



APOLLO: An accurate and independently validated prediction model of lower-grade gliomas overall survival and a comparative study of model performance

Jiajin Chen,^{a,†} Sipeng Shen,^{a,b,†} Yi Li,^{c,†} Juanjuan Fan,^{a,†} Shiyu Xiong,^d Jingtong Xu,^d Chenxu Zhu,^a Lijuan Lin,^a Xuesi Dong,^e Weiwei Duan,^f Yang Zhao,^a Xu Qian,^g Zhonghua Liu,^h Yongyue Wei,^{a,b} David C. Christiani,^{i,***} Ruyang Zhang,^{a,b,*} and Feng Chen^{a,b,j,**}

^aDepartment of Biostatistics, Center for Global Health, School of Public Health, Nanjing Medical University, Nanjing, Jiangsu, China, 211166

^bChina International Cooperation Center for Environment and Human Health, Nanjing Medical University, Nanjing, China, 211166

^cDepartment of Biostatistics, University of Michigan, Ann Arbor, MI, USA, 48109

^dDepartment of Clinical Medicine, The First Clinical Medical College, Nanjing Medical University, Nanjing, Jiangsu, China, 211166

^eOffice of Cancer Screening, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, 100021

^fDepartment of Bioinformatics, School of Biomedical Engineering and Informatics, Nanjing Medical University, Nanjing, Jiangsu, China 211166

^gDepartment of Nutrition and Food Hygiene, Institute for Brain Tumors, Center for Global Health, School of Public Health, Nanjing Medical University, Nanjing, China, 211166

^hDepartment of Statistics and Actuarial Science, the University of Hong Kong, Hong Kong, China, 999077

ⁱPulmonary and Critical Care Division, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA, 02114

^jJiangsu Key Lab of Cancer Biomarkers, Prevention and Treatment, Cancer Center, Collaborative Innovation Center for Cancer Personalized Medicine, Nanjing Medical University, Nanjing, Jiangsu, China, 211166

Summary

Background Virtually few accurate and robust prediction models of lower-grade gliomas (LGG) survival exist that may aid physicians in making clinical decisions. We aimed to develop a prognostic prediction model of LGG by incorporating demographic, clinical and transcriptional biomarkers with either main effects or gene-gene interactions.

Methods Based on gene expression profiles of 1,420 LGG patients from six independent cohorts comprising both European and Asian populations, we proposed a 3-D analysis strategy to develop and validate an Accurate Prediction mOdel of Lower-grade gLiomas Overall survival (APOLLO). We further conducted decision curve analysis to assess the net benefit (NB) of identifying true positives and the net reduction (NR) of unnecessary interventions. Finally, we compared the performance of APOLLO and the existing prediction models by the first systematic review.

Findings APOLLO possessed an excellent discriminative ability to identify patients at high mortality risk. Compared to those with less than the 20th percentile of APOLLO risk score, patients with more than the 90th percentile of APOLLO risk score had significantly worse overall survival ($HR=54.18$, 95% CI: 34.73-84.52, $P=2.66 \times 10^{-69}$).

eBioMedicine 2022;79:
104007
Published online xxx
<https://doi.org/10.1016/j.ebiom.2022.104007>

*Corresponding author: Ruyang Zhang, School of Public Health, Nanjing Medical University, SPH Building Room 406, 101 Longmian Avenue, Nanjing, Jiangsu 211166, China.

**Corresponding author: Feng Chen, School of Public Health, Nanjing Medical University, SPH Building Room 412, 101 Longmian Avenue, Nanjing, Jiangsu 211166, China.

***Corresponding author: David C. Christiani Harvard School of Public Health and Harvard Medical School, Building 1 Room 1401, 665 Huntington Avenue, Boston, MA 02115, USA.

E-mail addresses: dchris@hsph.harvard.edu (D.C. Christiani), zhangruiyang@njmu.edu.cn (R. Zhang), fengchen@njmu.edu.cn (F. Chen).

† These authors contributed equally to this work.

Senior author who supervised the work.

Further, APOLLO can accurately predict both 36- and 60-month survival in six independent cohorts with a pooled $AUC_{36\text{-month}}=0.901$ (95% CI: 0.879-0.923), $AUC_{60\text{-month}}=0.843$ (95% CI: 0.815-0.871) and $C\text{-index}=0.818$ (95% CI: 0.800-0.835). Moreover, APOLLO offered an effective screening strategy for detecting LGG patients susceptible to death ($NB_{36\text{-month}}=0.166$, $NR_{36\text{-month}}=40.1\%$ and $NB_{60\text{-month}}=0.258$, $NR_{60\text{-month}}=19.2\%$). The systematic comparisons revealed APOLLO outperformed the existing models in accuracy and robustness.

Interpretation APOLLO has the demonstrated feasibility and utility of predicting LGG survival (<http://bigdata.njmu.edu.cn/APOLLO>).

Funding National Key Research and Development Program of China (2016YFE0204900); Natural Science Foundation of Jiangsu Province (BK20191354); National Natural Science Foundation of China (81973142 and 82103946); China Postdoctoral Science Foundation (2020M681671); National Institutes of Health (CA209414, CA249096, CA092824 and ES000002).

Copyright © 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Keywords: Lower-grade gliomas; Survival; Prognostic prediction; Nomogram; Online tool; Systematic review

Research in Context

Evidence before this study

We searched PubMed, Embase, MEDLINE, Web of Science, and Cochrane Library for articles about prognostic prediction models of LGG published before Aug 30, 2021, using search term “((lower-grade glioma) OR (lgg)) AND ((progn*) OR (survival)) AND ((predict*) OR (auc) OR (area under the curve) OR (receiver operator characteristic curve) OR (c-index) OR (c statistic) OR (roc) OR (calibration))”. We found the existing models underwent limited prediction accuracy and model validation, as most of them either solely relied on training populations or retrained models in testing populations to assess the model performance, which might be overestimated due to overfitting. Additionally, most of these existing models have limited model robustness and transportability to accommodate independent population, impeding their wide applications.

Added value of this study

In this study, we collected 1,420 LGG patients from six European and Asian populations and proposed an effective modeling strategy to develop and validate an **A**ccurate **P**rediction **m**odel of **L**ower-grade **g**liomas **O**verall survival (APOLLO), which has the demonstrated feasibility and utility in distinguishing LGG patients at high risk of mortality and predicting their survival. Our systematic review revealed that APOLLO outperformed the existing models in accuracy and robustness.

Implications of all the available evidence

APOLLO has clinical benefits at identifying LGG patients at high mortality risk and presents a higher net benefit of identifying true positives and net reduction of unnecessary interventions. A convenient online tool to

implement APOLLO was developed at <http://bigdata.njmu.edu.cn/APOLLO>.

Introduction

Gliomas, the most common malignant cancer in the brain and central nervous system, account for over 80% of malignant brain tumors.¹ Lower-grade gliomas (LGG), consisting of diffuse low- and intermediate-grade gliomas, are graded II and III by World Health Organization (WHO).² Compared to those diagnosed as glioblastoma (GBM) with a WHO grade IV, LGG patients tend to have more favorable prognosis; however, 70% of them will progress to GBM within ten years.³ Thus, delaying tumor progression for LGG patients is critical. What is often overlooked is wide heterogeneity of LGG prognosis that is ubiquitous for those even with similar clinical features, indicating possible molecular underpinnings of the disease progression process.⁴ As a crucial milestone, the WHO Classification of Tumors of the Central Nervous System synthesized molecular and histological information to reclassify gliomas, by using well recognized molecular biomarkers.² Recent evidence has emerged that gene expressions may pose inducible and reversible effects on LGG prognosis via several channels, including immunity,^{5,6} stemness,⁷ and autophagy.^{8,9}

The prognostic prediction utilizing biomarkers can aid physicians in making clinical decisions or guiding adjuvant therapy.¹⁰ Recently, much effort has been shifting to the LGG prognostic prediction.¹¹⁻¹³ However, existing prediction models have various technical bottlenecks, impeding their wide applications. Specifically, these models

underwent limited model validation, as most of them either solely relied on training populations or retrained models in testing populations to assess the model performance, which might be overestimated due to overfitting.^{9,14} Therefore, most of these existing models have limited model robustness and transportability to accommodate independent populations.^{6,15,16}

Furthermore, almost all of the studies merely focused on predictors with main effects, but neglected predictors exhibiting gene-gene (G×G) interactions, which may provide pivotal clues regarding the biologic mechanisms of complex diseases¹⁷ and enhance prediction accuracy,^{18,19} as evidenced by our own study of lung cancer.²⁰

To address challenges in LGG survival prediction, we developed an Accurate and independently validated Prediction mOdel of Lower-grade gLiomas Overall survival (APOLLO) which identifies and includes biomarkers with significant main effects or G×G interactions, based on six cohorts with both European and Asian populations. Additionally, we have developed a free online tool implementing APOLLO to facilitate prediction of LGG survival.

Materials and Methods

Data collection and study population

We curated the clinical and gene expression data of LGG patients from six glioma cohorts, namely, the Cancer Genome Atlas (TCGA),²¹ the Chinese Glioma Genome Atlas (CGGA1),²² CGGA2,²³ Rembrandt (GSE108476),²⁴ Weller (GSE61374)²⁵ and Gravendeel (GSE16011) cohorts.²⁶ Only newly diagnosed LGG patients with complete overall survival time and transcriptomics data were retained. With the focus on biological functions and clinical utility, we considered a total of 723 pan-cancer driving genes defined by COSMIC;²⁷ among them, included in our study were 680 genes shared by all six cohorts. All gene expression levels were log₂-transformed and standardized before being passed into association analyses; see the Supplementary Methods for the details of sample quality control. Included in our subsequent analyses were a total of 1,420 LGG patients with 680 genes, whose demographic and clinical characteristics were summarized in Supplementary Table S1.

APOLLO construction and validation

Figure 1, depicting the study design and workflow, features a 3-D strategy (Double Types of Effects, Double Steps of Screening, and Double Steps of Modeling) for the development and validation of the APOLLO model.

- (i) **Double Types of Effects.** For selection of important main effects and G×G interactions, we considered

Cox Models 1 and 2, respectively:

$$\begin{aligned} \text{Model1: } h(t) &= h_0(t) \exp(\alpha \times \text{gene} + \sum \beta_i \times \text{covariate}_i), \\ \text{Model2: } h(t) &= h_0(t) \exp(\alpha_1 \times \text{gene}_1 + \alpha_2 \times \text{gene}_2 + \alpha_3 \\ &\quad \times \text{gene}_1 \times \text{gene}_2 + \sum \beta_i \\ &\quad \times \text{covariate}_i) \end{aligned}$$

which adjusted for covariates, including age, WHO grade, IDH mutation and 1p/19q status (Supplementary Table S2).

- (ii) **Double Steps of Screening.** We scanned the pan-cancer related genes to select candidate genes and interactions, and then validated them with an independent validation dataset. Specifically, on the TCGA cohort, we fitted Models 1 and 2 on each gene and interaction, respectively, and selected important genes and interactions by controlling the false positive rate at a 5% level ($q\text{-FDR} \leq 5\%$). On the CGGA1 cohort, we validated these selected genes or interactions; only those with $P \leq 0.05$ and with same effect directions as in the discovery step would be selected as candidate biomarkers to be passed onto the next modeling stage.
- (iii) **Double Steps of Modeling.** On the TCGA cohort and with the candidate genes and interactions identified from the previous screening stage, we used Cox models (adjusted for demographic and clinical predictors) to conduct forward stepwise regression, that is, using the likelihood ratio test with $P_{\text{entry}} \leq 0.05$ and $P_{\text{removal}} > 0.05$, to identify a final multivariable Cox model and construct APOLLO. As validation, we assessed the discriminative performance of the obtained APOLLO via area under the receiver operating characteristic curves (AUC) or concordance index (C-index) on one internal cohort (CGGA1) and four external cohorts, namely, CGGA2, Rembrandt, Weller and Gravendeel.

Bioinformatics analysis for transcriptional predictors

To understand the potential gene functions of the identified transcriptional predictors, we conducted a gene enrichment pathway analysis based on Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) database by using R package *clusterProfiler*. Estimation of Stromal and Immune Cells in Malignant Tumor Tissues Using Expression Data (ESTIMATE)²⁸ was used to predict the presence of stromal and immune cells in tumor tissue, and CIBERSORT was performed to determine the proportions of 22 immune cells from bulk tumors based on gene expression.²⁹ Finally, the gene network analysis of screened genes and immune checkpoint genes was performed using GeneMANIA,³⁰ a plugin of the Cytoscape application.

A systematic review of LGG survival prediction models

Following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (Supplementary

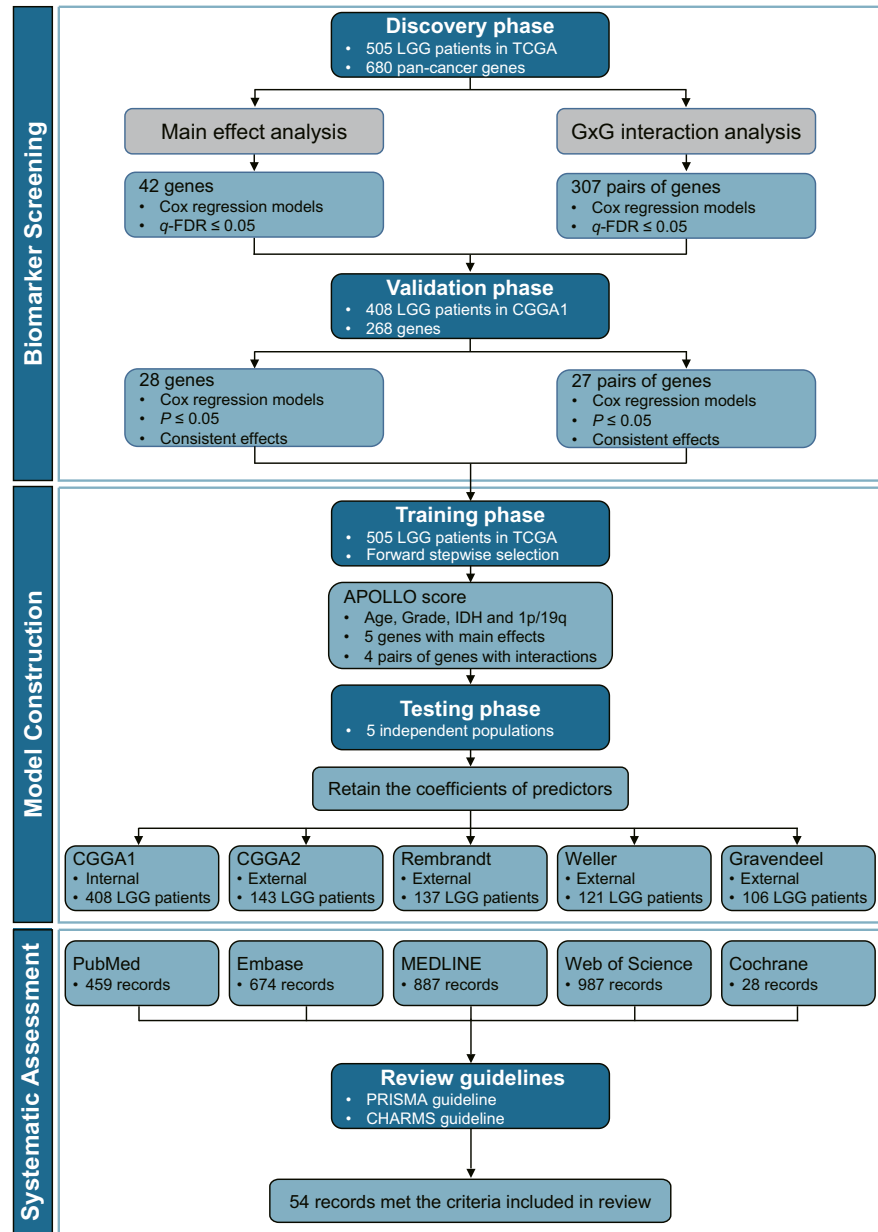


Figure 1. Flowchart of development and validation of APOLLO and a systematic assessment.

The APOLLO model was developed by using a 3-D analysis strategy, encompassing **Double Types of Effects**, **Double Steps of Screening**, and **Double Steps of Modeling**. For biomarker screening for double types of effects, we tested both main effects and G×G interactions, followed by double steps of screening (biomarkers were first identified using the TCGA cohort and then validated in the CGGA1 cohort); the double steps of modeling meant that the model was first trained in the TCGA cohort and then tested in the CGGA1, CGGA2, Rembrandt, Weller and Gravendeel cohorts, respectively.

Table S3), we conducted a systematic literature search on prognostic prediction models of LGG using five major databases, namely, PubMed, Embase, MEDLINE, Web of Science, and Cochrane Library. The literature search and data extraction were done independently by two researchers (S.X. and J.X.), and the discrepancies

were arbitrated by a third researcher (J.C.). Details of search strings, exclusion criteria and data extraction process were provided in Supplementary Methods. We totally retrieved 3,035 articles. After removing duplicates, 1,444 articles were included for further screening. Among them, 126 articles that met the criteria on title

or abstract, were eligible for a full-text review. Finally, 54 articles fully meeting our selection criteria were retained and used for data extraction.

APOLLO visualization and online software

We generated a nomogram for visualizing APOLLO by using R package *rms*, which can be accessed at <http://bigdata.njmu.edu.cn/APOLLO>. With input values of predictors for a LGG patient, the online calculator immediately returns predicted survival rates and 95% confidence intervals (CIs) at any time point between 0 and 120 months, based on an interactive web-based Kaplan-Meier survival curve.

Statistical analysis

Continuous variables were summarized as mean \pm standard deviation, and categorized variables were described by frequency (*n*) and proportion (%). Associations between characteristics and overall survival were evaluated by Cox models using R package *survival*. Study centers were adjusted for when analyzing the combined samples. Kaplan-Meier survival curves illustrated the survival differences across different risk groups. The prediction accuracy was presented by using a time-dependent receiver operating characteristic (ROC) curve³¹ and was assessed by the time-dependent AUC, which can be obtained from R package *timeROC*. We used calibration plots to evaluate the consistency between nomogram-predicted and observed risks, and conducted decision curve analysis (with details given in Supplementary Methods) to gauge the net benefit (NB) of identifying true high risk patients that ought to have intervention and the net reduction (NR) of unnecessary interventions, due to the use of APOLLO as a screening tool.^{32,33} Since these transcriptional predictors were validated by trans-ethnic populations, we assumed APOLLO had uniform and homogenous performance

Statistical analyses were performed using R (version 3.6.3). A two-sided *P* value less than or equal to 0.05 was considered statistically significant unless otherwise specified. The source code and data were deposited at <https://github.com/JiajinChen/APOLLO>.

Ethics

The study was performed in accordance with Good Clinical Practice guidelines and the World Medical Association Declaration of Helsinki. All patients provided informed consents. All data used in this study were de-identified and no protected health data was needed.

Role of the funding source

The sponsors had no role in the study design, data collection, data analyses, interpretation, or writing of the study.

Results

Development and construction of APOLLO

First, 42 genes with main effects and 307 pairs of genes with G×G interactions were identified ($q\text{-FDR} \leq 0.05$) to be possibly associated with overall survival in TCGA cohort. Of them, 28 genes with main effects and 27 pairs of genes with G×G interactions were validated in CGGA1 cohort to be candidate transcriptional predictors (Supplementary Tables S4-S5). Then, out of these candidate transcriptional predictors and on the TCGA training cohort, we used forward stepwise regression strategy to construct a final Cox model, which included 5 genes with main effects and 4 pairs of genes with G×G interactions (Supplementary Table S6). Using the coefficient estimates from this final Cox model, APOLLO, which integrated demographic, clinical and transcriptional predictors, was defined as:

$$\begin{aligned} \text{APOLLO_Score} &= 0.0312 \times \text{age} + 0.5276 \times \text{grade} - 0.5510 \times \text{IDH} - 0.5163 \times 1p/19q \\ &\quad + 0.7528 \times \text{Transcriptional_Score} \\ \text{Transcriptional_Score} &= 0.2976 \times \text{CHIC2} + 0.3500 \times \text{IGF2BP2} \\ &\quad + 0.2387 \times \text{ITGAV} + 0.5532 \times \text{MSN} + 0.4034 \times \text{PLCG1} \\ &\quad + 0.2361 \times \text{BCORL1} - 0.1082 \times \text{PRF1} - 0.2498 \times \text{BCORL1} \times \text{PRF1} \\ &\quad - 0.1674 \times \text{HMGA1} - 0.1058 \times \text{TFG} + 0.1930 \times \text{HMGA1} \times \text{TFG} \\ &\quad - 0.1922 \times \text{CTNND2} - 0.1814 \times \text{GOLGA5} - 0.2340 \times \text{CTNND2} \times \text{GOLGA5} \\ &\quad - 0.0888 \times \text{FAS} - 0.2073 \times \text{SMAD4} - 0.1724 \times \text{FAS} \times \text{SMAD4} \end{aligned}$$

across cohorts. Thus, we performed meta-analysis to pool prediction accuracy of APOLLO from six cohorts using the fixed-effect model, implemented by the R package *meta*. Stratified analyses were displayed by forest plots using the R package *forestplot*.

Transcriptional predictors of APOLLO and their immune relevance

KEGG enrichment analysis categorized gene probes into 30 pathways, including the glioma pathway, and GO annotation identified 279 biological process

pathways, 24 molecular function pathways and 20 cellular component pathways, suggesting potential biological functions (Supplementary Table S7). We compared the proportions of 22 types of immune cells between high- and low-risk groups defined by the median transcriptional score (-0.2689), and found that they were significantly different between the two groups (Supplementary Figure S1a). Further, the transcriptional score was correlated with the stromal, immune and ESTIMATE scores (Supplementary Figure S1b). Additionally, we observed high connectivity and large correlations between transcriptional predictors and immune checkpoint genes (Supplementary Figure S1c and Supplementary Figure S2), indicating that the transcriptional predictors may play a role in immune responses. Numerous immunity-related drugs targeting these transcriptional predictors have been documented in the DrugBank database (Supplementary Table S8), and, thereby, APOLLO may have potential roles in guiding immunotherapy.

Discriminative ability of APOLLO

Patients in each of the six cohorts were categorized into low- and high-risk groups using the median APOLLO score (0.6945) obtained from the TCGA training set. The APOLLO score had an adequate discriminative ability in both training and testing sets. Compared to the low-risk group in the corresponding cohort, the high-risk group was associated with worse survival in the TCGA cohort (the training set) and CGGA1 (the internal testing cohort), exhibiting a large hazard ratio (HR) ($HR_{TCGA}=8.51$, 95% CI: $5.10-14.18$, $P=2.14 \times 10^{-16}$; $HR_{CGGA1}=4.86$, 95% CI: $3.24-7.28$, $P=1.75 \times 10^{-14}$) (Figure 2a-b), and in the 4 external testing sets ($HR_{CGGA2}=6.26$, 95% CI: $2.86-13.68$, $P=4.41 \times 10^{-6}$; $HR_{Rembrandt}=3.49$, 95% CI: $2.06-5.91$, $P=3.32 \times 10^{-6}$; $HR_{Weller}=3.41$, 95% CI: $1.73-6.72$, $P=3.99 \times 10^{-4}$; $HR_{Gravendeel}=2.19$, 95% CI: $1.31-3.68$, $P=2.88 \times 10^{-3}$) (Figure 2c-f). We further illustrated the discriminative ability of the APOLLO score by classifying patients into 6 groups defined by the quintiles and the 90 percentile of the score in the combined cohort. The median survival months dramatically dropped from 192.6 in the 1st group (less than the 20th percentile) to 15.7 in the 6th group (above the 90th percentile). There appeared to exist a dose-response association: higher-percentile groups were associated with shorter survival and higher mortality risk (HR_6 vs $1=54.18$, 95% CI: $34.73-84.52$, $P=2.66 \times 10^{-69}$; HR_5 vs $1=16.28$, 95% CI: $10.57-25.07$, $P=1.07 \times 10^{-36}$; HR_4 vs $1=7.05$, 95% CI: $4.66-10.69$, $P=3.03 \times 10^{-20}$; HR_3 vs $1=3.88$, 95% CI: $2.51-6.00$, $P=9.78 \times 10^{-10}$; HR_2 vs $1=2.63$, 95% CI: $1.69-4.10$, $P=1.83 \times 10^{-5}$); see Figure 2g-h.

Predictive performance of APOLLO

APOLLO predicted the 36- and 60-month survival rates quite accurately in the TCGA training set and CGGA1 internal testing set ($AUC_{36\text{-month}}=0.933$ and 0.888 ;

$AUC_{60\text{-month}}=0.854$ and 0.851) (Figure 3a-b) and exhibited an excellent predictive ability in the CGGA2, Rembrandt, Weller and Gravendeel external testing sets ($AUC_{36\text{-month}}=0.898$, 0.893 , 0.844 and 0.861 , $AUC_{60\text{-month}}=0.896$, 0.817 , 0.806 and 0.790) (Figure 3c-f). In the meta-analysis, APOLLO presented an excellent accuracy in both training sets ($AUC_{36\text{-month}}=0.913$, $AUC_{60\text{-month}}=0.852$) and testing sets ($AUC_{36\text{-month}}=0.879$, $AUC_{60\text{-month}}=0.831$), and combined data ($AUC_{36\text{-month}}=0.901$, $AUC_{60\text{-month}}=0.843$). The calibration curve suggested a good accordance (Supplementary Figure S3). APOLLO significantly outperformed a basic model with the four covariates aforementioned (Supplementary Figure S4), improving AUC by 5.4% ($P < 2 \times 10^{-16}$) and 5.8% ($P < 2 \times 10^{-16}$) for the 36- and 60-month survival prediction, respectively (Supplementary Figures S5-S6). Additionally, APOLLO presented an excellent C-index in the TCGA training cohort (0.874) and CGGA1 (0.804) internal testing cohort and four external testing cohorts: CGGA2 (0.807), Rembrandt (0.772), Weller (0.787) and Gravendeel (0.759); and a pooled C-index of 0.818 (95% CI: $0.800-0.835$) (Figure 3i).

Clinical net benefits with APOLLO

With 36-month survival as the endpoint, DCA showed that APOLLO presented more clinical net benefits than several competing intervention strategies, namely, intervention for all, intervention for none, and intervention based on a basic model with only clinical and demographic indicators. Specifically, compared with the strategy of intervention for none and with a reasonable threshold probability (e.g., $P_t=0.4$), APOLLO presented a higher net benefit (NB) than the basic model ($NB_{APOLLO}=0.130$ vs $NB_{Basic}=0.111$). In other words, APOLLO identified 13.0 true positive patients per 100 patients that ought to have intervention, whereas only 11.1 for the basic model (Figure 4a). On the other hand, compared to the strategy of intervention for all, APOLLO presented a higher net reduction (NR) than the basic model ($NR_{APOLLO}=55.4\%$ vs $NR_{Basic}=52.5\%$). This means APOLLO can reduce the number of unnecessary clinical interventions by 55.4%, without missing interventions for any patients truly at high mortality risk; by comparison, only 52.5% for basic model (Figure 4b). As a sensitivity analysis and by varying the threshold probability from 0 to 0.5, the APOLLO decision curves were higher than those of the other strategies over a spectrum of threshold probability and APOLLO had the best average NB and NR in for 36- and 60-month survival ($NB_{36\text{-month}}=0.166$, $NR_{36\text{-month}}=40.1\%$ and $NB_{60\text{-month}}=0.258$, $NR_{60\text{-month}}=19.2\%$), indicating its uniform utility and suitability for clinical implementation (Figure 4a-d).³⁴ For individualized prognostic prediction and screening of high-risk patients, a nomogram of APOLLO is presented in Figure 4e.

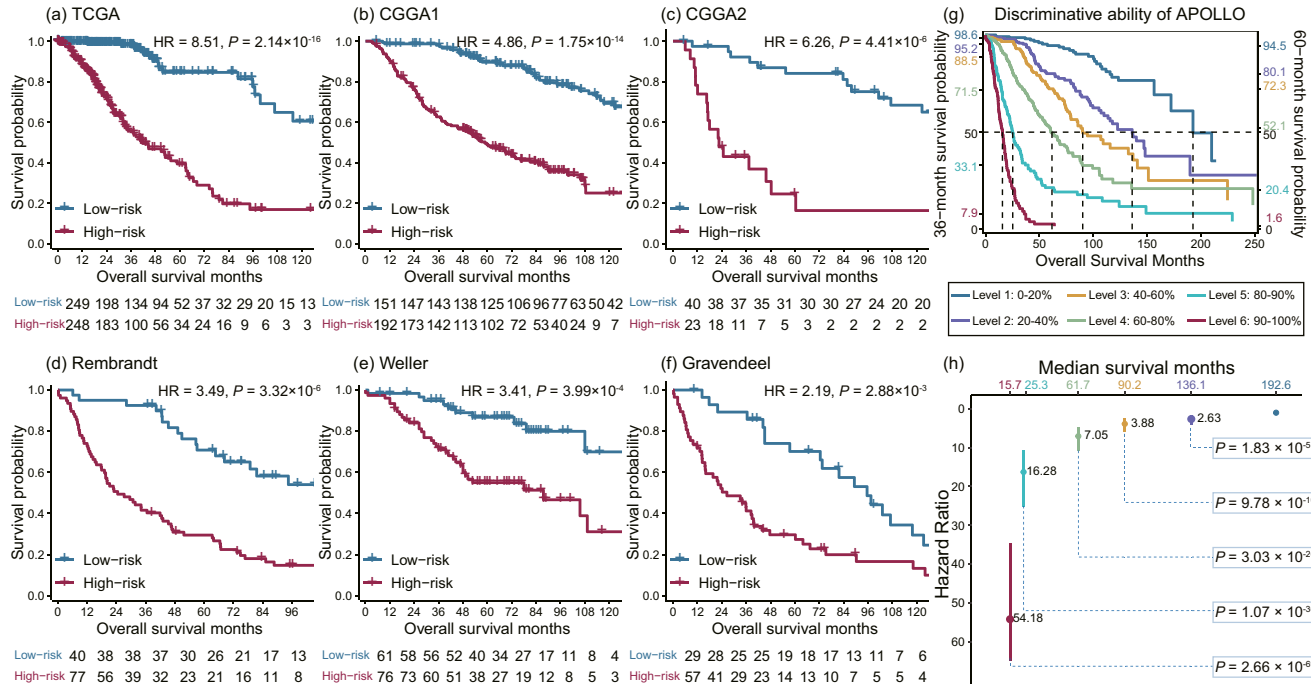


Figure 2. Kaplan-Meier survival curves of LGG patients stratified by APOLLO score.

Survival differences between high- and low-risk patients in (A) TCGA, (B) CGGA1, (C) CGGA2, (D) Rembrandt, (E) Weller and (F) Gravendeel cohorts. Patients in all six cohorts were categorized into two groups based on the same cutoff point: the median of APOLLO score defined in TCGA training set. (G) Discriminative ability of the APOLLO score by illustrating the 36- and 60-month survival rate, median survival month for six groups, defined by quantiles at 20%, 40%, 60%, 80% and 90% of APOLLO score as the cutoffs. (H) The hazard ratios (HRs) and P values for patients at different levels of APOLLO score (level 1 as reference), which were derived from a Cox proportional hazards model.

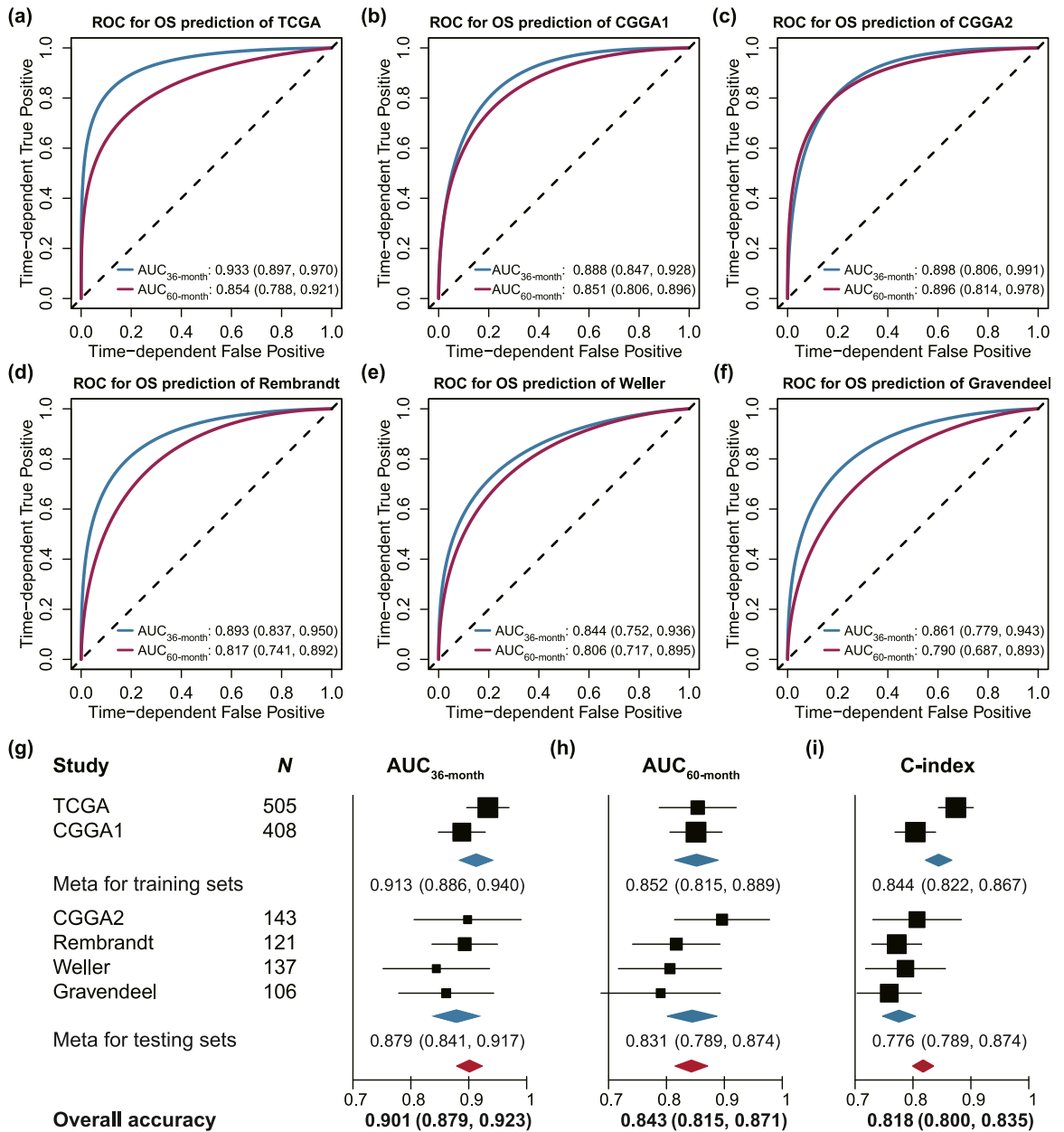


Figure 3. Time-dependent receiver operating characteristic curves of APOLLO for 36- and 60-month overall survival prediction.

The time-dependent ROC and AUC of APOLLO in (A) TCGA, (B) CGGA1, (C) CGGA2, (D) Rembrandt, (E) Weller and (F) Gravendeel cohorts, respectively. The pooled accuracy for (G) AUC_{36-month}, (H) AUC_{60-month} and (I) C-index of APOLLO across six independent cohorts.

Sensitivity analysis of APOLLO prediction

To assess the robustness of APOLLO, we performed a series of subgroup analyses with subgroups defined by age, gender, WHO grade, IDH mutation, 1p/19q status, MGMT promoter, radiotherapy and chemotherapy. In all the subpopulations examined, APOLLO presented good discriminative ability; the HRs that compare high- and low-risk groups within the subpopulations ranged

from 3.33 (95% CI: 2.45-4.52, $P=1.59 \times 10^{-14}$) to 8.77 (95% CI: 5.65-13.63, $P=4.54 \times 10^{-22}$) (Supplementary Figure S7a). Moreover, APOLLO had reasonable AUCs in all of these subpopulations, ranging from 0.829 (95% CI: 0.784-0.873) to 0.907 (95% CI: 0.875-0.940) for 36-month survival and 0.757 (95% CI: 0.705-0.810) to 0.921 (95% CI: 0.881-0.961) for 60-month survival (Supplementary Figure S7b-c).

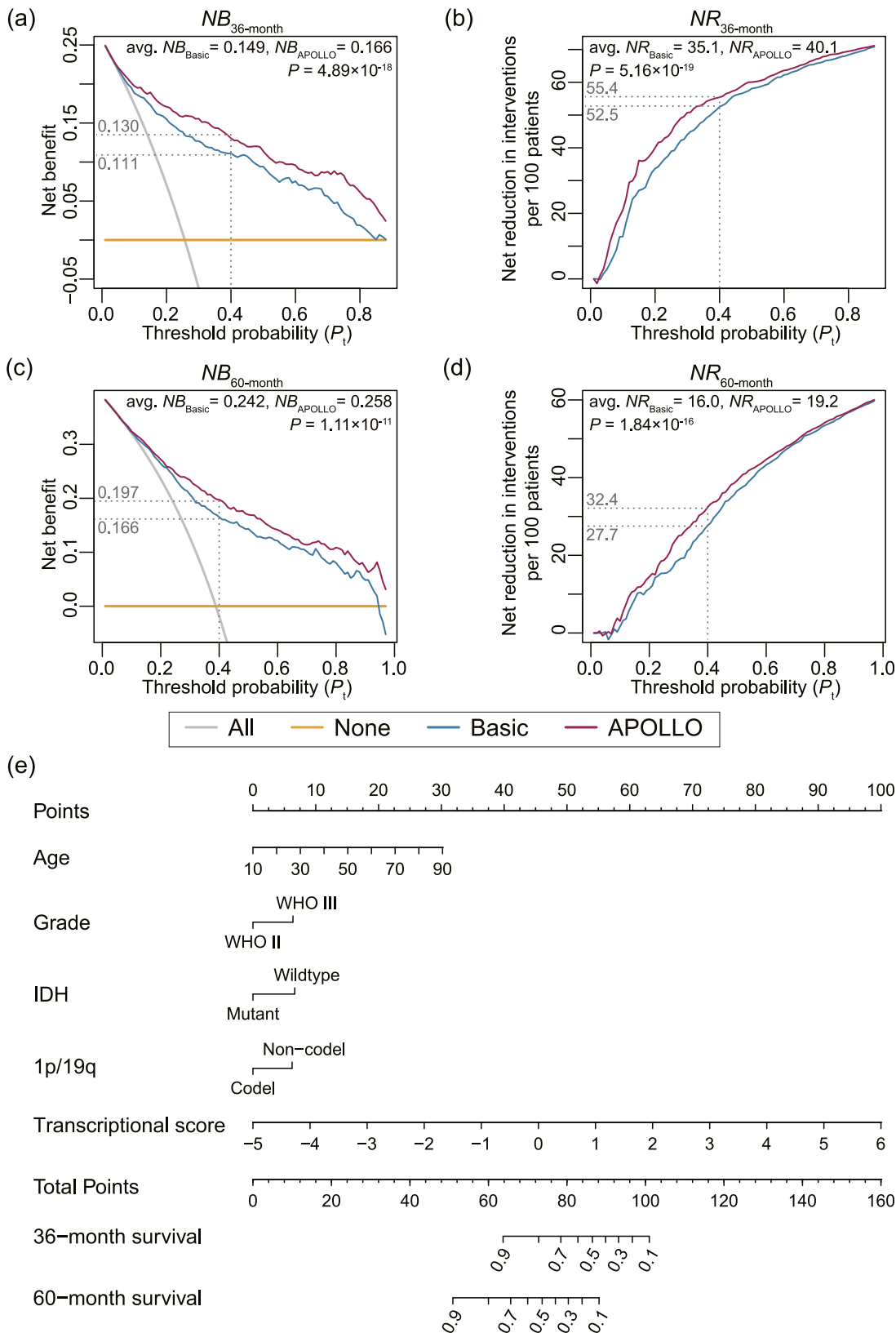


Figure 4. Decision curve analysis and nomogram for clinical application of APOLLO.

In real-world applications, missingness may happen, in which case we recommend to use the mean imputation to fill the missing values of genes before applying APOLLO. Our simulations verified the feasibility of mean imputation (Supplementary Table S9).

Comparison of APOLLO with existing models by a systematic review

Among the 54 screened articles (Figure 5), the prognostic models have various types of predictors: 31 (57.4%) models were developed based on gene expressions, 8 (14.8%) on lncRNA and 6 (11.1%) on radiomic features (Supplementary Table S10). A total of 30 (55.6%) models were constructed by integrating multi-level biomarkers, and 19 (35.2%) studies considered molecular mutations. Except for 4 models that were only applicable to LGG subgroups (2 for IDH-wild type LGG, 1 for Grade II LGG and 1 for LGG with epilepsy), all of the models were suitable for all LGG patients. While differing in biomarker selection methods, 52 (96.3%) models were derived using Cox models. Of the 35 models using clinical variables, age was the most common predictor ($n=34$), followed by grade ($n=27$), IDH mutation ($n=17$), gender ($n=8$) and 1p/19q status ($n=7$) (Supplementary Table S10).

The prediction accuracy of these published LGG prognostic models was extracted from the original paper and was summarized in Table 1 and Supplementary Table S11. While 8 studies had sample size $>1,000$, the rest only has small to modest sample sizes, which may not guarantee the reliability of the prediction model. The 24 (44.4%) models without any self-reported external validation should be used with caution; though the other 30 (55.6%) models were externally validated, 7 of which were not completely externally validated, as they used the validation sets to screen the predictors. Further, only 4 models had multiple validations (Supplementary Table S11). In general, among 22 models that were validated by completely external testing sets, their prediction accuracy varies (C-index=0.753, Range: 0.620-0.830; AUC_{3-year}=0.789, Range: 0.635-0.836 and AUC_{5-year}=0.720, Range: 0.594-0.807) and was in general smaller than that of APOLLO derived from four external testing sets (C-index=0.780, Range: 0.759-0.807; AUC_{3-year}=0.877, Range: 0.844-0.898 and AUC_{5-year}=0.812, Range: 0.790-0.896).

Discussion

Wide variation exists in LGG survival, ranging from 1 to over 10 years,^{1,21} and patients at high risk of mortality

may warrant close imaging monitoring and radical post-operative adjuvant therapy.⁵ Hence, there is an urgent need to develop accurate and robust prognostic prediction models for data-aided clinical decisions.³⁵ Leveraging available public LGG transcriptome data from six independent cohorts, we adopted a 3-D analysis strategy to screen biomarkers and developed APOLLO. Derived from a large LGG cohort (TCGA) and validated in 5 trans-ethnicity cohorts with European and Asian populations, APOLLO exhibited an excellent prediction accuracy in the training and testing sets. Further, it offered good clinical net benefits for screening patients with high risk of mortality. Our systematic review also confirmed that APOLLO outperformed existing prediction models.

As the utility and transportability of prediction models can be affected by gaps between the training population and the target population that the model is applied to,³⁶ we addressed this by proposing a 3-D analysis strategy, including Double types of effects, Double steps of screening and Double steps of modeling. The first one ensured the accuracy of the APOLLO by recognizing that G×G interactions which provided valuable insight into biological mechanisms of complex diseases.^{18,20} The latter two guaranteed the robustness of APOLLO. For example, our screening procedure identified biomarkers using a European population (TCGA) and validated those biomarkers using an Asian population (CGGA1). This trans-ethnic validation revealed robustness of the transcriptional predictors. In the ensuing modeling procedures, APOLLO was trained using a TCGA cohort and was later applied to one internal and 4 external cohorts (CGGA1, CGGA2, Rembrandt, Weller and Gravendeel), and retained excellent prediction accuracy regardless of stratification by age, gender, WHO grade, IDH mutation, 1p/19q status, MGMT promoter methylation level, and history of radiotherapy or chemotherapy.

According to Global Burden of Disease, there are over 1.71 million brain & nervous system cancer patients worldwide,¹ and 427.5 thousand (25%) are LGG. Assuming that LGG patients with probability of mortality ≥ 0.4 should be clinically intervened (Figure 5b and 5d), APOLLO yielded NR_{36-month}=55.4% and NR_{60-month}=32.4%, meaning that, compared to the most extreme strategy of offering interventions on every LGG patient, our model could help reduce 236.8 thousand (427.5 × 55.4%) and 138.5 thousand (427.5 × 32.4%) unnecessary interventions for short- and long-term survival outcome, respectively. In the future, APOLLO may, through customized biochips, offer maximized

The decision curve analysis for net benefit (NB) and net reduction (NR) of patients avoided unnecessary interventions at both 36-month (A-B) and 60-month (C-D) survival, respectively for APOLLO and the basic model composed of four common demographic and clinical predictors. (E) The nomogram for APOLLO. The value of each predictor can be converted into the corresponding points according to the axis in the top of nomogram. The sum of points for each predictor can correspond to the total points axis at the bottom of the nomogram and further used to estimate the patient's 36- and 60-month survival rate.

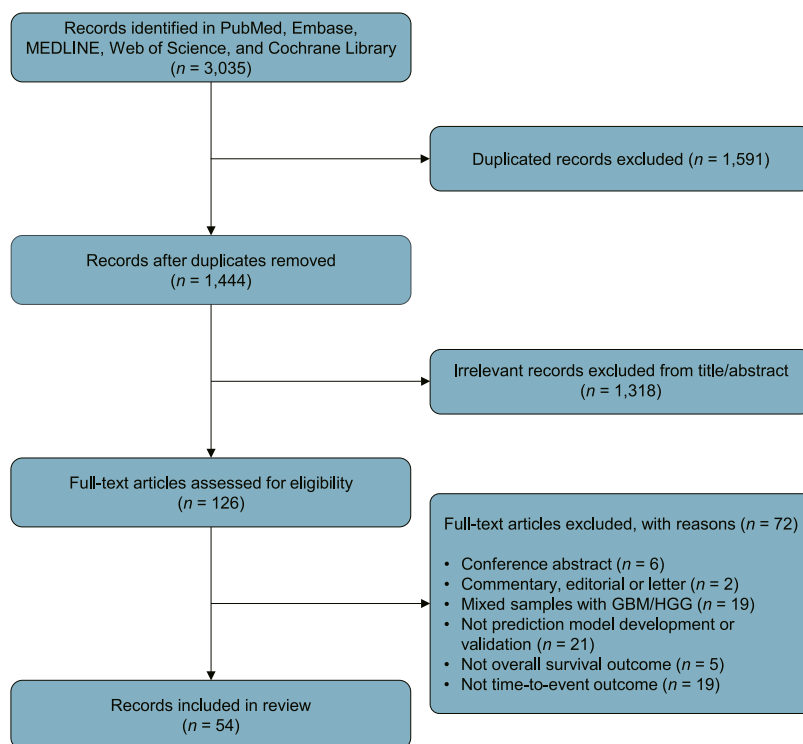


Figure 5. Flowchart of the systematic review of literature search and selection using five databases.

A total of relevant 3,035 records were obtained in PubMed, Embase, MEDLINE, Web of Science and Cochrane Library as of Aug 30, 2021. With the removal of duplicate records and irrelevant or ineligible records (based on title/abstract/text), retained were 54 records that met the criteria.

benefits to patients and provide cost-effective precision medicine. As such, our manuscript may present a proof of concept.

We found that APOLLO outperformed 54 models we reviewed in prediction accuracy and robustness. Further, we briefly summarized the biological functions of these transcriptional biomarkers in APOLLO. For the genes with significant main effects, the genetic variants of *CHIC2* are found in brain tumor tissues,³⁷ and *ITGAV*³⁸ is a prognostic factor of gliomas; *PLCG1* and *IGF2BP2* are related to SUMOylation and m6A methylation, and involved in the immune responses, occurrence and development of gliomas;³⁹⁻⁴¹ *MSN* is an active biomarker for glioma immune regulation and a drug target.⁴² For the pairs of genes with significant interactions, *PRF1* is strongly associated with anti-CTLA-4 or anti-PD-L1 immunotherapy, and is related to immune cell activities and survival of gliomas.⁴³ *BCORL1* is a transcriptional corepressor that can fuse with *ELF4*, and repress the activation of *PRF1*.⁴⁴ *HMGA1* and *TFG* are regulated by *NRF1*, and can affect the prognosis of gliomas.⁴⁵⁻⁴⁷ *FAS*⁴⁸ and *SMAD4*⁴⁹ are important members of the TNF-receptor superfamily and TGF- β signaling pathway, respectively, play a major role in tumor microenvironment and have antagonistic

interactions.⁵⁰ Though the biological function of the interaction between *CTNND2* and *GOLGA5* remains unclear, overexpressed *CTNND2* is likely to increase tumor invasion of gliomas.⁵¹

PRF1, *HMGA1*, *BCORL1*, *FAS*, and *MSN* are the top 5 transcriptional biomarkers that are the most correlated with immune checkpoint genes. Specifically, *PRF1* is viewed to be critically important for the immune cytolytic activity (CYT), reflecting the immune response of tumor cells, and is a well-established marker for cancer survival,⁵² including gliomas.⁵³ *HMGA1* contributes to the immunosuppressive microenvironment in tumors and the silencing of *HMGA1*, and can boost checkpoint blockade immunotherapy.⁵⁴ *BCORL1* is involved in the immune response pathway, impacting the response to immunochemotherapy.⁵⁵ *FAS* receptor signaling plays many important roles in the immune system, evidenced by that the tumoral *FAS* expression may predict the survival of CAR-T-treated patients.⁵⁶ *MSN*, a known target for cancer immunotherapy, regulates the migration of effector T cells.³⁷ Finally, genes included in APOLLO were transcriptional predictors with immune relevance, which can be immunotherapeutic targets.

Our study has several strengths. First, we performed, to our knowledge, the first systematic review of

No	PMID	Year	Method	ValidationType	Data source		Sample size			Performance in training set			Performance in testing set		
					Training	Testing	Training	Testing	Total	AUC _{3-year}	AUC _{5-year}	C-index	AUC _{3-year}	AUC _{5-year}	C-index
-	APOLLO	-	Cox	External	TCGA	CGGA2	505	143	1420	0.933	0.854	0.874	0.898	0.896	0.807
					CGGA1	Rembrandt	408	121		0.888	0.851	0.804	0.893	0.817	0.772
						GSE61374		137					0.844	0.806	0.787
						GSE16011		106					0.861	0.790	0.759
1	33665000	2021	Cox	External	TCGA	CGGA ^a	522	623	1145	0.766	0.763	-	0.744	0.764	-
2	34123829	2021	Cox	External	TCGA	CGGA ^a GSE16011 ^a + Rembrandt ^a	476	407 231	1114	-	-	0.878	-	-	0.734 0.748
3	33951297	2021	Cox	External	TCGA	CGGA	506	592	1098	0.710	0.601	-	0.655	0.655	-
4	33400376	2021	Cox	External	TCGA	CGGA	506	592	1098	0.782	-	-	0.734	-	-
5	33594759	2021	Cox	External	TCGA	CGGA	495	590	1085	0.875	0.816	-	0.756	0.728	-
6	34395274	2021	Cox	External	TCGA	CGGA1	474	407	1049	0.872	0.815	-	0.635	0.594	-
						CGGA2		168					0.775	0.807	
7	33381460	2020	Cox	External	TCGA	CGGA	525	420	945	-	0.633	-	-	0.671	-
8	34015817	2021	NN	External	TCGA	CGGA	493	408	901	0.925	0.871	-	0.795	0.767	-
9	33363544	2020	Cox	External	TCGA	CGGA	459	362	821	-	-	0.878	-	-	0.68
10	32519365	2021	Cox	External	TCGA	CGGA1	477	199	815	0.848	0.750	-	0.802	0.674	-
						CGGA2		139					0.828	0.755	
11	31824866	2019	Cox	External	TCGA	CGGA ^a	511	172	683	0.89	0.78	0.839			0.811
12	31803233	2019	Cox	External	TCGA	CGGA ^a	511	172	683	0.831	0.711	-	0.909	0.892	-
13	34408772	2021	Cox	External	TCGA	CGGA	495	172	667	0.84	0.74	-	0.74	0.71	-
14	32793593	2020	Cox	External	TCGA	CGGA ^a	476	170	646	0.860	0.806	0.817	0.783	0.759	0.642
15	31533943	2019	Cox	External	CGGA	TCGA	172	451	623	0.890	0.912	-	0.782	0.696	-
16	31921517	2020	Cox	External	TCGA	CGGA	456	159	615	0.878	0.827	-	0.806	0.807	-
17	24049111	2013	Cox	External	EORTC	RTOG+NCCTG	338	235	573	-	-	0.67	-	-	0.62

Table 1 (Continued)

No	PMID	Year	Method	ValidationType	Data source		Sample size			Performance in training set			Performance in testing set			
					Training	Testing	Training	Testing	Total	AUC _{3-year}	AUC _{5-year}	C-index	AUC _{3-year}	AUC _{5-year}	C-index	
18	29204839	2018	Cox	External	TCGA	CGGA	420	100	520	-	-	0.83	-	-	0.68	
19	32162004	2020	RSF	External	Local	TCIA	205	91	296	-	-	-	-	-	0.709 _{iAUC}	
20	30362964	2018	Cox	External	TCIA	CGGA	85	148	233	-	-	0.92	-	-	0.7	
21	33409797	2021	Cox	External	Local	Local	149	66	215	-	-	0.821	-	-	0.763	
22	32060714	2020	Cox	External	Local	TCIA	112	46	158	-	-	0.773 _{iAUC}	-	-	0.830 _{iAUC}	
23	32740813	2020	Cox	External	Local	TCIA	117	33	150	-	-	0.770 _{iAUC}	-	-	0.787 _{iAUC}	
24	32229719	2020	Cox	Internal	TCGA	CGGA1 ^a	329	405	1216	-	-	0.873	-	-	0.781	
				External		CGGA2	140	118				0.881				0.765
						GSE16011		88								0.721
						GSE61374		136								0.753
25	31853837	2020	Cox	Internal	TCGA	CGGA ^a	329	405	874	-	-	0.877	-	-	0.812	
				External			140					0.878				
26	32351547	2020	Cox	Internal	TCGA	CGGA	304	353	785	0.882	0.884	0.864	0.836	0.798	0.756	
				External			128					0.836				0.761
27	32431729	2020	Cox	Internal	TCGA	CGGA	297	353	774	0.905	0.837	0.870	0.798	0.740	0.753	
				External			124					0.915				0.828
28	33591634	2021	Cox	Internal	TCGA	CGGA	352	224	728	0.930	0.876	-	0.835	0.711	-	
				External			152					0.816				0.857

Table 1: Comparison of prediction accuracy between APOLLO and 28 models of LGG with self-reported external validation.

Abbreviations: NN: neural network; RSF: random survival forest; iAUC: integrated area under the time-dependent ROC curve; Internal: a model was cross validated by randomly splitting the original data. External: a model was externally validated by an independent external population. The performance for each model was extracted from the original paper.

^a Datasets were used for biomarker screening, which were not completely external validation.

prognostic prediction models for LGG and confirmed the good performance of APOLLO. Second, this is perhaps the largest molecular prognostic prediction study for LGG, and APOLLO was strongly overall, as well as trans-ethnically validated by several large LGG cohorts. Our extensive subgroup analysis suggested the robustness and transportability of APOLLO to different populations. Third, we proposed an effective 3-D strategy for biomarker screening and model construction, by focusing on biomarkers with important main effects or $G \times G$ interactions. The strategy struck a reasonable balance among statistical properties (false positive control vs statistical power gaining), model interpretations (main effects vs $G \times G$ interactions), and computational complexity (fast variable screening vs consistent model selection). Finally, we provided a web-based tool to facilitate the application of APOLLO.

We also acknowledge some limitations. First, heterogeneity existed across these cohorts with various sequencing or microarray platforms. To address this, we harmonized the data by performing standard normal transformation, which work to some degree. Second, some well recognized prognostic factors (e.g., tumor size and extent of surgical resection) were missing in several cohorts. We envision that there is much room for improvement with more available and complete clinical factors. Third, applications of APOLLO to the other ethnicity populations should be cautious, as APOLLO was trained and validated among the Asian and European populations. Forth, the improvement of accuracy was not uniform in all external validation datasets, possibly due to the population heterogeneity or the limited sample size in a single dataset. Finally, more biological experiments are needed to confirm gene functions of these transcriptional predictors used in APOLLO.

To conclude, we presented an Accurate and independently validated Prediction mOdel of Lower-grade gliomas Overall survival (APOLLO), which was demonstrated, by a systematic review, with the best prediction accuracy and robustness, and was a cost-effective strategy for screening LGG patients at high risk of mortality. A free and user-friendly online tool was developed at <http://bigdata.njmu.edu.cn/APOLLO>.

Contributors

Study design: J.C., R.Z., Y.W., S.S. and F.C.; Data collection and quality control: J.C., J.F. and D.C.C.; Analyses and interpretation: J.C., J.F., Y.W., S.S., W.D., X.D., Y.Z., L.L. Y.L. and Z.L.; Online tool: J.C., C.Z.; Systematic review: J. C., S.X. and J.X.; Manuscript draft: J.C., R.Z. and Y.L.; Manuscript revise: Y.L., Y.W., F.C., X.Q. and D.C.C.; All authors read and approved the final manuscript.

Data Sharing Statement

TCGA Database: <https://xenabrowser.net/>

CGGA Database: <http://www.cgga.org.cn/>

GEO Database: <https://www.ncbi.nlm.nih.gov/gds/>

COSMIC: <http://cancer.sanger.ac.uk/cosmic/>

Declaration of interests

All authors declare that they have no conflicts of interest.

Acknowledgements

The authors thank the patients and investigators who participated in TCGA, CGGA and GEO for providing the data. This study was supported by the National Key Research and Development Program of China (2016YFE0204900 to F.C.), National Science Foundation of Jiangsu Province (BK20191354 to R.Z.), National Natural Science Foundation of China (81973142 to Y.W. and 82103946 to S.S.), China Postdoctoral Science Foundation (2020M681671 to S.S.), the US National Institutes of Health (CA209414, CA249096, CA092824 and ES000002 to D.C.C., CA249096 and CA209414 to Y.L.), Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD). R.Z. was partially supported by the Qing Lan Project of the Higher Education Institutions of Jiangsu Province and the Outstanding Young Level Academic Leadership Training Program of Nanjing Medical University.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ebiom.2022.104007.

References

- Ostrom QT, Patil N, Cioffi G, Waite K, Kruchko C, Barnholtz-Sloan JS. CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2013–2017. *Neuro Oncol.* 2020;22(12 Suppl 2):iv1–iv96.
- Louis DN, Perry A, Reifenberger G, et al. The 2016 world health organization classification of tumors of the central nervous system: a summary. *Acta Neuropathol.* 2016;131(6):803–820.
- Kiran M, Chatrath A, Tang X, Keenan DM, Dutta A. A prognostic signature for lower grade gliomas based on expression of long non-coding RNAs. *Mol Neurobiol.* 2019;56(7):4786–4798.
- Nicholson JG, Fine HA. Diffuse glioma heterogeneity and its therapeutic implications. *Cancer Discov.* 2021;11(3):575–590.
- Song LR, Weng JC, Li CB, et al. Prognostic and predictive value of an immune infiltration signature in diffuse lower-grade gliomas. *JCI Insight.* 2020;5(8):e133811.
- Zhang H, Li X, Li Y, Chen B, Zong Z, Shen L. An immune-related signature for predicting the prognosis of lower-grade gliomas. *Front Immunol.* 2020;11:603341.
- Zhang M, Wang X, Chen X, Guo F, Hong J. Prognostic value of a stemness index-associated signature in primary lower-grade glioma. *Front Genet.* 2020;11:441.
- Wang C, Qiu J, Chen S, et al. Prognostic model and nomogram construction based on autophagy signatures in lower grade glioma. *J Cell Physiol.* 2021;236(1):235–248.
- Lin JZ, Lin N. A risk signature of three autophagy-related genes for predicting lower grade glioma survival is associated with tumor immune microenvironment. *Genomics.* 2021;113(1 Pt 2):767–777.

- 10 Wang JB, Li P, Liu XL, et al. An immune checkpoint score system for prognostic evaluation and adjuvant chemotherapy selection in gastric cancer. *Nat Commun.* 2020;11:6352.
- 11 Zhang Y, Ma W, Fan W, et al. Comprehensive transcriptomic characterization reveals core genes and module associated with immunological changes via 1619 samples of brain glioma. *Cell Death Dis.* 2021;12(12):1140.
- 12 Wang Z, Cheng W, Zhao Z, et al. Comparative profiling of immune genes improves the prognoses of lower grade gliomas. *Cancer Biol Med.* 2021. <https://doi.org/10.20892/j.issn.2095-3941.2021.0173>.
- 13 Gittleman H, Sloan AE, Barnholtz-Sloan JS. An independently validated survival nomogram for lower-grade glioma. *Neuro Oncol.* 2020;22(5):665–674.
- 14 Leu S, von Felten S, Frank S, et al. IDH/MGMT-driven molecular classification of low-grade glioma is a strong predictor for long-term survival. *Neuro Oncol.* 2013;15(4):469–479.
- 15 Qian Z, Li Y, Fan X, et al. Prognostic value of a microRNA signature as a novel biomarker in patients with lower-grade gliomas. *J Neurooncol.* 2018;137(1):127–137.
- 16 Zheng Y, Ji Q, Xie L, et al. Ferroptosis-related gene signature as a prognostic marker for lower-grade gliomas. *J Cell Mol Med.* 2021;25(6):3080–3090.
- 17 Zhang R, Lai L, He J, et al. EGN2 DNA methylation and expression interact with HIF1 α to affect survival of early-stage NSCLC. *Epigenetics.* 2019;14(2):118–129.
- 18 Chatterjee N, Shi J, Garcia-Closas M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nat Rev Genet.* 2016;17(7):392–406.
- 19 Li J, Li X, Zhang S, Snyder M. Gene-environment interaction in the era of precision medicine. *Cell.* 2019;177(1):38–44.
- 20 Zhang R, Chen C, Dong X, et al. Independent validation of early-stage non-small cell lung cancer prognostic scores incorporating epigenetic and transcriptional biomarkers with gene-gene interactions and main effects. *Chest.* 2020;158(2):808–819.
- 21 Brat DJ, Verhaak RG, Aldape KD, et al. Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. *N Engl J Med.* 2015;372(26):2481–2498.
- 22 Zhao Z, Meng F, Wang W, Wang Z, Zhang C, Jiang T. Comprehensive RNA-seq transcriptomic profiling in the malignant progression of gliomas. *Sci Data.* 2017;4: 170024.
- 23 Zhao Z, Zhang KN, Wang Q, et al. Chinese glioma genome atlas (CGGA): a comprehensive resource with functional genomic data from chinese glioma patients. *Genomics Proteom Bioinform.* 2021;19(1):1–12.
- 24 Gusev Y, Bhuvareshwar K, Song L, Zenklusen JC, Fine H, Madhavan S. The REMBRANDT study, a large collection of genomic data from brain cancer patients. *Sci Data.* 2018;5: 180158.
- 25 Weller M, Weber RG, Willscher E, et al. Molecular classification of diffuse cerebral WHO grade II/III gliomas using genome- and transcriptome-wide profiling improves stratification of prognostically distinct patient groups. *Acta Neuropathol.* 2015;129(5):679–693.
- 26 Gravendeel LA, Kouwenhoven MC, Gevaert O, et al. Intrinsic gene expression profiles of gliomas are a better predictor of survival than histology. *Cancer Res.* 2009;69(23):9065–9072.
- 27 Sondka Z, Bamford S, Cole CG, Ward SA, Dunham I, Forbes SA. The COSMIC cancer gene census: describing genetic dysfunction across all human cancers. *Nat Rev Cancer.* 2018;18(11):696–705.
- 28 Yoshihara K, Shahmoradgol M, Martínez E, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun.* 2013;4:2612.
- 29 Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods.* 2015;12(5):453–457.
- 30 Warde-Farley D, Donaldson SL, Comes O, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* 2010;38(suppl_2):W214–W220.
- 31 Zheng Y, Heagerty PJ. Semiparametric estimation of time-dependent ROC curves for longitudinal marker data. *Biostatistics.* 2004;5(4):615–632.
- 32 Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Mak.* 2006;26(6):565–574.
- 33 Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. *Diagn Progn Res.* 2019;3:18.
- 34 Fitzgerald M, Saville BR, Lewis RJ. Decision curve analysis. *JAMA.* 2015;313(4):409–410.
- 35 Duffau H. Paradoxes of evidence-based medicine in lower-grade glioma: to treat the tumor or the patient? *Neurology.* 2018;91(14):657–662.
- 36 Riley RD, Ensor J, Snell KI, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. *BMJ.* 2016;353:i140.
- 37 Holtkamp N, Ziegenhagen N, Malzer E, Hartmann C, Giese A, von Deimling A. Characterization of the amplicon on chromosomal segment 4q12 in glioblastoma multiforme. *Neuro Oncol.* 2007;9(3):291–297.
- 38 Schittenhelm J, Schwab EI, Sperveslage J, et al. Longitudinal expression analysis of αv integrins in human gliomas reveals upregulation of integrin $\alpha v \beta 3$ as a negative prognostic factor. *J Neuropathol Exp Neurol.* 2013;72(3):194–210.
- 39 Gouazé-Andersson V, Delmas C, Taurand M, et al. FGFR1 induces glioblastoma radioresistance through the PLC γ /Hif1 α Pathway. *Cancer Res.* 2016;76(10):3036–3044.
- 40 Guan S, He Y, Su Y, Zhou L. A risk signature consisting of eight m6A methylation regulators predicts the prognosis of glioma. *Cell Mol Neurobiol.* 2021. <https://doi.org/10.1007/s10571-021-01135-x>.
- 41 Li H, Wang D, Yi B, et al. SUMOylation of IGF2BP2 promotes vasculogenic mimicry of glioma via regulating OIP5-AS1/miR-495-3p axis. *Int J Biol Sci.* 2021;17(11):2912–2930.
- 42 Zhong H, Liu S, Cao F, et al. Dissecting tumor antigens and immune subtypes of glioma to develop mRNA vaccine. *Front Immunol.* 2021;12: 709986.
- 43 Roufas C, Chasiotis D, Makris A, Efstathiades C, Dimopoulos C, Zaravinos A. The expression and prognostic impact of immune cytolytic activity-related markers in human malignancies: a comprehensive meta-analysis. *Front Oncol.* 2018;8:27.
- 44 Suico MA, Shuto T, Kai H. Roles and regulations of the ETS transcription factor ELF4/MEF. *J Mol Cell Biol.* 2017;9(3):168–177.
- 45 Pang B, Fan H, Zhang IY, et al. HMGA1 expression in human gliomas and its correlation with tumor proliferation, invasion and angiogenesis. *J Neurooncol.* 2012;106(3):543–549.
- 46 Bhawe K, Felty Q, Yoo C, et al. Nuclear respiratory factor 1 (NRF1) transcriptional activity-driven gene signature association with severity of astrocytoma and poor prognosis of glioblastoma. *Mol Neurobiol.* 2020;57(9):3827–3845.
- 47 Wang L, Yang B, Xu Z, et al. NRF1-regulated CircNSUN2 promotes lymphoma progression through activating Wnt signaling pathway via stabilizing HMGA1. *Cell Cycle.* 2021;20(9):819–828.
- 48 Werner JM, Kuhl S, Stavrinou P, et al. Expression of FAS-L differs from primary to relapsed low-grade gliomas and predicts progression-free survival. *Anticancer Res.* 2017;37(12):6639–6648.
- 49 He SM, Zhao ZW, Wang Y, et al. Reduced expression of SMAD4 in gliomas correlates with progression and survival of patients. *J Exp Clin Cancer Res.* 2011;30(1):70.
- 50 Szondy Z, Pallai A. Transmembrane TNF-alpha reverse signaling leading to TGF-beta production is selectively activated by TNF targeting molecules: therapeutic implications. *Pharmacol Res.* 2017;115:124–132.
- 51 Shimizu T, Ishida J, Kurozumi K, et al. δ -Catenin promotes bevacizumab-induced glioma invasion. *Mol Cancer Ther.* 2019;18(4):812–822.
- 52 Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell.* 2015;160(1–2):48–61.
- 53 Wang ZL, Wang Z, Li GZ, et al. Immune cytolytic activity is associated with genetic and clinical properties of glioma. *Front Immunol.* 2019;10:1756.
- 54 Wang Y, Song W, Hu M, et al. Nanoparticle-mediated HMGA1 silencing promotes lymphocyte infiltration and boosts checkpoint blockade immunotherapy for cancer. *Adv Funct Mater.* 2018;28(36): 1802847.
- 55 Cai J, Liu P, Huang H, et al. Combination of anti-PD-1 antibody with P-GEMOX as a potentially effective immunotherapy for advanced natural killer/T cell lymphoma. *Signal Transduct Target Ther.* 2020;5(1):289.
- 56 Upadhyay R, Boiarsky JA, Pantsulaia G, et al. A critical role for fas-mediated off-target tumor killing in T-cell immunotherapy. *Cancer Discov.* 2021;11(3):599–613.
- 57 Ansa-Addo EA, Zhang Y, Yang Y, et al. Membrane-organizing protein moesin controls treg differentiation and antitumor immunity via TGF- β signaling. *J Clin Invest.* 2017;127(4):1321–1337.