Comprehensively analysis of the FGD family with potential effects on prognosis and immune infiltration in lung adenocarcinoma

Liangyu Zhang
Department of thoracic surgery, First Affiliated Hospital of Fujian Medical University
Fengqiang Yu
Department of thoracic surgery, First Affiliated Hospital of Fujian Medical University
Maohao Guan
Department of thoracic surgery, First Affiliated Hospital of Fujian Medical University
Xuan Huang
Department of thoracic surgery, First Affiliated Hospital of Fujian Medical University
Nanlong Lin
Department of thoracic surgery, First Affiliated Hospital of Fujian Medical University
Fei He
Department of Epidemiology and Health Statistics, School of Public Health, Fujian Medical University
Fancai Lai (laifancai@fjmu.edu.cn)
Department of thoracic surgery, First Affiliated Hospital of Fujian Medical University

Research Article

Keywords:

Posted Date: February 2nd, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2529550/v1

Additional Declarations: No competing interests reported.

Comprehensively analysis of the FGD family with potential effects on prognosis and immune infiltration in lung adenocarcinoma

Liangyu Zhang[1], Fengqiang Yu[1], Maohao Guan[1], Xuan Huang[1], Nanlong Lin[1], Fei He[2*], Fancai Lai[1*]

[1]Department of Thoracic Surgery, the First Affiliated Hospital, Fujian Medical University, Fuzhou, 350005, China

[2]Department of Epidemiology and Health Statistics, School of Public Health, Fujian Medical University, Fuzhou, China

*** Correspondence:**

Corresponding author:**Fancai Lai[1*]**; **Fei He[2*]**

E-mail-address:Laifancai@fjmu.edu.cn; i.fei.he@fjmu.edu.cn

Keywords: FGD family, lung adenocarcinoma, immune checkpoint blockade(ICB) therapy, immune infiltration, prognosis

ABSTRACT

Background： The FGD family consists of six genes, namely FGD1/2/3/4/5/6. Their roles in lung adenocarcinoma have been unidentified. This research focused on determining the diagnostic efficacy, prognostic value, and immune-related functions of them in lung adenocarcinoma(LUAD).

Methods： From the TCGA database, mRNA data for the FGD gene family and clinical data for the patients were obtained. Immunohistochemistry was performed to validate representative FGD gene's expression. A relationship between the FGD genes and immune system molecules was examined using the TIMER and GEPIA databases, the ssGSEA and the MCPcounter methods. Clinical prognosis in LUAD were analyzed by searching for TCGA, KMplotter and GEPIA databases. The TIDE algorithm, TCIA, KMplotter, ROCplotter and ICBatlas databases were used to analyze the value of FGD2 in predicting the efficacy of immunotherapy. TIGER database was used to analyze

single-cell RNA-sequencing data. The immune-related prognostic model was constructed using 3 machine learning algorithms: K-means clustering, LASSO regression, and WGCNA analysis.

Results: All the six FGD genes' protein and mRNA were aberrant expressed in the tissues of LUAD in contrast to healthy ones, and our external experiment confirmed FGD2's expression pattern. Low expression of FGD2, 3, 5 resulted in a shorter OS time and were determined as independent prognostic factors via multivariate analyses. FGD2, 3, 5 were markedly linked to immune infiltration while FGD1, 4, 6 were not. Sc-RNA sequencing analysis indicating that FGD2, 3, 5 were mainly expressed in immunocytes. NSCLC patients with higher FGD2 may more responsive to ICB therapy. The functions of FGD2, 3, 5 and FGD1, 4, 6 in LUAD are heterogeneous, and patients can be separated in to two groups based on these six FGDs' expression. A prognostic model constructed by immune-related DEGs between these two groups had good predictive value in one training set and 4 testing sets.

Conclusions: FGD2, FGD3 and FGD5 can be used as diagnostic, prognostic, and immune-implicated biomarkesr for patients with LUAD and FGD2 may help to predict the ICB therapy efficacy. The immune-related prognostic model had satisfactory predictive value.

INTRODUCTION

Worldwide, lung cancer is one of the most deadly types of cancer and is a universal malignancy(1). It is estimated that 85% of all lung cancers are NSCLC (non-small cell lung cancer), of which LUAD is the most universal subtype(2). The technology for treating lung cancer has improved, but long-term survival remains low, with less than 20% of patients surviving five years after diagnosis(3). Recent advances in targeted therapy and immunotherapy have revolutionized NSCLC treatment, and some are displacing traditional chemoradiotherapy(4). However, anti-PD1/PD-L1 therapy is only effective in 20% of patients with NSCLC(5), so the development of biomarkers that can predict the

51 prognosis and effectiveness of immunotherapy for NSCLC patients is therefore crucial.

52 The FGD (faciogenital dysplasia) gene family consists of 6 genes, including FGD1-FGD6. FGD1
53 was known as the cause of Aarskog-Scott syndrome, an X-chromosome-correlated condition
54 affecting multiple body parts(6), and the GTPase CDC42, which regulates cytoskeletal function and
55 organizes actin filaments, is activated by FGD1, which acts as a guanine nucleotide exchange
56 factor(7, 8). A number of cancers progress faster when FGD3 is abnormally expressed, including
57 pancreatic and breast cancer(9, 10), and high level of FGD3 can improve head and neck squamous
58 cell carcinoma patients' prognosis(11) The Charcot-Marie-Tooth disease and several disorders of the
59 peripheral nervous system are associated with the mutations of FGD4, and prostate cancer's
60 aggressive phenotype is positively correlated with FGD4 expression(12, 13). Additionally, FGD3,
61 FGD4 also been found to modulate actin cytoskeletons and exchange CdC42-Specific factors(14, 15).
62 According to FGD2, studies have shown that it affects the prognosis of melanoma patients and high
63 expression of it in HNSCC may improve patients' prognosis(11, 16). Besides, FGD2 has the ability to
64 positively regulate immune cell infiltration in LUAD(17). FGD6 is transcriptionally regulated by the
65 transcription factor EHF, which affects neutrophil recruitment in hepatocellular carcinoma(18). In
66 pancreatic ductal adenocarcinomas, high FGD6 level confers malignant phenotype(19). Studies on
67 lncRNA FGD5 antisense RNA 1 (FGD5-AS1) are various. However, it remains largely unknown
68 what expression patterns of FGD5 are associated with cancers and what the prognostic value of this
69 gene is. These six FGD genes in LUAD remain unknown as far as their expression pattern,
70 prognostic value, related function, and underlying mechanisms are concerned.

71 This study comprehensively analyzed the expression pattern, diagnostic and prognostic value of
72 these 6 FGD genes in LUAD and identified FGD2, FGD3 and FGD5 as independent prognostic
73 factors for LUAD. Data were retrieved from the TCGA database for comprehensive analyses. FGD2,
74 FGD3 and FGD5, their relationships with immune infiltration, and a comparison of their expression
75 among cancer patients who responded to ICB treatment and those who didn't was also determined.

We found FGD2, 3, 5 all positively regulate immune infiltration in LUAD and base on predicting immune checkpoint blockade(ICB) therapy efficacy, FGD2 exhibited the most value. Based on FGD1-6's expression, LUAD patients were separated into 2 groups and patients' prognosis, immune infiltration levels, immune checkpoint levels and mutation rates are significantly different between these 2 groups. Finally, an immune-related prognostic model constructed by DEGs between the 2 groups exhibited good predictive value.

METHODS

Source data

From the TCGA database, RNA sequencing data and clinical information for LUAD patients were downloaded in level 3 HTseq-FPKM format and log2 transformations were conducted. Duplicate data and samples without complete clinical data were deleted. In the UALCAN database, protein-level gene expression was obtained, and the HPA database helped us to analyze some immunohistochemistry data(20). Gene mutation data came from the Sangerbox tool and the GSCA database(21, 22). The miRNA-gene network was also acquired from GSCA. Drug sensitivity data was obtained from GDSC and predicted using the R package pRRophetic. 2 GEO sets, namely GSE19188 and GSE32863, were analyzed using GEO2R.

Analysis of diagnostic efficacy and prognostic value

ROC curves were visualized using the "pROC" R package. The "survival" package is used for statistical analysis of survival data and was used to conduct cox analysis (multivariate analysis included factors with p values less than 0.2 in univariate analysis) and the "survminer" package is used for visualization. To construct the nomogram prediction model, R package "rms" was used. Besides, GEPIA and KMplotter databases was also used for survival analysis(23, 24).

Immune infiltration analysis

101 In the “estimate” R package, the method ESTIMATE(25) was utilized to evaluate the stromal,
102 ESTIMATE, and immune scores for individual samples of LUAD. A gene-immunity cell infiltration
103 relationship was analyzed using ssGSEA and MCPcounter methods(26, 27). TIMER(28) and GEPIA
104 databases were used to analyze the correlation between genes and immune cell markers. The TIP
105 database was queried for gene sets' enrichment scores corresponding to different anti-cancer immune
106 cycles.
107

108 **Analysis of ICB therapy response**

109 LUAD patients' tumor immune dysfunction and exclusion (TIDE) scores were downloaded from
110 the TIDE website to predict ICB response(29). The TIDE score is higher when ICB therapy is less
111 effective. Immunophenoscore (IPS) was calculated without bias using effector cell,
112 immunosuppressive cell, MHC molecule, and immunomodulator expression data. A positive
113 relationship was found between ICB therapy response and IPS. IPSs for LUAD patients were
114 downloaded from TCIA database(30). The ICBatlas database was utilized to assess the expression of
115 the FGD genes in NSCLC patients treated with ICB(31). KMplotter and ROCplotter(32) were used
116 to analyze the expression of FGD genes in patients with other types of cancer treated with ICB.
117

118 **Single-cell RNA-sequencing analysis**

119 The TIGER online databases(33) was used to analyze sc-RNA sequencing sets of LUAD patients
120 from GEO database. Two single-cell RNA-sequencing sets for LUAD, namely GSE131907 and
121 GSE123904, were selected for specific analysis.

122 **Co-expressed genes and enrichment analysis**

123 GEPIA databases was used to identify genes co-expressed with 3 FGD genes. PPI network was
124 constructed using GENEMANIA(34) and GO_KEGG enrichment analysis was performed with the
125 help of the R package “clusterProfiler”(35). VennDiagram was drawn using the R package

126 “VennDiagram”.

127 **Prognostic model**

128 R package "ConsensusClusterplus" was used for unsupervised clustering. All the LUAD patients
129 were classified into 2 groups based on six FGDs’ level. Differential analysis was carried out by using
130 the R package “DESeq2” to identified differently expressed genes(DEGs). WGCNA and the Pearson
131 correlation analysis were carried out using the R package “WGCNA” on DEGs to explore the module
132 closely correlated with immune-related DEGs. The minimum module size was set at 30. Immune-
133 related DEGs which have prognostic value were subjected to LASSO cox regression analysis, which
134 was conducted by R package “glmnet” to acquire key genes and their coefficients. Risk score was
135 calculated based on formula as follows:

136 **Risk score = ∑ Gi*Ci** (Annotation: Gi=gene expression level, Ci = coefficient of each gene)

137 A package called "timeROC" was used to generate time-dependent ROC curves. We used the
138 TCGA data as training set, and four GEO data, including GSE36471, GSE72094, GSE8894 and
139 GSE37745, were used as validation sets. GEO data was analyzed using GEO2R. Since GSE8894 and
140 GSE37745 are sequencing sets for all NSCLC types, we deleted non-LUAD data, such as LUSC, on
141 them.

142 **Immunohistochemistry(IHC)**

143 Anatomical samples involving 10 LUAD and corresponding paracancerous tissues(Surgically
144 removed tissues within 2 cm of cancerous tissue) were arquired from the First Affiliated hospital of
145 Fujian Medical University (FJMU). All patients provided their informed consent. Tissues were fixed
146 in 10% formalin, embedded in paraffin, and cut into sections sequentially; and after dewaxing with
147 ethanol and blocking, rabbit anti-FGD2 (ab185968, 1:100, Abcam) was incubated overnight at 4°C,
148 followed by 50 minutes of incubation with goat anti-rabbit secondary antibody at room temperature.
149 After that, DAB (3,3′-diaminobenzidine) staining was performed, with positive expressions
150 appearing brownish yellow in color; and with hematoxylin, the nuclei of the cells were stained blue.
151 This study was approved by the Ethics Committee of the First Affiliated Hospital of FJMU.

152

153 **Statistical analysis**

154 Wilcoxon test was performed for comparison across two groups. Correlation analyses were
155 determined by spearman rank correlation. Ns- $P \geq 0.05$, *- $P < 0.05$, **- $P < 0.01$, ***- $P < 0.001$.
156 All statistical analysis was conducted in R (v 3.6.3).

157 **Database websites**

158 The following are links for all the online databases used in this study:

159 TCGA:https://portal.gdc.cancer.gov

160 UALCAN:http://ualcan.path.uab.edu/analysis.html

161 HPA:https://www.proteinatlas.org

162 Sangerbox:http://vip.sangerbox.com/login.html

163 GSCA:http://bioinfo.life.hust.edu.cn/GSCA/#/

164 GDSC:https://www.cancerrxgene.org/

165 GEPIA:http://gepia.cancer-pku.cn

166 KMplotter:https://kmplot.com/analysis/

167 TIMER:http://timer.cistrome.org

168 TIP:http://biocc.hrbmu.edu.cn/TIP/

169 TIDE:http://tide.dfci.harvard.edu

170 TCIA:https://tcia.at/home

171 ICBatlas:http://bioinfo.life.hust.edu.cn/ICBatlas/

172 ROCplotter:http://www.rocplot.org/site/index

173 GENEMania:http://genemania.org

174 TIGER:http://tiger.canceromics.org/#/home

175

176 **RESULTS**

177 **FGD family genes' expression pattern, prognostic value, and relationship with clinical**
178 **characteristics.**

179 We first analyzed the correlations among the six members of the FGD family. There is no strong
180 correlation between FGD1, FGD4, or FGD6 with any of the other members of the FGD family;
181 however, FGD2, FGD3, and FGD5 are strongly connected (Figure1 A). Then we specifically
182 analyzed the mRNA level of them in lung adenocarcinoma from TCGA database. Both the
183 comparison between healthy ones and LUAD patients, and the comparison between cancer tissues
184 and adjacent normal tissues from LUAD patients, showed that the levels of FGD2, 3, 4, 5 in LUAD
185 were decreased compared with normal ones, while the levels of FGD1, 6 were increased(Figure1
186 B,C). Besides, the UALCAN database showed that their protein expression pattern matched their
187 mRNA expression pattern(Figure1 D). For FGD1-6, the area under the ROC curves are 0.626, 0.779,
188 0.871, 0.883, 0.982, and 0.804, respectively(Figure1 E), which indicates that except FGD1, the other
189 5 members of the FGD family all have good diagnostic efficiency.

190 In the KM survival analysis of patients with LUAD in TCGA database, we found that prognosis
191 was poor when FGD2, FGD3 and FGD5 levels were low; while FGD1, FGD4, and FGD6 didn't
192 affect patients' survival(Figure1 F). Based on the KMplotter database, FGD2, 3, 5's prognostic value
193 was further confirmed (FigureS1 A-C). Additionally, Cox regression analysis showed that FGD2, 3,
194 5 could be used as independent prognostic factors for lung adenocarcinoma after adjusting for other
195 clinical features' effect (FigureS1 A-C).

196 After exploring the relationship between FGD1-6 and clinical features, we found that advanced T,
197 N, pathological stages and lethal OS event contributed to lower FGD2, 3, 5 levels, while after
198 primary therapy, PR&CR patients' FGD2, 3, 5 levels were higher than those with PD&SD outcomes
199 (Figure1 G). It is worth mentioning that FGD1, 4, 6 weren't associated with these clinical
200 features(FigureS1 D). We also analyzed the mutation frequency of FGD1-6 in LUAD, and found that
201 FGD5 had the highest mutation frequency, followed by FGD6, FGD1, FGD4, FGD3, and
202 FGD2(Figure1 H). Interestingly, patients with mutant FGD2, 3, 5 had a worse prognosis, while those
203 with FGD1, 4, 6 mutations had no difference in prognosis from those without(Figure1 I, J).
204 Finally, we constructed a miRNA-gene network related to FGD1-6, as shown in Figure 1K.

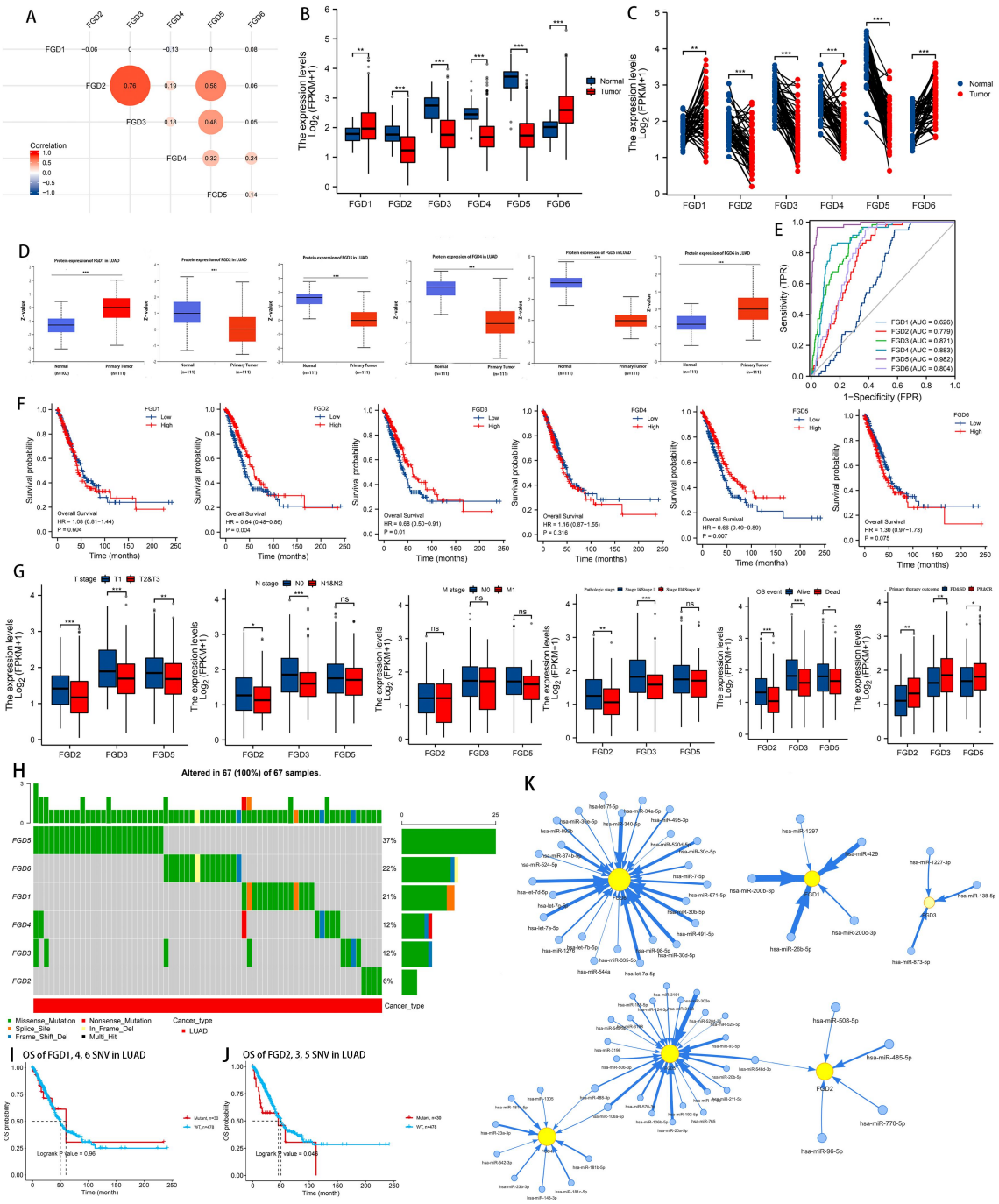


Figure 1. FGDs' expression pattern, diagnostic and prognostic value, relationship with clinical features, mutation data and miRNA-gene network. (A) Mutual correlations between six FGDs. (B) Comparison of FGDs' mRNA expression in LUAD and in healthy individuals. (C) Paired sample comparison of FGDs' mRNA level between LUAD and adjacent healthy ones. (D) The total FGDs' protein in LUAD and normal tissues. (E) The ROC curve for FGDs. (F) FGDs' prognostic value. (G) Relationship between FGDs and clinical characteristics of LUAD. (H) Mutation data of FGDs in LUAD. (I- J) Mutant FGD1, 4, 6(I) and FGD2, 3, 5(J) on LUAD patients' OS. (K) The miRNA-gene network.

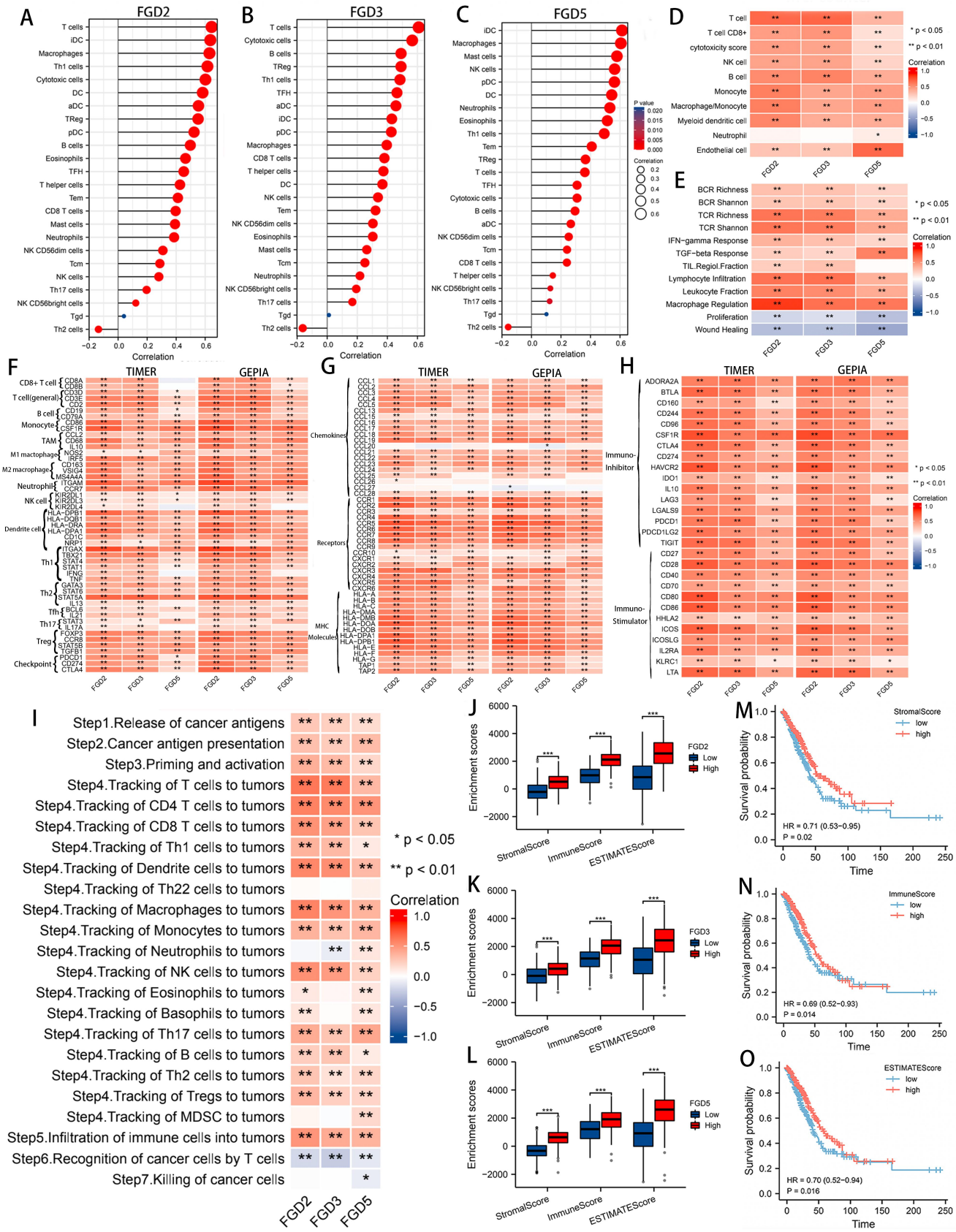
213 **FGD2, 3, 5 are significantly correlated with immunity**

214 Using ssGSEA, we found that FGD2, FGD3 and FGD5 were positively correlated with most
215 types of immunocytes. FGD2 and FGD3 are most correlated with T cells, while FGD5 is most
216 correlated with iDCs. Interestingly, all of them are weakly correlated with Tgd, and negatively
217 correlated with Th2 cells(Figure2 A-C). MCPcounter generated the similar results, indicating that
218 FGD2, 3, 5 are positively correlated with most types of immunocytes(Figure2 D). In addition, we
219 investigated the relationship between FGD2, 3, 5 and TCR/BCR diversity, two inflammation-related
220 pathways, four kinds of immune behaviors, in conjunction with two malignant phenotypes. It was
221 found that FGD2, 3, 5 are strongly positively correlated with immune characteristics, but negatively
222 correlated with two malignant phenotypes, namely proliferation and wound healing(Figure2 E).
223 Besides, by examining the correlation between these three FGD genes and markers of different
224 immune cells from the TIMER and GEPIA databases, we found that both databases showed that
225 these three FGD genes were positively associated with most immune cells' markers and the three
226 most common immune checkpoints: PD1(PDCD1), PD-L1(CD274) and CTLA4(Figure2 F).

227 We also examined the relationship between FGD2, 3, 5 and chemokines/receptors, MHC
228 molecules(MHCs), and immunomodulators, since they also play an important role in regulating the
229 TIME. FGD2, 3, 5 were shown to be positively correlated with most chemokines/receptors, MHCs,
230 and immunoinhibitors/immunostimulators, further supporting the theory that FGD2, 3, 5 are involved
231 in the regulation of the TIME (Figure 2 G, H). Besides, in line with our expectations, patients in the
232 high-FGD2, 3, 5 groups had higher stromal, immune, and ESTIMATE scores(Figure2 J- L).
233 However, these three scores differed insignificantly between high and low FGD1, 4, 6
234 groups(FigureS2 D), and FGD1, 4, 6's correlations with immunity are weak(FigureS2 A- C).

235 The stromal, immune, and ESTIMATE scores' effect on LUAD patients' OS are illustrated in
236 Figure2 M-O. As a result of the above analysis, we concluded that FGD2, FGD3 and FGD5 might
237 effect LUAD patients' prognosis by regulating the TIME. Finally, according to the results of

238 analyzing the relationship of FGD2, 3, 5 and seven stepwise anti-cancer immune cycles, we
239 hypothesized that FGD2, 3, 5 may regulating the TIME through recruiting immune cells to tumors
240 and guiding immune cells' infiltration(Figure2 I). However, the value of FGD1, 4, 6 needs to be
241 further studied.
242 .


A
FGD2
T cells
iDC
Macrophages
Th1 cells
Cytotoxic cells
DC
aDC
TReg
pDC
B cells
Eosinophils
TFH
T helper cells
Tem
CD8 T cells
Mast cells
Neutrophils
NK CD56dim cells
Tcm
NK cells
Th17 cells
NK CD56bright cells
Tgd
Th2 cells
Correlation
B
FGD3
Cytotoxic cells
B cells
TReg
Th1 cells
TFH
aDC
iDC
pDC
Macrophages
CD8 T cells
T helper cells
DC
NK cells
Tem
NK CD56dim cells
Eosinophils
Mast cells
Tcm
Neutrophils
NK CD56bright cells
Th17 cells
Tgd
Th2 cells
Correlation
C
FGD5
iDC
Macrophages
Mast cells
NK cells
pDC
DC
Neutrophils
Eosinophils
Th1 cells
Tem
TReg
T cells
TFH
Cytotoxic cells
B cells
aDC
NK CD56dim cells
Tcm
CD8 T cells
T helper cells
NK CD56bright cells
Th17 cells
Tgd
Th2 cells
Correlation
P value
Correlation
D
T cell
T cell CD8+
cytotoxicity score
NK cell
B cell
Monocyte
Macrophage/Monocyte
Myeloid dendritic cell
Neutrophil
Endothelial cell
FGD2
FGD3
FGD5
* p < 0.05
** p < 0.01
Correlation
E
BCR Richness
BCR Shannon
TCR Richness
TCR Shannon
IFN−gamma Response
TGF−beta Response
TIL.Regiol.Fraction
Lymphocyte Infiltration
Leukocyte Fraction
Macrophage Regulation
Proliferation
Wound Healing
FGD2
FGD3
FGD5
* p < 0.05
** p < 0.01
Correlation
F
TIMER
GEPIA
CD8+ T cell
T cell(general)
B cell
Monocyte
TAM
M1 mactophage
M2 macrophage
Neutrophil
NK cell
Dendrite cell
Th1
Th2
Tfh
Th17
Treg
Checkpoint
G
TIMER
GEPIA
Chemokines
Receptors
MHC Molecules
H
TIMER
GEPIA
Immuno-Inhibitor
Immuno-Stimulator
* p < 0.05
** p < 0.01
Correlation
I
Step1.Release of cancer antigens
Step2.Cancer antigen presentation
Step3.Priming and activation
Step4.Tracking of T cells to tumors
Step4.Tracking of CD4 T cells to tumors
Step4.Tracking of CD8 T cells to tumors
Step4.Tracking of Th1 cells to tumors
Step4.Tracking of Dendrite cells to tumors
Step4.Tracking of Th22 cells to tumors
Step4.Tracking of Macrophages to tumors
Step4.Tracking of Monocytes to tumors
Step4.Tracking of Neutrophils to tumors
Step4.Tracking of NK cells to tumors
Step4.Tracking of Eosinophils to tumors
Step4.Tracking of Basophils to tumors
Step4.Tracking of Th17 cells to tumors
Step4.Tracking of B cells to tumors
Step4.Tracking of Th2 cells to tumors
Step4.Tracking of Tregs to tumors
Step4.Tracking of MDSC to tumors
Step5.Infiltration of immune cells into tumors
Step6.Recognition of cancer cells by T cells
Step7.Killing of cancer cells
FGD2
FGD3
FGD5
* p < 0.05
** p < 0.01
Correlation
J
Enrichment scores
FGD2
Low
High
StromalScore
ImmuneScore
ESTIMATEScore
K
Enrichment scores
FGD3
Low
High
StromalScore
ImmuneScore
ESTIMATEScore
L
Enrichment scores
FGD5
Low
High
StromalScore
ImmuneScore
ESTIMATEScore
M
StromalScore
low
high
Survival probability
HR = 0.71 (0.53−0.95)
P = 0.02
Time
N
ImmuneScore
low
high
Survival probability
HR = 0.69 (0.52−0.93)
P = 0.014
Time
O
ESTIMATEScore
low
high
Survival probability
HR = 0.70 (0.52−0.94)
P = 0.016
Time

243

244 **Figure 2. Immune infiltration analysis.** (A-D) Correlation between FGD2, 3, 5 and immunocytes using ssGSEA(A, B, C) and
245 MCPcounter(D) methods. (E) FGD2, 3, 5's relationship with other immune features. (F) Relationship between three FGD genes and
246 immune cell markers and immune checkpoints in TIMER and GEPIA. (G) Relationship between three FGD genes and
247 chemokines/receptors and MHC molecules in TIMER and GEPIA. (H) Relationship between three FGD genes and Immunomodulators
248 in TIMER and GEPIA. (I)FGD2, 3, 5's relationship with seven anti-cancer immune cycles. (J-L) Stromal, Immune, and ESTIMATE
249 scores between high and low FGD2(J), FGD3(K), and FGD5(L) groups in LUAD. (M-O) Stromal(M) ,immune(N), and ESTIMATE(O)
250 scores' effect on LUAD patients' OS.

251

252 **Single-cell RNA sequencing analysis**

253 In view of the above analysis, we thought that FGD2, 3, 5 participate in the regulation of the TIME
254 in LUAD. To further demonstrate this, we investigated the expression of FGD2, 3, 5 at cellular level.
255 Two single-cell RNA-sequencing sets for LUAD, namely GSE131907 and GSE123904, were
256 included in this study. 208506 cells in the GSE131907 dataset can be divided into 11 cluster(Figure3
257 B). We analyzed the expression of FGD2, 3, 5 in subpopulations and found that they were mainly
258 expressed in B cell&myeloid cell, T cell&B cell, and endothelial cell&myeloid cell,
259 respectively(Figure3 C). Then, by performing the pseudo-time analysis, we found that FGD2, 3, 5
260 expressed more when cells differentiated toward immunocytes, and less when cells differentiated
261 toward other types(Figure3 D, E). 38758 cells in the GSE123904 dataset can be divided into 7
262 cluster(Figure3 G), and we found that FGD2, 3, 5's expression pattern in this dataset is the same with
263 GSE131907, indicating that they were mainly expressed in B cell&myeloid cell, T cell&B cell, and
264 endothelial cell&myeloid cell, respectively(Figure3 H). The result of pseudo-time analysis for cells
265 in the GSE123904 dataset is also consistent with that of GSE131907, showed that when cells tended
266 to differentiate into cells which are correlated with immunity, FGD2, 3, 5's expression
267 improved(Figure3 I, J). Finally, from the TISIDB database, we acknowledged that FGD2, 3, 5's level
268 significantly differed between 6 different immune subtypes, namely C1-C6(Figure3 K-M). Figure3 N
269 illustrates the subtypes' annotation.

270 The above analysis helps us understand the expression pattern of FGD2, 3, 5 better and further
271 proves that they participate in the regulation of the TIME.

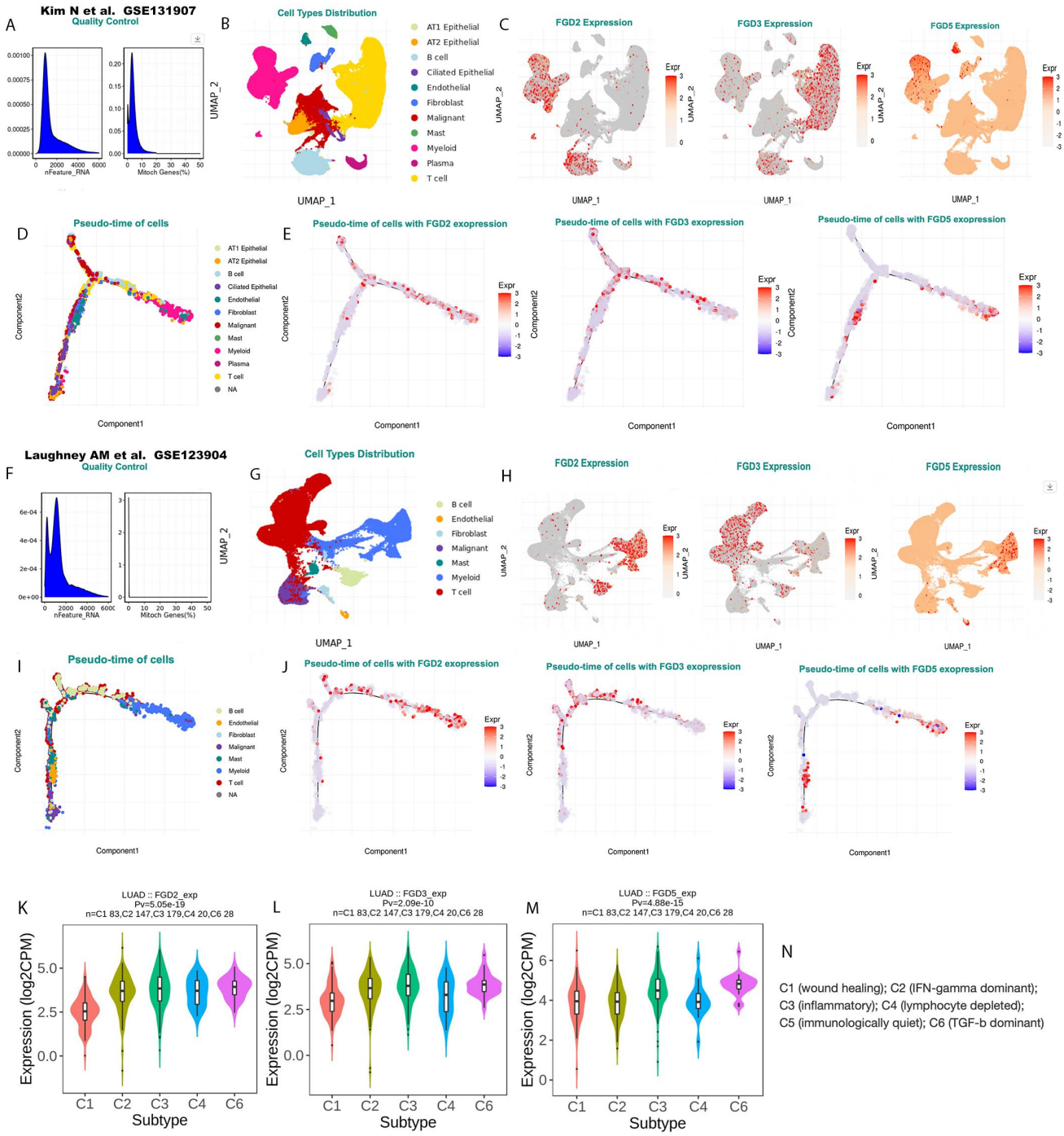


Figure3. Single-cell RNA-sequencing analysis. (A, F) Quality control map for sequencing data in GSE131907(A) and GSE123904(B) sets. (B, G) Gene ontology class in GSE131907(B) and GSE12904(G) sets. (C, H) FGD2, 3, 5's expression in different cell clusters in GSE131907(C) and GSE123904(H). (D, I) Cell differentiation tracks in GSE131907(D) and GSE123904(I). (E, J) Expression of FGD2, 3, 5 during cell differentiation locus in GSE131907(E) and GSE123904(J). (K-M) FGD2(K), FGD3(L), and FGD5(M)'s expression between different immune subtypes. (N) Annotation for immune subtypes.

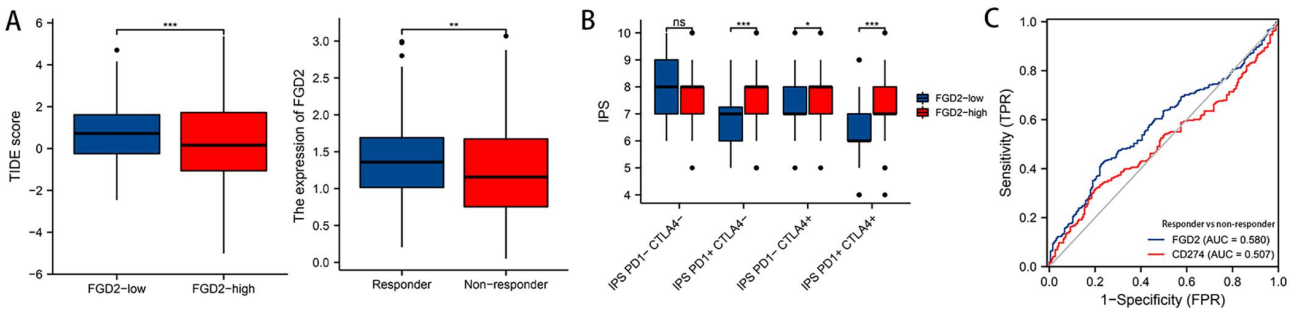
279 **FGD2 level can predict ICB therapy efficacy**

280 In LUAD, a positive correlation between FGD2, 3, 5 and the vast majority of TIME modulators,
281 including PDCD1 and CD274 , was found. Therefore, we surmise that LUAD patients with higher
282 FGD2, 3, 5 level may be more sensitive to ICB therapy. To further explore it, we first predicted the
283 potential ICB response using the TIDE algorithm. Consistent with our expectations, the TIDE score
284 in the high-FGD2 group is significantly lower than in the low-FGD2 group, which suggests that
285 LUAD patients with high FGD2 levels may be more responsive to ICB therapy (Figure4 A). It's
286 interestingly to note that the TIDE algorithm believed that FGD2 was more efficient than PD-L1 in
287 diagnosing ICB responders and non-responders (Figure4 C).

288 Next, with the help of the TCIA database, we validated our hypothesis again: we found that the
289 IPS-PD1 block, IPS-CTLA4 block and IPS-PD1-CTLA4 block scores were higher in the high FGD2
290 expression group (Figure4 B), which means LUAD patients with higher FGD2 levels are more
291 sensitive to anti-PD1, anti-CTLA4 and anti-PD1&CTLA4 therapy. Then, from the ICBatlas database,
292 we found that NSCLC patients who responded to anti-PD1 and anti-PD-L1 treatment have higher
293 FGD2 levels than non-responders, which further confirms our conjecture (Figure4 D). Interestingly,
294 we found that melanoma patients who responded to ICB therapy had higher FGD2 levels than non-
295 responders (Figure4 D). This pushed us to analyze the data of patients with other cancers (including
296 melanoma, gastric cancer, bladder cancer, breast cancer and glioblastoma et al.) receiving ICB
297 treatment from ROCplotter database. Based on pre-treatment samples, we found that responders to
298 anti-PD1, anti-PD-L1, and anti-CTLA4 therapy have higher FGD2 levels(Figure4 E). ROCplotter
299 also showed that compared with non-responders to ICB treatment, PD-L1 levels were higher in ICB-
300 treated responders (Figure4 F), and it still illustrated lower diagnostic efficacy of PD-L1 than FGD2
301 (Figure4 G). Similar results were observed in on-treatment samples(Figure4 H- J). Finally, we found
302 from KMplotter that when those ICB-treated cancer patients' FGD2 level was higher, their prognosis
303 was better (Figure4 K-M).

304 The aforementioned results illustrate the value of FGD2 in predicting ICB effectiveness, not only
305 for NSCLC patients, but for other cancer patients. However, because FGD2 was not significantly
306 better than PD-L1 in distinguishing ICB-treated responders from non-responders, whether FGD2 can
307 replace PD-L1 as an independent biomarker for predicting ICB response needs to be further verified.

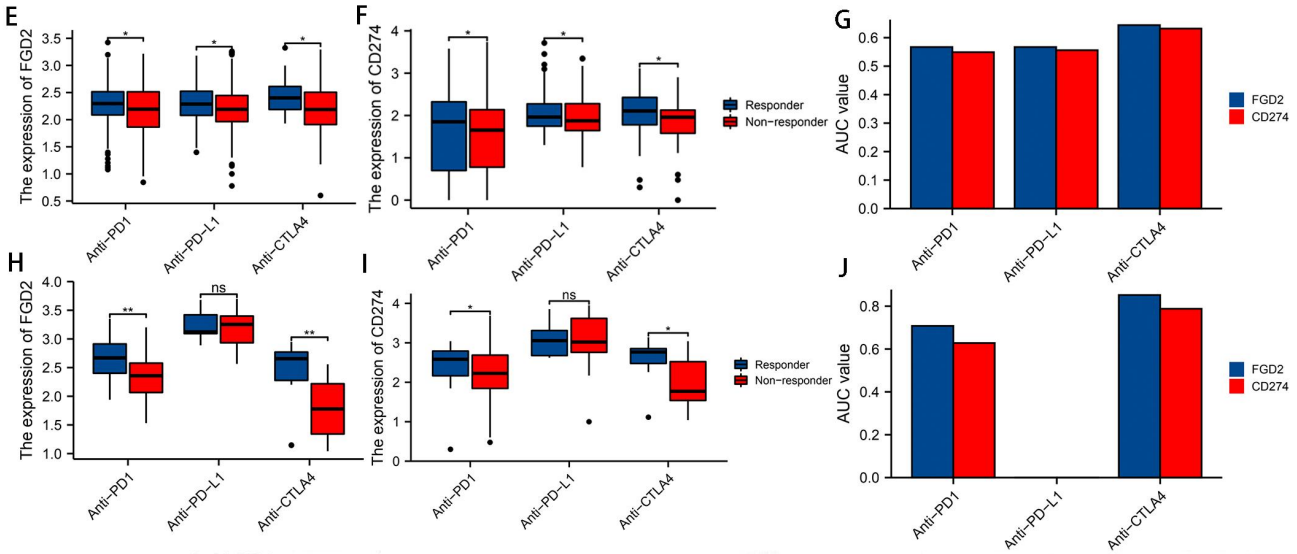
308 The high and low FGD3 expression groups did not significantly differ in TIDE score(FigureS3 A),
309 and contradicted with our expectations, patients with high FGD5 expression scored higher on TIDE
310 than patients with low FGD5 expression(FigureS3 B). In addition, we found that both FGD3 and
311 FGD5 were less effective than PD-L1 in differentiating immunotherapy responders from
312 nonresponders in the ROCplotter database(FigureS3 A, B). So FGD3 and FGD5's value for
313 predicting ICB efficacy is lower than FGD2.
314

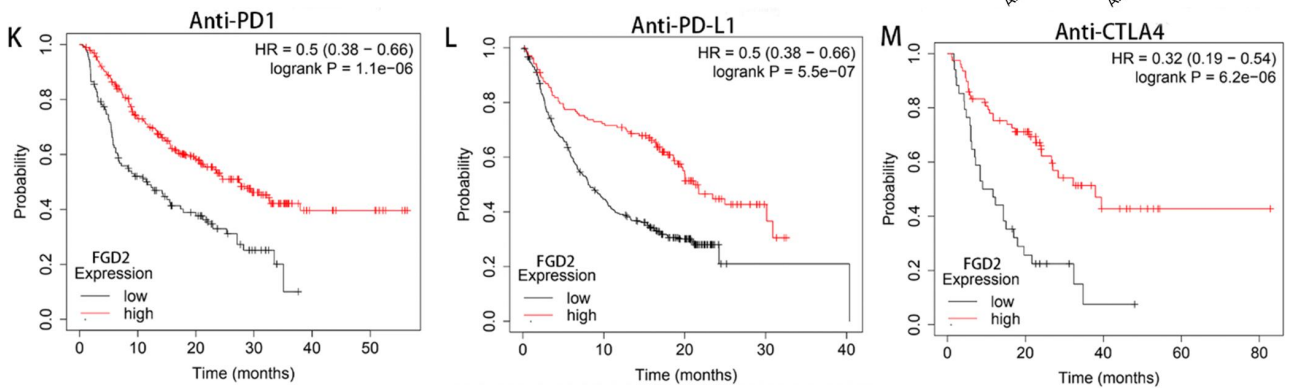


D

FGD2 expression for Response and Non-response based on Pre-treatment Samples in each dataset

Study	Cancers	Anti Target	Responders' FGD2 mean value	Non-responders' FGD2 mean value	Log2FC	FDR	P Value
anti-PD1	Melanoma,NSCLC,GBM,RCC,GC	anti-PD1	488.50	310.50	0.500 ↑	0.003	0.000
ERP105482,SRP011540,SRP070710,SRP094781,SRP150548,SRP230414,SRP250849,SRP302761	Melanoma	anti-PD1/anti-CTLA4/anti-PD1 + anti-CTLA4	430.00	357.50	0.483 ↑	0.022	0.001
SRP183455	Non-small Cell Lung Cancer	anti-PD1	780.50	226.50	1.854 ↑	0.058	0.002
ERP105482	Melanoma	anti-PD1	429.00	195.50	0.954 ↑	0.070	0.002
SRP183455,SRP217040	NSCLC	anti-PD1/PDL1	427.50	235.00	0.922 ↑	0.173	0.023
SRP011540	Melanoma	anti-PD1	510.50	342.50	0.490 ↑	0.252	0.048





315

316 **Figure 4. The value of FGD2 in predicting response to ICB treatment.** (A) Potential ICB treatment outcome predicted by TIDE
317 algorithm. (B) IPS comparison between high and low FGD2 expression groups. (C) Comparison of the efficacy of FGD2 and PD-L1 in
318 distinguishing ICB responders from non-responders by the TIDE predictive algorithm. (D) Differently expressed FGD2 between ICB
319 treated NSCLC responders and non-responders. (E-G) Differential expression of FGD2(E) and CD274(F) in anti-PD-1/PD-L1/CTLA4
320 reactive and non-reactive cancer patients and the comparison of their diagnostic efficacy(G) based on pre-treatment samples. (H-J)
321 Differential expression of FGD2(H) and CD274(I) in anti-PD-1/PD-L1/CTLA4 reactive and non-reactive cancer patients and the
322 comparison of their diagnostic efficacy(J) based on on-treatment samples. (K-M) FGD2's effect on anti-PD1(K)/PD-L1(L)/CTLA4(M)
323 treated patients' OS.

324

325 **GSEA**

326 According to our previous analysis, FGD2 can be used as an ICB-related biomarker, not only for
327 LUAD, but also for other cancers. To further prove that, we conduct the GSEA analysis with the help
328 of the CAMOIP database (http://www.camoip.net). As expected, GSEA analysis indicated that FGD2
329 is strongly correlated with PD-L1 expression and PD1 checkpoint pathway in cancer and the PD1
330 signaling. This result held true across 5 independent cohorts (Figure 5). Once again, FGD2
331 demonstrates its predictive ability.

332

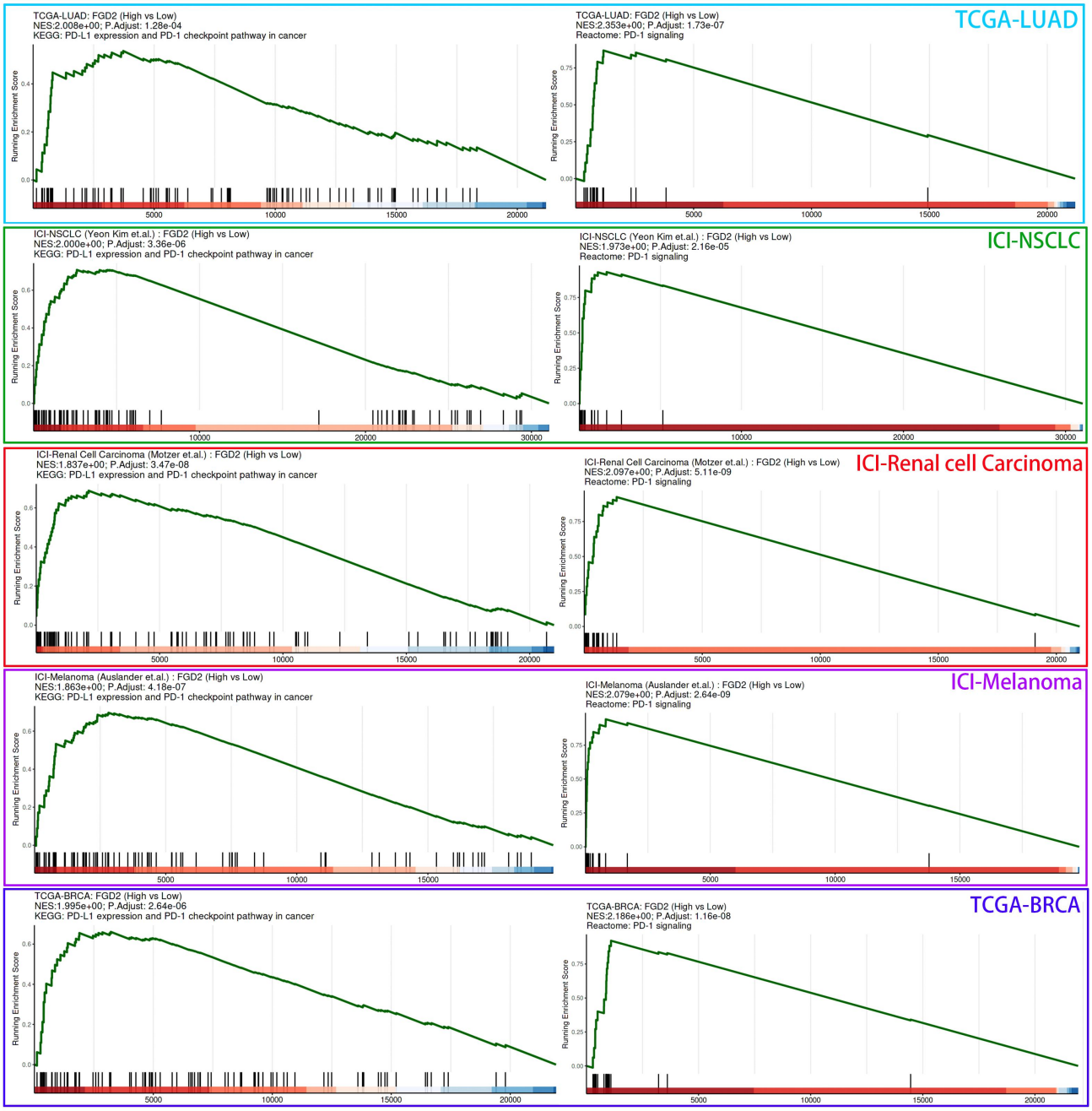


Figure 5. GSEA. This figure illustrates FGD2's relationship with immune checkpoint pathways in different cancer types based on the GSEA algorithm.

Gene co-expression analysis and enrichment analysis

We retrieved the top 10 genes that have the highest positive/negative correlation with FGD2, FGD3 and FGD5, respectively, and visualized them using heat maps (Figure6 A-C). Additionally, we

340 downloaded the top 100 genes from the GEPIA database that were positively correlated with FGD2,
341 FGD3 and FGD5, respectively, and used a Venn diagram to determine where their intersections
342 occurred and we obtained 18 genes co-expressed with FGD2, 3, 5, including MYO1F, ARHGAP30,
343 DOCK2, FAM78A, NCKAP1L, GIMAP1, DOK2, FERMT3, SPN, CD4, NFAM1, ARHGEF6,
344 NRROS, STK10, PLEKHO2, GIMAP8, GIMAP6, and SLC15A3 (Figure6 D). We used
345 GEMENANIA database to predict co-expression, shared protein domains, genetic intersections and
346 physical interactions among these 18 genes and found that their main functions are lymphocyte
347 differentiation, adaptive immune response, leukocyte migration, B cell activation, amd T cell
348 differentiation (Figure6 E). We also performed GO analysis on these 18 genes, and found the most
349 relevant biological functions among them were lymphocyte differentiation and leukocyte cell-cell
350 adhesion (Figure6 F).

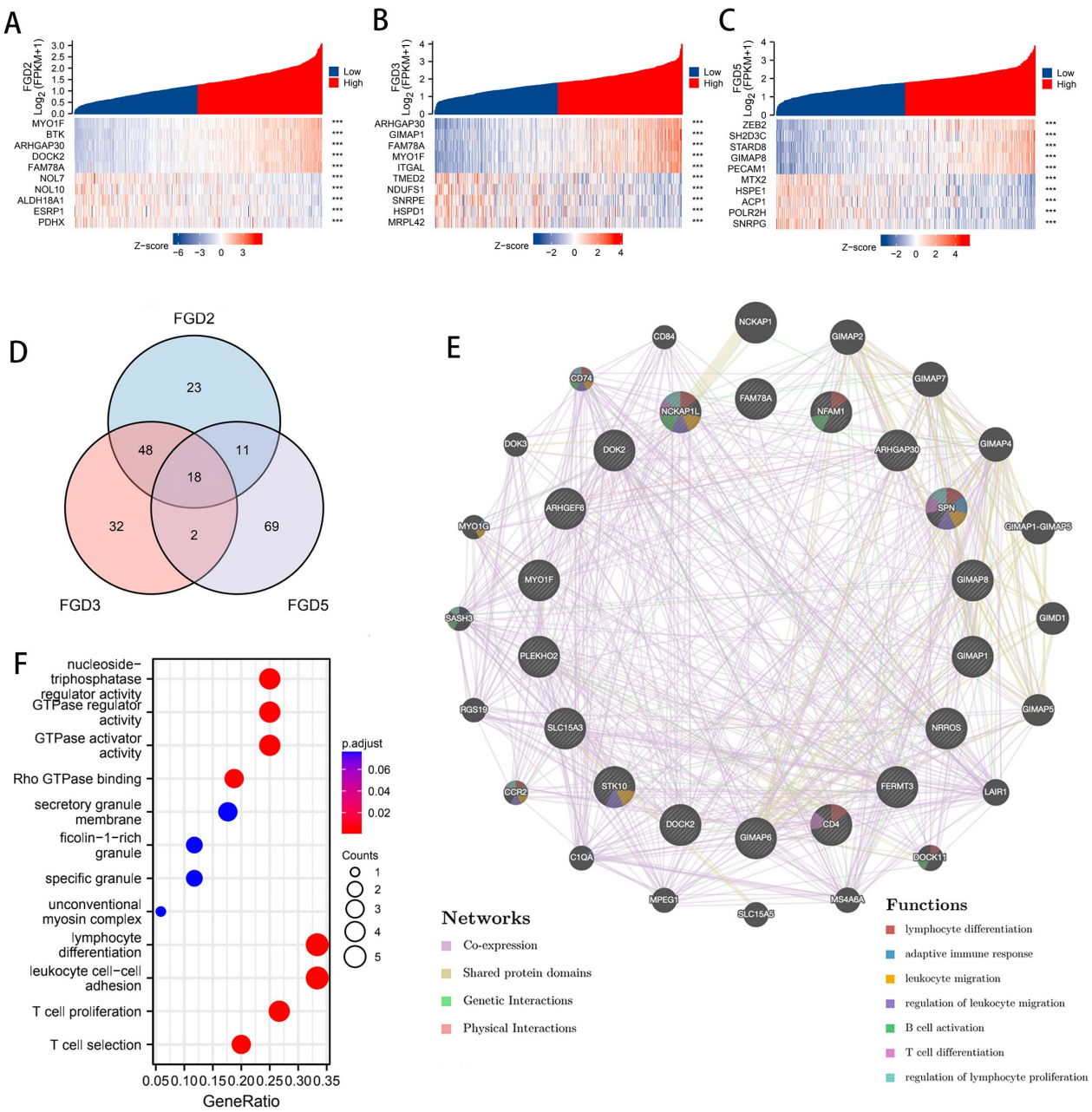
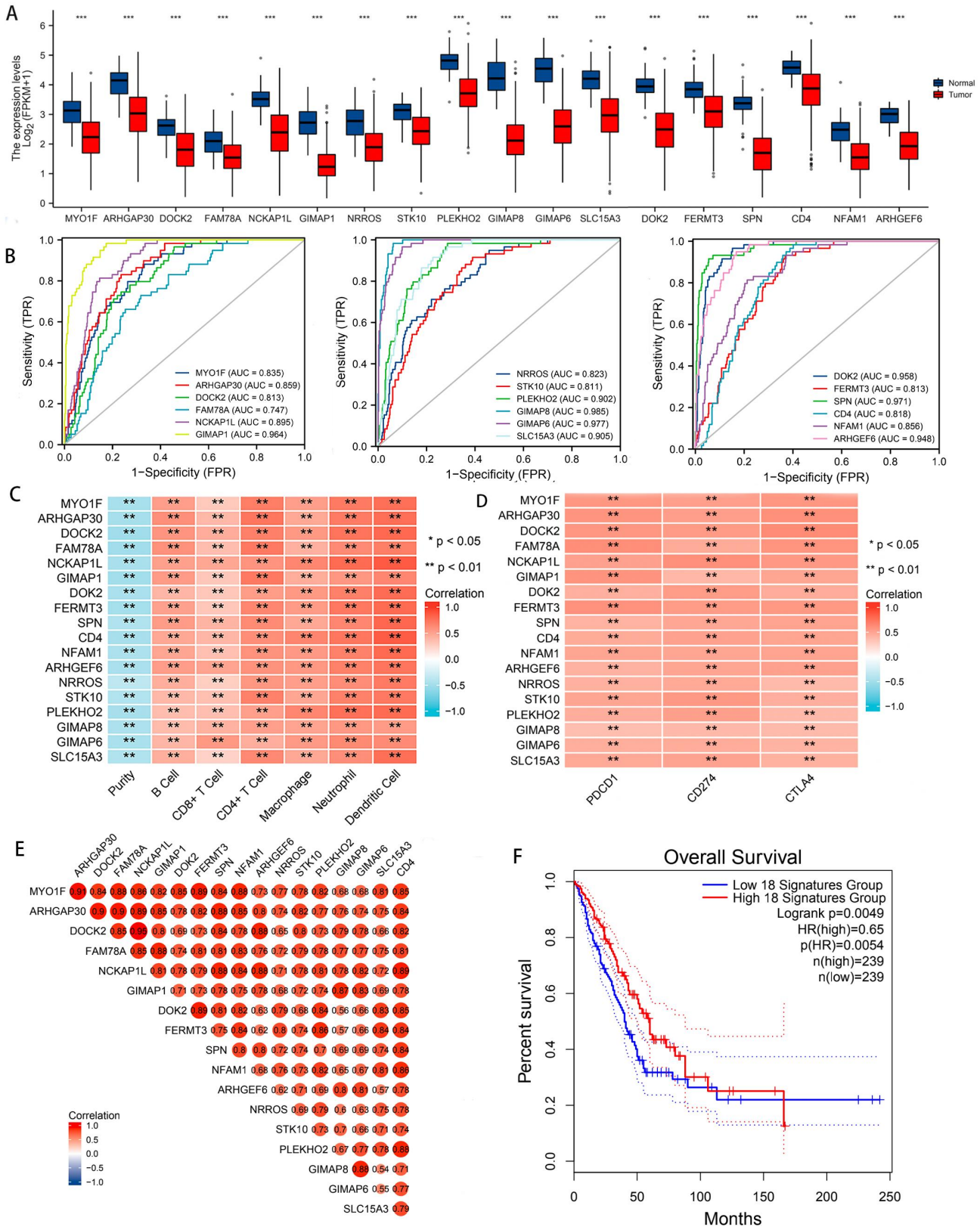


Figure 6. Genes co-expressed with these three FGD genes and enrichment analysis. (A-C) 10 genes with the most positive/negative correlation with FGD2 (A), FGD3 (B) and FGD5 (C). (D) Venn diagram of genes co-expressed with three FGD genes. (E) PPI network. (F) GO enrichment analysis.

Expression pattern and immune infiltration analysis of 18 co-expressed genes in LUAD

The expression of these 18 genes in LUAD was lower than normal ones, which was the same with the expression patterns of FGD2, FGD3 and FGD5 (Figure7 A), and all of them have satisfactory

359 diagnostic efficacy(Figure7 B). To determine if their relationship with immune infiltration is also
360 similar to that of FGD2, FGD3 and FGD5, we analyzed the relationship between these 18 genes and
361 tumor purity and 6 types of immune cells in TIMER. Consistent with our expectation, the 18 genes
362 were negatively correlated with tumor purity, while positively correlated with CD8+ T cell, CD4+ T
363 cell, B cell, macrophage, DC, and neutrophil (Figure7 C). We also analyzed the relationship between
364 them and PDCD1, CD274, and CTLA4. As expected, all of them were positively correlated with
365 these three immune checkpoints (Figure7 D). In addition, these 18 genes was positively correlated
366 with each other (Figure7 E), and the combined expression of them had a significant effect on OS in
367 LUAD patients (Figure7 F).


A
The expression levels Log2 (FPKM+1)
Normal
Tumor
MYO1F ARHGAP30 DOCK2 FAM78A NCKAP1L GIMAP1 NRROS STK10 PLEKHO2 GIMAP8 GIMAP6 SLC15A3 DOK2 FERMT3 SPN CD4 NFAM1 ARHGEF6
B
Sensitivity (TPR)
1−Specificity (FPR)
MYO1F (AUC = 0.835)
ARHGAP30 (AUC = 0.859)
DOCK2 (AUC = 0.813)
FAM78A (AUC = 0.747)
NCKAP1L (AUC = 0.895)
GIMAP1 (AUC = 0.964)
NRROS (AUC = 0.823)
STK10 (AUC = 0.811)
PLEKHO2 (AUC = 0.902)
GIMAP8 (AUC = 0.985)
GIMAP6 (AUC = 0.977)
SLC15A3 (AUC = 0.905)
DOK2 (AUC = 0.958)
FERMT3 (AUC = 0.813)
SPN (AUC = 0.971)
CD4 (AUC = 0.818)
NFAM1 (AUC = 0.856)
ARHGEF6 (AUC = 0.948)
C
Purity
B Cell
CD8+ T Cell
CD4+ T Cell
Macrophage
Neutrophil
Dendritic Cell
* p < 0.05
** p < 0.01
Correlation
D
PDCD1
CD274
CTLA4
E
Correlation
F
Overall Survival
Low 18 Signatures Group
High 18 Signatures Group
Logrank p=0.0049
HR(high)=0.65
p(HR)=0.0054
n(high)=239
n(low)=239
Percent survival
Months

370 **Figure 7. The expression patterns of 18 genes co-expressed with FGD2, 3, 5 and their correlation with immune infiltration in**
371 **LUAD.** (A-D) These 18 genes' expression pattern (A), their diagnostic efficacy (B) and their relationships with six immune cells (C)
372 and three immune checkpoints (D). (E) Correlation between these 18 genes. (F) Effect of combined expression of these 18 genes on OS
373 in LUAD patients.
374

375 **Five of the 18 genes affected the prognosis of patients with LUAD**

376 We were also interested in the impact of each of these 18 genes on the prognosis of LUAD patients
377 and from the GEPIA database, we found that all of their HR were less than 1, but only MYO1F,
378 ARHGAP30, DOCK2, FAM78A, NCKAP1L, GIMAP1, NRROS, GIMAP6, ARHGEF6 were
379 statistically significant (FigureS4 A). The prognostic value of these 9 genes was further verified from
380 the KMplotter database. Interestingly, the HR of MYO1F and NCKAP1L in KMplotter database was
381 greater than 1, contrary to that in GEPIA. ARHGAP30 and NRROS have no significant effect on the
382 prognosis of LUAD patients. Only DOCK2, FAM78A, GIMAP1, DOK2 and ARHGEF6 had the
383 same effect on the prognosis of LUAD patients as in the GEPIA database (FigureS4 B). Therefore,
384 we hypothesized that DOCK2, FAM78A, GIMAP1, DOK2 and ARHGEF6 may also be potential
385 prognostic factors for LUAD.

386 **FGD2, 3, 5 and FGD1, 4, 6 are functionally heterogeneous**

387 Based on the above analysis of the FGD family, we concluded that FGD2, 3, 5 were prognostic
388 factors of LUAD, and all of them, with the 18 genes co-expressed with them, may be the regulators
389 of the TIME. However, FGD1, 4, 6 did not affect the prognosis of LUAD patients and showed only a
390 weak correlation with TIME. Therefore, we hypothesized that the functions of these 6 members of
391 the FGD family in LUAD might be heterogeneous. To verify it, we clustered all LUAD patients
392 based on these six FGDs' expression. We found that when k=2 , the clustering effect was
393 best(Figure8 A, B). Then, all LUAD patients were divided into two groups by unsupervised
394 clustering. Group 1: high expression of FGD2, 3, 5; low expression of FGD1, 4, 6. Group 2: low

395 expression of FGD2, 3, 5 and high expression of FGD1, 4, 6 (Figure8 C). Patients in group1 had
396 better survival (Figure8 D), higher expression of immune checkpoint molecules (Figure8 F), and
397 higher levels of immune infiltration (Figure8 E, G). These results indicate that FGD2, 3, 5 and FGD1,
398 4, 6 play different roles in LUAD. Interestingly, we also found that patients in group1 are more
399 sensitive to gefitinib, a commonly used targeted drug for NSCLC (Figure8 H). Finally, the
400 differences in gene mutation frequencies between the two groups were evaluated. As shown in the
401 waterfall plot, the top 20 genes with the most significant differences in mutation frequencies were
402 visualized, and we found that KRAS topped the list (Figure8 I).

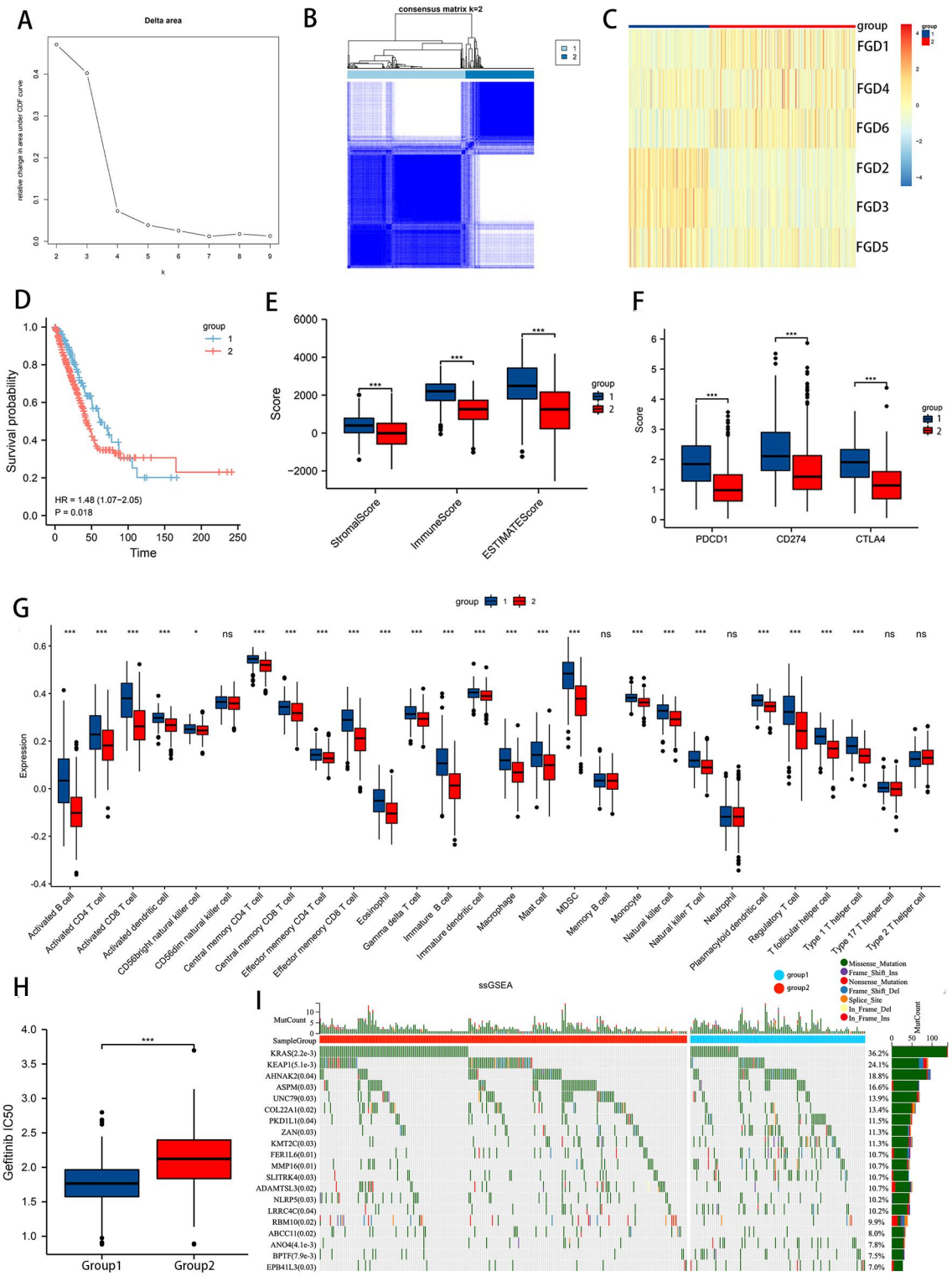


Figure 8. Heterogeneity of the FGD family. (A) Delta area. (B, C) Unsupervised clustering graph when k=2. (D) Survival curves of patients in different groups. (E, G) The immune infiltration level of patients in different groups calculated by ESTIMATE (E) and ssGSEA (G) algorithm. (F) Patients' immune checkpoint level in different groups. (H)Patients' sensitivity to gefitinib in different groups. (I) The waterfall plot.

408

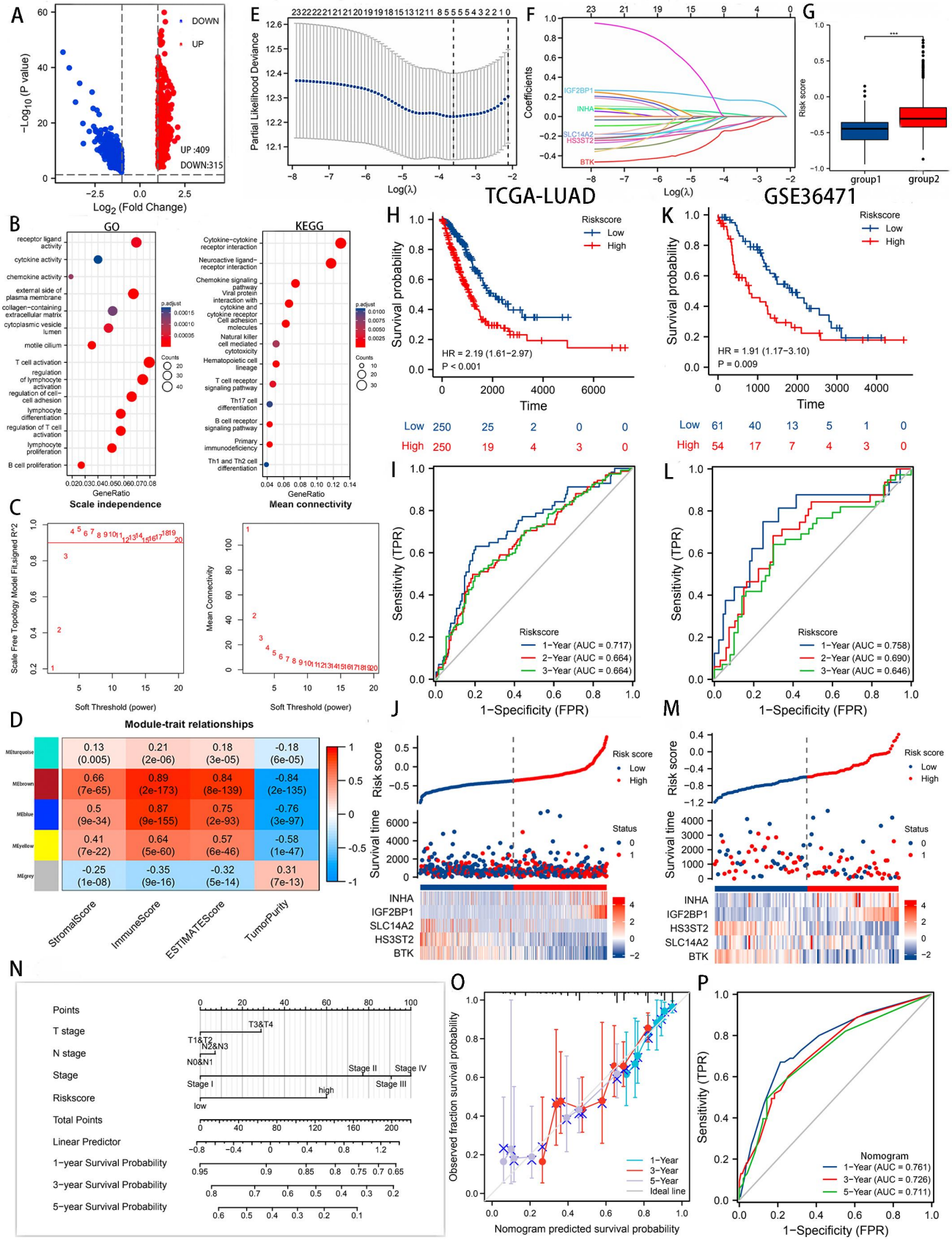
409 **Construction and validation of an immune-related prognostic model**

410 Differentially expressed genes (DEGs) were obtained according to a cutoff of |log2Fold change| >
411 1 and *P* adj < 0.05. We obtained the DEGs between the group1 and group2, and found that 409 were
412 up-regulated and 315 were down-regulated(Figure9 A). GO_KEGG enrichment analysis showed that
413 these 724 DEGs were mainly related to immune activity and immune-related signaling
414 pathways(Figure9 B). Then, WGCNA was performed on these DEGs, and we selected 4 as the
415 optimal soft threshold(Figure9 C). 116 DEGs in the brown module were thought to be most strongly
416 associated with immunity(Figure9 D). Univariate cox regression identified 24 of these 116 genes
417 influence the prognosis of patients with LUAD(FigureS5 A). Following the LASSO regression
418 analysis, we identified 5 key genes based on these 24 genes(Figure9 E, F). As a last step, we
419 constructed an immune-related prognostic model using five genes, specifically INHA, IGF2BP1,
420 SLC14A2, HS3ST2, and BTK. The following formula was used to calculate each patient's risk score:

421 *Risk score = INHA * (0.0218) + IGF2BP1 * (0.1661) + SLC14A2 * (- 0.0856) + HS3ST2 * (-*
422 *0.0787) + BTK * (- 0.1339).*

423 Interestingly, we found that patients in group1 (high FGD2, 3, 5 and low FGD1, 4, 6) had lower
424 risk score than patients in group2 (high FGD1, 4, 6 and low FGD2, 3, 5) (Figure9 G). Then based on
425 the optimal cut-off value of the risk score, LUAD patients were divided into high-risk (H-risk) and
426 low-risk (L-risk) groups. There were worse outcomes for patients in the H-risk group in both the
427 training set (TCGA-LUAD) and validation sets (GSE36471, GSE72094, GSE8894 and GSE37745)
428 (Figure9 H, K; FigureS5 D, E, F). Time-dependent ROC analysis showed that risk score had
429 satisfactory predictive efficiency in both training set and validation sets (Figure9 I, L; FigureS5 D, E,
430 F). As shown in Figure9 J, M, patients in the H-risk group had a higher mortality rate. In addition,
431 cox regression analysis using TCGA data showed that risk score is an independent prognostic
432 factor(FigureS5 B, C). It confirms that this immune-correlated prognostic model had good predictive
433 value.

434 Finally, we selected T, N, Stage and risk score to construct a nomogram prognosis model for OS.
435 In the nomogram, a higher score on the prognostic factor meant a lower survival rate, and higher total
436 points meant a worse prognosis. We found that higher risk score will lead to a higher score, and the
437 c-index of this model is 0.702 (Figure9 N). According to the calibration and ROC curves of this
438 model, we know that combining risk score with other clinical factors has a better ability to predict the
439 OS of LUAD patients(Figure9 O, P).


A
DOWN
UP
UP :409
DOWN:315
−Log10 (P value)
Log2 (Fold Change)
E
Partial Likelihood Deviance
Log(λ)
F
Coefficients
IGF2BP1
INHA
SLC14A2
HS3ST2
BTK
Log(λ)
G
Risk score
group1
group2
TCGA-LUAD
GSE36471
B
GO
KEGG
GeneRatio
Scale independence
Mean connectivity
C
Scale Free Topology Model Fit,signed R^2
Mean Connectivity
Soft Threshold (power)
H
Riskscore
Low
High
Survival probability
HR = 2.19 (1.61−2.97)
P < 0.001
Time
Low 250 25 2 0 0
High 250 19 4 3 0
K
HR = 1.91 (1.17−3.10)
P = 0.009
Low 61 40 13 5 1 0
High 54 17 7 4 3 0
I
Sensitivity (TPR)
1−Specificity (FPR)
1−Year (AUC = 0.717)
2−Year (AUC = 0.664)
3−Year (AUC = 0.664)
L
1−Year (AUC = 0.758)
2−Year (AUC = 0.690)
3−Year (AUC = 0.646)
D
Module-trait relationships
MEturquoise 0.13 (0.005) 0.21 (2e-06) 0.18 (3e-05) -0.18 (6e-05)
MEbrown 0.66 (7e-65) 0.89 (2e-173) 0.84 (8e-139) -0.84 (2e-135)
MEblue 0.5 (9e-34) 0.87 (9e-155) 0.75 (2e-93) -0.76 (3e-97)
MEyellow 0.41 (7e-22) 0.64 (5e-60) 0.57 (6e-46) -0.58 (1e-47)
MEgrey -0.25 (1e-08) -0.35 (9e-16) -0.32 (5e-14) 0.31 (7e-13)
StromalScore
ImmuneScore
ESTIMATEScore
TumorPurity
J
Risk score
Low
High
Survival time
Status
INHA
IGF2BP1
SLC14A2
HS3ST2
BTK
M
N
Points
T stage
N stage
Stage
Riskscore
Total Points
Linear Predictor
1−year Survival Probability
3−year Survival Probability
5−year Survival Probability
O
Observed fraction survival probability
Nomogram predicted survival probability
1−Year
3−Year
5−Year
Ideal line
P
Nomogram
1−Year (AUC = 0.761)
3−Year (AUC = 0.726)
5−Year (AUC = 0.711)

440

441 **Figure 9. The immune-related prognostic model.** (A) The volcano plot. (B) GO_KEGG analysis. (C) The best soft threshold is set as
442 4. (D) The gene-module correlation heatmap. (E, F) Lasso coefficient regression track. (G) Risk score between group1 and group2. (H-
443 J) The risk score's effect on LUAD patients' OS(H), time-dependent ROC curve for the risk score(I) and the risk factor distribution
444 map(J) in the TCGA-LUAD cohort. (K-M) The risk score's effect on LUAD patients' OS(K), time-dependent ROC curve for the risk
445 score(L) and the risk factor distribution map(M) in the GSE36471 cohort. (N-P) The nomogram model(N) and its calibration(O) and
446 time-dependent ROC(P) curves.

447

448 **External validation of FGD2, 3, 5's expression pattern**

449 In our research on the six members of the FGD family, only FGD2, 3 and 5 showed high value, so
450 we further verified their expression patterns in LUAD. Like TCGA, the GSE19188 set showed that
451 FGD2, 3, 5's level was lower in LUAD compared with normal ones(Figure10 A), and they all
452 showed good diagnostic efficacy, especially FGD5(Figure10 B). This is also true in GSE32863, a
453 sequencing set targeting LUAD tissues and adjacent paracancer tissues(Figure10 C). Representative
454 IHC images from the HPA database also illustrated that as compared to normal tissues, LUAD
455 tissues contained a notably lower level of FGD2(Figure10 D), FGD3(Figure10 E), and
456 FGD5(Figure10 F).

457 Then, for FGD2, 3, 5, FGD2 shows the highest value: it can predict the efficacy of ICB treatment.
458 Thus, we select FGD2 as the representative gene among FGD2, FGD3, and FGD5, and performed
459 IHC to validate its expression. The result showed that FGD2 is mainly expressed in the cytoplasm,
460 and paracancerous tissues expressed significantly higher levels of FGD2 than LUAD tissues, which
461 is in accordance with our bioinformatics analysis(Figure10 G).

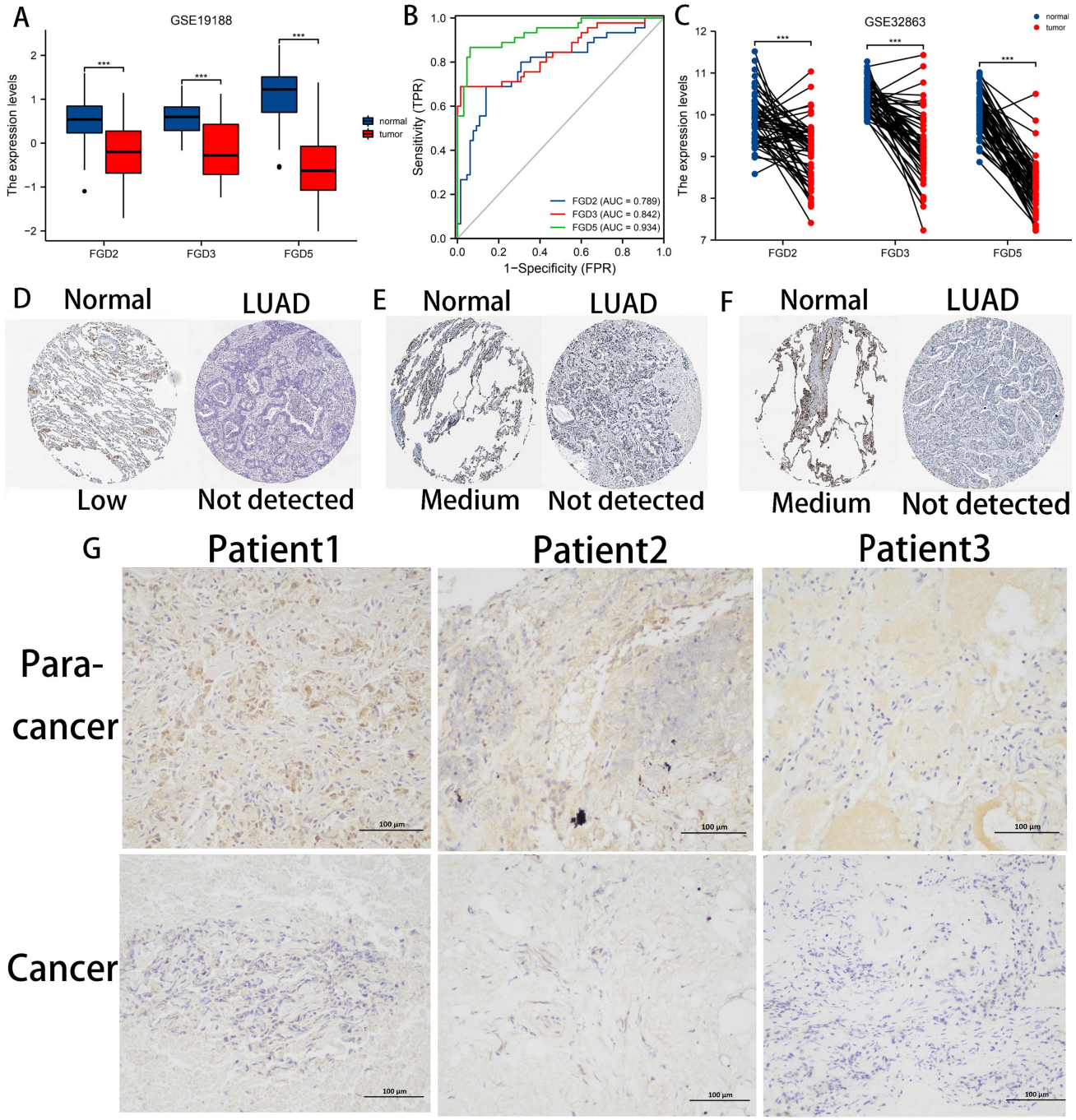


Figure 10. External validation of the expression pattern of three key FGDs. (A, C) FGD2, 3, 5's expression in GSE19188(A) and GSE32863(C) sets. (B) FGD2, 3, 5's diagnostic efficacy based on GSE19188. (D-F) Representative IHC image for FGD2(D), FGD3(E), and FGD5(F) from the HPA database. (G) Representative IHC staining pattern for FGD2 in paracancerous tissues as well as in LUAD tissues.

469 **DISCUSSION**

470 The FGD gene family consists of six genes namely FGD1, FGD2, FGD3, FGD4, FGD5 and FGD6.
471 Each one plays a different role, but they are also interconnected. Aberrant expression of some
472 members of them may result in malignant transformation, since they correlate with a number of
473 important biological processes. Considering that the role of these six FGD genes in LUAD is poorly
474 studied, we conducted a systematic analysis of them.

475 In our study, we found that FGD2, 3, 4, 5 expressed lower while FGD1, 6 expressed higher in
476 LUAD than in normal ones in the TCGA database. TCGA and KMplotter confirmed that high FGD2,
477 3, 5 level confers favourable prognosis of LUAD patients and using multivariate Cox regression
478 analysis, FGD2, 3, 5 showed independent prognostic significance. After primary therapy, PR&CR
479 patients' FGD2, 3, 5 levels were higher than those with PD&SD outcomes. Besides, advanced T, N,
480 pathological stages, and lethal OS events contributed to lower FGD2, 3, 5 levels. Thus a possible
481 protective effect of FGD2, 3, 5 on LUAD is suggested by these results. Interestingly, FGD1, 4, 6
482 don't correlate with these clinical characteristics and they don't have significant prognostic value.
483 Further, we examined the relationship between FGD1-6 and immune cells in LUAD using ssGSEA,
484 MCPcounter, and estimate methods, and found that FGD2, 3, 5 was positively correlated with most
485 types of immune cells and immune-related scores. Besides, we found that FGD2, 3, 5 are positively
486 correlated with the TCR/BCR richness and shannon diversity index, which means that when patients
487 expressed more FGD2, 3, 5, their T/B cells and their immune systems are more active(36); and we
488 unraveled FGD2, 3, 5's negative relationship with two malignant phenotypes, further suggesting their
489 protective role in LUAD. Using GEPIA and TIMER, we also found that FGD2, 3, 5 are positively
490 correlated with 49 immune cells' markers, 39 chemokines/receptors and 14 MHCs. However, FGD1,
491 4, 6's relationship with these immune characteristics are weak. Since the TIME plays a key role in
492 the carcinogenesis of assorted cancers, the significant correlation between FGD2, 3, 5 and these
493 TIME modulators suggests that they may be involved in the regulation of the TIME. We also

494 conducted single-cell sequencing analysis and pseudo-time analysis, and found that FGD2, 3, 5 was
495 mainly expressed in cells which are tightly correlated with immunity.
496 Recent advances in immunotherapy have made it an attractive treatment option for cancer,
497 particularly anti-PD1/PDL1 immunotherapy(37, 38). Several newly discovered immune checkpoints,
498 including LAG3, CTLA4, TIGIT, IDO, and BTLA, have been tested in clinical trials[(39, 40).
499 Besides, a large number of studies have shown that cancer patients with immune infiltrates have a
500 better ICB therapeutic effect than those without(41). In NSCLC, the latest guidelines from the NCCN
501 still point out that the level of PD-L1 expression determines the effectiveness of ICB therapy should
502 be applied (42). In spite of this, numerous clinical studies have shown that PD-L1 level does not
503 accurately predict the effectiveness of immunotherapy(43). Since the ICB therapy was only effective
504 in a minority of patients with LUAD, it is therefore essential to enhance tumor cells' response to
505 immune checkpoint inhibitors. The present study showed a strong positive correlation between FGD2,
506 3, 5 and PD1, PD-L1 and CTLA4. Initially, we hypothesized that each of FGD2, 3, 5 had a potential
507 function as a predictor of ICB efficacy, and GSEA analysis indicated that all of FGD2, 3, 5 are
508 correlated with the PD-1 checkpoint pathway in cancer. However, after prediction by TIDE algorithm,
509 only FGD2 showed the expected result: FGD2 high expression group had lower TIDE score. There
510 was no significant difference between TIDE scores for groups with high and low FGD3 expression,
511 whereas a higher TIDE score was obtained by the group with high expression of FGD5 than by the
512 group with low expression of FGD5. So to further confirm FGD2's value, we investigated the
513 ICBatlas and found that NSCLC patients with higher expression of FGD2 are more sensitive to anti-
514 PD1/PDL1/CTLA4 treatment. Then based on the TCIA database, we compared low and high FGD2
515 expression groups for IPS and found that the high FGD2 group had a higher IPS, indicating that
516 LUAD patients with high FGD2 were more responsive to ICB treatment. Besides, in the KMplotter
517 and ROCplotter databases, we comprehensively compared the expression level of FGD2 in patients
518 with other types of cancer who received immunotherapy, and found that FGD2 in responders was

519 significant higher expressed than non-responders, and high level of FGD2 could improve the
520 prognosis of cancer patients who received anti-PD1/PDL1/CTLA4 therapy. In addition, the TIDE
521 prediction algorithm and the ROCplotter database both indicated that FGD2 had better predictive
522 performance than PD-L1. This suggests that FGD2 is a good predictor of ICB efficacy, not only for
523 NSCLC patients, but also for other types of patients. However, since FGD2 was not significantly
524 better than PD-L1 in distinguishing ICB-treated responders from non-responders, whether it can be a
525 predictive marker independent to PD-L1 still needs further verification.

526 We also identified 18 genes that were co-expressed with FGD2, FGD3, and FGD5, and found that
527 their expression was decreased in LUAD as compared to normal tissues. Interestingly, these 18 genes
528 also positively regulated immune infiltration in LUAD, and GO analysis also showed that these
529 genes were mainly associated with immune activity. This further confirmed that FGD2, 3, 5 are
530 TIME modulators and these 18 genes may also be the regulators of the TIME. In addition, among the
531 18 genes, 5 genes including DOCK2, FAM78A, GIMAP1, DOK2 and ARHGEF6 had prognostic
532 value in both GEPIA and KMplotter databases. We found that except for DOK2, the role of the other
533 four genes in LUAD is still unclear, and FAM78A has been least studied so far. New insights into the
534 role of these four genes in LUAD may be provided by our study.

535 Since FGD2, 3, 5's prognostic value, their relationship with clinical characteristics and immunity
536 are significantly differed with FGD1, 4, 6, we systematically analyzed the six members of the FGD
537 family again. By consensus clustering, all patients were separated into two groups based on FGD
538 genes' level. Group1: high FGD2, 3, 5 and low FGD1, 4, 6. Group2: high FGD 1, 4, 6 and low FGD2,
539 3, 5. We found that patients in group 1 had better prognosis, higher immune cell infiltration, and
540 higher PDCD1, CD274 and CTLA4 level. This confirmed the strong link between FGD2, FGD3 and
541 FGD5, and their prognostic and predictive value in lung adenocarcinoma again, and unearthed the
542 heterogenity of the FGD family. In our study, FGD1, FGD4 and FGD6's role in the progression of
543 LUAD still remains unclear.

544 Finally, we obtained the DEGs between group1-2. GO_KEGG analysis indicated that they were
545 mainly participated in some immune process and immune-related signaling. After WGCNA's
546 selection, we obtained 116 DEGs considered most relevant to immunity. And then after lasso cox
547 regression, a prognostic model, consisted with 5 genes tightly related to immunity, including INHA,
548 IGF2BP1, HS3ST2, SLC14A2, and BTK, was constructed. One training cohort accompanied by 4
549 testing cohorts confirmed its good predictive performance, and cox analysis illustrated the riskscore
550 is an independent prognostic factor. A nomogram model constructed by riskscore and T, N, Stage
551 showed that combined with clinical features, the riskscore had better predictive power.

552 Our study is the first to systematically analyze the role of FGD family in LUAD. In this study, the
553 value of FGD2, FGD3 and FGD5, and the heterogenity of FGD1, 4, 6 and FGD2, 3, 5 in LUAD was
554 revealed for the first time. FGD2, 3, 5 all can be used as independent prognostic factors for LUAD,
555 they all positively regulate the immune cell infiltration in LUAD and FGD2 may be used to predict
556 the effectiveness of ICB therapy. FGD1, 4, 6 exhibited little value in this study.

557 However, our study has some limitations. The cell function and animal model experiments were
558 not conducted, and the molecular mechanism of FGD2, FGD3 and FGD5 in LUAD immunity is still
559 unclear. Our future experiments aim to validate these results and to do further research on identifying
560 the molecular mechanism of FGD2, FGD3 and FGD5 in LUAD.

561 Declarations

562 **Competing interests**

563 All authors declare that they have no conflict of interest.

564 **Ethical approval**

565 This study was approved by the Ethics Committee of the First Affiliated Hospital of Fujian Medical University. All
566 patients provided their informed consent.

567 **Author contributions**

FC Lai and F He conceived, designed and supervised the study.LY Zhang ,FQ Yu collected, analyzed the data and wrote the manuscript. MH Guan, X Huang and NL Lin reviewed and revised the manuscript. All authors read and approved the final manuscript.

Funding

This research was supported by National Natural Science Foundation of China (No. 82204457) and Young and Middle-aged Key Personnel Training Project of Fujian Provincial Health Commission (No. 2021GGA028).

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/supplementary material.

*** Correspondence address:**

Corresponding author: **Fancai Lai[1*]**; **Fei He[2*]**

E-mail-address:Laifancai@fjmu.edu.cn; i.fei.he@fjmu.edu.cn

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA Cancer J Clin. 2019;69(1):7-34.
2. Molina JR, Yang P, Cassivi SD, Schild SE, Adjei AA. Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. Mayo Clin Proc. 2008;83(5):584-94.
3. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. CA Cancer J Clin. 2016;66(1):7-30.
4. Anagnostou VK, Brahmer JR. Cancer immunotherapy: a future paradigm shift in the treatment of non-small cell lung cancer. Clin Cancer Res. 2015;21(5):976-84.
5. Brahmer JR, Tykodi SS, Chow LQ, Hwu WJ, Topalian SL, Hwu P, et al. Safety and activity of anti-PD-L1 antibody in patients with advanced cancer. N Engl J Med. 2012;366(26):2455-65.
6. Pasteris NG, Cadle A, Logie LJ, Porteous ME, Schwartz CE, Stevenson RE, et al. Isolation and characterization of the faciogenital dysplasia (Aarskog-Scott syndrome) gene: a putative Rho/Rac guanine nucleotide exchange factor. Cell. 1994;79(4):669-78.
7. Hou P, Estrada L, Kinley AW, Parsons JT, Vojtek AB, Gorski JL. Fgd1, the Cdc42 GEF responsible for Faciogenital Dysplasia, directly interacts with cortactin and mAbp1 to modulate cell shape. Hum Mol Genet. 2003;12(16):1981-93.
8. Zheng Y, Fischer DJ, Santos MF, Tigyi G, Pasteris NG, Gorski JL, et al. The faciogenital dysplasia gene product FGD1 functions as a Cdc42Hs-specific guanine-nucleotide exchange factor. J Biol Chem. 1996;271(52):33169-72.

9. Guo F, Cheng X, Jing B, Wu H, Jin X. FGD3 binds with HSF4 to suppress p65 expression and inhibit pancreatic cancer progression. Oncogene. 2022;41(6):838-51.
10. Susini T, Renda I. FGD3 Gene as a New Prognostic Factor in Breast Cancer. Anticancer Res. 2020;40(7):3645-9.
11. Ma C, Li H, Li X, Lu S, He J. The prognostic value of faciogenital dysplasias as biomarkers in head and neck squamous cell carcinoma. Biomark Med. 2019;13(16):1399-415.
12. Bird TD. Charcot-Marie-Tooth Hereditary Neuropathy Overview. In: Adam MP, Everman DB, Mirzaa GM, Pagon RA, Wallace SE, Bean LJH, et al., editors. GeneReviews(®). Seattle (WA): University of Washington, Seattle Copyright © 1993-2022, University of Washington, Seattle. GeneReviews is a registered trademark of the University of Washington, Seattle. All rights reserved.; 1993.
13. Bossan A, Ottman R, Andl T, Hasan MF, Mahajan N, Coppola D, et al. Expression of FGD4 positively correlates with the aggressive phenotype of prostate cancer. BMC Cancer. 2018;18(1):1257.
14. Ono Y, Nakanishi H, Nishimura M, Kakizaki M, Takahashi K, Miyahara M, et al. Two actions of frabin: direct activation of Cdc42 and indirect activation of Rac. Oncogene. 2000;19(27):3050-8.
15. Pasteris NG, Nagata K, Hall A, Gorski JL. Isolation, characterization, and mapping of the mouse Fgd3 gene, a new Faciogenital Dysplasia (FGD1; Aarskog Syndrome) gene homologue. Gene. 2000;242(1-2):237-47.
16. Ning X, Li R, Zhang B, Wang Y, Zhou Z, Ji Z, et al. Immune Score Indicator for the Survival of Melanoma Patients Based on Tumor Microenvironment. Int J Gen Med. 2021;14:10397-416.
17. Zhao S, Jin X, Xu S. [Expression of RASGRP2 in Lung Adenocarcinoma and Its Effect on Immune Microenvironment]. Zhongguo Fei Ai Za Zhi. 2021;24(6):404-11.
18. Luan M, Tian X, Zhang D, Sun X, Jiang M, Duan Y, et al. Identifying the potential regulators of neutrophils recruitment in hepatocellular carcinoma using bioinformatics method. Transl Cancer Res. 2021;10(2):724-37.
19. Zhang YF, Li Q, Huang PQ, Su T, Jiang SH, Hu LP, et al. A low amino acid environment promotes cell macropinocytosis through the YY1-FGD6 axis in Ras-mutant pancreatic ductal adenocarcinoma. Oncogene. 2022;41(8):1203-15.
20. Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarthi B, et al. UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. Neoplasia. 2017;19(8):649-58.
21. Liu CJ, Hu FF, Xia MX, Han L, Zhang Q, Guo AY. GSCALite: a web server for gene set cancer analysis. Bioinformatics. 2018;34(21):3771-2.
22. Shen W, Song Z, Zhong X, Huang M, Shen D, Gao P, et al. Sangerbox: A comprehensive, interaction-friendly clinical bioinformatics analysis platform. iMeta. 2022;1(3):e36.
23. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res. 2017;45(W1):W98-w102.
24. Györffy B, Lanczky A, Eklund AC, Denkert C, Budczies J, Li Q, et al. An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. Breast Cancer Res Treat. 2010;123(3):725-31.
25. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun. 2013;4:2612.
26. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics. 2013;14:7.
27. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. Genome Biol. 2016;17(1):218.
28. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, et al. TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. Cancer Res. 2017;77(21):e108-e10.
29. Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. Nat Med. 2018;24(10):1550-8.
30. Hugo W, Zaretsky JM, Sun L, Song C, Moreno BH, Hu-Lieskovan S, et al. Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma. Cell. 2016;165(1):35-44.
31. Yang M, Miao YR, Xie GY, Luo M, Hu H, Kwok HF, et al. ICBatlas: A comprehensive resource for depicting immune checkpoint blockade therapy characteristics from transcriptome profiles. Cancer Immunol Res. 2022.
32. Fekete JT, Győrffy B. ROCplot.org: Validating predictive biomarkers of chemotherapy/hormonal therapy/anti-HER2 therapy using transcriptomic data of 3,104 breast cancer patients. Int J Cancer. 2019;145(11):3140-51.
33. Chen Z, Luo Z, Zhang D, Li H, Liu X, Zhu K, et al. TIGER: A Web Portal of Tumor Immunotherapy Gene Expression Resource. Genomics Proteomics Bioinformatics. 2022.
34. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res. 2010;38(Web Server issue):W214-20.

657 35. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene
658 clusters. Omics. 2012;16(5):284-7.
659 36. Liu X, Wu J. History, applications, and challenges of immune repertoire research. Cell Biol Toxicol.
660 2018;34(6):441-57.
661 37. Chuan J, He S, Xie T, Wang G, Yang Z. Characterization of guanine nucleotide exchange activity of DH domain
662 of human FGD2. Protein Expr Purif. 2020;176:105693.
663 38. Wang C, Ma C, Gong L, Guo Y, Fu K, Zhang Y, et al. Macrophage Polarization and Its Role in Liver Disease.
664 Front Immunol. 2021;12:803037.
665 39. De Giglio A, Di Federico A, Nuvola G, Deiana C, Gelsomino F. The Landscape of Immunotherapy in Advanced
666 NSCLC: Driving Beyond PD-1/PD-L1 Inhibitors (CTLA-4, LAG3, IDO, OX40, TIGIT, Vaccines). Curr Oncol Rep.
667 2021;23(11):126.
668 40. Johnson DB, Rioth MJ, Horn L. Immune checkpoint inhibitors in NSCLC. Curr Treat Options Oncol.
669 2014;15(4):658-69.
670 41. Li X, Xu Z, Cui G, Yu L, Zhang X. BTLA Expression in Stage I-III Non-Small-Cell Lung Cancer and Its
671 Correlation with PD-1/PD-L1 and Clinical Outcomes. Onco Targets Ther. 2020;13:215-24.
672 42. Ettinger DS, Wood DE, Aisner DL, Akerley W, Bauman JR, Bharat A, et al. NCCN Guidelines Insights: Non-
673 Small Cell Lung Cancer, Version 2.2021. J Natl Compr Canc Netw. 2021;19(3):254-66.
674 43. Hong L, Negrao MV, Dibaj SS, Chen R, Reuben A, Bohac JM, et al. Programmed Death-Ligand 1
675 Heterogeneity and Its Impact on Benefit From Immune Checkpoint Inhibitors in NSCLC. J Thorac Oncol.
676 2020;15(9):1449-59.

677

678

679

680

681

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementary.pdf