Tumor mutation burden (TMB)-associated signature constructed to predict survival of lung squamous cell carcinoma patients

Dan Yan & Yi Chen

Lung squamous cell carcinoma (LUSC) is a common type of lung cancer with high incidence and mortality rate. Tumor mutational burden (TMB) is an emerging biomarker for selecting patients with non-small cell lung cancer (NSCLC) for immunotherapy. This study aimed to reveal TMB involved in the mechanisms of LUSC and develop a model to predict the overall survival of LUSC patients. The information of patients with LUSC were obtained from the cancer genome atlas database (TCGA). Differentially expressed genes (DEGs) between low- and high-TMB groups were identified and taken as nodes for the protein–protein interaction (PPI) network construction. Gene oncology (GO) enrichment analysis and gene set enrichment analysis (GSEA) were used to investigate the potential molecular mechanism. Then, we identified the factors affecting the prognosis of LUSC through cox analysis, and developed a risk score signature. Kaplan–Meier method was conducted to analyze the difference in survival between the high- and low-risk groups. We constructed a nomogram based on the risk score model and clinical characteristics to predict the overall survival of patients with LUSC. Finally, the signature and nomogram were further validated by using the gene expression data downloaded from the Gene Expression Omnibus (GEO) database. 30 DEGs between high- and low-TMB groups were identified. PPI analysis identified CD22, TLR10, PIGR and SELE as the hub genes. Cox analysis indicated that FAM107A, IGLL1, SELE and T stage were independent prognostic factors of LUSC. Low-risk scores group lived longer than that of patients with high-risk scores in LUSC. Finally, we built a nomogram that integrated the clinical characteristics (TMN stage, age, gender) with the three-gene signature to predict the survival probability of LUSC patients. Further verification in the GEO dataset. TMB might contribute to the pathogenesis of LUSC. TMB-associated genes can be used to develop a model to predict the OS of lung squamous cell carcinoma patients.

Lung cancer is the commonest cancer, and is the main cause of global tumor morbidity and mortality. Non-small cell lung cancer (NSCLC) is a common type of lung cancer, including lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC) and large cell lung carcinoma (LCLC). Among NSCLC, LUAD is the most common followed by LUSC which constitutes approximately 30% of all lung cancer cases. With the discovery of target genes and the development of targeted therapy and immunotherapy, the survival of patients with NSCLC has been extended, especially patients with LUAD. However, due to the lack of effective targeted treatments, the progress of LUSC therapy is very slow and the clinical outcome of LUSC remains unsatisfactory.

TMB is the number of mutations per million bases in tumor tissue, including base substitutions, gene insertion, gene coding and deletion errors. TMB has a vital role in tumor occurrence and development, and affects the immune response and survival prognosis of LUSC patients. And many studies suggest that TMB is a potential and emerging biomarker for selecting NSCLC patients suitable for immunotherapy, even better than PD-1/PDL-1 expression. In resected NSCLC patients, decreased TMB is believed to be associated with poor prognosis.

1Department of Respiratory, Jinhua Municipal Central Hospital, Jinhua Hospital of Zhejiang University, No. 365, East Renmin Road, Jinhua 321000, Zhejiang Province, People’s Republic of China. 2Department of Hepatology, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou 325000, People’s Republic of China. *email: 18367815791@163.com.
NSCLC patient prognosis is most often evaluated in light of the American Joint Committee on Cancer (AJCC) staging system (8th edition), with the stage of the cancer being used to guide treatment decision making. However, in many cases this system may fail to accurately predict the prognosis of a given patient, as a number of other factors can influence such outcomes. Several studies have suggested that TMB might be a potentially-useful clinical predictor in NSCLC patients undergoing immunotherapy and patients with resected NSCLC. However, there are few studies focusing on the relationship between TMB and LUSC. However, few studies have focused on the relationship between TMB and LUSC. Therefore, using data from the TCGA database, we built a novel predictive model able to predict the survival probability of LUSC patients.

Method

Data acquisition. The RNA transcriptome sequences and corresponding clinicopathological data for LUSC patients were collected from The Cancer Genome Atlas (TCGA, https://portal.gdc.cancer.gov/) and used for model training; this included 502 tumors and 49 paracancerous tissues. Information obtained from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) was used for external validation; this included 69 lung squamous carcinoma samples (GSE73403). Data was obtained solely from public databases, obviating any ethical conflict.

TMB calculation and differential expressed genes (DEGs) screening. TMB is the sum of mutations per megabase in tumor tissue. TMB for each organization can be detected using the VarScan method, as calculated by the R package “maftools”. The R package “limma” was used to identify differentially expressed genes (DEGs) between a high TMB population and a low TMB population. Classification used the median of the TMB score and a [log2 fold change] ≥ 1.0. A P value < 0.05 and a False Discovery Rate (FDR) < 0.05 were the screening criteria. TMB evaluation of different analysis pipelines (MuTect and Muse detected Mutation data) was undertaken using R package ‘maftools’. Kaplan–Meier analysis was used to show the survival difference between high and low TMB expression groups.

Protein–protein interactions (PPI) network construction. We construct a PPI network of DEGs using the STRING database (https://string-db.org/) and visualized using Cytoscape software. A confidence score of C ≥ 0.15 was defined as the threshold criterion.

Functional enrichment analysis. We used Gene Set Enrichment Analysis (GSEA) to analyze the signal path for two different TMB expression groups. Gene Ontology (GO) enrichment analysis was used to reveal potential biological processes of TMB-associated DEGs in LUSC.

Construct and validate a Cox proportional hazards model. Univariate and multivariate Cox analysis were used to identify factors affecting survival and prognosis of LUSC. The risk associated with TMB-related genes was found from multivariate Cox regression analysis. From the median calculated score, patients were classified into high- and low-risk groups. Kaplan–Meier curves showed the survival difference between groups for both training and validation groups. A nomogram utilizing the risk score and baseline clinical information was able to predict 1-, 3- and 5-year overall survival (OS) in LUSC patients. The predictive accuracy of the model was found using ROC curves. RNA transcriptome profiles of LUSC patients were downloaded from the GEO database and used for external validation.

Statistical analysis. DEGs were screened using the R package “Limma”. GO was conducted using “clusterProfiler”, “ggplot2”, “enrichplot”, “stringi”, and “DOSE” packages. By using the TMB value, GSEA was able to analyze the signal path. We choose the “c2.cp.kegg.v6.2.symbols.gmt” gene set downloaded from the MSigDB database (http://software.broadinstitute.org/gsea/msigdb/) as the reference. To assess prognostic value, Cox regression analysis was used to estimate the hazard ratio and 95% confidence interval (95% CI) for each variable in the LUSC cohort. Statistically significant variables (P < 0.05) in the initial univariate Cox regression analysis were used in the subsequent multivariate analysis. A predictive risk model was created using multivariate Cox regression analysis implemented in the “survival” package. The resulting model and clinical information were incorporated into a final predictive nomogram, able to predict 1-, 3-year and 5-year LUSC patient OS. Nomogram prediction was undertaken using the “survival” package. Predictive accuracy was evaluated using ROC curves. All statistical analysis was performed using R version 3.5.2, Cytoscape, GSEA, and perl. A P value < 0.05 was considered significant.

Result

DEGs screening and PPI network construction. 30 DEGs (including CCL19, BPIFB1, SCGB1A1, PIGR, SELE, NR5A1, PIP) were sorted into low and high expression TMB groups: threshold [Log2 FC] > 1.0, P value < 0.05 and FDR < 0.05 (in Table 1; Fig. 1a). Kaplan–Meier analysis indicated patient OS in the low expression group was significantly lower than in the high expression group. When evaluating TMBs generated by different analysis pipelines this remained true (Fig. 3). Figure 2 shows the distribution of DEGs in the two TMB expression groups. By using a STRING database comprising 25 nodes and 44 edges, we established a PPI network between DEGs, using a confidence threshold of C ≥ 0.15 (Fig. 1b). CD22, TLR10, PIGR, and SELE were identified as hub genes and visualized using Cytoscape (https://cytoscape.org/) (Fig. 1c). DEGs included in the model and hub genes were verified in TMB evaluation of different analysis pipelines (Table 2).
Enrichment analysis. GO analysis indicated that the identified DEGs participated in regulation of lymphocyte activation, lymphocyte, mononuclear cell, and leukocyte proliferation (Table 3; Fig. 3). MS4A1, TNFSF8, SCGB1A1, CD22, and CCL19 were associated with lymphocyte, Leukocyte and mononuclear cell proliferation. TNFSF8, SCGB1A1, CD22, and CCL19 were involved with regulating lymphocyte proliferation and leukocyte proliferation. GSEA analysis indicated that high expression of TMB correlated with DNA replication, cell cycle, oocyte meiosis, spliceosome, RNA degradation, base excision repair, and pyrimidine metabolism. Low-TMB was mainly associated with B cell receptor, T cell receptor, and chemokine signaling pathways (Fig. 4).

Cox proportional hazards model construction. Univariant analysis suggested that survival rate was correlated with FAM107A, IGLL1, SELE and T stage. Multivariate Cox analysis indicated that FAM107A, IGLL1, SELE and T stage were independent prognostic factors of LUSC (Table 4). The regression model we constructed using multivariate Cox analysis can be used to predict LUSC patient OS. According to the median score, patients were divided between high- and low-risk groups. In training, Kaplan–Meier indicated patient OS in the low-risk group was significantly lower than in the high-risk group, as confirmed by external validation using data from the GEO database (GSE73403) (Fig. 5).

Development of a nomogram. A nomogram able to predict LUSC patient survival at 1 year, 3 years, 5 years was constructed, incorporating the following information: T stage, M stage, N stage, age, gender and risk score model. In this nomogram, lines under each independent prognostic factor correspond to a score, with a combined score produced by summing all individual scores. This overall score allows prediction of patient prognosis after 1 years ("sur1year"), 3 years ("sur3year") and 5 years ("sur5year") (Fig. 6).

Nomogram validation. We then validated this nomogram using Receiver Operator Characteristic (ROC) curves (Fig. 7). We found that in training the area under the ROC curve was 0.672 (95%CI: 0.632–0.684) for 1-year survival, 0.659 (95%CI: 0.621–0.665) for 3-year survival, and 0.645 (95%CI: 0.621–0.665) for 5-year survival (Fig. 7). In external validation the corresponding areas were 0.648 (95%CI: 0.632–0.684) for 1-year survival, 0.681 (95%CI: 0.661–0.695) for 3-year survival, and 0.652 (95%CI: 0.621–0.665) for 5-year survival.

Table 1. Differential expressed genes between low TMB and high TMB groups.
**Discussion**

LUSC is a common type of lung cancer with high incidence and mortality rate. Despite the improved diagnosis and therapy of LUSC, the prognosis is still very poor. As the survival of NSCLC patients is influenced by many factors beyond just tumor stage, the use of the TNM staging system to estimate patient prognosis can often lead to inaccurate survival estimates. Many other studies have generated prognostic models aimed at more accurately estimating the survival of patients with NSCLC in a comprehensive manner.

TMB can alter responses to immunotherapy and affect the prognosis of many cancers, including breast cancer, lung cancer and colon cancer. We identified 30 genes differentially expressed between high- and low-expression TMB groups. This included 4 up-regulated genes and 26 down-regulated genes. Univariate and multivariate analyses were used to examine the influence of each gene and clinical characteristic on OS, allowing a combined model and nomogram to be developed to predict LUSC patient OS.
Figure 2. The expression of DEGs distributed in high-TMB and low-TMB groups.

| Gene   | SomaticSniper logFC | pValue | MuTect logFC | pValue | Muse logFC | pValue |
|--------|---------------------|--------|--------------|--------|------------|--------|
| CD22   | −0.738376           | 1.22E−09 | −0.8605324   | 0.00718201 | −0.4142631 | 0.00088519 |
| SELE   | −1.3754053          | 7.29E−05 | −1.0018508   | 0.01206434 | −0.5736792 | 0.00823645 |
| PIGR   | −1.1479969          | 1.06E−06 | −0.7435242   | 0.00391641 | −1.0187546 | 0.00018397 |
| TLR10  | −1.0942878          | 1.68E−07 | −0.9334027   | 0.03117443 | −0.9074117 | 0.01194325 |
| FAM107A| −1.2206673          | 1.58E−07 | −0.6545071   | 0.00402084 | −0.860165  | 0.00012132 |
| IGLL1  | −2.144051           | 2.21E−07 | −0.9500279   | 0.02153361 | −1.591     | 0.01274359 |

Table 2. DEGs are verified in TMB evaluation of different analysis pipelines.

Table 3. GO enrichment analysis of TMB associated DEGs in LUSC.
GO enrichment analysis indicated these DEGs were mainly involved in lymphocyte activation and lymphocyte/leukocyte/monocyte proliferation. These genes are likely to be closely related to the functioning of immune checkpoint inhibitors in the cancer microenvironment. The proliferation, activation, and differentiation of lymphocytes are critical to the immune system. Many studies have indicated TMB to be an independent predictor of immune checkpoint inhibitor therapy influencing the immune microenvironment. Our work supports the view that TMB is closely related to immunity.

Univariant analysis and multivariate Cox analysis revealed that FAM107A, IGLL1 and SELE were independent prognostic factors of LUSC. These genes reportedly participate in the pathological processes of the TMB.

Figure 3. GO enrichment analysis of the DEGs between high- and low-TMB groups.

Figure 4. GSEA analysis was performed to further screen the significant pathway between high TMB group and low TMB group. The q-value < 0.05 was considered as significance. (a) Significant pathway identified in the high-TMB group. (b) Significant pathway identified in the low-TMB group.

GO enrichment analysis indicated these DEGs were mainly involved in lymphocyte activation and lymphocyte/leukocyte/monocyte proliferation. These genes are likely to be closely related to the functioning of immune checkpoint inhibitors in the cancer microenvironment. The proliferation, activation, and differentiation of lymphocytes are critical to the immune system. Many studies have indicated TMB to be an independent predictor of immune checkpoint inhibitor therapy influencing the immune microenvironment. Our work supports the view that TMB is closely related to immunity.

Univariant analysis and multivariate Cox analysis revealed that FAM107A, IGLL1 and SELE were independent prognostic factors of LUSC. These genes reportedly participate in the pathological processes of the TMB.
Table 4. Cox proportional hazards model analysis of prognostic factors.

| Variables | Univariate Cox analysis | Multivariate Cox analysis |
|-----------|-------------------------|--------------------------|
|           | HR | 95% CI | P  | HR | 95% CI | P  |
| FAM107A   | 1.015 | 1.01–1.025 | 0.0035 | 1.0156 | 1.00–1.03 | 0.044 |
| IGL1      | 1.062 | 1.01–1.113 | 0.0124 | 1.0723 | 1.02–1.127 | 0.006 |
| SELE      | 1.025 | 0.99–1.054 | 0.05 | 1.0006 | 1.01–1.045 | 0.012 |
| PIGR      | 1.011 | 0.98–1.023 | 0.542 | 0.9977 | 0.99–1.004 | 0.66 |
| ADH1B     | 1.022 | 0.97–1.02 | 0.548 | 1.1697 | 0.99–1.007 | 0.62 |
| T         | 1.94 | 0.95–3.959 | 0.068 | 1.8135 | 0.767–1.78 | 0.03 |
| M         | 1.85 | 0.59–5.841 | 0.291 | 1.7957 | 1.05–3.12 | 0.47 |
| N         | 1.156 | 0.84–1.564 | 0.379 | 1.2321 | 0.84–3.86 | 0.12 |

Figure 5. Kaplan–Meier survival curves. (a/b) Patients from the TCGA and GSE73403 dataset are stratified into two groups according to median values for the risk scores calculated by three gene based on risk score signature. (a) Kaplan–Meier survival curves of the signature in TCGA dataset. (b) Kaplan–Meier survival curves of the signature in GSE73403 dataset. (c) Kaplan–Meier survival curves of different TMB groups calculated by VarScan. (d) Kaplan–Meier survival curves of different TMB groups calculated by MuTect. (Red means high-TMB group and blue means low-TMB group).
microenvironment. FAM107A, a putative tumor suppressor, was originally identified within common missing area of 3p21 in renal cell carcinoma. FAM107A plays a vital role in lung carcinogenesis. Katarzyna Kiwerska et al. found that the expression of FAM107A was reduced significantly in larynx squamous cell carcinoma (LSCC) and the recurrent inactivation of FAM107A may be involved in LSCC development. FAM107A expression was low or even absent in Hodgkin Reed-Sternberg (HRS) cells. IGLL1 is part of the immunoglobulin gene superfamily, and its expression is closely related to humoral immunity. It is associated with immune cell progression. IGLL1 is thought to be a component of pre-BCR (precursor B cell antigen receptor). IGLL1 Mutations can induce blood system disease. SELE, a selectin, is a cell adhesion molecule that contributes significantly to tumorigenesis and tumor progression. By studying TCGA data, we found that low SELE expression was associated with worse OS in female lung cancer patients who never smoked. Kang et al. found that E-selectin could act as a circulating signaling molecule and facilitate tumor progression and metastasis. SELE was associated with several cancers, including lung cancer, prostate cancer, and colon cancer.

The PPI network indicated that CD22, TLR10, PIGR, and SELE were hub genes. CD22, a transmembrane glycoprotein, is thought to be a regulator of autoimmunity and B cell responses. It is also expressed in human lung tumors. CD22 is an important drug target for ameliorating autoimmune diseases and acute lymphoblastic leukemia. TLR-10 is a pattern recognition receptor with anti-inflammatory properties. Kopp et al. found...
found that TLRs are associated with colorectal cancer via nuclear transcription factor-κB-initiated transcription of inflammatory genes. TLR-10 is particularly important in asthma genetics. It is an anti-inflammatory factor affecting the risk of tuberculosis. Polymeric immunoglobulin receptor (PIGR) is a component of the mucosal immune system correlated with several cancers, such as pancreatic cancer, colon cancer, hepatocellular carcinoma and bladder cancer.

TMB has a clear role in tumorigenesis and development. It is associated with the immune microenvironment and inflammation. In this study, using TMB-associated genes we developed a three gene risk model to predict patient survival. Patient mortality increases with increasing risk score. We also constructed a nomogram based on the TMB-associated genes and clinical characteristic to predict LUSD patient OS.

This study has shortcomings. First, there is no direct experimental verification of the identified DEG in LUSC. Secondly, there is not a large enough number of clinical samples to confirm the risk model and nomogram. Larger clinical trials will be needed to verify this in the future.

Conclusion
TMB correlated with tumourigenesis and the development of LUSC. We constructed a TMB-associated risk score and nomogram to predict LUSC patient survival.

References
1. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2019. CA Cancer J. Clin. 69(1), 7–34. https://doi.org/10.3322/caac.21551 (2019).
2. Zhang, X. C. et al. Comprehensive genomic and immunological characterization of Chinese non-small cell lung cancer patients. Nat. Commun. 10(1), 1772. https://doi.org/10.1038/s41467-019-09762-1 (2019).
3. Shroff, G. S., de Groot, P. M., Papadimitrakopoulou, V. A., Truong, M. T. & Carter, B. W. Targeted therapy and immunotherapy in the treatment of non-small cell lung cancer. Radiol. Clin. North Am. 56(3), 485–495. https://doi.org/10.1016/j.rcl.2018.01.012 (2018).
4. Kleczko, E. K., Kwak, J. W., Schenk, E. L. & Nemenoff, R. A. Targeting the complement pathway as a therapeutic strategy in lung cancer. Front. Immunol. 10, 954. https://doi.org/10.3389/fimmu.2019.00954 (2019).
5. Chae, Y. K. et al. Clinical implications of circulating tumor DNA tumor mutational burden (ctDNA TMB) in non-small cell lung cancer. Oncologist 24(6), 820–828. https://doi.org/10.1634/theoncologist.2018-0433 (2019).
6. Goodman, A. M. et al. Tumor mutational burden as an independent predictor of response to immunotherapy in diverse cancers. Mol. Cancer Ther. 16(11), 2598–2608. https://doi.org/10.1158/1535-7163.Mct-17-0386 (2017).
7. McFarland, D. C. et al. Tumor mutation burden and depression in lung cancer: Association with inflammation. J. Natl. Compr. Cancer Netw. 18(4), 434–442. https://doi.org/10.6004/jnccn.2019.7374 (2020).
8. Devarakonda, S. et al. Tumor mutation burden as a biomarker in resected non-small-cell lung cancer. J. Clin. Oncol. 36(30), 3095–3066. https://doi.org/10.1200/ jco.2018.78.1963 (2018).
9. Abdel-Rahman, O. Validation of the prognostic value of new sub-stages within the A/JCC 8th edition of non-small cell lung cancer. Clin. Transl. Oncol. 19(11), 1414–1420. https://doi.org/10.1007/s12094-017-1673-7 (2017).
10. Sholl, L. M. et al. The promises and challenges of tumor mutation burden as an immunotherapy biomarker: A perspective from the international association for the study of lung cancer pathology committee. J. Thorac. Oncol. 15(9), 1409–1424. https://doi.org/10.1016/j.jtho.2020.05.019 (2020).
11. Tafe, L. J. Non-small cell lung cancer as a precision oncology paradigm: Emerging targets and tumor mutational burden (TMB). Adv. Anat. Pathol. 27(1), 3–10. https://doi.org/10.1097/pai.0000000000000524 (2020).

12. Sun, F. et al. A nomogram to predict prognosis after surgery in early stage non-small cell lung cancer in elderly patients. Int. J. Surg. 42, 11–16. https://doi.org/10.1016/j.ijsu.2017.04.024 (2017).

13. Park, S. E. et al. Clinical implication of tumor mutational burden in patients with HER2-positive refractory metastatic breast cancer. Oncol. Rep. 70(6), e166768. https://doi.org/10.1080/10292034.2018.1466768 (2018).

14. Mitsudomi, T., Hamajima, N., Ogawa, M. & Takahashi, T. Prognostic significance of p53 alterations in patients with non-small cell lung cancer: a meta-analysis. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. 6(10), 4055–4063 (2000).

15. Conlin, A., Smith, G., Carey, F. A., Wolf, C. R. & Steele, R. J. The prognostic significance of K-ras, p53, and APC mutations in colorectal carcinoma. Gut 54(9), 1283–1286. https://doi.org/10.1136/gut.2005.066514 (2005).

16. Heinzel, S., Marchingo, J. M., Horton, M. B. & Hodgkin, P. D. The regulation of lymphocyte activation and proliferation. Curr. Opin. Immunol. 51, 32–38. https://doi.org/10.1016/j.coi.2018.01.002 (2018).

17. Klempner, S. J. et al. Tumor mutational burden as a predictive biomarker for response to immune checkpoint inhibitors: A review of current evidence. Oncologist 25(1), e147–e159. https://doi.org/10.1634/theoncologist.2019-0244 (2020).

18. van den Boom, J., Wolter, M., Blaschke, B., Knobbe, C. B. & Reifenberger, G. Identification of novel genes associated with astrocytoma progression using suppression subtractive hybridization and real-time reverse transcription-polymerase chain reaction. Int. J. Cancer 119(10), 2330–2338. https://doi.org/10.1002/ijc.22108 (2006).

19. Pastuszak-Lewadowska, D. et al. Decreased FAM107A expression in patients with non-small cell lung cancer. Adv. Exp. Med. Biol. 852, 39–48. https://doi.org/10.1007/978-3-319-31214-9_10 (2015).

20. Kiwerska, K. et al. Combined deletion and DNA methylation result in silencing of FAM107A gene in laryngeal tumors. Sci. Rep. 7(1), 5386. https://doi.org/10.1038/s41598-017-05857-1 (2017).

21. Lawrie, A. et al. Combined linkage and association analysis of classical Hodgkin lymphoma. Oncotarget 9(29), 20377–20385. https://doi.org/10.18632/oncotarget.24872 (2018).

22. Luo, X. et al. Identification of genes underlying the enhancement of immunity by a formula of lentinan, pachymaran and tremelia polysaccharides in immunosuppressive mice. Sci. Rep. 8(1), 10082. https://doi.org/10.1038/s41598-018-28414-w (2018).

23. Lu, X. et al. MT AuroraB regulates B cell development and cooperates with OCA-B in controlling the pre-B to immature B cell transition. Cell Rep. 28(2), 472–485.e475. https://doi.org/10.1016/j.celrep.2019.06.029 (2019).

24. Chen, D. et al. The expression pattern of the pre-B cell receptor components correlates with cellular stage and clinical outcome in acute lymphoblastic leukemia. PLoS ONE 11(9), e0162638. https://doi.org/10.1371/journal.pone.0162638 (2016).

25. Evans, R. J. & Hollis, G. F. Genomic structure of the human Iγ lambda 1 gene suggests that it may be expressed as an Ig lambda 14.1-like protein or as a canonical B cell Iγ lambda light chain: Implications for Ig lambda gene evolution. J. Exp. Med. 173(2), 305–311. https://doi.org/10.1084/jem.173.2.305 (1991).

26. Kourtides, A., Lu, R., Pence, L. J. & Anastasiadis, P. Z. A central role for cadherin signaling in cancer. Exp. Cell Res. 358(1), 78–85. https://doi.org/10.1016/j.yexcr.2017.04.014 (2017).

27. Shi, K., Li, N., Yang, M. et al. Identification of key genes and pathways in female lung cancer patients who never smoked by a high-throughput transcriptomics approach. Cancer Cell Rep. 3(7), 1165–1174. https://doi.org/10.1182/blood-2012-06-438002 (2013).

28. Rossel, M. et al. Anti-CD22-chimeric antigen receptors targeting B-cell precursor acute lymphoblastic leukemia. Science 349(6246), 977–982. https://doi.org/10.1126/science.1266527 (2015).

29. Enterina, J. R., Jung, J. & Macauley, M. S. Coordinated roles for glycans in regulating the inhibitory function of CD22 on B cells. Cellular Immunol. 372(1), 51–60. https://doi.org/10.1016/j.cellimm.2019.06.029 (2019).

30. Kopp, T. L., Vogel, U., Tjonneland, A. & Andersen, V. Meat and fiber intake and interaction with pattern recognition receptors (TLR1, TLR2, TLR4, and TLR10) in relation to colorectal cancer in a Danish prospective, case-cohort study. Am. J. Clin. Nutr. 107(3), 465–479. https://doi.org/10.1093/ajcn/nqx011 (2018).

31. Heinzmann, A. et al. Joint influences of Acidic-Mammalian-Chitinase with Interleukin-4 and Toll-like receptor-10 with Interleukin-13 in the genetics of asthma. Pediatr. Allergy Immunol. Off. Publ. Eur. Soc. Pediatr. Allergy Immunol. 21(4 Pt 2), e679-e686. https://doi.org/10.1111/pai.12959 (2010).

32. Bulat-Kardum, L. J., Etokebe, G. E., Lederer, P., Balan, S. & Dembic, Z. Genetic polymorphisms in the toll-like receptor 10, interleukin (IL)17A and IL17F genes differently affect the risk for tuberculosis in Croatian population. Scand. J. Immunol. 82(1), 63–69. https://doi.org/10.1111/sji.12300 (2015).

33. Ohkuma, R. et al. High expression levels of polymeric immunoglobulin receptor are correlated with chemoresistance and poor prognosis in pancreatic cancer. Oncol. Rep. 44(1), 252–262. https://doi.org/10.23929/or.2020.7610 (2020).

34. Poger, M. E., Hirsch, B. R. & Lamm, M. E. Synthesis of secretory component by colonic neoplasms. Am. J. Pathol. 82(2), 327–338 (1976).

35. At, I. et al. The role of polymeric immunoglobulin receptor in inflammation-induced tumor metastasis of human hepatocellular carcinoma. J. Natl. Cancer Inst. 103(22), 1696–1712. https://doi.org/10.1093/jnci/djt360 (2011).

36. Rosen, M. et al. Alterations in polymeric immunoglobulin receptor expression and secretory component levels in bladder carcinoma. Urol. Res. 19(6), 361–366. https://doi.org/10.1007/bf00310153 (1991).
Author contributions
Y.C. contributed to the conception of the study; D.Y. contributed significantly to analysis and wrote the main manuscript text and Y.C. prepared Figs. 1–7. All authors reviewed the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to D.Y.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2021