Starting from the influence of tobacco: AKR1C3 and COVID-19 receptor ACE2 are potential prognostic biomarkers for brain lower grade glioma, evidence from bioinformatics analyses

Heng-Yi Yang[1]*, Tian-Ni Mao[2]

[1] School of Marine Engineering, Dalian Maritime University, Dalian, 116026, China.
[2] School of Humanities, Nanchang University, NanChang, 330000, China
* Correspondence: yhyatesss@foxmail.com

Abstract

Background: Coronavirus disease (COVID-19) related pneumonia is leaded by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), which infects host cells through receptors named Angiotensin-converting enzyme 2 (ACE2). Smoking is thought to be related to poor disease prognosis, as a large amount of evidence highlights the negative effects of smoking on lung health and its causal relationship with various respiratory diseases. Methods: We first evaluated the role of ACE2 expression level with tobacco effect. And correlation analyses were used to identify the related genes of ACE2 both in smoker trait and human normal multi-tissues. After intersections, hub genes were later used to further GSEA and co-regulated mechanisms were explored by GeneMANIA. We used UALCAN to perform survival curves of pan-cancers included 33 types of cancers. New clinical model of top co-occurrence cancer type was constructed and validated. Results: Effected by cigarette smoking, the expression level of ACE2 expressed statistically upwards in current smokers compared with never smokers, upwards in groups after acute smoke exposed compared with normal control. But no strong evidence detected in third-hand smoke, as poor amounts of samples, only 4. A little trend of expressing upwards became in groups after third-hand smoke compared with groups after clean air exposing. 4 genes included PIR, ADH7, AKR1C2 and AKR1C3 were identified as ACE2 related genes in smoker trait and human normal multi-tissues. Then, we made the survival curves of pan-cancers in 4 genes. Brain lower grade glioma was the co-occurrence type as both ACE2, PIR and AKR1C3 had the significantly prognostic situation. Later, we made the new clinical prediction model as the C-index was 0.827 and the Area Under Curves (AUCs) of 1-year survival, 2-year survival and 3-year survival were 0.921(95%CI, 0.882-0.961), 0.911(95%CI, 0.860-0.962) and 0.878(95%CI, 0.818-0.939). In inner validation, the AUCs of 1-year survival, 2-year survival and 3-year survival were 0.777(95%CI, 0.640-0.914), 0.916(95%CI, 0.842-0.990) and 0.888(95%CI, 0.768-1.000). In outer validation one, the AUCs of 1-year survival, 2-year survival and 3-year survival were 0.764(95%CI, 0.672-0.856), 0.848(95%CI, 0.783-0.913) and 0.748(95%CI, 0.656-0.841). And in outer validation two the AUCs of 1-year survival, 2-year survival and 3-year survival were 0.740(95%CI, 0.654-0.826), 0.766(95%CI, 0.688-0.843) and 0.794(95%CI, 0.721-0.866). Conclusions: In our study, we compared the ACE2 expression level between smokers and non-smokers crowds and identified associated hub genes. Then we explored the further role of hub genes in tumor fields to identify the correlation and influence between the avoidable host factors such as smoking on COVID-19 contamination and tumor. ACE2 and AKR1C3 are potential prognostic genes for brain lower grade glioma, and we created the web dynamic nomogram and encapsulation app through validations.
Keywords: ACE2; COVID-19; Brain lower grade glioma; AKR1C3

Introduction

Coronavirus disease (COVID-19) related pneumonia is leaded by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), which infects host cells through receptors named Angiotensin-converting enzyme 2 (ACE2) [1]. Since 31 December 2019 and as of 11 July 2020, 12476028 cases of COVID-19 have been reported, including 559l998 deaths[2]. The main entrance way of virus is through mucosal tissues includes nose, mouth and upper respiratory tract. Researchers had verified that lung alveolar epithelial cells

are infected by receptor-mediated endocytosis of SARS-CoV-2, and through an entry receptor named Angiotensin-Converting Enzyme II (ACE2)[3]. Due to lung alveolar cells injury, the onset of disease may lead to progressive respiratory failure and even death. To date, smoking is thought to be related to poor disease prognosis, as a large amount of evidence highlights the negative effects of smoking on lung health and its causal relationship with various respiratory diseases[4]. Tobacco kills more than 8 million people every year and more than 7 million of these deaths were the result of direct smoking, while approximately 1.2 million people were the result of non-smokers' exposure to second-hand smoke[5]. Tobacco smoke exposure leads to inflammatory processes in the lung, and increased permeability in epithelial cells, mucus overproduction, and impaired mucociliary clearance, as the results of increasing mucosal inflammation[6]. Previous studies have shown that smokers were also noted to have higher mortality in the previous MERS-CoV outbreak[7]. In many countries, secondhand smoke exposure is very common, and in 2004 worldwide, 40% of children, non-smokers in 33% of male and 35% of female were exposed to second-hand smoke[8]. Second-Hand Smoke (SHS) defined as the tobacco combustion products inhaled by nonsmokers near burning tobacco, contains high concentrations of carcinogens and irritants such as nicotine, benzene and fine particles (PM2.5). For SHS, researchers identified that SHS significantly increased tumor size, weight and blood vessels[9]. Third-Hand Smoke (THS) is a subset of chemicals in SHS such as nicotine that sticks to indoor surfaces and can be reemitted into environment reacted with other compounds. In summary, smokers are more vulnerable to infectious diseases. Smoking may by an important host factor for reducing viral contamination[10]. In our study, we compared the ACE2 expression level between smokers and non-smokers crowds and identified associated hub genes. Then we explored the further role of hub genes in tumor fields to identify the correlation and influence between the avoidable host factors such as smoking on COVID-19 contamination and tumor.

Materials and methods

Data resource and processing

Gene Expression Omnibus database (GEO) is the platform that store chips, microarrays data and so on, organized by NCBI (US National Center for Biotechnology Information)[11]. We download the human airway epithelial cell effected by cigarette smoke transcriptome data in GSE994 with 75 samples[12] and GSE7895 with 104 samples[13]. GSE97010 was downloaded as human airway epithelial cell transcriptome data effected by second-hand smoke with 126 samples[14]. GSE129959 was downloaded as human nasal epithelial cell effected by third-hand transcriptome data with 16 samples[15]. GSE19667[16] and GSE52237[17] datasets were downloaded as input data of weighted correlation network analysis. Both 6 GEO datasets were received the processing of null value deleting, probe information merging and selecting the highest expression value in gene overlap situation. Limma package was used to normalize between arrays[18]. The Cancer Genome Atlas (TCGA) stores clinical, copy number, mutation, protein, methylation and mRNA data of many types of cancers. We downloaded the clinical and mRNA-Seq FPKM data of Brain Lower Grade Glioma (LGG) with 510 cases, and were matched and combined. Chinese Glioma Genome Atlas (CGGA) platform provides basic and clinical research of gliomas data. We downloaded the mRNAseq_325 (batch 2)[19, 20] and mRNA-array_301 datasets[21] with 182 and 174 LGG cases and combined clinical and expression data. Gene expression RNA-seq recompute tpm data (n=7,862) and phenotype data (n=9,783) of GTEx and gene expression RNA-seq recompute tpm data (n=10,535) and curated clinical data (n=12,591) of TCGA Pan-Cancer were downloaded at UCSC Xena platform[22]. Gene expression and phenotype data were matched and combined. Normal cases in TCGA Pan-Cancer data were abandoned. R3.6.3, R4.0 and RStudio software were used to perform the processing work[23, 24].

Analysis of ACE2 different expression between smokers and non-smokers

For the effect of cigarette, we divided into 3 classification, included direct, second-hand and third-hand smokers. In GSE994 and GSE7895 datasets, 3 groups were compared as former smokers, never smokers and current smokers. In GSE97010 datasets, acute smoke exposed groups were treated as second hand smoke effect and compared with control groups. In GSE129959 datasets, four groups contained before clean air (CA), after clean air, before third hand smoke (THS) and after third hand smoke was used to display the difference

of ACE2 expression level. R software was used to perform Shapiro-Wilk normality test and Bartlett test of homogeneity of variances. GraphPad prism 8.0.2 software was used to perform one-tailed t-test[25].

Weighted correlation network analysis and associated module identification of smokers

WGCNA package was used to construct 2 weighted correlation networks of GSE19667 and GSE52237[26, 27]. To check the sample quality, function of goodSamplesGenes was used and gplots package was employed to create heatmap[28]. Function of hclust made the sample dendrogram to eliminate the outliers. Then, pickSoftThreshold function calculated the average connectivity degrees of different modules and their independence in several powers, and assisted to recognize the best soft thresholding power β-value when the degree of independence reached 0.9. Function of blockwiseModules constructed the co-expression network and detect modules. with mergeCutHeight = 0.25 and minModuleSize = 20 in GSE19667 analysis and mergeCutHeight = 0.25 and minModuleSize = 200 in GSE52237. Later, phenotype traits were converted into numerical values and related to co-expression modules by regression analysis. Gene significance (GS) indicated the correlation between genes and clinical traits. And, module membership (MM) demonstrated the association between genes and specific modules. The genes in modules which showed the highest correlation with specific trait were all permit to further analysis. In our study, 4 modules in GSE19667 and GSE52237 were selected with 2 highest coefficient positive modules and 2 lowest coefficient negative modules in trait of smokers.

Functional enrichment analysis of correlation module genes of smokers

In GSE19667 and GSE52237, selected modules were combined into 2 groups included positive and negative. EnrichR is a platform for online enrichment analysis and was used to perform the Gene Ontology (GO) ontologies[29] and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways[30] analysis[31]. Term names, adjusted P-values and combined scores were the values used to construct the enrichment network. Combined score is computed by taking the log of the P-value from the Fisher exact test and multiplying that by the z-score of the deviation from the expected rank[32]. Cytoscape is a powerful software for visualizing network data[33]. Cytoscape yFiles Layout Algorithms made the layout of network[34].

ACE2 related hub genes identification and interaction analysis

We made the 2 intersections to identify the most convincing gens related ACE2 in overall situation. First, in GES19667 and GSE52237 datasets, Correlations between genes and ACE2 were calculated and P <0.05, absolute valued of coefficient >0.4 were the selection criterion. Later, 4 modules in highest correlation included positive and negative in smoker trait in GSE19667 and GSE52237 were intersected with the ACE2 correlation genes in smokers. Secondly, we identified the genes in one-way interaction included positive or negative with ACE2 in human normal multi-tissues of GTEx datasets. Finally, intersection of ACE2 correlation genes in multi-tissues and ACE2 correlation genes in trait of smokers was performed which screened the ACE2 related hub genes in overall application. The hub genes correlations plots in multi-tissues were performed by ggplot2[35], ggrepel[36] and pheatmap[37] packages. Then, GSE19667 and GSE52237 datasets were merged and we calculated the all genes correlation of hub genes as the traditional fold change in gene-set enrichment analysis (GSEA). Hallmark gene sets of MSigDB[38] was used to GSEA and performed by clusterProfiler package[39]. Finally, GeneMANIA was employed to predict the hub gens functions and interactions and construct the network with 20 related genes[40].

### *Related cancers screening by prognosis and clinical risk prediction model construction and validation*

In order to further explore the role or interaction of hub gens in tumors, UALCAN was used to perform the TCGA pan-cancer survival analysis of hub genes[41]. Survival curve status and P-value <0.05 were the main criteria in 33 types of cancers. We downloaded the 510 LGG patient's mRNA-Seq FPKM and clinical data and developed a new nomogram containing gene expression with clinical information to predict the 1-year, 2-year and 3-yaer survival of LGG patients. We separated the patient's information randomly in 8:2 as training and inner testing data. X-tile software calculated the cut-off points and divided the specific clinical factor into 2 or 3 groups [42]. After clinical factors values were converted into numerical values, R package name survival used the Cox proportional hazard regression model to estimate hazard ratio and its 95% confidence interval (CI) for each factors[43, 44]. In the multivariate Cox regression analysis, to measure the

goodness of fit of statistical models, R function named extractAIC indicated the Akaike information criterion (AIC) of candidate models[45]. And Concordance index (C-index) was used to validate the ability of distinguishing events and non-events correctly. Previous researches identified the age at diagnosis was the largest contributor to patient survival, followed by WHO grade within our variables[46, 47]. In our study, tumor grade and ages were determined the basic and key factors. During the model genes factors selecting, we used the correlation analysis between candidate genes in pan-cancers, LGG and related type glioblastoma multiforme (GBM) were especially focused to verify the choice. R package named timeROC created the receiver operating characteristic curve (ROC) curves to predict summarize the predictive capacity[48] and confint of R indicated the 95%CI value of Area Under Curve (AUC) of ROC. R packages named rms[49] and Hmisc[50] created the corresponding nomogram model. Later, 2 CGGA datasets were used to perform 2 outer validations. Risk scores of outer datasets were calculated by the established nomogram model and separated into 2 groups of low or high by median value. And finally, ROC of 1-year, 2-year and 3-year survival of outer validation were performed. To improve the convenient of nomogram, DynNom package [51] produced the web nomogram model and uploaded on Shinyapp platform[52]. We also encapsulated the web nomogram into an Android app named LGG Tracker[53].

Results

### *Analysis of ACE2 different expression between smokers and non-smokers*

Fig.1A-B indicated that in tobacco smoke, the ACE2 of current smokers significantly expressed upwards compared with never smokers, both in GSE994 and GSE7895. And the ACE2 relative expression between former smokers and never smokers, current smokers and former smokers did not show the significant difference. Later for second-hand smoke, the group after Acute Smoke Exposing (ASE) showed the significant upward expression difference (**Fig.1C**). For third-smoke, the ACE2 relative expression between group after Third-Hand Smoke (THS) and group after Clean Air (CA) showed little significant, as the amount of sample was poor with 4, but it still indicated the upward trend of ACE2 expression after THS treated (**Fig.1D**).
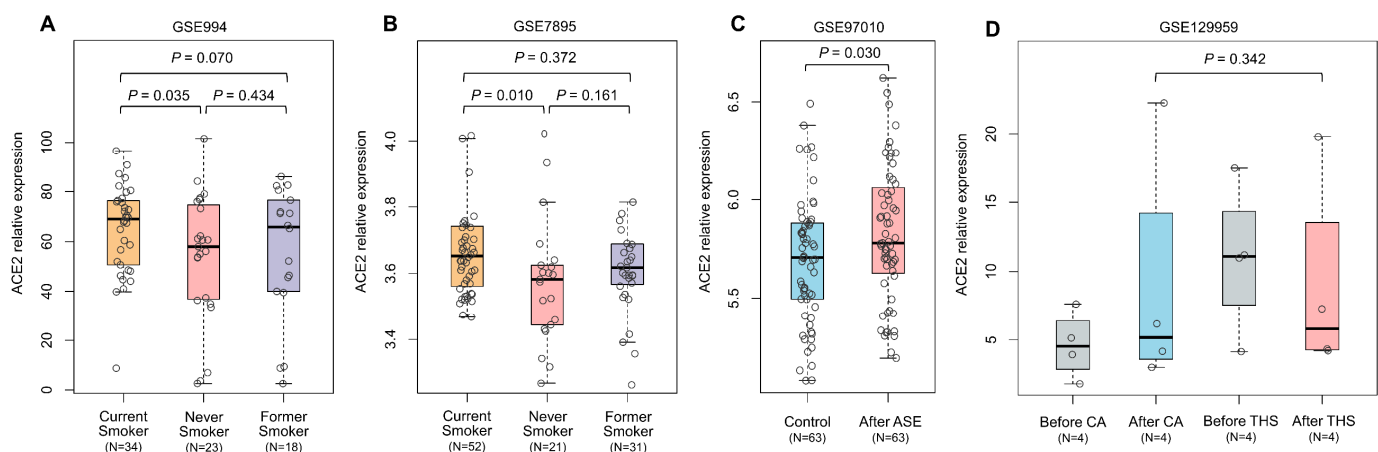

Fig.1 The ACE2 relative expression between different smoke influence, included tobacco smoke (A and B), second-hand smoke (C) and third-hand smoke (D). ASE, Acute Smoke Exposed; CA, Clean Air; THS, Third Hand Smoker

Weighted correlation network analysis and associated module identification of smokers

Fig.2A-B pointed that gene expression tidiness of both GSE19667 and GSE52237 origin data was crude. After normalization by Limma package, both gene expression tidiness showed good performance (**Fig.2C-D**). Then, 8 non-smokers (GSM469989, GSM434049, GSM469995, GSM350956, GSM350955, GSM298225, GSM350873, GSM458582) and 8 smokers (GSM298246, GSM470005, GSM252871, GSM298237, GSM254159, GSM434054, GSM298239, GSM254157) were eliminated in GSE19667. 3 smokers (GSM1074741, GSM1261681, GSM1074727) were abandoned in GSE52237, as showed in **Fig3**. Later, WGCNA goodSamplesGenes function results of both GSE19667 and GSE52237 were ok.
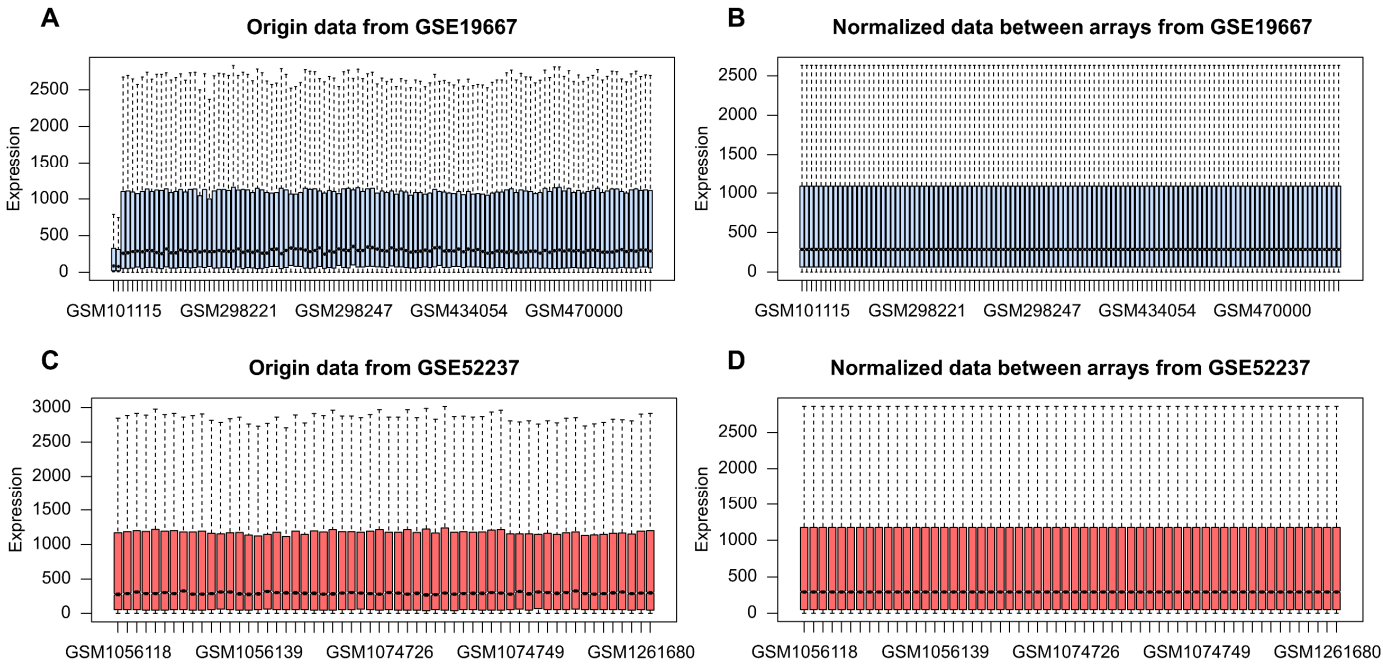
Fig.2 Origin data of GSE19667 (A) and GSE52237 (C) and Normalized data between arrays of GSE19667 (B) and GSE52237 (D)
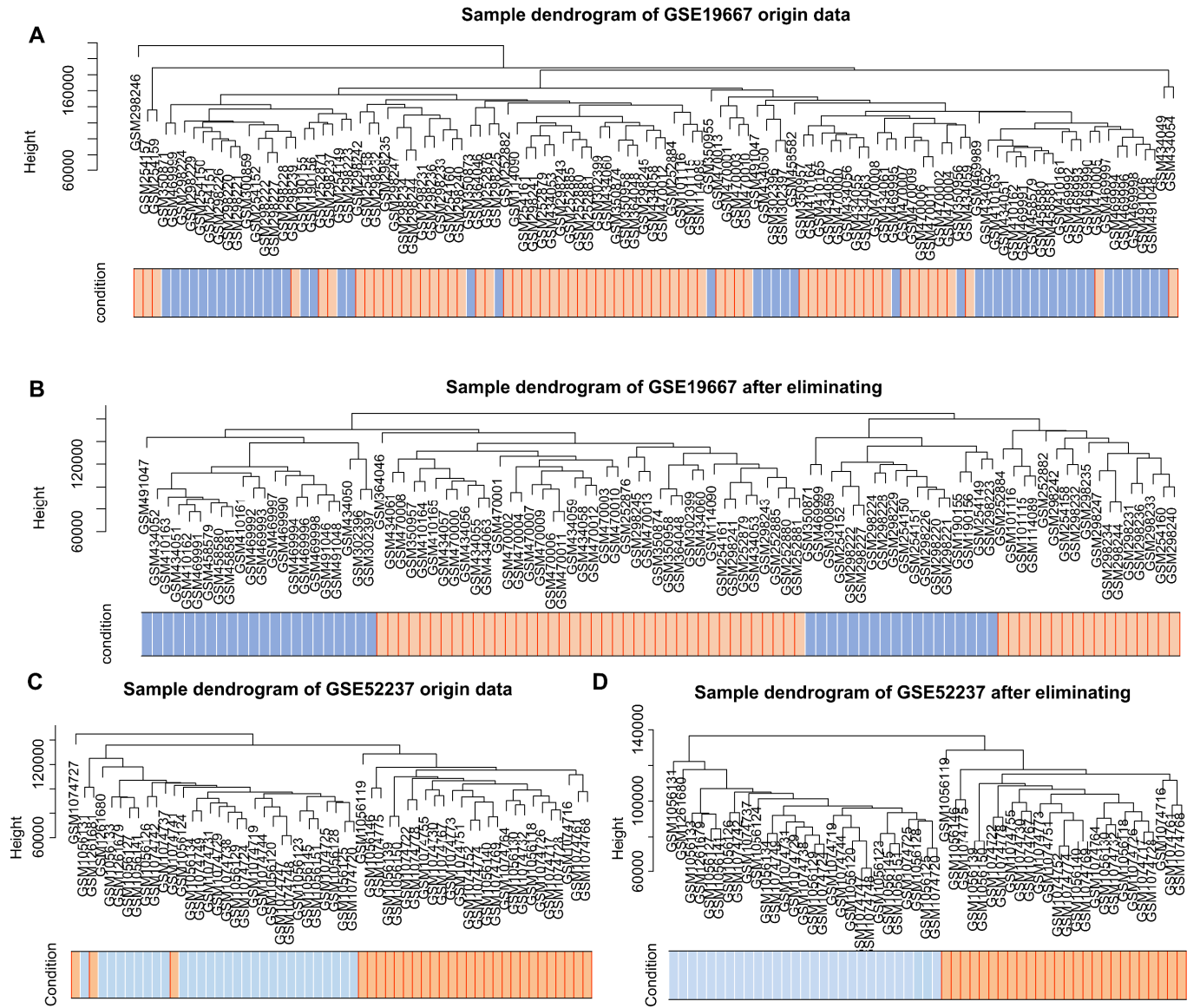

Fig.3 Sample dendrogram of GSE19667 and GSE52237 origin data (A and C) and corresponding pure data after outlier eliminated (B and D), blue and orange color corresponded non-smoker and smoker condition

Before WGCNA network were constructed, pickSoftThreshold function of WGCNA was used to calculate the soft thresholding power. As shown in **Fig.4A**, in GSE19667, the soft-thresholding power was set at 10 as the first time to reach the 0.9 scale independence and with a low mean connectivity. For GSE52237, 6 was selected to set soft-thresholding power for its initially reached 0.9 scale independence with a low mean connectivity (**Fig.4B**). Later, one-step network construction function of WGCNA was employed to identify gene modules. Parameter of minModuleSize was set as 30 and mergeCutHeight was set as 0.25 in GSE19667. In GSE52237, minModuleSize was set as 200 and mergeCutHeight was set as 0.25. Finally, 28 modules were Identified in GSE19667 (**Fig.5A**) and 19 modules were Identified in GSE52237 (**Fig.5B**).
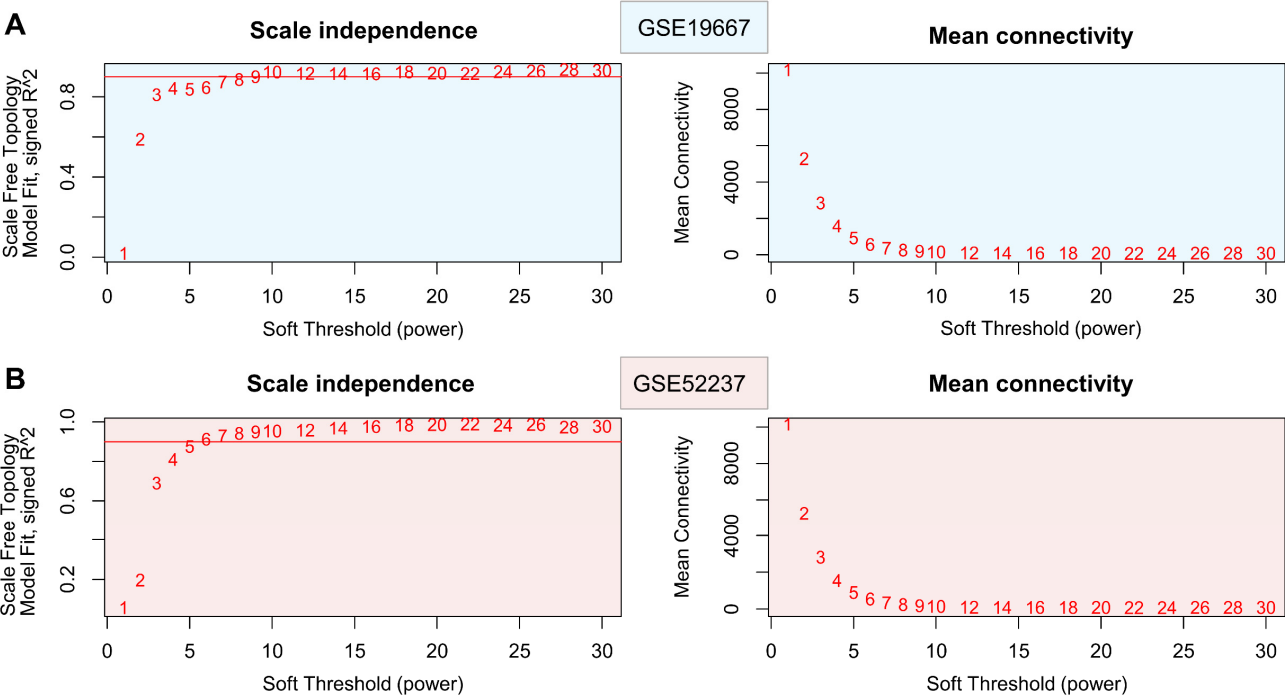
Fig.4 Scale independence and mean connectivity of various soft-thresholding values of GSE19667（A）and GSE52237 (B)
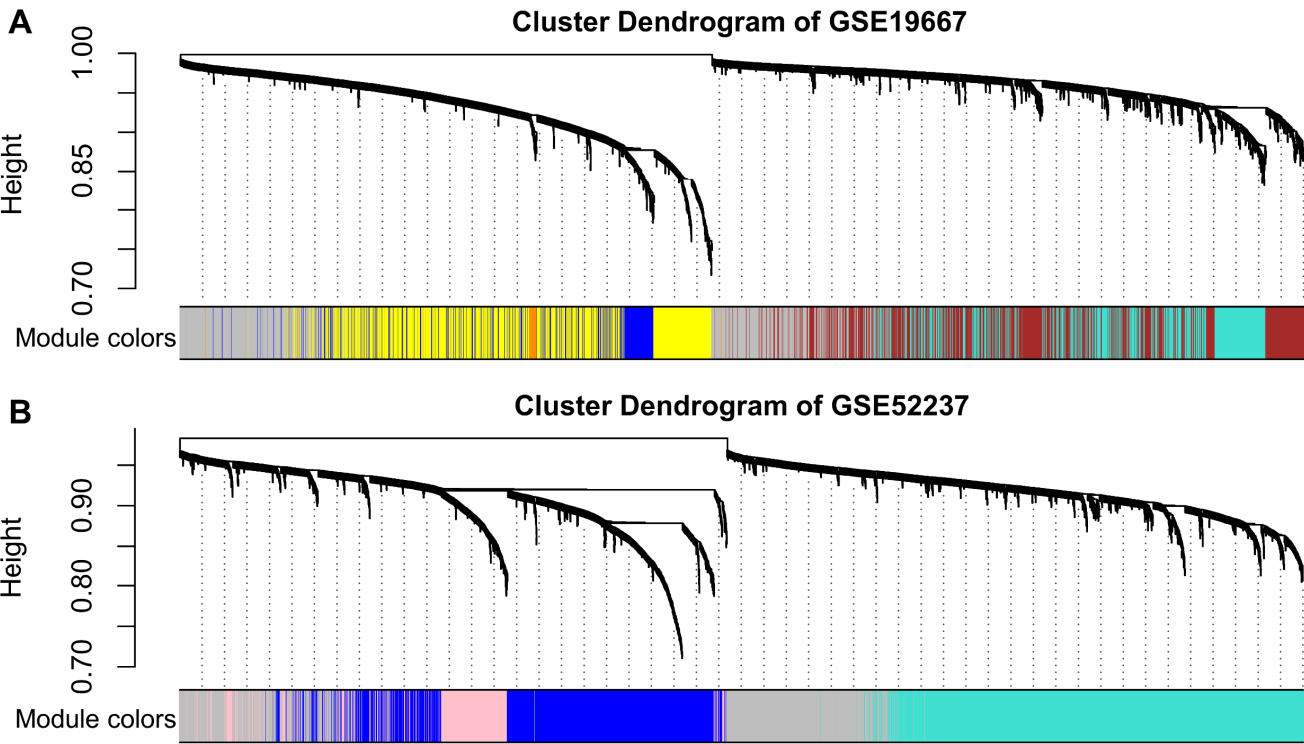
Fig.5 Clustering dendrogram of genes with dissimilarity based on topological overlap, together with assigned module colors. 28 modules in GSE19667 (A) and 19 modules in GSE52237 (B)

As shown in **Fig.6A**, In GSE19667, the magenta module showed the highest positive relationship with smoker group as correlation coefficient was 0.85 and pink module indicated the highest negative relationship with smoker group as -0.8 correlation coefficient. Blue and turquoise module correspond to the highest positive and negative relationship with smoker group in GSE52237 (**Fig. 6B**). Then, the four modules were used to further analyses. 1697 genes and 2330 genes consisted in the positive and negative of smoker which
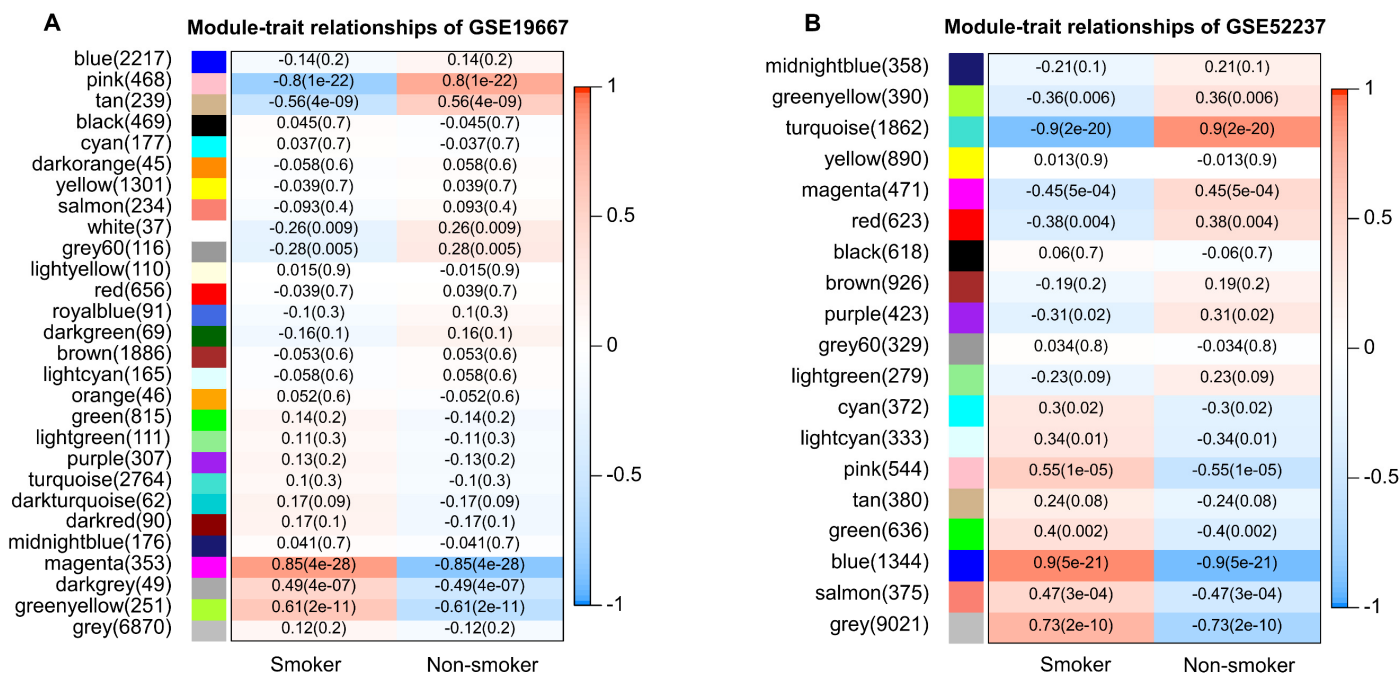
Fig.6 Module-trait relationships of GSE19667 (A) and GSE52237 (B). Each row shows a module eigengene also included the gene number, and each column corresponds to a clinical trait contains Smoker and Non-smoker. Each cell contains the corresponding correlation and P-value. The table is color-coded by correlation according to the color legend.
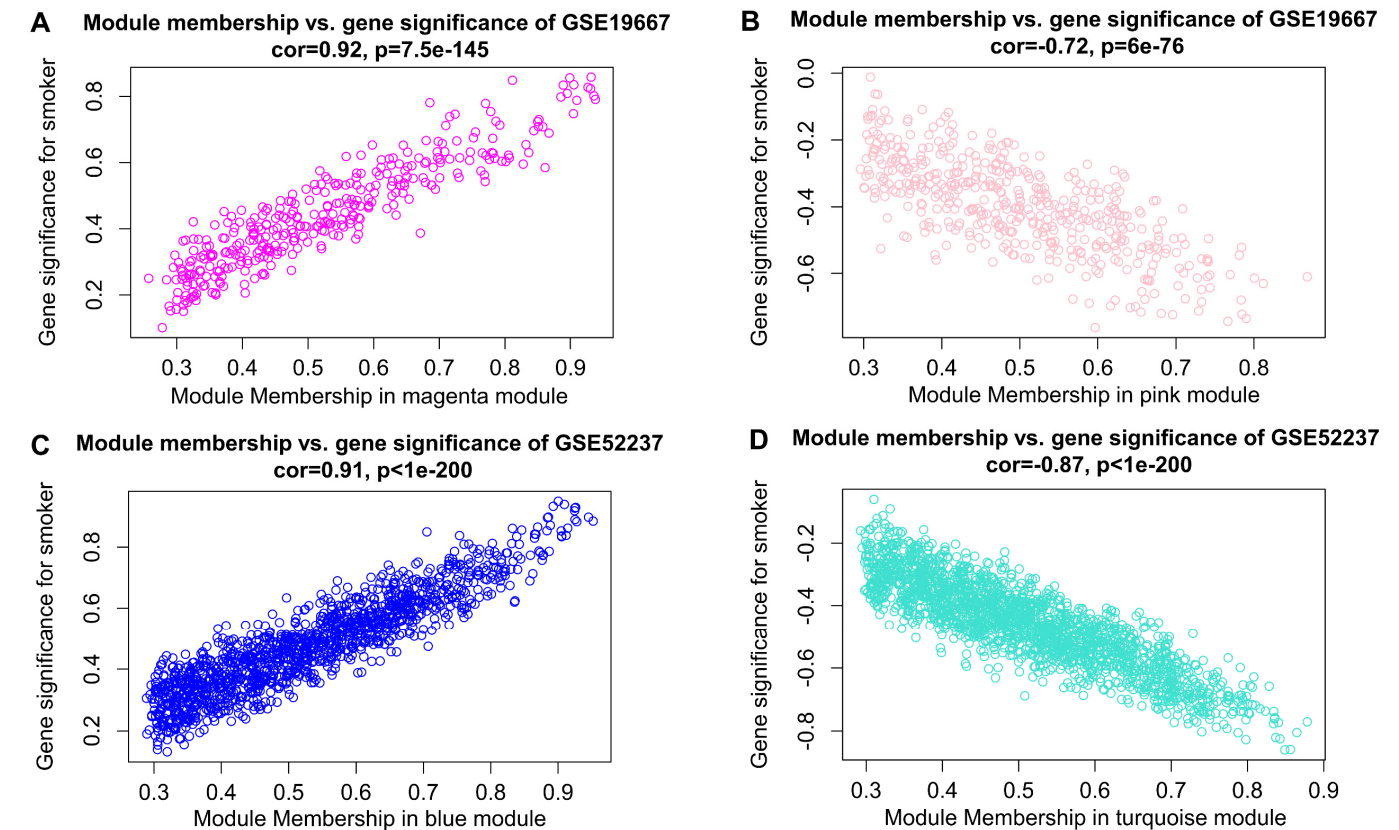
Fig.7 The scatter plot of 4 selected modules between module membership and gene significance for smoker. Magenta and pink module were in GSE19667 while blue and turquoise module were in GSE52237

may overlapped. P-values of both all selected modules <0.05. The correlations between module membership and the gene significance of the four modules included magenta (**Fig.7A**), pink (**Fig.7B**) in GSE19667 and blue (**Fig.7C**) and turquoise (**Fig.7D**) in GSE52237 were high relative. And, P-values of both 4 modules <0.05.

Functional enrichment analysis of correlation module genes of smokers

Evidence indicated that the four selected modules were convincing to identify the genes associated with smoker. Pink and turquoise modules were combined as positive correlation genes of smoker. Blue and magenta were combined as the negative correlation group. The results of EnrichR were reproduced by Cytoscape, as shown in **Fig.8A** of positive correlation and **Fig.8B** of negative correlation. In positive correlation of smoker, the GO enrichment indicated that aminoglycoside antibiotic metabolic process, doxorubicin metabolic process and daunorubicin metabolic process were enriched in biological process, pyrimidine nucleotide-sugar transmembrane transporter activity, protein-disulfide reductase activity and alcohol dehydrogenase (NADP+) activity were enriched in molecular function, Golgi subcompartment, Golgi membrane and vesicle coat were enriched in cellular component. KEGG enrichment indicated that the positive correlation enriched in Mucin type O-glycan biosynthesis, Protein export and Protein processing in endoplasmic reticulum. In negative correlation group of smokers, SRP-dependent co-translational protein targeting to membrane, co-translational protein targeting to membrane and viral gene expression were enriched in biological process, protein kinase C activity, small ribosomal subunit rRNA binding and leucine binding were enriched in molecular function, cytosolic ribosome, polysomal ribosome and ribosome were enriched in cellular component. KEGG enrichment indicated that the negative correlation enriched in Ribosome, Pantothenate and CoA biosynthesis and Sulfur metabolism.
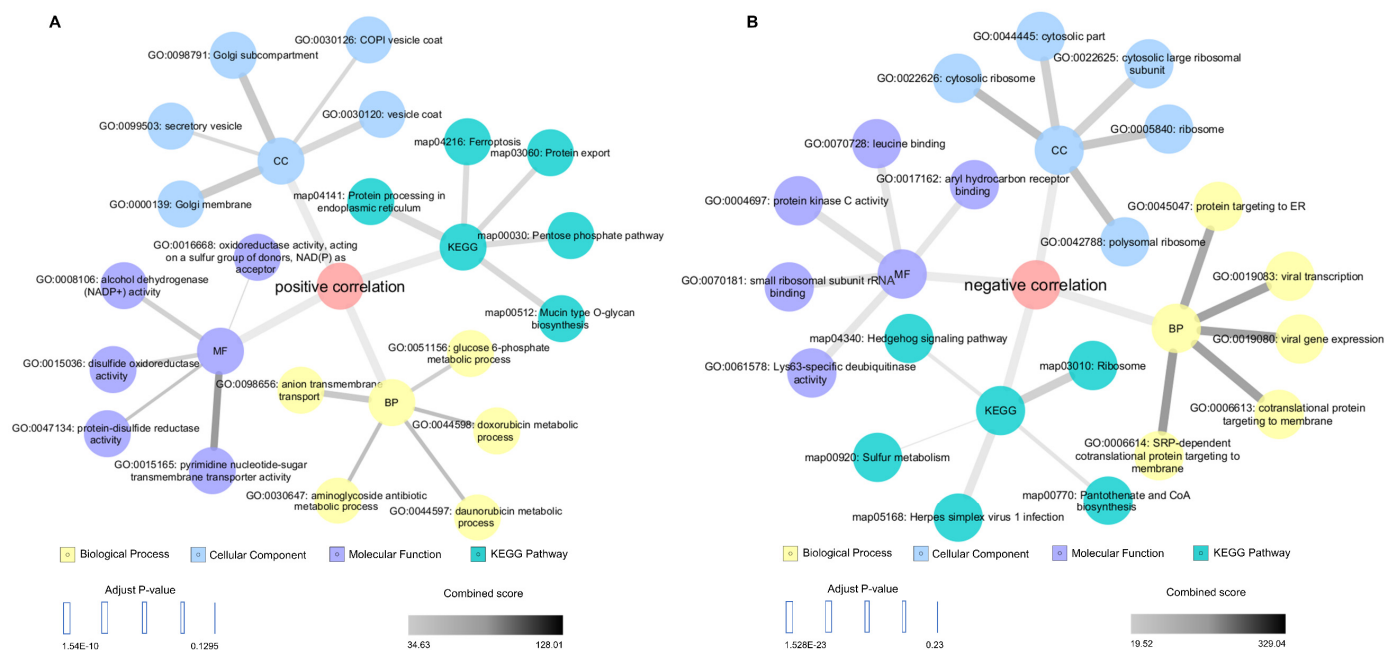

Fig.8 The GO and KEGG enrichment of combined modules divided positive (A) and negative (B) correlation group, achieved by Cytoscape.

*ACE2 related hub genes identification and interaction analysis*

After intersection between selected WGCNA modules genes in smoker group and ACE2 correlation genes in GSE19667 and GSE52237, as shown in **Fig.9A**, 36 genes were identified, included NDUFS1, PIR, MSMB, SLC6A6, CDC42EP3, FAM3B, TALDO1, CLDN10, SLC2A1, GSEC, MUC12-AS1, SRPX2, MUCL1, TXNRD1, PHEX, CABYR, AKR1C1, MIPOL1, ME1, DTNA, COA6, AKR1C2, ADH7, VPS13D, NMRAL2P, ATP6V0A4, TMEM45B, NQO1, HTATIP2, FRK, LINC01697, AKR1C3, TXN, PHTF2, SAR1B and SLC51A. Then GTEx database was used to identify the related genes about ACE2, from correlation evidence of multi-tissues. Then 528 genes in singe correlation direction that positive or negative with 31 human tissues were recognized. The 31 human tissues contained Adipose Tissue, Adrenal Gland, Bladder, Blood, Blood Vessel, Bone Marrow, Brain, Breast, Cervix Uteri, Colon, Esophagus, Fallopian Tube, Heart,

Kidney, Liver, Lung, Muscle, Nerve, Ovary, Pancreas, Pituitary, Prostate, Salivary Gland, Skin Small Intestine, Spleen, Stomach, Testis, Thyroid, Uterus and Vagina. Genes which owned the both correlation direction with human tissues were abandoned. We made the intersection between 528 ACE2 correlation genes in multi-tissues and ACE2 correlation genes in smoker after WGCNA analysis. Finally, 4 genes were identified, included PIR, AKR1C2, AKR1C3 and ADH7 (**Fig.9B**).
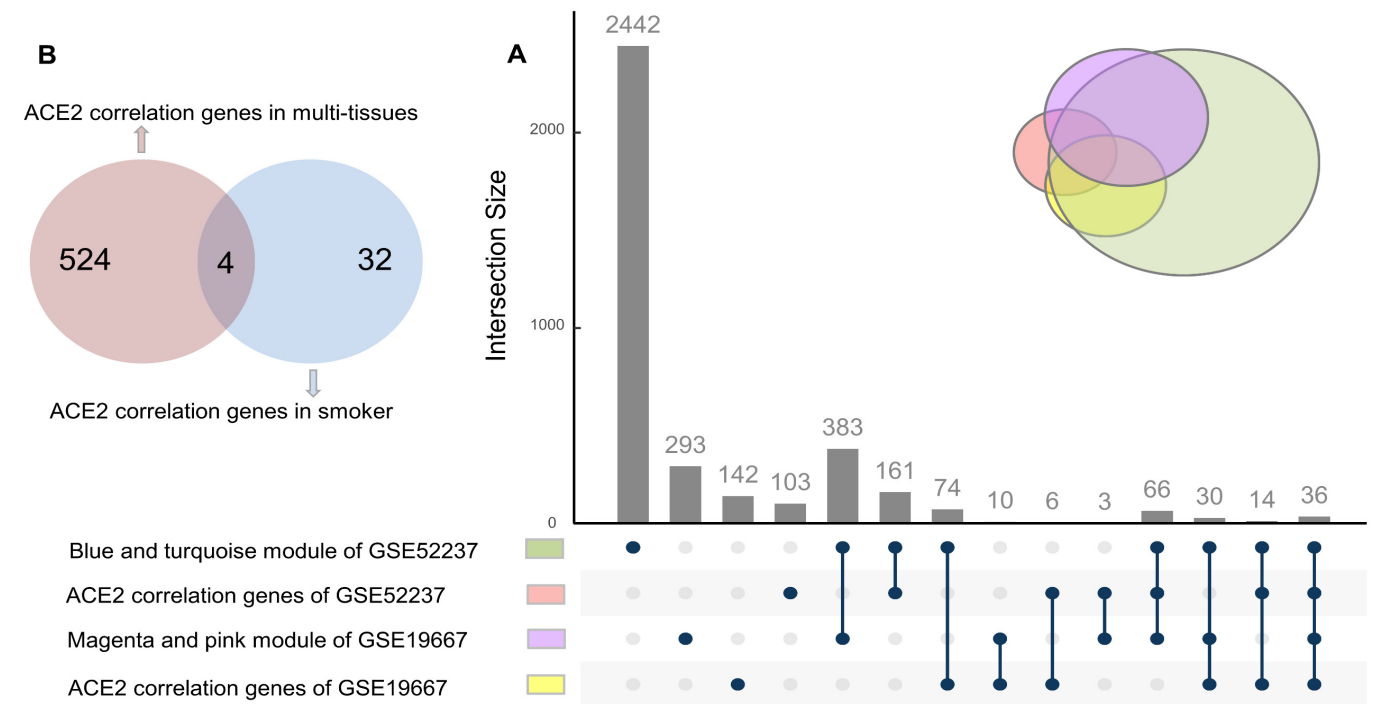

Fig.9 The upset plot between ACE2 correlation gens in smokers and WGCNA selected modules gens (A) and Venn plot between ACE2 correlation genes in multi-tissues of GTEx and ACE2 correlation genes in smoker by WGCNA analysis (B)

As shown in **Fig.10A**, both all 4 genes that PIR, AKR1C2, ADH7 and AKR1C3 indicated the positive correlation with ACE2. For the detail correlation coefficient, **Fig.10B** showed that in 31 human normal tissues, PIR and ACE2 had a 0.89 positive correlation in kidney, 0.80 in pancreas and 0.74 in stomach, AKR1C2 and ACE2 had a 0.92 positive correlation in fallopian tube, 0.68 in kidney and 0.66 in cervix uteri, ADH7 and ACE2 had a 0.69 positive correlation in cervix uteri, and 0.68 in bladder, 0.58 in vagina, AKR1C3 and ACE2 had a 0.91 positive correlation in kidney, 0.86 in small intestine and 0.84 in colon.
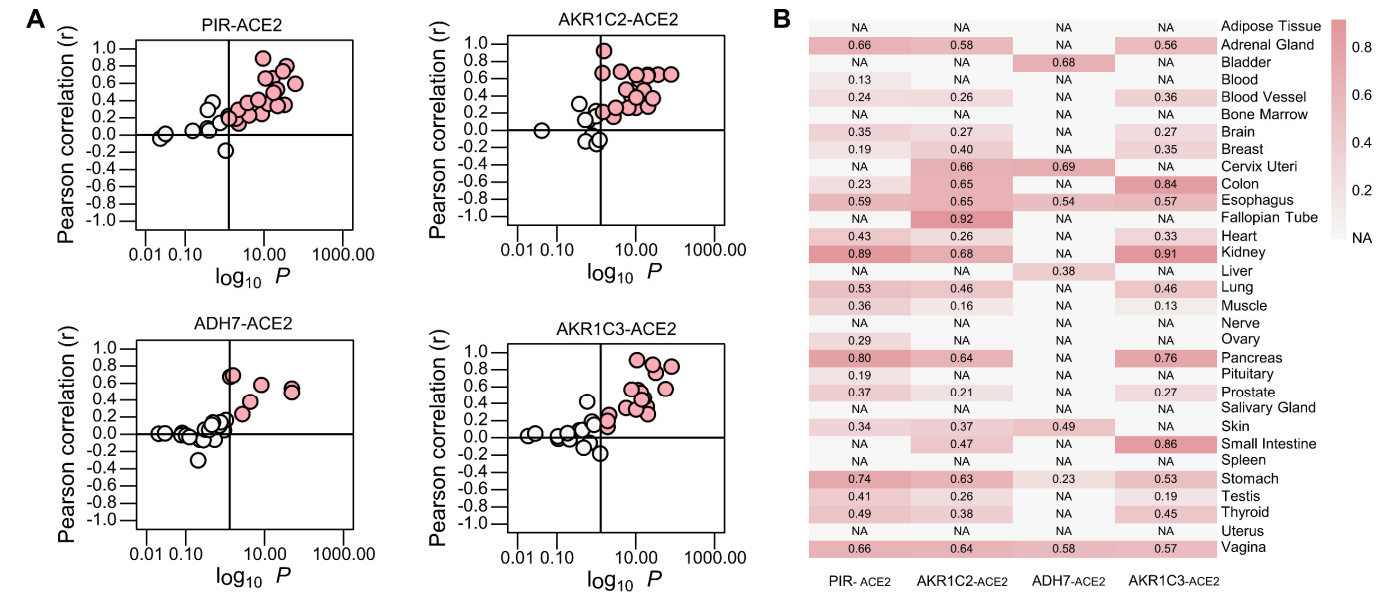

Fig.10 The correlation between ACE2 and 4 identified genes in normal multi-issues (A) and detail situation heatmap in 31 human normal tissues of 4 genes matched with ACE2 (B)
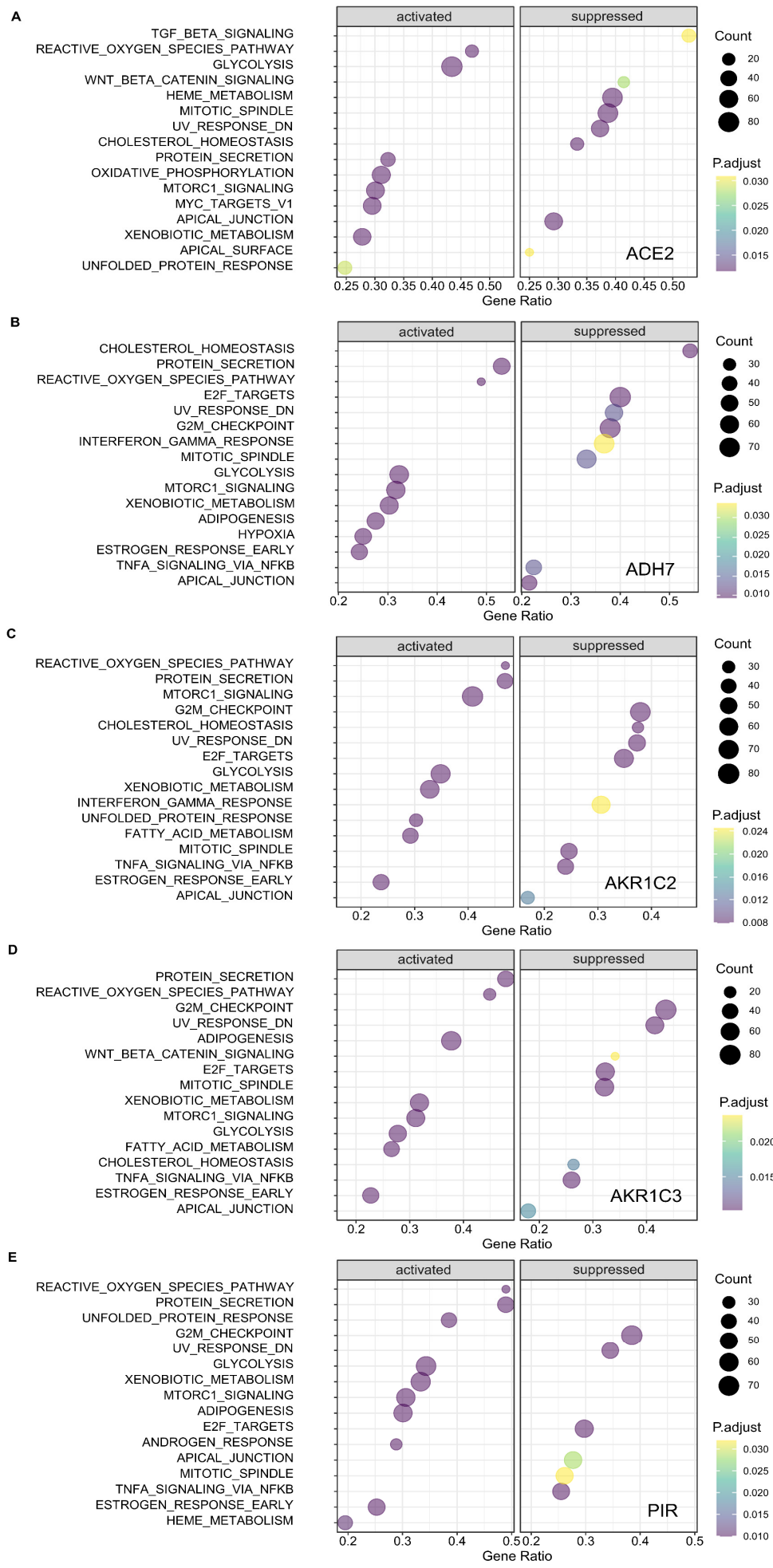
Fig.11 GSEA bubble plots of 5 genes of ACE2 (A), ADH7 (B), AKR1C2 (C), AKR1C3 (D) and PIR (E)

To explore the potential functions of five genes in smoker condition, we combined the GSE19667 and GSE52237 normalized data as the single gene GSEA input data. After correlations were calculated, functions of one gene were reversed by analyzing the most relevant gene in positive and negative direction, the correlation coefficients represented the log fold change in traditional GSEA analysis. The results indicated ACE2 suppressed in BETA SIGNALING, WNT BETA CATENIN SIGNALING and HEME METABOLISM, activated in REACTIVE OXYGEN SPECIES PATHWAY, GLYCOLYSIS and PROTEIN_SECRETION, as shown in **Fig.11A**. For ADH7 as shown in **Fig.11B**, PROTEIN SECRETION, REACTIVE OXYGEN SPECIES PATHWAY and GLYCOLYSIS were activated, CHOLESTEROL HOMEOSTASIS, E2F TARGETS and UV RESPONSE_DN were suppressed. The results of AKR1C2 as shown in **Fig.11C** indicated that REACTIVE OXYGEN SPECIES PATHWAY, PROTEIN SECRETION and MTORC1 SIGNALING were activated, G2M CHECKPOINT, CHOLESTEROL HOMEOSTASIS and UV RESPONSE DN were suppressed. For AKR1C3 as shown in **Fig.11D**, PROTEIN SECRETION, REACTIVE OXYGEN SPECIES PATHWAY and ADIPOGENESIS were activated, G2M CHECKPOINT, UV RESPONSE DN and WNT BETA CATENIN SIGNALING were suppressed. In **Fig.11E**, results indicated that PIR activated in PROTEIN SECRETION, REACTIVE OXYGEN SPECIES PATHWAY and ADIPOGENESIS, G2M CHECKPOINT, UV RESPONSE DN and WNT BETA CATENIN SIGNALING were suppressed. Finally, in the top 3 items, REACTIVE OXYGEN SPECIES PATHWAY was activated in ACE2, AKR1C3, AKR1C2, PIR and ADH7. PROTEIN SECRETION and REACTIVE OXYGEN SPECIES PATHWAY were activated in ADH7, AKR1C2, AKR1C3 and PIR. For the possible co-regulated mechanisms between 5 genes, as shown in **Fig.12**, results of GeneMANIA showed that regulation of hormone levels and hormone metabolic process were particularly involved in ACE2, oxidoreductase activity, acting on CH-OH group of donors and oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor were involved in ADH7. AKRIC2 and AKR1C3 were involved in 7 functions included aldo-keto reductase (NADP) activity, oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor, oxidoreductase activity, acting on CH-OH group of donors, glycoside metabolic process, hormone metabolic process, regulation of hormone levels and quinone metabolic process.
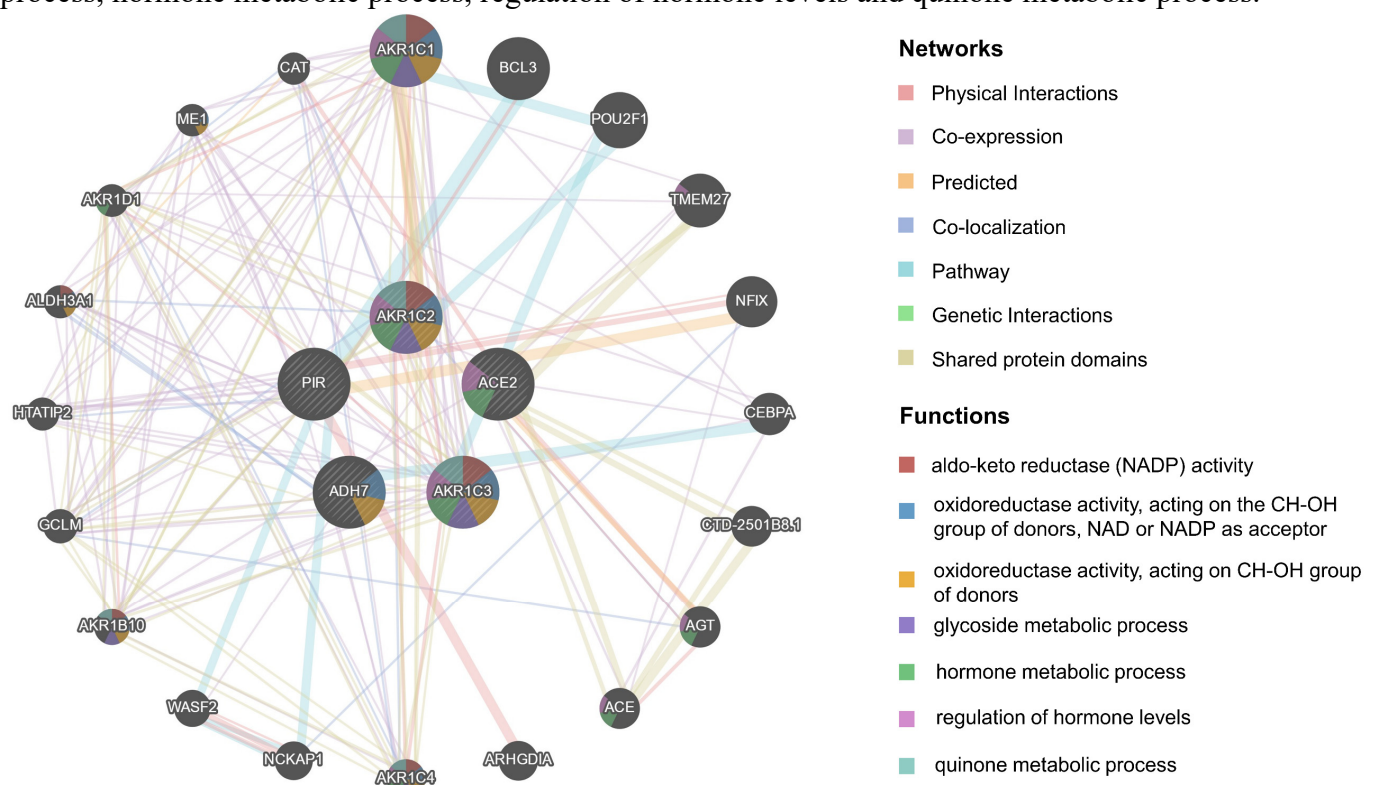
Fig.12 The network between ACE2, PIR, ADH7, AKR1C2, AKR1C3 and 20 interacted proteins
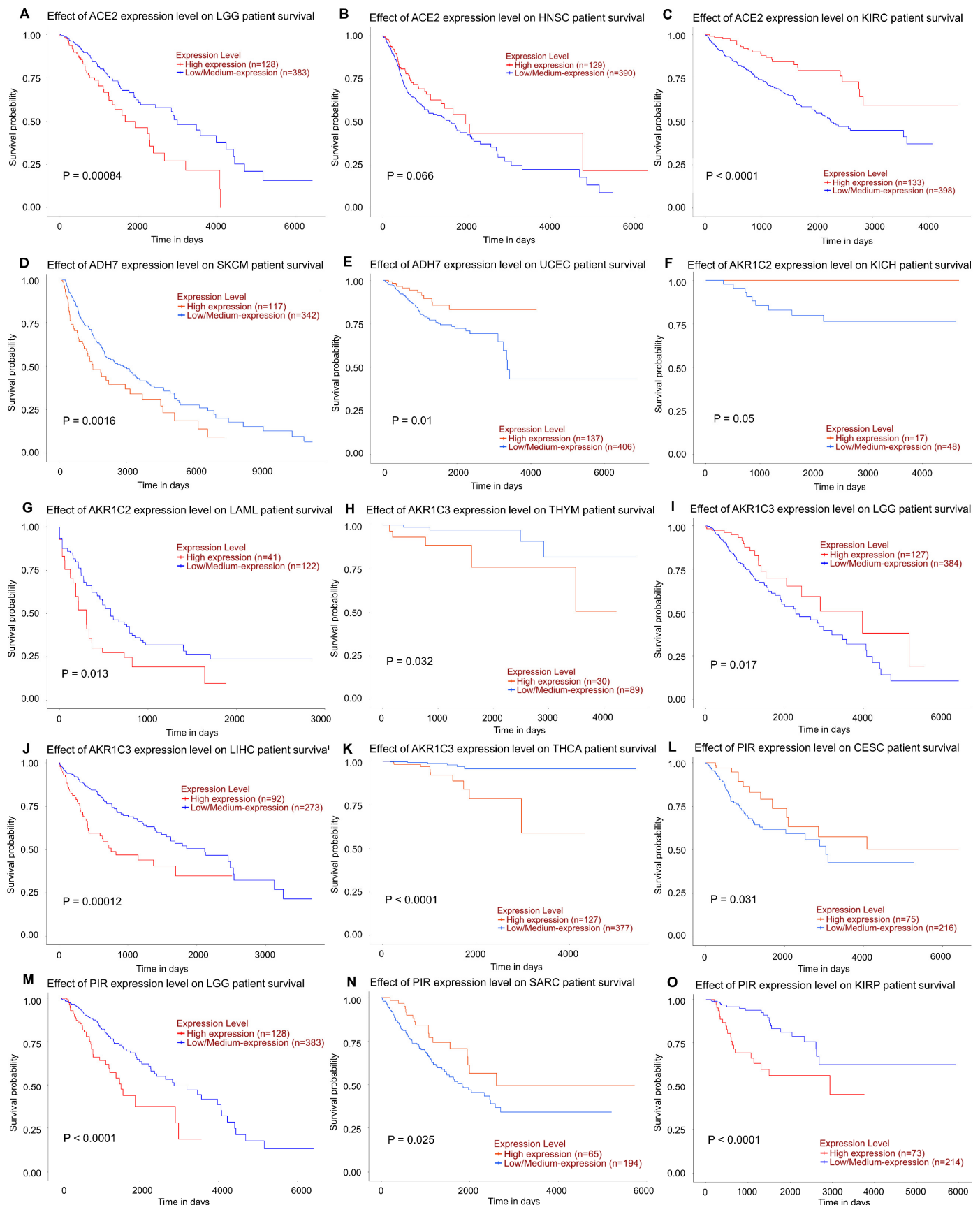
Fig.13 The survival probability of multi-cancers patients based on the 5 selected genes expression level, screening by P-value and survival curve status

Related cancers screening by prognosis and clinical risk prediction model construction and validation

After 5 genes were selected in smoker condition, the survival probability curves of all 33 type cancers were performed by ULCAN. After P-value and survival difference screening, as shown in **Fig.13**, 15 curves were satisfying included ACE2 in LGG (**Fig.13A**), HNSC (**Fig.13B**) and KIRC (**Fig.13C**), ADH7 in SKCM (**Fig.13D**) and UCEC (**Fig.13E**), AKR1C2 in KICH (**Fig.13F**) and LAML (**Fig.13G**), AKR1C3 in THYM

(**Fig.13H**), LGG (**Fig.13I**), LIHC (**Fig.13J**) and THCA (**Fig.13K**), PIR in CESC (**Fig.13L**), LGG (**Fig.13M**), SARC (**Fig.13N**) and KIRP (**Fig.13O**). P-values of both survival probability curves were less than 0.05 except ACE2 gene in HNSC had a margin at the edge of significance. Results indicated that these genes may play prognostic biomarkers in cancer. Brain lower grade glioma (LGG) was the highest co-occurrence type as both ACE2, PIR and AKR1C3 had the significantly prognostic situation. So, we further explored these 3 genes in LGG.
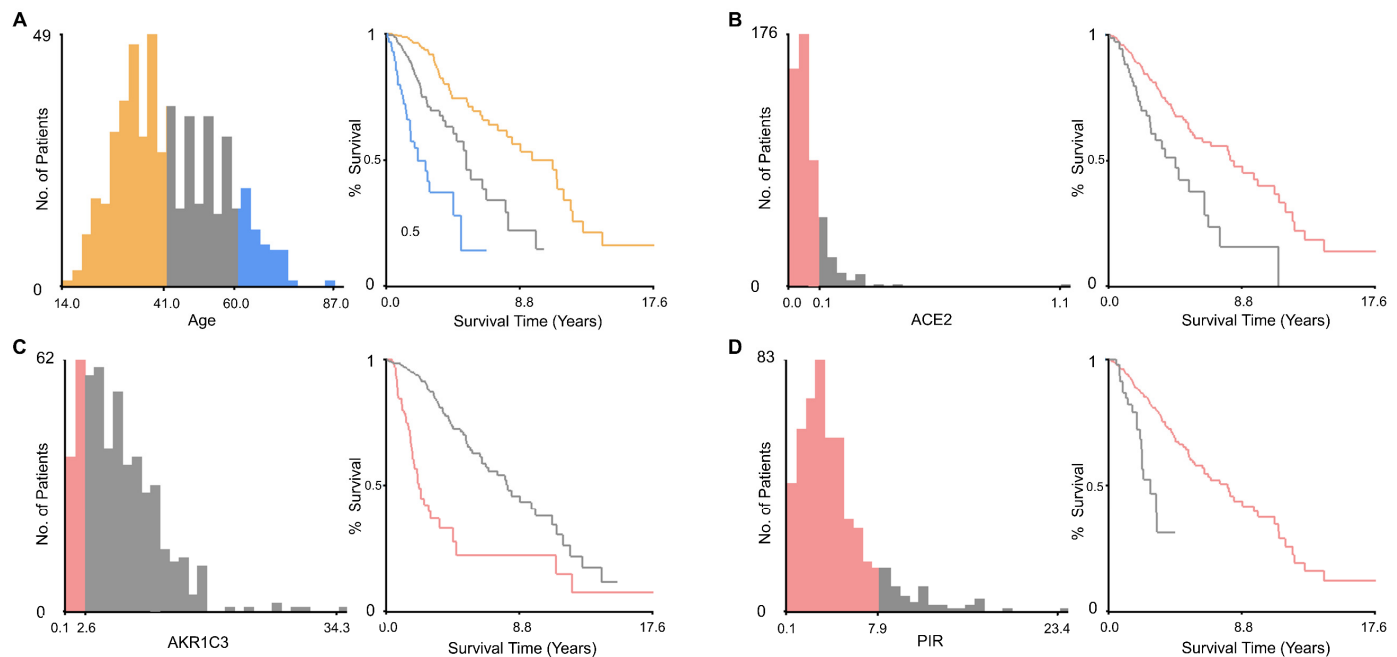
Fig.14 The cut-off points of Age (A), ACE2 (B), AKR1C3 (C) and PIR (D) detected by X-tile software

Three genes included ACE2, AKR1C3 and PIR expression levels were inserted in clinicopathological data. X-tile software made the cut-off points analysis of age, ACE2, PIR and AKR1C3 factors, instead of as the median level. As shown in **Fig.14A**, ternary classifications were in age with 14-41 years, 42-60 years and 61-87 years. Binary classifications were in ACE2, PIR and AKR1C3 with low and high expression status, as shown in **Fig.14B**, the cut-off point between low expression level of ACE2 is 0.12, and the cut-off point of AKR1C3 is 2.6 (**Fig.14C**) and the cut-off point of PIR is 7.9 (**Fig.14D**). After screening, patients with NA values in anyone of factors were excluded. Finally, 510 patients of TCGA database and 328 patients of CGGA database were obtained. Detailed baseline characteristics are summarized in **Table 1**. After the univariate Cox analysis, age, tumor grade, pharmaceutical, radiation, ACE2, PIR, AKR1C3 were detected to be statistically associated with the mortality of LGG (**Table 2**). Later, in the multivariable Cox proportion hazard regression models in which factors with p-value less than 0.05 were included based on the univariate analysis, to simplify the model, the most meaningful factors were included as the results indicated that the concordance index (C-index) of multi-factors permutation and combination (**Table 3**). Results indicated that the C-indexes of ACE2 and AKR1C3 factor joined the model individually were the highest value, and the Akaike information criterion (AIC) value of the only AKR1C3 joined model was slightly less than the both ACE2 and AKR1C3 joined (**Table 3**). AIC is to find the model that can best interpret the data but contains the least free parameters. The less AIC value, the better model. To further investigate and proof the rationality of ACE2 factor joined, we made the correlation relationship analysis of 3 genes in UCSC recomputed TCGA datasets contained 33 cancers. As shown in **Fig.15A-C**, in LGG and related cancer type, only AKR1C3 and ACE2 indicated the highest coefficient. The related correlation coefficient of AKR1C2 and ACE2 in LGG was 0.51 (**Fig.15D**) and in GBM was 0.79 (**Fig.15E**). In our study, age and tumor grade were determined the key and necessary factors, results indicated that the C-index improved when ACE2 and AKR1C3 were inserted. Finally, multivariable Cox analysis indicated that age, tumor grade and AKRIC3 were significant in statistical, and ACE2 were still brought as the P-value of 0.364 and its special important status in COVID-19 (**Table 2**).

Table 1. Clinicopathological characteristics of three datasets included training set and inner validation set of TCGA and two outer validation sets of CGGA

Factors	Subgroup	TCGA DATA		CGGA DATA	
		Training set N=410	Inner validation set N=100	Outer validation set 1 N=157	Outer validation set 2 N=171
Age	14-41	215 (0.5)	51 (0.5)	93 (0.6)	98 (0.6)
	42-60	146 (0.4)	37 (0.4)	56 (0.4)	63 (0.4)
	61-87	49 (0.1)	12 (0.1)	8 (0.1)	10 (0.1)
Gender	Female	191 (0.5)	37 (0.4)		
	Male	219 (0.5)	63 (0.6)		
Tumor grade	G2	205 (0.5)	43 (0.4)	105 (0.7)	98 (0.6)
	G3	205 (0.5)	57 (0.6)	52 (0.3)	73 (0.4)
Pharmaceutical	NO	176 (0.4)	36 (0.4)		
	YES	234 (0.6)	64 (0.6)		
Radiation	NO	160 (0.4)	34 (0.3)		
	YES	250 (0.6)	66 (0.7)		
ACE2	Low	345 (0.8)	86 (0.9)	87 (0.6)	70 (0.4)
	High	65 (0.2)	14 (0.1)	70 (0.4)	101 (0.6)
PIR	Low	366 (0.9)	88 (0.9)		
	High	44 (0.1)	12 (0.1)		
AKR1C3	Low	84 (0.2)	19 (0.2)	155 (1.0)	9 (0.1)
	High	326 (0.8)	81 (0.8)	2 (0.0)	162 (0.9)

Abbreviations: TCGA, The Cancer Genome Atlas; CGGA, Chinese Glioma Genome Atlas.

Table 2. Univariate and multivariable Cox hazards analysis of the training set

Factors	Subgroup	Univariate			Multivariate		
		HR	95%CI	P	HR	95%CI	P
Age	14-41	1.000			1.000		
	42-60	2.423	1.498-3.920	0.000	2.015	1.228-3.305	0.005
	61-87	7.063	4.062-12.280	0.000	5.291	2.978-9.400	0.000
Gender	Female	1.000					
	Male	1.108	0.746-1.646	0.612			
Tumor grade	G2	1.000			1.000		
	G3	3.635	2.323-5.689	0.000	2.409	1.509-3.845	0.000
Pharmaceutical	NO	1.000					
	YES	1.575	1.035-2.420	0.038			
Radiation	NO	1.000					
	YES	2.224	1.359-3.638	0.001			
ACE2	Low	1.000			1.000		
	High	2.113	1.344-3.322	0.001	1.241	0.778-1.981	0.364
PIR	Low	1.000					
	High	2.735	1.490-5.024	0.001			
AKR1C3	Low	1.000			1.000		
	High	0.214	0.143-0.321	0.000	0.295	0.193-0.453	0.000

**Table 3.** The C-index and AIC of multi-factors permutation and combination

Age	Tumor grade	Radiation	ACE2	PIR	AKR1C3	C-index	AIC
√	√					0.777	894.935
√	√	√				0.780	896.028
√	√		√			0.785	895.442
√	√			√		0.780	891.025
√	√				√	0.827	867.309
√	√			√	√	0.827	868.510

Abbreviations: C-index, Concordance index; AIC, Akaike information criterion.
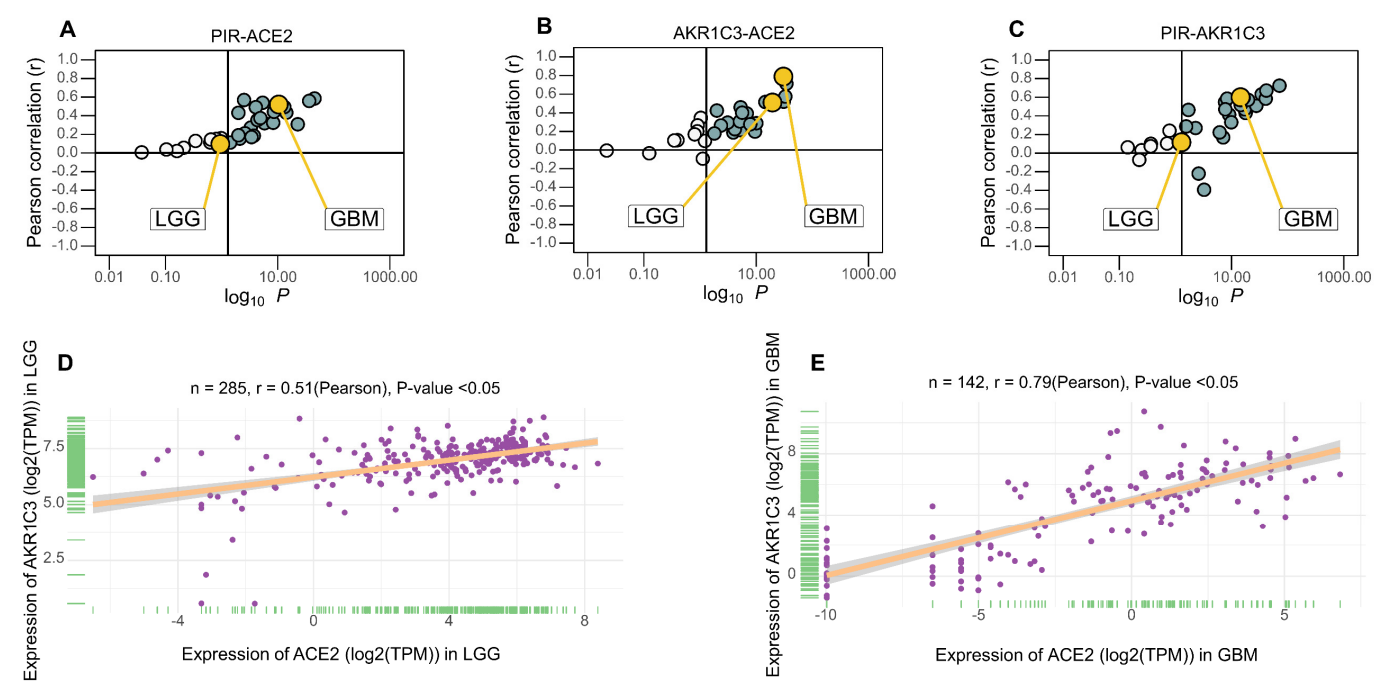
Fig.15 The correlation of between ACE2 and PIR (A), AKR1C3 and ACE2 (B), PIR and AKR1C3 (C) in multi-cancers, and detail coefficient between AKR1C3 and ACE2 in LGG cancer (D) and GBM caner (E)

Then, we established a nomogram to predict the 1-year, 2-year and 3-year survival, as shown in **Fig.16A**. Furthermore, a web calculator was constructed to improve the accuracy, the available URL is https://yhy9.shinyapps.io/LGG_Nomogram and the standby URL is https://figshare.com/articles/online_resource/LGG_nomogram/12619157. And the corresponding app of Android platform was achieved to improve the convenient and can be downloaded at https://figshare.com/articles/software/LGG_Tracker_1_0-signed_apk/12619121. The C-index of training model was 0.827, which indicated that about 82.7% probability of nomogram model to predict patient mortality correctly. As shown in **Fig.16B-C**, results pointed that the Area Under Curves (AUCs) of 1-year survival, 2-year survival and 3-year survival were 0.921(95%CI, 0.882-0.961), 0.911(95%CI, 0.860-0.962) and 0.878(95%CI, 0.818-0.939). In inner validation, the AUCs of 1-year survival, 2-year survival and 3-year survival were 0.777(95%CI, 0.640-0.914), 0.916(95%CI, 0.842-0.990) and 0.888(95%CI, 0.768-1.000). In outer validation one (**Fig.16D**), the AUCs of 1-year survival, 2-year survival and 3-year survival were 0.764(95%CI, 0.672-0.856), 0.848(95%CI, 0.783-0.913) and 0.748(95%CI, 0.656-0.841). And **Fig.16E** indicated that in outer validation two the AUCs of 1-year survival, 2-year survival and 3-year survival were 0.740(95%CI, 0.654-0.826), 0.766(95%CI, 0.688-0.843) and 0.794(95%CI, 0.721-0.866).
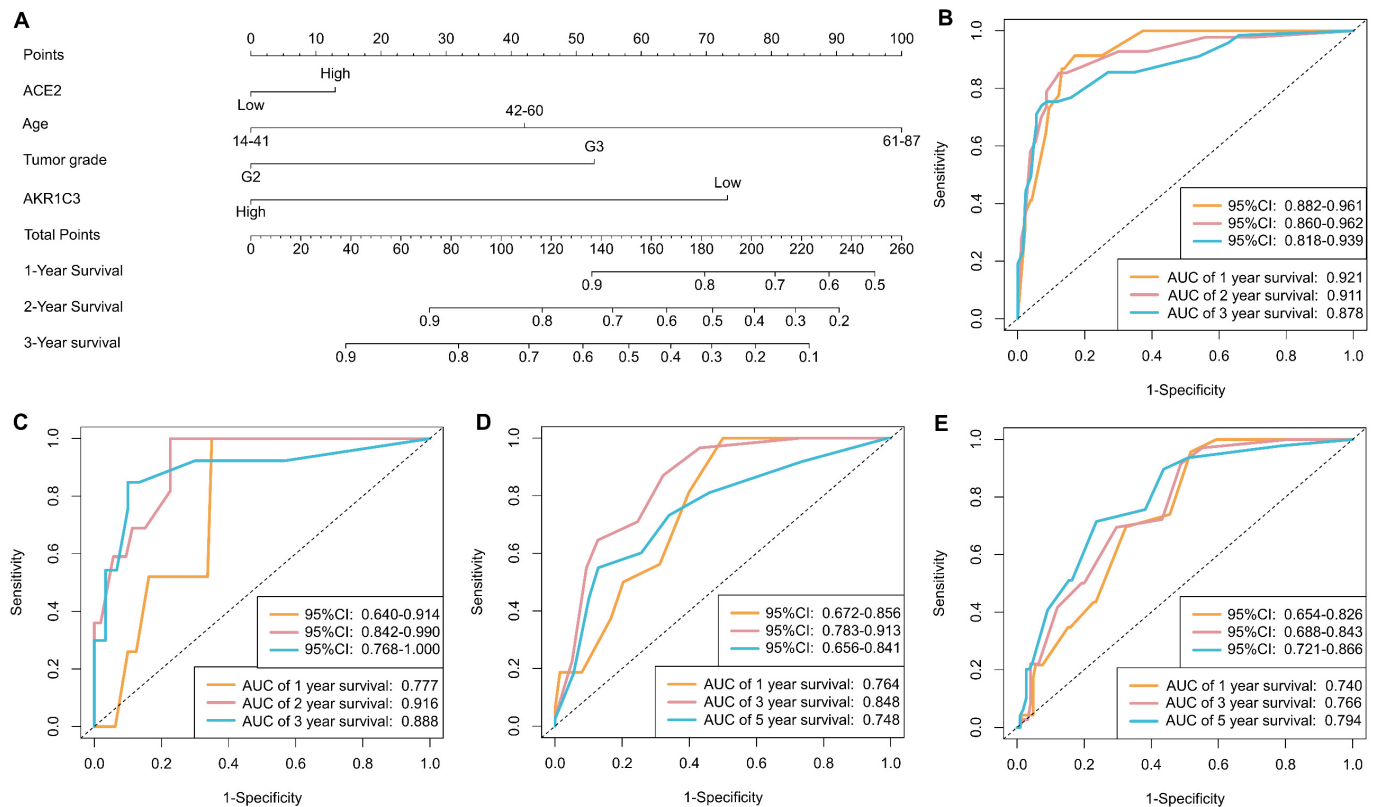
Fig.16 Nomogram of predicting 1-year, 2-year and 3-year survival of LGG patients (A), and the four ROC curves of training model (B), inner testing (C), outer testing one (D) and another outer testing two (E). For nomogram, each risk factor corresponded to a point by drawing a line straight upward to the points axis. The sum of the points located on the total points axis represented the probability of 1-year, 2-year and 3-year survival by drawing a line straight down to the survival axis

Discussions

We evaluated the ACE2 expression level of smokers included direct, second-hand and third-hand. ACE2 expression level significantly expressed upward in current smokers compared with never smokers, while other comparisons included former smokers and never smokers, current smokers and former smokers did not show the significant difference. Then, we explore the effect of Acute Smoke Exposing (ASE) in ACE2 expression level, results showed the significant upward expression difference. In Third-Hand Smoke (THS) groups, due to the poor amounts of samples, there was no significant expression difference and a little upward expression of ACE2 in after THS showed a trend. Total results pointed out that cigarette smoking effected the ACE2 expression level. To further identify the related genes about smoker trait, then, we chose GSE19667 and GSE52237 for WGCNA. Several modules were detected. And in most convincing positive correlation modules of smoker trait, the GO enrichment indicated that aminoglycoside antibiotic metabolic process, doxorubicin metabolic process and daunorubicin metabolic process were enriched in biological process, pyrimidine nucleotide-sugar transmembrane transporter activity, protein-disulfide reductase activity and alcohol dehydrogenase (NADP+) activity were enriched in molecular function, Golgi subcompartment, Golgi membrane and vesicle coat were enriched in cellular component. In negative correlation group of smokers, SRP-dependent co-translational protein targeting to membrane, co-translational protein targeting to membrane and viral gene expression were enriched in biological process, protein kinase C activity, small ribosomal subunit rRNA binding and leucine binding were enriched in molecular function, cytosolic ribosome, polysomal ribosome and ribosome were enriched in cellular component. KEGG enrichment indicated that the positive correlation groups enriched in Mucin type O-glycan biosynthesis, Protein export and Protein processing in endoplasmic reticulum and the negative correlation groups enriched in Ribosome, Pantothenate and CoA biosynthesis and Sulfur metabolism. Later, we made 2 steps intersections contained

the first intersection between ACE2 correlation genes and related genes about smoker trait in GSE52237 and GSE19667, and the second intersection between ACE2 correlation genes in human multi-tissues and gens both related ACE2 and smoker trait. 5 hub genes included ACE2 with PIR, ADH7, AKR1C2 and AKR1C3 were identified. GeneCards and Entrez gene summaries indicated that[54-56]: ACE2 (Angiotensin-converting enzyme 2) may play a role in the regulation of cardiovascular and renal function, as well as fertility. Associated diseases with ACE2 include Severe Acute Respiratory Syndrome and Coronavirus Infection. Related pathways contain Agents Acting on the Renin-Angiotensin System Pathway, Pharmacodynamics and Peptide hormone metabolism. Gene Ontology (GO) annotations related to ACE2 include metallopeptidase activity and peptide binding; PIR (Pirin), the encoded protein may act as a transcriptional cofactor and be involved in the regulation of DNA transcription and replication. Associated diseases include Exhibitionism and Obsessive-Compulsive Personality Disorder. Related pathways include Cell Receptor Signaling Pathway and Apoptosis and Autophagy. GO annotations related to PIR include transcription coregulator activity and quercetin 2,3-dioxygenase activity; ADH7 (Alcohol Dehydrogenase 7), the encoded enzyme may participate in the synthesis of retinoic acid, a hormone important for cellular differentiation. Associated diseases with ADH7 include Alcohol Dependence and Atrophic Gastritis. Related pathways are Drug metabolism-cytochrome P450 and Cytochrome P450-arranged by substrate type. GO annotations related to ADH7 include oxidoreductase activity and retinol binding; AKR1C2 (Aldo-keto Reductase Family 1, member C2), associated Diseases with AKR1C2 include 46, Xy Sex Reversal 8 and Colorectal Cancer, Hereditary Nonpolyposis, Type 2. Related pathways are Drug metabolism - cytochrome P450 and Naphthalene metabolism. GO annotations related to AKR1C2 include oxidoreductase activity and carboxylic acid binding; AKR1C3 (Aldo-keto Reductase Family 1, member C3) may play an important role in the pathogenesis of allergic diseases such as asthma, and in controlling cell growth and/or differentiation. Associated diseases with AKR1C3 include Prostate Disease and Endometrial Cancer. Related pathways are Signaling by Retinoic Acid and Naphthalene metabolism. GO annotations related to AKR1C3 include oxidoreductase activity and aldo-keto reductase (NADP) activity. GSEA results indicated that in smokers and non-smokers groups, ACE2 suppressed in BETA SIGNALING, WNT BETA CATENIN SIGNALING and HEME METABOLISM, activated in REACTIVE OXYGEN SPECIES PATHWAY, GLYCOLYSIS and PROTEIN_SECRETION. For ADH7, PROTEIN SECRETION, REACTIVE OXYGEN SPECIES PATHWAY and GLYCOLYSIS were activated, CHOLESTEROL HOMEOSTASIS, E2F TARGETS and UV RESPONSE_DN were suppressed. For AKR1C2, REACTIVE OXYGEN SPECIES PATHWAY, PROTEIN SECRETION and MTORC1 SIGNALING were activated, G2M CHECKPOINT, CHOLESTEROL HOMEOSTASIS and UV RESPONSE DN were suppressed. For AKR1C3, PROTEIN SECRETION, REACTIVE OXYGEN SPECIES PATHWAY and ADIPOGENESIS were activated, G2M CHECKPOINT, UV RESPONSE DN and WNT BETA CATENIN SIGNALING were suppressed. For PIR, activated in PROTEIN SECRETION, REACTIVE OXYGEN SPECIES PATHWAY and ADIPOGENESIS, G2M CHECKPOINT, UV RESPONSE DN and WNT BETA CATENIN SIGNALING were suppressed. GeneMANIA showed that regulation of hormone levels and hormone metabolic process were particularly involved in ACE2, oxidoreductase activity, acting on CH-OH group of donors and oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor were involved in ADH7. AKRIC2 and AKR1C3 were involved in 7 functions included aldo-keto reductase (NADP) activity, oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor, oxidoreductase activity, acting on CH-OH group of donors, glycoside metabolic process, hormone metabolic process, regulation of hormone levels and quinone metabolic process. Later, we made the pan-cancers survival curves of 5 hub genes, and Brain Lower Grade Glioma (LGG) was the highest co-occurrence type as both ACE2, PIR and AKR1C3 had the significantly prognostic situation. Further clinical predicting model was constructed with the factors of ACE2, age, AKR1C3 and tumor grade. And related web dynamic nomogram and encapsulation APP were created as well. Some limitations occurred in current study, 3 prognostic genes included ACE2, PIR and AKR1C3 need to be verified in laboratory experiments. And the prediction model based on the retrospective data and the factors in clinical data were limited that can miss something more

important. In conclusions, ACE2 and AKR1C3 play a prospective role in brain lower grade glioma and corresponding clinical nomogram improves the prognostic capability. But the mechanism of ACE2 and AKR1C3 in LGG need to be deeply investigated in the future.

Funding

Acknowledgements

Conflicts of Interest

All authors declare no conflict of interest.

References

1. Zheng Y-Y, Ma Y-T, Zhang J-Y, Xie X: **COVID-19 and the cardiovascular system.** *Nature Reviews Cardiology* 2020, **17:**259-260.
2. **COVID-19 situation update worldwide** [https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases]
3. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, Schiergens TS, Herrler G, Wu N-H, Nitsche A, et al: **SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor.** *Cell* 2020, **181:**271-280.e278.
4. Tonnesen P, Marott JL, Nordestgaard B, Bojesen SE, Lange P: **Secular trends in smoking in relation to prevalent and incident smoking-related disease: A prospective population-based study.** *Tob Induc Dis* 2019, **17:**72.
5. **Tobacco** [https://www.who.int/news-room/fact-sheets/detail/tobacco]
6. Strzelak A, Ratajczak A, Adamiec A, Feleszko W: **Tobacco Smoke Induces and Alters Immune Responses in the Lung Triggering Inflammation, Allergy, Asthma and Other Lung Diseases: A Mechanistic Review.** *International Journal of Environmental Research and Public Health* 2018, **15**.
7. Park J-E, Jung S, Kim A: **MERS transmission and risk factors: a systematic review.** *BMC Public Health* 2018, **18:**574.
8. Öberg M, Jaakkola MS, Woodward A, Peruga A, Prüss-Ustün A: **Worldwide burden of disease from exposure to second-hand smoke: a retrospective analysis of data from 192 countries.** *The Lancet* 2011, **377:**139-146.
9. Zhu B-q, Heeschen C, Sievers RE, Karliner JS, Parmley WW, Glantz SA, Cooke JP: **Second hand smoke stimulates tumor angiogenesis and growth.** *Cancer Cell* 2003, **4:**191-196.
10. Berlin I, Thomas D, Le Faou A-L, Cornuz J: **COVID-19 and Smoking.** *Nicotine & Tobacco Research* 2020.
11. Edgar R, Domrachev M, Lash AE: **Gene Expression Omnibus: NCBI gene expression and hybridization array data repository.** *Nucleic Acids Res* 2002, **30:**207-210.
12. Spira A, Beane J, Shah V, Liu G, Schembri F, Yang X, Palma J, Brody JS: **Effects of cigarette smoke on the human airway epithelial cell transcriptome.** *Proc Natl Acad Sci U S A* 2004, **101:**10143-10148.
13. Beane J, Sebastiani P, Liu G, Brody JS, Lenburg ME, Spira A: **Reversible and permanent effects of tobacco smoke exposure on airway epithelial gene expression.** *Genome Biol* 2007, **8:**R201.
14. Billatos E, Faiz A, Gesthalter Y, LeClerc A, Alekseyev YO, Xiao X, Liu G, Ten Hacken NHT, Heijink IH, Timens W, et al: **Impact of acute exposure to cigarette smoke on airway gene expression.** *Physiol Genomics* 2018, **50:**705-713.
15. Pozuelos GL, Kagda MS, Schick S, Girke T, Volz DC, Talbot P: **Experimental Acute Exposure to Thirdhand Smoke and Changes in the Human Nasal Epithelial Transcriptome: A Randomized Clinical Trial.** *JAMA Network Open* 2019, **2:**e196362-e196362.

16.   Strulovici-Barel Y, Omberg L, O'Mahony M, Gordon C, Hollmann C, Tilley AE, Salit J, Mezey J, Harvey BG, Crystal RG: **Threshold of biologic responses of the small airway epithelium to low levels of tobacco smoke.** *Am J Respir Crit Care Med* 2010, **182:**1524-1532.

17.   Walters MS, De BP, Salit J, Buro-Auriemma LJ, Wilson T, Rogalski AM, Lief L, Hackett NR, Staudt MR, Tilley AE, et al: **Smoking accelerates aging of the small airway epithelium.** *Respir Res* 2014, **15:**94.

18.   Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK: **limma powers differential expression analyses for RNA-sequencing and microarray studies.** *Nucleic Acids Research* 2015, **43:**e47-e47.

19.   Zhao Z, Meng F, Wang W, Wang Z, Zhang C, Jiang T: **Comprehensive RNA-seq transcriptomic profiling in the malignant progression of gliomas.** *Scientific Data* 2017, **4:**170024.

20.   Bao Z-S, Chen H-M, Yang M-Y, Zhang C-B, Yu K, Ye W-L, Hu B-Q, Yan W, Zhang W, Akers J, et al: **RNA-seq of 272 gliomas revealed a novel, recurrent PTPRZ1-MET fusion transcript in secondary glioblastomas.** *Genome Research* 2014, **24:**1765-1773.

21.   Fang S, Liang J, Qian T, Wang Y, Liu X, Fan X, Li S, Wang Y, Jiang T: **Anatomic Location of Tumor Predicts the Accuracy of Motor Function Localization in Diffuse Lower-Grade Gliomas Involving the Hand Knob Area.** *American Journal of Neuroradiology* 2017, **38:**1990-1997.

22.   Goldman MJ, Craft B, Hastie M, Repečka K, McDade F, Kamath A, Banerjee A, Luo Y, Rogers D, Brooks AN, et al: **Visualizing and interpreting cancer genomics data via the Xena platform.** *Nature Biotechnology* 2020, **38:**675-678.

23.   Team RC: **R: A Language and Environment for Statistical Computing.** 2016.

24.   Team R: **RStudio: Integrated Development for R.** 2020.

25.   Software G: **One-tailed T-test with unpaired was performed using GraphPad Prism version 8.0.2 for Windows.** 2020.

26.   Langfelder P, Horvath S: **WGCNA: an R package for weighted correlation network analysis.** *BMC Bioinformatics* 2008, **9:**559.

27.   Langfelder P, Horvath S: **Fast R Functions for Robust Correlations and Hierarchical Clustering.** *J Stat Softw* 2012, **46**.

28.   and GRWaBBaLBaRGaWHaALaTLaMMaAMaSM: **gplots: Various R Programming Tools for Plotting Data.** 2020.

29.   The Gene Ontology Consortium: **The Gene Ontology Resource: 20 years and still GOing strong.** *Nucleic Acids Research* 2018, **47:**D330-D338.

30.   Kanehisa M, Goto S: **KEGG: kyoto encyclopedia of genes and genomes.** *Nucleic acids research* 2000, **28:**27-30.

31.   Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, et al: **Enrichr: a comprehensive gene set enrichment analysis web server 2016 update.** *Nucleic Acids Res* 2016, **44:**W90-97.

32.   Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR, Ma'ayan A: **Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool.** *BMC Bioinformatics* 2013, **14:**128.

33.   Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: **Cytoscape: a software environment for integrated models of biomolecular interaction networks.** *Genome Res* 2003, **13:**2498-2504.

34.   GmbH y: **cytoscape yFiles Layout Algorithms.** 2020.

35.   Wickham H: **ggplot2: Elegant Graphics for Data Analysis.** 2016.

36.   Slowikowski K: **ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2'.** 2020.

37.   Kolde R: **pheatmap: Pretty Heatmaps.** 2019.

38.   Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP: **Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles.** *Proceedings of the National Academy of Sciences* 2005, **102:**15545-15550.

39.   Yu G, Wang L-G, Han Y, He Q-Y: **clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters.** *OMICS: A Journal of Integrative Biology* 2012, **16:**284-287.

40.   Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, Franz M, Grouios C, Kazi F, Lopes CT, et al: **The GeneMANIA prediction server: biological network integration for gene prioritization and predicting**

gene function. *Nucleic Acids Res* 2010, **38:**W214-220.

41.   Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarthi B, Varambally S: **UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses.** *Neoplasia* 2017, **19:**649-658.

42.   Camp RL, Dolled-Filhart M, Rimm DL: **X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization.** *Clin Cancer Res* 2004, **10:**7252-7259.

43.   T T: **A Package for Survival Analysis in R.** 2020.

44.   Grambsch TMTM: **The Cox Model.** *Modeling Survival Data: Extending the Cox Model Statistics for Biology and Health* 2000**:**39-77.

45.   Akaike H: **A new look at the statistical model identification.** *IEEE Transactions on Automatic Control* 1974, **19:**716-723.

46.   Gittleman H, Sloan AE, Barnholtz-Sloan JS: **An independently validated survival nomogram for lower-grade glioma.** *Neuro-Oncology* 2019, **22:**665-674.

47.   Han M-Z, Huang B, Ni S-L, Wang J, Li X-G, Bjerkvig R: **A validated prognostic nomogram for patients with newly diagnosed lower-grade gliomas in a large-scale Asian cohort.** *Neuro-Oncology* 2020, **22:**729-731.

48.   Blanche P, Dartigues J-F, Jacqmin-Gadda H: **Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks.** *Statistics in Medicine* 2013, **32:**5381-5397.

49.   Jr FEH: **rms: Regression Modeling Strategies.** 2020.

50.   others FEHJawcfCDam: **Hmisc: Harrell Miscellaneous.** 2020.

51.   Newell AJaDRaAA-laJ: **DynNom: Visualising Statistical Models using Dynamic Nomograms.** 2019.

52.   McPherson WCaJCaJAaYXaJ: **shiny: Web Application Framework for R.** 2020.

53.   Yang H-Y: **LGG Tracker.apk.** *figshareSoftware* 2020.

54.   Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, Stein TI, Nudel R, Lieder I, Mazor Y, et al: **The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses.** *Current Protocols in Bioinformatics* 2016, **54:**1.30.31-31.30.33.

55.   Maglott D, Ostell J, Pruitt KD, Tatusova T: **Entrez Gene: gene-centered information at NCBI.** *Nucleic Acids Res* 2007, **35:**D26-31.

56.   O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, et al: **Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation.** *Nucleic Acids Res* 2016, **44:**D733-745.