

Identification of immune prognostic signature of lung carcinoma based on Genome-wide immune expression profile

Bo Ling (✉ lingbo268@163.com)

Youjiang Medical University for Nationalities

Guangbin Ye

Youjiang Medical University for Nationalities

Qihua Zhao

Youjiang Medical University for Nationalities

Yan Jiang

Medical College of Guangxi University

Lingling Liang

Youjiang Medical University for Nationalities

Qianli Tang

Youjiang Medical University for Nationalities

Research article

Keywords: lung carcinoma; immune checkpoint genes; immune subtypes; immune gene expression profile; prognosis

Posted Date: March 3rd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-15898/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background : Lung cancer is one of the most common types of cancer with low early diagnosis rate and poor prognosis. The integration of immune checkpoint gene expression data and patient prognosis information can help identify the immune subtypes of lung cancer and provide reference for individualized gene immunotherapy in patients with lung cancer. **Methods :** The data of immune gene expression for lung cancer patients were obtained from TCGA and GEO databases. The relationship between the expressions of 45 immune checkpoint genes (ICGs) and prognosis were analysed. In the other hand, the correlation between the expressions of 45 biomarkers , tumor mutation load (TMB), MMRs, neoantigens and other immunotherapy biomarkers were been identified. Ultimately, prognosis-related ICGs were combined with IDO1, CD274, and CTLA4 to divide lung cancer immune subgroups and the prognostic differences between lung cancer immune subgroups were identified. **Results:** Based on TCGA database and GEO database, 9 and 11 ICGs were obtained respectively, which were closely related to prognosis. There was a certain synergistic relationship between them. The expression of CD200R1 had a significant negative correlation with TMB and neoantigens. CD200R1 showed a significant positive correlation with CD8A, CD68 and GZMB genes, indicating that it may cause the expression disorder of adaptive immune resistance pathway genes. Based on CD200R1 and combination with IDO1, CD274 and CTLA4, the group with high expression of CD200R1 and low expression of IDO1, CD274 and CTLA4 had the best prognosis among the immune subtypes. **Conclusion :** Our research provides a method of integrating immune checkpoint gene expression profile and clinical prognosis information to identify immune subtypes of lung cancer, which can provide a unique reference for gene immunotherapy of lung cancer patients.

Background

Lung cancer accounts for more than one-tenth of all cancer cases worldwide and is one of the most common types of cancer [1], smoking is a major risk factor for lung cancer. Exposure to high carcinogens in tobacco smoke can induce the accumulation of mutation load in lung cells [2–3]. The prognosis of lung cancer is rather poor, with a 5-year survival rate of 18% [4]. The prognosis of lung cancer is related to the clinical pathological characteristics of the patients, and the early diagnosis rate is low. Therefore, the formation and pathological grade of lung cancer could not be accurately judged, thus limiting the operation [5]. At the time of diagnosis of lung cancer, most patients were already in the stage of extensive metastasis. The main therapeutic strategy is to reduce the malignant phenotype by suppressing the oncogenic gene pathway. Although response to chemotherapy and radiotherapy in the early stage is relatively good, this treatment often leads to drug resistance in the treatment of lung cancer. In view of this situation, the treatment is relatively limited [6–7].

The progress of biomedical big data, including the mature technology of new generation sequencing and multi-generation sequencing, has deepened our understanding of lung cancer development, cellular heterogeneity and interaction at the precise molecular level [8], on which the immunotherapy development has shown great potential in treating various malignant tumors [9]. At present, monoclonal antibody targeted drugs are innovating the treatment of lung cancer, especially the immunotherapy developed with PD-1 and CD274 as targets has achieved some great results in clinical trials of lung cancer treatment [10–11]. Currently, these targeted drugs can be used as first-line treatment or second-line drugs for lung cancer since there are

over 50% of tumor cells in first-line patients express PD-L1 [12]. However, this therapy does not work for all patients due to the dynamic and isomeric expressions, which suggests other therapeutic targets will be needed for specific patients [13–14]. On the whole, the effect of most targeted monoclonal antibodies is still limited among all lung cancer patients for the reason that beneficial responses are observed from some patients, while many others are not [15]. This indicates that not all patients show the same sensitivity to therapeutic targets because of tumor heterogeneity. Immunotherapy is not a panacea for all patients. So it is highly necessary to further differentiate tumor immune subtypes for improving clinical trial design and identifying patients who may benefit from immunotherapy.

Traditional cancer research usually focuses on the role of a particular target gene in cancer. Because the occurrence and development of cancer is rather complex, it involves the abnormality of multiple genes and multiple signaling pathways. Therefore, studying the role of a single gene in cancer has its limitations [16]. At present, biological studies have entered the multi-omics data age, which provides possibility for understanding and studying tumors from genome level. The establishment of several large cancer databases of TCGA and GEO enables researchers to obtain the gene expression data of multiple immune checkpoints and the corresponding clinical data of patients. On this basis, the integrated analysis can divide different immune subtypes in lung cancer samples in more detail and guide the "individualized" treatment of patients [17–18]. In the existing reports related to immunotherapy for lung cancer, most of them are drug design studies based on a single immune checkpoint gene as a target. There is still a lack of research on gene expression profile of multiple immune check points [19]. Therefore, based on the gene expression data of 47 immune checkpoints such as PD1 (PDCD1), PD-L1 (CD274) and IDO1, the immune subgroup and prognosis of 761 lung cancer samples from TCGA and GEO database were identified.

In our research, the expression of 45 immune checkpoint genes (ICGs) and its relationship with prognosis were analyzed. Then the relationship between immunotherapy biomarkers and ICGs, such as mismatch repair defect (MMRd) and tumor mutation load (TMB), was studied by integrated cell mutation data to determine the relationship between the expression of ICGs (PD1/PD-L1 and CTLA4) in biomarkers, which has been widely used in immunotherapy, and other biomarkers, which are not used in immunotherapy. Ultimately, the relationship between ICGs expression and immune activation related signature genes was studied in order to determine the relationship between immune activation and immunosuppression in tumor microenvironment, and to provide reference for individualized gene immunotherapy in patients with lung cancer.

Methods

Data downloading

47 genes of immune checkpoints were downloaded (S_ Table 1). The latest clinical follow-up informations were downloaded on 14 June, 2019 from TCGA GDC API. which were contained 515 samples with RNA-seq data. GSE31210 data of microarray expression was downloaded from NCBI, it was contained 246 samples with clinical features.

Table 1

Pre-processed clinical information of the two data sets

Characteristic	TCGA	Characteristic	GSE31210
Event		Event	
Alive	314	Alive	191
Dead	178	Dead	35
Sex		Sex	
Male	230	Male	105
Female	262	Female	121
Survival times		Survival times	
33 ~ 1476	417	221 ~ 950	32
1476 ~ 2919	58	950 ~ 1679	72
2919 ~ 4362	11	1679 ~ 2408	86
4362 ~ 5805	3	2408 ~ 3137	28
5805 ~ 7248	2	3137 ~ 3863	8
Pathologic N			
N0	319	-	-
N1	92	-	-
N2	68	-	-
N3	2	-	-
NX	11	-	-
Pathologic T			
T1	164	-	-
T2	264	-	-
T3	43	-	-
T4	18	-	-
TX	3	-	-
Pathologic M			
M0	362	-	-
M1	24	-	-
MX	142	-	-

Characteristic	TCGA	Characteristic	GSE31210
TNM Stage		Pathological stages	
I	261	I	168
II	114	II	58
III	79	III	0
IV	25	IV	0
Smoke		Smoke	
Non- smoking	68	Non- smoking	115
Smoker	408	Smoker	111

Data preprocessing

The following pre-steps were conducted in the processing of 515 RNA-seq data samples:1) Samples without clinical information or OS < 30 days were removed;2) The data of normal tissue samples was removed;3) Remove genes whose FPKM value is equal to 0; thus more than half of the samples were removed.

Furthermore, GSE31210 data were pre-processed as follows:1) The data of normal tissue samples was removed and we retained the only primary tumor data; 2) Year/month-based OS data were converted into days-based; 3) The microarray probe was mapped to human gene SYMBOL by using bioconductor package; 4) Only the expression profile of immune-related genes was retained. Statistical information for the three data sets was shown in Table 1.

Relationship between ICGs expression and prognosis

The ICGs with expression in TCGA dataset were screened. The patients were divided into groups based on different levels of expression. The relationship between these ICGs with expression and patient prognosis was analyzed using univariate cox regression analysis (log rank $p < 0.05$). The expression and prognostic patterns of the ICGs were analyzed by the same method in GEO database.

Relationship between ICGs and other immune checkpoint biomarkers

Spearman method was used to evaluate the correlation between TMB and ICGs. The relationship between neoantigens and ICGs expression was further analyzed based on the somatic cell mutation data of TCGA.

Relationship between prognosis and subtypes defined by ICGs

Therefore, the high expression group (H) and low expression group (L) were divided according to the horizontal density distribution of these genes. Taking the deviation from the first main distribution interval (density peak) as the threshold, the genes were divided into high expression group and low expression group. The High/Low expression groups of IDO1, CD274, CTLA4 and CD8A were integrated, respectively. Then the HCC samples were divided into four categories to analyze the survival of the three gene classification samples.

Statistical analysis

For comparison between the two groups, the statistical significance of the normal distribution variables was estimated by the nonpaired student's t test, and the non-normal distribution variables were analyzed using the Mann-Whitney U test. For comparisons between more than two groups, Kruskal-Wallis test and one-way ANOVA were used as nonparametric and parametric methods, respectively [20]. The correlation coefficients were calculated by Spearman correlation analysis. Kaplan-Meier method was used to generate a survival curve for the subgroups in each data set, and logrank test was used to determine the statistical significance at the level of $P < 0.05$. All these analyses were performed using R 3.4.3, based on defaulted parameters unless otherwise specified.

Results

The relationship between ICGs and prognosis

The expression of ICGs

Firstly, a total of 45 ICGs with expression in the TCGA dataset were selected. Depending on their expression level, the 45 ICGs were obviously divided into three expression patterns: high expression group (red) was represented by CD44, CD40, TNFRSF14 and NRP1, with generally high expression level in all samples; medium expression group (green) was represented by IDO1, LAG3 and PDCD1, with distinct difference in different samples; and low expression group (blue) was represented by KIR3DL1, BTNL2 and so forth, with lower expression level in most samples (Fig. 1).

Prognostic analysis of ICGs

Using univariate COX regression analysis, we calculated the relationship between the 45 ICGs with expression quantity in the TCGA data set and prognosis, and obtained altogether 9 genes that were significantly correlated with prognosis: TNFRSF14, CD70, TNFRSF14, CD27, LAIR1, CTLA4, CD28, CD40LG and CD200R1 (Fig. 2A, log rank $p < 0.05$). It could be found that these nine genes were mainly positively correlated and showed obvious aggregation effect (Fig. 2B) through correlation analysis of ICGs expression levels, indicating a coordinated expression relationship between ICGs.

ICGs analysis on the GEO data set

Among the 47 selected ICGs on the GSE31210 data set, there were 44 ICGs with expression, and their expression patterns were also significantly clustered into three categories: high expression group, medium expression group, low expression group (Fig. 3A). Moreover, we found that the genes in the three expression groups of GSE31210 were highly consistent with those in the three expression groups of TCGA. For instance, LGALS9, TNFRSF14, CD40 and CD44 showed high expression levels in the two data sets. The expression levels of HAVCR2, LAIR1, TNFRSF9 and other genes were medium in both data sets. Such low expression genes in GEO as HHLA2, CD200R1, CD244, CD28 and BTNL2 showed the same low expression level in TCGA data set.

Univariate survival analysis of the relationship between 44 ICGs and overall survival (OS) revealed that 11 ICGs are auspiciously correlated with the prognosis of OS (Fig. 3B), and there was a significant correlation of CD40LG gene with better prognosis in both TCGA and GSE31210 data sets (HR < 1, log rank $p < 0.05$). On the GSE31210 data set, ICGs expression levels were also largely in a positive correlation and had an obvious aggregation effect (Fig. 3C), showing no difference with TCGA results.

Relationship between ICGs and biomarker for other immune checkpoints treatment

The relationship between ICGs and TMB

Our study calculated the TMB of LAUD according to the somatic mutation data of TCGA, and removed the intron interval and the mutation annotated as silent. First, we selected 9 ICGs with expression quantity that were strongly correlated with OS prognosis, and used Spearman method to evaluate the correlation between TMB and the 9 ICGs (Shapiro.test $p < 2.2e-16$). We observed highly negative correlations between TMB and CD40LG, CD200R1, TNFRSF14 and TNFSF14 expression (Fig. 4, $R < 0$ and $FDR < 0.005$). The TNFSF14 was the potential factor, that was causing unfavourable prognosis. Thus, we speculated that the high expression TNFSF14 corresponds to the lower tumor mutation burden. And low TMB generally did not apply to immune checkpoint inhibitor treatment.

The relationship between ICGs and MMRs

MMRs were intracellular mismatch repair mechanisms, and the function loss of key genes in this mechanism will lead to irreparable DNA replication errors, producing much more somatic cell mutations. Here we used LUAD somatic mutation data of TCGA to evaluate the relations between the five MMRs genes: MLH1, MSH2, MSH6, PMS2, EPCAM mutations and ICGs expression. It turned out that there were 34 samples with different number of MMRs mutant gene, and they were positively related to the expression levels of ICGs and had higher correlation with TNFRSF8 and CD70 (Fig. 5). While the CD70, which was a prominently prognosis-correlated gene, could be regarded as potential prediction of MMR gene mutations.

The relationship between ICGs and neoantigens

The occurrence of tumor somatic mutations in protein-coding genes produced incorrect proteins (neoantigens), which could be presented by the major histocompatibility class I complex (MHC I). These new peptides induced anti-tumor adaptive immune responses by binding to TCR (recognized by TCR). Based on TCGA somatic mutation data, we further analyzed the relationship between neoantigens and the 9 ICGs expressions that were largely affected prognosis. We found an existence of high negative correlation between the four genes of CD40LG, CD200R1, TNFSF14 and TNFRSF14 and the expression of neoantigens (Fig. 6, $R < 0$ and $FDR < 0.05$) This results consisted with aforementioned results, whereas TMB was extremely negatively correlated with the expression of the four genes.

The relationship between ICGs and of adaptive immune-resistance pathway genes

CD8 T cells can produce IFN γ and lead to up-regulated expression (activation of immune pathways) of adaptive immune-resistance pathway genes (including PD-1/PD-L1 axis and IDO1, etc.). Therefore, we had analyzed the correlation between CD8A, GZMB, CD68, NOS2 and ICGs. In the genes of adaptive immune-resistance pathways, there were four genes except NOS2 with very and mostly positively correlated with ICGs expressions (Fig. 7A). The significance test of correlation coefficients showed that there were at least half of significant correlations between these genes ($p < 1e-3$, Fig. 7B).

We had also analyzed the expression relationship between adaptive immune-resistance pathway's gene and ICGs by using GSE31210 data. This relationship was positively correlated with (72.9%), as well a significant positive correlation was shown in CD8A and GZMB (S1_Figure), which coincided with the result of TCGA data set.

The relationship between ICGs and clinical features

Univariate survival analysis had identified 9 ICGs that were highly related to OS prognosis (need ref.). Combined with the clinical information provided by TCGA, we analyzed the relationship between these 9 genes and clinical features. Due to the large deviation in the distribution of the samples of M stage of LUAD (M0/M1:326/24), this study focused on the analysis of the expression patterns of these 9 ICGs on stages of T, N and Stage. From the expression level, these 9 ICGs were divided into three high expression group and six low expression group (S2_Figure). These genes displayed relatively consistent expression patterns on stages of T, N and Stage, that is, the expression level of tumor samples in early stage is higher than that of advanced tumor samples (Fig. 8A-C). Significance test indicates that TNFRSF14, CD27 and CD200R1 genes show obvious differences on T, Grade and Stage stages (kw test $p < 0.05$). No significant relationship is observed between sex, new events and the 9 ICGs (Supplementary Fig. 3).

The relationship between ICGs subtypes and prognosis

Although PD1 (PDCD1), PD-L1 (CD274) and IDO1 are important genes in immune checkpoints and important biomarkers. Univariate analysis showed that there was no significant prognostic relationship between them. Based on the comprehensive analysis above, we found that CD200R1 was not only strongly associated with the prognosis of the TCGA samples, but also has highly negative correlations with TMB and neoantigens, and positive correlations with CD8A, CD68 and GZMB. This indicates that CD200R1 can cause expression dysregulation of adaptive immune-resistance pathway genes. Therefore we analyzed the relationship between such gene combinations as CD200R1, PD-L1 (CD274), IDO1 and prognosis.

Integrating the H/L expression groups of IDO1, CD274, CTLA4 and CD200R1, respectively, we sorted the LUAD samples into four cluster according to the median of gene expression. The survival analysis of the three gene classification samples showed that there exist significant differences in the prognosis of OS among the four cluster of samples. They all revealed that the group with the high expression of CD200R1 and the low expressions including IDO1, CD274 and CTLA4 has the best prognosis. The group with low expression of CD200R1 and low expression of IDO1, CD274 and CTLA4 had the worst prognosis (Fig. 9A-C). Significant differences can be seen through prognostic analysis between the two cluster with the best and the worst prognosis (Fig. 9D-F).

On the GSE31210 data set, we also observed that the three genes and several expression patterns of CD200R1 were significantly correlated with different prognosis, and that CD200R1 was highly expressed and IDO1, CD274, and CTLA4 low expression groups had the best prognosis (S4_Figure).

Discussion

This study aims to identify the relationship between immune subgroup and prognosis of lung cancer samples based on gene expression data from multiple immune checkpoints. The gene expression data from TCGA and GEO databases were divided into high - medium - low expression groups according to the gene expression levels of 47 immune checkpoints. The gene expression of the three expression groups from the two databases are of high consistency. In other words, genes with either high or low expression in TCGA database are also similarly expressed in GEO database. It indicates a relatively consistent distribution of genes in the collected samples of the two databases. Therefore, the analysis results can be mutually verified in subsequent studies.

Then by using univariate COX regression analysis to screen for genes associated with patients' overall survival, we obtained nine genes significantly related to prognosis from TCGA database and 11 prognosis-related genes from GEO database. Among them, CD40LG, a leukocytic coding gene of the cluster of differentiation 40 ligand - CD40L, is significantly correlated with better prognosis in both databases (HR < 1, log rank $p < 0.05$) [21]. CD40L, which is a type II membrane-related glycoprotein and a member of the tumor necrosis factor (TNF) superfamily [22], is conducive to regulate the immune response and inhibit tumor growth. It produces direct growth-inhibiting effect through apoptosis in CD40L over-expressed malignant tumor cells (such as breast cancer) [23]. The correlation of ICGs expression level in both databases is positive and is showing an obvious aggregation effect, indicating a possibility of coordinated expression relationship between ICGs.

To make clear the relationship between ICGs and other biomarkers for immune checkpoint treatment, we analyzed the relationship between ICGs and tumor mutational burden (TMB), MMRs, and neoantigens, respectively. TMB is a reliable indicator for predicting clinical efficacy of PD-1 inhibitor [24]. The more mutant genes in the tumor tissue, the more likely abnormal proteins will be largely produced. These abnormal proteins are able to activate the body's anti-tumor immune response, improving the sensitivity of immunotherapy. So it is more suitable for patients with high TMB level to continue immunotherapy with PD1 inhibitors [25]. Through analyzing the relationship between ICGs and TMB, we found that the expression of TMB is strongly inversely correlated with that of CD40LG, CD200R1, TNFRSF14 and TNFSF14 ($R < 0$ and $FDR < 0.005$). That is, the high expression of CD40LG, CD200R1, TNFRSF14 and TNFSF14 corresponds to the low expression of TMB. Among them, TNFSF14 is a potential cause of unfavourable prognosis. It is inferred that the high expression of TNFSF14 corresponds to the low tumor mutational burden, so the high expression of TNFSF14 may not be suitable for the treatment of immune checkpoint inhibitors. MMRs is an intracellular mismatch repair mechanism. Its loss of function will lead to irreparable DNA replication errors, thus, leading to the high generation of gene mutations [26]. The 5 MMRs gene mutations evaluated in this study are positively correlated with ICGs expression. Among them, CD70, which is also significantly correlated with prognosis which might be a potential predictor of MMR gene mutations. New proteins produced by gene mutation that may activate the immune system and further trigger the immune system to attack cancer cells are called neoantigens [27]. Our study found that CD40LG, CD200R1, TNFSF14 and TNFRSF14 were also significantly

negatively correlated with the expressions of neoantigens ($R < 0$ and $FDR < 0.05$), which was in line with the extremely correlation of negative expression between TMB and the four genes mentioned above. Our experiment showed that TNFSF14, the unfavourable prognostic gene, was of high expression in patients. While the high expression of TNFSF14 corresponds to low level neoantigens. So in this case, it was not appropriate for patients to be treated with personalized neoantigens [28]. In general, the greater total number of gene mutations (the higher the TMB) in the tumor tissues of patients, the more neoantigens they carry [29]. According to our results, the expressions of CD40LG, CD200R1, TNFSF14 and TNFRSF14 were negatively correlated with TMB and neoantigens. This indicated TMB and neoantigens share the same variation trend, which coincides with what was reported in literature.

In order to reveal the relationship between ICGs and adaptive immune-resistance pathway genes, we analyzed the correlation between CD8A, GZMB, CD68, NOS2 and ICGs. Most of them were found out to be positively and significantly correlated. We obtained the same results in TCGA and GEO databases. This indicated that adaptive immune pathway genes might have a certain regulatory effect on ICGs expressions.

In the relationship between ICGs and clinical features, we found that the expression levels of ICGs in early tumor samples were higher than that in advanced tumor samples, CD200R1 in particular. In the above analysis, we found that CD200R1 were highly correlated with the overall survive of TCGA, and negatively correlated with TMB and neoantigens. In addition, in correlation with the adaptive immune-resistance pathway genes, we found CD200R1 were positively correlated with CD8A, CD68 and GZMB, suggesting that CD200R1 might be the possible cause of expression dysregulation of adaptive immune-resistance pathway genes. Then we analyzed the relationship between IDO1, CD274, CTLA4 and the H/L expression groups and prognosis of CD200R1. based on TCGA database our analysis results showed that prognosis will be the best when it is the group with high expression of CD200R1 and low expressions of IDO1, CD274 and CTLA4, while prognosis will be the worst when it is the group with CD200R1 low expression and low expressions of IDO1, CD274 and CTLA4. Moreover, there was a significant difference between of two groups. It was also confirmed in GEO database that the group with high expression of CD200R1 and low expressions of IDO1, CD274 and CTLA4 had the best prognosis. However, in the previous univariate COX analysis, we did not find a direct correlation between such ICGs as IDO1, CD274 and CTLA4 and patient prognosis. Therefore, integrating the expression levels of PD-I1, IDO1, CTLA4 and CD200R1 could be better evaluated for the immune status of lung cancer patients and explore the prognostic information that cannot be acquired by a single immune checkpoint gene.

However, here are some limitations existing in this study. For example, there is insufficient large-scale lung cancer samples for calculation and analysis, and most of the samples are no more than 1000 [30]. Our research results still need to be verified by more experiments and clinical observation, and on this basis, to develop and design potential immune checkpoint inhibitors [31].

Conclusion

In this study, we analyzed the correlation between ICGs and prognosis, other immune checkpoint biomarker, adaptive immune pathway genes and clinical features based on the gene expressions of immune checkpoints of 761 lung cancer samples from TCGA and GEO databases. We then conducted combinatory analyses between the key gene CD200R1 and ICGs including IDO1, CD274, CTLA4, and identified the prognostic

differences among lung cancer immune subgroup and other groups so as to provide references for personalized treatment for lung cancer patients and new ideas for drug design for immune checkpoints.

Abbreviations

TCGA: The Cancer Genome Atlas

GEO: Gene Expression Omnibus database

ICGs: immune checkpoint genes

TMB: Tumor mutation burden

MMRd: mismatch repair defect

PD-1: Programmed death-1

PD-L1: Programmed death-1 ligand

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

The data used to support the findings of this study are available from the corresponding author on reasonable request

Competing interests

The authors declare that there are no conflicts of interest

Funding

Special Funding for Guangxi Special Experts [NO: Gui Rencai Tongzi [2019] No.13]; Guangxi Medical High-level backbone Talents Training "139" Project [NO: Guiwei Science and Education [2018] No. 22]

Acknowledgments

Thank numerous individuals participated in this study

Authors' contributions

Bo Ling designed the current study. Guangbin Ye and Qihua Zhao analyzed and interpreted the data. Bo Ling, Yan Jiang and Lingling Liang wrote the manuscript. Qianli Tang supervised the study. All authors read and approved the final version of the manuscript and agreed to be accountable for all aspects of the research in ensuring that the accuracy or integrity of any part of the work are appropriately investigated and resolved.

References

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer*. 2015; Mar 01:136(5).
2. Pedersen JH, Tønnesen P, Ashraf H. Smoking cessation and lung cancer screening. *Ann Transl Med*. 2016; Apr:4(8).
3. Cataldo JK, Dubey S, Prochaska JJ. Smoking Cessation: An Integral Part of Lung Cancer Treatment. 2010;78.
4. Parascandola M, Xiao L. Tobacco and the lung cancer epidemic in China. *JAMA Netw Open*. 2019; 03 01:2(3)
5. Ergonul A, Akcam T, Ozdil A, et al. PUB126 The Retrospective Analysis of Asymptomatic Lung Cancer Patients Diagnosed Incidentally. *Journal of Thoracic Oncology*. 2017; 12(1): S1520.
6. Zhang Q, Yang J, Bai J, Ren J. Reverse of non-small cell lung cancer drug resistance induced by cancer-associated fibroblasts via a paracrine pathway. *Cancer Sci*. 2018; Apr:109(4).
7. Yeung K T, More S, Woodward B, et al. Circulating Tumor DNA for Mutation Detection and Identification of Mechanisms of Resistance in Non-Small Cell Lung Cancer. *Molecular Diagnosis & Therapy*. 2017; 21(5).
8. Xu M, Wang DC, Wang X, Zhang Y. Correlation between mucin biology and tumor heterogeneity in lung cancer. *Semin. Cell Dev. Biol*. 2017; 04: 64.
9. Ahn, Myung J, Sun, Jong M, Lee, Se H, et al. EGFR TKI combination with immunotherapy in non-small cell lung cancer. *Expert Opinion on Drug Safety*. 2017; 16(4):465-469.
10. Katz SI, Hammer M, Bagley S, et al. Radiologic pseudo progression during anti-PD1 therapy for advanced non-small cell lung cancer. *Journal of Thoracic Oncology*, 2018: S1556086418305367.
11. Herbst, Roy S. Targeting PD1 and PDL1 in lung cancer treatment: Where are we now? *Journal of Thoracic Oncology*, 2016, 11(2):S8.
12. Loo K, Daud A. Emerging biomarkers as predictors to anti-PD1/PD-L1 therapies in advanced melanoma. *Immunotherapy*, 2016, 8(7):775-784.
13. Kasahara N, Sunaga N, Tsukagoshi Y, Miura Y, Sakurai R, Kitahara S, Yokobori T, Kaira K, Mogi A, Maeno T, et al. Post-treatment Glasgow Prognostic Score Predicts Efficacy in Advanced Non-small-cell Lung Cancer Treated With Anti-PD1. *Anticancer Res*. 2019; Mar:39(3).
14. Merelli B, Massi D, Cattaneo L, Mandalà M. Targeting the PD1/PD-L1 axis in melanoma: Biological rationale, clinical challenges and opportunities. *Crit Rev Oncol Hematol*. 2014; 89(1):140-65.
15. Wang C, Sun W, Wright G, Wang AZ, Gu Z. Inflammation-Triggered Cancer Immunotherapy by Programmed Delivery of CpG and Anti-PD1 Antibody. *Adv Mater*. 2016; 28(40):8912-8920.

16. Xia WX, Yu Q, Li GH, Liu YW, Xiao FH, Yang LQ, Rahman ZU, Wang HT, Kong QP. Identification of four hub genes associated with adrenocortical carcinoma progression by WGCNA. *Peer J*. 2019 Mar 14;7: e6555.
17. Yuan A, Hsiao YJ, Chen HY, Chen HW, Ho CC, Chen YY, Liu YC, Hong TH, Yu SL, Chen JJ, Yang PC. Opposite Effects of M1 and M2 Macrophage Subtypes on Lung Cancer Progression. *Sci Rep*. 2015; Sep 24:5.
18. Busch SE, Hanke ML, Kargl J, Metz HE, MacPherson D, Houghton AM. Lung Cancer Subtypes Generate Unique Immune Responses. *J Immunol*. 2016;197(11):4493-4503.
19. Long L, Zhao C, Ozarina M, Zhao X, Yang J, Chen H. Targeting Immune Checkpoints in Lung Cancer: Current Landscape and Future Prospects. *Clin Drug Investig*. 2019; 4:39(4).
20. Hazra, A and Gogtay N. Biostatistics Series Module 3: Comparing Groups: Numerical Variables. *Indian J Dermatol*. 2016; 61(3).
21. Chen Y, Peubez C, Jayne S, Kocsis-Fodor G, Dyer MJS, Macip S. Differential activation of pro-survival pathways in response to CD40LG/IL4 stimulation in chronic lymphocytic leukaemia cells. *Br. J. Haematol*. 2019; 03:184(5).
22. Wieckowski, Sébastien, Trouche, Nathalie, Chaloin, Olivier, et al. Cooperativity in the Interaction of Synthetic CD40L Mimetics with CD40 and Its Implication in Cell Signaling. *Biochemistry*. 2007; 46(11):3482-3493.
23. Pan W, Gong J, Yang C, Feng R, Guo F, Sun Y, Chen H. Peripheral blood CD40-CD40L expression in human breast cancer. *Ir J Med Sci*. 2013; Dec:182(4).
24. Jafarnejad M, Gong C, Gabrielson E, Bartelink IH, Vicini P, Wang B, Narwal R, Roskos L, Popel AS. A Computational Model of Neoadjuvant PD-1 Inhibition in Non-Small Cell Lung Cancer. *AAPS J*. 2019; 4:21(5).
25. Keith M. Kerr, Marianne C. Nicolson. Non-Small Cell Lung Cancer, PD-L1, and the Pathologist. *Archives of Pathology & Laboratory Medicine*. 2016; 140(3):249-254.
26. Pezza R J, Villarreal M A, Montich G G , et al. Vanadate inhibits the ATPase activity and DNA binding capability of bacterial MutS. A structural model for the vanadate-MutS interaction at the Walker A motif. *Nucleic Acids Research*. 2002; 30(21):4700.
27. Julie Denoeud, Muriel Moser. Role of CD27/CD70 pathway of activation in immunity and tolerance. *Journal of Leukocyte Biology*. 2011; 89(2).
28. Hong-Min Kim, Choon-Soo Jeong, Hye-Sun Choi, et al. LIGHT/TNFSF14 enhances adipose tissue inflammatory responses through its interaction with HVEM. *Febs Letters*. 2011; 585(3): 579-84.
29. Jia Q, Wu W, Wang Y, Alexander PB, Sun C, Gong Z, Cheng JN, Sun H, Guan Y, Xia X, et al. Local mutational diversity drives intratumoral immune heterogeneity in non-small cell lung cancer. *Nat Commun*. 2018; 18:9(1).
30. Chang JT, Lee YM, Huang RS. The impact of the Cancer Genome Atlas on lung cancer. *Transl Res*. 2015; Dec:166(6).
31. Tuma, Rabiya S. Research round-up: AACR-IASLC combine forces on molecular origins of lung cancer. *Oncology Times Uk*. 2010, 7(4):13-14.

Figures

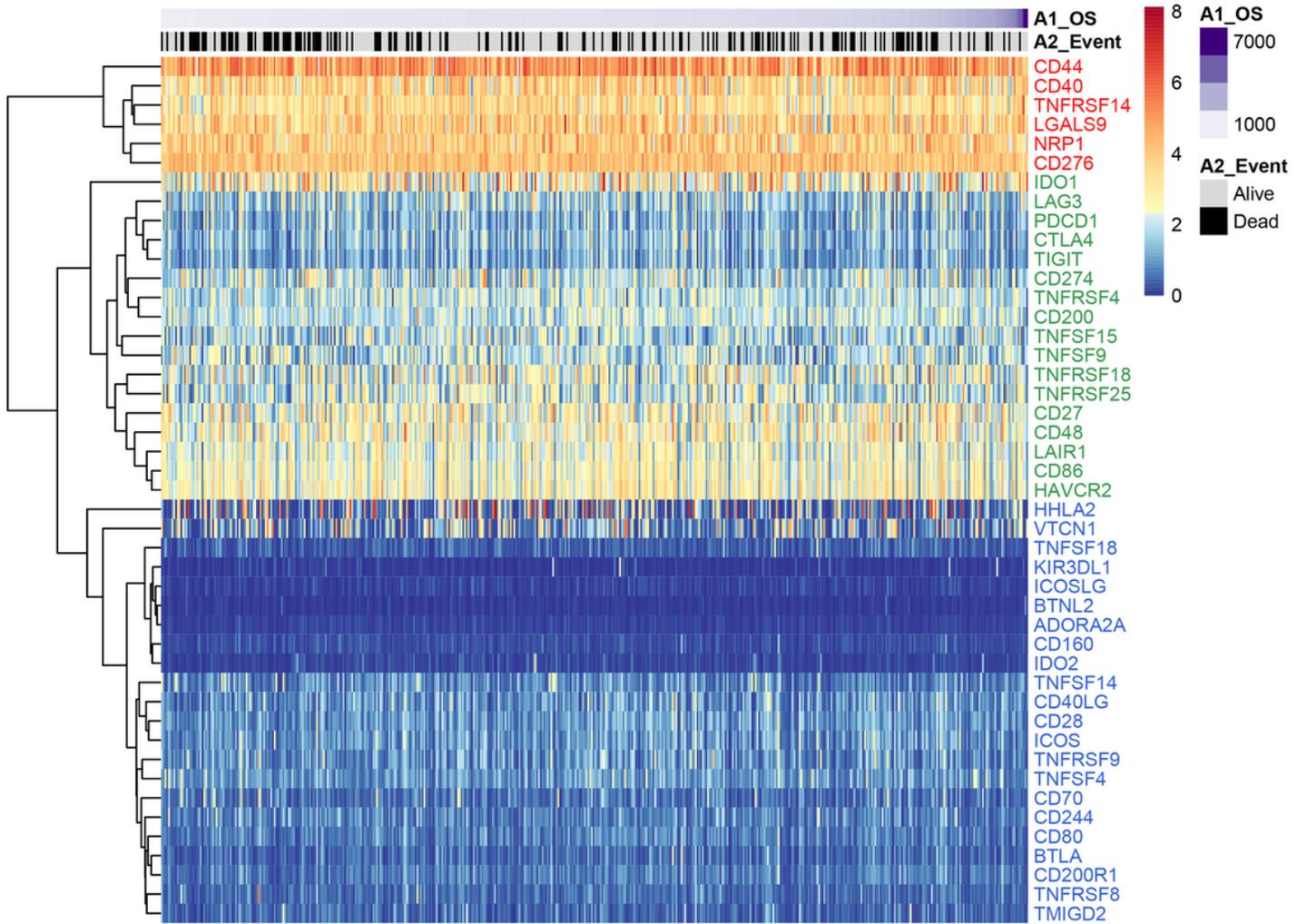


Figure 1

The heatmap of ICGs on TCGA dataset. red: high expression group, green: medium expression group; blue: low expression group.

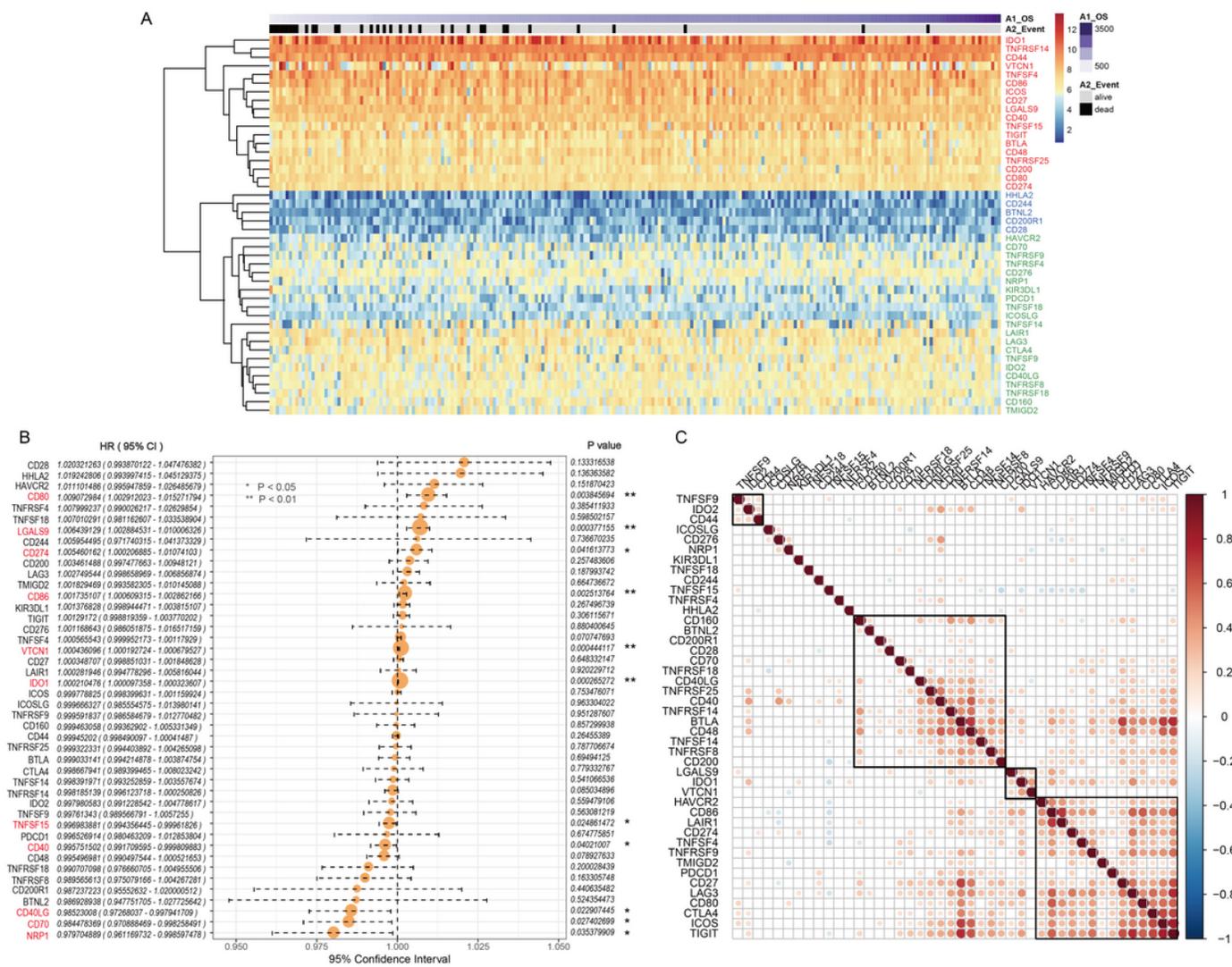


Figure 3

A: ICGs heatmap on GSE31210 data set. red: high expression group, green: medium expression group; blue: low expression group. B: relationship between ICGs expression and prognosis; C: correlation of ICGs expression levels. Description: only the gene pairs with significant correlation test are shown, and the blank shows non-significant correlation.

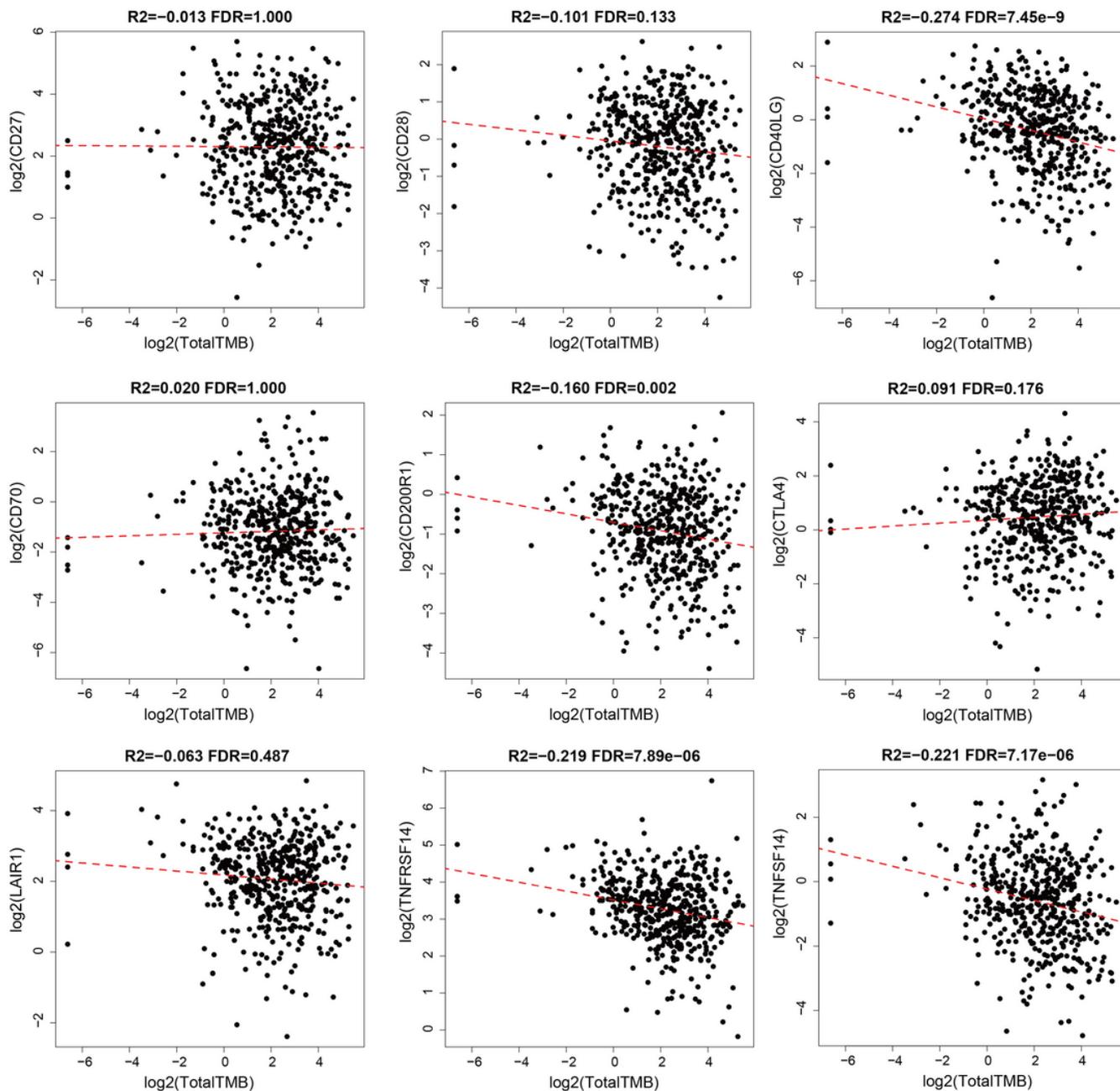


Figure 4

Scatter diagram of IGCs expression levels and TMB. R2 is the correlation coefficient, and FDR is false discovery rate.

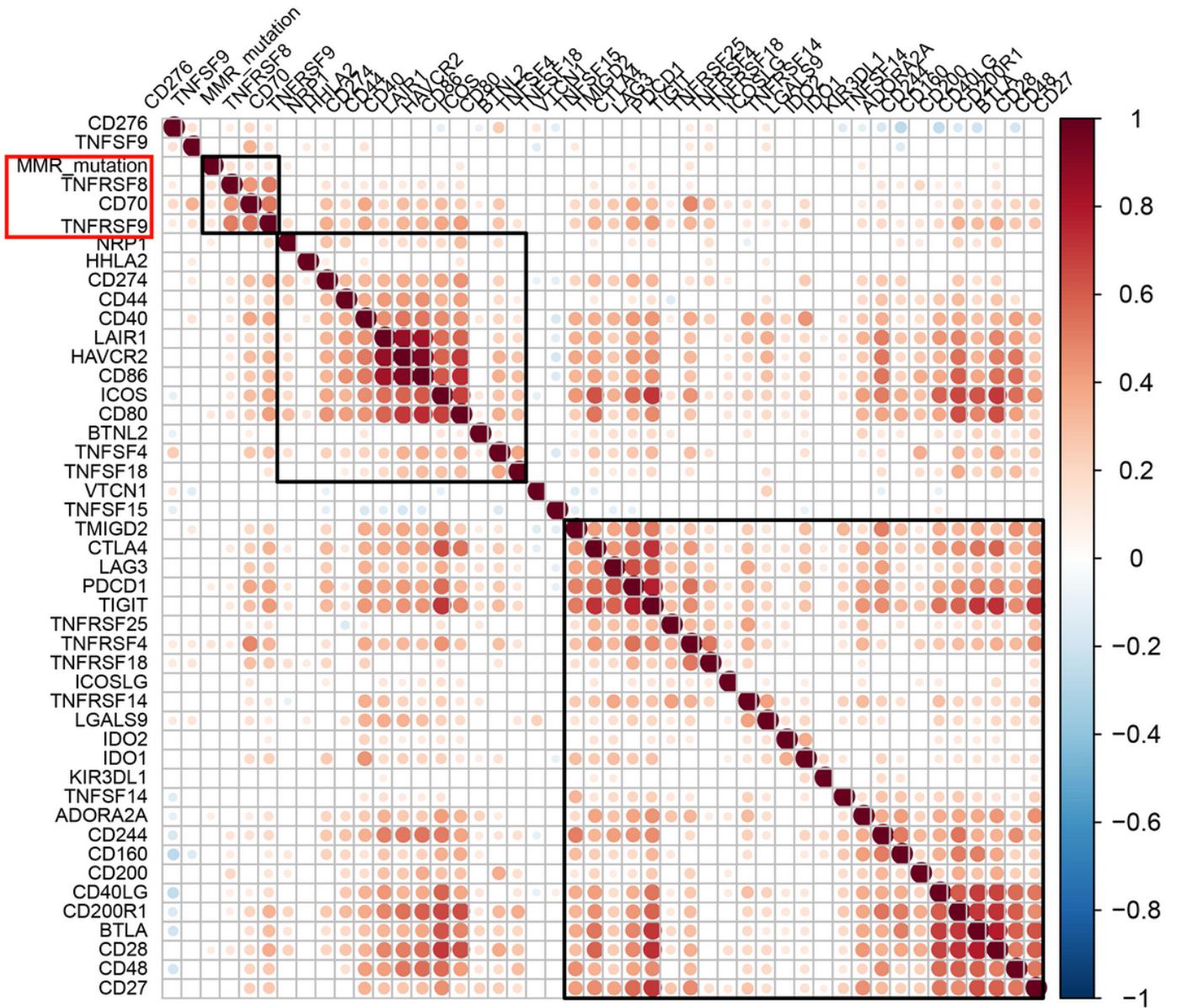


Figure 5

correlation between IGCs expression levels and MMR gene mutation. The larger the point, the stronger the correlation.

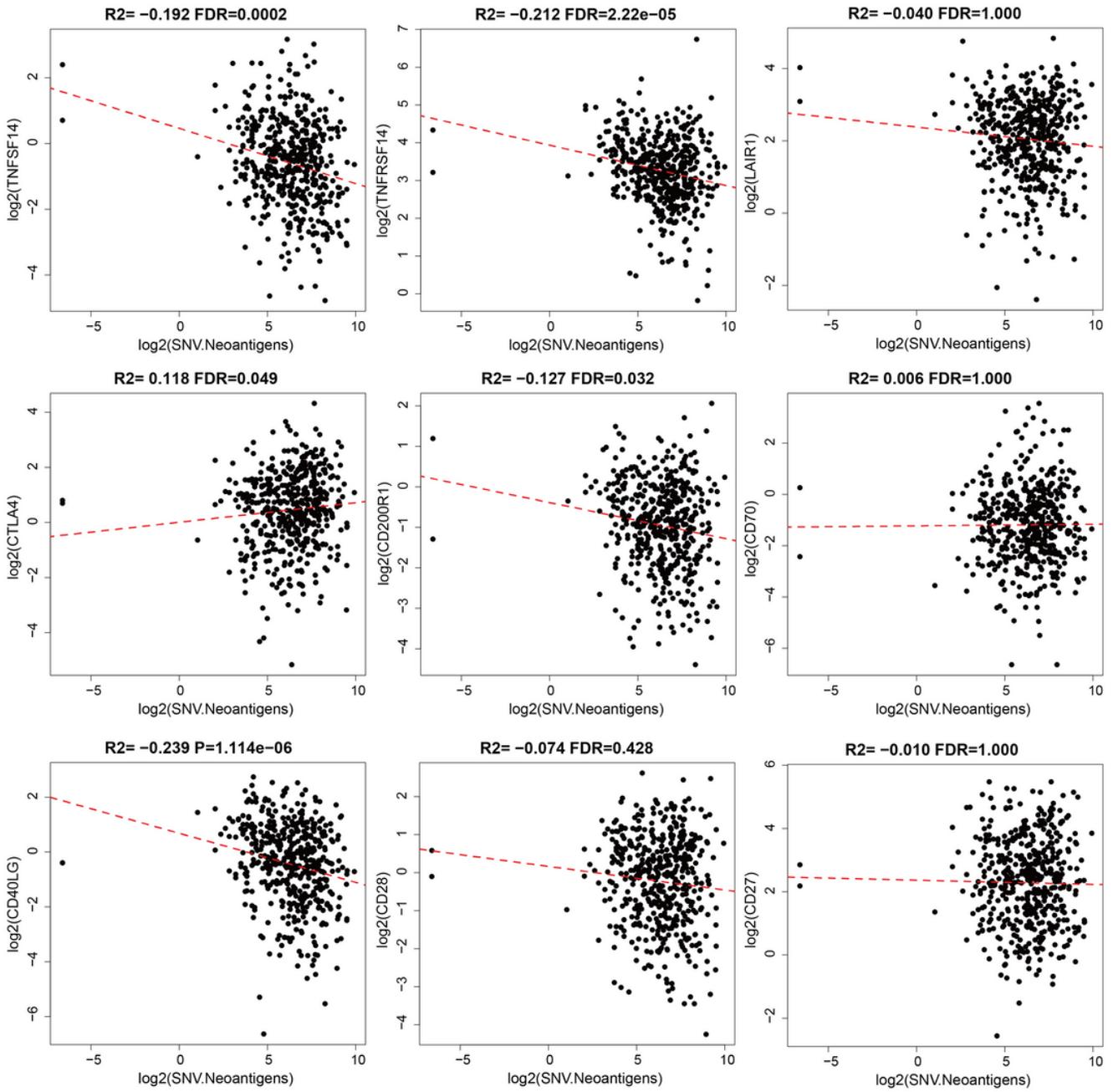


Figure 6

scatter diagram of IGCs expression levels and neoantigens. R^2 is the correlation coefficient, and FDR is false discovery rate.

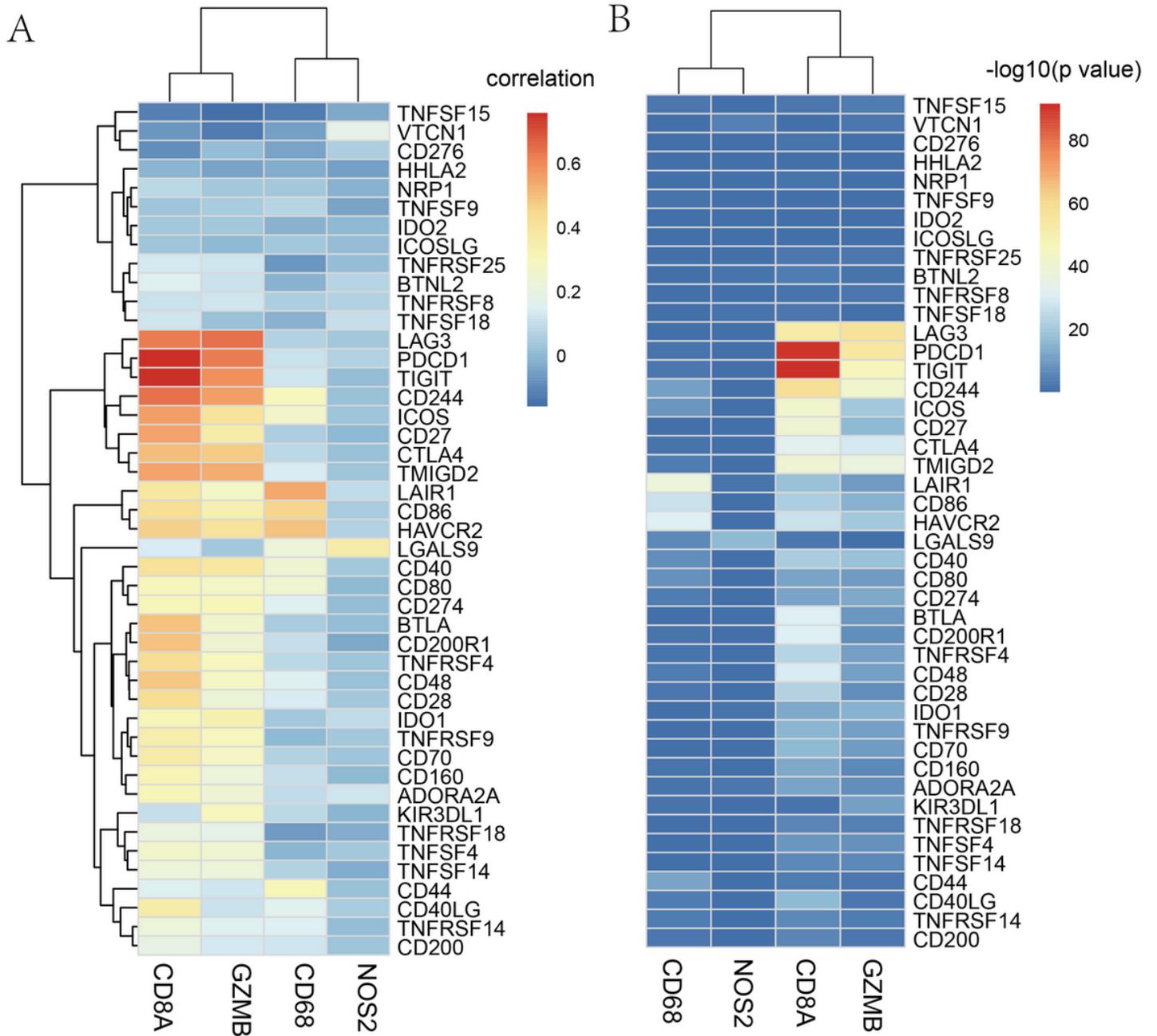


Figure 7

A: heatmap of the correlation coefficients between ICGs and adaptive immune-resistance pathway genes; B: $-\log_{10}(P\text{-value})$ is tested for the correlation coefficients between ICGs and adaptive immune-resistance pathway genes.

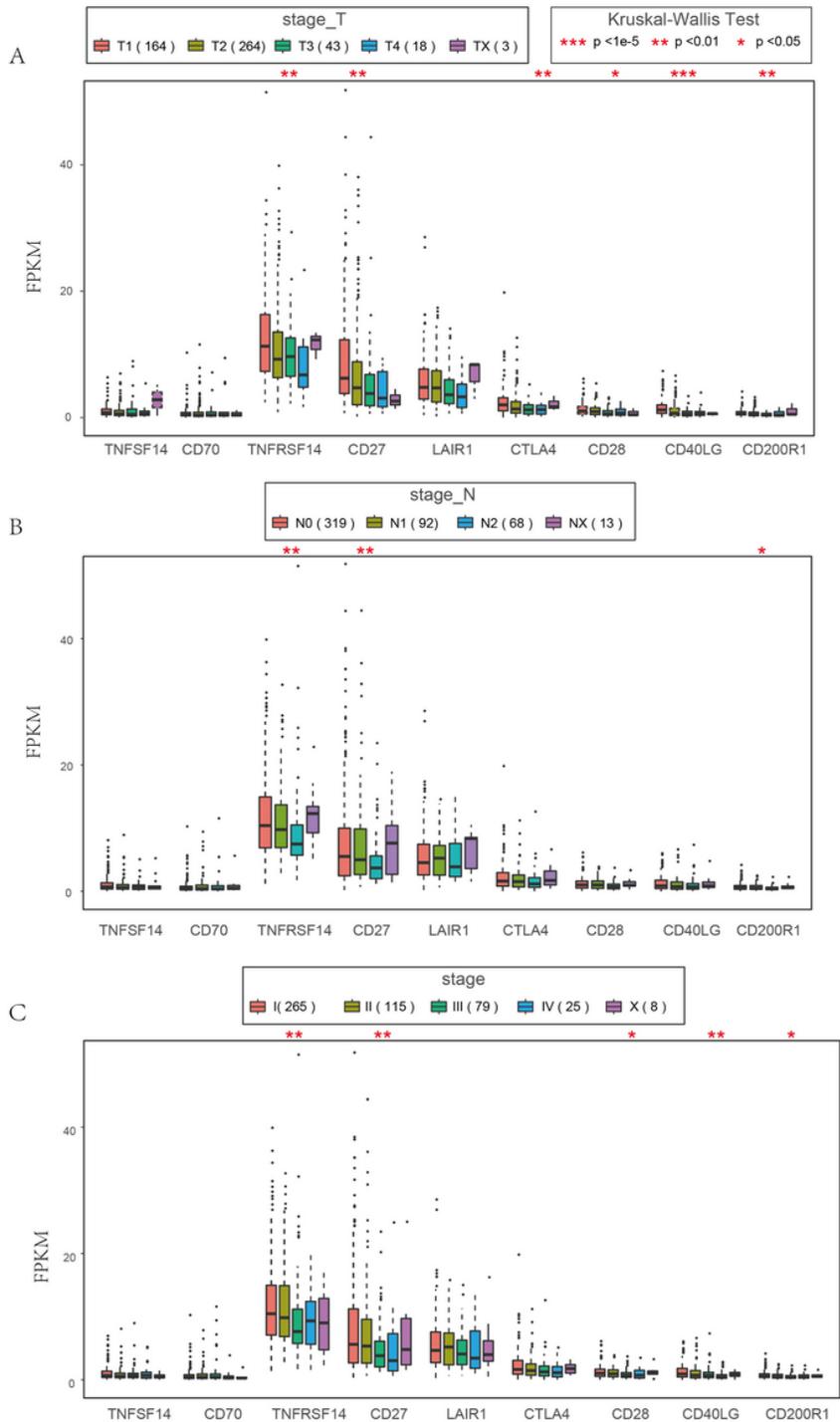


Figure 8

A: Box diagram of 9 ICGs expression in T stages (FPKM); B: Box diagram of 9 ICGs expression in N stages (FPKM); C: Box diagram of 9 ICGs expression in Stages (FPKM).

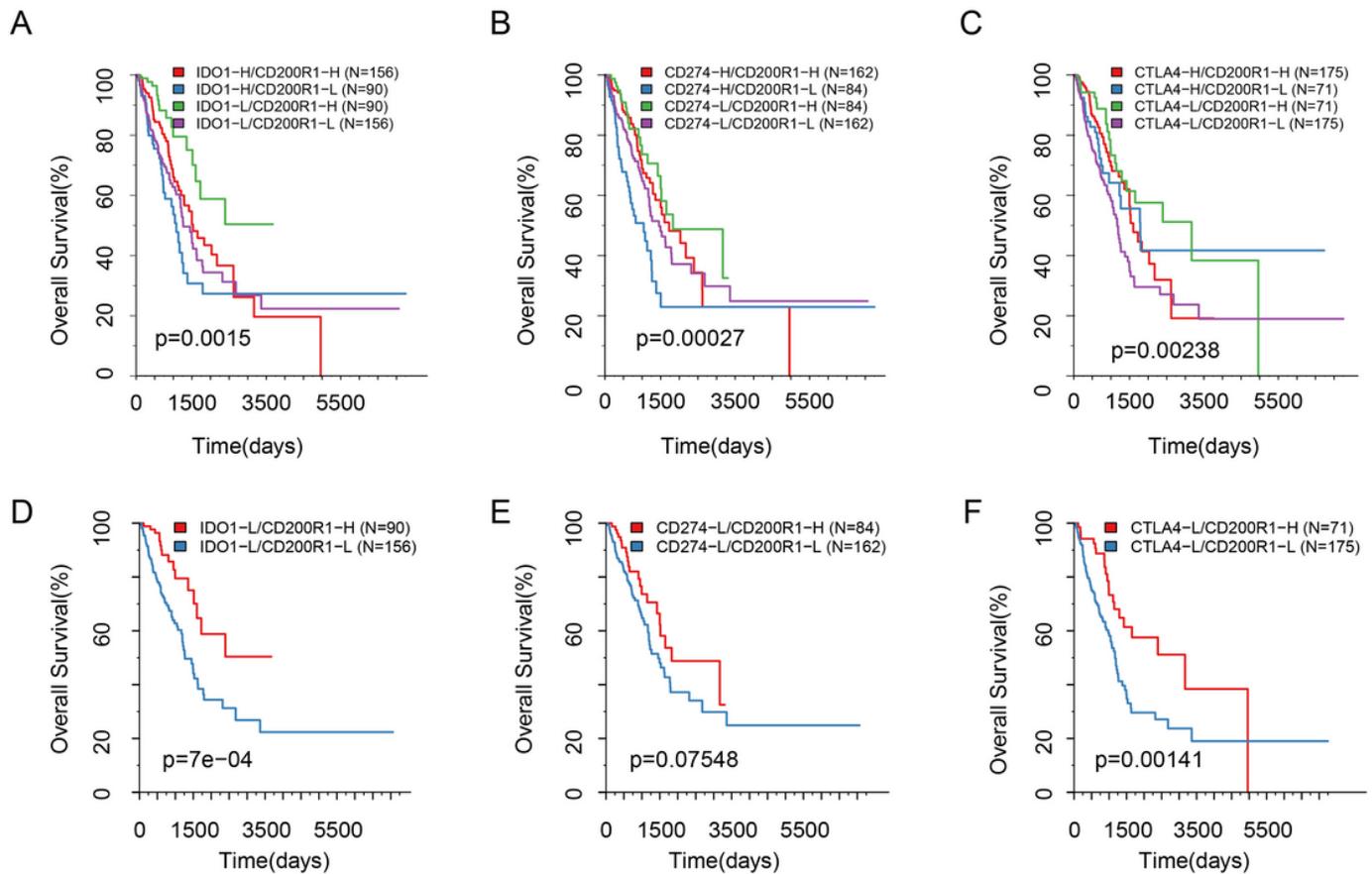


Figure 9

A: KM survival curve of grouped samples based on IDO1 High/Low expression and CD200R1 High/Low expression; B: KM survival curve of grouped samples based on High/Low expression of PD-L1 (CD274) and High/Low expression of CD200R1; C: KM survival curve of grouped samples based on CTLA4 High/Low expression and CD200R1 High/Low expression; D: KM survival curve of grouped samples with IDO1 Low expression + CD200R1 High expression and IDO1 Low expression + CD200R1 Low expression. E: KM survival curve of grouped samples with PD-L1 (CD274) Low expression + CD200R1 High expression and PD-L1 (CD274) low expression + CD200R1 Low expression. F: KM survival curve of grouped samples with CTLA4 Low expression + CD200R1 High expression and CTLA4 Low expression + CD200R1 Low expression.

Supplementary Files

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

- [S2.tif](#)
- [S1.tif](#)
- [S1Table.txt](#)
- [S4.tif](#)

- S3.tif