Analysis of human aldehyde dehydrogenases (ALDH) gene expression pattern in breast cancer tissue samples: rutin-copper complex inhibit the breast cancer cell proliferation

Shridhar V. Pattar1, Manisha R. Mirjankar1, Suyamindra Kulkarni2, Pramod B. Gai2, Namadev K. Pujar1, H. G. Premakshi3, Sikandar I. Mulla4, R. L. Babu5 and Chandrappa M. Kamanavalli1*

Abstract

**Background:** Higher aldehyde dehydrogenases (ALDH) activity is one of the important signatures of breast cancer appearance and has been associated with poor prognosis. ALDH1A3 has been over-expressed in breast cancer patients. This study aims to analyze gene expression patterns of 18 ALDH isozymes in breast cancer tissue samples. It is carried out using a chip-based microarray, next-generation DNA sequencing of ALDH2 gene following in silico study to identify the natural products which act as inhibitors for over-expressed ALDH isoforms. The synthesis of rutin-copper complex and cell viability assay is carried out on MDA-MB-468 cell line.

**Results:** ALDH1A3 and ALDH18A1 have shown the highest positive mean fold of variation; whereas, ALDH2 and ALDH1A2 have shown the highest negative mean fold variation. In silico studies revealed that rutin has the highest binding affinity with both ALDH1A3 and ALDH18A1 and supported with IC50 value of rutin against MDA-MB-468 cells (144.50 μg/ml). Chemically synthesized rutin-copper complex significantly lowered the IC50 value to 119.40 μg/ml. The next-generation sequencing study provides the novel single nucleotide polymorphism (SNP) from T to G in the ALDH2 gene.

**Conclusion:** The present study signifies that, along with ALDH1A3, ALDH18A1 also acts as a marker for breast cancer. Apart from that, inhibitors of ALDH1A3 and ALDH18A1 were attained. Perhaps the single nucleotide polymorphism (SNP) obtained during the mutation analysis may be the probable cause of the highest downregulation of ALDH2 in breast cancer.

**Keywords:** Aldehyde dehydrogenases, Breast cancer, Molecular docking, Rutin
Background
Breast cancer (BC) is the second most common cancer that leads to the death of women globally [1]. In India and other Asian countries, the incidence of breast cancer continues to be on the rise and every year nearly half a lakh women lose their lives due to this deadly cancer [2]. The BC tumors, which were not expressing estrogen (ER), human epidermal growth factor receptor 2 (HER2) and progesterone receptor (PR) denote as triple-negative breast cancer (TNBC) [3]. A variety of factors associated with breast cancer like age, family history, lifestyle, hormonal regulation, alcohol consumption, and smoking are chiefly responsible for this pathogenesis [4].

There are several biomarkers used for diagnosis of breast cancer [5]. Over-expressed genes in early stages of breast cancer, such as tumor protein 53 (TP53), phosphoinositide-3-kinase-catalytic-alpha polypeptide (PIK3CA), human epidermal growth factor receptor 2 (HER2) ERBB2, fibroblast growth factor receptor 1 (FGFR1), myelocytomatosis oncogene (MYC), and phosphatase and tensin homolog (PTEN) [6] given in (Table 1) are extensively studied. Humans, aldehyde dehydrogenases (ALDHs) are mainly involved in aldehyde detoxification and are also important in the synthesis of retinoic acid [7]. They also play a vital role in alcohol metabolism by removing aldehydes [8]. Among 18 different isoforms of ALDH, the activity of each isoform differs from the other according to the localization and tissue distribution.

Several studies have reported that ALDH is a specific cancer stem cell (CSC) marker in solid tumors like breast, colon, and lung. The isoforms ALDH1L1, ALDH1A1, ALDH1A3, ALDH2, ALDH1A2, ALDH3A1, and ALDH7A1 are associated with different types of tumors. Among these genes, ALDH1A3 which is predominantly expressed in kidney, salivary glands, stomach, and thus over-expressed in breast cancer [9]. Diverse chemotherapeutic treatments also interfere with increased activity of ALDH. Certain biomarkers that lead to breast cancer can be used for proper analysis and further treatment [10]. Recent data suggest that there is a high demand for naturally extracted bioactive compounds for their anticancer properties worldwide [11]. Few findings also suggest that some phytochemicals can stimulate apoptosis in malignant cells in both in vivo and in vitro [12].

The rutin (quercetin glycoside) has tremendous clinical applications which include anti-inflammatory, anti-oxidant, and inhibition of platelet aggregation [13]. Rutin acts as an adjuvant agent that increases the cytotoxic efficiency of two important chemotherapeutic drugs viz., cyclophosphamide (CYC) and methotrexate (MTX) in MDA-MB-231 breast cancer cells [14]. In our present study, we investigated the differential expression patterns of all isoforms of ALDH genes in human breast cancer tissue samples and also synthesized the rutin-copper complex to understand the differences in the efficiency of anti-cancer potentials. In silico molecular modelling studies lead to new innovation in designing novel drugs as this study requires less time and can be performed with large quantities of ligands. Besides, it is easy to compare and is found effective in screening [15]. In this study, we have used 12 natural bioactive compounds (Table 2) against ALDH1A3 and ALDH18A1 proteins through computational in silico docking. Next-generation sequencing (NGS) was carried out for ALDH2 gene using illumina platform.

Methods
Selection of patient sample and materials used
RNA isolation kit RNA easy Minikit (Cat-74104), Agilent’s Quick-Amp labeling Kit (p/n 5190-0442), TrueSeq DNA sample preparation kit (Illumina, #FC-121-1001), Agencourt AMPURE XP beads (Beckman Coulter, #A63881), 7500 Bioanalyzer Chips High Sensitivity Bioanalyzer Chips (Agilent, #5067-4626), Oligo aCGH hybridization kit (Agilent, # 5188-5380), Microarray slide backings (Agilent, # G2534-6005), Human Cot-1 DNA (Invitrogen, #15279-011), Oligo aCGH Wash buffer 1 and 2 set (Agilent, #5188-5226), and Nuclease free water (Ambion, #AM9939),

Table 1 Overexpressed genes in early breast cancer

| Abbreviations | Gene | Function |
|---------------|------|----------|
| TP53          | Tumor protein 53 | Tumor suppressor |
| PIK3CA        | Phosphoinositide-3-kinase-catalytic-alpha polypeptide | Regulates PI3K signalling. PI3K signalling is important for cell proliferation, migration of cells |
| ERBB2 (HER2)  | Human epidermal growth factor receptor 2 (HER2) | Play an important role in breast cancer progression |
| FGFR1         | Fibroblast growth factor receptor 1 | Cell differentiation, growth, proliferation, migration |
| MYC           | Myelocytomatosis oncogene | Acts as protooncogene |
| PTEN          | Phosphatase and tensin homolog | Control cell movement, adhesion of cells to surrounding tissues, and helps in angiogenesis |

Several studies have reported that ALDH1L1, ALDH1A1, ALDH1A3, ALDH2, ALDH1A2, ALDH3A1, and ALDH7A1 are involved in human breast cancer tissue samples and also synthesized the rutin-copper complex to understand the differences in the efficiency of anti-cancer potentials. In silico molecular modelling studies lead to new innovation in designing novel drugs as this study requires less time and can be performed with large quantities of ligands. Besides, it is easy to compare and is found effective in screening [15]. In this study, we have used 12 natural bioactive compounds (Table 2) against ALDH1A3 and ALDH18A1 proteins through computational in silico docking. Next-generation sequencing (NGS) was carried out for ALDH2 gene using illumina platform.
Bioruptor (Diagenode). Data of the breast cancer patients without any major clinical issues were included and nine tissue samples were collected from Dharwad, India. The tissue samples were stored at –70 °C in RNAlater solution for further experimental use.

**Isolation of RNA and labeling the target**

RNA easy Minikit (Cat-74104) from Qiagen was used to isolate total RNA from tissue samples. RNA integrity analysis, cDNA synthesis, labeling, and microarray methods were carried out sequentially according to Kulkarni et al. [16].

**Next-generation DNA sequencing**

The next-generation DNA sequencing was performed at genotypic technologies private laboratory Bangalore. For this study nine breast cancer tissue samples were collected and used for sequencing of chromosome 12 of gene ID ENSG00000111275 using illumina sequencing platform. 100 bp reads of illumina reads were sequenced using Illumina GAIIx Analyzer. The alignment software Burrows-Wheeler Aligner (BWA version 0.5.9 r16) was used to perform the gapped alignment of illumina sequences against the reference sequence. Alignment software bowtie (version 0.12.7) was used to perform ungapped alignment of illumina sequences against the reference sequence. Samtools (version samtools-0.1.7a) was used for calling SNPs. The procedure flow chart is given in Fig. 1.

**Synthesis of rutin–copper complex**

The stoichiometric mixture of rutin and copper chloride dehydrate of 1 mmol and 1.5 mmol were used respectively in methanol solvent and reflux for 5 h. The resulting pale yellow precipitate was dried and washed three times with a mixture of 1:3 (ethanol-water) and was dried. The resultant was used for further experimentation and analysis [17].

**Cell culture and treatments**

Breast cancer cells (MDA-MB-468) were cultured in Dulbecco’s modified eagle medium (DMEM) which was supplemented with 10% heat-inactivated fetal bovine serum (FBS) and 1x anti-myocotic solution. In each well of 96 wells flat bottom plate 200 μl of the cell suspension containing approximately 10,000 cells were seeded and cultured in CO2 incubator at 37 °C in 95% humidity by supplying 5% CO2. The 100–600 μg/ml concentration of rutin and rutin-copper complex samples was used in the treatment of cells. The IC50 value for both rutin and rutin-copper complex was calculated after 24 h of experimental incubation time using GraphPad prism version 5.1 [18].

---

**Table 2** Name of natural bioactive compounds and their structures

| Name of the compound | Structure |
|----------------------|----------|
| Rutin                | ![Structure of Rutin] |
| Epigallocatechin     | ![Structure of Epigallocatechin] |
| Resveretrol          | ![Structure of Resveretrol] |
| Arctigenin           | ![Structure of Arctigenin] |
| Camptothecin         | ![Structure of Camptothecin] |
| 3,3′-diindolylmethane | ![Structure of 3,3′-diindolylmethane] |
| Genistein            | ![Structure of Genistein] |
| Silibinin            | ![Structure of Silibinin] |
| Thymoquinone         | ![Structure of Thymoquinone] |
| 28-homocastasterone  | ![Structure of 28-homocastasterone] |
| 24-Epibrassinolide   | ![Structure of 24-Epibrassinolide] |
| Ajouene              | ![Structure of Ajouene] |
Fig. 1 Systematic workflow of NGS
Molecular docking studies

The molecular docking studies were performed on 12 natural compounds against ALDH1A3 (5FHZ) and ALDH18A1 (Protein data bank PDBID – 2H5G) using Schrödinger Maestro 11.2 version. The 3D structure of the protein ALDH1A3 (PDB ID – 5FHZ) was retrieved from PDB. The protein was optimized accordingly, hydrogen was added and water components surrounding greater than 5 Å were removed. Finally, it was minimized to the default root mean square deviation (RMSD) value of 0.30 using protein preparation wizard. Receptor grid generation was performed for minimized protein to identify the binding site of the ligand to the protein. The structures of natural compounds were drawn using 2D sketcher. They were performed using the Ligprep module to generate the 3D tautomer of the compound and to generate all combinations up to 32 per ligand. Ligand docking was performed according to the method of Kalirajan [19]. In Maestro 11.2 version dock, the binding of natural ligands to the active site of the human ALDH1A3 is an automated process with extra precision (XP) descriptor information and added epik state penalties to docking score. The best bioactive conformation was finalized based on the highest G score and highest number of interactions with the protein, mostly the number of hydrogen bonds. The XP visualizer was used as a tool to scrutinize the interactions. The scoring method was the most accepted design, with different parameters like polar, salvation, repulsive, hydrophobic, electrostatic, activity, and G score.

Results

The microarray analysis has given an insight on differential expressions of ALDH isoforms in breast cancer tissues. ALDH1A3 and ALDH18A1 with highest positive mean fold variation and ALDH2 with highest negative mean fold variation were noted as depicted in Table 3. These results suggest that ALDH1A3 and ALDH18A1 act as biomarkers for breast cancer. Thus, these may be targeted further for designing inhibitors.

| Gene name | Mean fold variation in breast cancer |
|-----------|-------------------------------------|
| ALDH3B1   | − 0.60                              |
| ALDH4A1   | − 2.0                               |
| ALDH8A1   | − 0.18                              |
| ALDH6A1   | − 2.22                              |
| ALDH9A2   | − 2.22                              |
| ALDH5A1   | − 1.26                              |
| ALDH9A1   | − 1.03                              |
| ALDH16A1  | − 0.31                              |
| ALDH1B1   | − 0.52                              |
| ALDH1A2   | − 3.12                              |
| ALDH1L1   | − 2.60                              |
| ALDH18A1  | 1.93                                |
| ALDH1A3   | 2.22                                |
| ALDH3A1   | − 1.03                              |
| ALDH3B2   | 0.42                                |
| ALDH2     | − 4.19                              |
| ALDH7A1   | 0.24                                |
| ALDH1A1   | − 2.13                              |

Negative (−) sign indicates down regulation of the gene

| Chromosome Chromosome position | Gene ID       | Gene name | Biotype protein coding (PC) | Transcript ID | Variation base sample B | SNP quality sample B | Old codon/new codon sample B |
|-------------------------------|---------------|-----------|----------------------------|---------------|-------------------------|----------------------|-----------------------------|
| chr12 1.12E + 08              | ENSG00000111275 | ALDH2     | PC                         | ENST00000552234 | G                       | 30                   | gcT/gcG                      |
| chr12 1.12E + 08              | ENSG00000111275 | ALDH2     | PC                         | ENST00000548536 | G                       | 30                   | gcT/gcG                      |
| chr12 1.12E + 08              | ENSG00000111275 | ALDH2     | PC                         | ENST00000549106 | G                       | 30                   | −                           |
| chr12 1.12E + 08              | ENSG00000111275 | ALDH2     | PC                         | ENST00000416293 | G                       | 30                   | gcT/gcG                      |
| chr12 1.12E + 08              | ENSG00000111275 | ALDH2     | PC                         | ENST00000553044 | G                       | 30                   | −                           |
| chr12 1.12E + 08              | ENSG00000111275 | ALDH2     | PC                         | ENST00000546840 | G                       | 30                   | gcT/gcG                      |
| chr12 1.12E + 08              | ENSG00000111275 | ALDH2     | PC                         | ENST00000261733 | G                       | 30                   | gcT/gcG                      |
| chr12 1.12E + 08              | ENSG00000111275 | ALDH2     | PC                         | ENST00000551906 | G                       | 30                   | Tgg/Ggg                      |

Table 4 Single-nucleotide polymorphism (SNP) with codon mutation from T to G in ALDH2 gene in breast cancer samples

Table 3 Showing the list of fold variation in terms of gene expression profiling in ALDH isoforms
Next-generation sequencing (NGS)
To date, hardly any studies examined the role of ALDH2*2 in the development of breast cancer and none of them found any association between ALDH*2 in correlation with the risk of breast cancer [20]. These studies have shown G to A point mutation in ALDH2*2. However, in our study, a different novel SNP 'T to G point mutation' was observed in breast cancer samples and is depicted in Table 4.

Effect of rutin and rutin-copper complex on cell viability was analyzed through in vitro studies. The analysis was carried out on MDA-MB-468 cells using cell viability assay 3-(4,5-Dimethylthiazol-2-yl]-2,5 diphenyl tetrazolium bromide (MTT assay) for different time periods. The IC<sub>50</sub> value of rutin was 144.50 μg/ml and the rutin-copper complex was 119.40 μg/ml in MDA-MB-468 cells, with a difference of 25.10 μg/ml (Fig. 2). The lesser IC<sub>50</sub> value of the rutin-copper complex may suggest a suitable replacement of the rutin to act as an adjuvant.

Molecular docking studies have revealed that rutin had a best binding affinity with both ALDH1A3 and ALDH18A1 among different natural bioactive compounds as shown in Figs. 3 and 4. Schrödinger maestro docking scores of ligands against ALDH1A3 and ALDH18A1 are given in (Tables 5 and 6), respectively.

Discussion
Aldehyde dehydrogenases have a variety of biological effects. It has specifically been linked to different types of cancer, Parkinson’s disease, cataract, obesity, etc. ALDH1A1 is highly expressed in tissues but not true in case of breast cancer. ALDH1A3 is a marker for breast cancer patients [21]. In the present study, ALDH1A3 gene was highly expressed compared to normal as reported earlier [22] but ALDH18A1 also exhibited high expression in breast cancer. This ultimately proves that ALDH18A1 also acts as a biomarker along with ALDH1A3. Hence, identifying inhibitors against these may help in better management of breast cancer.

Currently, chemotherapy is widely used to treat a variety of cancers which has its limitations with high toxic side effects to the patients [23]. Hence, research on an alternative approach to treat cancer without side effects is in high demand. Few natural products like flavonoids and polyphenols have better anticancer activity [24]. In silico molecular docking study revealed that rutin has higher binding affinity to both over-expressed ALDH1A3 and ALDH18A1 gene among 12 selected natural compounds.

As previously reported, the synergistic action of rutin (glycoside) with other chemo drugs, manifests its potential role as an adjuvant for the treatment of breast cancer patients. Recent study reports on flavonoids complexing with macromolecules to enhance their bio-availability and chemo-preventive efficacy. Rutin (Ru)-fucoidan (Fu) complex was used to overcome the limitations of bio-availability of rutin molecule [25]. Hence, in our current study, a rutin-copper complex was synthesized to study the effects
**Fig. 3**  

a The probable binding mode of all ligands to ALDH1A3.  
b Interaction diagram of rutin ligand with ALDH1A3 protein: the yellow dotted lines in 3D figure indicates the hydrogen bond between the ligand and protein receptor.  
c Two dimensional interaction diagram indicates the rutin interaction with ALDH1A3, one hydrogen bond is between the ligand and amino acid 471 leucine, 469 asparagine, 92 aspartic acid, 159 aspartic acid, and 139 arginine residue in receptor of protein.
Fig. 4  

a The probable binding mode of all ligands to ALDH18A1.  
b Interaction diagram of rutin ligand with ALDH18A1 protein: the yellow dotted lines in 3D figure indicates the hydrogen bond between the ligand and protein receptor.  
c Two-dimensional interaction diagram indicates the rutin interaction with ALDH18A1, one hydrogen bond is between the ligand and amino acid 735 leucine, two hydrogen bond with 749 aspargine, 755 glutamic acid, and 747 histidine residue in receptor of protein.
of both rutin and its complex with copper for their potential bioactivity against breast cancer cells. We noticed that the IC\textsubscript{50} value decreased to a large extent for rutin-copper complex compared to rutin. Hence, it suggests that the rutin-copper complex is a lead molecule for further studies in replacing rutin to act as an adjuvant in treating breast cancer patients. Further in silico molecular dynamics simulation study and in vivo studies also act as supportive information to understand the nature of rutin-copper complex.

A rare mutation lys/lys genotype of rare SNP rs671, which is associated with increased breast cancer in Asian women, has been recently reported [26]. However, in this study, the expression of ALDH2 was very low in breast cancer tissue samples compared to normal breast tissues. Further, it showed different novel SNP with T to G rather than G to A.

### Conclusion

Like ALDH1A3, ALDH18A1 also acts as a biomarker for breast cancer and a novel ALDH2 single-nucleotide polymorphism may also be a prognostic indicator for breast cancer in south Indian women. The computational study reveals that Rutin has greater binding effect on both ALDH1A3 and ALDH18A1 proteins. The rutin-copper complex has a greater effect on breast cancer cell lines as compared to rutin alone. These results should be further validated with in vivo studies and this study needs further investigation with a large number of patient samples.

### Table 5 Schrödinger maestro docking scores of ligands against ALDH1A3

| Ligand             | G score | Dock score | Lipophilic EvdW | PhobEn | Hydrogen bond | Electro | Site map |
|--------------------|---------|------------|----------------|--------|---------------|---------|----------|
| Rutin              | 13.68   | 13.67      | 4.28           | 0.4    | 7.17          | 1.93    | 0.05     |
| Epigallocatechin   | 10.41   | 10.35      | 3.96           | 0.2    | 5.32          | 1.07    | 0.32     |
| Silbibin           | 8.94    | 8.94       | 4.49           | 0.19   | 3.53          | 0.84    | 0.17     |
| Genistein          | 8.28    | 8.26       | 4.14           | 0.8    | 2.23          | 0.72    | 0        |
| 24epibrassinolide  | 7.79    | 7.79       | 2.76           | 0.68   | 3.81          | 0.89    | 0        |
| 28homocastasterone | 7.74    | 7.74       | 5              | 0.44   | 2.21          | 0.21    | 0.14     |
| Arctigenin         | 7.51    | 7.51       | 5.02           | 1.59   | 0.7           | 0.23    | 0        |
| 33’diindolylmethane| 7.44    | 7.44       | 4.97           | 1.11   | 0.7           | 0.4     | 0        |
| Resvertrol         | 7.2     | 7.2        | 3.51           | 0.96   | 1.61          | 0.89    | 0        |
| Camptothecin       | 7.13    | 7.13       | 5.05           | 0.54   | 0.96          | 0.2     | 0.12     |
| Thymoquinone       | 5.1     | 5.1        | 2.56           | 0.64   | 0.7           | 0.1     | 0.6      |
| Ajoene             | 4.09    | 4.09       | 3.4            | 0.98   | 0.7           | 0.27    | 0.38     |

### Table 6 Schrödinger maestro docking scores of ligands against ALDH18A1

| Ligand             | G score | Dock score | Lipophilic EvdW | PhobEn | Hydrogen bond | Electro | Site map |
|--------------------|---------|------------|----------------|--------|---------------|---------|----------|
| Rutin              | 11.66   | 11.65      | 3.75           | 0      | 6.38          | 1.58    | 0.11     |
| Epigallocatechin   | 8.37    | 8.31       | 3.42           | 0      | 4.34          | 1.04    | 0        |
| Silbibin           | 6.6     | 6.6        | 2.49           | 0      | 3.28          | 1.15    | 0.09     |
| Resvertrol         | 6.35    | 6.35       | 2.33           | 0      | 2.8           | 1       | 0        |
| 28homocastasterone | 5.73    | 5.73       | 2.18           | 0      | 3.29          | 0.72    | 0        |
| Genistein          | 5.13    | 5.11       | 2.41           | 0      | 2.12          | 0.99    | 0        |
| 24epibrassinolide  | 4.97    | 4.97       | 1.61           | 0      | 2.82          | 1.68    | 0        |
| Arctigenin         | 4.64    | 4.64       | 2.79           | 0      | 1.62          | 0.53    | 0        |
| 33’diindolylmethane| 2.66    | 2.66       | 2.8            | 0      | 0.7           | 0.15    | 0        |
| Camptothecin       | 2.35    | 2.35       | 2.11           | 0      | 0.83          | 0.29    | 0.24     |
| Ajoene             | 2.03    | 2.03       | 2.57           | 0      | 0.62          | 0.31    | 0.18     |
| Thymoquinone       | 1.39    | 1.39       | 1.2            | 0      | 0.35          | 0.21    | 0.09     |
