



# Functional clustering analysis identifies specific subtypes of aldehyde dehydrogenase associated with glioma immunity

Di Wang<sup>1#</sup>, Haoyu Jiang<sup>1#</sup>, Zhiliang Wang<sup>1</sup>, Ruoyu Huang<sup>2</sup>, Tao Jiang<sup>1,2,3,4,5</sup>, Ulf Dietrich Kahlert<sup>2,6,7,8</sup>, Guanzhang Li<sup>1</sup>, Wei Zhang<sup>1,5</sup>

<sup>1</sup>Department of Neurosurgery, Beijing Tiantan Hospital, Capital Medical University, Beijing, China; <sup>2</sup>Beijing Neurosurgical Institute, Capital Medical University, Beijing, China; <sup>3</sup>Center of Brain Tumor, Beijing Institute for Brain Disorders, Beijing, China; <sup>4</sup>China National Clinical Research Center for Neurological Diseases, Beijing, China; <sup>5</sup>Chinese Glioma Genome Atlas (CGGA), Beijing, China; <sup>6</sup>Neurosurgical Clinic, Medical Faculty, Heinrich-Heine University Duesseldorf, Duesseldorf, Germany; <sup>7</sup>German Cancer Consortium, DKTK, Essen/Duesseldorf, Germany; <sup>8</sup>Medizinische Fakultät, Otto-von-Guericke-Universität Magdeburg, Magdeburg, Germany

*Contributions:* (I) Conception and design: UD Kahlert, G Li, W Zhang; (II) Administrative support: T Jiang; (III) Provision of study materials or patients: Z Wang, R Huang; (IV) Collection and assembly of data: D Wang, H Jiang; (V) Data analysis and interpretation: D Wang, H Jiang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

<sup>#</sup>These authors contributed equally to this work.

*Correspondence to:* Wei Zhang. Department of Neurosurgery, Beijing Tiantan Hospital, Capital Medical University, 119 South Fourth Ring Western Road, Fengtai District, Beijing 100070, China. Email: zhangwei\_vincent@126.com; Guanzhang Li. Department of Neurosurgery, Beijing Tiantan Hospital, Capital Medical University, 119 South Fourth Ring Western Road, Fengtai District, Beijing 100070, China. Email: liguanzhang122@163.com; Ulf Dietrich Kahlert. Medizinische Fakultät, Otto-von-Guericke-Universität Magdeburg, Magdeburg, Germany. Email: ulf.kahlert@med.ovgu.de.

**Background:** As a novel cancer stem cell marker, the biological functions of aldehyde dehydrogenases (ALDH enzymes) are a hot topic in current cancer research. However, there is a lack of systematic understanding of ALDH enzymes, which has hindered the translation of targeted therapies from bench work into the clinic.

**Methods:** Based on transcriptome data from 999 glioma patients, functional clustering analysis was performed to reveal the functional phenotypes of ALDH isoforms. Subsequently, ALDH subgroups closely related to gliomas were identified by Cox survival analysis. Finally, gene set variation analysis (GSVA) and Pearson correlation analysis were used to explore the biological functions of ALDH enzymes.

**Results:** Our study found that ALDH enzymes could be classified into 5 subgroups, among which 3 groups were closely related to malignant progression and the prognosis of gliomas. We found that ALDH enzymes were closely related to gene mutations, which were most likely caused by changes in DNA repair functions. Further studies revealed that ALDH enzymes affect tumor immune functions, especially the expression of immune checkpoints. The effectiveness of immune checkpoint inhibitors in treating glioma might be improved by altering the expression of ALDH enzymes in specific subgroups.

**Conclusions:** This study comprehensively revealed the biological functions of ALDH enzymes in glioma and provided details about the potential clinical application of targeted therapy for ALDH enzymes.

**Keywords:** Acetaldehyde dehydrogenase (ALDH); glioma; DNA repair; tumor immunity; prognosis

Submitted Jul 03, 2021. Accepted for publication Nov 11, 2021.

doi: 10.21037/tcr-21-1160

View this article at: <https://dx.doi.org/10.21037/tcr-21-1160>

## Introduction

Gliomas are one of the most aggressive and infiltrative tumors in the central nervous system (CNS). The 5-year survival rate of patients with glioblastoma multiforme (GBM), which is classified as a grade IV glioma by the World Health Organization (WHO), is less than 5% (1). Therefore, new effective therapeutic methods for gliomas are an expanding focus of research, including immunotherapy and small molecule inhibitors therapy (2,3). Isocitrate dehydrogenase (IDH) mutation, which involves the metabolism of cancer cells, is one of the most important mutations of gliomas (4). Clinical and basic research about IDH mutation has made encouraging progress, such as identification of the inhibitor of IDH (5). Emerging evidence has shown that 2-hydroxyglutarate (2-HG), a metabolite of mutated IDH, contributes to the inhibitory immune microenvironment in glioblastoma (6). Research has implied that distinct subtypes of metabolic enzymes may have crucial roles in the crosstalk between tumors and their immune microenvironments, rather than acting independently within the tumor.

Aldehyde dehydrogenases (ALDH enzymes), as oxidizing enzymes which are composed of 19 isoforms, play a crucial part in cell proliferation, differentiation, metabolism, and other functions, and recent evidence has suggested that a high expression of ALDH is related to cancer malignancy (7). Moreover, mounting evidence shows that ALDH, particularly ALDH1A3, has close ties to tumorigenesis and tumor progression (8-10). Interrupting the ALDH metabolic pathway has become a promising therapeutic strategy. However, side effects induced by the inhibition of ALDH enzymes remain unclear. We hypothesized that a target specific ALDH isozyme would cause an increased expression of one or several ALDH isozymes for functional compensation. At present, classification of ALDH enzymes is mainly based on the molecular structure of the enzymes, but this has not been confirmed in clinical application (11). Therefore, reclassification of ALDH isoforms based on their functions is urgently needed.

In this study, we developed a new approach for the reclassification of ALDH enzymes. We divided ALDH enzymes into 5 subgroups based on their functional evaluation across 2 databases, the Chinese Glioma Genome Atlas (CGGA) and The Cancer Genome Atlas (TCGA). These 5 subgroups had different characteristics, and all ALDH enzymes in the same subgroup had similar

functions. Importantly, we identified 3 subgroups of ALDH enzymes that relate to tumor immune functions. This new classification displayed the biological functions of different ALDH enzymes and could improve the clinical application of ALDH inhibitors in glioma treatment. This reclassification process enabled us to better understand the functions and characteristics of ALDH enzymes in gliomas. Furthermore, our study provided a theoretical basis for ALDH functional research and ALDH targeted immunotherapy of glioma patients. We present the following article in accordance with the Materials Design Analysis Reporting (MDAR) checklist (available at <https://dx.doi.org/10.21037/tcr-21-1160>).

## Methods

This study included messenger RNA (mRNA) sequencing data from 999 patients with gliomas of all grades. The sequencing data of the test group were from the CGGA. The sequencing was performed with the Illumina HiSeq 2000 platform (Illumina, San Diego, CA, USA). Only samples that consisted of more than 80% of tumor cells were used. Overall survival (OS) was calculated from the date of diagnosis to death or the last follow-up. Transcriptome sequencing data and patient information can be freely downloaded from the CGGA portal (<http://www.cgga.org.cn>). The sequencing data of the validation group were from TCGA mRNA-seq database downloaded from public databases (<https://cancergenome.nih.gov>). The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the Ethics Committee of Beijing Tiantan Hospital, Capital Medical University (KY2020-093-02). The institutional review board decided that informed consent was not required.

### *Functional clustering analysis*

The biological function and signal pathway activation scores of each participant were calculated by gene set variation analysis (GSVA). The calculation process selected default settings. The Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) gene sets were downloaded from gene set enrichment analysis (GSEA) web portals (<http://software.broadinstitute.org/gsea/index.jsp>). Subsequently, the correlation between biological function scores and gene expressions was calculated using Pearson correlation analysis. A matrix of

correlation coefficients of all patients with 4,653 GO terms and 186 KEGG terms was obtained. Finally, functional clustering results were acquired using unsupervised cluster analysis based on the data above. The integrated clustering analysis was performed by Cluster-of-Cluster Assignments analysis as previously described (12). All statistical computations and figure drawing were performed with the statistical software environment R version 3.5.0 (<http://www.r-project.org/>).

### ***Chromosome stability and DNA repair functions***

The chromosome stability was evaluated by aneuploidy scores, which reflect the alterations of chromosome arm-level copy-numbers in a sample (13). This score can be further divided into chromosome arm-level loss and gain scores. The GSVA scores of DNA repair gene set evaluated DNA repair functions. The relationship between ALDH enzymes and the above functions was analyzed using Pearson correlation analysis. Furthermore, the correlation analyses between ALDH enzymes and leukocyte, tumor purity, mutation number, and stemness index were also performed with the same method.

### ***Immune functions analysis***

The relationship between ALDH enzymes and immune functions was evaluated using the Pearson correlation analysis of ALDH enzymes' mRNA expression and immune function scores. Immune function scores were calculated using GSVA analysis, and the immune function gene set was downloaded from AmiGO 2 web portals (<http://amigo.geneontology.org/amigo/landing>). The classification of immune functions was based on the classification in the AmiGO2 website.

### ***Prognostic analysis***

The prognostic value of ALDH enzymes or ALDH groups was estimated by Kaplan-Meier analysis or Cox analysis. Kaplan-Meier analysis was performed with the statistical software environment R version 3.5.0, and Cox analysis was performed with the statistical software SPSS 25.0 (IBM Corp., Armonk, NY, USA). The median was used as the cutoff of each group in all prognostic analyses. When more than 2 genes were involved in the grouping, each gene group's intersection was used as a grouping standard.

### ***Statistical analysis***

All statistical computations and figure drawing were performed with R packages (ggplot, pheatmap, and survival), GraphPad Prism 7.0 (GraphPad Software, La Jolla, CA, USA), and Microsoft Office 2016 pro (Microsoft, Redmond, WA, USA). Statistical differences between 2 populations were calculated by *t*-test. The log-rank test was used to assess the statistical significance of stratified survival groups. For all statistical methods,  $P < 0.05$  was considered a significant difference.

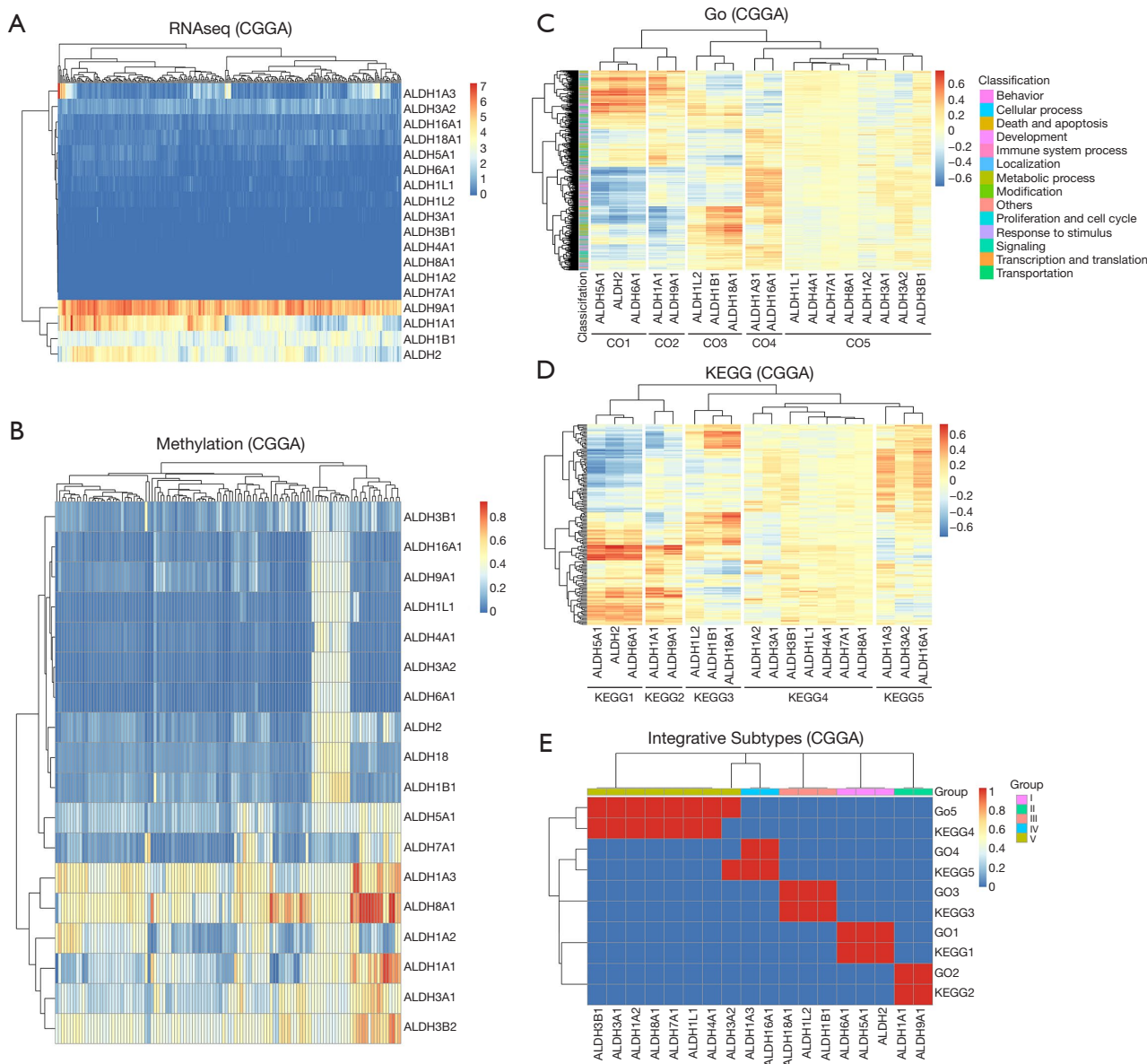
## **Results**

### ***A new classification of ALDH enzymes based on their biological functions and signaling activation***

To reclassify ALDH isoforms based on genome, we performed unsupervised cluster analysis on their mRNA expression level and gene methylation level in the CGGA database (*Figure 1A,1B*). However, no significant classification trend was observed. A similar situation was seen in TCGA database (*Figure S1A,S1B*). Therefore, we developed a new approach to reclassify ALDH enzymes based on their functions. When using the new functional clustering method (details in methods), the isoforms of ALDH enzymes could be divided into 5 groups according to their biological functions (GO groups 1–5) and signal pathway activations (KEGG groups 1–5), respectively (*Figure 1C,1D*; <https://cdn.amegroups.com/static/public/tcr-21-1160-1.xlsx>). Subsequently, we saw similar results in the integrated cluster analysis (*Figure 1E*). As expected, functional clustering analysis also worked well in TCGA database (*Figure S1C-S1E*). Notably, the classification results were highly consistent in both databases.

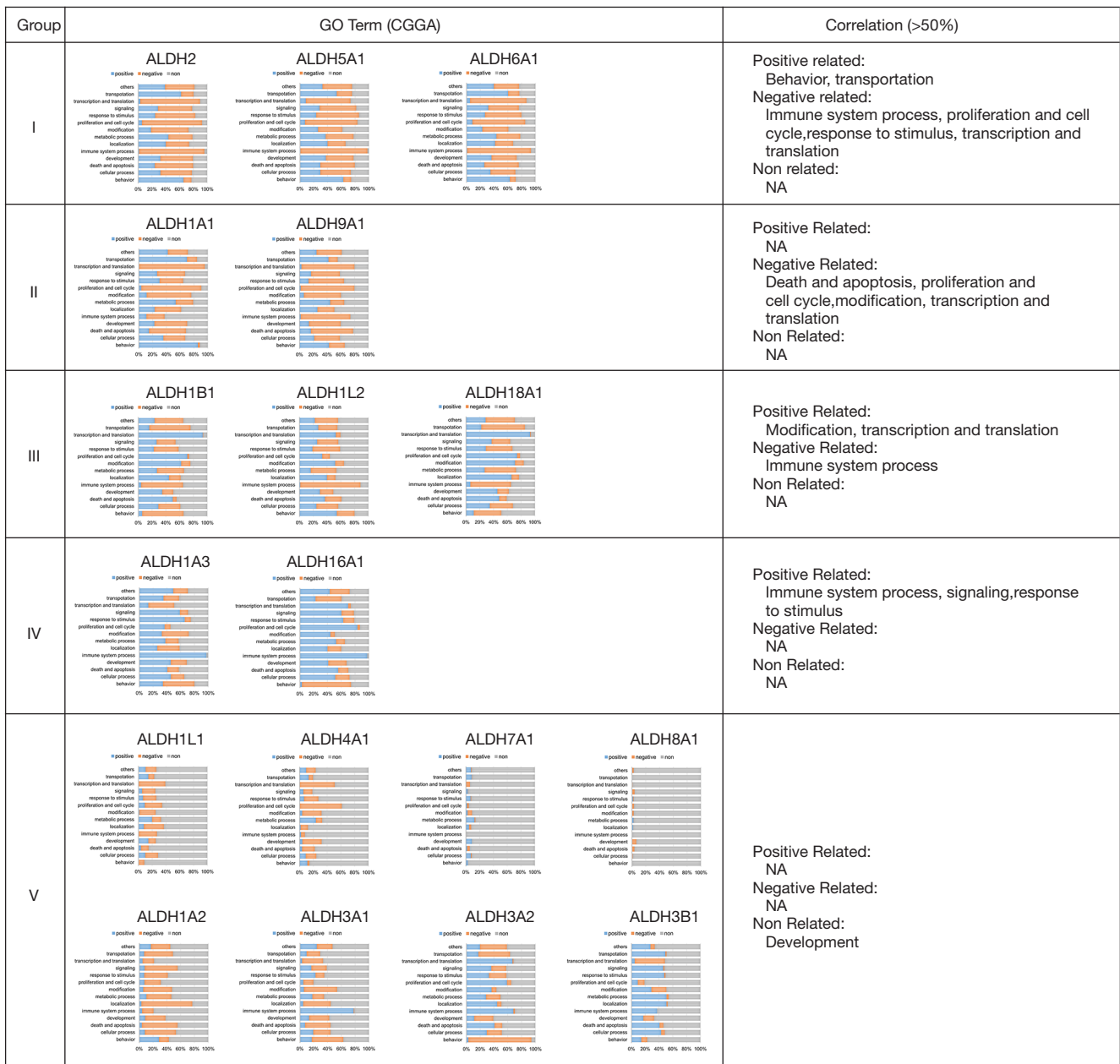
### ***Different ALDH enzymes in the same subgroup most likely play similar biological roles in glioma***

The landscape of the biological functions of ALDH enzymes revealed that each group has its specific functional spectrum in CGGA (*Figure 2*; <https://cdn.amegroups.com/static/public/tcr-21-1160-2.xlsx>) and TCGA databases (*Figure S2*; <https://cdn.amegroups.com/static/public/tcr-21-1160-3.xlsx>). In Group I, all 3 isoforms were positively correlated with most GO terms (>50%) related to behavior and transportation. Meanwhile, they were significantly negatively correlated with most immune system processes,



**Figure 1** Multi-platform cluster analysis of ALDH in CGGA database. (A) Unsupervised clustering of all ALDH isoforms by gene mRNA expression in the CGGA database. (B) Unsupervised clustering of all ALDH isoforms by gene methylation level in the CGGA database. (C) Unsupervised clustering of all ALDH isoforms by the correlation coefficient between gene mRNA expression and 4,653 biological function scores. ALDH isoforms can be classified into 5 groups according to their biological functions. (D) Unsupervised clustering of all ALDH isoforms by the correlation coefficient between gene mRNA expression and 171 signal pathway activation scores. ALDH isoforms can be classified into 5 groups according to their signal pathway activations. (E) Integrated Cluster-Of-Cluster Assignments analysis of all ALDH isoforms reveals that ALDH isoforms can be classified into 5 groups according to their biological functions and signal pathway activations. ALDH, aldehyde dehydrogenase; CGGA, Chinese Glioma Genome Atlas; mRNA, messenger RNA.





**Figure 2** Landscape of biological functions of all ALDH isoforms in 5 groups in CGGA database. Most related biological functions of ALDH isoforms in each group. Only when all ALDH isoforms in a group are significantly related to more than 50% of the biological functions of one classification is the group defined as most related. ALDH, aldehyde dehydrogenase.

proliferation, and cell cycle, response to stimulus, and transcription and translation GO terms. Like Group I, ALDH enzymes in Group II were also negatively related to the majority of death and apoptosis, proliferation and cell cycle, response to stimulus and transcription, and translation GO terms, but did not show a positive relationship with any

biological function category. Group III and ALDH enzymes were highly positively correlated with modification, transcription, and translation, and negatively correlated with the immune system process. Unlike other groups, ALDH enzymes in Group IV were positively correlated with most immune system processes, signaling, and

response to stimulus GO terms. Nevertheless, for Group V, the situation was different; ALDH enzymes showed almost no specific functions.

### ***ALDH functional subgroups are related to clinical characteristics in glioma patients***

To explore the relationship between clinical features and our new classification, we analyzed the ALDH expression of each isoform in different pathological grades based on the 2021 WHO Classification of Tumors of the Central Nervous System. Compared to WHO II grade gliomas, the expression of ALDH enzymes in Group I and Group II was significantly lower than the expression in WHO III, but all ALDH enzymes in Group IV showed the opposite trend. Moreover, a similar situation was observed in Group I and Group IV between WHO III and WHO IV, while ALDH enzymes in other Groups showed no apparent difference (Figure 3A-3E; Figure S3). Univariate Cox analysis was performed to further explore the prognostic values of ALDH enzymes in glioma. Consistent with the above clinical results, high expression of ALDH isoforms in Group I and Group II was a more effective prognostic indicator, and high expression in Group IV was a less favorable prognostic indicator (Figure 3F). In summary, we found that ALDH enzymes, especially those in Group I, Group II, and Group IV, were closely related to the clinical phenotype in glioma patients.

### ***ALDH functional subgroups are related to chromosome stability and DNA repair function***

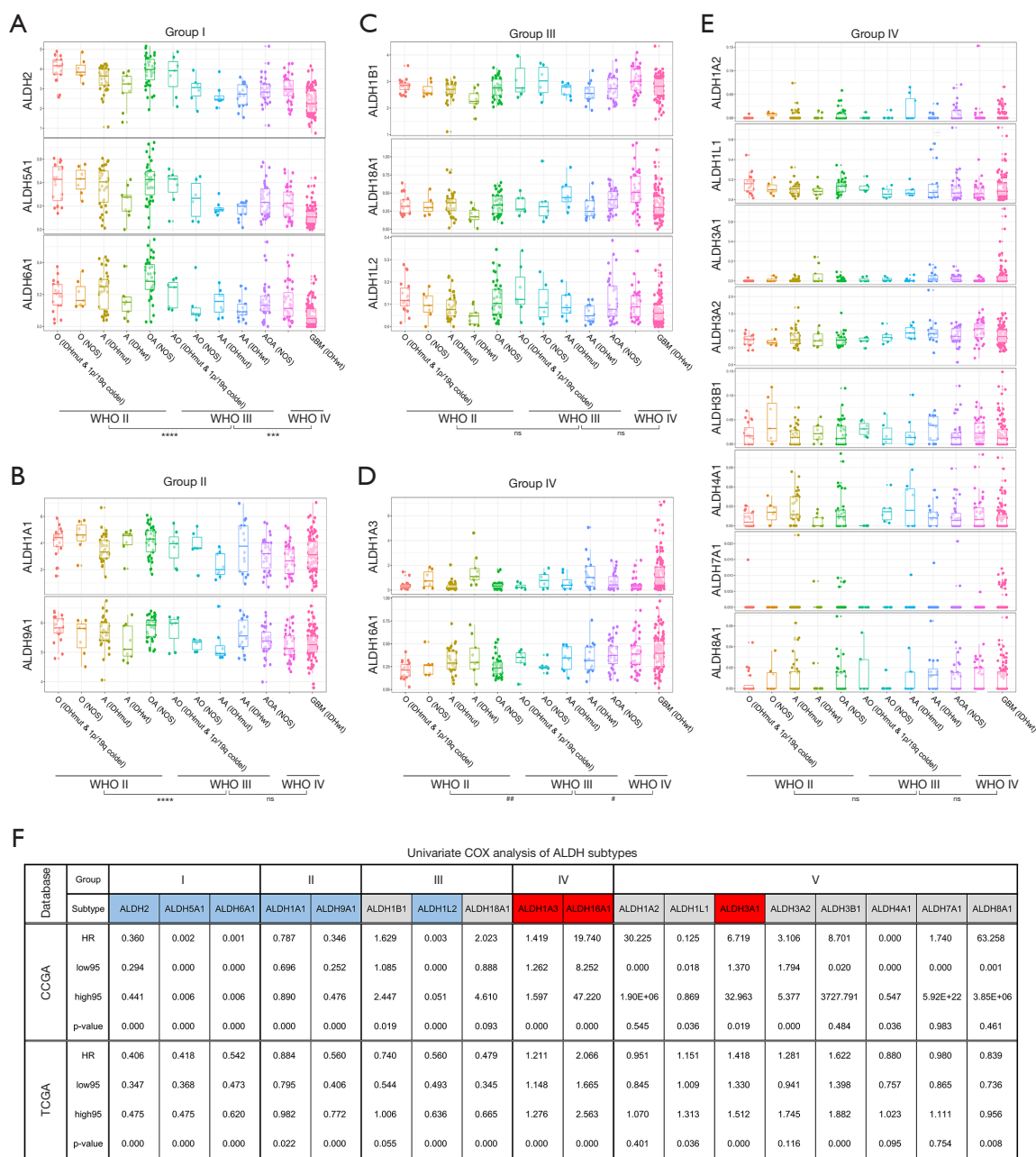
Gene mutation and stability of chromosomes are vital factors that affect patient prognosis. Further studies were performed to verify whether ALDH enzymes in each group were associated with gene mutations (<https://cdn.amegroups.com/static/public/tcr-21-1160-4.xlsx>). The results showed that total mutation counts were significantly lower in the higher expression isoforms of Group I and Group II and significantly higher in the higher expression isoforms of Group IV. No such difference was observed in Groups III and V (Figure 4A). As for the chromosome stability, the trends of differences were consistent with total mutations, whether for chromosome amplification or deletion (Figure 4B,4C). To further explore the causes of DNA mutation, we measured the relationship between all ALDH enzymes and the DNA repair function scores through the correlation coefficient. We found that 2 kinds of DNA repair functions were most

positively related to the expression of ALDH enzymes in Group I and Group II and negatively related to the expression of ALDH enzymes in Group IV.

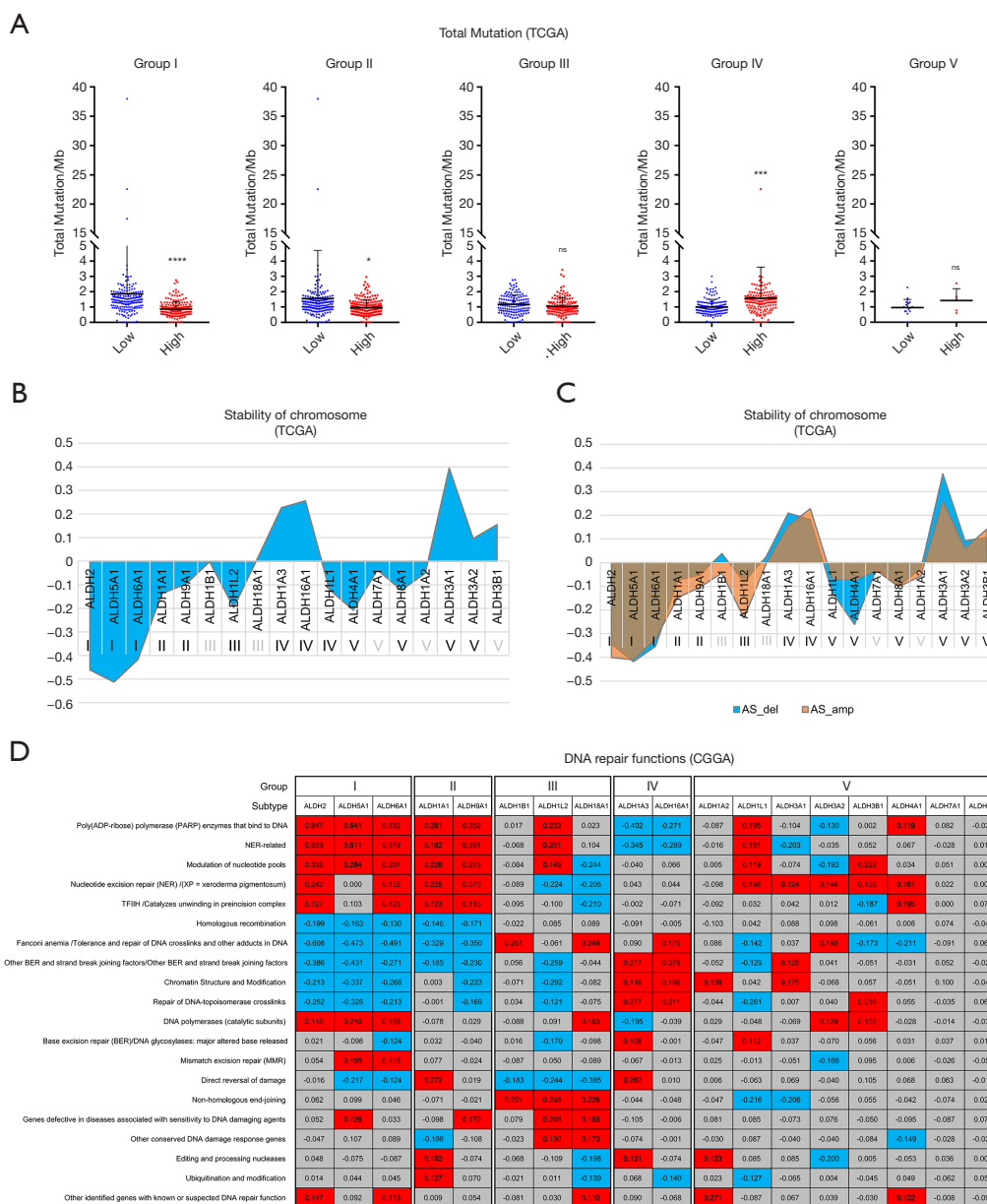
Conversely, a DNA repair function was negatively related to ALDH isoforms in Group I or Group II and positively related to the expression of ALDH isoforms in Group IV (Figure 4D; Figure S4A). Altogether, we demonstrated that ALDH enzymes, particularly those in Group I, II, and IV, were closely related to gene mutation. The critical factor in this process may be DNA repair functions.

### ***ALDH functional subgroups are closely related to the tumor immune microenvironment***

The function of DNA repair is likely to be a crucial reason for the malignant process of glioma, but how it influences the biological function is still unknown. To investigate this, we assessed the correlation between DNA-repair functions and GO terms in CGGA and TCGA databases (Figure 5A; Figure S4B). We found that more than 95% of DNA repair functions had a strong connection with the immune system processes, much more than any other kind of GO term. To understand the differences in immune status between different groups, we performed a correlation coefficient analysis in CGGA and TCGA databases (Figure 5A; Figure S4C). We observed that most immune functions, except for the T cell-mediated immune response to tumor cell (T cell response), were negatively correlated with most ALDH isoforms in Group I, II, and III (Figure 5B). However, the situation in Group IV was different. Most of the immune functions except for T cell response had a positive correlation with ALDH1A3 and ALDH16A1. This abnormal phenomenon made us think of the immune depletion in tumors caused by immune checkpoints. We investigated the correlation between all ALDH enzymes and immune checkpoints in CGGA and TCGA databases (Figure 5C-5H). The results showed that ALDH enzymes in the same group always had a highly similar immune checkpoint expression status. This suggested that immune functions, particularly checkpoints, may have significant influence over the classification of ALDH function. To explore which cell-types express different subtypes of ALDH enzymes. We analysed different sub-classify ALDH enzyme in distinct cell types based on a public scRNA-seq dataset from 28 patients with GBM (GSE131928). These 24,131 single cells were divided into four cell types including macrophages, malignant cells, oligodendrocytes and T cells. The dot plot showed that a majority of ALDH

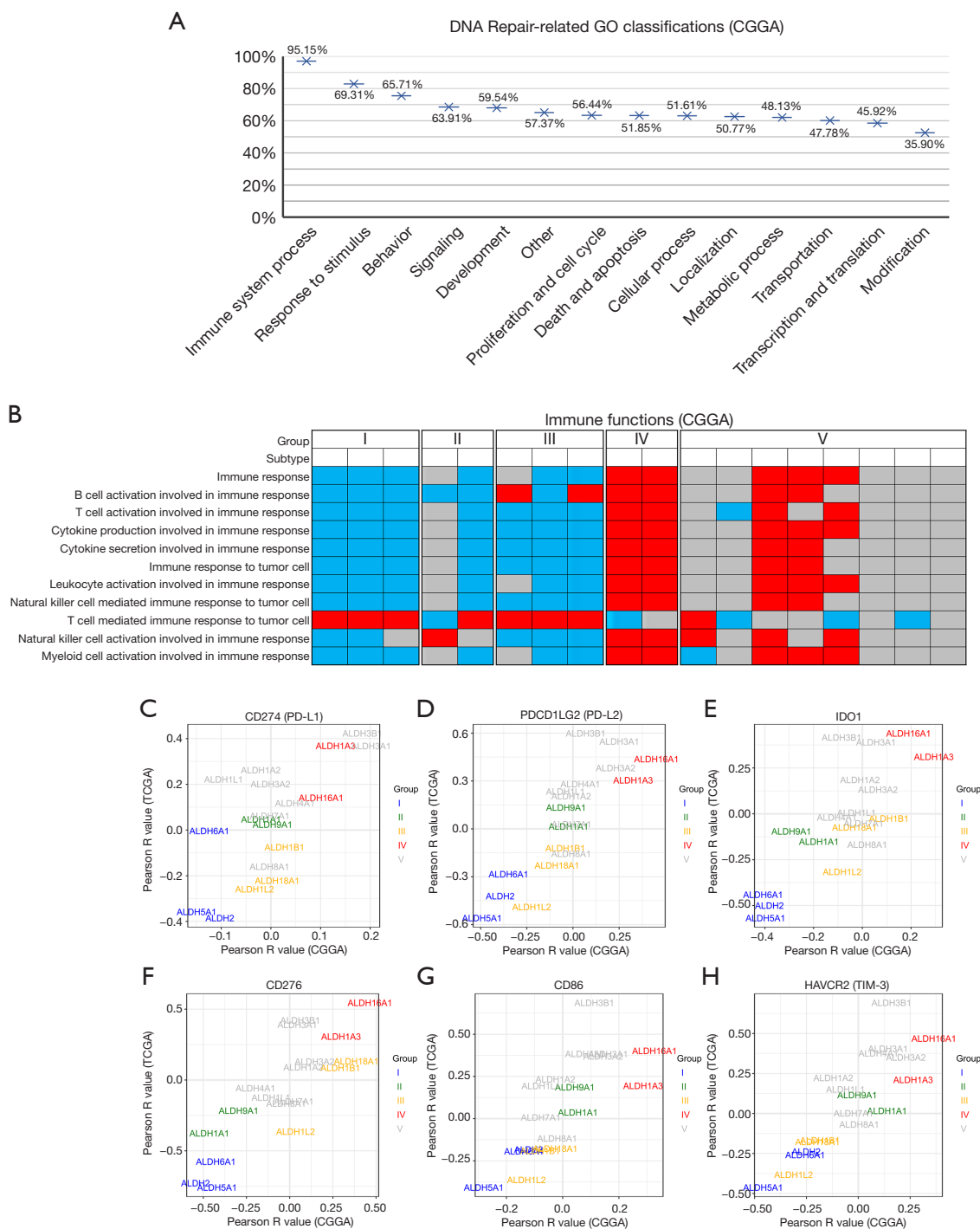


**Figure 3** The clinical features of all ALDH isoforms in 5 groups. (A) In the CGGA database, the expression pattern of ALDH isoforms in Group I. The expression of all ALDH isoforms in Group I decreases as the WHO grade increases. \*\*\*\*,  $P < 0.0001$ , \*\*\*,  $P < 0.001$ . (B) In the CGGA database, the expression pattern of ALDH isoforms in Group II. The expression of all ALDH isoforms in Group II decreases as the WHO grade increases. \*\*\*\*,  $P < 0.0001$ ; ns,  $P > 0.05$ . (C) In the CGGA database, the expression pattern of ALDH isoforms in Group III. The expression of ALDH isoforms in Group III does not share the same changing trend with the increase of WHO grade. ns,  $P > 0.05$ . (D) In the CGGA database, the expression pattern of ALDH isoforms in Group IV. The expression of all ALDH isoforms in Group IV increases as the WHO grade increases. #,  $P < 0.01$ , #,  $P < 0.05$ . (E) In the CGGA database, the expression pattern of ALDH isoforms in group V. The expression of ALDH isoforms in group V does not share the same changing trend with the increase of WHO grade. ns,  $P > 0.05$ . (F) Univariate Cox analysis of ALDH isoforms in each group in both CGGA and TCGA databases. ALDH, aldehyde dehydrogenase; WHO, World Health Organization; CGGA, Chinese Glioma Genome Atlas.

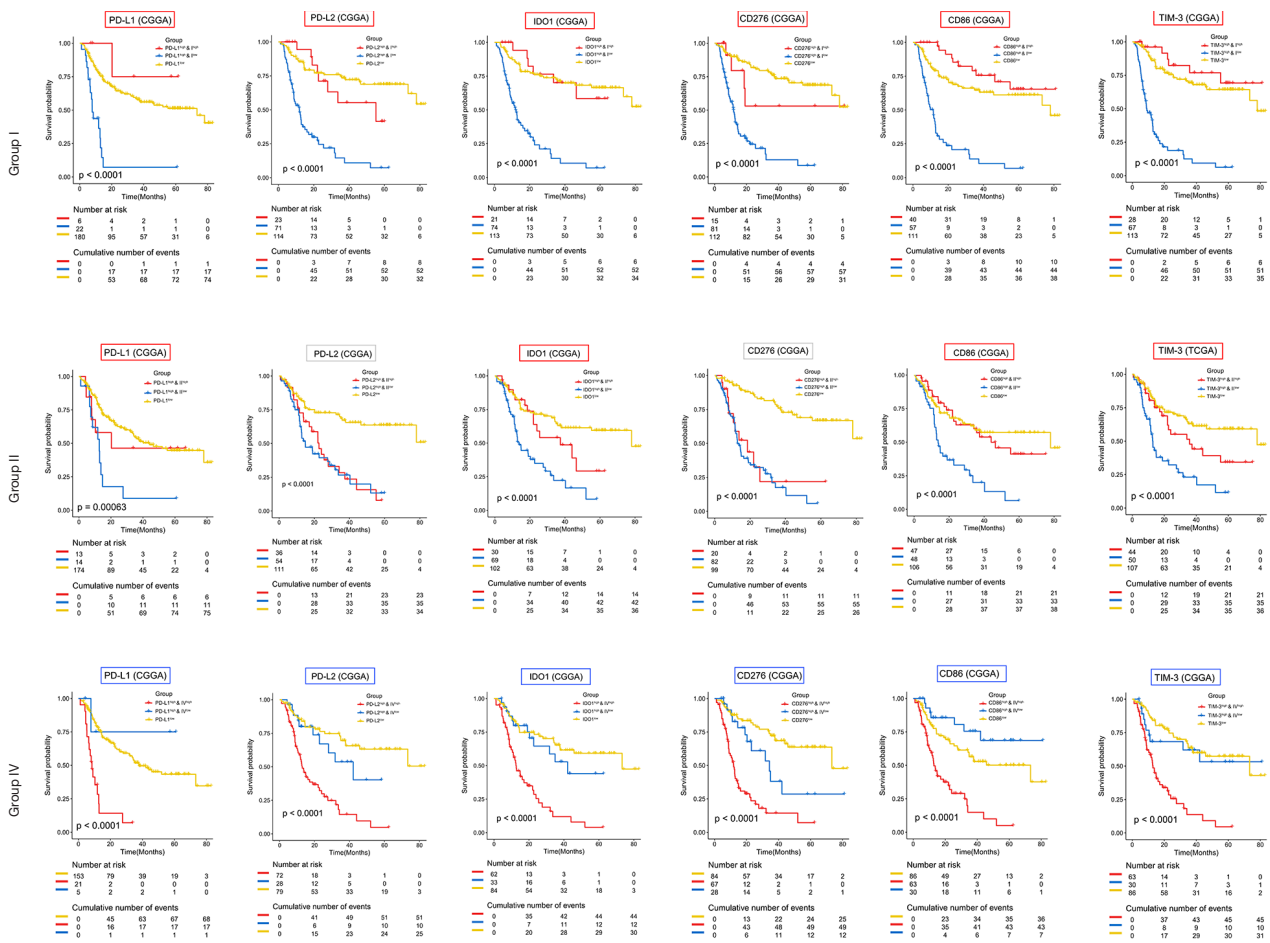


**Figure 4** The ALDH groups are closely related to the state of genomic mutation. (A) Total mutation counts per Mb are significantly fewer in Group I and II isoforms and with higher expression groups than in Group IV isoforms. Expression of all ALDH isoforms in one group in the top 50% of patients is defined as a higher expression group. \*\*\*\*, P<0.0001; \*\*\*, P<0.001; \*, P<0.05; ns, P>0.05. (B) The correlation coefficient between all ALDH isoforms and aneuploidy scores in the TCGA database. Gray group names indicate no statistical difference between them. (C) The correlation coefficient between all ALDH isoforms and aneuploidy amplification (deletion) scores in the TCGA database. Gray group names indicate no statistical difference between them. (D) The correlation coefficient between all ALDH isoforms and DNA repair function scores reveals that PARP enzymes bind to DNA and NER-related are most positively related to ALDH isoforms in Group I and Group II and negative related to the expression of ALDH isoforms in Group IV. Other BER and strand break joining factors are most negatively related to the expression of ALDH isoforms in Group I and Group II and positively related to ALDH isoforms in Group IV. The red background represents a positive correlation. The blue background represents a negative correlation. A grey background represents no significant correlation. ALDH, aldehyde dehydrogenase; PARP, poly ADP-ribose polymerase; TCGA, The Cancer Genome Atlas; BER, base excision repair; NER, nucleotide excision repair.





**Figure 5** The ALDH groups are closely related to the state of tumor immune functions. (A) The 3 most ALDH group-related DNA repair functions show significant correlation, with 95.15% biological functions in the immune system process classification. The scale values in the graph represent the proportions of significant correlation with biological functions in each biological function classification. (B) The correlation coefficient between all ALDH isoforms and immune function scores. The red background represents a positive correlation. The blue background represents a negative correlation. A grey background represents no significant correlation. (C-H) The correlation analysis of all ALDH isoforms and immune checkpoints in both CGGA and TCGA databases. Colors represent different groups. ALDH, aldehyde dehydrogenase.



**Figure 6** Prediction of the therapeutic effect of immunological and enzymatic combined therapy. Kaplan-Meier survival curves reveal the therapeutic effect of immunological and enzymatic combined therapy. A red frame means that elevating the expression of ALDH isoforms in the corresponding group can improve the prognosis of patients. A blue frame means knocking down the expression of ALDH isoforms in the corresponding group can improve the prognosis of patients. A grey frame means changes in ALDH isoforms expression in the corresponding group do not affect patient prognosis. ALDH, aldehyde dehydrogenase.

isoforms, including ALDH16A1, ALDH18A1, ALDH1B1, ALDH1L1/2, ALDH3A2, ALDH4A1, ALDH5A1, ALDH6A1, ALDH7A1, ALDH8A1 and ALDH9A1 were enriched in malignant cells, while ALDH16A1, ALDH2, ALDH3B1 tended to enrich in tumor infiltrating macrophages. ALDH3A2 and ALDH6A1 were enriched in oligodendrocytes. Interestingly, all of the ALDH isoforms expression in tumor infiltrating T cells were lower than those in malignant cells (Figure S5).

**ALDH functional classification can be used to improve the efficacy of tumor immunotherapy**

As a new star in tumor immunotherapy, immune checkpoint

inhibitors have entered clinical practice and shown good clinical benefits in several solid tumors. However, they have not shown an excellent therapeutic effect in the treatment of gliomas. Based on the above results, we wondered if ALDH functional subgroups can improve the immunotherapy effects for glioma treatment. As shown in Figure 6 and Figure S6, in patients with high immune checkpoint gene expression and high expression of ALDH enzymes in Groups I and II, their prognosis was not worse than that of patients with low immune checkpoint gene expression. Furthermore, patients with high immune checkpoint gene expression and low expression of ALDH enzymes in Group IV also had a worse survival rate. This result provided a theoretical basis for the clinical application of

specific inhibitors or activators of the corresponding ALDH subgroups, rather than pan-ALDH activity activators or inhibitors.

## Discussion

The vital biochemical functions of ALDH in cancer has become a focal point of cancer research. Recent studies have also confirmed ALDH as a novel stem cell marker for various solid tumors, such as glioma, lung cancer, breast cancer, and cholangiocarcinoma (9,10,14-16). There are also many targeted treatments for ALDH in cancers (17). However, no efforts have yet to succeed in improving patient outcomes in the clinic. There may be 2 reasons for this lack of progress: (I) in cancer research, almost all research has focused on enzyme activity or only one specific ALDH isoform. As the ALDH enzyme contains 19 isoforms, some different isoforms are likely to have similar biological functions in cancer. Therefore, it is not surprising that treatments targeting a single isoform were unsuccessful. (II) In the study of isoforms of enzymes, biochemists have classified ALDH isoforms based on their protein structure or substrates in normal cells (18). However, in the unique metabolic environment of tumors, the isoforms of ALDH are likely to exert completely different biological functions to those of normal cells. Therefore, the functional classification of ALDH isoforms in tumors is imperative.

Unlike the methods used by biochemists, we grouped ALDH isoforms based on their biological functions and signaling activations in glioma. This integrated clustering result was mutually verifiable in a large cohort of 999 glioma patients from 2 independent databases. Traditional clusters based on mRNA and methylation cannot be repeated in 2 databases. Our analysis also found that functional and pathway clustering analysis was inherently flawed, which was different from that of traditional clustering analysis. Isoforms in the same subgroup had similar functional characteristics, which were different from those of other subgroups. Although almost all previous studies have shown that ALDH enzymes could promote the malignant progression of tumors, our study surprisingly found that isoforms in some subgroups may inhibit glioma progression. Subsequently, the predictive analysis also confirmed that higher expression of isoforms in subgroups for which glioma was inhibited led to a better prognosis. These results further supported the reliability of the results from the functional analysis. Unsurprisingly, we found that isoforms in different subgroups affected the prognosis of

glioma patients by regulating proliferation, transcription, and apoptosis of tumor cells (8,19). Similar results can be seen in previous ALDH related cancer research (16,20,21).

To investigate whether ALDH affected glioma progression by another unexpected function, we explored the relationship between ALDH subgroups and gene mutation counts. Interestingly, we found a close relationship between some subgroups of ALDH and genomic stability. The further in-depth analysis showed that altered DNA repair functions caused by some ALDH subgroups were among the reasons for this phenomenon. However, only a few existing studies have encompassed the relationship between ALDH enzymes and DNA repair functions. One study examined the DNA repair functions of ALDH in alcoholic liver damage, and another analyzed the effects of ALDH1A1 on DNA repair functions in ovarian cancer stem cells (22,23). Our study found a close relationship between ALDH enzymes and DNA repair functions in gliomas and revealed that this mechanism might be an essential cause of the poor prognosis of patients with glioma. This finding provides a new consideration for the use of ALDH functional study in glioma treatment.

We also found that DNA repair functions most relevant to ALDH subgroups were closely related to the immune functions in glioma. Different subgroups had significantly different immune statuses, one of the reasons for which was that ALDH regulated the expression of immune checkpoints. Although few reports have covered the relationship between ALDH and immune functions in other cancers, all studies have regarded ALDH as a stem marker for targeted elimination (24,25). However, our study found that a ALDH enzymes had a robust immune regulation function. We also confirmed that only a small number of glioma patients could benefit from pan-ALDH targeted therapy. However, based on ALDH functional classification, almost all glioma patients could benefit from immunotherapy. This is another innovative discovery of this research.

This study has provided a novel approach for functional analysis of multi-subtype metabolic enzymes in diseases and yielded supporting data for basic and clinical studies of ALDH in glioma. Furthermore, we also revealed several new potential functions and regulatory mechanisms of ALDH in glioma, even providing a possible solution to the dilemma of immunotherapy. However, this functional analysis approach was based on mRNA sequencing data, and more experimental data is needed to valid its accuracy. The clinical application of this research depends on the development of ALDH isoform-specific activators and

inhibitors (26–29). Considering ALDH is an essential metabolic enzyme, it also plays an essential role in the average human brain. We were excited to find that different ALDH isoforms have different levels of importance in the metabolism of the average human brain and glioma cells. Therefore, the application of a future ALDH drug cocktail will have better efficacy and fewer side effects. Such an application will enhance the meaningfulness and clinical application of our research.

In conclusion, this study comprehensively provided a novel approach to reveal different biological functions of distinct subtypes of ALDH enzymes in glioma. Our data showed a strong relationship between distinct subtypes of ALDH enzymes, immune regulation functions, and DNA repair function of glioma, which provided a basis for developing relevant immunotherapy strategies.

### Acknowledgments

We thank Ms. Shuqing Sun and Hua Huang for tissue sample collection and clinical data retrieval. We thank C. Mullens and J. Gray for modifying the manuscript.

*Funding:* This research was funded by grants from the National Natural Science Foundation of China (No. 82072768), Construction of the Genomics Platform for Chinese People's Brain Diseases (No. PXM2019\_026280\_000002), Sino-German Center Cooperation and Exchanges Program (M-0020), The Medical and Health Technology Innovation Project of the Chinese Academy of Medical Sciences (2020-I2M-C&T-A-024) and National Natural Science Foundation of China (No. 82002994).

### Footnote

*Reporting Checklist:* The authors have completed the MDAR checklist. Available at <https://dx.doi.org/10.21037/tcr-21-1160>

*Peer Review File:* Available at <https://dx.doi.org/10.21037/tcr-21-1160>

*Conflicts of Interest:* All authors have completed the ICMJE uniform disclosure form (available at <https://dx.doi.org/10.21037/tcr-21-1160>). The authors have no conflicts of interest to declare.

*Ethical Statement:* The authors are accountable for all

aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the Ethics Committee of Beijing Tiantan Hospital, Capital Medical University (KY2020-093-02). The institutional review board decided that informed consent was not required.

*Open Access Statement:* This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

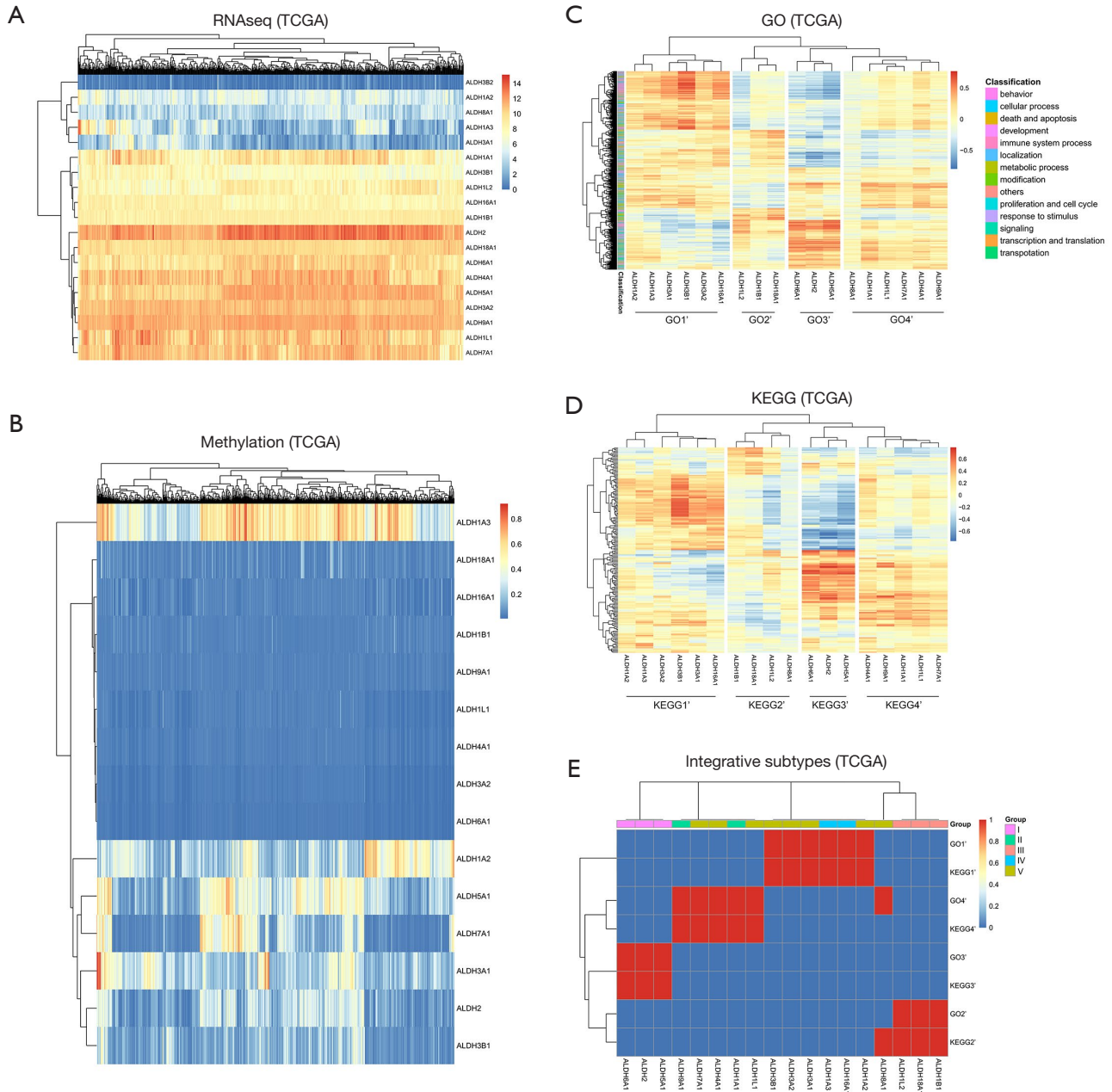
### References

1. Stupp R, Mason WP, van den Bent MJ, et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N Engl J Med* 2005;352:987-96.
2. Bouffet E, Larouche V, Campbell BB, et al. Immune Checkpoint Inhibition for Hypermutant Glioblastoma Multiforme Resulting From Germline Biallelic Mismatch Repair Deficiency. *J Clin Oncol* 2016;34:2206-11.
3. Hu H, Mu Q, Bao Z, et al. Mutational Landscape of Secondary Glioblastoma Guides MET-Targeted Trial in Brain Tumor. *Cell* 2018;175:1665-1678.e18.
4. Yan H, Parsons DW, Jin G, et al. IDH1 and IDH2 mutations in gliomas. *N Engl J Med* 2009;360:765-73.
5. Pusch S, Krausert S, Fischer V, et al. Pan-mutant IDH1 inhibitor BAY 1436032 for effective treatment of IDH1 mutant astrocytoma in vivo. *Acta Neuropathol* 2017;133:629-44.
6. Johanns TM, Miller CA, Dorward IG, et al. Immunogenomics of Hypermutated Glioblastoma: A Patient with Germline POLE Deficiency Treated with Checkpoint Blockade Immunotherapy. *Cancer Discov* 2016;6:1230-6.
7. Ginestier C, Hur MH, Charafe-Jauffret E, 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;1:555-67.
8. Mao P, Joshi K, Li J, et al. Mesenchymal glioma stem cells are maintained by activated glycolytic metabolism

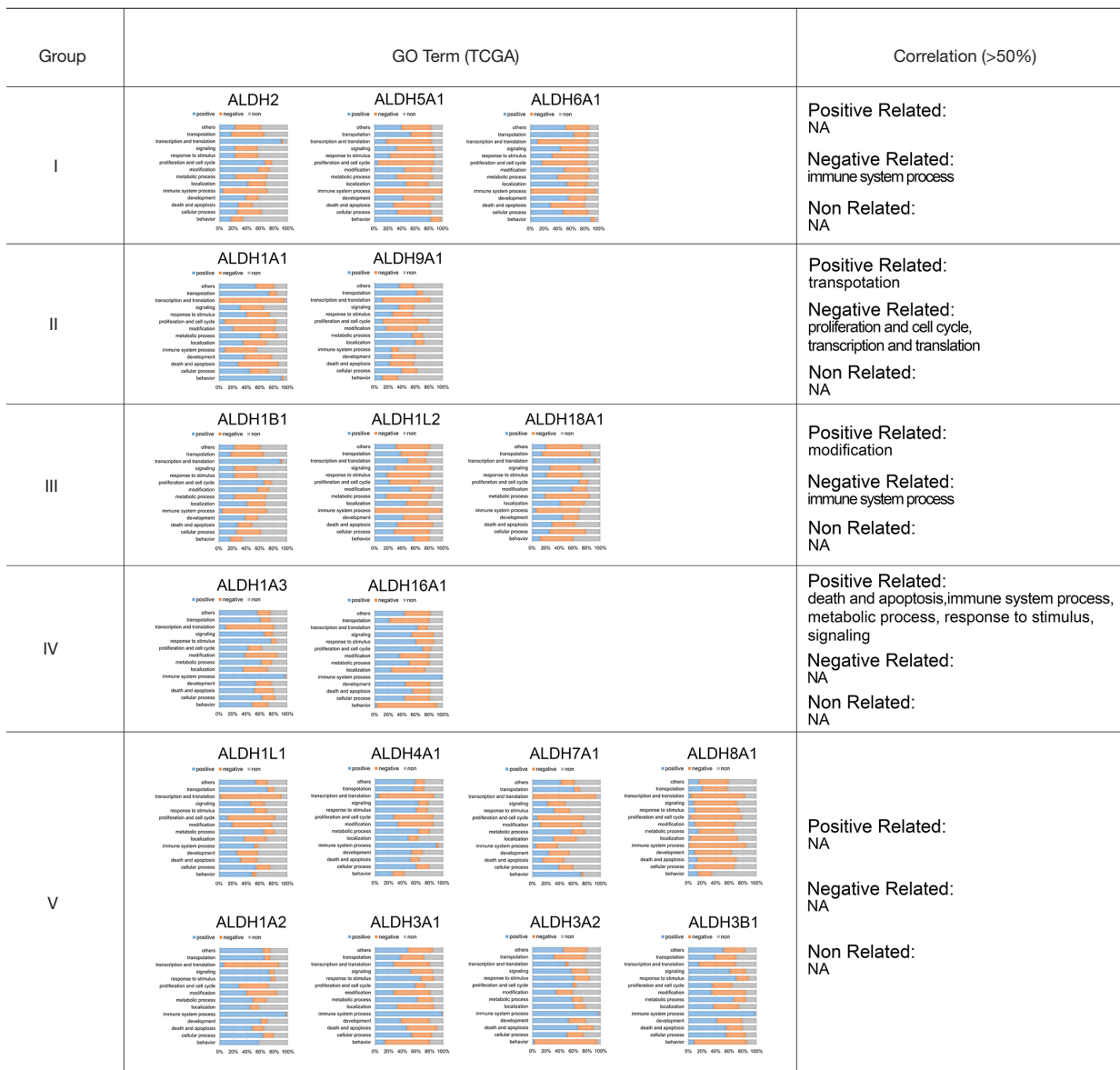


- involving aldehyde dehydrogenase 1A3. *Proc Natl Acad Sci U S A* 2013;110:8644-9.
9. Chen MH, Weng JJ, Cheng CT, et al. ALDH1A3, the Major Aldehyde Dehydrogenase Isoform in Human Cholangiocarcinoma Cells, Affects Prognosis and Gemcitabine Resistance in Cholangiocarcinoma Patients. *Clin Cancer Res* 2016;22:4225-35.
  10. Shao C, Sullivan JP, Girard L, et al. Essential role of aldehyde dehydrogenase 1A3 for the maintenance of non-small cell lung cancer stem cells is associated with the STAT3 pathway. *Clin Cancer Res* 2014;20:4154-66.
  11. Butowski N, Prados MD, Lamborn KR, et al. A phase II study of concurrent temozolomide and cis-retinoic acid with radiation for adult patients with newly diagnosed supratentorial glioblastoma. *Int J Radiat Oncol Biol Phys* 2005;61:1454-9.
  12. Hoadley KA, Yau C, Wolf DM, et al. Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell* 2014;158:929-44.
  13. Taylor AM, Shih J, Ha G, et al. Genomic and Functional Approaches to Understanding Cancer Aneuploidy. *Cancer Cell* 2018;33:676-689.e3.
  14. Rasper M, Schäfer A, Piontek G, et al. Aldehyde dehydrogenase 1 positive glioblastoma cells show brain tumor stem cell capacity. *Neuro Oncol* 2010;12:1024-33.
  15. Xu X, Chai S, Wang P, et al. Aldehyde dehydrogenases and cancer stem cells. *Cancer Lett* 2015;369:50-7.
  16. Charafe-Jauffret E, Ginestier C, Iovino F, et al. Aldehyde dehydrogenase 1-positive cancer stem cells mediate metastasis and poor clinical outcome in inflammatory breast cancer. *Clin Cancer Res* 2010;16:45-55.
  17. Cheng P, Wang J, Waghmare I, et al. FOXD1-ALDH1A3 Signaling Is a Determinant for the Self-Renewal and Tumorigenicity of Mesenchymal Glioma Stem Cells. *Cancer Res* 2016;76:7219-30.
  18. Marchitti SA, Deitrich RA, Vasiliou V. Neurotoxicity and metabolism of the catecholamine-derived 3,4-dihydroxyphenylacetaldehyde and 3,4-dihydroxyphenylglycolaldehyde: the role of aldehyde dehydrogenase. *Pharmacol Rev* 2007;59:125-50.
  19. Wu W, Schecker J, Würstle S, et al. Aldehyde dehydrogenase 1A3 (ALDH1A3) is regulated by autophagy in human glioblastoma cells. *Cancer Lett* 2018;417:112-23.
  20. Yang L, Ren Y, Yu X, et al. ALDH1A1 defines invasive cancer stem-like cells and predicts poor prognosis in patients with esophageal squamous cell carcinoma. *Mod Pathol* 2014;27:775-83.
  21. Pérez-Alea M, McGrail K, Sánchez-Redondo S, et al. ALDH1A3 is epigenetically regulated during melanocyte transformation and is a target for melanoma treatment. *Oncogene* 2017;36:5695-708.
  22. Smith C, Gasparetto M, Jordan C, et al. The effects of alcohol and aldehyde dehydrogenases on disorders of hematopoiesis. *Adv Exp Med Biol* 2015;815:349-59.
  23. Meng E, Mitra A, Tripathi K, 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:e107142.
  24. Bazewicz CG, Dinavahi SS, Schell TD, et al. Aldehyde dehydrogenase in regulatory T-cell development, immunity and cancer. *Immunology* 2019;156:47-55.
  25. Maccalli C, Parmiani G, Ferrone S. Immunomodulating and Immuno-resistance Properties of Cancer-Initiating Cells: Implications for the Clinical Success of Immunotherapy. *Immunol Invest* 2017;46:221-38.
  26. Morgan CA, Parajuli B, Buchman CD, et al. N,N-diethylaminobenzaldehyde (DEAB) as a substrate and mechanism-based inhibitor for human ALDH isoenzymes. *Chem Biol Interact* 2015;234:18-28.
  27. Yang SM, Yasgar A, Miller B, et al. Discovery of NCT-501, a Potent and Selective Theophylline-Based Inhibitor of Aldehyde Dehydrogenase 1A1 (ALDH1A1). *J Med Chem* 2015;58:5967-78.
  28. Canuto RA, Muzio G, Salvo RA, et al. The effect of a novel irreversible inhibitor of aldehyde dehydrogenases 1 and 3 on tumour cell growth and death. *Chem Biol Interact* 2001;130-132:209-18.
  29. Thomas ML, de Antueno R, Coyle KM, et al. Citral reduces breast tumor growth by inhibiting the cancer stem cell marker ALDH1A3. *Mol Oncol* 2016;10:1485-96.

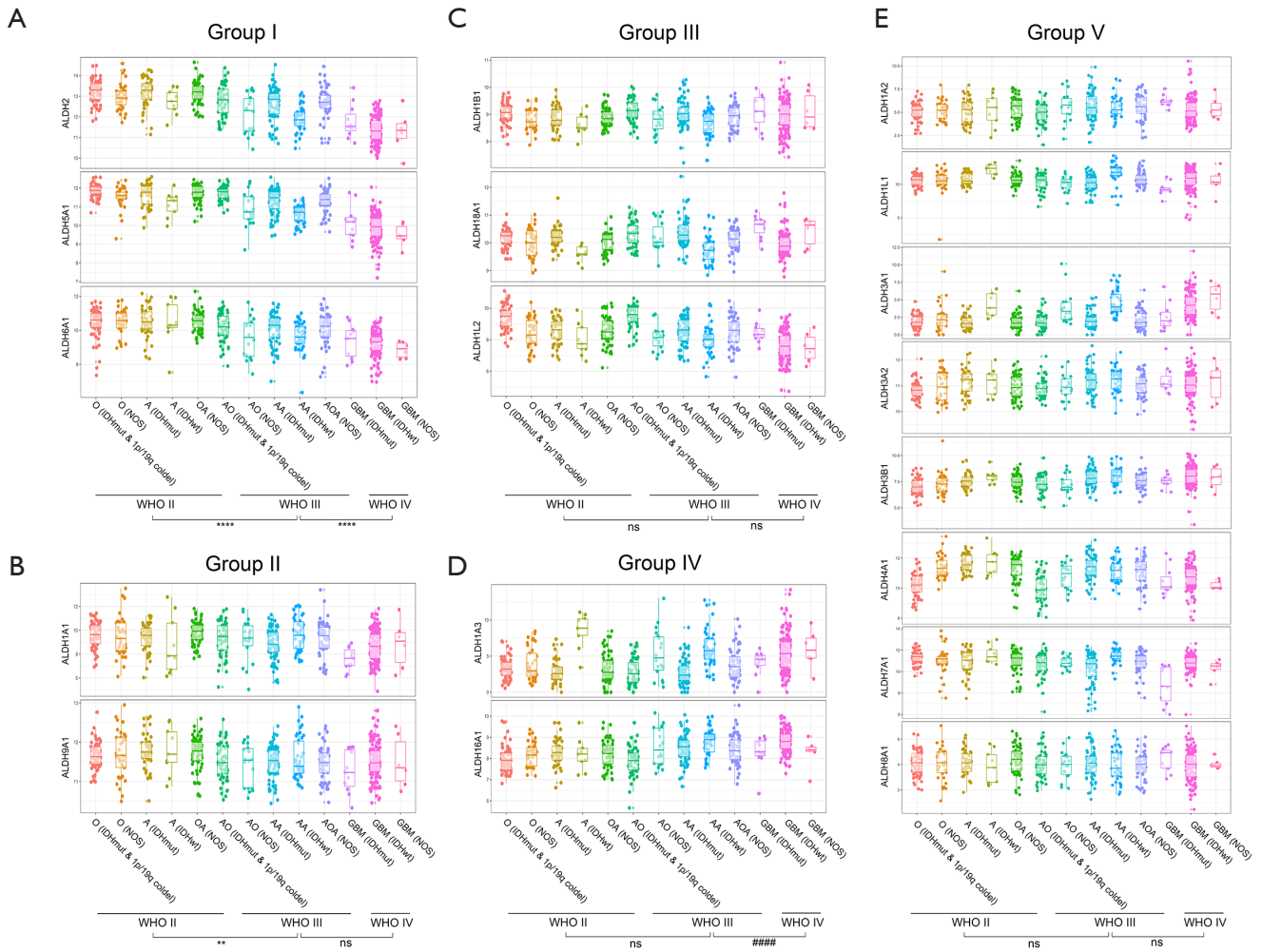
**Cite this article as:** Wang D, Jiang H, Wang Z, Huang R, Jiang T, Kahlert UD, Li G, Zhang W. Functional clustering analysis identifies specific subtypes of aldehyde dehydrogenase associated with glioma immunity. *Transl Cancer Res* 2021;10(12):5052-5064. doi: 10.21037/tcr-21-1160



**Figure S1** Multi-platform cluster analysis of ALDH in TCGA database. (A,B) Unsupervised clustering of all ALDH isoforms by gene mRNA expression and gene methylation level in TCGA database. (C,D) Unsupervised biological clustering of all ALDH isoforms in TCGA database. ALDH isoforms can be classified into 4 groups according to their biological functions and 4 groups according to their signal pathway activations. (E) ALDH isoforms can be classified into 5 groups by integrated Cluster-Of-Cluster assignments analysis according to their biological functions and signal pathway activations.



**Figure S2** Landscape of biological functions of all ALDH isoforms in 5 groups in CGGA database. Most related biological functions of ALDH isoforms in each group. Only all ALDH isoforms in a group are significantly (not) related to more than 50% biological functions of one classification is defined as most (non) related.



**Figure S3** The clinical features of all ALDH isoforms in 5 groups. (A) In TCGA database, the expression of all ALDH isoforms in group I decreases as the WHO grade increases. \*\*\*\*:  $P < 0.0001$ . (B) In TCGA database, the expression of all ALDH isoforms in group II decreases as the WHO grade increases. \*\*:  $P < 0.01$ , ns:  $P > 0.05$ . (C) In TCGA database, the expression of ALDH isoforms in group III does not share the same changing trend with the increase of WHO grade. ns:  $P > 0.05$ . (D) In TCGA database, the expression of all ALDH isoforms in group IV increases as the WHO grade increases. #####:  $P < 0.0001$ , ns:  $P > 0.05$ . (E) In TCGA database, the expression of ALDH isoforms in group V does not share the same changing trend with the increase of WHO grade. ns:  $P > 0.05$ .

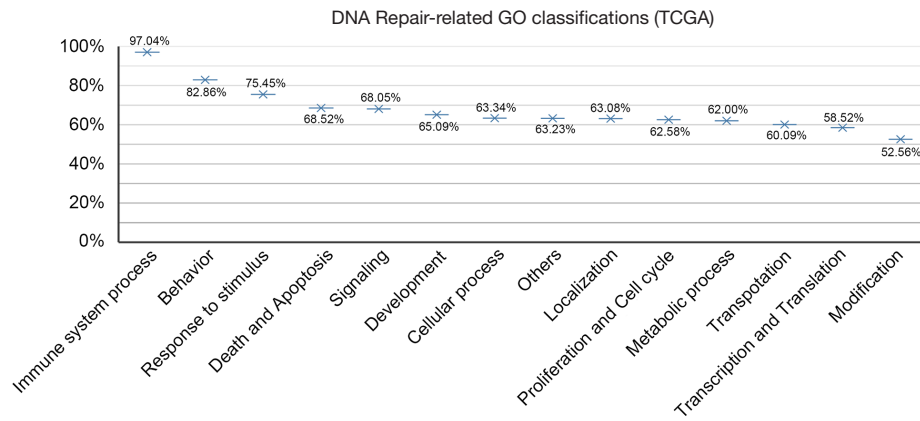


A

**DNA repair functions (TCGA)**

Group	I			II		III			IV		V							
Subtype	ALDH2	ALDH5A1	ALDH6A1	ALDH1A1	ALDH8A1	ALDH1B1	ALDH1L2	ALDH18A1	ALDH1A3	ALDH16A1	ALDH1A2	ALDH1L1	ALDH3A1	ALDH3A2	ALDH3B1	ALDH4A1	ALDH7A1	ALDH8A1
Poly(ADP-ribose) polymerase (PARP) enzymes that bind to DNA	0.632	0.684	0.515	0.301	0.281	0.053	0.346	-0.013	-0.292	-0.386	-0.090	-0.083	-0.334	-0.140	-0.358	0.030	0.167	0.073
NER-related	0.622	0.709	0.573	0.111	0.203	0.100	0.453	0.224	-0.395	-0.445	-0.021	-0.063	-0.575	-0.071	-0.293	0.021	0.006	0.006
TFIIH/Catalyzes unwinding in precision complex	0.278	0.258	0.286	0.094	0.199	-0.114	0.003	-0.161	-0.137	-0.141	0.060	0.071	-0.261	-0.037	-0.113	0.141	0.001	0.050
Nucleotide excision repair (NER) (XP = xeroderma pigmentosum)	0.217	0.084	0.114	0.201	0.401	-0.170	0.316	-0.281	-0.149	-0.060	0.077	0.408	0.336	0.397	0.257	0.419	0.237	0.170
Ubiquitination and modification	0.216	0.278	0.453	0.181	0.197	-0.032	0.158	-0.048	0.099	0.076	0.184	0.068	-0.116	-0.048	-0.223	0.051	0.060	-0.209
Base excision repair (BER)/DNA glycosylases: major altered base released	0.186	-0.070	-0.237	0.163	0.199	-0.116	-0.310	-0.302	-0.001	0.249	-0.089	0.289	0.174	-0.010	0.168	0.163	0.169	0.075
Other identified genes with known or suspected DNA repair function	0.193	0.098	0.089	0.059	0.182	-0.022	-0.042	0.035	-0.001	-0.231	0.050	0.046	0.030	0.168	0.062	0.172	-0.047	-0.227
Other BER and strand break joining factors/Other BER and strand break joining factors	-0.244	-0.207	-0.273	-0.195	-0.095	0.040	-0.280	0.094	0.021	0.512	0.072	-0.296	0.127	0.128	0.291	0.190	-0.111	-0.095
Fanconi anemia/Tolerance and repair of DNA crosslinks and other adducts in DNA	-0.590	-0.345	-0.374	-0.323	-0.413	0.273	0.066	0.235	0.099	0.148	-0.137	-0.329	0.118	-0.148	-0.172	-0.393	-0.185	-0.064
Homologous recombination	-0.492	-0.490	-0.259	-0.142	-0.186	-0.161	-0.244	-0.156	0.199	0.409	0.008	0.009	0.384	0.080	0.218	0.026	0.028	0.138
Repair of DNA-topoisomerase crosslinks	-0.215	-0.258	-0.111	0.012	-0.044	-0.076	-0.314	-0.180	0.347	-0.060	0.188	0.055	0.258	-0.008	0.116	-0.039	-0.019	-0.062
Chromatin Structure and Modification	-0.220	-0.372	-0.427	-0.063	-0.178	-0.088	-0.379	-0.113	0.074	0.000	-0.073	0.025	0.282	0.014	0.132	0.104	-0.067	0.138
Direct reversal of damage	-0.242	-0.062	-0.229	-0.108	0.129	-0.062	-0.199	-0.225	0.042	0.145	0.025	0.142	0.199	0.013	0.203	0.142	0.108	0.068
Mismatch excision repair (MMR)	-0.098	0.091	0.328	-0.008	-0.193	-0.121	0.121	-0.009	-0.070	-0.115	-0.104	-0.058	-0.053	-0.163	-0.249	-0.220	0.037	0.198
Non-homologous end-joining	0.017	0.100	0.144	-0.100	0.099	0.285	0.281	0.335	-0.056	-0.288	0.093	-0.231	-0.230	0.101	0.004	-0.073	-0.170	-0.257
Modulation of nucleotide pools	0.126	-0.010	-0.194	0.172	0.228	-0.007	-0.064	-0.220	0.081	0.050	-0.118	0.228	0.182	0.033	0.163	-0.154	0.230	-0.098
DNA polymerases (catalytic subunits)	-0.046	0.012	-0.008	-0.211	-0.278	-0.031	0.128	0.148	-0.189	0.223	0.007	-0.157	-0.266	-0.041	0.077	-0.040	-0.135	0.382
Genes defective in diseases associated with sensitivity to DNA damaging agents	0.052	0.228	0.226	-0.038	-0.069	0.069	0.449	0.183	-0.056	-0.153	-0.042	-0.223	-0.229	-0.213	-0.278	0.048	0.052	
Editing and processing nucleases	0.023	-0.196	-0.402	0.155	-0.053	0.111	-0.169	-0.115	0.198	0.026	-0.049	0.052	0.246	-0.093	0.106	-0.007	0.021	0.140
Other conserved DNA damage response genes	-0.206	0.050	0.118	-0.165	-0.188	0.058	0.299	0.241	-0.110	-0.113	0.019	-0.172	-0.179	-0.022	-0.146	-0.195	-0.118	-0.014

B

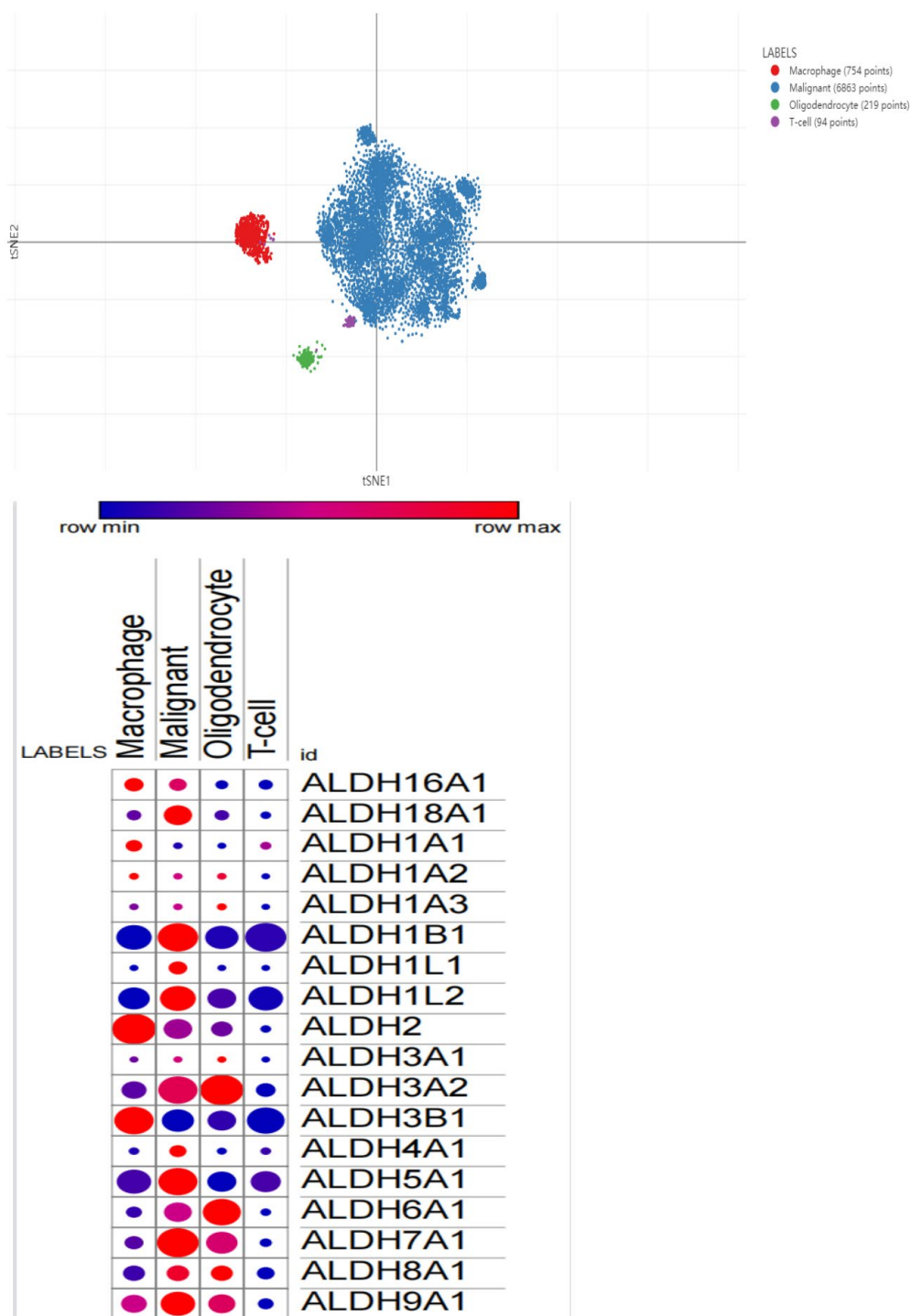


C

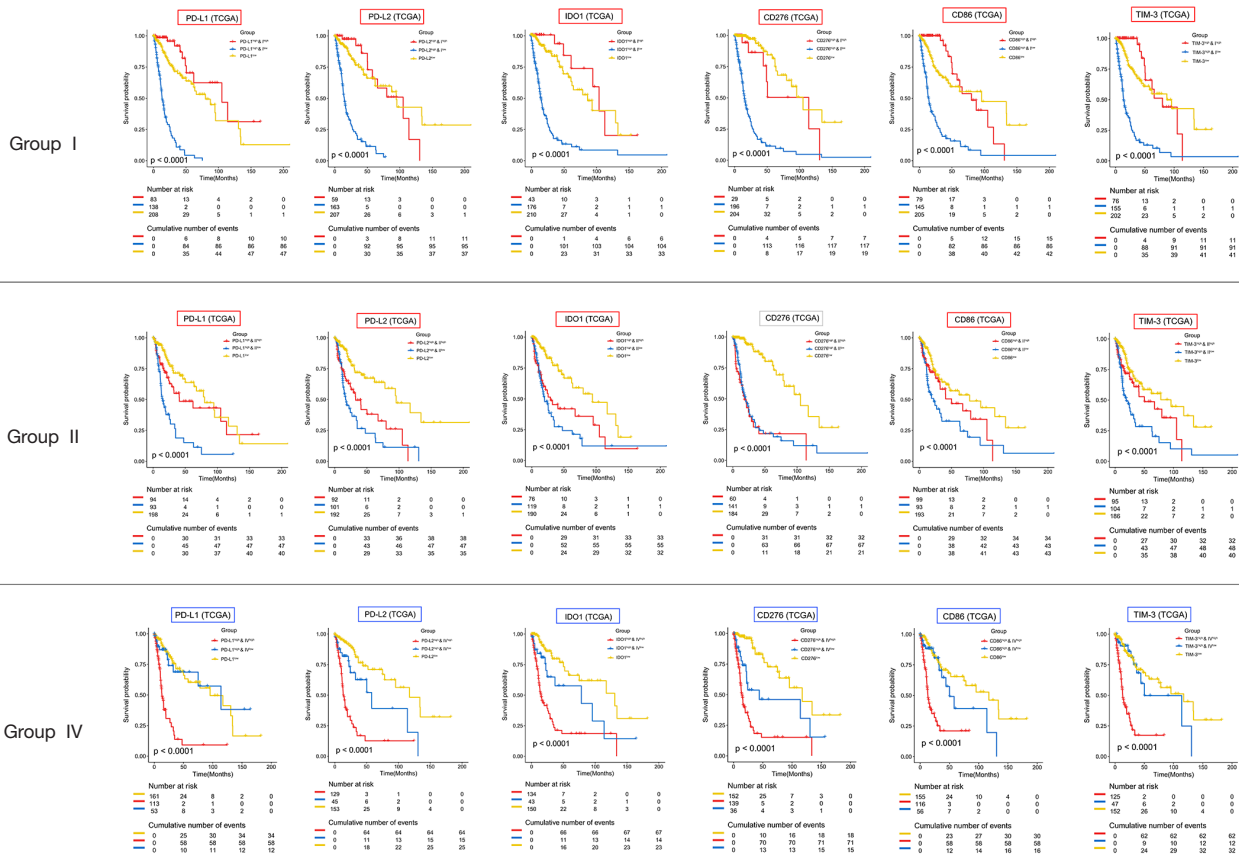
**Immune functions (TCGA)**

Group	I			II		III			IV		V							
Subtype	ALDH2	ALDH5A1	ALDH6A1	ALDH1A1	ALDH8A1	ALDH1B1	ALDH1L2	ALDH18A1	ALDH1A3	ALDH16A1	ALDH1A2	ALDH1L1	ALDH3A1	ALDH3A2	ALDH3B1	ALDH4A1	ALDH7A1	ALDH8A1
Immune response	-0.417	-0.591	-0.383	-0.078	0.029	-0.095	-0.421	-0.137	0.353	0.473	0.288	0.116	0.487	0.354	0.694	0.217	-0.029	-0.205
B cell activation involved in immune response	-0.642	-0.878	-0.434	-0.229	-0.105	-0.008	-0.341	0.019	0.312	0.491	0.175	-0.042	0.469	0.317	0.624	0.043	-0.116	-0.227
T cell activation involved in immune response	-0.329	-0.519	-0.302	-0.022	0.027	-0.134	-0.383	-0.120	0.300	0.468	0.327	0.073	0.411	0.528	0.680	0.267	-0.085	-0.090
cytokine production involved in immune response	-0.404	-0.583	-0.320	-0.028	0.061	-0.169	-0.436	-0.198	0.371	0.487	0.317	0.112	0.502	0.346	0.652	0.272	0.001	-0.122
cytokine secretion involved in immune response	-0.383	-0.524	-0.284	-0.043	0.052	-0.128	-0.387	-0.150	0.309	0.439	0.296	0.084	0.453	0.525	0.688	0.268	-0.039	-0.107
Immune response to tumor cell	-0.385	-0.548	-0.425	-0.045	0.039	-0.033	-0.427	-0.109	0.265	0.538	0.222	0.033	0.417	0.298	0.528	0.116	-0.034	-0.054
leukocyte activation involved in immune response	-0.396	-0.584	-0.355	-0.047	0.055	-0.098	-0.444	-0.159	0.359	0.439	0.274	0.142	0.497	0.344	0.697	0.246	-0.008	-0.204
natural killer cell mediated immune response to tumor cell	-0.324	-0.505	-0.349	0.020	-0.010	-0.112	-0.443	-0.191	0.276	0.560	0.248	0.083	0.367	0.234	0.609	0.289	-0.035	0.024
T cell mediated immune response to tumor cell	0.125	0.325	-0.006	-0.133	0.108	0.203	0.529	0.305	-0.417	-0.061	-0.207	-0.200	-0.264	0.171	-0.281	-0.110	-0.116	0.004
natural killer cell activation involved in immune response	-0.104	-0.312	-0.078	0.178	0.074	-0.269	-0.311	-0.276	0.340	0.286	0.294	0.192	0.330	0.193	0.563	0.317	0.084	0.045
myeloid cell activation involved in immune response	-0.357	-0.567	-0.346	-0.034	0.074	-0.101	-0.457	-0.178	0.353	0.422	0.259	0.168	0.494	0.339	0.683	0.258	0.011	-0.206

**Figure S4** The ALDH groups are closely related to the state of genomic mutation. (A) The correlation coefficient between all ALDH isoforms and DNA repair function scores reveals that PARP enzymes that bind to DNA, NER-related and H/Catalyzes unwinding in precision complex are most positive related to the expression of ALDH isoforms in group I and group II and negative related to the expression of ALDH isoforms in group IV. In addition, Fanconi anemia/Tolerance and repair of DNA crosslinks and other adducts in DNA and Homologous recombination are most negative related to the expression of ALDH isoforms in group I and group II and positive related to the expression of ALDH isoforms in group IV. Red background represents a positive correlation. Blue background represents a negative correlation. Grey background represents no significant correlation. (B) Biological functions, most related DNA repair functions, is immune system process classification. The scale values in the graph represent the proportions of significant correlation biological functions in each biological function classifications. (C) The correlation coefficient matrix of ALDH isoforms and immune function scores. Red background represents a positive correlation. Blue background represents a negative correlation. Grey background represents no significant correlation.



**Figure S5** Analysis on different sub-classify ALDH enzyme in distinct cell types based on scRNA-seq dataset from 28 patients with GBM. Cell subtypes clustering via t SNE method and distinct isoforms of ALDHs expression in different cell clusters based on single cell mRNA sequencing data. The color shows mRNA expression level compared to the average level in all cells. The size of plots represent the ratio of specific subtype of ALDH positive cells in distinct cell cluster.



**Figure S6** Prediction of therapeutic effect of immunological and enzymatic combined therapy. Kaplan-Meier survival curves reveal the therapeutic effect of immunological and enzymatic combined therapy. Red frame means elevating the expression of ALDH isoforms in the corresponding group can improve the prognosis of patients. Blue frame means knocking down the expression of ALDH isoforms in the corresponding group can improve the prognosis of patients. Grey frame means changes in ALDH isoforms expression in the corresponding group do not affect patient prognosis.