ANTICANCER RESEARCH *40*: 6733-6742 (2020)
doi:10.21873/anticanres.14696

Aldehyde Dehydrogenase 1-related Genes in Triple-negative Breast Cancer Investigated Using Network Analysis

AKIMITSU YAMADA[1], CHIHO SUZUKI[1], HIDETAKA SHIMA[1], KUMIKO KIDA[2], SHOKO ADACHI[2], SHINYA YAMAMOTO[2], KAZUTAKA NARUI[2], MIKIKO TANABE[3], DAISUKE SHIMIZU[4], RIE TANIGUCHI[5], MASANORI OSHI[1,6], KAZUAKI TAKABE[1,6,7], YOHEI MIYAGI[8], YASUSHI ICHIKAWA[9], TAKASHI ISHIKAWA[7] and ITARU ENDO[1]

[1]*Department of Gastroenterological Surgery, Yokohama City University Graduate School of Medicine, Kanagawa, Japan;*
[2]*Department of Breast and Thyroid Surgery, Yokohama City University Medical Center, Kanagawa, Japan;*
[3]*Department of Pathology, Yokohama City University Medical Center, Kanagawa, Japan;*
[4]*Department of Breast Surgery, Yokohama City Minato Red Cross Hospital, Kanagawa, Japan;*
[5]*KM data Inc., Tokyo, Japan;*
[6]*Division of Breast Surgery, Department of Surgical Oncology, Roswell Park Comprehensive Cancer Center, Buffalo, NY, U.S.A.;*
[7]*Department of Breast Disease, Tokyo Medical University Hospital, Tokyo, Japan;*
[8]*Molecular Pathology and Genetics Division, Kanagawa Cancer Center Research Institute, Kanagawa, Japan;*
[9]*Department of Oncology, Yokohama City University, Kanagawa, Japan*

Abstract. *Background/Aim: Aldehyde dehydrogenase 1 (ALDH1) is known as a breast cancer stem cell (CSC) marker. This study aimed to identify genes associated with ALDH1. Materials and Methods: ALDH1-positive and -negative breast cancer cells were isolated using laser capture microdissection from five tissue samples of ALDH1-positive breast cancer patients. Messenger RNA expression levels were compared between ALDH1-positive and -negative cells. Results: We found 104 differentially expressed genes between ALDH1-positive and -negative cells. Gene ontology and pathway analysis revealed that these genes were correlated with CSC functions and pathways. Network analyses identified 10 genes that were closely associated with ALDH1. We validated these 10 genes utilizing The Cancer Genome Atlas and the Molecular Taxonomy of Breast Cancer International Consortium cohort, and found that they were associated with ALDH1 expression and correlated with Wnt pathway signaling. Conclusion: The 10 genes we identified could be potential targets for CSC therapy of breast cancer.*

Correspondence to: Akimitsu Yamada, MD, Ph.D., Lecturer, Department of Gastroenterological Surgery, Yokohama City University Graduate School of Medicine, 3-9 Fukuura, Kanazawa-ward, Yokohama, Kanagawa, 236-0004, Japan. Tel: +81 457872650, Fax: +81 457829161, e-mail: ayamada@yokohama-cu.ac.jp

Key Words: Triple-negative breast cancer, aldehyde dehydrogenase 1, microarray analysis, network analysis.

Aldehyde dehydrogenase 1 (ALDH1) has been identified as a marker of breast cancer stem cells (CSCs) (1). Two meta-analyses on ALDH1 function in breast cancer have been reported (2, 3). One of these studies analyzed 15 publications on ALDH1A1 and revealed that ALDH1A1 expression was significantly associated with tumor size, nodal status, histological grade, estrogen receptor (ER)- and progesterone receptor (PR)-negativity, and epidermal growth factor receptor 2 (HER2)-positivity. The prognosis in patients with ALDH1A1-positive tumors was worse than that in patients with ALDH1-negative tumors (2). In the other meta-analysis on 12 eligible studies, the results were similar except for tumor size and nodal status (3).

We also previously examined ALDH1A1 expression in 653 invasive breast cancer cases using core needle biopsy specimens at diagnosis (4). ALDH1 expression was examined in tumor cells and detected in 139 of the 653 cases (21.3%). The association of ALDH1 expression with clinicopathological features was consistent with that shown in previous meta-analyses. According to intrinsic subtypes, ALDH1-positive cases were found in the luminal type (12.2%), luminal-HER2 type (36.5%), HER2-enriched type (37.9%), and triple-negative type (30.0%).

Based on these results, it is clear that ALDH1 is associated with poor clinical outcomes in breast cancer patients, probably through regulating CSC features. ALDH1 is known as an enzyme that catalyzes biosynthesis of retinoic acid (RA) by oxidizing retinal and aliphatic aldehydes and plays a role in detoxification (5). However, questions remain

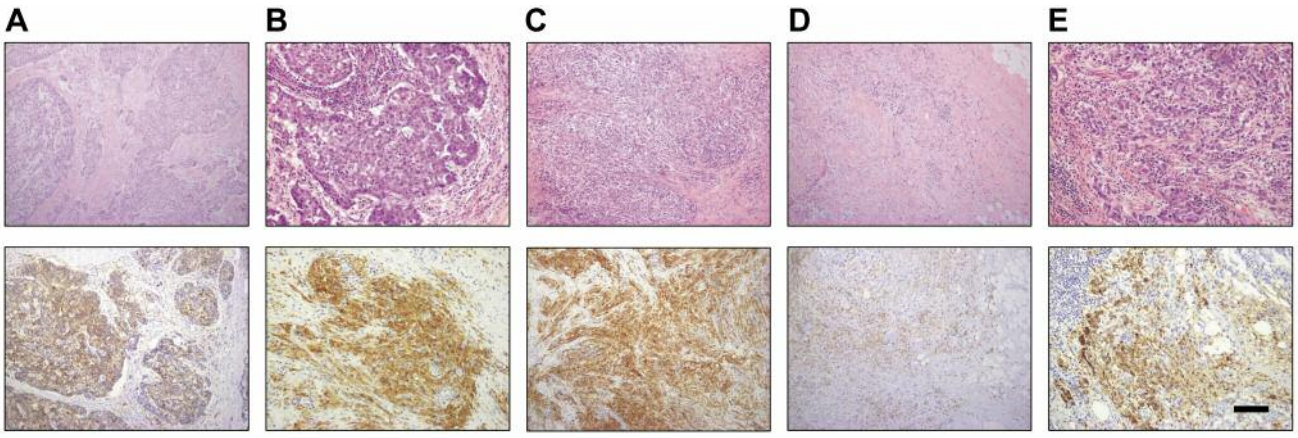

Figure 1. *Representative images of histology of 5 triple negative breast cancer. Hematoxylin-eosin staining (upper) and immunohistochemistry of ALDH1A1 staining (bottom) from 5 patient samples are shown. All patients were diagnosed as invasive ductal breast cancer by H&E. Scale bar=200 μm.*

as to how ALDH1 affects biological features of breast cancer cells and why this gene acts as a marker of CSCs.

In this study, we focused on triple-negative breast cancer (TNBC) because some cellular populations of TNBC were shown to possess stem cell features in comprehensive molecular analysis (6, 7). We aimed to identify genes associated with *ALDH1* function as potential target genes in CSC that could be used to develop treatment for TNBC.

Materials and Methods

Patients and samples. Tissue samples were obtained from patients who underwent surgery at the Yokohama City University Medical Center. Five patients with triple-negative breast cancer (TNBC) and ALDH1A1 expression were enrolled in this study. The patients did not receive any preoperative treatments to avoid potential gene modification. This study was approved by the Institutional Review Board of Yokohama City University (D1207027). All procedures performed on human participants were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The patients provided informed consent prior to inclusion in the study.

Histopathological and immunohistological staining. Hematoxylin and eosin (H&E)-stained sections from each block were prepared to determine the histological examination and diagnosis. To determine the breast cancer subtype, immunohistochemistry (IHC) of paraffin-embedded breast cancer tissues was performed to detect ER, PgR, and HER2. ER-negative, PgR-negative, and HER2-negative tumors were considered as TNBC. IHC was performed with an anti-ALDH1A1 (EP1933Y, ab52492, Abcam, Cambridge, UK) antibody. The IHC protocol with anti-ALDH1A1 was as previously described (4). Representative images of the H&E and ALDH1A1 staining are shown in Figure 1.

Laser micro dissection of ALDH1-positive and ALDH1-negative tumor cells for RNA extraction. ALDH1-positive and -negative cells were dissected separately from the five TNBC tissue samples using laser capture microdissection (LCM; PALM MicroBeam, Zeiss, Germany). Representative images pre- and post-LCM are shown in Figure 2. Then, the RNA was isolated from tumor tissue specimens after LCM according to a proprietary procedure from Response Genetics (Los Angeles, CA, USA) (8). Total RNA was analyzed using Affymetrix GeneChip microarrays (Affymetrix Human Genome U133 Plus 2.0 Array Thermo Fisher Scientific, Waltham, MA, USA). We performed a microarray analysis of five ALDH1-positive TNBC samples.

Microarray analysis to identify differentially expressed mRNAs between ALDH1-positive and ALDH1-negative tumor cells. The data were calibrated and standardized using Microarray Suite version 5.0 (MAS 5.0) (9, 10). MAS5 is the most commonly used and suitable method for microarray normalization. Following standardization, we excluded genes with unreliable values or values <300 for the quality of microarray data. We calculated the fold change (FC) of gene expression (ALDH1-positive area *vs.* ALDH1-negative area) and identified 104 genes with FC values >2.0 or <0.5.

Molecular network and statistical analyses. The 104 identified genes were analyzed using the KeyMolnet knowledge database (viewer program version 6.2, contents version 9.7.20180921161102) (KM Data Inc.; www.km-data.jp) (11). KeyMolnet has manually curated content on numerous associations among genes, proteins, metabolites, microRNAs, and molecular annotations such as diseases, pathological events, drug targets, and biomarker information. The list of differentially expressed genes was imported into KeyMolnet. The "start points and end points" network search algorithm was performed using differentially expressed genes as the start points and *ALDH1* as the end-point to generate the network and identify candidate regulatory molecules causing *ALDH1* induction. The statistical significance in concordance between the canonical pathways and the extracted network was evaluated using an algorithm that counts the number of overlapping molecular relations shared by both. This made it possible to identify the

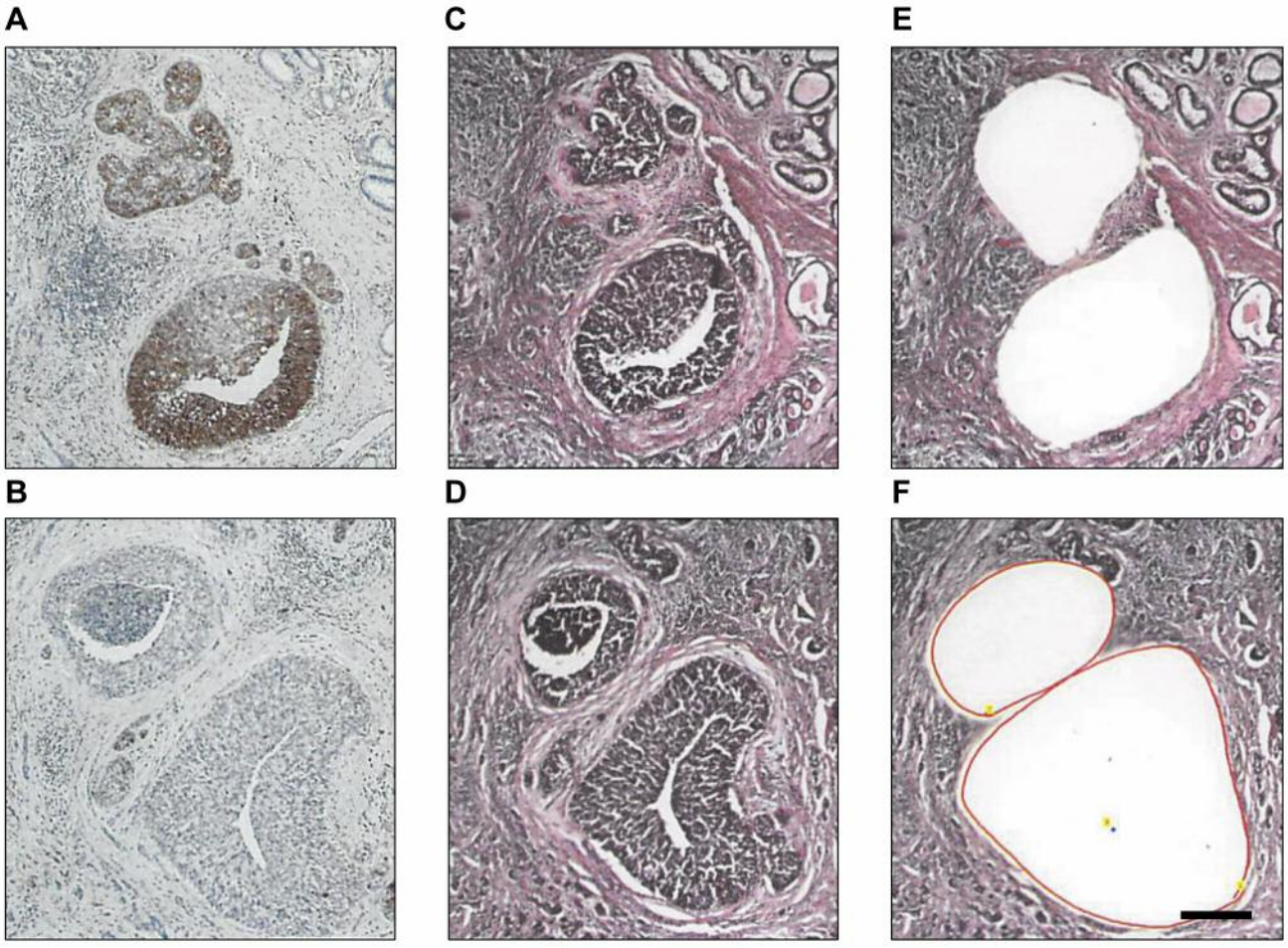

Figure 2. *Representative images of laser microdissection of ALDH1-positive and -negative breast cancer cells. Immunohistochemistry analysis revealed ALDH1A1-positive cells (A) and -negative cells (B). The slide with hematoxylin-eosin staining was marked according to IHC (C, D). Laser microdissection was performed on ALDH1A1-positive (E) and -negative cells (F).*

canonical pathway exhibiting the most significant contribution to the extracted network.

Gene expression analyses of the TCGA-BRCA and METABRIC cohorts. We used two large publicly available cohorts, The Cancer Genome Atlas (TCGA) (12) and the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) (13), to confirm the clinical relevance of the identified genes. Normalized gene expression data (log2 transcripts per million values) of primary breast cancer tumors from female patients in the two cohorts were obtained from the cBio Cancer Genomics data portal. Gene set variation analysis (GSVA) was used to transform the gene expression values into enrichment scores for the pathway (14). The GSVA score for the HALLMARK_WNT_BETA_CATENIN_SIGNALING mSigDb Hallmark gene set (15) was calculated for each tumor from its gene expression. For each of the *ALDH1*-associated genes of interest, patients from both cohorts were grouped into high- and low-expression groups based on the within-cohort 10th percentile gene expression value. The boxplots depicted median, inter-quartile range, and outliers using the Tukey method. The Hallmark gene set scores, as well as the ALDH1 gene expression values of the two groups were compared using one-way ANOVA.

Results

Identification of genes associated with ALDH1A1. The total RNA isolated from ALDH1A1-positive and -negative cells dissected using LCM was subjected to gene expression analysis using Affymetrix GeneChip microarrays (Figure 2). The data on up-regulation and down-regulation of genes were recorded. Initially 54,682 genes were extracted, and 32,264 genes were selected after background noise elimination. The *GAPDH* as a housekeeping gene and *ALDH1A1* from our microarray datasets are shown in Tables I and II. High expression of *GAPDH* was detected in all samples (Table I). On the other hand, the expression of

Table I. *GAPDH signaling in microarray data.*

ID	Patients	Condition	Detection	Ch1 raw signal	Processed signal
1	1	ALDH1A1-positive	P	3730.377	–1.9939966
2	1	ALDH1A1-negative	P	14619.6	0
3	2	ALDH1A1-positive	P	10622.08	0.40078735
4	2	ALDH1A1-negative	P	7824.224	0
5	3	ALDH1A1-positive	P	11867.65	1.4744778
6	3	ALDH1A1-negative	P	4345.466	0
7	4	ALDH1A1-positive	P	5395.896	–0.8161192
8	4	ALDH1A1-negative	P	9130.295	0
9	5	ALDH1A1-positive	P	2573.483	0.50331974
10	5	ALDH1A1-negative	P	1781.484	0

P: Present, M: marginal, A: absent.

Table II. *ALDH1A1 signaling in microarray data.*

ID	Patients	Condition	Detection	Ch1 raw signal	Processed signal
1	1	ALDH1A1-positive	P	1340.314	5.8000393
2	1	ALDH1A1-negative	A	23.66741	0
3	2	ALDH1A1-positive	P	788.4848	1.681387
4	2	ALDH1A1-negative	P	239.0706	0
5	3	ALDH1A1-positive	A	29.62089	-0.02503395
6	3	ALDH1A1-negative	A	30.66673	0
7	4	ALDH1A1-positive	M	257.9763	-0.6529236
8	4	ALDH1A1-negative	A	389.8291	0
9	5	ALDH1A1-positive	A	15.1019	-1.109621
10	5	ALDH1A1-negative	A	31.97685	0

P: Present, M: marginal, A: absent.

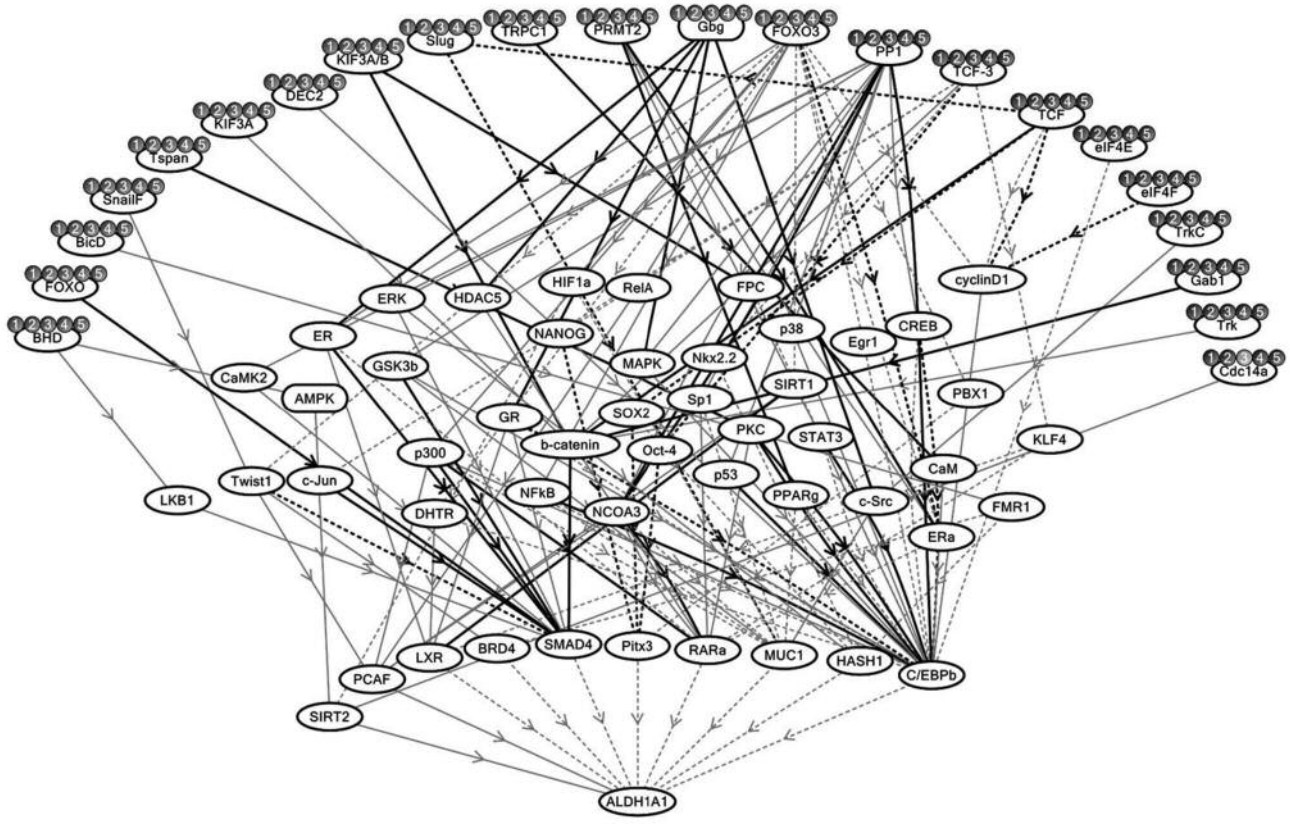

Figure 3. *Network analysis of genes associated with ALDH1. The list of the 104 differentially expressed genes was imported into KeyMolnet, and then the "start points and end points" network search algorithm was performed using differentially expressed genes as the start points and ALDH1 as the end-point, to generate the network and identify candidate regulatory molecules causing ALDH1 induction.*

ALDH1A1 varied among samples, and not all ALDH1A1-positive cells expressed *ALDH1A1* compared to ALDH1A1-negative cells (Table II).

The fold change (FC) in gene expression (ALDH1A1-positive area *vs*. ALDH1 A1-negative area) was calculated, and genes with FC values >2 or <0.5 in each of the five cases were identified (Table III). Among them, genes that were commonly different between the ADH1A1-positive and ALDH1A1-negative cells in the five cases were extracted. With regard to the FC in gene expression, 63 genes showed two-fold higher

Table III. *Number of up/down-regulated genes in 5 cases.*

	Case 1		Case 2		Case 3		Case 4		Case 5	
	Up	Down	Up	Down	Up	Down	Up	Down	Up	Down
Genes	9,106	9,281	9,253	8,584	7,015	6,830	9,866	10,030	7,330	7,694

Table IV. *Gene ontology analysis.*

Category	#in category	#overlaps	*p*-Value
Organ morphogenesis // non-traceable author statement	21	2	3.25E-04
Smooth muscle contraction // inferred from electronic annotation	22	2	3.54E-04
Cell differentiation // non-traceable author statement	53	2	0.001861
Ectoderm and mesoderm interaction // traceable author statement	1	1	0.0023
Regulation of tor signaling pathway // inferred from direct assay	1	1	0.0023
Glycosylphosphatidylinositol anchor biosynthetic process // inferred from electronic annotation	61	2	0.002433
Circadian rhythm // inferred from electronic annotation	71	2	0.00325
Cellular membrane fusion // inferred from mutant phenotype	2	1	0.003448
Positive regulation of fibroblast proliferation // non-traceable author statement	2	1	0.003448
Platelet alpha granule organization // inferred from mutant phenotype	2	1	0.003448
B cell costimulation // inferred from electronic annotation	3	1	0.004595
Cell proliferation // non-traceable author statement	90	2	0.005104
Glucose homeostasis // inferred from electronic annotation	90	2	0.005104
Melanosome localization // inferred from direct assay	4	1	0.005741
Lysosome localization // inferred from direct assay	4	1	0.005741
Immunoglobulin secretion // inferred from electronic annotation	5	1	0.006885
Positive regulation of germinal center formation // inferred from electronic annotation	5	1	0.006885
Positive regulation of membrane potential // inferred from electronic annotation	6	1	0.008028
Initiation of primordial ovarian follicle growth // inferred from electronic annotation	7	1	0.00917
Positive regulation of cd4-positive, cd25-positive, alpha-beta regulatory t cell differentiation // inferred from electronic annotation	7	1	0.00917

and 41 genes showed two-fold lower expression in ALDH1A1-positive cells compared to ALDH1-negative cells.

Gene ontology and pathway analysis. Gene Ontology (GO) analysis revealed that the identified genes were associated with stem cell function such as organ morphogenesis, cell differentiation, metabolic homeostasis, and regulation of TOR signaling pathways (Table IV). The results of pathway analysis are shown in Table V. It also revealed genes associated with metabolism alteration including cyanoamino acid, steroid, and fatty acid biosynthesis pathways. The ABC transporters and nucleotide excision repair that are associated with stemness were also altered among ALDH1-positive and -negative cells.

Network analysis of genes related to ALDH1A1. The list of the 104 differentially expressed genes was imported into KeyMolnet. Then, the "start points and end points" network search algorithm was performed using the differentially expressed genes as the start points and *ALDH1A1* as the end point to generate the network and identify candidate regulatory molecules causing *ALDH1A1* induction (Figure 3). Network analysis extracted 10 transcription factors: *SMAD4, RARα, MUC1, HASH1, C/EBPβ, PITX3, BRD4, LXR, PCAF,* and *SIRT2*. These factors were directly or indirectly associated with ALDH1A1 expression.

Gene expression analyses of TCGA-BRCA and METABRIC cohorts. We validated our data using two large publicly available cohorts, The Cancer Genome Atlas (TCGA) and the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) to identify the association between the 10 genes, *ALDH1A1,* and the Wnt signaling pathway that are related to cancer stem cell function (16). The results are shown in Figure 4. Indeed, several genes including *C/EBPβ, NR1H3 (LXR), MUC1*, and *SIRT2* were associated with ALDH1A1 expression in both datasets. The expression levels

Table V. *Pathway analysis*.

Up-regulated pathways	Total genes	Up regulating genes	%
Cyanoamino acid metabolism	7	1	14.3%
Glycosylphosphatidylinositol-anchor biosynthesis	34	4	11.8%
Steroid biosynthesis	27	3	11.1%
Fatty acid biosynthesis	10	1	10.0%
SNARE interactions in vesicular transport	71	7	9.9%
ABC transporters	43	4	9.3%
Nucleotide excision repair	57	5	8.8%
Glycosaminoglycan biosynthesis - chondroitin sulfate	23	2	8.7%
Long-term potentiation	108	9	8.3%
Taurine and hypotaurine metabolism	13	1	7.7%
B cell receptor signaling pathway	140	10	7.1%
African trypanosomiasis	43	3	7.0%
TGF-beta signaling pathway	147	10	6.8%
Salmonella infection	135	9	6.7%
RIG-I-like receptor signaling pathway	90	6	6.7%
Lysine degradation	91	6	6.6%
MAPK signaling pathway	420	27	6.4%
Allograft rejection	63	4	6.3%
Osteoclast differentiation	224	14	6.3%
Autoimmune thyroid disease	64	4	6.3%
Down regulating pathways	**Total genes**	**Down regulating genes**	**%**
Caffeine metabolism	5	2	40.0%
Fatty acid biosynthesis	10	2	20.0%
Steroid biosynthesis	27	5	18.5%
Cyanoamino acid metabolism	7	1	14.3%
Linoleic acid metabolism	23	3	13.0%
Nucleotide excision repair	57	6	10.5%
Vitamin digestion and absorption	19	2	10.5%
Glycosaminoglycan biosynthesis - heparan sulfate	30	3	10.0%
alpha-Linolenic acid metabolism	20	2	10.0%
mRNA surveillance pathway	153	15	9.8%
Cardiac muscle contraction	81	7	8.6%
Ribosome	125	10	8.0%
Metabolism of xenobiotics by cytochrome P450	88	7	8.0%
Fat digestion and absorption	39	3	7.7%
Mismatch repair	26	2	7.7%
Retinol metabolism	69	5	7.2%
ABC transporters	43	3	7.0%
DNA replication	44	3	6.8%
Endocrine and other factor-regulated calcium reabsorption	79	5	6.3%
Oxidative phosphorylation	159	10	6.3%

SNARE: Soluble N-ethylmaleimide-sensitive factor attachment protein receptor, ABC: ATP-binding cassette, TGF: transforming growth factor, RIG-I: retinoic acid-inducible gene-I, MAPK: mitogen-activated protein kinase, RNA: ribonucleic acid, DNA: deoxyribonucleic acid.

of *BRD4, C/EBPβ, ASCL1, NR1H2 (LXR), MUC1, PITX3, RARα*, *SIRT2*, and *SMAD4* were correlated with Wnt pathway signaling.

Discussion

In this study, we identified differentially expressed genes between ALDH1-positive and -negative breast cancer tissue samples. The difference in gene expression between ALDH1A1-positive and -negative cells in the same tumor may provide an explanation regarding the mechanism behind *ALDH1A1* function in cancer stemness. Notably, 63 genes were up-regulated whereas 41 genes were down-regulated in ALDH1A1-positive cells compared to ALDH1A1-negative cells. CSCs exhibited self-renewal and tumor initiating properties, and treatment resistance (17). Furthermore, CSCs showed metabolic alterations in glycolytic (18), lipid (19), and steroid biosynthesis (20). Indeed, GO analysis revealed stemness related categories such as organ morphogenesis, cell differentiation, metabolic alterations, and regulation of TOR signaling pathways. Likewise, the pathway analysis also revealed altered gene expression in stemness-related pathways, such as several metabolic and treatment resistance mechanisms including ABC transporters and nucleotide excision repair among ALDH1A1-positive cells compared to ALDH1A1-negative cells.

Network analysis identified 10 transcription factors (*e.g.*, *SMAD4, RARα, MUC1, HASH1, C/EBPβ, Pitx3, BRD4, LXR, PCAF,* and *SIRT2*) that were associated with *ALDH1A1*. For example, *SMAD4* is the main mediator of TGF-β signaling pathway that is involved in many biological activities including fibrosis, embryonic development, wound healing, tumor development, cell differentiation, apoptosis, homeostasis and immune response regulation. In the complex with other transcription factors, SMAD4 acts as a regulator of the expression of target genes such as *Twist1*, *Snail*, and *Slug* that are associated with stemness (21). We then validated the association between these 10 factors and ALDH1A1 expression or the CSC-related signaling pathway by utilizing two large publicly available cohorts, The Cancer Genome Atlas (TCGA) (12) and the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) (13). These two cohorts include all subtypes of breast cancer. We have used these cohorts to demonstrate the clinical relevance of several studies (22-30). Indeed, several genes, including *C/EBPβ* (31), *NR1H3 (LXR)* (32), *MUC1* (33), and *SIRT2* (34) were associated with *ALDH1A1* expression in both datasets. The expression of *BRD4* (35), C/EBPβ, ASCL1 (hASH1) (36), *NR1H2 (LXR), MUC1, PITX3* (37), *RARα* (5), *SIRT2,* and *SMAD4* (38) were correlated with the Wnt signaling pathway, which plays an

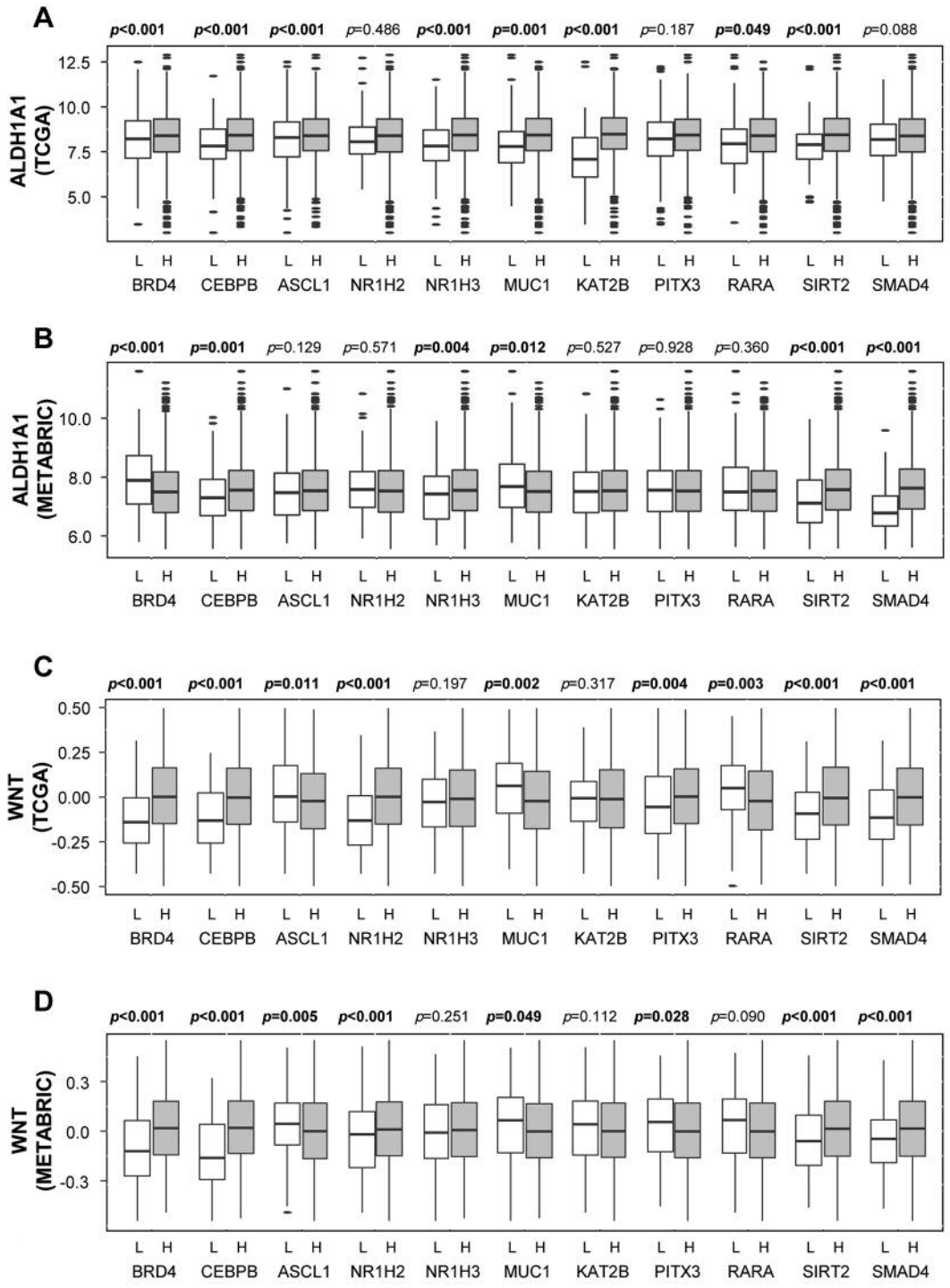

Figure 4. *Gene expression analyses of the TCGA-BRCA and METABRIC cohorts. For each of the 10 ALDH1-associated genes, breast cancers in the TCGA-BRCA (n=1,065) or METABRIC (n=1,903) cohorts were grouped into high and low expression groups based on the within-cohort 10th percentile value of gene expression. The two groups were compared for tumor expression of ALDH1 and of genes up-regulated by the Wnt-β catenin signaling pathway. Boxplots of ALDH1 gene expression among low and high expression groups of the 11 genes are shown for the TCGA-BRCA (Figure 3A) and METABRIC (B) cohorts. Boxplots of GSVA scores for the Hallmark of Wnt-β catenin signaling pathway among low and high expression groups of the 11 genes are shown for the TCGA-BRCA (C) and METABRIC (D) cohorts. ALDH1A1, aldehyde dehydrogenase 1 family, member A1; BRD4, bromodomain-containing protein 4; C/EBPβ, CCAAT/enhancer-binding protein beta; ASCL1, achaete-scute homolog 1; NR1H2, liver X receptor beta; NR1H3, liver X receptor alpha; MUC1, mucin 1, cell surface associated; KAT2B, K lysine acetyltransferase 2B; PITX3, pituitary homeobox 3; RARα, retinoic acid receptor alpha; SIRT2, NAD-dependent deacetylase sirtuin 2; and SMAD4, SMAD family member 4.*

important role in self-renewal and differentiation of stem cells (16). Interestingly, most of these 10 factors were associated with poor survival outcome in TCGA cohorts (data not shown).

Among the ALDH1A1-positive samples, some showed low expression levels of *ALDH1A1* in our microarray data. The discordance of the protein and mRNA expression levels was presumably derived from the difference in transcriptional activity of the cells or changes in transcriptional efficacy due to post-transcriptional modification (39). For example, microRNAs are recognized as one of the key mechanisms of the mRNA transcription regulatory network (40). As we have previously demonstrated the importance of ALDH1A1 protein expression in breast cancer patients (4), we have conducted microarray and network analyses based on the expression of the ALDH1A1 protein.

Although we validated our data by utilizing two large publicly available cohorts, subsequent studies involving the latest techniques such as single-cell sequencing are warranted to provide more specific information regarding the mechanisms of the regulation of breast CSCs (41). The specific mechanisms of regulation of ALDH1 in CSCs remain unclear. However, regulation of RA, reactive oxygen species (ROS), and detoxification by reactive aldehyde metabolism are considered to be closely related to functional roles of CSCs. ALDH1 has 19 human isozymes subdivided among 11 families and 4 subfamilies. Among them, ALDH1A1 and ALDH1A3 isoforms are particularly associated with CSCs owing to their roles mentioned above to exert resistance to radiotherapy and chemotherapy (5, 42). We only examined the ALDH1A1 isoform in this study. Thus, it is intriguing to perform the same analysis with ALDH1A3 as we did with ALDH1A1 in this study.

In conclusion, we found alterations of expression of 104 genes among ALDH1- positive and -negative cells that were associated with CSC functions. Network analysis showed that 10 genes were associated with *ALDH1* expression. Most of these 10 genes have already been shown to reinforce their critical roles in maintaining stem cell features, providing a rationale for ALDH1A1 being a stem cell marker of breast cancer. These genes can be potential targets for cancer stem cell therapy, particularly for treating incurable breast cancer.

Funding

This work was supported by National Institutes of Health (NIH) grant R01CA160688 to KT.

Conflicts of Interest

The Authors declare that they have no conflicts of interest in regard to this study.

Authors' Contributions

Conception and design: AY and TI. Acquisition of data: AY, CS, SA, HS, SY, MT, DS, MO, and KK. Drafting the manuscript: AY. Analyzed and interpreted data: KN, RT, KT, and YM. Supervised the project YI and EI. All Authors read and approved the final article.

Acknowledgements

The Authors thank Dr. Edward Barroga (https://orcid.org/0000-0002-8920-2607), Medical Editor and Professor of Academic Writing at St. Luke's International University, and Editage for editing the manuscript.

References

1 Ginestier C, Hur MH, Charafe-Jauffret E, Monville F, Dutcher J, Brown M, Jacquemier J, Viens P, Kleer CG, Liu S, Schott A, Hayes D, Birnbaum D, Wicha MS and Dontu G: Aldh1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. Cell Stem Cell *1(5)*: 555-567, 2007. PMID: 18371393. DOI: 10.1016/j.stem.2007.08.014

2 Liu Y, Lv DL, Duan JJ, Xu SL, Zhang JF, Yang XJ, Zhang X, Cui YH, Bian XW and Yu SC: Aldh1a1 expression correlates with clinicopathologic features and poor prognosis of breast cancer patients: A systematic review and meta-analysis. BMC Cancer *14*. 444, 2014. PMID: 24938375. DOI: 10.1186/1471-2407-14-444

3 Zhou L, Jiang Y, Yan T, Di G, Shen Z, Shao Z and Lu J: The prognostic role of cancer stem cells in breast cancer: A meta-analysis of published literatures. Breast Cancer Res Treat *122(3)*: 795-801, 2010. PMID: 20571867. DOI: 10.1007/s10549-010-0999-4

4 Kida K, Ishikawa T, Yamada A, Shimada K, Narui K, Sugae S, Shimizu D, Tanabe M, Sasaki T, Ichikawa Y and Endo I: Effect of aldh1 on prognosis and chemoresistance by breast cancer subtype. Breast Cancer Res Treat *156(2)*: 261-269, 2016. PMID: 26975188. DOI: 10.1007/s10549-016-3738-7

5 Tomita H, Tanaka K, Tanaka T and Hara A: Aldehyde dehydrogenase 1a1 in stem cells and cancer. Oncotarget *7(10)*: 11018-11032, 2016. PMID: 26783961. DOI: 10.18632/oncotarget.6920

6 Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y and Pietenpol JA: Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J Clin Invest *121(7)*: 2750-2767, 2011. PMID: 21633166. DOI: 10.1172/jci45014

7 Burstein MD, Tsimelzon A, Poage GM, Covington KR, Contreras A, Fuqua SA, Savage MI, Osborne CK, Hilsenbeck SG, Chang JC, Mills GB, Lau CC and Brown PH: Comprehensive genomic analysis identifies novel subtypes and targets of triple-negative breast cancer. Clin Cancer Res *21(7)*: 1688-1698, 2015. PMID: 25208879. DOI: 10.1158/1078-0432.ccr-14-0432

8 Lord RV, Brabender J, Gandara D, Alberola V, Camps C, Domine M, Cardenal F, Sanchez JM, Gumerlock PH, Taron M, Sanchez JJ, Danenberg KD, Danenberg PV and Rosell R: Low ercc1 expression correlates with prolonged survival after cisplatin plus gemcitabine chemotherapy in non-small cell lung cancer. Clin Cancer Res *8(7)*: 2286-2291, 2002. PMID: 12114432.

9 Hubbell E, Liu WM and Mei R: Robust estimators for expression analysis. Bioinformatics *18(12)*: 1585-1592, 2002. PMID: 12490442. DOI: 10.1093/bioinformatics/18.12.1585

10 Liu WM, Mei R, Di X, Ryder TB, Hubbell E, Dee S, Webster TA, Harrington CA, Ho MH, Baid J and Smeekens SP: Analysis of high density expression microarrays with signed-rank call algorithms. Bioinformatics *18(12)*: 1593-1599, 2002. PMID: 12490443. DOI: 10.1093/bioinformatics/18.12.1593

11 Sato H, Ishida S, Toda K, Matsuda R, Hayashi Y, Shigetaka M, Fukuda M, Wakamatsu Y and Itai A: New approaches to mechanism analysis for drug discovery using DNA microarray data combined with keymolnet. Curr Drug Discov Technol *2(2)*: 89-98, 2005. PMID: 16472233. DOI: 10.2174/1570163054064701

12 Colaprico A, Silva TC, Olsen C and Garofano L: Tcgabiolinks: An r/bioconductor package for integrative analysis of tcga data. Nucleic Acids Res *44(8)*: e71, 2016. PMID: 26704973. DOI: 10.1093/nar/gkv1507

13 Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, Graf S, Ha G, Haffari G, Bashashati A, Russell R, McKinney S, Langerod A, Green A, Provenzano E, Wishart G, Pinder S, Watson P, Markowetz F, Murphy L, Ellis I, Purushotham A, Borresen-Dale AL, Brenton JD, Tavare S, Caldas C and Aparicio S: The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature *486(7403)*: 346-352, 2012. PMID: 22522925. DOI: 10.1038/nature10983

14 Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES and Mesirov JP: Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA *102(43)*: 15545-15550, 2005. PMID: 16199517. DOI: 10.1073/pnas.0506580102

15 Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP and Tamayo P: The molecular signatures database (msigdb) hallmark gene set collection. Cell Syst *1(6)*: 417-425, 2015. PMID: 26771021. DOI: 10.1016/j.cels.2015.12.004

16 Ling L, Nurcombe V and Cool SM: Wnt signaling controls the fate of mesenchymal stem cells. Gene *433(1-2)*: 1-7, 2009. PMID: 19135507. DOI: 10.1016/j.gene.2008.12.008

17 Vermeulen L, de Sousa e Melo F, Richel DJ and Medema JP: The developing cancer stem-cell model: Clinical challenges and opportunities. Lancet Oncol *13(2)*: e83-89, 2012. PMID: 22300863. DOI: 10.1016/s1470-2045(11)70257-1

18 Peiris-Pagès M, Martinez-Outschoorn UE, Pestell RG, Sotgia F and Lisanti MP: Cancer stem cell metabolism. Breast Cancer Res *18(1)*: 55, 2016. PMID: 27220421. DOI: 10.1186/s13058-016-0712-6

19 Yi M, Li J, Chen S, Cai J, Ban Y, Peng Q, Zhou Y, Zeng Z, Peng S, Li X, Xiong W, Li G and Xiang B: Emerging role of lipid metabolism alterations in cancer stem cells. J Exp Clin Cancer Res *37(1)*: 118, 2018. PMID: 29907133. DOI: 10.1186/s13046-018-0784-5

20 Alferez DG, Simões BM, Howell SJ and Clarke RB: The role of steroid hormones in breast and effects on cancer stem cells. Curr Stem Cell Rep *4(1)*: 81-94, 2018. PMID: 29600163. DOI: 10.1007/s40778-018-0114-z

21 Ahmadi A, Najafi M, Farhood B and Mortezaee K: Transforming growth factor-β signaling: Tumorigenesis and targeting for cancer therapy. J Cell Physiol *234(8)*: 12173-12187, 2019. PMID: 30537043. DOI: 10.1002/jcp.27955

22 Katsuta E, Yan L, Takeshita T, McDonald KA, Dasgupta S, Opyrchal M and Takabe K: High myc mRNA expression is more clinically relevant than myc DNA amplification in triple-negative breast cancer. Int J Mol Sci *21(1)*: 217, 2019. PMID: 31905596. DOI: 10.3390/ijms21010217

23 Takeshita T, Asaoka M, Katsuta E, Photiadis SJ, Narayanan S, Yan L and Takabe K: High expression of polo-like kinase 1 is associated with tp53 inactivation, DNA repair deficiency, and worse prognosis in er positive her2 negative breast cancer. Am J Transl Res *11(10)*: 6507-6521, 2019. PMID: 31737202.

24 Takahashi H, Katsuta E, Yan L, Dasgupta S and Takabe K: High expression of annexin a2 is associated with DNA repair, metabolic alteration, and worse survival in pancreatic ductal adenocarcinoma. Surgery *166(2)*: 150-156, 2019. PMID: 31171367. DOI: 10.1016/j.surg.2019.04.011

25 Okano M, Oshi M, Butash AL, Asaoka M, Katsuta E, Peng X, Qi Q, Yan L and Takabe K: Estrogen receptor positive breast cancer with high expression of androgen receptor has less cytolytic activity and worse response to neoadjuvant chemotherapy but better survival. Int J Mol Sci *20(11)*: 2019. PMID: 31151151. DOI: 10.3390/ijms20112655

26 McDonald KA, Kawaguchi T, Qi Q, Peng X, Asaoka M, Young J, Opyrchal M, Yan L, Patnaik S, Otsuji E and Takabe K: Tumor heterogeneity correlates with less immune response and worse survival in breast cancer patients. Ann Surg Oncol *26(7)*: 2191-2199, 2019. PMID: 30963401. DOI: 10.1245/s10434-019-07338-3

27 Katsuta E, Qi Q, Peng X, Hochwald SN, Yan L and Takabe K: Pancreatic adenocarcinomas with mature blood vessels have better overall survival. Sci Rep *9(1)*: 1310, 2019. PMID: 30718678. DOI: 10.1038/s41598-018-37909-5

28 Sporn JC, Katsuta E, Yan L and Takabe K: Expression of microrna-9 is associated with overall survival in breast cancer patients. J Surg Res *233*: 426-435, 2019. PMID: 30502282. DOI: 10.1016/j.jss.2018.08.020

29 Hirose Y, Nagahashi M, Katsuta E, Yuza K, Miura K, Sakata J, Kobayashi T, Ichikawa H, Shimada Y, Kameyama H, McDonald KA, Takabe K and Wakai T: Generation of sphingosine-1-phosphate is enhanced in biliary tract cancer patients and is associated with lymphatic metastasis. Sci Rep *8(1)*: 10814, 2018. PMID: 30018456. DOI: 10.1038/s41598-018-29144-9

30 Yamada A, Nagahashi M, Aoyagi T, Huang WC, Lima S, Hait NC, Maiti A, Kida K, Terracina KP, Miyazaki H, Ishikawa T, Endo I, Waters MR, Qi Q, Yan L, Milstien S, Spiegel S and Takabe K: Abcc1-exported sphingosine-1-phosphate, produced by sphingosine kinase 1, shortens survival of mice and patients with breast cancer. Mol Cancer Res *16(6)*: 1059-1070, 2018. PMID: 29523764. DOI: 10.1158/1541-7786.mcr-17-0353

31 Ding S, Gan T, Song M, Dai Q, Huang H, Xu Y and Zhong C: C/ebpb-cited4 in exercised heart. Adv Exp Med Biol *1000*: 247-259, 2017. PMID: 29098625. DOI: 10.1007/978-981-10-4304-8_14

32 Wang B and Tontonoz P: Liver x receptors in lipid signalling and membrane homeostasis. Nat Rev Endocrinol *14(8)*: 452-463, 2018. PMID: 29904174. DOI: 10.1038/s41574-018-0037-x

33 Alam M, Ahmad R, Rajabi H, Kharbanda A and Kufe D: Muc1-c oncoprotein activates erk—>c/ebpbeta signaling and induction of aldehyde dehydrogenase 1a1 in breast cancer cells. J Biol Chem *288(43)*: 30892-30903, 2013. PMID: 24043631. DOI: 10.1074/jbc.M113.477158

34 Masoud GN and Li W: Hif-1alpha pathway: Role, regulation and intervention for cancer therapy. Acta Pharm Sin B *5(5)*: 378-389, 2015. PMID: 26579469. DOI: 10.1016/j.apsb.2015.05.007

35 Filippakopoulos P: What is the bet on solid tumors? J Clin Oncol *36(30)*: 3040-3042, 2018. PMID: 29847297. DOI: 10.1200/jco.2018.78.8695

36 Ball DW: Achaete-scute homolog-1 and notch in lung neuroendocrine development and cancer. Cancer Lett *204(2)*: 159-169, 2004. PMID: 15013215. DOI: 10.1016/s0304-3835(03)00452-x

37 Jimenez-Jimenez FJ, Garcia-Martin E, Alonso-Navarro H and Agundez JA: Pitx3 and risk for parkinson's disease: A systematic review and meta-analysis. Eur Neurol *71(1-2)*: 49-56, 2014. PMID: 24525476. DOI: 10.1159/000353981

38 Wang Y, Chu J, Yi P, Dong W, Saultz J, Wang Y, Wang H, Scoville S, Zhang J, Wu LC, Deng Y, He X, Mundy-Bosse B, Freud AG, Wang LS, Caligiuri MA and Yu J: Smad4 promotes tgf-beta-independent nk cell homeostasis and maturation and antitumor immunity. J Clin Invest *128(11)*: 5123-5136, 2018. PMID: 30183689. DOI: 10.1172/jci121227

39 Chen K and Rajewsky N: The evolution of gene regulation by transcription factors and microrna. Nat Rev Genet *8(2)*: 93-103, 2007. PMID: 17230196. DOI: 10.1038/nrg1990

40 He L and Hannon GJ: Microrna: Small rnas with a big role in gene regulation. Nat Rev Genet *5(7)*: 522-531, 2004. PMID: 15211354. DOI: 10.1038/nrg1379

41 Liu J, Adhav R and Xu X: Current progresses of single cell DNA sequencing in breast cancer research. Int J Biol Sci *13(8)*: 949-960, 2017. PMID: 28924377. DOI: 10.7150/ijbs.19627

42 Vassalli G: Aldehyde dehydrogenases: Not just markers, but functional regulators of stem cells. Stem Cells Int *2019*. 3904645, 2019. PMID: 30733805. DOI: 10.1155/2019/3904645

Received October 15, 2020
Revised November 3, 2020
Accepted November 13, 2020