


ORIGINAL ARTICLE

Collagen Type III Alpha 1 chain regulated by GATA-Binding Protein 6 affects Type II IFN response and propanoate metabolism in the recurrence of lower grade glioma

Runzhi Huang^{1,2,3} | Zhenyu Li³ | Xiaolong Zhu¹ | Penghui Yan¹ | Dianwen Song⁴ | Huabin Yin⁴ | Peng Hu¹ | Ruoyi Lin³ | Shengyu Wu³ | Tong Meng⁴ | Jie Zhang^{3,5} | Zongqiang Huang¹ 

¹Department of Orthopedics, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China

²Division of Spine, Department of Orthopedics, Tongji Hospital affiliated to Tongji University School of Medicine, Shanghai, China

³Tongji University School of Medicine, Shanghai, China

⁴Department of Orthopedics, Shanghai General Hospital, School of Medicine, Shanghai Jiaotong University, Shanghai, China

⁵Shanghai First Maternity and Infant Hospital, Tongji University School of Medicine, Shanghai, China

Correspondence

Zongqiang Huang, Department of Orthopaedics, The First Affiliated Hospital of Zhengzhou University, Zhengzhou 450052, China.
Email: gzhuangzq@163.com

Tong Meng, Department of Orthopedics, Shanghai General Hospital, School of Medicine, Shanghai Jiaotong University, Shanghai, China.
Email: mengtong@medmail.com.cn

Jie Zhang, Shanghai First Maternity and Infant Hospital, Tongji University School of Medicine, 201204, Shanghai, China.
Email: jiezhang@tongji.edu.cn

Funding information

This research was funded by the National Natural Science Foundation of China (Grant No. 81702659; 81772856; 81501203). Youth Fund of Shanghai Municipal Health Planning Commission (No.2017YQ054); Henan medical science and technology research project (Grant No. 201602031). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Abstract

Some studies suggested the prognosis value of immune gene in lower grade glioma (LGG). Recurrence in LGG is a tough clinical problem for many LGG patients. Therefore, prognosis biomarker is required. Multivariate prognosis Cox model was constructed and then calculated the risk score. And differential expressed transcription factors (TFs) and differential expressed immune genes (DEIGs) were co-analysed. Besides, significant immune cells/pathways were identified by single sample gene set enrichment analysis (ssGSEA). Moreover, gene set variation analysis (GSVA) and univariate Cox regression were applied to filter prognostic signalling pathways. Additionally, significant DEIG and immune cells/pathways, and significant DEIG and pathways were co-analysed. Further, differential enriched pathways were identified by GSEA. In sum, a scientific hypothesis for recurrence LGG including TF, immune gene and immune cell/pathway was established. In our study, a total of 536 primary LGG samples, 2,498 immune genes and 318 TFs were acquired. Based on edgeR method, 2,164 DEGs, 2,498 DEIGs and 31 differentials expressed TFs were identified. A total of 106 DEIGs were integrated into multivariate prognostic model. Additionally, the AUC of the ROC curve was 0.860, and *P* value of Kaplan-Meier curve < 0.001. GATA6 (TF) and COL3A1 (DEIG) were selected ($R = 0.900$, $P < 0.001$, positive) as significant TF-immune gene links. Type II IFN response ($P < 0.001$) was the significant immune pathway. Propanoate metabolism ($P < 0.001$) was the significant KEGG pathway. We

Runzhi Huang, Zhenyu Li, Xiaolong Zhu have contributed equally to this work.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Journal of Cellular and Molecular Medicine* published by Foundation for Cellular and Molecular Medicine and John Wiley & Sons Ltd

proposed that COL3A1 was positively regulated by GATA6, and by effecting type II IFN response and propanoate metabolism, COL3A1 involved in LGG recurrence.

KEYWORDS

low-grade glioma, prognosis, propanoate metabolism, recurrence, type II IFN response

1 | INTRODUCTION

Immune genes involve in tumour genesis, development and prognosis,^{1,2} and these genes can be applied in mechanism exploring, diagnosis, therapy target identification and prognosis.^{2,3} Some studies indicated that immune-related lncRNA showed its prognosis value in lower grade glioma (LGG).²

LGG is a serial common centre nervous system tumour. According to classification of WHO, glioma can be divided into I, II, III and IV grade, and grade I and II are included in LGG. And estimated by the American Cancer Society, it will be 23,890 new cases and 18,020 death in America.⁴ Generally, the prognosis of primary LGG is better than higher grade glioma, and the median overall survival is 5-10 years.⁵ However, the recurrence, progression and deterioration make the therapy tough. And the recurrence rates of grade I and II are 84.5%⁶ and 57.6%,⁶ respectively. Besides, 50%-75% LGG patients died of progression and deterioration.⁷ Therefore, for optimizing treatment, the biomarker for recurrence is required in the early stage. And we proposed immune gene might participate in LGG recurrence.

In our study, differential expressed immune genes (DEIGs) were identified, and univariate Cox regression was utilized to access the prognostic value. Additionally, significant prognosis DEIGs were integrated into multivariate prognosis Cox model and then calculated the risk score. Univariate and multivariate Cox regression were applied to validate risk score as the independent prognosis factor. Similarly, differential expressed transcription factors (TFs) were identified. Then, co-analysis for differential expressed TFs and DEIGs to identify the significant TF-DEIG relationship was performed. Based on CIBERSORT, immune cells/pathways were identified, and significant immune cells/pathways were identified by ssGSEA. And significant DEIG and immune cells/pathways to identify the significant immune cell/pathway were co-analysed. Moreover, GSEA was utilized to identify the expression level in clinical LGG sample, and univariate Cox regression evaluated the prognosis value of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. Further, differential expressed KEGG pathways were identified by GSEA. Additionally, significant DEIG and KEGG pathways to identify the significant KEGG pathway were co-analysed. In sum, a scientific hypothesis for LGG recurrence based on TF, immune gene, immune cell/pathway and KEGG pathway was established.

2 | MATERIAL AND METHOD

2.1 | Data acquisition

Based on TCGA database (<https://tcga-data.nci.nih.gov>), gene expression profiles and clinical information of 536 primary LGG

samples were acquired. Data of 2498 immune genes were available from ImmPort (<http://www.immport.org/>) and MsigDB (<http://software.broadinstitute.org/gsea/msigdb/>) database. Besides, data of 318 TFs were downloaded from Cistrome database (<http://cistrome.org>).

2.2 | Differential expressed genes identification

Gene expression level of clinical samples in TCGA database was performed by edgeR method, differential expressed genes (DEGs) based on whether recurrence or not were identified by log₂ FC (Fold change) and false discovery rate (FDR) *P* value. When log₂ FC >1.0 or <-1.0, and FDR *P* value < 0.05, DEGs were defined. Further, Gene Oncology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis were utilized to explore clinical significance of DEGs. Similarly, immune genes were performed by edgeR method, when log₂ FC >1.0 or <-1.0, and FDR *P* value < 0.05, DEIGs were defined. In addition, to access the prognostic value, DEIGs were screened by univariate Cox regression. Also, TFs were screened by edgeR method, differential expressed TFs were identified as log₂ >1.0 or <-1.0, and FDR *P* value < 0.05.

2.3 | Multivariate model construction and independent prognosis analysis

To predict the prognosis of LGG, differential expressed prognostic significance immune genes were integrated into multivariate model. Additionally, the accuracy of the model was tested by receiver operator characteristic (ROC) curve. And for each LGG sample, risk score was calculated by this following formula:

$$\text{Risk Score} = \beta_1 \times \text{DEIG}_1 + \beta_2 \times \text{DEIG}_2 + \beta_3 \times \text{DEIG}_3 + \dots + \beta_n \times \text{DEIG}_n.$$

For each single sample, 'n' represented the number of DEIG in the multivariate model, 'β' represented regression coefficient of each DEIG, and DEIG_n represented the expression level of the Nth DEIG in the corresponding sample. Then, median of risk score was selected to distinguish the high and low-risk score, and samples were divided into the high- and low-risk group, respectively. Besides, Kaplan-Meier curve was utilized to predict the prognosis value of the multivariate model based on risk score.

Further, the value of independently prognostic for risk score was evaluated by univariate and multivariate Cox regression analysis. And other demographic information like age, gender and race was applied for correction.

2.4 | Regulatory network for immune genes and TFs

To validate the clinical relationship with recurrence/new tumour event, non-parametric test for DEIGs was utilized. The Pearson correlation analysis was utilized to co-analyse differential TFs and DEIGs. When correlation coefficient more than 0.300 and P value < 0.001 , the corresponding regulatory relationships based on TFs and DEIGs were considered to be statistically significant. Besides, the most significant regulatory relationship was selected, and the most significant TF and DEIG were selected, respectively.

2.5 | Co-analysis the relationship of immune gene with immune cells/pathways and KEGG pathway

To identify the characteristic of immune cells and pathways in the downstream of DEIGs in clinical samples, single sample gene set enrichment analysis (ssGSEA) was utilized to filter significant immune cells and pathways. Besides, the Pearson analysis was used to identify the correlation between the most significant DEIG and immune cells/pathways. When P value < 0.05 , the interplay was regarded as significant. And the most significant immune cell or pathway was decided by the largest correlation coefficient.

Aiming to filter prognostic KEGG pathway, gene set variation analysis (GSVA) was applied to access the expression level of KEGG pathways, and all KEGG pathways were in the downstream of the most significant DEIG. In addition, univariate Cox regression was utilized to calculate the prognostic value of KEGG pathways. And survival-related KEGG pathways were identified. What's more, to co-analyse the most significant DEIG and survival-related KEGG pathways, the Pearson analysis was utilized, and P value < 0.05 was defined as significant.

2.6 | Regulatory network for TFs, immune genes, immune cell and KEGG pathways

Gene set enrichment analysis (GSEA) was utilized to identify significant KEGG pathway between primary and recurrence LGG samples in clinical data. According to the result of GSVA and GSEA, the most significant survival-related and recurrent-related KEGG pathway was selected.

The regulatory network was constructed based on the interaction of the most significant TF, DEIG, top 9 immune cells/pathways and top three KEGG pathways.

2.7 | Multidimensional validation

To validate our scientific hypothesis and decrease inherent deficient in silico, multidimensional validation was applied based on the following online databases. Immune gene set was utilized to find related genes of corresponding immune cells/pathways.⁸ Besides, to find top 5 genes

represent KEGG pathway, Pathway Card (<https://pathcards.genecards.org/>) was applied. Gene Expression Profiling Interactive Analysis (GEPIA),⁹ Oncomine,¹⁰ PROGeneV2,¹¹ UALCAN,¹² Linkedomics,¹³ cBioportal,¹⁴ Genotype-Tissue Expression (GTEx),¹⁵ UCSC xena,¹⁶ Cancer Cell Line Encyclopedia (CCLE),¹⁷ Expression atlas,¹⁸ The Human Protein Atlas,¹⁹ CGGA²⁰ and String²¹ were utilized to validate the hypothesis.

2.8 | Statistics analysis

For all analysis process, tow-sided P value < 0.05 was defined as statistically significant. And R version 3.6.1 was utilized for analysis in silico (Institute for Statistics and Mathematics, Vienna, Austria; www.r-project.org) (Package: e1071, parallel, preprocessCore, sva, limma, edgeR, ggplot2, survminer, survival, rms, randomForest, pROC, glmnet, pheatmap, timeROC, vioplot, corrplot, ConsensusClusterPlus, forestplot, survivalROC, beeswarm, chromVAR, Biostrings, BSgenome.Hsapiens, UCSC.hg38, TxDb.Hsapiens.UCSC.hg38.knownGene, clusterProfiler, org.Hs.eg.db, ggplot2, karyoploteR, limma, pheatmap, TCGAAbiolinks, dplyr, kableExtra, ggplot2, ggrepel, ggthemes, gridExtra, msigdb, data.table, GSVA and stringr).

3 | RESULT

3.1 | Differential immune gene identification

All analysis process was illustrated in Figure S1. And all baseline information of LGG samples were summarized in Table 1. Gene expression level from 536 LGG samples in TCGA database was analysed, \log_2 FC > 1 or < -1 , and P value < 0.05 was defined as expressed significantly and differentially. Heatmap plot (Figure 1A) and volcano plot (Figure 2B) demonstrated a total of 2,164 DEGs based on primary and recurrence were identified. Additionally, GO (Figure 1C) and KEGG (Figure 1D) enrichment analyses were performed for exploring the clinical link of DEGs. Most significant GO for BP, CC and MF was skeletal system development, extracellular matrix and DNA-binding transcription activator activity, RNA polymerase II-specific; most significant KEGG pathway was neuroactive ligand-receptor interaction. Moreover, based on edgeR method for 2498 immune genes, when \log_2 FC > 1 or < -1 and P value < 0.05 , DEIGs were defined. And heatmap plot (Figure 2A) and volcano plot (Figure 2B) showed 269 DEIGs included in 2164 DEGs. In addition, to evaluate the prognosis value of DEIGs, univariate Cox regression was utilized and 99 DEIGs were significant as risk factor and 7 DEIGs were significant as protective factor (Figure 2C).

3.2 | Multivariate prognostic model and independent prognostic factor identification

A total of 106 DEIGs were integrated into multivariate prognostic model to access the prognosis of LGG, and ROC curve (Figure 3A)

TABLE 1 Baseline information of 536 patients diagnosed with lower grade glioma

Variables	Total patients (N = 536)
Age, years	
Mean \pm SD	42.67 \pm 13.32
Median (Range)	40.00 (14-87)
Gender	
Female	242 (45.15%)
Male	293 (54.66%)
NA	1 (0.19%)
Race	
White	494 (92.16%)
Black or African-American	22 (4.11%)
American-Indian or Alaska native	1 (0.19%)
Asian	8 (1.49%)
NA	11 (2.05%)
Sample type	
Primary tumour	516 (96.27%)
Recurrent tumour	20 (3.73%)
New tumour event	
Yes	148 (27.61%)
No	279 (52.05%)
NA	109 (20.34%)

Abbreviation: NA, not available; SD, Standard deviation.

was applied to evaluate the efficiency of the multivariate model (AUC = 0.860). Besides, to validate the prognostic value of the risk score, Kaplan-Meier curve (Figure 3B) was utilized, and P value < 0.001. Additionally, scatter plot (Figure 3C) and risk curve (Figure 3D) were also demonstrated the clinical link of risk score based on clinical samples. Moreover, in the heatmap plot (Figure 3E), green and red represented the high- and low-risk group samples, and 21 DEIGs lower than low-risk group, 32 DEIGs higher than the high-risk group and 53 DEGs were not significant.

To explore whether risk score was an independent prognostic factor or not, univariate (HR = 1.007, 95% CI (1.004-1.009), P < 0.001) (Figure 3F) and multivariate (HR = 1.006, 95% CI (1.003-1.008), P < 0.001) (Figure 3G) Cox regression models were utilized. Therefore, risk score was defined as an independent prognostic factor.

3.3 | Co-analysis DEIGs and TFs

EdgeR method for identified differential expressed TFs, and log₂ FC >1 or <-1 and P value < 0.05 was set. Heatmap (Figure 4A) and volcano (Figure 4B) plots illustrated 31 differential expressed TFs from 318 TFs. Then, co-analysed differential expressed TFs and DEIGs, links of 11 TFs and 23 DEIGs were significant. Further, to identify DEIGs related to clinical and TFs, the Venn plot (Figure 4C) of

recurrent (54 immunogenes), TFs-related (23 immune genes) and new tumour event (60 immune genes) was applied, and 14 DEIGs were selected. Besides, the Pearson analysis results (Figure 4D) for regulatory relationship of TFs and immune genes, and the link of GATA6 (TF) and COL3A1 (DEIG) was selected (R = 0.900, P < 0.001, positive). What's more, Beeswam plots for new tumour event (P = 0.006) (Figure 4E) and recurrent (P = 0.003) (Figure 4F) showed the clinical relationship of COL3A1.

3.4 | Co-analysis DEIG with immune cells/ pathways and KEGG pathways

Based on CIBERSORT for LGG samples, 29 types immune cells or pathways were defined. Further, ssGSEA was utilized to find significant immune cells and pathways in overall survival, and 4 immune cells and 5 immune pathways were significant. Then, co-analysed COL3A1 and 29 types immune cells or pathways, heatmap plot (Figure 5A) demonstrated the top three immune cells or pathways: type II IFN response (R = 0.300, P < 0.001), Treg (R = 0.270, P < 0.001) and CCR (R = 0.260, P < 0.001). The detail correlation of COL3A1 and type II IFN response (P < 0.001) (Figure 5C), Treg (P < 0.001) (Figure 5D) and CCR (P < 0.001) (Figure 5E) based on ssGSEA also showed. Eventually, type II IFN response was selected.

To find overall survival-related KEGG pathways, 58 significant KEGG pathways were filtered by GSVA based on 184 KEGG pathways. And co-analysed COL3A1 and 184 KEGG pathways, heatmap plot (Figure 5B) showed the top four KEGG pathways were ECM receptor interaction (R = 0.21, P < 0.001), N-glycan biosynthesis (R = 0.19, P < 0.001), small cell lung cancer (R = 0.18, P < 0.001) and glycosaminoglycan degradation (R = 0.18, P < 0.001).

3.5 | Significant KEGG pathway identification

A total of 184 KEGG pathways were screened by GSEA, and 12 KEGG pathways were significant differential expressed between primary and recurrent LGG samples. And Venn plot (Figure 6A) showed the 3 KEGG pathways significant in GSEA (12 KEGG pathways) and GSVA (58 KEGG pathways): propanoate metabolism, systemic lupus erythematosus and proteasome. And all significant KEGG pathways were analysed by GSEA (Figure 6B). Besides, for the relationship with COL3A1, propanoate metabolism (P < 0.001) (Figure 6C), systemic lupus erythematosus (P = 0.003) (Figure 6D) and proteasome (P = 0.027) (Figure 6E) were all significant. Moreover, plots of propanoate metabolism (Figure 6F), systemic lupus erythematosus (Figure 6G) and proteasome (Figure 6H) based on GSEA were also demonstrated. Additionally, according to co-analysis with COL3A1 in heatmap, propanoate metabolism was selected from propanoate metabolism (R = -0.140, P < 0.001), systemic lupus erythematosus (R = 0.130, P < 0.001) and proteasome (R = 0.100, P < 0.001).

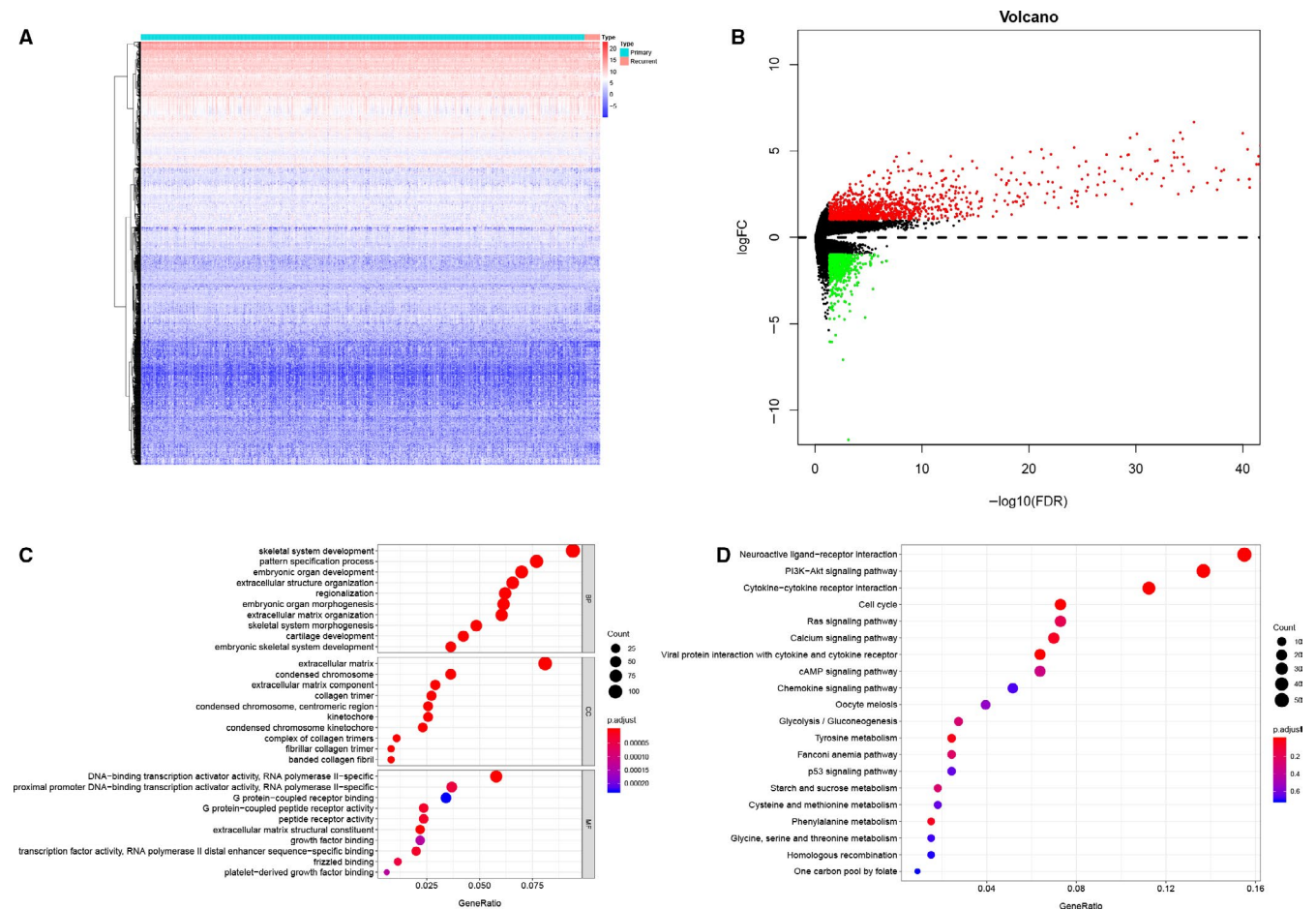


FIGURE 1 Heatmap for genes from samples (A), blue and red represent primary and recurrent samples. Volcano plot for genes from samples (B), green and red dots represent significant differential expression. Bubble plot for GO (C) and KEGG (D) enrichment analysis. GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes

Therefore, the regulatory network (Figure 7A) based on GATA6 (the most significant TF), COL3A1 (DEIG), CCR (immune pathway), Treg (immune cell), APC-co-stimulation (immune pathways), DC (immune cell), check-point (immune pathway), T-cell co-stimulation (immune pathway), type II IFN response (immune pathway), macrophage (immune cell), Th2 cell (immune cell), propanoate metabolism (KEGG pathway), systemic lupus erythematosus (KEGG pathway) and proteasome (KEGG pathway) was established. Co-expressed circle plot (Figure 7B) of COL3A1, propanoate metabolism, systemic lupus erythematosus and proteasome illustrated the interplay relationship.

Ultimately, the scientific hypothesis (Figure 7C): COL3A1 was positively regulated by GATA6 and affect type II IFN response and propanoate metabolism in recurrent LGG.

3.6 | Multidimensional validation

To decrease the bias of different platform and validate our scientific hypothesis, multivariate validation was utilized. Based on

immune gene set and Pathway Card, GPR146, SELP and AHR represented type II IFN response, SELP, AHR, BCKDHA, BCKDHB, DLD, SUCLG1 and DBT represented propanoate metabolism, respectively. GEPIA (Figure S2), Oncomine (Figure S3), PROGeneV2 (Figure S4), UALCAN (Figure S5), Linkedomics (Figure S6), cBioportal (Figure S7), GTEx (Figure S8), UCSC xena (Figure S9), CCLE (Figure S10), Expression atlas, The human protein atlas (Figure S11), CGGA (Figure S12) and String (Figure S13) database were validated. Gene relationships were summarized in Table S1, all expression levels were summarized in Table S2, and all prognosis information were summarized in Table S3. All these figures and tables were mentioned in supplement materials.

GATA6 and SELP low-expressed and AHR, BCKDHA, BCKDHB, DLD, SUCLG1 and DBT high-expressed in LGG.

GATA6 ($P = 0.007$, Linkedomics, Figure S6A), COL3A1 ($P < 0.001$, GEPIA, Figure S2L; $P = 0.007$, PROGeneV2, Figure S4B; $P < 0.001$, Linkedomics, Figure S6B), AHR ($P = 0.014$, Linkedomics, Figure S6E), BCKDHA ($P = 0.011$, PROGeneV2, Figure S4E; $P = 0.041$, Linkedomics, Figure S6F; $P = 0.013$, CGGA, Figure S12F), SUCLG1 ($P < 0.001$, GEPIA, Figure S2S; $P < 0.001$, Linkedomics, Figure S6I) and

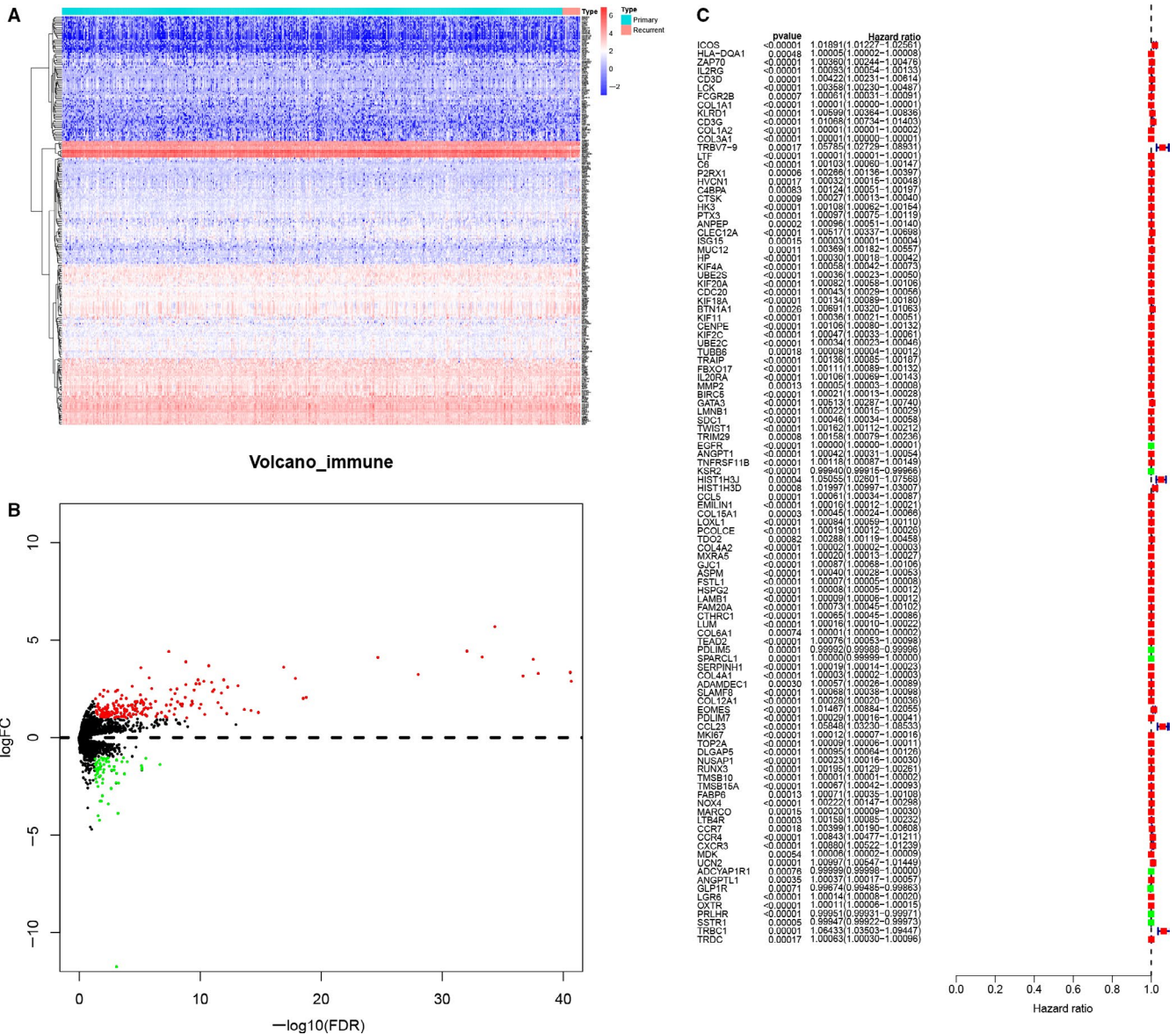


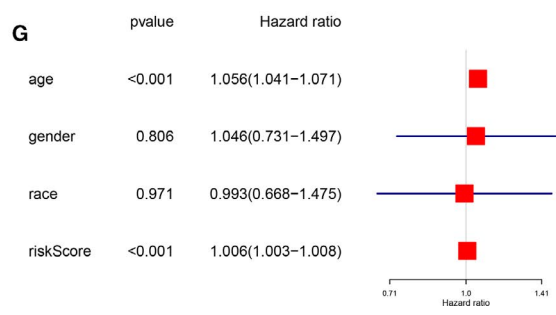
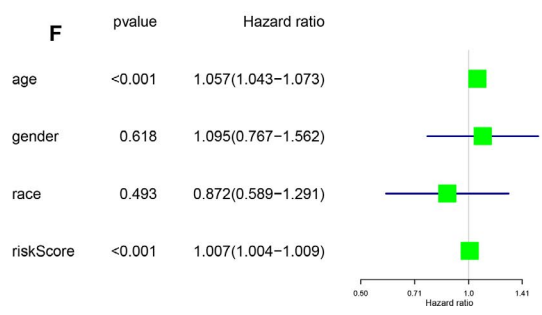
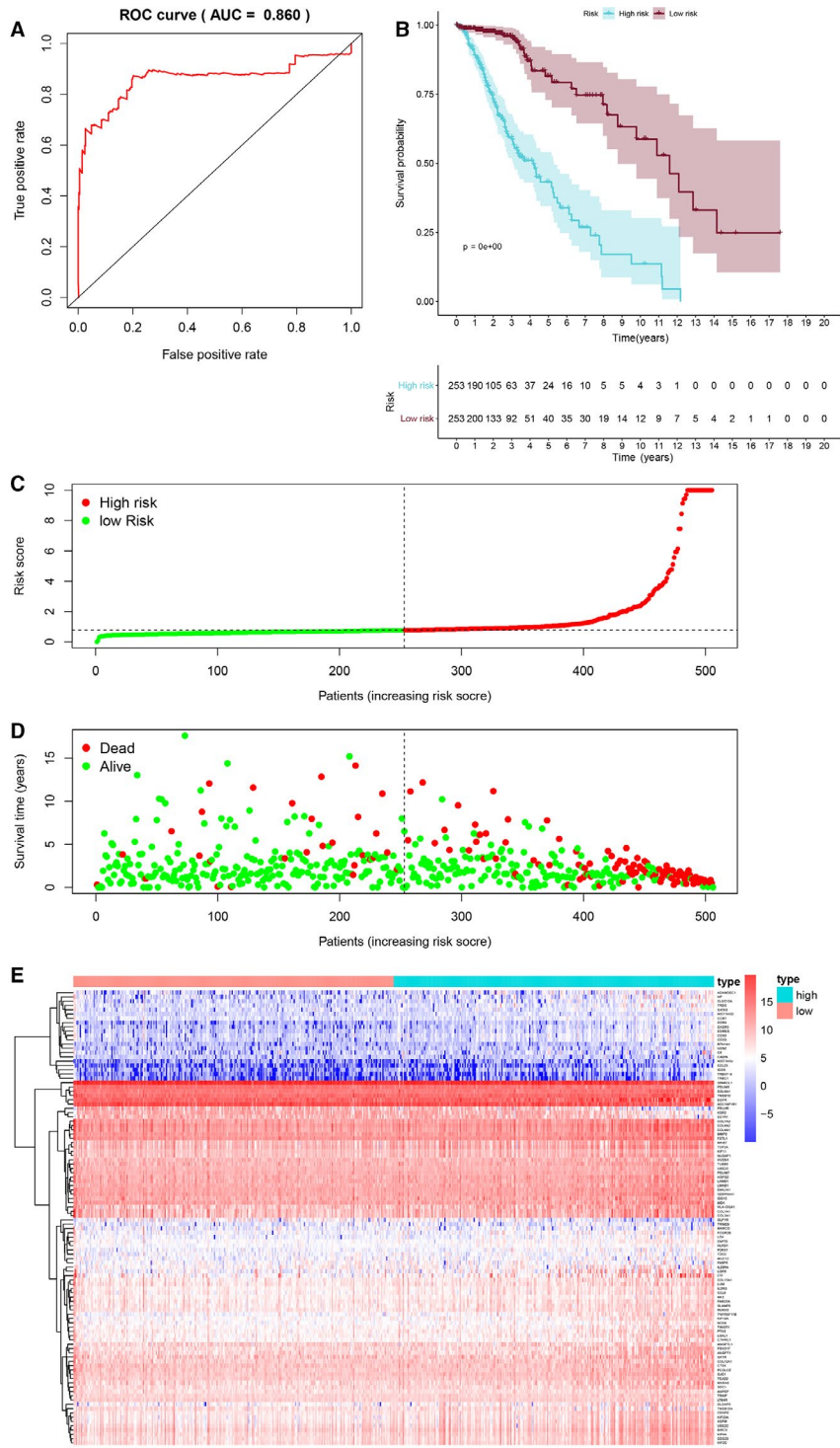
FIGURE 2 Heatmap for immune genes (A), blue and red represent primary and recurrent samples. Volcano plot for immune genes (B), green and red dots represent significant differential expression. Univariate Cox regression for immune genes (C)

DBT ($P = 0.032$, cBioportal, Figure S7T; $P = 0.008$, CGGA, Figure S12T) showed significant related to overall survival; BCKDHA ($P = 0.021$, cBioportal, Figure S7P) and DBT ($P < 0.001$, cBioportal, Figure S8R) showed significant related to prognosis; GPR146 ($P = 0.021$, CGGA, Figure S12M), SELP ($P = 0.021$, CGGA, Figure S12N) and BCKHDA ($P = 0.012$, CGGA, Figure S12P) showed significant related to recurrence.

4 | DISCUSSION

Primary LGG develops slowly and has a well prognosis. However, the prognosis becomes poor when LGG recurs, progresses and deteriorates.⁷ And the progression-free survival of 5-, 10- and 15- years were 38%, 18% and 1%, respectively.⁶ Besides, the rate of malignant conversion at the first recurrence was 36.4%.⁶ Thus, a biomarker for

FIGURE 3 ROC curve for (AUC = 0.860) prognostic immune genes (A). Overall survival Kaplan-Meier curve for prognostic immune genes (B). Scatterplot (C) for the high and low-risk group samples, red and green represent high-risk and low-risk groups. Risk plot (D) for dead and alive samples, red and green represent dead and alive samples. Heatmap for prognostic immune genes (E), red and green represent low-risk and high-risk groups. Univariate (F) and multivariate (G) Cox regression model. Green and red represent univariate and multivariate Cox regression model. AUC, area under the curve; ROC, The receiver operator characteristic



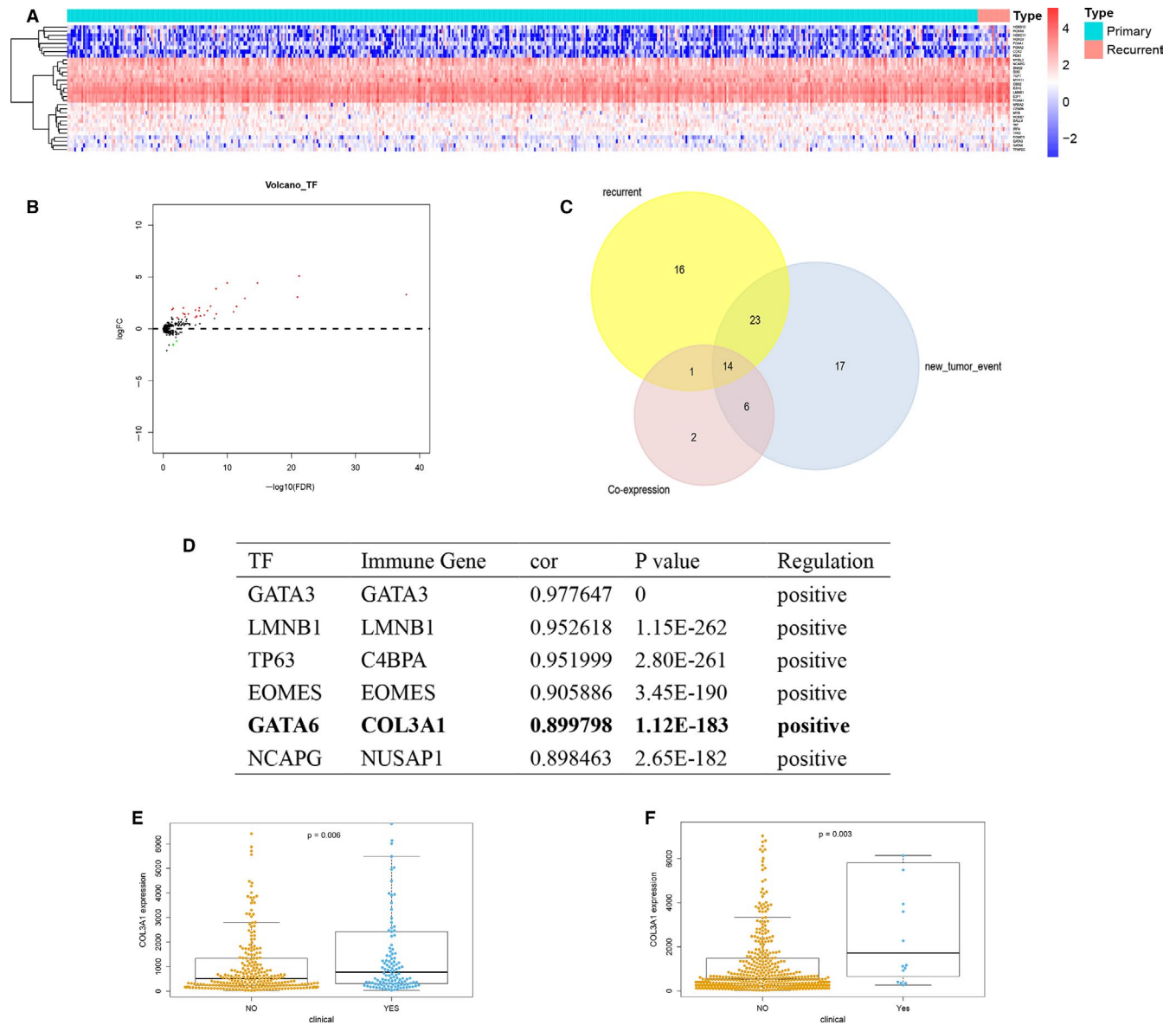


FIGURE 4 Heatmap for TFs (A), blue and red represent primary and recurrent samples. Volcano plot for TFs (B), green and red dots represent significant differential expression. Venn plot for co-expressed, recurrent and new tumour event immune genes (C). Result for co-analysis immune genes and TFs (D). Non-parametric test beeswarm plot for new tumour event (E) and recurrent (F) and COL3A1. TF, transcription factor

predicting the recurrence in LGG is essential. We proposed immune genes play a role in prognosis.²²

In our study based on comprehensive bioinformatics, a total of 536 LGG samples, 2,498 immune genes and 318 TFs were acquired. Based on edgeR method, 2,164 DEGs, 2,498 DEIGs and 31 differential expressed TFs were identified. A total of 106 DEIGs from univariate Cox regression were integrated into multivariate prognostic model. Additionally, the AUC of the ROC curve was 0.860, and Kaplan-Meier curve P value < 0.001 . And univariate ($HR = 1.007$, 95% CI (1.004-1.009), $P < 0.001$) and multivariate ($HR = 1.006$, 95% CI (1.003-1.008), $P < 0.001$) Cox regression model showed risk score from multivariate prognostic model was an independent prognostic factor. Further, according to 14 clinical-related DEIGs and significant TF-immune gene links, GATA6

(TF) and COL3A1 (DEIG) were selected ($R = 0.900$, $P < 0.001$, positive). Moreover, 4 immune cells and 5 immune pathways were selected by ssGSEA. And based on co-analysed COL3A1 and 29 types of immune cells or pathways, type II IFN response ($R = 0.300$, $P < 0.001$) was significant. What's more, based on the comprehensive consideration of GSEA, GSVA and co-analysis result of COL3A1 and 184 KEGG pathways, propanoate metabolism ($R = -0.140$, $P < 0.001$) was selected. Eventually, COL3A1 was positively regulated by GATA6, and type II IFN response and propanoate metabolism were the downstream of COL3A1 in recurrent LGG.

GATA6 is a TF named GATA-binding protein 6, which belongs to the GATA family, and it expressed in neuron, astrocyte, choroid plexus epithelial cells and endothelial cells.²³ GATA6 plays an

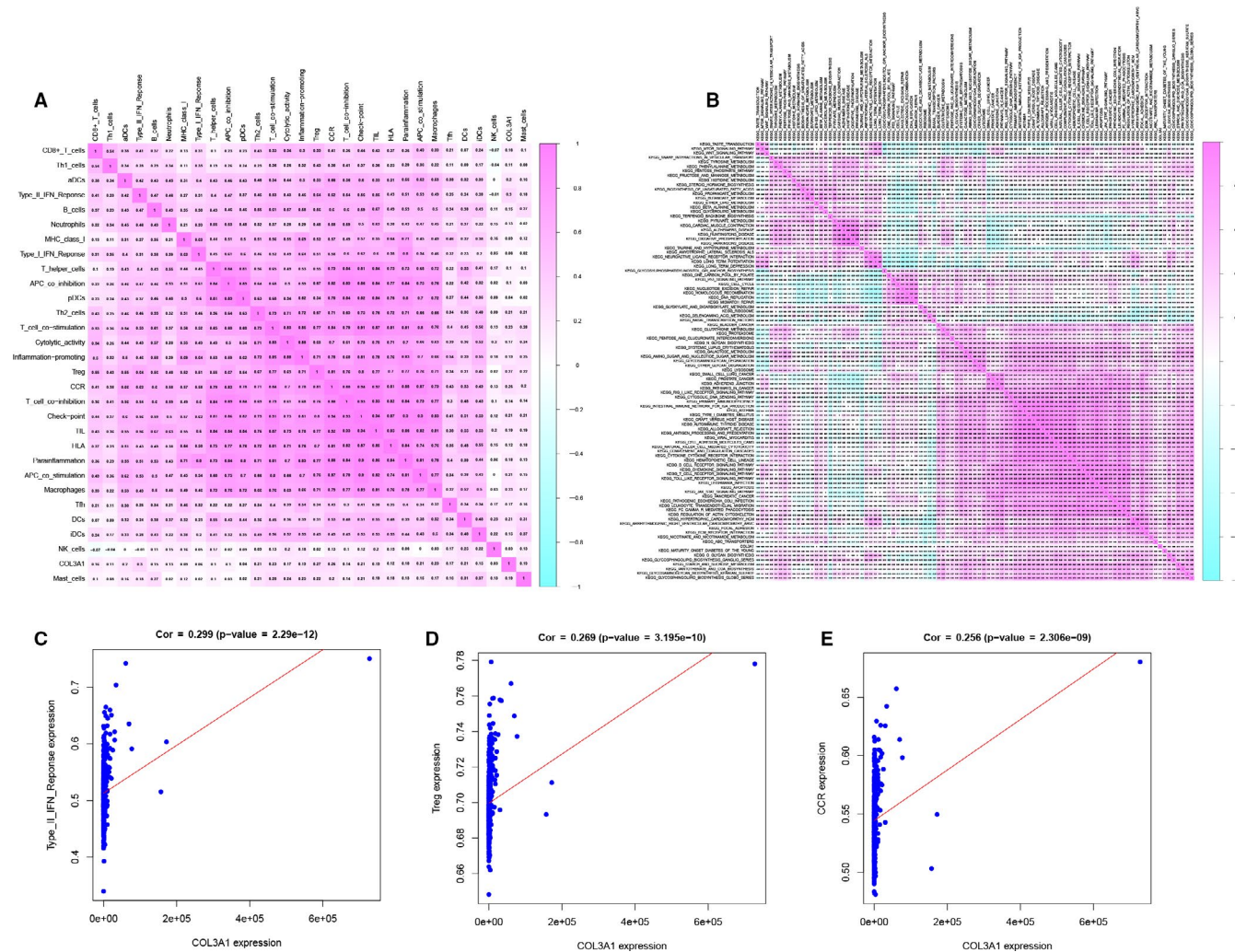


FIGURE 5 Co-analysis immune cells/pathways and COL3A1 by ssGSEA analysis (A), co-analysis KEGG pathways and COL3A1 by GSVA analysis (B). Correlation analysis for Type II IFN Reponse and COL3A1 (C), correlation analysis for Treg cell and COL3A1 (D), correlation analysis for CCR expression and COL3A1 (E) by ssGSEA analysis. GSVA, Gene Set Variation Analysis; ssGSEA, Single Sample Gene Set Enrichment Analysis

important role in differentiation during the process of embryogenesis, in proliferation during the process of development, and in inhibition apoptosis during the process tumour genesis.²⁴ In addition, the frequency of methylation in GATA6 was 68.4%, and hypermethylation in GATA6 was positively related to the prognosis of glioblastoma multiforme (GBM).²⁵ Besides, based on expression, loss of heterozygosity and loss function analysis, GATA6 low-expressed in tumour, and researcher indicated that GATA6 was a new tumour suppressor gene in astrogloma.²⁶ However, GATA6 also plays a different role in promoting recurrence in cholangiocarcinoma and related to poor prognosis.²⁷

COL3A1 is an immune gene named collagen type III alpha 1 chain, and it codes the pro-alpha 1 chains of type III collagen in the vascular system and other tissue.²⁸ COL3A1 highly expressed in glioma.¹ Some study indicated that it promoted tumour genesis by regulating the angiogenesis process, and it was defined as an oncogene.^{28,29} By validation based on qPCR, COL3A1 correlated with

prognosis of GBM, and the expression level was related to the grade of glioma (normal vs LGG, $P < 0.001$; normal vs higher grade glioma (HGG), $P = 0.002$; LGG vs HGG, $P = 0.005$).¹ And some studies showed it might relate to recurrence in LGG by genome-scale integrated analysis.²⁹

However, there was no study reported the direct interaction between GATA6 and COL3A1. Based on bioinformation analysis and other studies,^{27,29} we speculated GATA6 played a role in eliciting the transcription of COL3A1, and high-expressed COL3A1 coding more pro-alpha 1 chains of type III collagen in vessel, then activated angiogenesis process which participate in recurrence of LGG.

Type II IFN response is a kind of defence mechanism, and the main molecule was IFN- γ . However, IFN- γ can positively and negatively regulate immune system; thus, it is an antitumour cytokine and also a promoter for tumour genesis.³⁰ And in centre nervous system, IFN- γ mediated neurogenesis and differentiation through ERK1/ 2, and it also plays a role in apoptosis in brain cells and blastoma.³¹

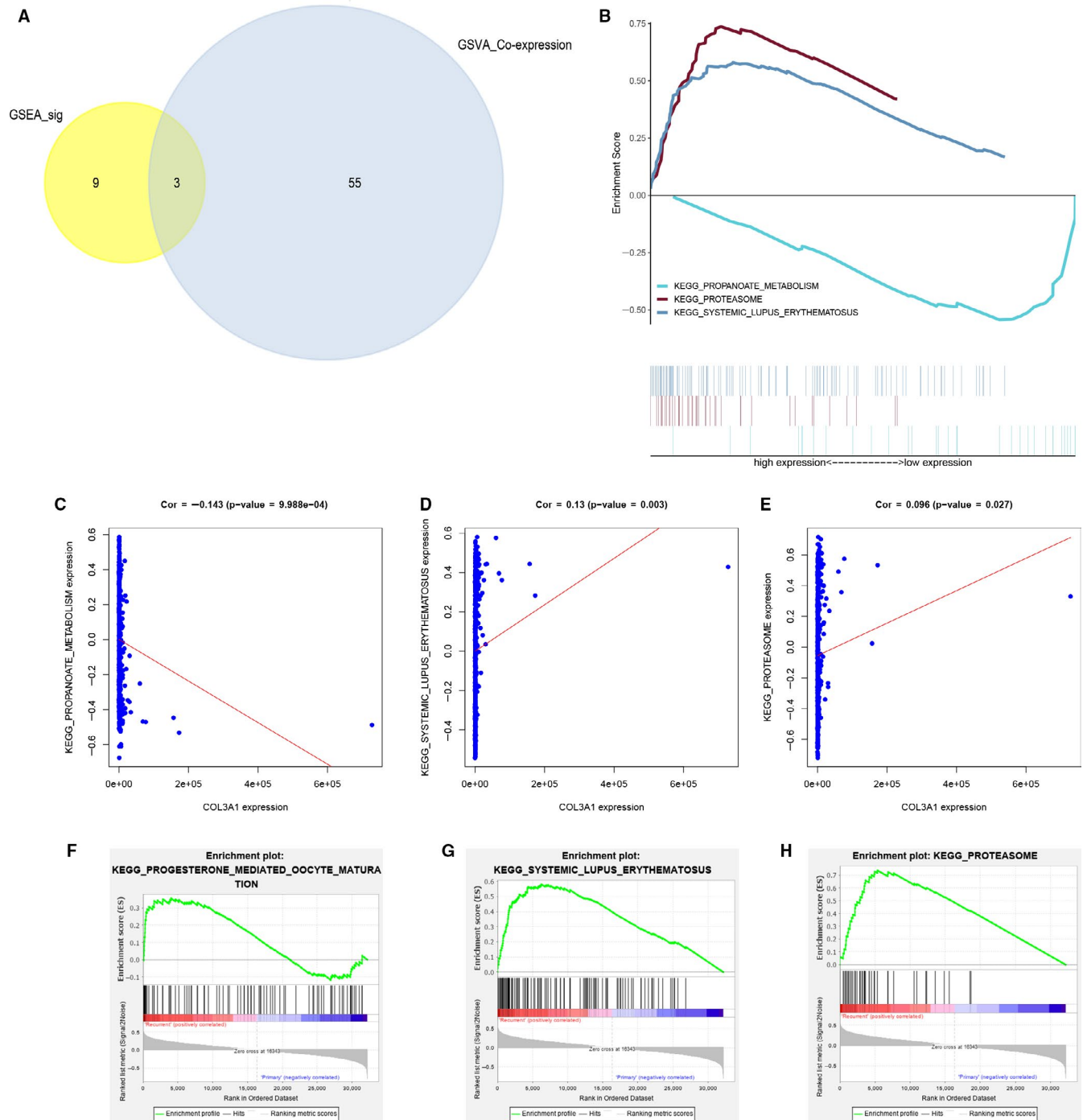


FIGURE 6 Venn plot for KEGG pathway selected by GSEA and GSVa analysis (A), GSEA analysis for top three KEGG pathways (B). Correlation analysis for propanoate metabolism pathway and COL3A1 (C), correlation analysis for systemic lupus erythematosus pathway and COL3A1 (D), correlation analysis for proteasome pathway and COL3A1 (E) by GSEA analysis. GSEA analysis for propanoate metabolism pathway (F), systemic lupus erythematosus pathway (G) and proteasome pathway (H). GSEA, Gene Set Enrichment Analysis; GSVa, Gene Set Variation Analysis; KEGG, Kyoto Encyclopedia of Genes and Genomes

Besides, it promoted PD-L1 expression and inhibited the immune system, which inhibited tumour eradication by T cells and promoted immune escape.³²

IFN- γ can inhibit tumour angiogenesis to inhibit tumour genesis, and ICAM1 low-expressed when knockout IFN- γ .³³ By String database²¹ analysis, COL3A1 regulates biomarkers of type II IFN

response (GPR146, SELP and AHR) by ICAM1 and VCAM1. No study has shown the relationship between COL3A1 and IFN- γ , but we assume COL3A1 down-regulated type II IFN response to promote angiogenesis in tumour.

Propanoate metabolism is downstream of the lipid metabolism, and propanoate will convert into succinyl-coenzyme A; then, the

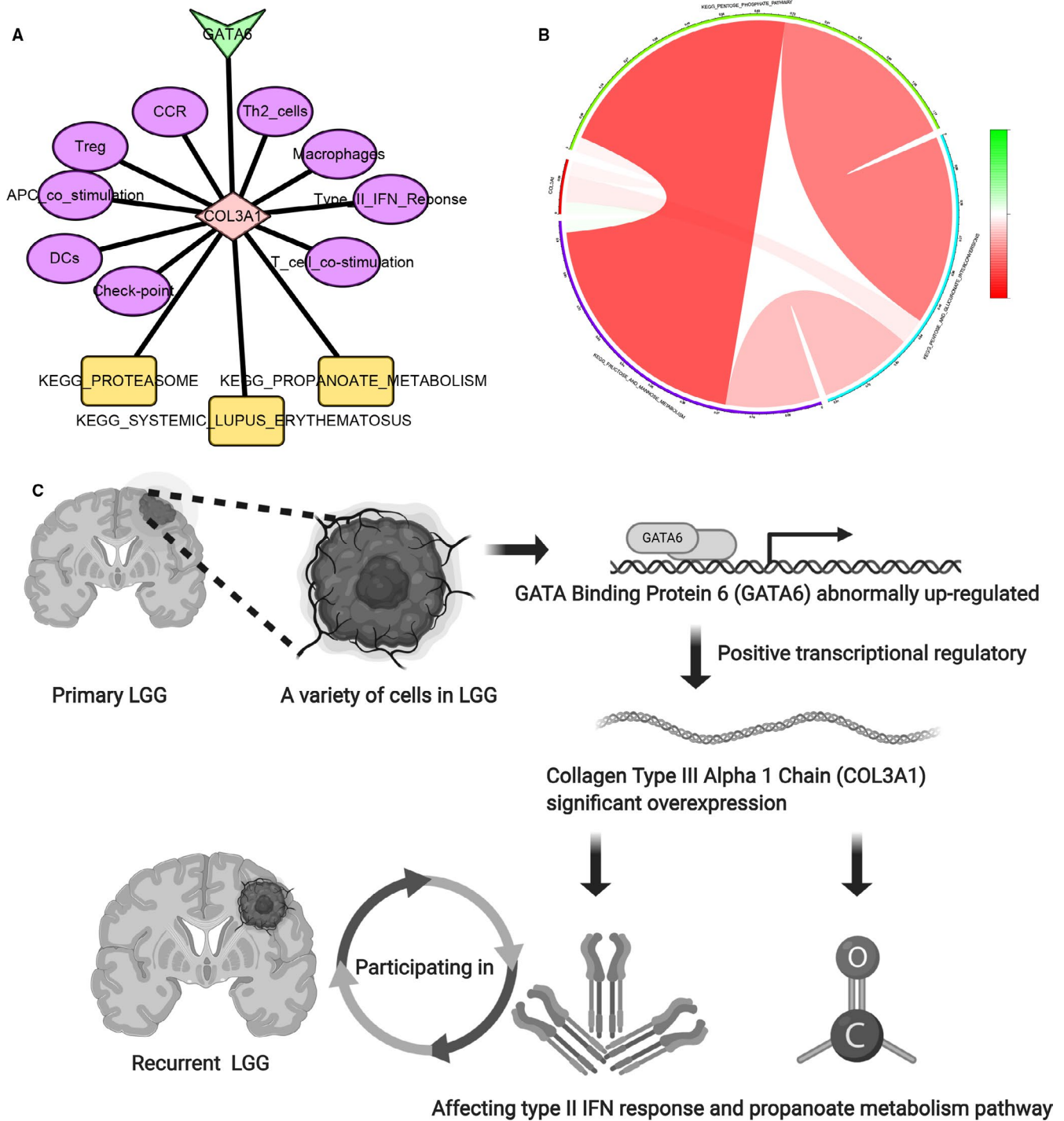


FIGURE 7 Network of TF, immune gene, immune cells/pathways and KEGG pathways (A), arrow represents TF, diamond represents immune genes, ellipses represent immune cells/pathways and rectangles represent KEGG pathways. Co-expressed circle plot of KEGG pathways (B). Schematic diagram of all scientific hypothesis about COL3A1 Regulated by GATA6 affects type II IFN response and propanoate metabolism in the recurrence of lower grade glioma based on analysis in silico (C). KEGG, Kyoto Encyclopedia of Genes and Genomes; TF, transcription factor

latter will enter the tricarboxylic acid cycle to supply energy or enhance lipogenesis. Based on Pathway Card (<https://pathcards.genecards.org/>), the main genes in propanoate metabolism are BCKDHA, BCKDHB, DLD, SUCLG1 and DBT. And the propanoate metabolism can be a carbon source for fatty acid biosynthesis and promote cell multiplication in pancreatic carcinoma.³⁴

In tumour genesis, propanoate metabolism plays a role in lipogenesis, and by down-regulating propanoate metabolism and lipogenesis, leading to inhibition of angiogenesis in vitro/in vivo.³⁵ However, COL3A1 elicits the process of angiogenesis.²⁸ Besides, ICAM1 is induced by IFN- γ , and IFN- γ inhibits angiogenesis.³³ And String database²¹ analysis, COL3A1 regulates biomarkers of propanoate

metabolism (BCKDHA, BCKDHB, DLD, SUCLG1 and DBT) by ICAM1 and VCAM1. Although no study indicated COL3A1 regulated propanoate metabolism, we speculated COL3A1 down-regulated propanoate metabolism to promote lipogenesis, thus promoting angiogenesis in tumour genesis and recurrence.

All in all, in our hypothesis, GATA6 was the most significant TF and COL3A1 was the most significant DEIG. And COL3A1 was positively regulated by GATA6, and have an influence on recurrence of LGG through two potential downstream pathways: type II IFN response and propanoate metabolism. However, the inherent vice of analysis in silico was bias base on different platform, and our scientific hypothesis only based on bioinformation instead of mechanism exploring.

Therefore, multidimensional validation was applied to validate our hypothesis in different databases. Additionally, a serial experiment according to some studies to validate our hypothesis will be launched.³⁶ Firstly, expression level of GATA6, COL3A1 and biomarker of type II IFN response and propanoate metabolism will be detected by immunohistochemistry (IHC) in normal brain tissue, adjacent tissue and LGG tissue. Secondly, gain/loss of function experiment was utilized to explore the regulatory relationship between these genes and phenotype of recurrence. Thirdly, to explore direct interplay of COL3A1 and GATA6, COL3A1 and type II IFN response, COL3A1 and propanoate metabolism, chromatin immunoprecipitation, luciferase reporter assay, RNA immunoprecipitation and co-immunoprecipitation will be applied. Besides, to explore the regulatory relationship, gain/loss of function experiment was utilized. Ultimately, the indirect regulatory relationship including GATA6-COL3A1-type II IFN response and GATA6-COL3A1-propanoate metabolism was validated by rescue assays. All our effort might offer new biomarkers to prognosis the recurrence in LGG.

Based on analysis in silico and multidimensional validation, we proposed that COL3A1 was positively regulated by GATA6, and by affecting type II IFN response and propanoate metabolism, COL3A1 involved in recurrence LGG. And the prognosis value of GATA6, COL3A1, type II IFN response and propanoate metabolism might play a role in recurrence mechanism.

ACKNOWLEDGEMENTS

We thank the The Cancer Genome Atlas (TCGA) team for using their data.

CONFLICT OF INTEREST

The authors declare that there is no potential conflict of interest and no commercial or financial relationships in this study.

AUTHOR CONTRIBUTIONS

Conception/design: Runzhi Huang, Zhenyu Li, Xiaolong Zhu, Penghui Yan, Dianwen Song, Huabin Yin, Peng Hu, Ruoyi Lin, Shengyu Wu, Tong Meng, Jie Zhang, Zongqiang Huang. Collection and/or assembly of data: Runzhi Huang, Zhenyu Li, Xiaolong Zhu, Penghui Yan, Dianwen Song, Huabin Yin, Peng Hu, Ruoyi Lin, Shengyu Wu. Data analysis and interpretation: Runzhi Huang, Zhenyu Li, Xiaolong Zhu, Penghui

Yan, Dianwen Song, Huabin Yin, Peng Hu, Ruoyi Lin, Shengyu Wu, Tong Meng, Jie Zhang, Zongqiang Huang. Manuscript writing: Runzhi Huang, Zhenyu Li, Xiaolong Zhu, Penghui Yan/Tong Meng, Jie Zhang, Zongqiang Huang. Final approval of manuscript: Runzhi Huang, Zhenyu Li, Xiaolong Zhu, Penghui Yan, Dianwen Song, Huabin Yin, Peng Hu, Ruoyi Lin, Shengyu Wu, Tong Meng, Jie Zhang, Zongqiang Huang.

ETHICAL APPROVAL

The study was approved by the Ethics Committee of First Affiliated Hospital of Zhengzhou University. The authors declare that there is no conflict of interests.

DATA AVAILABILITY STATEMENT

The datasets generated and/or analysed during the current study are available in the in the Supplementary Material and TCGA-LGG program (<https://portal.gdc.cancer.gov>).

ORCID

Zongqiang Huang  <https://orcid.org/0000-0002-2787-1629>

REFERENCES

- Gao Y-F, Mao X-Y, Zhu T, et al. COL3A1 and SNAP91: novel glioblastoma markers with diagnostic and prognostic value. *Oncotarget*. 2016;7(43):70494-70503.
- Li X, Meng Y. Survival analysis of immune-related lncRNA in low-grade glioma. *BMC Cancer*. 2019;19(1):813.
- Zhou M, Zhang Z, Zhao H, Bao S, Cheng L, Sun J. An immune-related six-lncRNA signature to improve prognosis prediction of glioblastoma multiforme. *Mol Neurobiol*. 2018;55(5):3684-3697.
- Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin*. 2020;70(1):7-30.
- Yu W, Ren X, Hu C, et al. Glioma SOX2 expression decreased after adjuvant therapy. *BMC Cancer*. 2019;19(1):1087.
- Jansen E, Hamisch C, Ruess D, et al. Observation after surgery for low grade glioma: long-term outcome in the light of the 2016 WHO classification. *J Neurooncol*. 2019;145(3):501-507.
- Keles GE, Lamborn KR, Berger MS. Low-grade hemispheric gliomas in adults: a critical review of extent of resection as a factor influencing outcome. *J Neurosurg*. 2001;95(5):735-745.
- Charoentong P, Finotello F, Angelova M, et al. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep*. 2017;18(1):248-262.
- 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.
- Rhodes DR, Yu J, Shanker K, et al. ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia*. 2004;6(1):1-6.
- Goswami CP, Nakshatri H. PROGeneV2: enhancements on the existing database. *BMC Cancer*. 2014;14:970.
- Chandrashekar DS, Bashel B, Balasubramanya SAH, et al. UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses. *Neoplasia*. 2017;19(8):649-658.
- Vasaikar SV, Straub P, Wang J, Zhang B. LinkedOmics: analyzing multi-omics data within and across 32 cancer types. *Nucleic Acids Res*. 2018;46(D1):D956-D963.
- Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012;2(5):401-404.

15. Consortium G. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science (New York, NY)*. 2015;348(6235):648-660.
16. Goldman M, Craft B, Swatloski T, et al. The UCSC Cancer Genomics Browser: update 2015. *Nucleic Acids Res*. 2015;43(D1):D812-D817.
17. Ghandi M, Huang FW, Jane-Valbuena J, et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature*. 2019;569(7757):503-508.
18. Papatheodorou I, Fonseca NA, Keays M, et al. Expression Atlas: gene and protein expression across multiple studies and organisms. *Nucleic Acids Res*. 2018;46(D1):D246-D251.
19. Uhlen M, Fagerberg L, Hallstrom BM, et al. Proteomics. Tissue-based map of the human proteome. *Science (New York, NY)*. 2015;347(6220):1260419.
20. Hu H, Mu Q, Bao Z, et al. Mutational landscape of secondary glioblastoma guides MET-targeted trial in brain tumor. *Cell*. 2018;175(6):1665-1678.e18.
21. Snel B, Lehmann G, Bork P, Huynen MA. STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Res*. 2000;28(18):3442-3444.
22. Zhou M, Zhao H, Xu W, Bao S, Cheng L, Sun J. Discovery and validation of immune-associated long non-coding RNA biomarkers associated with clinically molecular subtype and prognosis in diffuse large B cell lymphoma. *Mol Cancer*. 2017;16(1):16.
23. Kamnasaran D, Guha A. Expression of GATA6 in the human and mouse central nervous system. *Brain Res Dev Brain Res*. 2005;160(1):90-95.
24. Shureiqi I, Jiang W, Fischer SM, et al. GATA-6 transcriptional regulation of 15-lipoxygenase-1 during NSAID-induced apoptosis in colorectal cancer cells. *Cancer Res*. 2002;62(4):1178-1183.
25. Skiriute D, Vaitkiene P, Saferis V, et al. MGMT, GATA6, CD81, DR4, and CASP8 gene promoter methylation in glioblastoma. *BMC Cancer*. 2012;12:218.
26. Kamnasaran D, Qian B, Hawkins C, Stanford WL, Guha A. GATA6 is an astrocytoma tumor suppressor gene identified by gene trapping of mouse glioma model. *Proc Natl Acad Sci U S A*. 2007;104(19):8053-8058.
27. Tian F, Li D, Chen J, et al. Aberrant expression of GATA binding protein 6 correlates with poor prognosis and promotes metastasis in cholangiocarcinoma. *Eur J Cancer*. 2013;49(7):1771-1780.
28. Gao YF, Zhu T, Chen J, Liu L, Ouyang R. Knockdown of collagen alpha-1(III) inhibits glioma cell proliferation and migration and is regulated by miR128-3p. *Oncol Lett*. 2018;16(2):1917-1923.
29. Deng T, Gong Y-Z, Wang X-K, et al. Use of genome-scale integrated analysis to identify key genes and potential molecular mechanisms in recurrence of lower-grade brain glioma. *Med Sci Monit*. 2019;25:3716-3727.
30. Bhat MY, Solanki HS, Advani J, et al. Comprehensive network map of interferon gamma signaling. *J Cell Commun Signal*. 2018;12(4):745-751.
31. Song JH, Wang CX, Song DK, Wang P, Shuaib A, Hao C. Interferon gamma induces neurite outgrowth by up-regulation of p35 neuron-specific cyclin-dependent kinase 5 activator via activation of ERK1/2 pathway. *J Biol Chem*. 2005;280(13):12896-12901.
32. Qian J, Wang C, Wang B, et al. The IFN-gamma/PD-L1 axis between T cells and tumor microenvironment: hints for glioma anti-PD-1/PD-L1 therapy. *J Neuroinflammation*. 2018;15(1):290.
33. Mitra-Kaushik S, Harding J, Hess J, Schreiber R, Ratner L. Enhanced tumorigenesis in HTLV-1 tax-transgenic mice deficient in interferon-gamma. *Blood*. 2004;104(10):3305-3311.
34. Lee JH, Cho Y-R, Kim JH, et al. Branched-chain amino acids sustain pancreatic cancer growth by regulating lipid metabolism. *Exp Mol Med*. 2019;51(11):146.
35. Hussain A, Qazi AK, Mupparapu N, et al. Modulation of glycolysis and lipogenesis by novel PI3K selective molecule represses tumor angiogenesis and decreases colorectal cancer growth. *Cancer Lett*. 2016;374(2):250-260.
36. Qi Y, Yu J, Han W, et al. A splicing isoform of TEAD4 attenuates the Hippo-YAP signalling to inhibit tumour proliferation. *Nat Commun*. 2016;7(1):ncomms11840.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Huang R, Li Z, Zhu X, et al. Collagen Type III Alpha 1 chain regulated by GATA-Binding Protein 6 affects Type II IFN response and propanoate metabolism in the recurrence of lower grade glioma. *J Cell Mol Med*. 2020;24:10803-10815. <https://doi.org/10.1111/jcmm.15705>