screening and reviewing were before February 29, 2016.

Definitions and Standardizations

This study used a priori defined standardized outcomes and definitions for *ALDH1* status to avoid subjective selection of outcomes and definitions across studies as much as possible. Expression of *ALDH1* is measured by immunohistochemistry as part of the large gene analysis. For immunohistochemistry, we define *ALDH1*-positive status as nuclear staining in tumor cells or at least moderate staining in qualitative scales. The cut-off point may vary across included publications. If different *ALDH1*-positive status were used, the cut-off point is recorded according the original papers. The comparison groups for Hazard Ratio or Relative Risk in the survival analysis were transformed and standardized to *ALDH1*-positive group vs. *ALDH1*-negative group, with negative expression as the reference. The main outcome was Hazard Ratio for Overall Survival or Disease-Free Survival by Cox Regression Analysis. To avoid bias that may arise, if investigators select the follow-up period to report according to the results at each follow-up interval, we standardized definitions to include 24 months of follow-up in all studies. Cox models that allow estimation of a hazard ratio for the entire follow-up survival analysis are not routinely presented in *ALDH1* studies.

Data Extraction

For each individual research, we recorded author name, journal and year of publication, sample size, cancer type, demographics, gene and *ALDH1* status for immunohistochemistry analyses, definition of a *ALDH1*- positive status, cox model analysis used, outcome, HR, 95%CI, p-value during the analysis, overall survival, disease-free survival. For papers didn't include Hazard Ratio, we used Kaplan-Meier survival curve and methods based on Tierney to calculate the Hazard Ratio for the study. The PlotDigitizer was used to extract the data from Kaplan-Meier curve. According to Tierney[67], the Hazard Ratio can be calculated in each time interval and two groups. The data taken into calculation

include following steps:1) event-free at the start of the interval, 2) censored during the interval, 3) at risk during the interval, 4) the number of events during each time interval, 5) O-E, V and HR for each time interval, 6) O-E, V and HR for the entire Kaplan-Meier curve. PlotDigitizer can extract data for numbers event-free at the start of specific time interval for *ALDH1*-positive group and *ALDH1*-negative group. Overall HR and 95%CI can be calculated by the Calculations Spreadsheet provided by Tierney [67].

Statistical Analysis

To determine poor clinical outcome associated with each category of *ALDH1* genes and cancer types, hazard ratio (HR) from time-to-event analyses was extracted along with the 95% confidence interval as well as *ALDH1* expression level. When the 95% confidence interval was not available from original papers, two methods were used to validate the data of such studies. 95%CI can be calculated by HR and p-value, or 95%CI can be approximated from Kaplan-Meier curve. The association of the expression of *ALDH1* and the general prognostic markers is assessed for breast cancer, colon cancer, esophageal squamous cell, head and neck squamous carcinoma, non-small cell lung cancer, ovarian cancer and rectal cancer, and the survival outcomes including overall survival, disease-free survival or other prognosis outcomes. The published data and figures from original papers were used to assess the HR according to the methods described by Parmar et al[68]. Adjusted Hazard Ratio was calculated by multivariate analysis based on Cox Regression Model, and unadjusted Hazard Ratio was calculated by univariate analysis. The Hazard Ratio calculated from Kaplan-Meier curve was unadjusted Hazard Ratio. Both adjusted and unadjusted Hazard Ratio were included in this meta-analysis and were categorized into different subgroups. The original research articles that meet the inclusion criteria of this meta-analysis should be included, even if unadjusted Hazard Ratio for such articles needs to be calculated from Kaplan-Meier curve. To exclude such articles may introduce bias for this meta-analysis. Forest plots of Hazard Ratios of survival analysis were constructed to show the association between *ALDH1* gene expression and overall survival or disease-free survival, the outcomes of which were end points in this meta-analysis. The p-values for Hazard Ratios were two-sided, with the significance cut-off point setting at smaller than 0.05.

The heterogeneity assumption was calculated by using a Q-test, and P-values greater than 0.05 indicated a lack of heterogeneity among studies. Hence the differences for subgroup studies were due to chance and fixed-effect model was used. Otherwise, a random-effect model (DerSimonian-Laird method) was used. I^2 was chosen as the indicator for subgroup heterogeneity study and the cutoff standard for choosing fixed-effect model or random-effect model was based on whether $I^2>50\%$ and whether $p>0.05$. In addition, in order to see whether individual studies will influence on the pooled effect, a sensitivity analysis was performed. The sensitivity analysis sequentially excluded each individual study in each meta-analysis and examined whether the pooled HRs were significantly changed. Funnel plots and Egger's test were constructed to estimate the possible evidence for publication bias. The funnel plots included each individual studies with each point positioning in different X-axis(Hazard Ratio of the study) and Y-axis(standard error of LogHR). The expected findings of smaller studies will distribute randomly centered around the pooled Hazard ratio. And in comparison, larger studies will show tighter cluster around the pooled Hazard Ratio. If there's publication bias, the funnel plots will show an asymmetric distribution. If no significant publication bias exists, all the studies will show a symmetric triangular funnel on funnel plots. All statistics are processed by Review Manager 5.3(The Nordic Cochrane Center, The Cochrane Collaboration, Copenhagen, Denmark). Generic Inverse Variance Method is applied in Review Manager to conduct Hazard Ratio meta-analysis.

Results

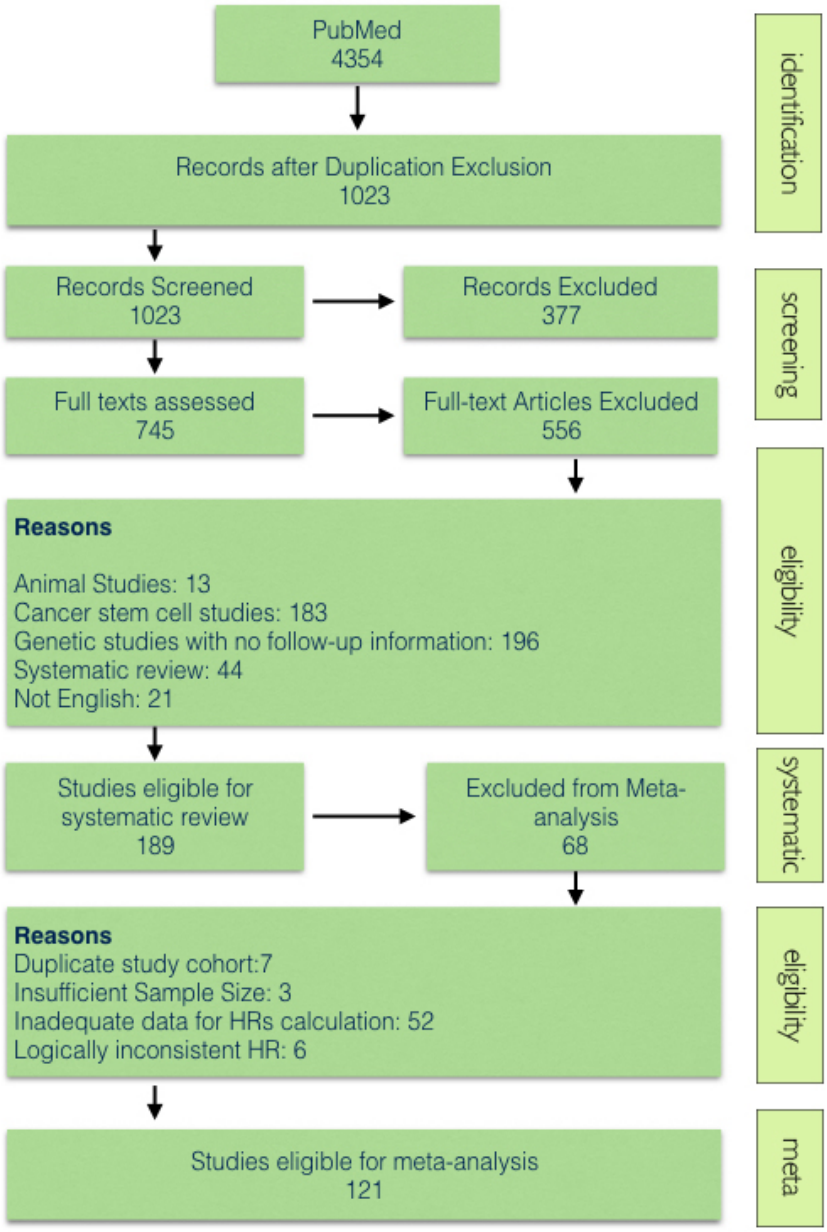

Figure 1. Flow diagram showing the inclusion and exclusion of studies on *ALDH1* genes and all possible cancer prognosis.

Figure 1 and Table 1 show the flow of inclusion and exclusion process of studies on *ALDH1* genes and all possible cancer patients' prognosis. A total number of one hundred and twenty one papers met the inclusion criteria for this meta-analysis after screening abstracts and reviewing original papers. Four thousand three hundred and fifty four papers were identified from PubMed, EMBASE or Web of Science database. After excluding duplicated papers, one thousand and twenty three papers were

screened for abstract. Three hundred and seventy seven papers were excluded based on screening result, because such studies were cellular mechanism study rather than patients survival analysis. Seven hundred and forty five papers were fully read and five hundred and fifty six full-text papers were excluded. The exclusion reasons include: thirteen papers were animal studies with no population data, one hundred and eighty three papers were cancer stem cell studies, one hundred and ninety-six papers lacked follow-up information and only reported gene expression of *ALDH1* genes, forty-four papers were reviews, and twenty one papers were published in other languages. Among the remaining one hundred and eighty nine eligible papers for systematic review, sixty-eight papers were excluded from meta-analysis. The reasons include same study cohorts across researches, insufficient sample size (<20), inadequate data for Hazard Ratio calculation and only reported p-value, logically inconsistent Hazard Ratio. In the end, a total number of one hundred and twenty one original researches were eligible for this meta-analysis.

Table 1. Flow table showing number of inclusion and exclusion of this meta-analysis studies.

	ALDH1	***ALDH1A1***	***ALDH1A2***	***ALDH1A3***	***ALDH1B1***	***ALDH1L1***	***ALDH1L2***
Total published paper (PubMed, EMBase and Web of Science)	3064	860	111	186	63	60	10
Remaining papers after excluding duplicates	595	252	48	68	22	28	10
Remaining papers after screening (abstract only)	396	241	33	46	10	19	0
Remaining papers after full-text review	111	51	5	13	1	8	0
Number of unique papers for each category (after excluding the ones with insufficient data)	68	40	5	10	1	4	0
Total number of unique papers for the meta-analysis				121			

Table 2.1 Eligible studies for *ALDH1* on cancer prognosis (not specifying the *ALDH1* family). First author, cancer type, patient size, cut-off point for ALDH expression, analysis method and outcomes are recorded.

First author	Cancer type	Patient Number	gene	Cut-off point for ALDH expression	Analysis method	Outcomes
Liu[25]	astrocytoma	76	*ALDH1*	0	adjusted	OS, DFS
Goudarzi[69]	astrocytoma	36	*ALDH1*	0	unadjusted	CS
Ito[70]	axillary lymph node metastases	47	*ALDH1*	0	adjusted	DFS
Xu[71]	bladder cancer	227	*ALDH1*	0	adjusted	OS, DFS

Mieog[72]	breast cancer	193	*ALDH1*	0	adjusted	OS, DFS
Mieog[72]	breast cancer	61	*ALDH1*	0	unadjusted	OS, DFS
Yasuyo[73]**	breast cancer	257	*ALDH1*	0	adjusted	OS, DFS
Yasuyo[73]**	breast cancer	106	*ALDH1*	0	adjusted	DFS
Charafe-Jauffret[74]	breast cancer	77	*ALDH1*	0	adjusted	MFS, SS
Sakakibara[75]	breast cancer	115	*ALDH1*	0	adjusted	OS
Tan[76]*	breast cancer	139	*ALDH1*	0	unadjusted	DFS
Zhong[77]	breast cancer	121	*ALDH1*	0	adjusted	DFS
Genestier[28]	breast cancer	577	*ALDH1*	0	adjusted	OS
Kang[78]*	breast cancer	390	*ALDH1*	0	unadjusted	OS, DFS
Bane[79]*	breast cancer	255	*ALDH1*	NA	unadjusted	OS
Ito[73]**	breast cancer	47	*ALDH1*	0	adjusted	DFS
Zheng[80]	breast cancer	65	*ALDH1*	0	adjusted	OS
Nogami[81]	breast cancer	40	*ALDH1*	0	unadjusted	DFS
Neumeister[82]	breast cancer	642	*ALDH1*	NA	adjusted	OS
Huang[83]	breast cancer	552	*ALDH1*	0	unadjusted	DFS
Zhou[84]	breast cancer	61	*ALDH1*	0	adjusted	OS
Yu[85]*	breast cancer	96	*ALDH1*	0	unadjusted	DFS
Morimoto[86]	breast cancer	203	*ALDH1*	0	adjusted	OS
Bednarz-Knoll[87]	breast cancer	330	*ALDH1*	0	adjusted	OS, DFS
Dong[88]	breast cancer	161	*ALDH1*	0	adjusted	OS, DFS
Kim[89]	breast cancer	227	*ALDH1*	0	unadjusted	OS
Brot[90]	breast cancer	140	*ALDH1*	0	adjusted	OS
Hashimoto[91]*	breast cancer	92	*ALDH1*	0	unadjusted	OS, DFS
Lee[92]	breast cancer	184	*ALDH1*	0	adjusted	DFS
Santilli[93]	breast cancer	110	*ALDH1*	NA	adjusted	MFS, SS
Pistelli[94]	breast cancer	81	*ALDH1*	0	unadjusted	OS
Yao[95]	cervical cancer	198	*ALDH1*	0	adjusted	CS
Xie[96]	cervical cancer	52	*ALDH1*	0	adjusted	OS, DFS
Hou[97]	cervical cancer	179	*ALDH1*	NA	adjusted	OS, DFS
Goossens-Beumer[98]	colon cancer	232	*ALDH1*	NA	adjusted	OS, DFS, DSS
Zhou[99]*	colon cancer	60	*ALDH1*	20%	unadjusted	CS
O'Dwyer[100]	colon cancer	28	*ALDH1*	0	unadjusted	CS
Vogler[101]	colon cancer	60	*ALDH1*	0	adjusted	OS
Rahadiani[102]	endometrioid adenocarcinoma	98	*ALDH1*	0	adjusted	OS, DFS

Honing[102]	esophageal squamous cancer	94	*ALDH1*	NA	unadjusted	OS, DFS
Minato[103]**	esophageal squamous cancer	56	*ALDH1*	0	adjusted	DFS
Minato[103]**	esophageal squamous cancer	40	*ALDH1*	0	adjusted	DFS
Minato[103]**	esophageal squamous cancer	56	*ALDH1*	0	adjusted	DFS
Wang[104]	esophageal squamous cancer	79	*ALDH1*	0	adjusted	OS
Hwang[105]*	esophageal squamous cancer	41	*ALDH1*	0%	unadjusted	CS
Ji[106]	esophageal squamous cancer	138	*ALDH1*	NA	adjusted	OS
Ajani[107]	esophageal squamous cancer	167	*ALDH1*	0	unadjusted	OS, DFS
Kim[108]	Eyelid Sebaceous Gland Carcinoma	50	*ALDH1*	NA	adjusted	MFS, SS
Suzuki[109]	liver cancer	49	*ALDH1*	NA	unadjusted	DFS
Morise[110]	lung cancer	105	*ALDH1*	0	adjusted	OS, DFS
Okudela[111]	lung cancer	177	*ALDH1*	85%	adjusted	DFS
Zenke[112]	lung cancer	52	*ALDH1*	NA	adjusted	DFS
Liu[113]	oral cancer	141	*ALDH1*	0	adjusted	OS
Ayub[114]	ovarian cancer	55	*ALDH1*	0	adjusted	OS, DFS
Liebscher[115]	ovarian cancer	112	*ALDH1*	0	adjusted	OS
Kuroda[116]*	ovarian cancer	123	*ALDH1*	0	unadjusted	DFS
Chen[117]	ovarian cancer	80	*ALDH1*	0	adjusted	OS
Mizuno[118]*	ovarian cancer	81	*ALDH1*	10%	adjusted	OS
Huang[119]	ovarian cancer	232	*ALDH1*	NA	adjusted	OS
Wang[120]	ovarian cancer	84	*ALDH1*	50%	adjusted	OS
Chang[121]*	ovarian cancer	442	*ALDH1*	20%	adjusted	OS
Avoranta[122]	rectal cancer	197	*ALDH1*	0	adjusted	DFS
Seung[123]	rectal cancer	51	*ALDH1*	NA	adjusted	OS, DFS
Deng[124]*	rectal cancer	64	*ALDH1*	21%	unadjusted	DFS
Yoon[125]	rectal cancer	145	*ALDH1*	0	adjusted	DFS, CSS
Goossens-Beumer[101]**	rectum cancer	73	*ALDH1*	NA	adjusted	OS, DFS, DSS
Liu[126]	renal pelvis carcinoma	114	*ALDH1*	0	unadjusted	OS, DFS
Huang[127]	tongue squamous cell carcinoma	66	*ALDH1*	>1	unadjusted	OS
Kitamura[128]	urinary cancer	226	*ALDH1*	0	adjusted	CSS
Wu[129]*	volvar squamous cancer	154	*ALDH1*	0	unadjusted	DFS

NA: not applicable
*: The Hazard Ratio of these articles were calculated from Kaplan-Meier curve.
**: These articles included more than one cohort of analysis.

Table 2.2: Eligible studies for *ALDH1A1, ALDH1A2, ALDH1A3, ALDH1B1 and ALDH1L1* on cancer prognosis. First author, cancer type, patient size, cut-off point for ALDH expression, analysis method and outcomes are recorded.

First author	Cancer type	Patient Number	Gene	Cut-off point for ALDH expression	Analysis method	Outcomes
Su[30]*	bladder cancer	216	*ALDH1A1*	NA	unadjusted	OS
Khoury[130]	breast cancer	513	*ALDH1A1*	NA	adjusted	OS, DFS
Liu[131]	breast cancer	596	*ALDH1A1*	0	adjusted	OS, DFS
Wei[132]	breast cancer	92	*ALDH1A1*	NA	adjusted	OS, DFS
Ali[133]	breast cancer	2392	*ALDH1A1*	4	adjusted	OS
Wu[134]	breast cancer	3455	*ALDH1A1*	NA	unadjusted	OS
Zhong[134]	breast cancer	147	*ALDH1A1*	0	adjusted	DFS
Zhou[135]	breast cancer	119	*ALDH1A1*	10%	adjusted	OS
Sjöström[136]	breast cancer	426	*ALDH1A1*	10%	adjusted	DFS
Kahlert[137]	colon and rectal cancer	996	*ALDH1A1*	NS	unadjusted	OS, DFS
Xu[138]	colon cancer	107	*ALDH1A1*	>1	adjusted	OS
Yang[31]*	esophageal squamous carcinoma	134	*ALDH1A1*	NA	unadjusted	OS
Li[139]	gastric cancer	216	*ALDH1A1*	0	adjusted	OS, DFS
Shen[35]	gastric canceer	876	*ALDH1A1*	0	unadjusted	OS
Adam[140]	glioblastoma	93	*ALDH1A1*	0	adjusted	OS
Xu[141]	glioma	237	*ALDH1A1*	5	adjusted	OS
Qian[142]	HNSCC	81	*ALDH1A1*	0	adjusted	DFS
Koukourakis[143]	HNSCC	74	*ALDH1A1*	5%	adjusted	DFS
Xu[144]*	HNSCC	96	*ALDH1A1*	1.3	unadjusted	OS, DFS
Leinung[145]*	HNSCC	48	*ALDH1A1*	0	unadjusted	OS
Martin[146]	larygeal cancer	84	*ALDH1A1*	0%	adjusted	DFS
Tanaka[147]	liver cancer	60	*ALDH1A1*	>1	adjusted	DFS
Jiang[148]	lung cancer	303	*ALDH1A1*	10%	unadjusted	OS
Li[149]	lung cancer	179	*ALDH1A1*	0	adjusted	OS
Sullivan[150]	lung cancer	282	*ALDH1A1*	NA	unadjusted	OS
Shimada[151]	lung cancer	103	*ALDH1A1*	5%	adjusted	OS
		134		0	unadjusted	DFS
Dimou[152]**	lung cancer	296	*ALDH1A1*	0	unadjusted	DFS
You[36]**	lung cancer	1926	*ALDH1A1*	NA	unadjusted	OS
Alamgeer[153]	lung cancer	205	*ALDH1A1*	NA	adjusted	OS, DFS
Gao[154]	lung cancer	133	*ALDH1A1*	0	adjusted	OS
Kaminagakura[155]**	oral squamous cell carcinoma	100	*ALDH1A1*	10%	unadjusted	DFS
Ishiguro[156]	ovarian cancer	90	*ALDH1A1*	NA	unadjusted	OS, DFS

Chui[157]	ovarian cancer	558	*ALDH1A1*	0	unadjusted	OS
Kahlert[158]	pancreatic cancer	97	*ALDH1A1*	4	adjusted	OS
Xing[159]	papillary thyroid carcinoma	247	*ALDH1A1*	NA	adjusted	DFS
Li[160]	prostate cancer	163	*ALDH1A1*	10%	adjusted	OS, CSS
Magnen[161]*	prostate cancer	85	*ALDH1A1*	NA	unadjusted	OS
Sung[162]*	pulmonary adenocarcinoma	97	*ALDH1A1*	NA	unadjusted	OS, DFS
Wang[163]	renal cancer	95	*ALDH1A1*	NA	unadjusted	OS, DFS
Aguilera[164]*	Sporadic colorectal cancer	699	*ALDH1A1*	0	unadjusted	OS
Wu[138]**	breast cancer	3455	*ALDH1A2*	NA	unadjusted	OS
Shen[35]	gastric canceer	876	*ALDH1A2*	0	unadjusted	OS
Seidensaal[165]	HNSCC	101	*ALDH1A2*	NA	adjusted	OS, DFS
You[38]**	non-small cell lung cancer	1926	*ALDH1A2*	NA	unadjusted	OS
Kostareli[166]*	oropharyngeal squamous cancer	115	*ALDH1A2*	NA	unadjusted	OS, DFS
Kim[47]*	bladder cancer	163	*ALDH1A3*	0%	adjusted	DFS
Marcato[167]	breast cancer	176	*ALDH1A3*	NA	adjusted	OS
Jiang[168]	breast cancer	144	*ALDH1A3*	0%	adjusted	OS
Liu[131]	breast cancer	596	*ALDH1A3*	0	adjusted	OS, DFS
Wu[138]	breast cancer	3455	*ALDH1A3*	NA	unadjusted	OS
Qiu[169]*	breast cancer	125	*ALDH1A3*	0	unadjusted	OS
Shen[37]	gastric canceer	876	*ALDH1A3*	0	unadjusted	OS
		177		NA	unadjusted	OS
Chen[170]*	glioma	443	*ALDH1A3*	NA	unadjusted	OS
You[38]**	non-small cell lung cancer	1926	*ALDH1A3*	NA	unadjusted	OS
		46		0	adjusted	OS
Casanova-Salas[171]	prostate cancer	80	*ALDH1A3*	0	adjusted	OS
You[38]**	lung	1926	*ALDH1B1*	NA	unadjusted	OS
Wu[138]**	breast cancer	3455	*ALDH1L1*	NA	unadjusted	OS
Shen[37]**	gastric canceer	876	*ALDH1L1*	0	unadjusted	OS
Chen[172]	liver cancer	112	*ALDH1L1*	0	adjusted	OS
You[38]**	lung	1926	*ALDH1L1*	NA	unadjusted	OS

NA: not applicable
*: The Hazard Ratio of these articles were calculated from Kaplan-Meier curve.
**: These articles included more than one cohort of analysis.

Table 2.1 listed the characteristics of eligible studies to examine association of *ALDH1* genes and cancer prognosis. Sixty eight papers were eligible and it covers nineteen different cancer types. This sixty eight papers didn't specify the *ALDH1* genes. Through the inclusion flow process, we can find the majority of original researches didn't specify *ALDH1* genes isozymes. This might be due to the limitation of experiment design, the budget for identifying gene marker. Some paper incorporated *ALDH1* as ALDH1A1, perhaps ALDH1A1 was the main and largest gene type for *ALDH1* genes. However, in this meta-analysis, we separate the original researches of *ALDH1* as a specific subgroup. Table 2.2 listed the characteristics of eligible studies for ALDH1A1, ALDH1A2, ALDH1A3, ALDH1B1, and ADLH1L1. Forty papers were available for ALDH1A1 and prognosis studies and they covered eighteen cancer types. Five papers examined prognosis effect of ALDH1A2 and five different cancers. Ten papers targeted on ALDH1A3 and six cancer types were studied. Only one paper was available for ALDH1B1. Four papers studied ALDH1L1 and four cancer types. Two papers [38, 139] categorized *ALDH1* genes into each specific subgroups, thus records were repeated for available study number and patient number included.

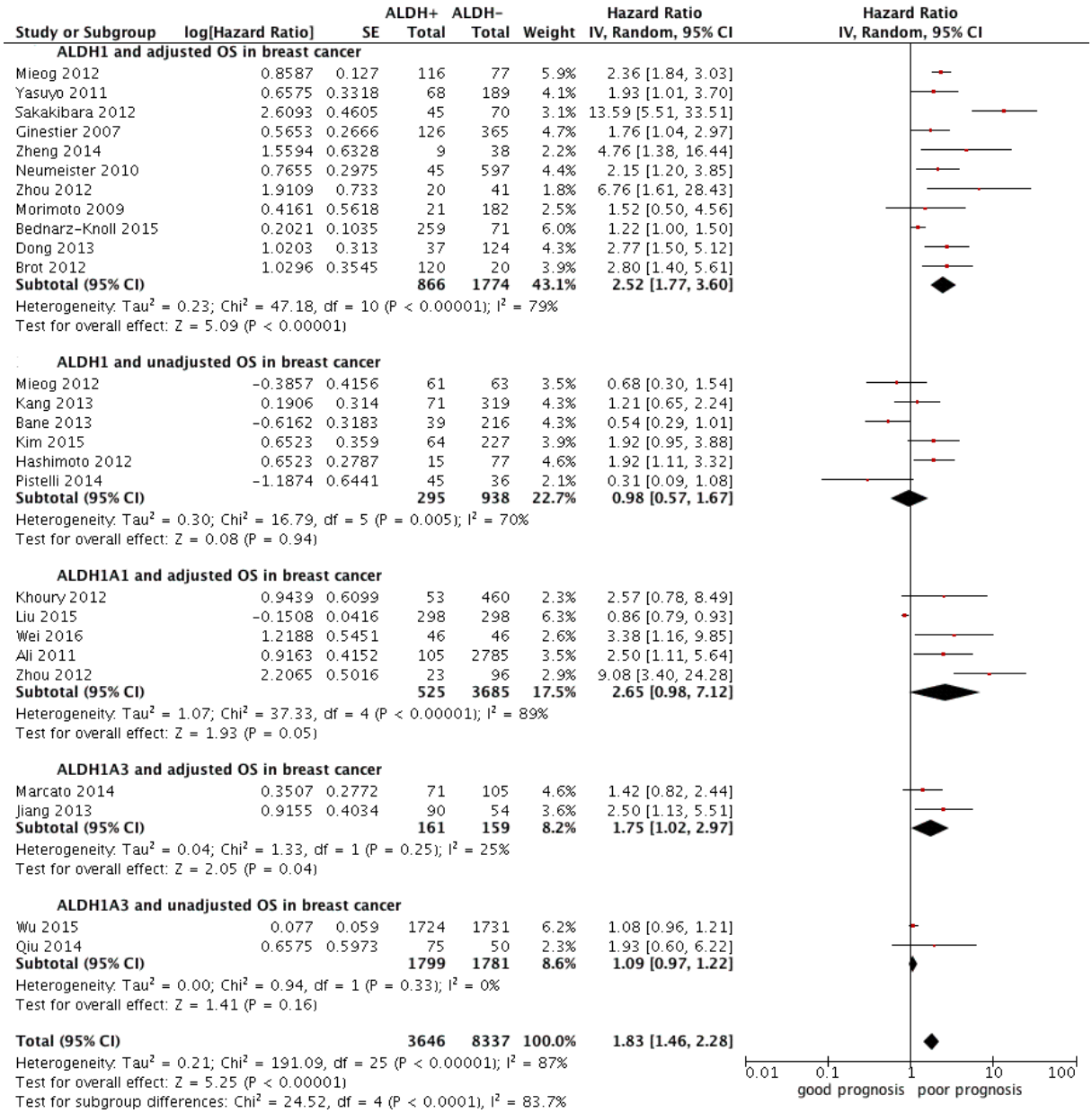

Figure 2.1 Association between *ALDH1* genes and Overall Survival for breast cancer

Figure 2.1 shows the results of the meta-analysis of *ALDH1*, *ALDH1A1, ALDH1A3* expression in prognosis of overall survival in breast cancer patients, for which sixteen papers were available for association with *ALDH1*, six papers were available for association with *ALDH1A1* and five papers were available for association with *ALDH1A3*. A total number of 11,983 breast cancer patients were identified by three *ALDH1* genes and were evaluated for overall survival as a clinical outcome. The pooled hazard ratio for overall survival using Random Effect Model between *ALDH1* genes positive patients and *ALDH1* genes negative patients is 1.83, with 95%CI of (1.46, 2.28), which suggested *ALDH1* genes were a significant poor prognosis predictor for overall survival in breast cancer

patients($p<0.001$). *ALDH1* genes show different prognosis effect when the HR is calculated by multivariate analysis of Cox Regression model with controlling covariates and univariate analysis without controlling covariates, because adjusted HRs are less biased. The Hazard Ratio for adjusted Overall Survival is 2.52(95%CI: 1.77, 3.60) between breast cancer patients with *ALDH1* positive expression and breast cancer patients with *ALDH1* negative expression ($p<0.001$). In comparison, the Hazard Ratio for unadjusted Overall Survival falls to 0.98(95%CI: 0.57, 1.67) between the *ALDH1* positive and negative caner patients. Similarly, the Hazard Ratio for adjusted Overall Survival is 2.65(95%CI: 0.98, 7.12) for breast cancer patients *ALDH1A1* positive expression compared to patients with negative expression. For *ALDH1A3*, the Hazard Ratio for multivariate Overall Survival is 1.75(95%CI: 1.02, 2.97) for breast cancer patients with positive expression versus patients with negative expression. The Hazard Ratio falls to 1.09(95%CI: 0.97, 1.22) for univariate Overall Survival for breast cancer patients with positive *ALDH1A3* expression and negative *ALDH1A3* expression. The I^2 of Heterogeneity analysis indicated the overall studies and subgroup studies have significant heterogeneity for the association between *ALDH1* genes and breast cancer overall survival ($I^2>50\%$, $p<0.05$), thus random-effect model is applied for this meta-analysis. This forest plot indicated that Overall Survival with adjusted HR of *ALDH1* genes predicts a significant poor prognosis for breast cancer patients, while Overall Survival with unadjusted HR may not be a significant prognosis predictor for breast cancer patients. In sensitivity analysis, each study was sequentially excluded to examine if change in the pooled Hazard Ratio was significant. No significant change was found for meta-analysis between *ALDH1* genes and overall survival in breast cancer.

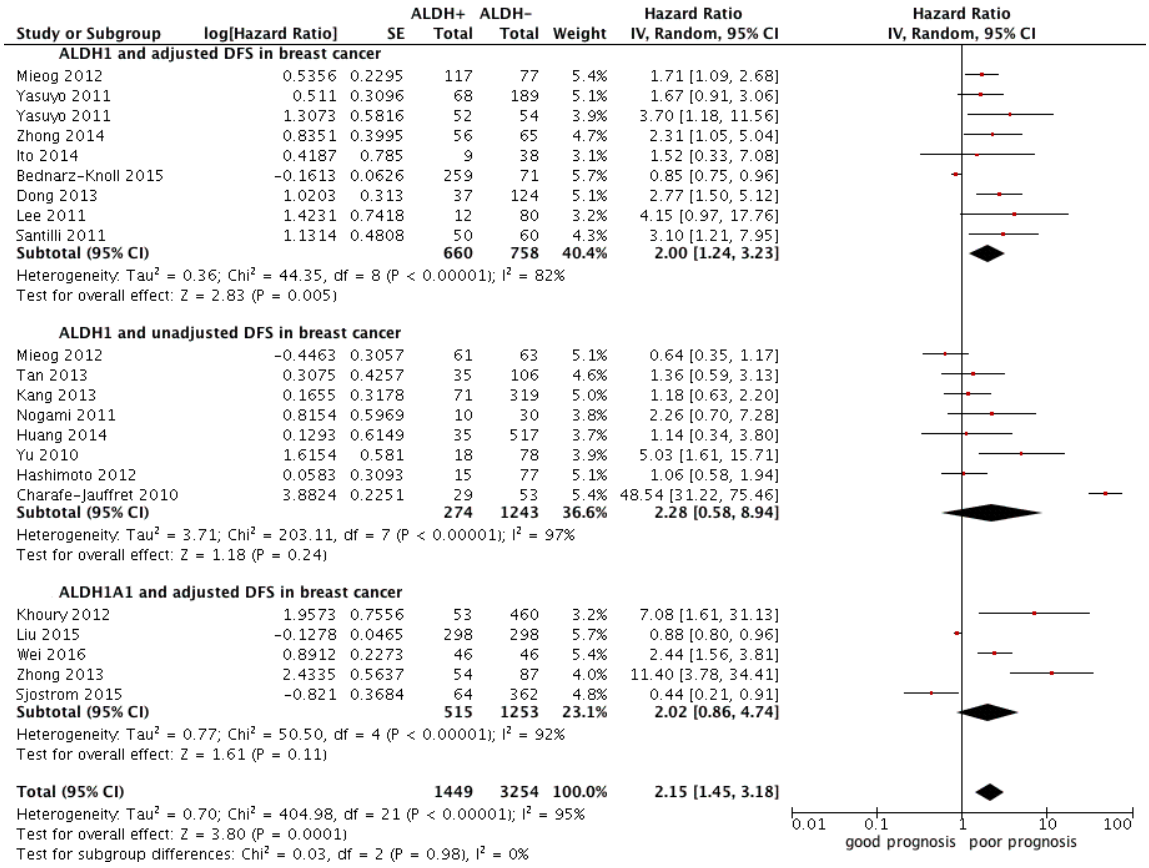

Figure 2.2. Association between *ALDH1* genes and Disease-Free Survival in Breast Cancer

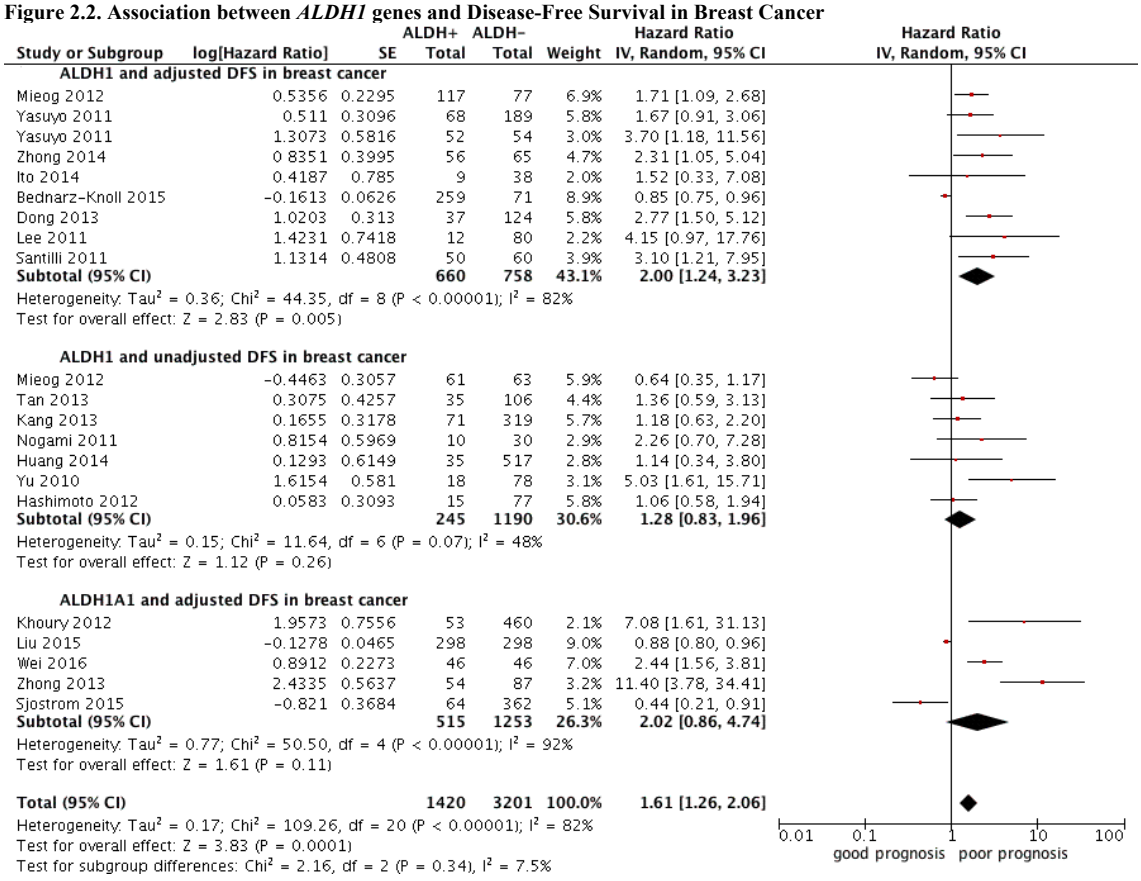

Figure 2.3 Association between *ALDH1* genes and Disease-Free Survival in Breast Cancer after sensitivity analysis.

Figure 2.2 is the forest plot of indicating expression of *ALDH1*, *ALDH1A1, ALDH1A3* in prognosis of disease-free survival (DFS) in breast cancer patients, for which twenty papers were included for this analysis. A total number of 4,073 breast cancer patients were identified by *ALDH1*, *ALDH1A1* and *ALDH1A3* in subgroup analysis. The pooled Hazard Ratio for disease-free survival between breast cancer patients with *ALDH1* genes positive expression and patients with negative expression is 2.15 with 95%CI of (1.45, 3.18) and the poor prognosis for disease-free survival is significant ($p<0.001$). For the *ALDH1* genes subgroup, the hazard ratio for adjusted DFS is 2.00(1.24, 3.23) in breast cancer patients with positive *ALDH1* expression versus patients with negative expression. The hazard ratio for DFS in breast cancer patients was 2.28(0.58, 8.94) when the multivariate analysis is switched to univariate analysis. The adjusted HR for DFS is 2.02(0.86, 4.74) for breast cancer patients with positive *ALDH1A1* expression in comparison to patients with negative *ALDH1A1* expression. Based on the heterogeneity results, the subgroup for *ALDH1* and unadjusted DFS in breast cancer patients showed insignificant heterogeneity ($I^2<50$, $p=0.11$). But the pooled HR analysis and other subgroups show significant heterogeneity ($I^2>50$, $p<0.05$). So random-effect model is used for meta-analysis between *ALDH1* genes and prognosis of disease-free survival in breast cancer patients. This forest plot indicated that adjusted Disease-Free Survival of *ALDH1* genes predicts a significant poor prognosis for breast cancer patients, while unadjusted Disease-Free Survival may not be a significant prognosis predictor for breast cancer patients. In Mieog's study [69], patients were categorized into two groups based on their age (.>65ys) and the disease-free survival were conducted respectively in these two cohorts. When conducting sensitivity analysis for this meta-analysis, we found research by Charafe-Jauffret [71] for association of *ALDH1* and adjusted DFS among breast cancer patients exerted significant changes for the pooled Hazard Ratio. This original research concluded by Kaplan Meier univariate analysis for metastasis free survival of 74 inflammatory breast carcinomas, without controlling the covariates that may contribute to disease-free survival. After deleting the result of this study, the pooled unadjusted Hazard Ratio of disease-free survival related to *ALDH1* genes among breast cancer patients falls to 1.28 (95%CI: 0.83, 1.96). The overall Hazard Ratio of disease-free survival is 1.61 (95%CI: 1.26, 2.06) between breast cancer patients who expressed *ALDH1* genes and who didn't, as showed in Figure 2.3 after conducting sensitivity analysis.

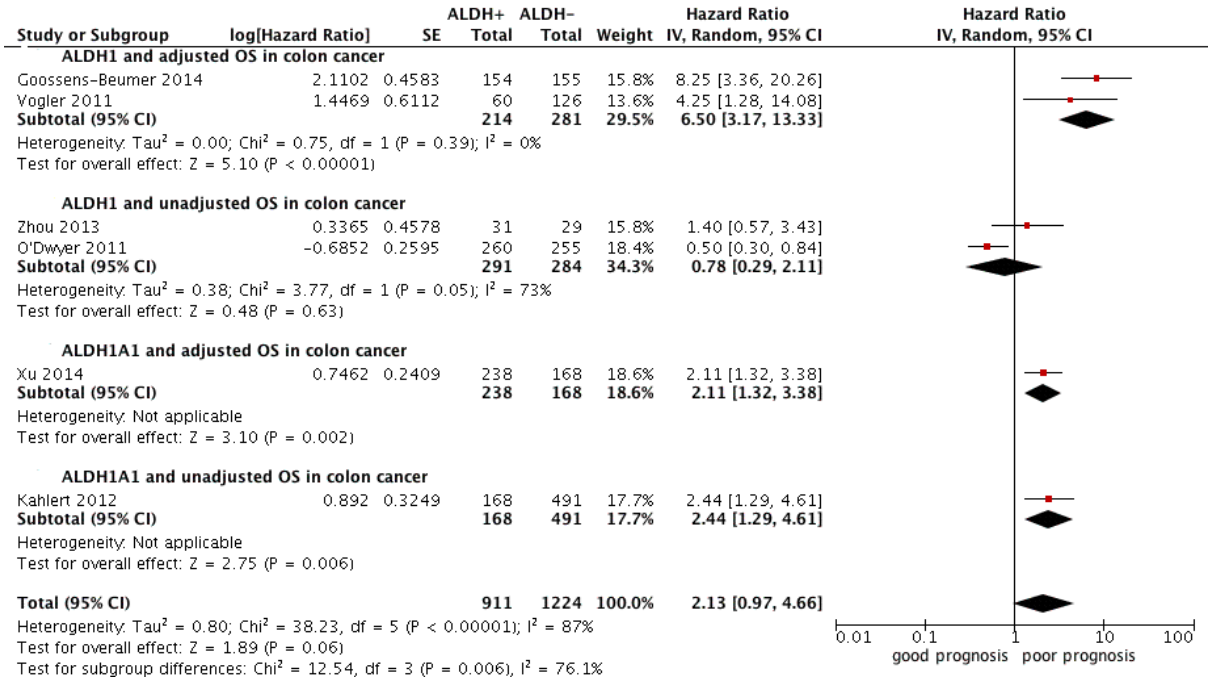

Figure 3. Association between *ALDH1* genes and Overall Survival in Colon Cancer

Figure 3 is the forest plot to show the association of *ALDH1* genes expression and prognosis of overall survival in colon cancer patients. Six studies were included into this meta-analysis and a total number of 2,135 colon cancer patients were followed up for overall survival analysis. The pooled Hazard Ratio for Overall Survival in Colon Cancer between patients with positive expression of *ALDH1* genes and patients with negative expression of *ALDH1* genes is 2.13(95%CI: 0.97, 4.66) and HR of OS is significant predictor for poor prognosis. The adjusted Hazard Ratio of overall survival in colon cancer patients with positive *ALDH1* expression is 6.50(95%CI: 3.17, 13.33) in comparison to colon cancer patients with negative *ALDH1* expression. This Hazard Ratio calculated by Cox Regression Model is significant ($p<0.001$). The unadjusted Hazard Ratio for Overall Survival in colon cancer patients with positive *ALDH1* gene is 0.78(95%CI: 0.29, 2.11) in comparison to patients with negative *ALDH1* expression and this unadjusted HR is not significant ($p=0.631$). The adjusted Hazard Ratio for Overall survival in colon cancer patients with positive *ALDH1A1* expression is 2.11(95%CI: 1.32, 3.38) versus patients with negative *ALDH1A1* expression. And the unadjusted Hazard Ratio for Overall Survival between positive *ALDH1A1* expression patients and negative expression patients is 2.44(95%CI: 1.29, 4.61). For *ALDH1A1* subgroups, only one paper was included specifically for each of the two subgroups. The test for subgroup differences indicated the heterogeneity was significant among the papers in

different subgroups (I^2>50%, p=0.006). Random-effect model was conducted in this meta-analysis to examine association between *ALDH1* genes and Overall Survival in Colon Cancer. This forest plot indicated that *ALDH1* genes for adjusted Overall Survival among colon cancer patients predict a significant poor prognosis, while *ALDH1* genes may not be a significant prognosis predictor for colon cancer patients with unadjusted Overall Survival. In the sensitivity analysis, the original researches from Goossens-Beumer [95] and Zhou [96] would change the pooled Hazard Ratio significantly. But based on the limited availability of researches on colon cancer prognosis, more investigation was required for prognosis effect. The cut-off points in these two studies were 50% and 20% respectively, which means after standardizing expression cut-off points, the hazard ratio can be larger than this estimation.

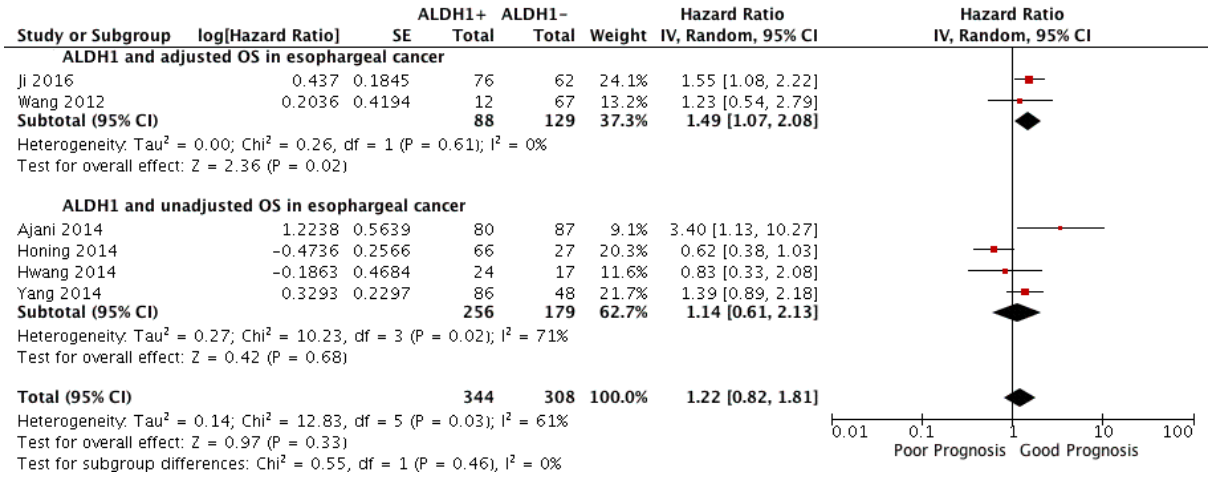

Figure 4.1 Association between *ALDH1* genes and Overall Survival in Esophageal Squamous Carcinoma

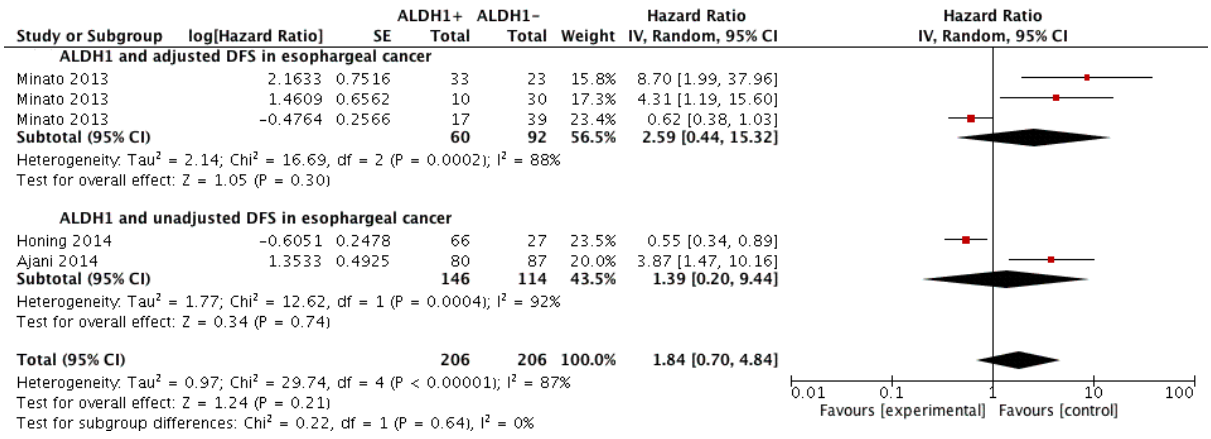

Figure 4.2 Association between *ALDH1* genes and Disease-Free Survival in Esophageal Squamous Carcinoma

Figure 4.1 and Figure 4.2 are the forest plots that examine association between *ALDH1* genes and Overall Survival as well as Disease-Free Survival among patients with Esophageal Squamous Carcinoma. Six papers were available for overall survival meta-analysis for esophageal squamous carcinoma and a total number of 652 patients were followed up for overall survival. The pooled Hazard Ratio of overall survival is 1.22 (95%CI: 0.82, 1.81) between esophageal squamous carcinoma patients who had positive *ALDH1* gene expression and who had negative *ALDH1* gene expression. This poor prognosis prediction of overall survival is not significant (p=0.33). The heterogeneity among papers for overall survival analysis is significant (I^2>50%, p=0.03), thus random-effect model was implemented in this analysis. For subgroup differences test, the heterogeneity among papers that reported overall survival using Cox Regression Model is not significant (I^2<50%, p=0.61) and heterogeneity among papers that reported univariate overall survival is significant(I^2>50%, p=0.02). The pooled adjusted Hazard Ratio of overall survival is 1.49 (95%CI: 1.07, 2.08) between esophageal squamous carcinoma patients who had positive *ALDH1* expression and who had negative *ALDH1* expression. The poor prognosis prediction of *ALDH1* for adjusted overall survival is significant (p=0.02), however only two papers were analyzed for this prognosis. The pooled unadjusted Hazard Ratio of overall survival is 1.14(95%CI: 0.61, 2.13) for esophageal squamous carcinoma between patients with positive *ALDH1* and patients with negative *ALDH1* expression. Four papers and 260 patients were included and the association is not significant (p=0.74). Based on such findings, it's ideal for original researches to carry out multivariate analysis for adjusted Hazard Ratio to evaluate the association between *ALDH1* genes and prognosis for cancer patients. The sensitivity analysis didn't find any individual research would significantly alter the pooled Hazard Ratio.

Three papers and 412 esophageal squamous carcinoma patients were included for meta-analysis of association of *ALDH1* expression and disease-free survival prognosis. The heterogeneity is significant in this meta-analysis (I^2>50%, p<0.001) and significant for both subgroup analysis (p<0.001), thus random-effect model was used for pooled Hazard Ratio analysis. The pooled Hazard Ratio for disease-free survival including adjusted HR and unadjusted HR for esophageal squamous carcinoma is 1.84 (95%CI: 0.70, 4.84) between patients with positive *ALDH1* expression and patients with negative

ALDH1 expression. From the pooled HR, *ALDH1* did not predict a significant poor prognosis for recurrence of esophageal squamous carcinoma (p=0.64). The adjusted disease-free survival of esophageal cancer patients included three different cohorts from one paper included surgery without induction therapy group (OP), surgery with neoadjuvant chemotherapy group (NAC) and initial systemic chemotherapy group (CT). In the OP and NAC groups, multivariate analysis found that *ALDH1* was independently associated with postoperative recurrence and prognosis (OP group, P = 0.004 and 0.016, respectively; NAC group, P = 0.026 and 0.014, respectively). Among the *ALDH1*-negative clinical stage II/III patients, the OP and NAC groups displayed better prognoses than the CT group (P<0.001). However, among the *ALDH1*-positive clinical stage II/III patients, the OP and NAC groups displayed poorer prognoses than the CT group (P = 0.049). For the unadjusted disease-free survival analysis, two papers indicated two different prognosis effect, a significant poor prognosis with Hazard Ratio 3.87 (95%CI: 1.47, 10.16) and a significant good prognosis with Hazard Ratio 0.55 (95%CI: 0.34, 0.89). The patients who showed poor prognosis of disease-free survival with esophageal squamous carcinoma underwent preoperative chemoradiation, so the different prognosis effects might be due to the diversified treatment for patients. The treatment patients received can be an important covariate that will influence the hazard ratio if not controlled by Cox Regression Model. Based on the limited researches found for disease free survival of esophageal squamous carcinoma patients and *ALDH1* prognosis effect, the sensitivity analysis may not fit in this case.

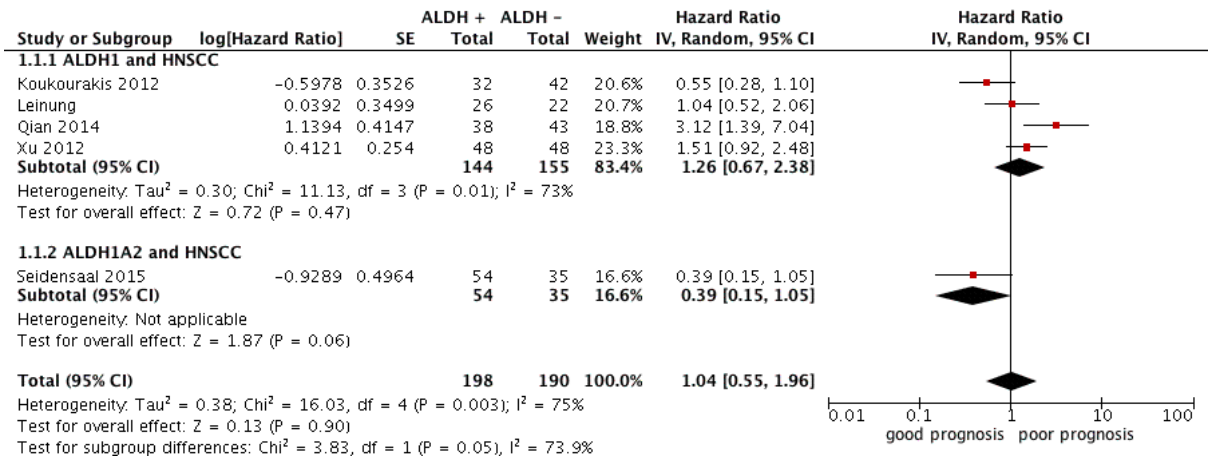

Figure 5. Association between *ALDH1* genes and Disease-Free Survival in Head and Neck Squamous Cell Carcinoma

Five papers were included for meta-analysis of association between *ALDH1* genes expression and disease-free survival in Head and Neck Squamous Cell Carcinoma(HNSCC). A total number of 388 HNSCC patients were included in this association analysis. The pooled Hazard Ratio for disease-free survival in HNSCC patients with positive *ALDH1* genes expression is 1.04(95%CI: 0.55, 1.96) in comparison to patients with negative expression. The Hazard Ratio for disease-free survival included both adjusted HR and unadjusted HR [143]. The heterogeneity of the five papers is significant (I^2>50%, p=0.05), indicating more investigation should be conducted to examine the association of *ALDH1* genes expression and HNSCC prognosis.. The pooled HR didn't show that *ALDH1* genes are significant poor prognosis of disease-free survival among HNSCC patients. This meta-analysis didn't separate univariate disease-free survival and multivariate disease-free survival as subgroups based on papers identified. Based on the limited availability of researches found and the subgroups analysis, sensitivity analysis was not conducted in this meta-analysis.

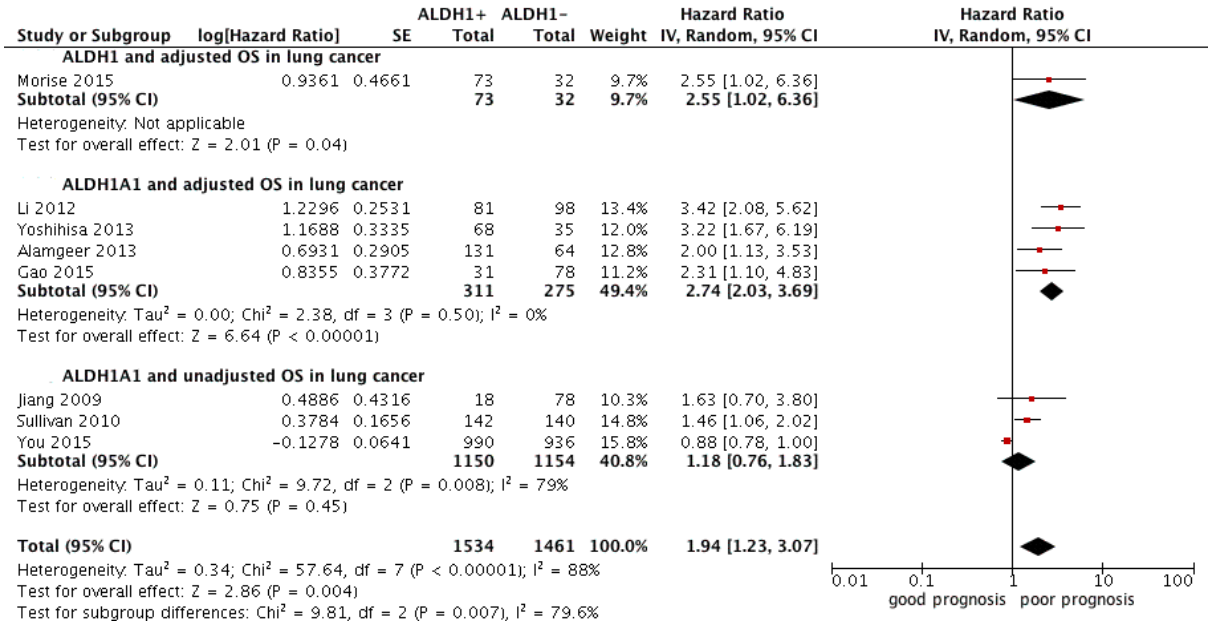

Figure6.1 Association between *ALDH1* genes and Overall Survival in Non-Small Cell Lung Cancer

Figure 6.1 is the forest plot to show the association of *ALDH1* genes expression and overall survival in non-small cell lung cancer(NSCLC) patients. Seven papers were included into this meta-analysis and a total number of 2995 NSCLC patients were followed up for overall survival. The pooled Hazard Ratio for Overall Survival is 1.94 (95%CI: 1.23, 3.07) between NSCLC patients with positive *ALDH1* genes

expression versus patients with negative *ALDH1* genes expression, and the prediction for poor prognosis is significant (p=0.004). The heterogeneity result indicated a significant heterogeneity for subgroup differences (I^2>50%, p<0.01), with the heterogeneity in subgroup of *ALDH1A1* and adjusted OS not significant (I^2=0%, p=0.50), heterogeneity in subgroup of *ALDH1A1* and unadjusted OS significant (I^2>50%, p<0.01). Random-effect model is conducted for this meta-analysis in *ALDH1* genes and OS in NSCLC patients. The adjusted Hazard Ratio by multivariate analysis of overall survival in NSCLC is 2.74 (95%CI: 2.03, 3.69) between patients with positive *ALDH1A1* expression and patients with negative *ALDH1A1* expression (p<0.001). The unadjusted Hazard Ratio by univariate analysis of overall survival in NSCLC is 1.18(95%CI: 1.23, 3.07) between patients with positive *ALDH1A1* expression and patients with negative ALDH1A1 expression (p=0.45). The difference in prediction of NSCLC overall survival can be attributed to other covariates that may contribute to this association. Based on sensitivity analysis, no individual researches significantly altered the pooled Hazard Ratio.

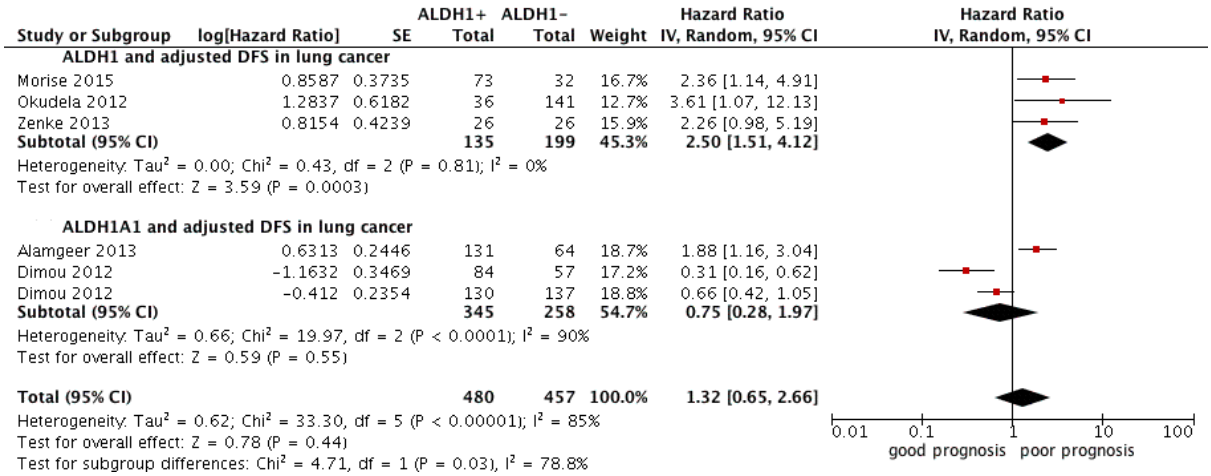

Figure 6.2 Association between *ALDH1* genes and Disease-Free Survival in Non-Small Cell Lung Cancer

Figure 6.2 is the forest plot to show the association of *ALDH1* genes expression and disease-free survival prognosis among NSCLC patients. Five papers were included in this meta-analysis and a total of 937 patients were followed-up for disease-free survival analysis. This meta-analysis was categorized into two subgroups, association of *ALDH1* genes and adjusted Disease-free survival in NSCLC and association of *ALDH1A1* expression and adjusted DFS in NSCLC. The pooled Hazard Ratio of disease-free survival is 1.32(95%CI: 0.65, 2.66) for NSCLC patients with positive *ALDH1* genes expression in

comparison to NSCLC patients with negative *ALDH1* expression. This pooled Hazard Ratio is not significant and the pooled Hazard Ratio between the two subgroups indicated different prognosis effects. The Hazard Ratio for disease-free survival in NSCLC patients by Cox Regression Model is 2.50(95%CI: 1.51, 4.12) between patients who expressed *ALDH1* and patients who didn't express *ALDH1*. The association is significant (p<0.001) and heterogeneity is not significant (p=0.81). The Hazard Ratio for disease-free survival in NSCLC patients by Cox Regression Model is 0.75(95%CI: 0.28, 1.97) between patients who expressed *ALDH1A1* and patients who didn't express *ALDH1A1*. This prognosis association is insignificant (p=0.55) and heterogeneity is significant (p<0.001). Random-Effect model is used to conduct this meta-analysis. The sensitivity analysis didn't find significant change of individual research on pooled Hazard Ratio. The original research by Dimou [151] included two cohort of Yale cohort and Sotirial/Patras cohort, and followed-up the cohort separately and provided two hazard ratio for disease-free survival for NSCLC. To sum up, *ALDH1* might be a poor prognosis for disease-free survival among patients suffering non-small cell lung cancer.

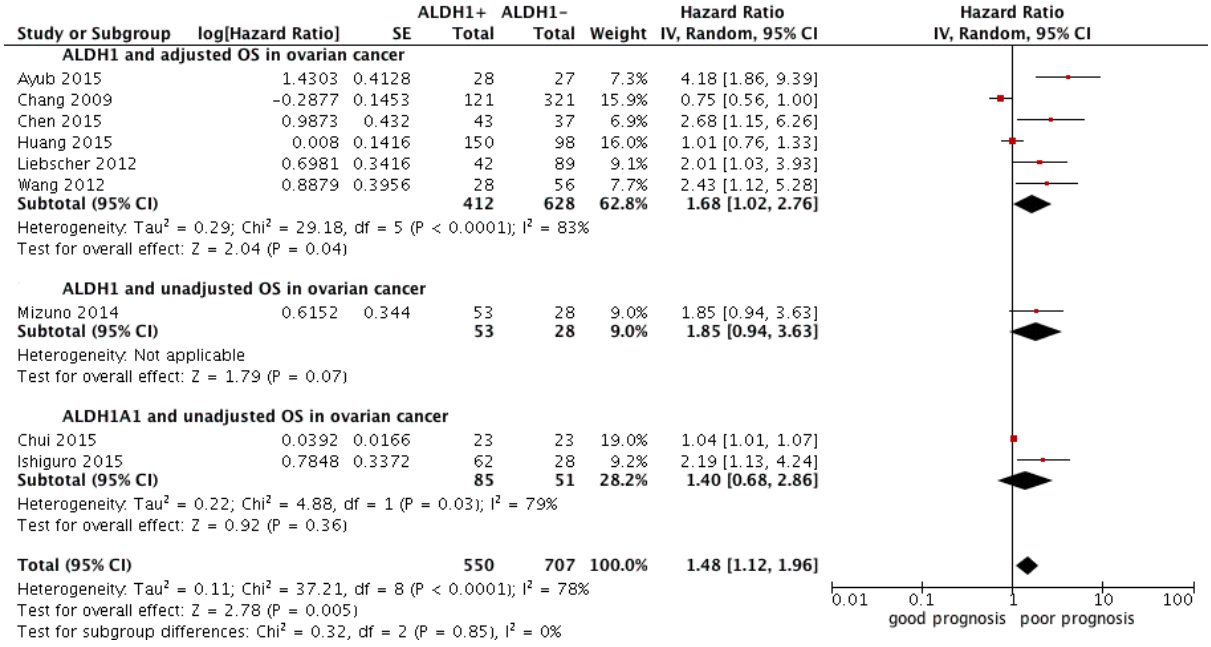

Figure 7. Association between *ALDH1* genes and OS prognosis in ovarian cancer

Figure 7 is the forest plot to examine association between *ALDH1* genes expression and overall survival prognosis in ovarian cancer. Nine papers and a total of 1,257 ovarian cancer patients were included for

this overall survival meta-analysis. Random-effect model was used because the heterogeneity is significant for subgroups ($p<0.05$). The pooled Hazard Ratio for overall survival of ovarian cancer is 1.48 (95%CI: 1.12, 1.96) between patients who expressed *ALDH1* genes and who didn't *ALDH1* genes and the overall survival prognosis is significant ($p<0.001$). For patients who expressed *ALDH1* genes and who didn't express, the hazard ratio for overall survival is 1.68 (95%CI: 1.02, 2.76) and the association is significant ($p=0.04$). The unadjusted Hazard Ratios for overall survival of ovarian cancer between patients who expressed ALDH1 and who didn't, between patients who expressed *ALDH1A1* and who didn't are 1.85(95%CI: 0.94, 3.63) and 1.40(95%CI: 0.68, 2.86) and they are not significant ($p<0.05$). The cut-off points in deciding positive and negative *ALDH1* genes expression are not uniform, with some original researches setting 0%, and some setting 10%, or 50%. The univariate analysis for overall survival among ovarian cancer patients might be influenced by other factors. The sensitivity analysis didn't find a significant change in pooled Hazard Ratio of overall survival for ovarian cancer when taking away each individual study.

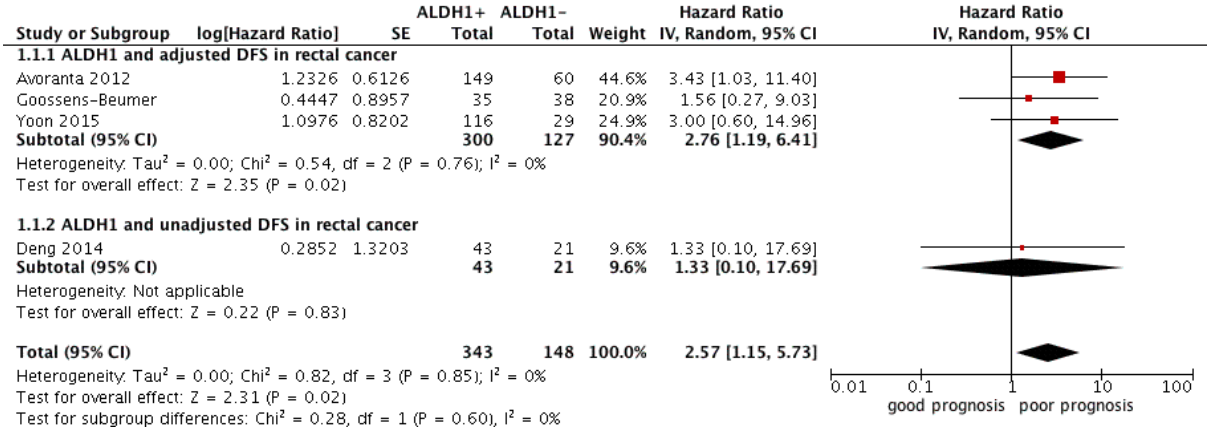

Figure 8. Association between *ALDH1* genes and prognosis in Rectal Cancer

Figure 8 is the forest plot to examine the association of *ALDH1* genes and prognosis in rectal cancer patients. Four papers were included in the meta-analysis and a total of 481 rectal cancer patients were analyzed for disease-free survival prognosis. The heterogeneity of overall researches and subgroup researches was not significant ($p>0.05$), so fixed-model model was used in this meta-analysis. The papers we found for rectal cancer analysis only included *ALDH1* genes expression, and the prognosis

included overall survival and disease-free survival. The pooled Hazard Ratio of disease-free survival for rectal cancer patients is 2.57 (95%CI: 1.15, 5.73) between patients who expressed *ALDH1* genes and who didn't express it. This pooled Hazard Ratio for disease-free survival is significant (p=0.02). In this meta-analysis, the unadjusted DFS hazard ratio was calculated from Kaplan-Meier curve. But in the original paper, multivariate Cox analysis showed postoperative *ALDH1* independently predicted poor prognosis in patients with rectal cancer who received radiochemotherapy (P=0.0095). The pooled disease-free survival of rectal cancer patients indicated that *ALDH1* is a significant poor prognosis predictor for disease-free survival. This result provides a potential prognosis research field for rectal cancer, which hasn't been investigated much by researchers. The sensitivity analysis didn't find significant change in pooled Hazard Ratio of disease free survival when sequentially excluding each study from the meta-analysis.

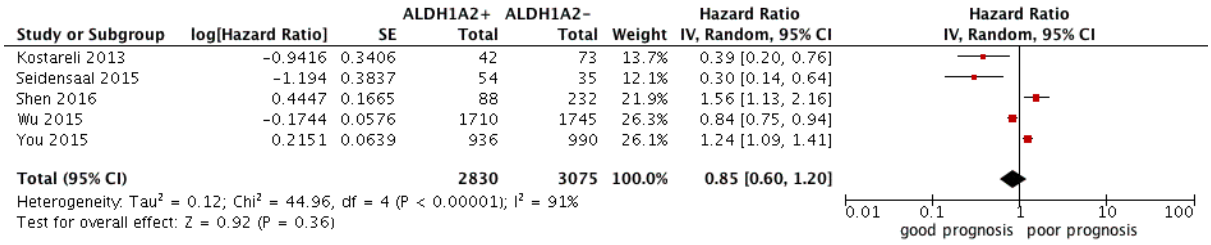

Figure 9.1 Association between *ALDH1A2* and prognosis in different cancers

Figure 9.1 is the forest plot to examine the association of *ALDH1A2* expression status and overall survival in cancer patients. Five papers specified *ALDH1A2* expression in five different cancer types including breast cancer, gastric cancer, head and neck squamous cell carcinoma, non-small cell lung cancer and oropharyngeal squamous cancer and a total number of 5905 patients were analyzed for overall survival. The heterogeneity is significant in this meta-analysis (I^2>50%, p<0.001), and the prognosis effect vary across the different cancer types. Four papers reported overall survival by univariate analysis, and the pooled Hazard Ratio might be influenced by other covariates that were not controlled in the analysis. The pooled Hazard Ratio of overall survival in cancer patients is 0.85 (95%CI: 0.60, 1.20) between patients with positive *ALDH1A2* expression and negative *ALDH1A2* expression and the association is not significant (p=0.36).

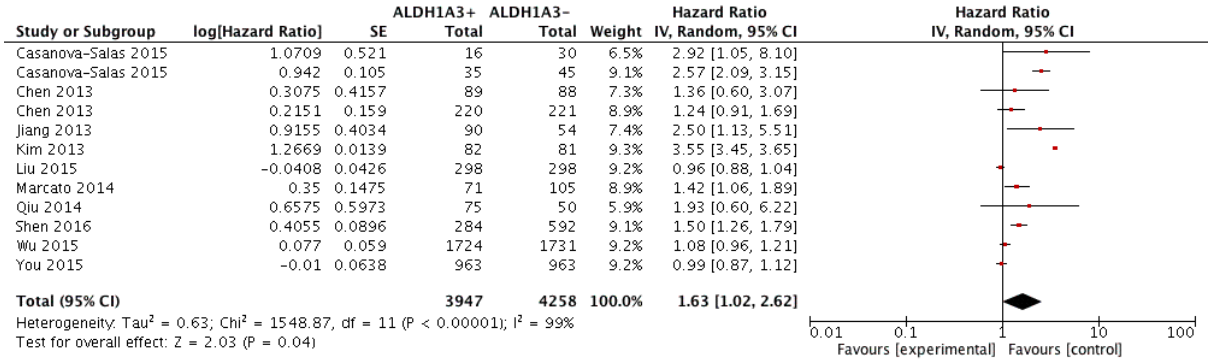

Figure 9.2 Association between *ALDH1A3* and prognosis in different cancers

This forest plot examines the association between overall survival of different cancer patients and *ALDH1A3* expression. Ten papers analyzed *ALDH1A3* expression association and were included in this meta-analysis, with a total number of 8205 patients. The ten papers analyzed six types of cancers including bladder cancer, breast cancer, gastric cancer, glioma, non-small cell lung cancer and prostate cancer. The heterogeneity of the ten papers was significant (I^2>50%, p<0.001) and random-effect model was used to calculate the pooled Hazard Ratio. The pooled Hazard Ratio for overall survival of cancer patients is 1.63 (95%CI: 1.02, 2.62) between patients with positive *ALDH1A3* expression and patients with negative *ALDH1A3* expression. And the pooled Hazard Ratio is statistically significant (p<0.001). Based on the availability of original researches found, subgroups analysis wasn't conducted to examine the prognosis effect for each cancer types. The pooled Hazard Ratio indicated a significant poor prognosis effect in cancer patients who expressed *ALDH1A3*. Therefore, *ALDH1A3* might be a potential poor prognosis predictor for cancer patients.

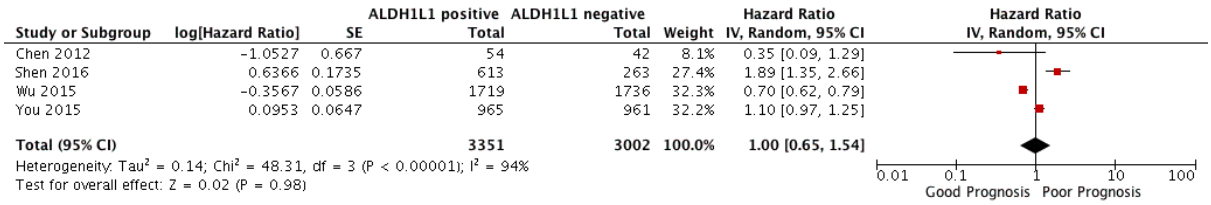

Figure 9.3 Association between *ALDH1L1* and prognosis in different cancers

This forest plot is an attempt to see if there's association between *ALDH1L1* expression and overall survival in cancer patients including liver cancer, gastric cancer, breast cancer and non-small cell lung

cancer. Four papers were included in this analysis with a total number of 6353 patients of four different cancer types. The heterogeneity is significant among the four original researches (I^2>50%, p<0.001), and the different cancer types may lead to various overall survival. A better scrutiny of each Hazard Ratio for overall survival, we can see the prediction of good prognosis or poor prognosis for different cancer types vary. The pooled Hazard Ratio of overall survival between cancer patients with positive *ALDH1L1* expression and patients with negative *ALDH1L1* expression is 1.00 (95%CI: 0.65, 1.54). Even if the pooled HR didn't show prognosis effect, more investigation is worthwhile to disclose the association.

Table 3 Hazard Ratio for association between *ALDH1* gene expression status and prognosis of cancers, specified by Random effect model and Fixed effect model calculation

Levels of Synthesized Information	No. of Studies (No. of Patients)	Random Effects Hazard Ratio Estimates (95%CI)	Fixed Effects Hazard Ratio Estimates (95%CI)
Overall Survival of Breast Cancer	25(11,983)	1.83(1.46, 2.28)	1.09(1.03, 1.16)
ALDH1 and adjusted OS in breast cancer	11(2,640)	2.52(1.77, 3.60)	1.84(1.61, 2.10)
ALDH1 and unadjusted OS in breast cancer	6(1,233)	0.98(0.57, 1.67)	1.09(0.82, 1.44)
ALDH1A1 and adjusted OS in breast cancer	5(4,210)	2.65(0.98, 7.12)	0.89(0.83, 0.97)
ALDH1A3 and adjusted OS in breast cancer	2(320)	1.75(1.02, 2.97)	1.70(1.09, 2.66)
ALDH1A3 and unadjusted OS in breast cancer	2(3,580)	1.09(0.97, 1.22)	1.09(0.97, 1.22)
Disease Free Survival of Breast Cancer	21(4,703)	2.15(1.45, 3.18)	1.06(0.99, 1.13)
ALDH1 and adjusted DFS in breast cancer	8(1,418)	2.00(1.24, 3.23)	1.01(0.90, 1.12)
ALDH1 and unadjusted DFS in breast cancer	8(1,517)	2.28(0.58, 8.94)	3.62(2.83, 4.61)
ALDH1A1 and adjusted DFS in breast cancer	5(1,768)	2.02(0.86, 4.74)	0.93(0.85, 1.01)
Overvall Survival of Colon Cancer	9(2,135)	2.13(0.97, 4.66)	1.64(1.26, 2.15)
Overall Survival of esophageal squamous carcinoma	6(652)	1.22(0.82, 1.81)	1.23(0.99, 1.54)
ALDH1 and adjusted OS in esophageal carcinoma	2(217)	1.49(1.07, 2.08)	1.49(1.07, 2.08)
ALDH1 and unadjusted OS in esophageal carcinoma	4(435)	1.14(0.61, 2.13)	1.05(0.78, 1.42)
Disease Free Survival of Esophageal Carcinoma	3(412)	1.84(0.70, 4.84)	0.90(0.66, 1.23)
Disease Free Survival of HNSCC	5(388)	1.04(0.55, 1.96)	1.12(0.83, 1.52)
Overall Survival of NSCLC	8(2995)	1.94(1.23, 3.07)	1.11(0.99, 1.23)
Disease Free Survival of NSCLC	5(937)	1.32(0.65, 2.66)	1.32(0.65, 2.66)
Overall Survival of Ovarian Cancer	9(1,257)	1.48(1.12, 1.96)	1.10(0.92, 1.32)
Disease Free Survival of Rectal Cancer	4(491)	2.57(1.15, 5.73)	2.57(1.15, 5.73)

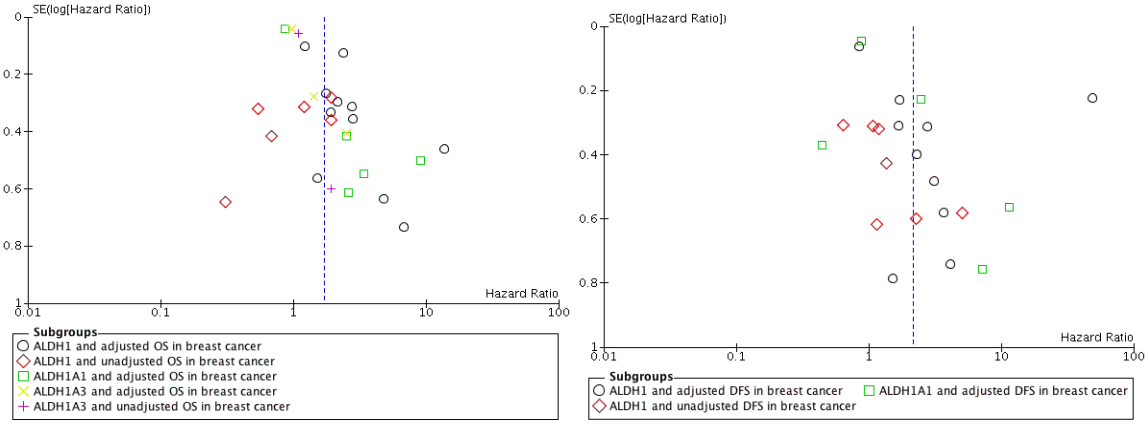

Figure 10.1 Funnel plot of *ALDH1* genes expression and breast cancer overall survival, as a means of assessing publication bias.
Figure 10.2 Funnel plot of *ALDH1* genes expression and breast cancer disease-free survival

Figure 10.1 and Figure 10.2 are funnel plots to check the existence of publication bias for meta-analysis. Y-axis, logHR, represents size of studies and X-axis represents the Hazard Ratio for each individual study. The ideal situation of no publication bias will present large studies plotting near pooled Hazard Ratio and small studies spreading randomly on both sides, which creates a funnel-shaped distribution. Figure 10.1 indicated no obvious evidence of publication bias for association between *ALDH1* and overall survival in breast cancer. Figure 10.2 showed minor evidence of publication bias for association between *ALDH1* and disease-free survival in breast cancer as there was one outlier point in the funnel plot. The findings suggested *ALDH1* genes can be a strong prognosis marker for breast cancer patients.

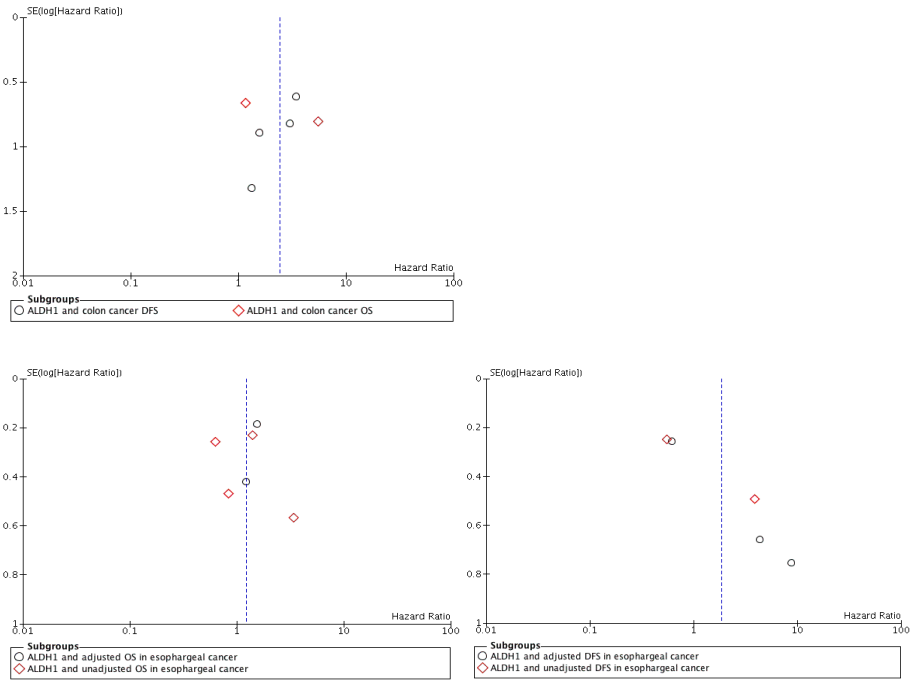

Figure 10.3 Funnel plot of *ALDH1* genes expression and overall survival in colon cancer
Figure 10.4 Funnel plot of *ALDH1* genes expression and overall survival in esophageal squamous carcinoma
Figure 10.5 Funnel plot of *ALDH1* genes expression and disease-free survival in esophageal squamous carcinoma

No significant publication bias was found from the three studies, which suggested the pooled Hazard Ratio of prognosis of *ALDH1* genes can be a strong predictor for colon cancer and esophageal squamous carcinoma.

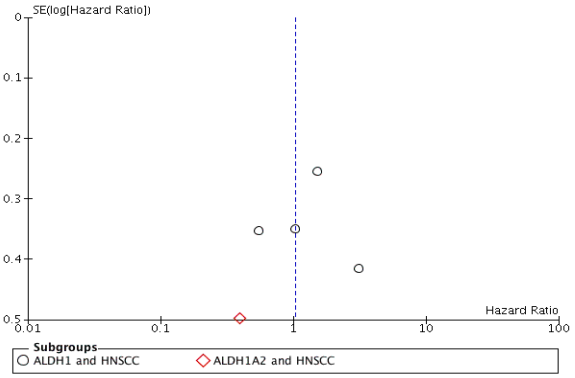

Figure 10.6 Funnel plot of *ALDH1* genes expression and disease-free survival in HNSCC

This funnel plot indicated no obvious publication bias for disease-free survival of head and neck squamous carcinoma cancer with *ALDH1* genes expression.

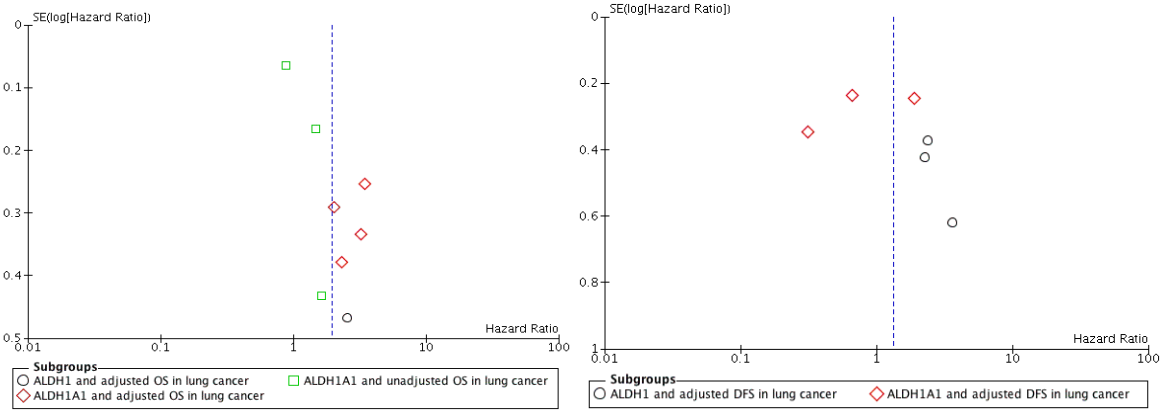

Figure 10.7 Funnel plot of *ALDH1* genes expression and overall survival in non-small cell lung cancer
Figure 10.8 Funnel plot of *ALDH1* genes expression and disease-free survival in non-small cell lung cancer

Figure 10.7 suggested minor evidence for publication bias in checking association of *ALDH1* and overall survival in non-small cell lung cancer. There was outlier study that made the funnel asymmetric.

Figure 10.8 suggested no obvious evidence for publication bias in association of *ALDH1* and disease-free survival in non-small cell lung cancer. The funnel was in a symmetric shape and studies scattered around the pooled Hazard Ratio. But within each subgroup, the studies were not distributed symmetrically.

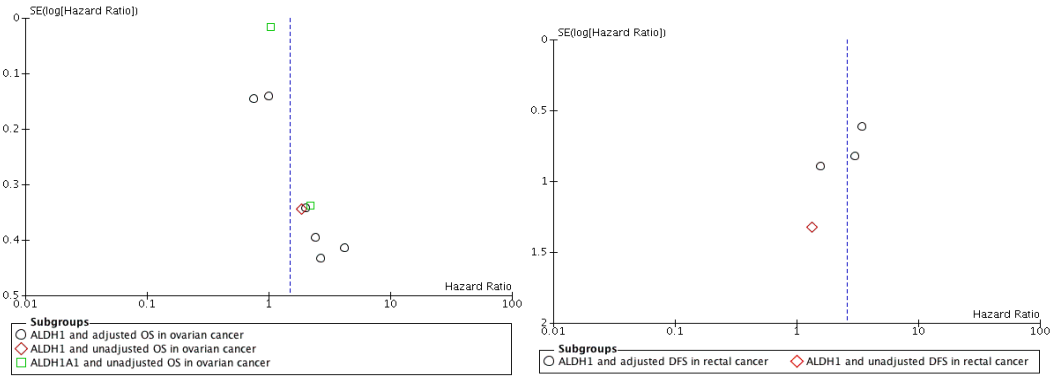

Figure 10.9 Funnel plot of *ALDH1* genes expression and overall survival in ovarian cancer
Figure 10.10 Funnel plot of *ALDH1* genes expression and disease-free survival in rectal cancer

Funnel plots 10.9 and 10.10 didn't show significant publication bias, which were strong evidence of *ALDH1* being a prognosis predictor for patients with ovarian cancer and rectal cancer.

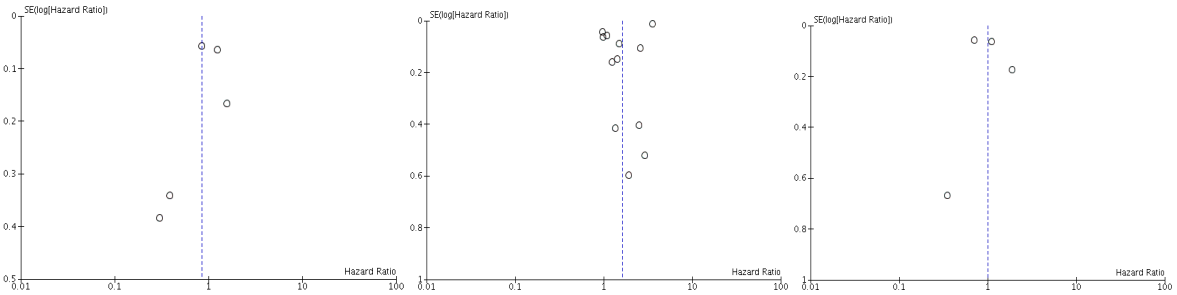

Figure 10.11 Funnel plot of ALDH1A2 and prognosis in cancer patients
Figure 10.12 Funnel plot of ALDH1A3 and prognosis in cancer patients
Figure 10.13 Funnel plot of ALDH1L1 and prognosis in cancer patients

The three funnel plots provided information as whether more investigation was required to examine the prognosis effect of *ALDH1A2, ALDH1A3, and ALDH1L1* subfamilies.

Discussion

In recent studies, a particular sub-group of tumor cells are believed to play a critical role in cancer, which is called cancer stem cells(CSCs) or tumor initiating cells(TICs). The most important characteristics of CSCs are enhanced tumorigenicity and the capacity for self renewal and self differentiation. The ALDH activity has been identified and can separate CSCs from a series of cancer types [23, 173]. The ALDH isozymes actively participate in various physiological responses including drug resistance and RA formation, also ALDH isozymes can protect stem cells from toxic endogenous and exogenous aldehydes. Hence ALDHs can be a potential stem cell marker, or cancer stem cell predictor [174]. Among the nineteen ALDH isoenzymes, *ALDH1A1* was extensively considered to interact with cancer stem cells including breast cancer and non-small cell lung cancer. Studies on murine hematopoietic stem cells, murine progenitor pancreatic cells, and breast cancer stem cells demonstrated that *ALDH1A3* expression may result in aldefluor positivity, which exerted influence in regulation CSCs [63]. Previous researches have indicated the potential of ALDHs to predict cancer patients' outcome because of its role in CSCs.

Unlike Relative Risk or Odds Ratio, Hazard Ratio is the time-to-event analysis instead of event analysis. In order to study the prognosis effect of ALDH in cancer patients, estimation is conducted to evaluate the proportion of cancer patient group who would survive in a given length of time under the same ALDH status from a set of observed survival time interval. And Kaplan-Meier curve is constructed in the already published papers to display the survival functions. The Cox model is used to simultaneously explore the effects of different risk factors related to cancer patients' survival, or different combinations of covariates to cancer patients death [175]. As to the clinical outcomes of overall survival or disease-free survival, the Cox Regression Model is based on the assumption that the predictor variable are constant over time and additive in log scale. The Cox model can allow isolation the ALDH expression status from other contributable variables to survival outcome, by adjusting other covariates effects.

Meta-analysis can provide a more accurate estimation of researched effect, because meta gives weight to each studies based on the sample size and include individual researches into meta-analysis. The

reason why a meta-analysis is conducted on ALDH prognosis effect on cancer patient is that it can address certain practical difficulties that may beset anyone trying to make sense of prediction of ALDH prognosis influence. The validity of this meta-analysis study depends on the quality of the systematic review on the survival analysis considering ALDH expression in cancer patients. This meta-analysis study aims to assess all relevant studies on ALDH and cancer patients' survival analysis, presents a decent summary of existing researches, looks for the presence of heterogeneity and unbiased synthesis among these published studies, and explore the robustness of the main findings using sensitivity analysis. To overcome bias, a rigorous systematic review is conducted to quantitatively evaluate survival outcomes and *ALDH1* expression status. A well-executed meta-analysis requires a complete unbiased collection of all the original studies of acceptable quality that examine prognosis of *ALDH1* on cancer patients. Sensitivity analysis will help explore the effect of excluding various categories of studies and how consistent the results are among studies[176].

This meta-analysis found *ALDH1* genes expression is association with poor overall survival of breast cancer (HR: 1.83) and disease-free survival of breast cancer (HR: 1.61). The significant association provided evidence of ALDH1 families as prognosis predictor for breast cancer patients. The adjusted Hazard Ratio provided an even stronger association for *ALDH1* genes prognosis. This conclusion is consistent with most of published studies, but some studies did conclude different way. The inconsistency with conclusion from Liu [135] might be attributed to specific breast cancer, triple-negative breast cancer, and also because of the experiment design of gene expression from stromal cells or cancer cells, or the analysis with different cohort effect sizes.

This meta-analysis also found *ALDH1* genes expression is associated with poor overall survival in colon cancer (HR: 2.13). This association is not significant, however the adjusted overall survival is significantly associated with *ALDH1* genes expression status. The conclusion that *ALDH1* genes are poor prognosis of colon cancer is consistent with previous researches. For patients with esophageal squamous carcinoma, this meta-analysis discovered the insignificant association of *ALDH1* genes with poor overall survival (HR: 1.22) and with disease-free survival (HR: 1.84). The researches included

proposed controversial prognosis effect of *ALDH1*. The controversy of conclusion may be related to the different treatment received by esophageal squamous carcinoma [106] or age group [105]. The expression of *ALDH1* in esophageal squamous carcinoma patients required further investigation in order to draw meaningful conclusion and check the prognosis effect. A most unique and important finding in this meta-analysis study is that *ALDH1* is a significant poor prognosis of disease-free survival in rectal cancer patients (HR: 2.57). The result is consistent with each of the four original researches found. Researches can be conducted simultaneously for colon cancer and rectal cancer, which is believed to originate from normal stem cells.

Another important finding of this study is the association of *ALDH1* genes expression with poor prognosis in non-small cell lung cancer patients, both for overall survival (HR: 1.94) and disease-free survival (HR: 1.32). And the adjusted HR indicated a stronger association for poor prognosis. Previous studies that concluded in different prognosis effect may not use multivariate analysis by Cox Regression Model as reported in You [38]. The opposite conclusion of good prognosis of disease-free survival may also be due to AQUA score-defined threshold of detecting *ALDH1* genes expression [157]. Because of the limited researches found on *ALDH1A2, ALDH1A3, ALDH1B1*, and *ALDH1L1*, this meta-analysis for non-small cell lung cancer prognosis didn't include results from such *ALDH1* genes. According to You [38], high expression of *ALDH1A2* and *ALDH1B1* was significantly associated with poor overall survival in NSCLC patients. Thus *ALDH1A2* and *ALDH1B1* might be good potential drug targets and overall survival predictor for NSCLC patients.

This meta-analysis also concluded that there's association of *ALDH1* genes especially *ALDH1A1* expression with poor prognosis of overall survival in ovarian cancer patients (HR: 1.48). The significant poor prognosis effect is consistent with most of what original researches found [117, 120]. But some research concluded different way as favorable prognosis of *ALDH1* for ovarian cancer [124]. In Chang's study [124], high levels of *ALDH1* expression was associated with endometrioid adenocarcinoma, early disease stage, complete response to chemotherapy and favorable survival. The cut-off point in determine

high and low *ALDH1* expression was 20%. The prognosis markers for identifying one type cancer cells may not always be useful for predicting other types of cancer cells [177-179].

The forest plots and funnel plots were mapped by Review Manager 5.3. To conduct a meta-analysis for Hazard Ratio in RevMan, logHR and se(logHR) are needed to be transformed from reported Hazard Ratio. Papers that either reported HR with 95%CI or HR with p-value fit the transformation calculation and the calculation was based on Hazard Ratio Meta-analysis Spreadsheet. This spreadsheet was developed by Hans Messersmith using the methods in Parmar [71] in Statistics in Medicine. Unlike other softwares including R, SAS, SPSS, RevMan didn't require the 95%CI for HR reported in the original researches in order to meet the inclusion criteria. Because some papers may not report 95%CI and only reported p-value for HR if the prognosis association was not significant, this transformation provided method to be incorporated into meta-analysis. The method was also proposed by Parmar [71]. STATA and RevMan are useful in processing subgroup meta-analysis, while other softwares were more strict in subgroup meta. Especially RevMan is easier for sensitivity analysis to sequentially exclude each individual research from meta-analysis and repeat the whole process to see if pooled Hazard Ratio will be significantly altered.

The limitations of this meta-analysis study concerns with four parts. In the first place, even if the data extraction method from Kaplan-Meier curve can include more validated researches, the HR calculated this way is unadjusted HR. Other published papers which already include the HR and 95%CI mainly use Cox regression model and calculate adjusted hazard ratio. This study didn't separate adjusted HR and unadjusted HR into different categories. Hence the interpretation of overall HR cannot be arbitrarily concluded as controlling other covariates. The influence of covariates in survival in cancer patients remain unclear, and this can reduce the validation of overall results. There might be a situation when some covariates contribute more to prognosis than *ALDH1* genes expression status and not properly controlling the covariates can affect HR in both directions. Another problem with this data extraction method is the calculation of 95%CI is dependent on sample size. The studies with small sample size tend to have a broader 95%CI range, even if the original paper provided a significant p-value based on

either log-rank test or Cox regression model. There's a tradeoff in using this data extraction method to include validated outcomes, and it can reduce the selection bias from inclusion process while not guarantee the uniform data analysis. In addition, the cut-off points for *ALDH1* genes expression status is not uniform, with some studies deciding positive and negative status based on their specific immunology results. This meta-analysis didn't transform such criteria and standardize it. The choice of cut-off points will also exert influence the final result, especially when it's not merely expression versus non-expression. A higher cut-off point will weaken the calculation of HR and draw it to the direction towards 1.This meta-analysis study didn't cover all the cancer types found from database. Further original researches should be conducted on the prognosis effect and the less-studied cancer types. Next when interpreting the hazard ratio for a survival analysis, it's better to take into consideration of time such as median survival time under scrutiny, comparison of two groups at the time point that half of patients experienced the event. This meta-analysis didn't include the information about time in each study. Finally, the Hazard Ratio meta-analysis using Review Manager needs the log transformation for Hazard Ratio, 95%CI, or p-value. As such calculation is taken into consideration of sample size, some studies may include more than one cohort for prognosis analysis. Thus the calculation by Review Manager may differentiate the original HR and 95%CI provided by the original papers.

Conclusions

ALDH1 genes expression is associated with poor overall survival of breast cancer (HR: 1.83), poor disease-free survival of breast cancer (HR: 1.61), poor overall survival of colon cancer (HR: 2.13), poor overall survival of non-small cell lung cancer (HR: 1.94), poor overall survival of ovarian cancer (HR: 1.48) and poor disease-free survival of rectal cancer (HR: 2.57). This study also found *ALDH1* genes expression is not associated with disease-free survival of non-small cell lung cancer, overall survival and disease-free survival of esophageal squamous carcinoma, as well as disease-free survival of Head and Neck Squamous Carcinoma Cancer. Expression of *ALDH1* genes predicts poor prognosis for breast cancer, colon cancer, non-small cell lung cancer, ovarian cancer and rectal cancer.

Competencies

- Describe the mechanisms of toxicity of biological, chemical, and physical stressors, including absorption, distribution, metabolic transformation, elimination, and genetic susceptibility.
  - This thesis describes aldehyde dehydrogenases family and its use as prognosis factors for certain human cancer types. Unlike mechanistic studies on *ALDH1* genes, this study attempts to discover clinical utility of *ALDH1* genes for cancer patients.
- Review, critique, and evaluate environmental epidemiology research articles.
  - The meta-analysis is a comprehensive review of already published papers on *ALDH1* genes families and human cancers.
- Synthesize information from a variety of environmental health and related studies
  - The meta-analysis analyzed and organized data from each related papers and made a comprehensive summary to assess all possible original researches concerning survival analysis of *ALDH1* genes among cancer patients.
- Use epidemiological, exposure assessment, toxicological and statistical techniques in assessing the risks associated with environmental hazards in the working, residential, and community environment.
  - The statistical analysis helps in discovering the *ALDH1A1, ALDH1A2, ALDH1A3, ALDH1B1, ALDH1L1* level and activity in association with breast cancer, lung cancer and esophageal cancer or other cancer types among humans. It shed light on how enzymes may work as prognosis biomarkers for these cancers.
- Explain the interrelationships among a multitude of factors that can influence a public health problem.
  - The study aims to find out the correlation or association between *ALDH1* genes expression and prognosis for cancer patients. It is more important to improve the life quality of cancer patients, including preventing the recurrence of cancer for humans. By identifying the prognosis effect, this study can disclose the clinical potential of *ALDH1 families* as marker for prognosis prediction. This can provide recommendation for personalized treatment for cancer patients.

Innovation of This Meta-Analysis

- Previous Systematic Review and Meta-Analysis for prognosis effect of *ALDH1* genes on cancer studies didn't provide enough information about categorical difference in six *ALDH1* genes. The majority of papers put the six *ALDH1* genes together without specifying the isozymes. This meta-analysis specified *ALDH1* isozymes into six subtypes as *ALDH1*, *ALDH1A1, ALDH1A2, ALDH1A3, ALDH1B1, ALDH1L1, ALDH1L2* (no prognosis research) and then combine isozymes together and carry out the analysis as ALDH1 family. It's a more comprehensive systematic review compared to targeting at only one isozyme and its prognosis effect.
- The HR calculation based on Kaplan-Meier curve data extraction provides a more validated result for meta-analysis study. Excluding the survival analysis paper that didn't calculate HR can bring in bias for the entire study results. Including the unadjusted calculated HR can give weight to such studies that meet inclusion criteria for this meta-analysis studies. This estimation method extends validated study sample for prognosis effect of *ALDH1* genes.
- This study shed enlightening light on exploration and examination of whether *ALDH1* can be a clinical biomarker in predicting prognosis as well as metastasis. Its clinical significance for survival prediction caters to the trend of personalized medicine and arising genetic sequencing skill utility. The conclusion of the study enriches the clinical utility of *ALDH1* genes, and can work as new mechanism for drug treatment and cancer progression pathway. This meta-analysis also provides basis for prognosis effect of other ALDH genes like *ALDH2*, and the cancer types cover some of the most common one. By categorizing cancer type, the prognosis prediction has even more clinical significance in personalized treatment for cancer patients.

ACKNOWLEDGMENTS

I am grateful to my thesis advisor, Professor Vasilis Vasiliou, whose expertise, profound insight, generous guidance and devotion made it possible for me to work on such an exciting and meaningful topic. This study also inspired me for mapping out my future career path. It was my greatest privilege to have this opportunity to work with him.

I am hugely indebted to Dr. Georgia Charkoftaki for providing her precious and kind advice regarding the topic of my thesis study, and for her timely meeting and reply for my detailed questions throughout my thesis.

I would especially express my gratitude to my thesis advisor, Dr. Feng Dai, for providing me with materials and knowledge pool so that I could continue my thesis study more confidently, for his kind words and suggestions and for your interest in the topic as well as immense support.

I am also grateful to my academic advisor, Professor Yawei Zhang, for your suggestions and support during my master study in Yale School of Public Health. I'm also thankful to Professor Maria Ciarleglio for guidance on survival analysis.

I would also thank my friends, Rui Ye, Weixiong Dang, Dr. Zhuihui, Dr. Yuanli and Half-Moon Kitchen for your inspiration and timely encouragement, for your diversified talents.

I am especially grateful to my parents, for always trusting me and supporting me, for your tolerance and sacrifice, for all the love.

References

1. O'Brien, P.J., A.G. Siraki, and N. Shangari, *Aldehyde sources, metabolism, molecular toxicity mechanisms, and possible effects on human health.* Crit Rev Toxicol, 2005. **35**(7): p. 609-62.
2. Perez-Miller, S.J. and T.D. Hurley, *Coenzyme isomerization is integral to catalysis in aldehyde dehydrogenase.* Biochemistry, 2003. **42**(23): p. 7100-9.
3. Vasiliou, V., A. Pappa, and D.R. Petersen, *Role of aldehyde dehydrogenases in endogenous and xenobiotic metabolism.* Chem Biol Interact, 2000. **129**(1-2): p. 1-19.
4. Marchitti, S.A., et al., *Non-P450 aldehyde oxidizing enzymes: the aldehyde dehydrogenase superfamily.* Expert Opin Drug Metab Toxicol, 2008. **4**(6): p. 697-720.
5. Sladek, N.E., *Human aldehyde dehydrogenases: potential pathological, pharmacological, and toxicological impact.* J Biochem Mol Toxicol, 2003. **17**(1): p. 7-23.
6. Kitagawa, K., et al., *Aldehyde dehydrogenase (ALDH) 2 associates with oxidation of methoxyacetaldehyde; in vitro analysis with liver subcellular fraction derived from human and Aldh2 gene targeting mouse.* FEBS Lett, 2000. **476**(3): p. 306-11.
7. Vasiliou, V. and A. Pappa, *Polymorphisms of human aldehyde dehydrogenases - Consequences for drug metabolism and disease.* Pharmacology, 2000. **61**(3): p. 192-198.
8. Enomoto, N., et al., *Alcoholic Liver-Disease in Heterozygotes of Mutant and Normal Aldehyde Dehydrogenase-2 Genes.* Hepatology, 1991. **13**(6): p. 1071-1075.
9. Yokoyama, A., et al., *Alcohol and aldehyde dehydrogenase gene polymorphisms and oropharyngolaryngeal, esophageal and stomach cancers in Japanese alcoholics.* Carcinogenesis, 2001. **22**(3): p. 433-439.
10. Kamino, K., et al., *Deficiency in mitochondrial aldehyde dehydrogenase increases the risk for late-onset Alzheimer's disease in the Japanese population.* Biochemical and Biophysical Research Communications, 2000. **273**(1): p. 192-196.
11. Baumgartner, M.R., et al., *Hyperammonemia with reduced ornithine, citrulline, arginine and proline: a new inborn error caused by a mutation in the gene encoding Delta(1)-pyrroline-5-carboxylate synthase.* Human Molecular Genetics, 2000. **9**(19): p. 2853-2858.
12. Dupe, V., et al., *A newborn lethal defect due to inactivation of retinaldehyde dehydrogenase type 3 is prevented by maternal retinoic acid treatment.* Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(24): p. 14036-14041.
13. Yoshida, A., et al., *Human aldehyde dehydrogenase gene family.* European Journal of Biochemistry, 1998. **251**(3): p. 549-557.
14. Duester, G., *Families of retinoid dehydrogenases regulating vitamin A function - Production of visual pigment and retinoic acid.* European Journal of Biochemistry, 2000. **267**(14): p. 4315-4324.
15. Burkert, J., N.A. Wright, and M.R. Alison, *Stem cells and cancer: an intimate relationship.* J Pathol, 2006. **209**(3): p. 287-97.
16. Reya, T., et al., *Stem cells, cancer, and cancer stem cells.* Nature, 2001. **414**(6859): p. 105-11.
17. Kastan, M.B., et al., *Direct demonstration of elevated aldehyde dehydrogenase in human hematopoietic progenitor cells.* Blood, 1990. **75**(10): p. 1947-50.
18. Pearce, D.J., et al., *Characterization of cells with a high aldehyde dehydrogenase activity from cord blood and acute myeloid leukemia samples.* Stem Cells, 2005. **23**(6): p. 752-60.
19. Seigel, G.M., et al., *Cancer stem cell characteristics in retinoblastoma.* Mol Vis, 2005. **11**: p. 729-37.
20. Corti, S., et al., *Identification of a primitive brain-derived neural stem cell population based on aldehyde dehydrogenase activity.* Stem Cells, 2006. **24**(4): p. 975-85.

21. Sladek, N.E., et al., *Cellular levels of aldehyde dehydrogenases (ALDH1A1 and ALDH3A1) as predictors of therapeutic responses to cyclophosphamide-based chemotherapy of breast cancer: a retrospective study. Rational individualization of oxazaphosphorine-based cancer chemotherapeutic regimens.* Cancer Chemother Pharmacol, 2002. **49**(4): p. 309-21.
22. Douville, J., R. Beaulieu, and D. Balicki, *ALDH1 as a functional marker of cancer stem and progenitor cells.* Stem Cells Dev, 2009. **18**(1): p. 17-25.
23. Ginestier, C., et al., *ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome.* Cell Stem Cell, 2007. **1**(5): p. 555-67.
24. Glinsky, G.V., O. Berezovska, and A.B. Glinskii, *Microarray analysis identifies a death-from-cancer signature predicting therapy failure in patients with multiple types of cancer.* J Clin Invest, 2005. **115**(6): p. 1503-21.
25. Liu, D.Y., et al., *ALDH1 expression is correlated with pathologic grade and poor clinical outcome in patients with astrocytoma.* Journal of Clinical Neuroscience, 2012. **19**(12): p. 1700-1705.
26. Moreb, J.S., et al., *RNAi-mediated knockdown of aldehyde dehydrogenase class-1A1 and class-3A1 is specific and reveals that each contributes equally to the resistance against 4-hydroperoxycyclophosphamide.* Cancer Chemother Pharmacol, 2007. **59**(1): p. 127-36.
27. Sladek, N.E., *Aldehyde dehydrogenase-mediated cellular relative insensitivity to the oxazaphosphorines.* Curr Pharm Des, 1999. **5**(8): p. 607-25.
28. Meng, E., et al., *ALDH1A1 maintains ovarian cancer stem cell-like properties by altered regulation of cell cycle checkpoint and DNA repair network signaling.* PLoS One, 2014. **9**(9): p. e107142.
29. Chen, C., et al., *Evidence for epithelial-mesenchymal transition in cancer stem cells of head and neck squamous cell carcinoma.* PLoS One, 2011. **6**(1): p. e16466.
30. Su, Y., et al., *Aldehyde dehydrogenase 1 A1-positive cell population is enriched in tumor-initiating cells and associated with progression of bladder cancer.* Cancer Epidemiol Biomarkers Prev, 2010. **19**(2): p. 327-37.
31. Yang, L., et al., *ALDH1A1 defines invasive cancer stem-like cells and predicts poor prognosis in patients with esophageal squamous cell carcinoma.* Mod Pathol, 2014. **27**(5): p. 775-83.
32. Schnier, J.B., et al., *Identification of cytosolic aldehyde dehydrogenase 1 from non-small cell lung carcinomas as a flavopiridol-binding protein.* FEBS Lett, 1999. **454**(1-2): p. 100-4.
33. Niederreither, K., et al., *Differential expression of retinoic acid-synthesizing (RALDH) enzymes during fetal development and organ differentiation in the mouse.* Mech Dev, 2002. **110**(1-2): p. 165-71.
34. Kim, H., et al., *The retinoic acid synthesis gene ALDH1a2 is a candidate tumor suppressor in prostate cancer.* Cancer Res, 2005. **65**(18): p. 8118-24.
35. Shen, J.X., et al., *Mining distinct aldehyde dehydrogenase 1 (ALDH1) isoenzymes in gastric cancer.* Oncotarget, 2016.
36. You, Q., H. Guo, and D. Xu, *Distinct prognostic values and potential drug targets of ALDH1 isoenzymes in non-small-cell lung cancer.* Drug Des Devel Ther, 2015. **9**: p. 5087-97.
37. Liu, Z.H., et al., *[Effect of Gegen Qinlian decoction on hepatic cytochrome CYP450 isozymes in rats by HPLC-MS/MS].* Zhongguo Zhong Yao Za Zhi, 2015. **40**(15): p. 3072-80.
38. Molotkov, A., N. Molotkova, and G. Duester, *Retinoic acid guides eye morphogenetic movements via paracrine signaling but is unnecessary for retinal dorsoventral patterning.* Development, 2006. **133**(10): p. 1901-10.
39. Wagner, E., T. Luo, and U.C. Drager, *Retinoic acid synthesis in the postnatal mouse brain marks distinct developmental stages and functional systems.* Cereb Cortex, 2002. **12**(12): p. 1244-53.

40. Everts, H.B., et al., *Immunolocalization of enzymes, binding proteins, and receptors sufficient for retinoic acid synthesis and signaling during the hair cycle.* J Invest Dermatol, 2007. **127**(7): p. 1593-604.
41. Molotkova, N., A. Molotkov, and G. Duester, *Role of retinoic acid during forebrain development begins late when Raldh3 generates retinoic acid in the ventral subventricular zone.* Dev Biol, 2007. **303**(2): p. 601-10.
42. Wagner, E., et al., *Retinoic acid delineates the topography of neuronal plasticity in postnatal cerebral cortex.* Eur J Neurosci, 2006. **24**(2): p. 329-40.
43. Graham, C.E., et al., *Characterization of retinaldehyde dehydrogenase 3.* Biochem J, 2006. **394**(Pt 1): p. 67-75.
44. Dupe, V., et al., *A newborn lethal defect due to inactivation of retinaldehyde dehydrogenase type 3 is prevented by maternal retinoic acid treatment.* Proc Natl Acad Sci U S A, 2003. **100**(24): p. 14036-41.
45. Rexer, B.N., W.L. Zheng, and D.E. Ong, *Retinoic acid biosynthesis by normal human breast epithelium is via aldehyde dehydrogenase 6, absent in MCF-7 cells.* Cancer Res, 2001. **61**(19): p. 7065-70.
46. Okamura, S., et al., *Identification of seven genes regulated by wild-type p53 in a colon cancer cell line carrying a well-controlled wild-type p53 expression system.* Oncol Res, 1999. **11**(6): p. 281-5.
47. Kim, Y.J., et al., *HOXA9, ISL1 and ALDH1A3 methylation patterns as prognostic markers for nonmuscle invasive bladder cancer: array-based DNA methylation and expression profiling.* Int J Cancer, 2013. **133**(5): p. 1135-42.
48. Yamashita, S., et al., *Chemical genomic screening for methylation-silenced genes in gastric cancer cell lines using 5-aza-2'-deoxycytidine treatment and oligonucleotide microarray.* Cancer Sci, 2006. **97**(1): p. 64-71.
49. Han, J., L. Yang, and R.K. Puri, *Analysis of target genes induced by IL-13 cytotoxin in human glioblastoma cells.* J Neurooncol, 2005. **72**(1): p. 35-46.
50. Koch, J.G., et al., *Mammary tumor modifiers in BALB/cJ mice heterozygous for p53.* Mamm Genome, 2007. **18**(5): p. 300-9.
51. Chen, M.H., et al., *ALDH1A3, the major aldehyde dehydrogenase isoform in human cholangiocarcinoma cells, affects prognosis and gemcitabine resistance in cholangiocarcinoma patients.* Clin Cancer Res, 2016.
52. Hsu, L.C. and W.C. Chang, *Cloning and characterization of a new functional human aldehyde dehydrogenase gene.* J Biol Chem, 1991. **266**(19): p. 12257-65.
53. Singh, S., et al., *ALDH1B1 Is Crucial for Colon Tumorigenesis by Modulating Wnt/beta-Catenin, Notch and PI3K/Akt Signaling Pathways.* PLoS One, 2015. **10**(5): p. e0121648.
54. Stewart, M.J., et al., *The novel aldehyde dehydrogenase gene, ALDH5, encodes an active aldehyde dehydrogenase enzyme.* Biochem Biophys Res Commun, 1995. **211**(1): p. 144-51.
55. Kutzbach, C. and E.L. Stokstad, *Mammalian methylenetetrahydrofolate reductase. Partial purification, properties, and inhibition by S-adenosylmethionine.* Biochim Biophys Acta, 1971. **250**(3): p. 459-77.
56. Krupenko, S.A. and N.V. Oleinik, *10-formyltetrahydrofolate dehydrogenase, one of the major folate enzymes, is down-regulated in tumor tissues and possesses suppressor effects on cancer cells.* Cell Growth Differ, 2002. **13**(5): p. 227-36.
57. Johlin, F.C., et al., *Studies on the mechanism of methanol poisoning: purification and comparison of rat and human liver 10-formyltetrahydrofolate dehydrogenase.* Mol Pharmacol, 1989. **35**(6): p. 745-50.
58. Stevens, V.L., et al., *Association of polymorphisms in one-carbon metabolism genes and postmenopausal breast cancer incidence.* Cancer Epidemiol Biomarkers Prev, 2007. **16**(6): p. 1140-7.
59. Garner, C.D., E.W. Lee, and R.T. Louis-Ferdinand, *Muller cell involvement in methanol-induced retinal toxicity.* Toxicol Appl Pharmacol, 1995. **130**(1): p. 101-7.

60. Jongeneel, C.V., et al., *An atlas of human gene expression from massively parallel signature sequencing (MPSS).* Genome Res, 2005. **15**(7): p. 1007-14.
61. Ackerstaff, E., et al., *Anti-inflammatory agent indomethacin reduces invasion and alters metabolism in a human breast cancer cell line.* Neoplasia, 2007. **9**(3): p. 222-35.
62. Marcato, P., et al., *Aldehyde dehydrogenase: its role as a cancer stem cell marker comes down to the specific isoform.* Cell Cycle, 2011. **10**(9): p. 1378-84.
63. Spruance, S.L., et al., *Hazard ratio in clinical trials.* Antimicrob Agents Chemother, 2004. **48**(8): p. 2787-92.
64. Bradburn, M.J., et al., *Survival analysis part II: multivariate data analysis--an introduction to concepts and methods.* Br J Cancer, 2003. **89**(3): p. 431-6.
65. Clark, T.G., et al., *Survival analysis part I: basic concepts and first analyses.* Br J Cancer, 2003. **89**(2): p. 232-8.
66. Kneale, G.W., T.F. Mancuso, and A.M. Stewart, *Hanford radiation study III: a cohort study of the cancer risks from radiation to workers at Hanford (1944-77 deaths) by the method of regression models in life-tables.* Br J Ind Med, 1981. **38**(2): p. 156-66.
67. Tierney, J.F., et al., *Practical methods for incorporating summary time-to-event data into meta-analysis.* Trials, 2007. **8**: p. 16.
68. Stewart, L.A., M.K. Parmar, and J.F. Tierney, *Meta-analyses and large randomized, controlled trials.* N Engl J Med, 1998. **338**(1): p. 61; author reply 61-2.
69. Goudarzi, P.K., et al., *Expression and prognostic value of the aldehyde dehydrogenase 1 (ALDH1) and N-myc downstream regulated gene 2 (NDRG2) as potential markers in human astrocytomas.* Tumour Biol, 2015.
70. Ito, M., et al., *Evaluation of aldehyde dehydrogenase 1 and transcription factors in both primary breast cancer and axillary lymph node metastases as a prognostic factor.* Breast Cancer, 2016. **23**(3): p. 437-44.
71. Xu, N., et al., *Aldehyde dehydrogenase 1 (ALDH1) expression is associated with a poor prognosis of bladder cancer.* Cancer Epidemiol, 2015. **39**(3): p. 375-81.
72. Mieog, J.S., et al., *Age determines the prognostic role of the cancer stem cell marker aldehyde dehydrogenase-1 in breast cancer.* BMC Cancer, 2012. **12**: p. 42.
73. Yoshioka, T., et al., *Aldehyde dehydrogenase 1 expression is a predictor of poor prognosis in node-positive breast cancers: a long-term follow-up study.* Histopathology, 2011. **58**(4): p. 608-16.
74. Charafe-Jauffret, E., et al., *Aldehyde dehydrogenase 1-positive cancer stem cells mediate metastasis and poor clinical outcome in inflammatory breast cancer.* Clin Cancer Res, 2010. **16**(1): p. 45-55.
75. Sakakibara, M., et al., *Aldehyde dehydrogenase 1-positive cells in axillary lymph node metastases after chemotherapy as a prognostic factor in patients with lymph node-positive breast cancer.* Cancer, 2012. **118**(16): p. 3899-910.
76. Tan, E.Y., et al., *ALDH1 expression is enriched in breast cancers arising in young women but does not predict outcome.* Br J Cancer, 2013. **109**(1): p. 109-13.
77. Zhong, Y., et al., *ALDH1 is a better clinical indicator for relapse of invasive ductal breast cancer than the CD44+/CD24- phenotype.* Med Oncol, 2014. **31**(3): p. 864.
78. Kang, E.J., et al., *Association of aldehyde dehydrogenase 1 expression and biologically aggressive features in breast cancer.* Neoplasma, 2014. **61**(3): p. 352-62.
79. Bane, A., et al., *Clinical-pathologic significance of cancer stem cell marker expression in familial breast cancers.* Breast Cancer Res Treat, 2013. **140**(1): p. 195-205.
80. Zheng, R., et al., *Expression of ALDH1 and TGFbeta2 in benign and malignant breast tumors and their prognostic implications.* Int J Clin Exp Pathol, 2014. **7**(7): p. 4173-83.
81. Nogami, T., et al., *Expression of ALDH1 in axillary lymph node metastases is a prognostic factor of poor clinical outcome in breast cancer patients with 1-3 lymph node metastases.* Breast Cancer, 2014. **21**(1): p. 58-65.
82. Neumeister, V., et al., *In situ identification of putative cancer stem cells by multiplexing ALDH1, CD44, and cytokeratin identifies breast cancer patients with poor prognosis.* Am J Pathol, 2010. **176**(5): p. 2131-8.

83. Huang, Y.H., et al., *Increased SOX2 expression in less differentiated breast carcinomas and their lymph node metastases.* Histopathology, 2014. **64**(4): p. 494-503.
84. Zhou, L., et al., *Molecular markers of therapeutic resistance in breast cancer.* Hum Pathol, 2013. **44**(7): p. 1421-8.
85. Lv, Y.G., et al., *The role of survivin in diagnosis, prognosis and treatment of breast cancer.* J Thorac Dis, 2010. **2**(2): p. 100-10.
86. Morimoto, K., et al., *Stem cell marker aldehyde dehydrogenase 1-positive breast cancers are characterized by negative estrogen receptor, positive human epidermal growth factor receptor type 2, and high Ki67 expression.* Cancer Sci, 2009. **100**(6): p. 1062-8.
87. Bednarz-Knoll, N., et al., *Stromal expression of ALDH1 in human breast carcinomas indicates reduced tumor progression.* Oncotarget, 2015. **6**(29): p. 26789-803.
88. Dong, Y., et al., *The expression of aldehyde dehydrogenase 1 in invasive primary breast tumors and axillary lymph node metastases is associated with poor clinical prognosis.* Pathol Res Pract, 2013. **209**(9): p. 555-61.
89. Kim, S.J., et al., *Prognostic Impact and Clinicopathological Correlation of CD133 and ALDH1 Expression in Invasive Breast Cancer.* J Breast Cancer, 2015. **18**(4): p. 347-55.
90. De Brot, M., et al., *Prognostic impact of the cancer stem cell related markers ALDH1 and EZH2 in triple negative and basal-like breast cancers.* Pathology, 2012. **44**(4): p. 303-12.
91. Hashimoto, K., et al., *Immunohistochemical detection of breast cancer stem cells in hormone receptor-positive breast cancer and their role in response to endocrine therapy and clinical outcome.* Oncology, 2012. **82**(3): p. 168-74.
92. Lee, H.E., et al., *An increase in cancer stem cell population after primary systemic therapy is a poor prognostic factor in breast cancer.* Br J Cancer, 2011. **104**(11): p. 1730-8.
93. Santilli, G., et al., *Breast cancer-initiating cells: insights into novel treatment strategies.* Cancers (Basel), 2011. **3**(1): p. 1405-25.
94. Pistelli, M., et al., *Androgen receptor expression in early triple-negative breast cancer: clinical significance and prognostic associations.* Cancers (Basel), 2014. **6**(3): p. 1351-62.
95. Yao, T., et al., *Aldehyde dehydrogenase 1 (ALDH1) positivity correlates with poor prognosis in cervical cancer.* J Int Med Res, 2014. **42**(4): p. 1038-1042.
96. Xie, Q., et al., *Aldehyde Dehydrogenase 1 Expression Predicts Chemoresistance and Poor Clinical Outcomes in Patients with Locally Advanced Cervical Cancer Treated with Neoadjuvant Chemotherapy Prior to Radical Hysterectomy.* Ann Surg Oncol, 2016. **23**(1): p. 163-70.
97. Hou, T., et al., *Putative stem cell markers in cervical squamous cell carcinoma are correlated with poor clinical outcome.* BMC Cancer, 2015. **15**: p. 785.
98. Goossens-Beumer, I.J., et al., *Clinical prognostic value of combined analysis of Aldh1, Survivin, and EpCAM expression in colorectal cancer.* Br J Cancer, 2014. **110**(12): p. 2935-44.
99. Zhou, F., et al., *Expression and prognostic value of tumor stem cell markers ALDH1 and CD133 in colorectal carcinoma.* Oncol Lett, 2014. **7**(2): p. 507-512.
100. O'Dwyer, D., et al., *The proteomics of colorectal cancer: identification of a protein signature associated with prognosis.* PLoS One, 2011. **6**(11): p. e27718.
101. Vogler, T., et al., *The expression pattern of aldehyde dehydrogenase 1 (ALDH1) is an independent prognostic marker for low survival in colorectal tumors.* Exp Mol Pathol, 2012. **92**(1): p. 111-7.
102. Rahadiani, N., et al., *Expression of aldehyde dehydrogenase 1 (ALDH1) in endometrioid adenocarcinoma and its clinical implications.* Cancer Sci, 2011. **102**(4): p. 903-8.

103. Minato, T., et al., *Aldehyde dehydrogenase 1 expression is associated with poor prognosis in patients with esophageal squamous cell carcinoma.* Ann Surg Oncol, 2013. **20**(1): p. 209-17.
104. Wang, Y., et al., *Cancer stem cell marker ALDH1 expression is associated with lymph node metastasis and poor survival in esophageal squamous cell carcinoma: a study from high incidence area of northern China.* Dis Esophagus, 2012. **25**(6): p. 560-5.
105. Hwang, C.C., et al., *A retrospective review of the prognostic value of ALDH-1, Bmi-1 and Nanog stem cell markers in esophageal squamous cell carcinoma.* PLoS One, 2014. **9**(8): p. e105676.
106. Ji, Y., et al., *Aldehyde Dehydrogenase-1 Expression Predicts Unfavorable Outcomes in Patients with Esophageal Squamous Cell Carcinoma.* Anticancer Res, 2016. **36**(1): p. 343-9.
107. Ajani, J.A., et al., *ALDH-1 expression levels predict response or resistance to preoperative chemoradiation in resectable esophageal cancer patients.* Mol Oncol, 2014. **8**(1): p. 142-9.
108. Kim, N., et al., *Cancer Stem Cell Markers in Eyelid Sebaceous Gland Carcinoma: High Expression of ALDH1, CD133, and ABCG2 Correlates With Poor Prognosis.* Invest Ophthalmol Vis Sci, 2015. **56**(3): p. 1813-9.
109. Suzuki, E., et al., *Aldehyde dehydrogenase 1 is associated with recurrence-free survival but not stem cell-like properties in hepatocellular carcinoma.* Hepatol Res, 2012. **42**(11): p. 1100-11.
110. Morise, M., et al., *Clinicopathological significance of cancer stem-like cell markers in high-grade neuroendocrine carcinoma of the lung.* J Cancer Res Clin Oncol, 2015. **141**(12): p. 2121-30.
111. Okudela, K., et al., *Expression of the potential cancer stem cell markers, CD133, CD44, ALDH1, and beta-catenin, in primary lung adenocarcinoma--their prognostic significance.* Pathol Int, 2012. **62**(12): p. 792-801.
112. Zenke, Y., et al., *Aldehyde dehydrogenase 1 expression in cancer cells could have prognostic value for patients with non-small cell lung cancer who are treated with neoadjuvant therapy: identification of prognostic microenvironmental factors after chemoradiation.* Pathol Int, 2013. **63**(12): p. 599-606.
113. Liu, W., et al., *Expression patterns of cancer stem cell markers ALDH1 and CD133 correlate with a high risk of malignant transformation of oral leukoplakia.* Int J Cancer, 2013. **132**(4): p. 868-74.
114. Ayub, T.H., et al., *Accumulation of ALDH1-positive cells after neoadjuvant chemotherapy predicts treatment resistance and prognosticates poor outcome in ovarian cancer.* Oncotarget, 2015. **6**(18): p. 16437-48.
115. Liebscher, C.A., et al., *Aldehyde dehydrogenase 1/epidermal growth factor receptor coexpression is characteristic of a highly aggressive, poor-prognosis subgroup of high-grade serous ovarian carcinoma.* Hum Pathol, 2013. **44**(8): p. 1465-71.
116. Kuroda, T., et al., *ALDH1-high ovarian cancer stem-like cells can be isolated from serous and clear cell adenocarcinoma cells, and ALDH1 high expression is associated with poor prognosis.* PLoS One, 2013. **8**(6): p. e65158.
117. Chen, P.X., Q.Y. Li, and Z. Yang, *Musashi-1 Expression is a Prognostic Factor in Ovarian Adenocarcinoma and Correlates with ALDH-1 Expression.* Pathol Oncol Res, 2015. **21**(4): p. 1133-40.
118. Mizuno, T., et al., *Cancer stem-like cells of ovarian clear cell carcinoma are enriched in the ALDH-high population associated with an accelerated scavenging system in reactive oxygen species.* Gynecol Oncol, 2015. **137**(2): p. 299-305.
119. Huang, R., et al., *The expression of aldehyde dehydrogenase 1 (ALDH1) in ovarian carcinomas and its clinicopathological associations: a retrospective study.* BMC Cancer, 2015. **15**: p. 502.
120. Wang, Y.C., et al., *ALDH1-bright epithelial ovarian cancer cells are associated with CD44 expression, drug resistance, and poor clinical outcome.* Am J Pathol, 2012. **180**(3): p. 1159-69.

121. Chang, B., et al., *ALDH1 expression correlates with favorable prognosis in ovarian cancers.* Mod Pathol, 2009. **22**(6): p. 817-23.
122. Avoranta, S.T., et al., *ALDH1 expression indicates chemotherapy resistance and poor outcome in node-negative rectal cancer.* Hum Pathol, 2013. **44**(6): p. 966-74.
123. Oh, S.Y., et al., *ALDH1 is a prognostic factor for patients treated with neoadjuvant chemoradiotherapy and radical resection for stage III rectal cancer.* J Surg Oncol, 2015. **111**(2): p. 243-7.
124. Deng, Y., et al., *ALDH1 is an independent prognostic factor for patients with stages II-III rectal cancer after receiving radiochemotherapy.* Br J Cancer, 2014. **110**(2): p. 430-4.
125. Yoon, G., et al., *Clinical influence of cancer stem cells on residual disease after preoperative chemoradiotherapy for rectal cancer.* Tumour Biol, 2015.
126. Liu, J., et al., *Prognostic Impact of ALDH1 Expression in Transitional Cell Carcinoma of the Renal Pelvis.* Anticancer Res, 2015. **35**(9): p. 4829-36.
127. Huang, C.F., et al., *Correlation of ALDH1, CD44, OCT4 and SOX2 in tongue squamous cell carcinoma and their association with disease progression and prognosis.* J Oral Pathol Med, 2014. **43**(7): p. 492-8.
128. Kitamura, H., et al., *Prognostic impact of the expression of ALDH1 and SOX2 in urothelial cancer of the upper urinary tract.* Mod Pathol, 2013. **26**(1): p. 117-24.
129. Wu, Q., et al., *Aldehyde dehydrogenase-1 predicts favorable prognosis in patients with vulvar squamous cell carcinoma.* Anticancer Res, 2014. **34**(2): p. 859-65.
130. Khoury, T., et al., *Aldehyde dehydrogenase 1A1 expression in breast cancer is associated with stage, triple negativity, and outcome to neoadjuvant chemotherapy.* Mod Pathol, 2012. **25**(3): p. 388-97.
131. Liu, Y., et al., *ALDH1A1 mRNA expression in association with prognosis of triple-negative breast cancer.* Oncotarget, 2015. **6**(38): p. 41360-9.
132. Wei, H., et al., *Breast cancer stem cells phenotype and plasma cell-predominant breast cancer independently indicate poor survival.* Pathol Res Pract, 2016. **212**(4): p. 294-301.
133. Ali, H.R., et al., *Cancer stem cell markers in breast cancer: pathological, clinical and prognostic significance.* Breast Cancer Res, 2011. **13**(6): p. R118.
134. Wu, S., et al., *Distinct prognostic values of ALDH1 isoenzymes in breast cancer.* Tumour Biol, 2015. **36**(4): p. 2421-6.
135. Zhou, L., et al., *Novel prognostic markers for patients with triple-negative breast cancer.* Hum Pathol, 2013. **44**(10): p. 2180-7.
136. Sjostrom, M., et al., *Stem cell biomarker ALDH1A1 in breast cancer shows an association with prognosis and clinicopathological variables that is highly cut-off dependent.* J Clin Pathol, 2015. **68**(12): p. 1012-9.
137. Kahlert, C., et al., *Expression analysis of aldehyde dehydrogenase 1A1 (ALDH1A1) in colon and rectal cancer in association with prognosis and response to chemotherapy.* Ann Surg Oncol, 2012. **19**(13): p. 4193-201.
138. Xu, S.L., et al., *Distinct patterns of ALDH1A1 expression predict metastasis and poor outcome of colorectal carcinoma.* Int J Clin Exp Pathol, 2014. **7**(6): p. 2976-86.
139. Li, X.S., et al., *ALDH1A1 overexpression is associated with the progression and prognosis in gastric cancer.* BMC Cancer, 2014. **14**: p. 705.
140. Adam, S.A., et al., *ALDH1A1 is a marker of astrocytic differentiation during brain development and correlates with better survival in glioblastoma patients.* Brain Pathol, 2012. **22**(6): p. 788-97.
141. Xu, S.L., et al., *Aldehyde dehydrogenase 1A1 circumscribes high invasive glioma cells and predicts poor prognosis.* Am J Cancer Res, 2015. **5**(4): p. 1471-83.
142. Qian, X., et al., *Prognostic significance of ALDH1A1-positive cancer stem cells in patients with locally advanced, metastasized head and neck squamous cell carcinoma.* J Cancer Res Clin Oncol, 2014. **140**(7): p. 1151-8.

143. Koukourakis, M.I., et al., *Cancer stem cell phenotype relates to radio-chemotherapy outcome in locally advanced squamous cell head-neck cancer.* Br J Cancer, 2012. **106**(5): p. 846-53.
144. Xu, J., et al., *Comparison of quantum dot technology with conventional immunohistochemistry in examining aldehyde dehydrogenase 1A1 as a potential biomarker for lymph node metastasis of head and neck cancer.* Eur J Cancer, 2012. **48**(11): p. 1682-91.
145. Leinung, M., et al., *Expression of ALDH1A1 and CD44 in primary head and neck squamous cell carcinoma and their value for carcinogenesis, tumor progression and cancer stem cell identification.* Oncol Lett, 2015. **10**(4): p. 2289-2294.
146. Martin, M., et al., *Aldehyde dehydrogenase isoform 1 (ALDH1) expression as a predictor of radiosensitivity in laryngeal cancer.* Clin Transl Oncol, 2015.
147. Tanaka, K., et al., *ALDH1A1-overexpressing cells are differentiated cells but not cancer stem or progenitor cells in human hepatocellular carcinoma.* Oncotarget, 2015. **6**(28): p. 24722-32.
148. Jiang, F., et al., *Aldehyde dehydrogenase 1 is a tumor stem cell-associated marker in lung cancer.* Mol Cancer Res, 2009. **7**(3): p. 330-8.
149. Li, X., et al., *Aldehyde dehydrogenase 1A1 possesses stem-like properties and predicts lung cancer patient outcome.* J Thorac Oncol, 2012. **7**(8): p. 1235-45.
150. Sullivan, J.P., et al., *Aldehyde dehydrogenase activity selects for lung adenocarcinoma stem cells dependent on notch signaling.* Cancer Res, 2010. **70**(23): p. 9937-48.
151. Shimada, Y., et al., *Cancer stem cell-related marker expression in lung adenocarcinoma and relevance of histologic subtypes based on IASLC/ATS/ERS classification.* Onco Targets Ther, 2013. **6**: p. 1597-604.
152. Dimou, A., et al., *Measurement of aldehyde dehydrogenase 1 expression defines a group with better prognosis in patients with non-small cell lung cancer.* Am J Pathol, 2012. **181**(4): p. 1436-42.
153. Alamgeer, M., et al., *The prognostic significance of aldehyde dehydrogenase 1A1 (ALDH1A1) and CD133 expression in early stage non-small cell lung cancer.* Thorax, 2013. **68**(12): p. 1095-104.
154. Gao, F., et al., *The role of LGR5 and ALDH1A1 in non-small cell lung cancer: Cancer progression and prognosis.* Biochem Biophys Res Commun, 2015. **462**(2): p. 91-8.
155. Kaminagakura, E., et al., *Protein expression of CYP1A1, CYP1B1, ALDH1A1, and ALDH2 in young patients with oral squamous cell carcinoma.* Int J Oral Maxillofac Surg, 2016.
156. Ishiguro, T., et al., *Establishment and Characterization of an In Vitro Model of Ovarian Cancer Stem-like Cells with an Enhanced Proliferative Capacity.* Cancer Res, 2016. **76**(1): p. 150-60.
157. Chui, M.H., et al., *Loss of ALDH1A1 expression is an early event in the pathogenesis of ovarian high-grade serous carcinoma.* Mod Pathol, 2015. **28**(3): p. 437-45.
158. Kahlert, C., et al., *Low expression of aldehyde dehydrogenase 1A1 (ALDH1A1) is a prognostic marker for poor survival in pancreatic cancer.* BMC Cancer, 2011. **11**: p. 275.
159. Xing, Y., et al., *High ALDH1A1 expression correlates with poor survival in papillary thyroid carcinoma.* World J Surg Oncol, 2014. **12**: p. 29.
160. Li, T., et al., *ALDH1A1 is a marker for malignant prostate stem cells and predictor of prostate cancer patients' outcome.* Lab Invest, 2010. **90**(2): p. 234-44.
161. Le Magnen, C., et al., *Characterization and clinical relevance of ALDHbright populations in prostate cancer.* Clin Cancer Res, 2013. **19**(19): p. 5361-71.
162. Sung, W.J., et al., *Epithelial-mesenchymal transition in patients of pulmonary adenocarcinoma: correlation with cancer stem cell markers and prognosis.* Int J Clin Exp Pathol, 2015. **8**(8): p. 8997-9009.
163. Wang, K., et al., *Increased expression of ALDH1A1 protein is associated with poor prognosis in clear cell renal cell carcinoma.* Med Oncol, 2013. **30**(2): p. 574.

164. Aguilera, O., et al., *Nuclear DICKKOPF-1 as a biomarker of chemoresistance and poor clinical outcome in colorectal cancer.* Oncotarget, 2015. **6**(8): p. 5903-17.
165. Seidensaal, K., et al., *Impaired aldehyde dehydrogenase 1 subfamily member 2A-dependent retinoic acid signaling is related with a mesenchymal-like phenotype and an unfavorable prognosis of head and neck squamous cell carcinoma.* Mol Cancer, 2015. **14**: p. 204.
166. Kostareli, E., et al., *HPV-related methylation signature predicts survival in oropharyngeal squamous cell carcinomas.* J Clin Invest, 2013. **123**(6): p. 2488-501.
167. Marcato, P., et al., *Aldehyde dehydrogenase 1A3 influences breast cancer progression via differential retinoic acid signaling.* Mol Oncol, 2015. **9**(1): p. 17-31.
168. Qiu, Y., et al., *ALDH(+)/CD44(+) cells in breast cancer are associated with worse prognosis and poor clinical outcome.* Exp Mol Pathol, 2016. **100**(1): p. 145-50.
169. Qiu, Y., et al., *The expression of aldehyde dehydrogenase family in breast cancer.* J Breast Cancer, 2014. **17**(1): p. 54-60.
170. Zhang, W., et al., *ALDH1A3: A Marker of Mesenchymal Phenotype in Gliomas Associated with Cell Invasion.* PLoS One, 2015. **10**(11): p. e0142856.
171. Yang, Z.L., et al., *Positive ALDH1A3 and negative GPX3 expressions are biomarkers for poor prognosis of gallbladder cancer.* Dis Markers, 2013. **35**(3): p. 163-72.
172. Chen, X.Q., J.R. He, and H.Y. Wang, *Decreased expression of ALDH1L1 is associated with a poor prognosis in hepatocellular carcinoma.* Med Oncol, 2012. **29**(3): p. 1843-9.
173. Ran, D., et al., *Aldehyde dehydrogenase activity among primary leukemia cells is associated with stem cell features and correlates with adverse clinical outcomes.* Exp Hematol, 2009. **37**(12): p. 1423-34.
174. Russo, J., et al., *4-(N,N-dipropylamino)benzaldehyde inhibits the oxidation of all-trans retinal to all-trans retinoic acid by ALDH1A1, but not the differentiation of HL-60 promyelocytic leukemia cells exposed to all-trans retinal.* BMC Pharmacol, 2002. **2**: p. 4.
175. Escuin-Ordinas, H., et al., *COX-2 inhibition prevents the appearance of cutaneous squamous cell carcinomas accelerated by BRAF inhibitors.* Mol Oncol, 2014. **8**(2): p. 250-60.
176. MacGillivray, S., et al., *Efficacy and tolerability of selective serotonin reuptake inhibitors compared with tricyclic antidepressants in depression treated in primary care: systematic review and meta-analysis.* BMJ, 2003. **326**(7397): p. 1014.
177. Clarke, M.F., et al., *Cancer stem cells--perspectives on current status and future directions: AACR Workshop on cancer stem cells.* Cancer Res, 2006. **66**(19): p. 9339-44.
178. Bradfute, S.B., T.A. Graubert, and M.A. Goodell, *Roles of Sca-1 in hematopoietic stem/progenitor cell function.* Exp Hematol, 2005. **33**(7): p. 836-43.
179. Burger, P.E., et al., *Sca-1 expression identifies stem cells in the proximal region of prostatic ducts with high capacity to reconstitute prostatic tissue.* Proc Natl Acad Sci U S A, 2005. **102**(20): p. 7180-5.