Briefings in Bioinformatics, 00(00), 2019, 1–18

doi: 10.1093/bib/bbz040
Advance Access Publication Date: 18 April 2019
Problem Solving Protocol

Single-cell transcriptome-based multilayer network biomarker for predicting prognosis and therapeutic response of gliomas

Ji Zhang, Meige Guan, Qianliang Wang, Jiajun Zhang, Tianshou Zhou and Xiaoqiang Sun

Corresponding author: Xiaoqiang Sun, Zhong-shan School of Medicine, Sun Yat-Sen University, Guangzhou 510080, China. Tel.: +86-020-87330128; Fax: +86-020-87331209; E-mail: sunxq6@mail.sysu.edu.cn; xiaoqiangsun88@gmail.com

Abstract

Occurrence and development of cancers are governed by complex networks of interacting intercellular and intracellular signals. The technology of single-cell RNA sequencing (scRNA-seq) provides an unprecedented opportunity for dissecting the interplay between the cancer cells and the associated microenvironment. Here we combined scRNA-seq data with clinical bulk gene expression data to develop a computational pipeline for identifying the prognostic and predictive signature that connects cancer cells and microenvironmental cells. The pipeline was applied to glioma scRNA-seq data and revealed a tumor-associated microglia/macrophage-mediated EGFR/ERBB2 feedback-crosstalk signaling module, which was defined as a multilayer network biomarker (MNB) to predict survival outcome and therapeutic response of glioma patients. We used publicly available clinical data sets from large cohorts of glioma patients to examine the prognostic significance and predictive accuracy of the MNB, which outperformed conventional gene biomarkers and other methods. Additionally, the MNB was found to be predictive of the sensitivity or resistance of glioma patients to molecularly targeted therapeutics. Moreover, the MNB was an independent and the strongest prognostic factor when adjusted for clinicopathologic risk factors and other existing gene signatures. The robustness of the MNB was further tested on additional data sets. Our study presents a promising scRNA-seq transcriptome-based multilayer network approach to elucidate the interactions between tumor cell and tumor-associated microenvironment and to identify prognostic and predictive signatures of cancer patients. The proposed MNB method may facilitate the design of more effective biomarkers for predicting prognosis and therapeutic resistance of cancer patients.

Key words: single-cell RNA-seq; multilayer network biomarker; tumor microenvironment; cancer prognosis

Ji Zhang, PhD, is a research fellow and doctor at Department of Neurosurgery, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Sun Yat-Sen University Cancer Center.
Meige Guan, was a bachelor student at School of Mathematics, Sun Yat-Sen University.
Qianliang Wang, is a postdoctoral research fellow at School of Life Science, Sun Yat-Sen University.
Jiajun Zhang, is an associate professor at School of Mathematics, Sun Yat-Sen University.
Tianshou Zhou, is a professor at School of Mathematics, Sun Yat-Sen University.
Xiaoqiang Sun, PhD, is an associate professor at Zhong-shan School of Medicine, Sun Yat-Sen University. His research interests include computational systems biology and bioinformatics. He is a member of American Society for Microbiology (ASM) and China Society for Industrial and Applied Mathematics (CSIAM).
Submitted: 19 December 2018; **Received (in revised form):** 22 February 2019

Introduction

Multiple cell types interactively coexist in a multicellular tissue or organism [1]. The cellular functioning and fate is not only determined by the cell itself but is also coordinated by its local environment and its neighboring cells [2]. The mechanisms by which extracellular signals regulate intracellular gene expression are essentially important to understand cell–cell interactions and cellular fate determination in a heterogeneous cell environment [3, 4]. Defects in the cell microenvironment and dysregulation of microenvironmental signals [5] have been found to play significant roles in the occurrence and development of various diseases, such as cancer [6], inflammation and metabolic diseases [7].

An increasing amount of experimental studies have demonstrated that tumor–microenvironment interactions impose profound impacts on tumor growth, progression and drug resistance [8–12]. The tumor microenvironment is the exterior environment of a living tumor cell and includes various types of cells (e.g. fibroblasts, endothelial cells and immune cells) and biochemical factors (e.g. cytokines, growth factors and chemokines) [13]. The tumor–microenvironment interaction occurs in two ways: physical interaction through direct contact, such as cell adhesion [14], and cell–cell communication through the exchange of secretory molecular factors, primarily consisting of ligand–receptor interactions [15].

Traditional experimental studies have primarily focused on a few or linear signaling pathways that are involved in cell–cell interactions. How best to systematically investigate the signaling network of cell communication is a challenging but invaluable task. Moreover, the signaling network involved in cell–cell interactions contains not only intercellular signaling pathways but also intracellular signaling transduction and gene expression [16]. Therefore, a systematic, multilayer, intercellular and intracellular signaling network is required to elucidate the signaling mechanisms that underlie the interplay between the microenvironment and tumor cells.

The recently emerging single-cell RNA sequencing (scRNA-seq) technologies provide a powerful tool to fulfill this goal [17–20]. The scRNA-seq transcriptome data can be used to identify cell types and to quantify cell type-specific gene expression in mixed cell populations, enabling us to resolve tumor–microenvironment interactions and dissect the microenvironment-mediated intercellular and intracellular signaling pathways of tumor cells. Recently, Boisset *et al.* [21] developed a cellular network approach to investigate physical cell interactions based on single-cell mRNA sequencing data. This approach, however, did not account for ligand–receptor-mediated cell communications. Zhou *et al.* [22] constructed a map of the cell–cell communication network of melanoma using scRNA-seq data that specified the gene expression of ligand–receptor interactions [23]. Similarly, Skelly *et al.* [24] used scRNA-seq data to construct large-scale networks of intercellular communication in the mouse heart. However, these studies did not consider the downstream signaling cascades that follow receptor activation, thus falling short in the intracellular signaling network mechanisms.

In this study, we developed a novel scRNA-seq data-based approach to reconstruct a multilayer signaling network that contains pathways from intercellular ligand–receptor interactions, intracellular transcriptional factors (TFs) and their target genes. We combined scRNA-seq data [19] with bulk gene expression data and clinical information of glioma patients as a case study to illustrate our method. Our approach was to build an immune cell-mediated multilayer signaling network of tumor cells, which revealed crosstalk and feedback loops of the EGFR signaling module and identified a novel multilayer network biomarker (MNB) that was proven to be of prognostic and predictive value for glioma patients. Moreover, we compared the predictive performance of the MNB with that of other existing biomarkers used for gliomas and other computational methods of biomarker identification. We also tested the robustness of the MNB using additional data sets and bootstrapping approach. Furthermore, we developed a multiscale model to mechanistically depict the intrinsic interplay between intercellular and intracellular signals in the MNB to interpret its superiority in comparison to the conventional gene signatures.

Methods

A schematic illustration of the multilayer intercellular/intracellular signaling network reconstruction, modeling and prediction is shown in Figure 1. We first analyzed the scRNA-seq data to infer the cell types based on the biological properties of marker genes. Based on the specificity of gene expression in different cell types, we constructed a multilayer intercellular/intracellular signaling network to identify MNB as a prognostic signature and therapeutic biomarker of glioma patients. Furthermore, a multiscale model was developed to mechanistically understand the functional role of dynamic interaction and regulation of components in the MNB in determining cellular fate and driving heterogeneous response kinetics.

Dimensionality reduction and clustering of scRNA-seq data

The scRNA-seq data of glioma patients [19] were downloaded from the GEO database with accession number GSE89567. The data included 6341 cell samples, each with 23 687 genes. We represented the data as an $m \times n$ matrix ($m = 23{,}687$, $n = 6341$), with columns corresponding to each cell and rows corresponding to the expression of each gene.

Since the scRNA-seq data are high dimensional and encompassing various types of noise, so, in this study, we employed a kernel-based manifold learning method, diffusion map (DM) [25, 26], to reduce the dimensions of the scRNA-seq data. DM uses the Gaussian kernel function to define the diffusion distance between any two nodes (data points) and then constructs random walk process on graphs using the graph Laplacian normalization method. DM can filter noisy and nonessential information during the diffusion process. Massive cells can be embedded and visualized into a low-dimensional (i.e. 2D–3D) space based on their gene expression similarity.

Based on the dimension-reduced data, we employed the shared-nearest-neighbor (SNN) algorithm [27] to cluster the cell samples. SNN is a density-based unsupervised clustering algorithm that does not require prior knowledge of the appropriate number of clusters and can effectively eliminate the noise data points. We used the 'Seurat' R package to perform the dimension reduction and cell clustering. In the current study, most parameters in functions 'RunDiffusion' (for DM-based dimension reduction) and 'FindClusters' (for SNN clustering) in 'Seurat' package were set as default, except that parameter 'resolution' in 'FindClusters' was set to 0.03 for getting the consistent clustering results with the previous study [19]. A larger value of 'resolution' would result in an increasing number of clusters.

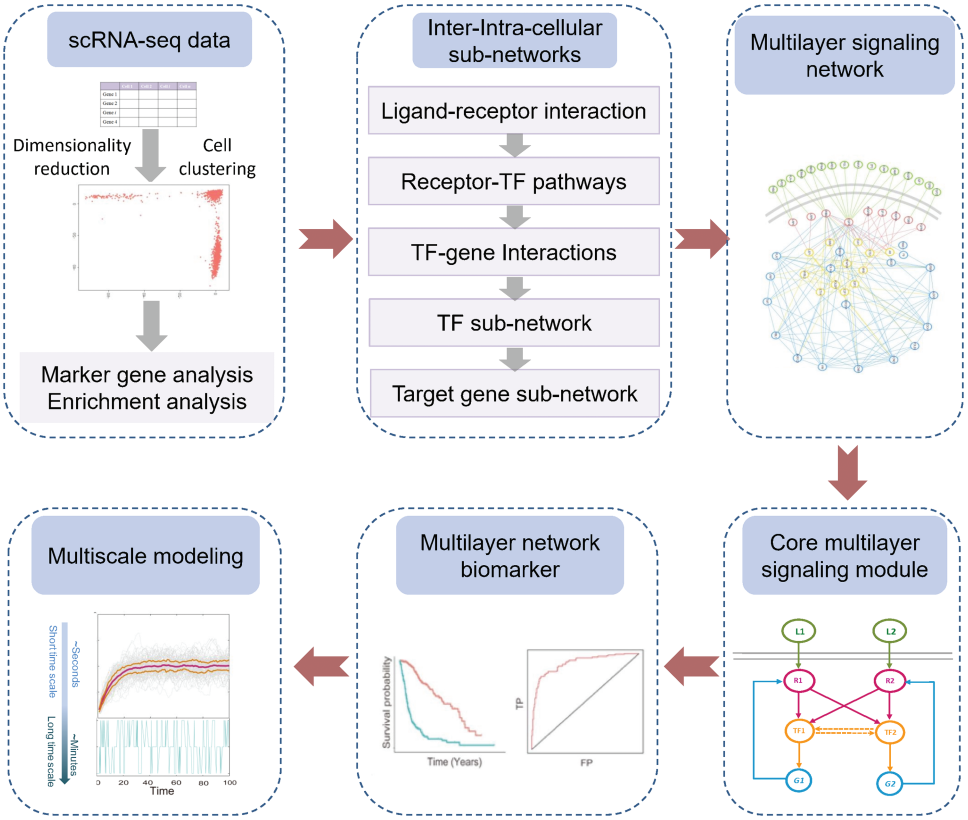

Figure 1. Schematic illustration of the scRNA-seq data-based multilayer intercellular/intracellular signaling network reconstruction, prediction and modeling. (**A**) We first analyzed the scRNA-seq data and conducted unsupervised clustering to infer cell types based on the biological properties of marker genes of each cell cluster. (**B**) Based on cell type-specific gene expression, we then constructed the intercellular pathways (ligand–receptor interactions) and intracellular sub-networks (receptor-TF pathways, TF–target gene interactions, TF sub-network and target gene sub-network). (**C**) As a result, a multilayer intercellular/intracellular signaling network was built by integrating the above sub-networks. (**D**) A core multilayer signaling module with crosstalk and feedback loops was extracted and analyzed. (**E**) An MNB was defined based on bulk gene expression data and clinical information to predict survival outcome and therapeutic response of cancer patients. (**F**) A multiscale simulation model was developed to mechanistically understand the dynamic interplay between intercellular and intracellular signals in the MNB.

Marker gene selection and cell type identification

Following unsupervised clustering, we selected and analyzed cell cluster-specific marker genes to infer cell types. A marker gene or differentially expressed gene (DEG) for each of the cell clusters can be viewed as a binary classifier, distinguishing this cell type from the other cells. The area under the curve (AUC) of the receiver operating characteristic (ROC) was used to evaluate the performance of each classifier and to select marker genes. We used the 'FindMarkers' function with parameters logfc.threshold = 2 and test.use = 'roc' in the 'Seurat' R package [28, 29] to identify cluster-specific marker genes. Several known marker genes of some cell types were also used to infer specific cell types. The functional enrichment of DEGs was also used to verify the cell type identification results based on the functional association network of DEGs in the STRING database (https://string-db.org/).

Constructing primary intercellular signaling pathways

We collected ligand–receptor pairing information from databases, including DLRP, IUPHAR, HPMR, HPRD and STRING and previous studies [30], and formed them into a list containing 2558 ligand–receptor directed pairs (Supplementary Table S1), denoted as $E = \{(ligand_i, receptor_i)\}$. This list provides the basis for constructing the ligand–receptor-mediated intercellular signaling network.

Based on the above data clustering and cell type identification, we then analyzed cell type-specific gene expression. We extracted the expression of each secretory ligand gene and cell surface receptor gene in each cell type. To reduce the complexity and false positives of the intercellular signaling network construction, we only considered highly expressed ligand and receptor genes. In this work, a gene was considered to be highly expressed in a specific cell type if the fold-change value (log scale) was greater than a threshold (e.g. 2) and the P-value from the t-test was less than 0.05. The set of highly expressed ligand genes in cell type A and receptor genes in cell type B were denoted as L_H^A and R_H^B, respectively. The primary intercellular signaling pathways from cell A to cell B were defined as $E_\cap\left(L_H^A \times R_H^B\right)$ based on the above ligand–receptor interaction paths and cell type-specific gene expression. These ligand–receptor interaction paths in $E_\cap\left(L_H^A \times R_H^B\right)$ were assumed to be the most probabilistically highly activated and physiologically important [23].

Constructing the intracellular signaling network

The intracellular signaling network includes three layers: receptors, TFs and target genes. The Fisher's exact test and LASSO regression models were used to evaluate the significance of TF activation and regulation relationships among TFs and their target genes. The TF–target gene networks were then connected

to the receptor layer by evaluating the significance of the connected pathways.

Constructing TF–target gene interactions

We collected human TF–target interactions from the TRED and KEGG databases, resulting in a TF–target gene pairing list (Supplementary Table S2). All upregulated genes in tumor cells (versus other cells) were denoted as T_{up}. For each TF, L_{TF} denoted a collection of its target genes. In addition, L_{all} denoted a collection of all annotated genes. We then evaluated the significance of the TF activation using the Fisher's exact test by calculating the following probability:

$$P = \binom{a+b}{a}\binom{c+d}{c}\Big/\binom{a+b+c+d}{a+c}, \quad (1)$$

where $\binom{n}{k}$ represents the binomial coefficient, $a = |T_{up} \cap L_{TF}|$ represents activated target genes, $b = |T_{up}| - a$ represents activated nontarget genes, $c = |L_{TF}| - a$represents inactivated target genes and $d = |L_{all}|$-a-b-c represents inactivated nontarget genes. A TF was considered to be significantly activated if the above P-value was less than 0.05.

Constructing TF sub-network and target gene sub-network

The TF regulatory sub-network was modeled using the following linear weighted model:

$$x_i = \sum_{j=1, j\neq i}^{L} a_{ij} x_j + c_i, \quad i = 1, 2, \cdots, L, \quad (2)$$

where x_i represents the expression level of each TF and L is the number of all activated TFs. The parameters (a_{ij}, c_i) in the above model were estimated by fitting the model to the experimental data (x_i^{exp}) using the LASSO regression method as follows:

$$\min_{a_{ij}, c_i} \left(\left\| x_i^{exp} - \sum_{j=1, j\neq i}^{L} a_{ij} x_j - c_i \right\|_{L^2}^2 + \lambda_i \|a_{ij}\|_{L^1}^2 \right), \quad i = 1, 2, \cdots, L, \quad (3)$$

where λ_i is a weight parameter for the penalty term. The R package 'glmnet' [31] was used to optimize the above objective function. A 10-fold cross validation was performed to select the optimal values of λ_i by minimizing the mean cross-validated error. The estimation of parameters was then obtained as the regression coefficients at the optimal penalty weight. The nonzero coefficient of a_{ij} was used to construct a sparse TF sub-network.

The target gene regulatory sub-network was constructed using a similar approach, i.e. a linear weighted model and the LASSO regression method.

Connecting receptor-TF pathways

To connect intracellular sub-networks of TFs and target genes to the intercellular signaling network, we needed to find pathways that connected the receptors to the TFs. We used the information from the KEGG PATHWAY database to extract the shortest pathways that connected receptors to the downstream TFs. The significance of the connected pathways' activation was evaluated using the Fisher's exact test, similar to that in Equation (1).

Multilayer intercellular/intracellular signaling network

The above five sub-networks (i.e. intercellular pathways, TF–target gene interaction, TF sub-network, target gene sub-network and receptor-TF pathways) were denoted as N_1, N_2, N_3, N_4 and N_5, respectively. The multilayer intercellular/intracellular signaling network could then be constructed as $N_{total} = N_1 \cup N_2 \cup N_3 \cup N_4 \cup N_5$. The cytoscape software was used to visualize and characterize each of the above sub-networks and the integrated network.

Evaluation of prognostic significance and accuracy of the MNB

The EGFR-mediated MNB consists of 14 genes: ligand genes (CAMP, ZP3, AREG, ANXA1, HBEGF, SPINK1, ICAM1, FGL1, HLA-A and SEMA4D), receptor genes (EGFR and ERBB2) and TFs (ESR2 and ELK1). To assess the prognostic significance of the proposed MNB, we collected clinical information and RNA-seq data of glioma patients from The Cancer Genome Atlas (TCGA) database (https://cancergenome.nih.gov/) and Chinese Glioma Genome Atlas (CGGA) database (http://www.cgga.org.cn/). By matching patient sample IDs from the clinical information and the gene expression data, we prepared a TCGA data set (N = 689) and a CGGA data set (N = 310) for survival analysis. A multivariate Cox proportional hazards (PH) model [32] was built to compute the hazard function at time t for each patient as follows:

$$H(t|X) = H_0(t) \cdot \exp\left(\sum_{i=1}^{m} \lambda_i \cdot y_i\right), \quad (4)$$

where y_i is the expression level of gene i involved in the MNB and m is number of genes in the MNB.λ_i is the regression coefficient of gene i in the Cox PH model. We then formulated the following risk score (RS) for each patient based on the expression level of the MNB:

$$RS = \sum_{i=1}^{m} \lambda_i \cdot y_i. \quad (5)$$

An MNB-based risk signature was trained from the TCGA data set: RS = 0.0256 × CAMP + 0.2140 × ZP3 + 0.1258 × AREG + 0.5908 × ANXA1 − 0.0922 × HBEGF + 0.1946 × SPINK1 + 0.0353 × ICAM1 + 0.3163 × FGL1 + 0.1571 × HLA-A + 0.0895 × SEMA4D − 0.0637 × EGFR + 0.0799 × ERBB2 + 0.1117 × ESR2 − 0.1106 × ELK1. The same risk signature was used to compute the RSs for patients in the CGGA data set. The patients in each data set were classified into a high-risk group and a low-risk group according to the optimal cut-off value of their RS that maximizes the sum of sensitivity and specificity using the ROC method. K-M survival curves were plotted for patients in the high-risk and low-risk groups. The statistical significance of difference between two K-M curves was assessed using the two-sided log-rank test.

To further investigate the predictive accuracy of prognostic classification with MNB, we used time-dependent ROC analysis [33]. The above RS was used to predict the 1 year, 3 year and 5 year survival of patients in the TCGA data set and CGGA data set.

Comparisons with other related signatures and other methods

To compare the prognostic accuracy of the MNB with other risk signatures for predicting the overall survival of glioma patients,

we calculated the AUC of ROC of the MNB, EGFR and ERBB2 signatures and a signature of Cheng *et al.* [34] that consists of the immune-related genes FOXO3, IL6, IL10, ZBTB16, CCL18, AIMP1, FCGR2B and MMP9. The statistical significance of the difference between the areas under the two ROC curves was assessed using DeLong's test [35].

To assess whether the MNB was independently correlated with the prognosis of glioma patients, we conducted univariate and multivariate Cox regression analyses of clinicopathologic factors and available gene signatures. Clinicopathologic information, including age, gender and grade, was available for glioma patients in both the TCGA, while information of age, gender and IDH mutation status was available for the CGGA data set. We also included the following gene signatures in the multivariate Cox regression analyses: the MNB signature newly proposed in this study; an EGFR + ERBB2 gene signature studied by many groups [36, 37]; and a recently published signature for predicting the prognosis of glioma patients, i.e. the immune-related signature [34].

The above risk factors that were significantly correlated with the overall and 5 year survival of glioma patients in both the TCGA and CGGA data sets were extracted for construction of a combined signature using the LASSO method for variable selection. As a result, we defined the combined signature as follows: $CS = (0.87213085 \times \text{Age}) + (0.73262929 \times \text{Grade}) + (0.63292736 \times \text{MNB}) + (0.04822249 \times \text{Immune-related signature})$. Here, Grade = 1 stands for lower-grade glioma (LGG), and Grade = 2 stands for glioblastoma multiforme (GBM). The prognostic accuracy of the MNB was compared with that of the combined signature.

In addition, we compared the predictive accuracy and robustness of the MNB with that of LASSO Cox PH model [38] and random forest methods for survival prediction (such as RF-SRC [39]).

Multiscale modeling and simulation of EGFR signaling module

To mechanistically understand the interplay between intercellular signaling pathways and intracellular gene expression regulation dynamics, we developed a multiscale model to simulate the fast signaling transduction and the relatively slow gene expression across multiple temporal scales, ranging from seconds to minutes. The multiscale model employed a hybrid approach that coupled ordinary differential equations (ODEs) and stochastic simulation to describe the ligand–receptor-TF signaling transduction and the gene expression, respectively. The details of the modeling development were described in Supplementary Text S1.

Results

Cell type identification and gene expression specificity based on scRNA-seq

We employed the DM method and SNN algorithm to conduct an unsupervised clustering of cell samples based on single-cell gene expression data (see Methods section). A sample of 6341 cells was clearly divided into 3 main clusters (Figure 2A) containing 104 cells, 5168 cells and 1069 cells. Supplementary Figure S1A shows the expression profile of marker genes in each cell cluster. Figure 2B shows the specificity of the expression of marker genes of cluster 1 and cluster 3 (Supplementary Figure S1B and C), demonstrating the effectiveness of the selection of the marker genes for identifying cell types.

Human brain glioma tissue comprises at least tumor cells, immune cells, such as tumor-associated microglia/macrophages (TAMs), and normal neural/glial cells such as oligodendrocytes [13]. There are some known significant marker genes for TAMs (e.g. CD14, CD163 and CX3CR1 [40]) and oligodendrocytes (e.g. MBP, MOBP, PLLP and CLDN11 [41]). Figure 1C shows the expression of oligodendrocytic marker genes, which were preferentially expressed in cluster 1 cells, leading to the inference of the cluster 1 cells as oligodendrocytes. TAM-specific marker genes were preferentially expressed in cluster 3 cells (Figure 1D), suggesting that cluster 3 cells consisted of TAMs.

To verify the inference of cell type, we performed a functional enrichment analysis of the marker genes in each cell type. The marker genes of cluster 1 cells were significantly enriched in axon ensheathment, myelination, oligodendrocyte differentiation, central nervous system myelination, oligodendrocyte development etc. (Figure 1E). These biological functions support the inference of cluster 1 cells as oligodendrocytes, a type of neuroglia that functions to provide support and insulation to axons in the central nervous system [42]. Meanwhile, the marker genes of cluster 3 cells were significantly enriched primarily in inflammatory response and immune response functions (Figure 1F), supporting the inference of the cluster 3 cells as TAMs, a known type of immune cell.

Taken together, we identified that cluster 1 cells were oligodendrocytes and cluster 3 cells were TAMs. As such, the remaining cluster 2 cells were considered mainly as tumor cells, in accordance with [19]. In this way, we obtained the cell type-specific gene expression from the scRNA-seq data, which provided a basis for the following multilayer intercellular/intracellular signaling network construction.

Multilayer signaling network reveals a TAM-mediated EGFR signaling module

TAMs are a type of abundant immune cell in the tumor-associated microenvironment, playing important roles in tumor growth and drug resistance [43–46]. We therefore investigated the interactions between glioma cells and TAMs by constructing the primary intercellular signaling pathways between them, as described in the Methods section. We further constructed the intracellular signaling networks of tumor cells in response to the stimulus of cytokines secreted by TAMs. The highly expressed genes, including 62 TAM-secreted ligand genes, 68 receptor genes as well as 12 TFs and 21 target genes in tumor cells, were included in the construction of the intercellular and intracellular networks. The intracellular signaling network included three layers: receptor layer, TF layer and target gene layer. The Fisher's exact test was used to evaluate the significance of TF activation, and LASSO regression models were built to construct sub-networks for TFs and their target genes (see the Methods section). The TF layer and target gene layer sub-networks were then connected to the receptor layer to construct the intracellular signaling transduction network. Supplementary Figure S2A–E shows the constructed primary intercellular signaling pathways, TF–target gene interactions, TF sub-network, target gene sub-network and receptor-TF pathways, respectively.

A multilayer signaling network (Figure 2G) was constructed by integrating the above intercellular pathways and intracellular sub-networks. The reconstructed network represented

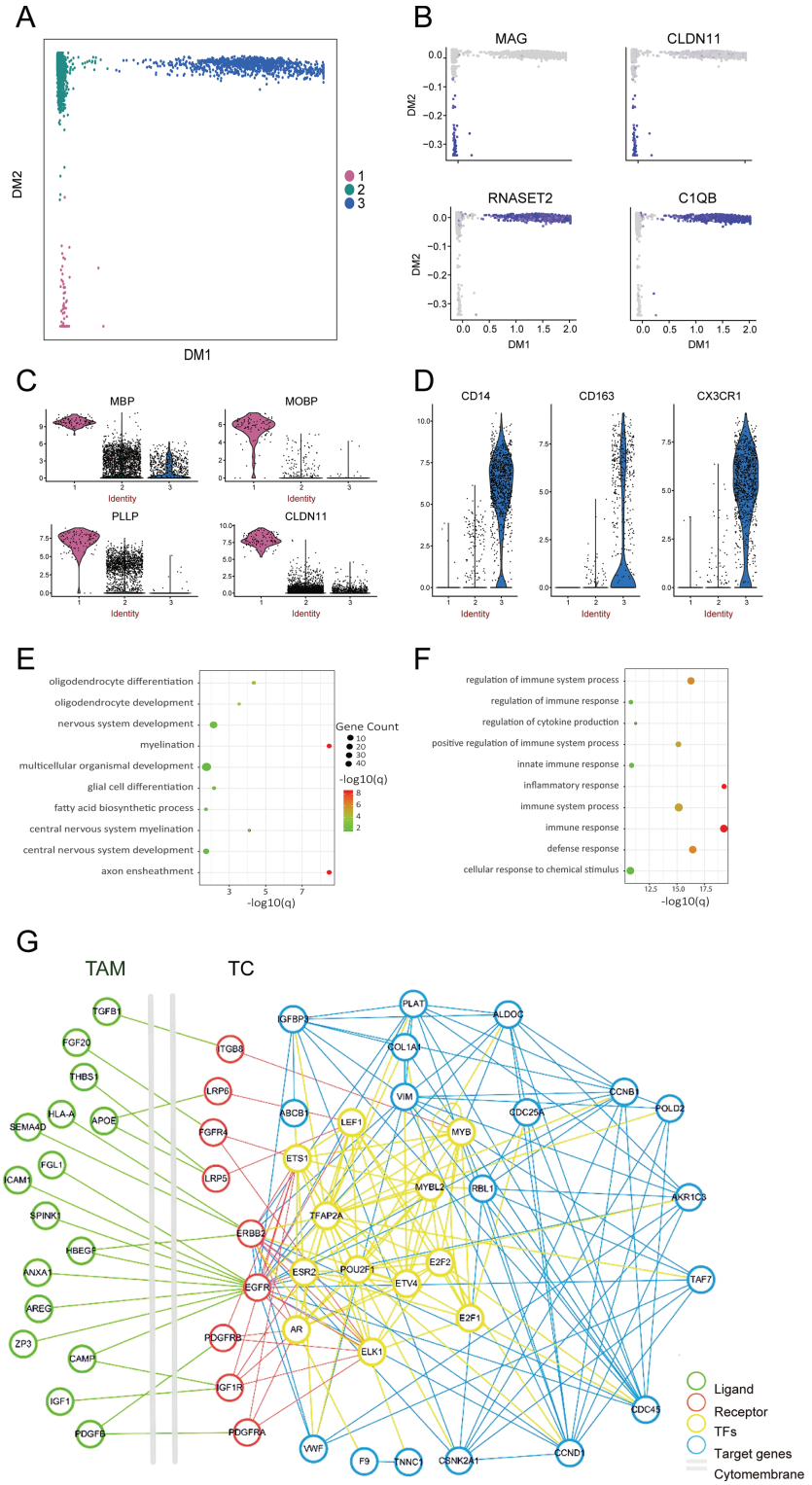

Figure 2. Cell type identification, marker gene analysis and multilayer network construction. (**A**) The cell clustering shows that 6341 cells were divided into 3 clusters, containing 104 cells, 5168 cells and 1069 cells, respectively. (**B**) The expression of the first two marker genes in cluster 1 cells (i.e. MAG and CLDN11) and cluster 3 cells (i.e. RNASET2 and C1QB), respectively. (**C**) Violin plots showing the expression of oligodendrocytic marker genes, including MBP, MOBP, PLLP and CLDN11 [41], which were highly expressed in cluster 1 cells, indicating cluster 1 cells as oligodendrocytes. (**D**) The expression of known marker genes of TAMs, including CD14, CD163 and CX3CR1 [40]. These marker genes were highly expressed in cluster 3 cells, indicating cluster 3 cells as TAMs. (**E** and **F**) Functional enrichment of marker genes of cluster 1 cells and cluster 3 cells, respectively. The top enriched gene ontology terms were listed with enrichment scores, shown as $-\log_{10}(q)$, where q is the significance of the enrichment after multiple testing. (**G**) Multilayer intercellular/intracellular signaling network of tumor cells (TC) activated by TAMs. The nodes with different colors represent ligands (green), receptors (red), TFs (yellow) or target genes (blue). Green lines represent ligand–receptor signaling pathways from TAMs to tumor cells. Red lines represent receptor-TF signaling pathways within tumor cells. Yellow lines represent correlations between TFs. Blue lines represent correlations between target genes.

Table 1. Experimental evidences for functional roles of the MNB genes in cancer progression. The table lists the gene name, the biological functions in cancers and the corresponding PubMed IDs and literature references for each TAM-related gene

Symbol	Gene name	Functional roles in cancers	PMIDs	References
EGFR	Epidermal growth factor receptor	Promotes glioma growth and angiogenesis	22139077	[48]
ERBB2	Erb-B2 receptor tyrosine kinase 2	Regulation of cell death	19578738	[49]
ESR2	Estrogen receptor 2	Tumor suppressor; suppression of GBM	27126081	[50]
ELK1	ELK1, ETS transcription factor	Promotes glioma proliferation, migration and invasion	25707769	[51]
CAMP	Cyclic adenosine monophosphate	Inhibits proliferation and proapoptotic of glioma stem cells	30171713	[52]
ZP3	Zona pellucida glycoprotein 3	Not reported	–	–
AREG	Amphiregulin	Promotes immunosuppression, dampens local inflammation and mediates type 2 immune resistance	28741529	[53]
ANXA1	Annexin A1	Enhances cancer growth and migration	29263330	[54]
HBEGF	Heparin-binding EGF like growth factor	Promotes glioblastoma invasion through activation of EGFR	22294205	[55]
SPINK1	Serine peptidase inhibitor, Kazal type 1	Promotes epithelial–mesenchymal transition through EGFR signaling pathway	24619958	[56]
ICAM1	Intercellular adhesion molecule 1	Antitumor immunity	10463777	[57]
FGL1	Fibrinogen-like 1	Promotes cell proliferation, invasion and migration	29845203	[58]
HLA-A	Major histocompatibility complex, class I, A	Adaptive immune responses against tumors	23852952	[59]
SEMA4D	Semaphorin 4D	Induces glioma invasiveness and angiogenesis	29133239	[60]

intercellular/intracellular signaling transduction from the TAM-secreted ligands to the tumor cell receptors and then to downstream TFs and target genes. The multilayer network correctly predicted some known important TAM-mediated pathways in gliomas that have been validated by experimental studies, such as the EGF pathway, TGF pathway and IGF1 pathway [13, 47]. Our approach also discovered some new ligand–receptor interactions in gliomas, such as THBS1-LRP5 and APOE-LRP6, which generated a new hypothesis of TAM-tumor cell interaction for further experimental testing.

The multilayer signaling network showed that EGFR and ERBB2 were two hub receptors connecting the extracellular TAM-secreted ligands and intracellular TFs. Moreover, we observed two feedback loops involved in the EGFR and ERBB2 signal axes, including EGFR-ESR2-*egfr*-EGFR and ERBB2-ELK1-*erbb2*-ERBB2. In addition, the TF sub-network indicated that correlations between ESR2 and ELK1 existed, suggesting a crosstalk between the EGFR-ESR2 pathway and the ERBB2-ELK1 pathway. The TAM-mediated EGFR feedback-crosstalk signaling module contains ligand genes (CAMP, ZP3, AREG, ANXA1, HBEGF, SPINK1, ICAM1, FGL1, HLA-A and SEMA4D), receptor genes (EGFR and ERBB2) and TFs (ESR2 and ELK1). Table 1 lists the experimental evidences for functional roles of the EGFR module genes in cancer progression [48–60]. Furthermore, the above 14 genes in the EGFR module exhibited significantly differential expression profiles between LGG (N = 172) and high-grade GBM patients (N = 530) in TCGA database (Figure 3A), indicating their potential correlation with glioma progression.

The prognostic significance and accuracy of an MNB

Based on the above multilayer signaling network, we proposed a novel MNB to predict the prognosis of cancer patients. We formulated the above TAM-mediated EGFR feedback-crosstalk signaling module as an MNB-based risk signature (see the Methods section). The identified MNB included 10 ligand genes (CAMP, ZP3, AREG, ANXA1, HBEGF, SPINK1, ICAM1, FGL1, HLA-A and SEMA4D), 2 receptor genes (EGFR and ERBB2) and 2 TFs (ESR2 and ELK1).

We first evaluated the prognostic significance of the MNB by using the TCGA data set (N = 598) and independent data set of glioma patients (N = 310) in CGGA database (Figure 4A and B). The statistical significance of the difference between the K-M survival curves for patients in the high-risk group (blue) and low-risk group (red) was assessed using the log-rank test, with P-values less than 0.0001 for both the TCGA and CGGA data sets. The high-risk group of patients had shorter overall survival time than the low-risk group. In addition, most of the 14 signature genes exhibited significantly differential expression profiles in high-risk group patients (N = 397) and low-risk group patients (N = 305) (Figure 3B).

We then evaluated the prognostic accuracy of the MNB by calculating the AUC of the time-dependent ROC with respect to the 1 year, 3 year and 5 year survival of glioma patients. The MNB showed good prognostic accuracy for both the TCGA data set (AUC at 1 year, 0.887; AUC at 3 years, 0.901; AUC at 5 years, 0.854; Figure 4C) and CGGA data set (AUC at 1 year, 0.774; AUC at 3 years, 0.867; AUC at 5 years, 0.931; Figure 4D). More evaluation indexes, including sensitivity, specificity, precision, accuracy, F1 score etc., of the predictive performance of the MNB in the validation set were listed in Supplementary Table S3.

We further compared the prognostic accuracy of the MNB with other prognostic biomarkers, including EGFR/ERBB2 genes and an immune-related gene signature (Cheng *et al.* [34]). The areas under the ROC curves of these biomarkers for predicting the overall survival of glioma patients in the TCGA and CGGA data sets were calculated and compared (Figure 4E and F). The results demonstrated a superior prognostic accuracy of the MNB in comparison with that of EGFR/ERBB2 genes and the immune-related gene signature (see also the Discussion section).

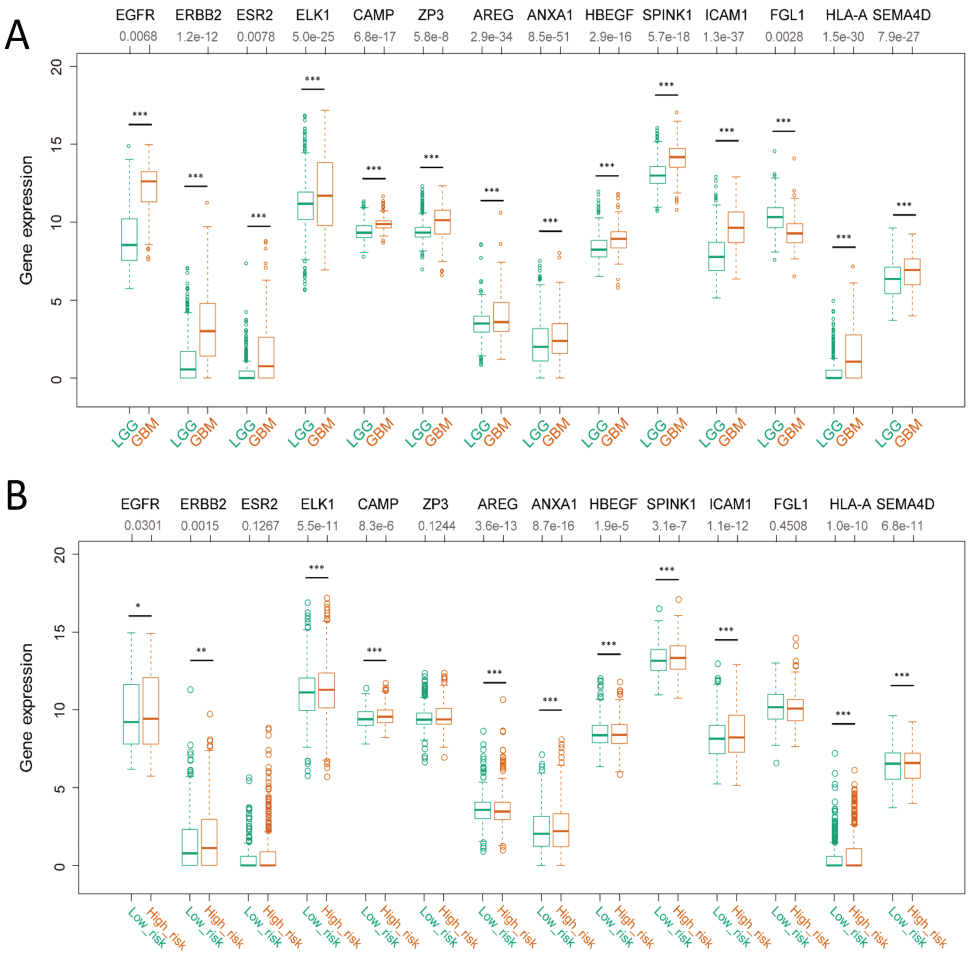

Figure 3. Expression profiles of the 14 signature genes involved in the MNB in different samples of glioma patients. (**A**) The distribution of 14 signature genes in GMB (green) samples and LGG samples (brown), respectively. (**B**) The distribution of 14 signature genes in low-risk group (green) and high-risk group (brown) samples of glioma patients, respectively. Mann–Whitney–Wilcoxon test was used to assess the significance of the difference of each gene's expression between two groups (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$). Gene names and the corresponding *P*-values were listed.

The MNB is associated with long-term response of targeted therapies in gliomas

We further selected the patients who received molecularly targeted therapy to examine the association between the above MNB-based risk signature and the drug sensitivity or resistance of glioma patients. The 3 year or 5 year survival status (alive or dead) of each patient following the molecular targeted therapy was used to evaluate the long-term response of the drug treatment (sensitive or resistant). Based on the above MNB-based risk signature, each patient was classified into sensitive group (i.e. low-risk group) or resistant group (i.e. high-risk group) according to the optimal cut-off value of the RS using ROC method. To assess the accuracy of the MNB for predicting the response of the targeted therapies, we compared the AUC of the ROC of the MNB with other gene biomarkers including EGFR/ERBB2 genes and an immune-related gene signature (Cheng *et al.* [34]). Figure 5A shows the AUC of the ROC of these signatures for predicting targeted-therapeutic response evaluated by the 3 year survival (AUC of MNB, 0.897; AUC of EGFR/ERBB2, 0.712; AUC of Cheng *et al.* signature, 0.823). Figure 5B shows the AUC of the ROC of these signatures for predicting targeted-therapeutic response evaluated by the 3 year survival (AUC of MNB, 0.854; AUC of EGFR/ERBB2, 0.662; AUC of Cheng *et al.* signature, 0.779). These results demonstrated a significantly higher accuracy of the MNB-based signature compared with other signatures for predicting the long-term response of the molecularly targeted therapies in glioma patients.

The MNB is an independent and robust prognostic signature

We conducted univariate and multivariate Cox regression analyses and found that the MNB was independently correlated with the overall and 5 year survival of glioma patients in both TCGA data set (Table 2) and CGGA data set (Supplementary Figure S3) with adjustment for clinicopathologic factors (e.g. age, gender and grade), IDH mutation status and other existing gene signatures.

To explore the prognostic value of the MNB in stratified cohorts, patients were first classified by two important clinicopathologic factors, age and grade, that significantly correlated with the prognosis of glioma patients (Table 2; Supplementary Figure S3). Figure 6A–D shows the prognostic significance of the MNB in different glioma cohorts stratified by age (Figure 6A and B) or grade (Figure 6C and D). In all of these cohorts, patients were classified as high-risk versus low-risk groups using cut point from ROCs, and the high-

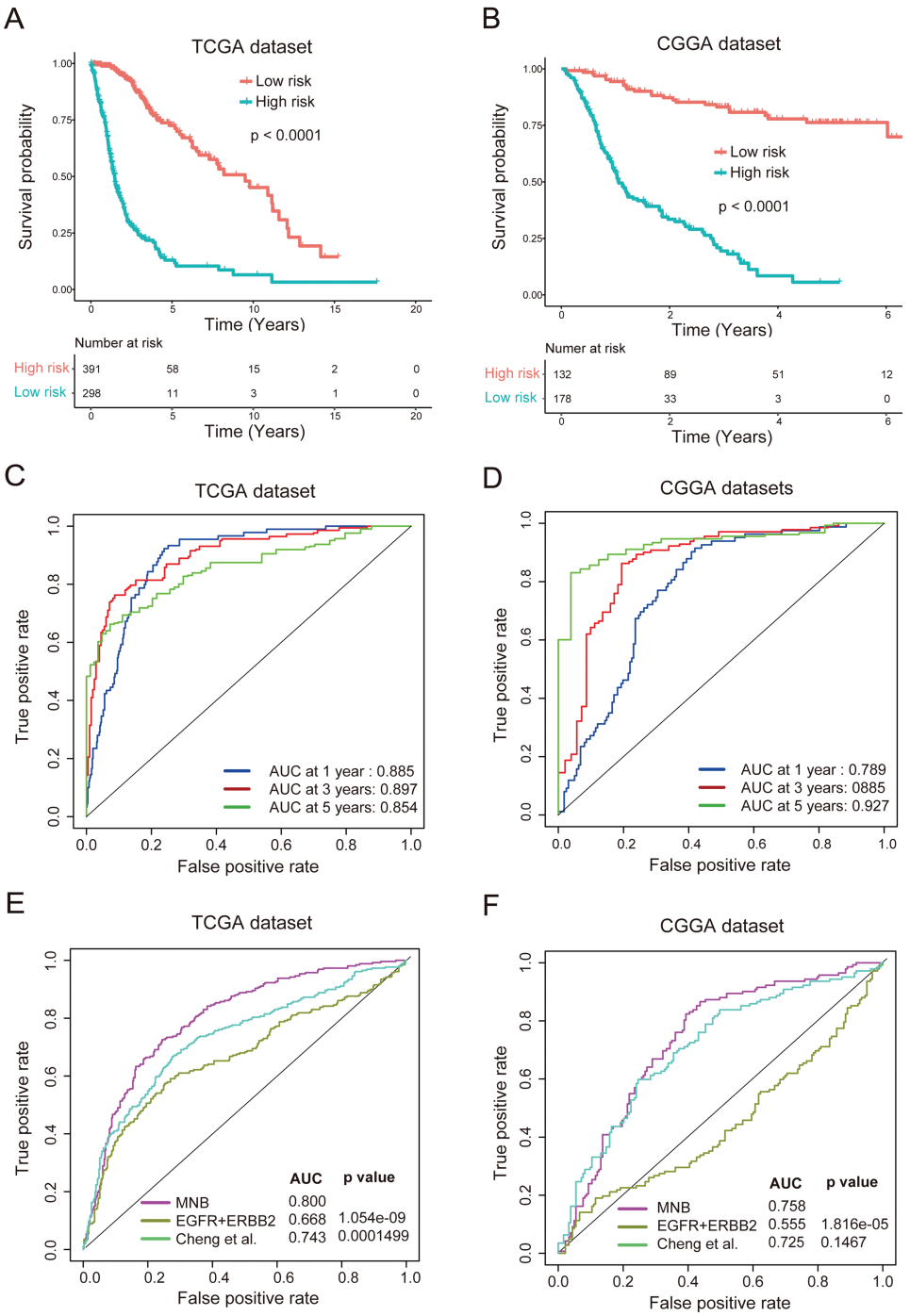

Figure 4. The prognostic significance and accuracy of the MNB. (**A** and **B**) Prognostic significance of the MNB assessed by the TCGA data set ($N = 598$) (A) and CGGA data set ($N = 310$) (B) of glioma patients. An RS was formulated from the Cox PH model based on the gene expression levels of the MNB. The statistical significance of the difference between K-M survival curves for patients in the high-risk group (blue) and in the low-risk group (red) was assessed using the log-rank test. (**C** and **D**) Prognostic accuracy of the MNB evaluated by the AUC of the time-dependent ROC with respect to 1 year, 3 year and 5 year survival of glioma patients in the TCGA data set (C) and CGGA data set (D). The MNB showed good prognostic accuracy. (**E** and **F**) Comparison of ROC curves of the MNB with that of EGFR and ERBB2 genes and an immune-related gene signature (Cheng *et al.* [34]) for the TCGA data set (E) and CGGA data set (F). The statistical significance of the difference between areas under the two correlated ROC curves was assessed using DeLong's test [35]. The MNB showed superior predictive power.

risk patients had a significantly shorter overall survival than low-risk patients. Subsequently, patients who received chemotherapy and radiotherapy were utilized to validate the prognostic significance of the MNB. Figure 6E–H demonstrates that the MNB retained prognostic significance for glioma patients treated with or without chemotherapy (Figure 6E and F) and radiotherapy (Figure 6G and H). These results indicated that the MNB could accurately identify patients with an unfavorable prognosis regardless of their clinicopathologic and treatment characteristics. The analysis in the stratified cohorts of CGGA data set showed the consistent results (Supplementary Figure S4).

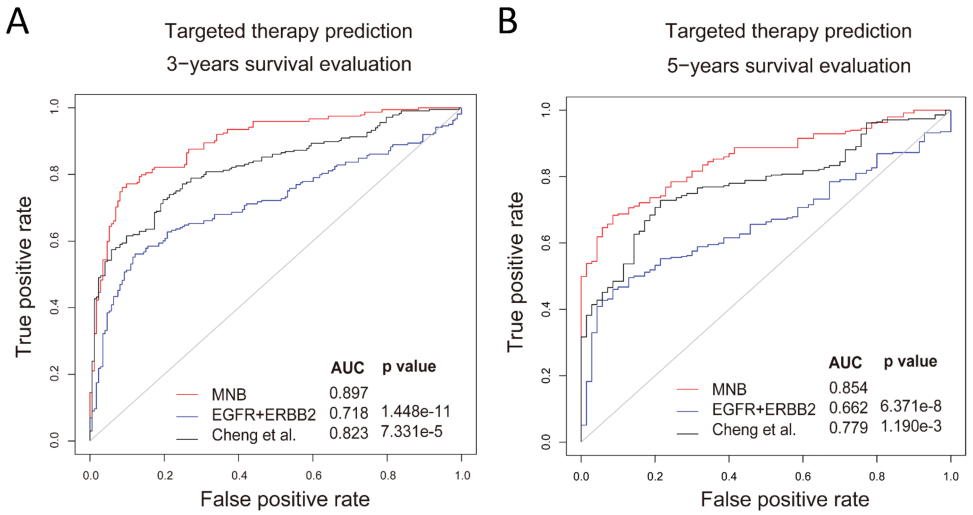

Figure 5. The accuracy of the MNB in predicting long-term response of the targeted therapy in glioma patients in comparison with other signatures [EGFR/ERBB2 genes and an immune-related gene signature (Cheng *et al.* [34])]. The predictive accuracies of these signatures evaluated by the 3 year survival (**A**) or 5 year survival (**B**) were assessed by the AUC under the ROC curves. The statistical significance was assessed using DeLong's test [35]. The MNB showed more accuracy.

Table 2. Multivariate Cox regression analysis of clinicopathologic factors (age, gender and grade) and three gene signatures for predicting overall survival and 5 year survival in TCGA glioma patients. The MNB signature showed prognostic significance for both the overall survival and 5 year survival (*P*-values are denoted in red) of glioma patients, indicating that the MNB signature is an independent risk factor for glioma patients. See also Supplementary Figure S3 for multivariate Cox regression analysis using CGGA data set

Variable	Overall survival				5 year survival			
	Univariate COX		Multivariate COX		Univariate COX		Multivariate COX	
	P-value	HR (95% CI)	*P*-value	HR (95% CI)	*P*-value	HR (95% CI)	*P*-value	HR (95% CI)
Age								
≥40 versus <40	<2e-16	4.342 (3.193–5.904)	1.50e-06	2.406 (1.682–3.440)	<2e-16	5.159 (3.62–7.352)	1.13e-05	2.440 (1.638–3.634)
Gender								
Male versus female	0.111	1.224 (0.954–1.571)	0.343173	1.133 (0.876–1.465)	0.0699	1.279 (0.980–1.669)	0.36602	1.136 (0.862–1.496)
Grade								
GBM versus LGG	<2e-16	9.499 (7.212–12.51)	0.000381	2.080 (1.389–3.116)	<2e-16	9.499 (7.212–12.51)	0.00114	1.975 (1.311–2.975)
MNB								
High-risk versus low-risk	**<2e-16**	2.718 (2.420–3.054)	**6.07e-10**	1.893 (1.547–2.318)	**<2e-16**	2.832 (2.511–3.195)	**1.88e-10**	2.011 (1.622–2.493)
EGFR + ERBB2								
High-risk versus low-risk	<2e-16	2.718 (2.264–3.263)	0.995660	0.999 (0.800–1.248)	<2e-16	2.82 (2.335–3.405)	0.60139	0.940 (0.745–1.186)
Immune-related signature								
High-risk versus low-risk	<2e-16	1.559 (1.469–1.655)	0.424095	1.042 (0.942–1.154)	<2e-16	1.581 (1.488–1.679)	0.52072	1.035 (0.932–1.149)

We also examined whether combining the MNB with clinicopathologic risk factors and other existing gene signatures could significantly improve the prognostic accuracy of the risk signature. The time-dependent ROC curves (Figure 7) compared the prognostic accuracy by age, grade (LGG or GBM), MNB, an immune-related gene signature (i.e. Cheng *et al.* signature) and the combined signature (see the definition in the Methods section). The AUCs of ROC curves in both the TCGA data set (Figure 7A and B) and the CGGA data set (Figure 7C and D) showed that the combined signature greatly outperformed other risk signatures except the MNB for predicting the 3 and 5 year survival rates of glioma patients. Noticeably, the AUC of the ROC of the MNB was rather close to that of the combined signature. These results indicated that the MNB possessed convincingly strong prognostic power and was almost as accurate as the combination of all the considered clinicopathologic factors and existing risk signatures.

We tested the robustness of the MNB in comparison with LASSO Cox PH model using a bootstrapping approach (Figure 8). We generated 100 random data sets by randomly taking 60% of the samples from the CGGA data sets (i.e. validation set). The AUC values of ROC with respect to overall survival (Figure 8A), 3 year survival (Figure 8B) and 5 year survival (Figure 8C) were computed. Kolmogorov–Smirnov test (K-S test) *P*-values were computed to assess the significance of the difference between the probability distributions of AUC values of the MNB and LASSO. Although LASSO was fitted well with the TCGA data set (i.e. training set; Supplementary Figure S5), its predictive accuracy on the validation set was significantly lower than the MNB.

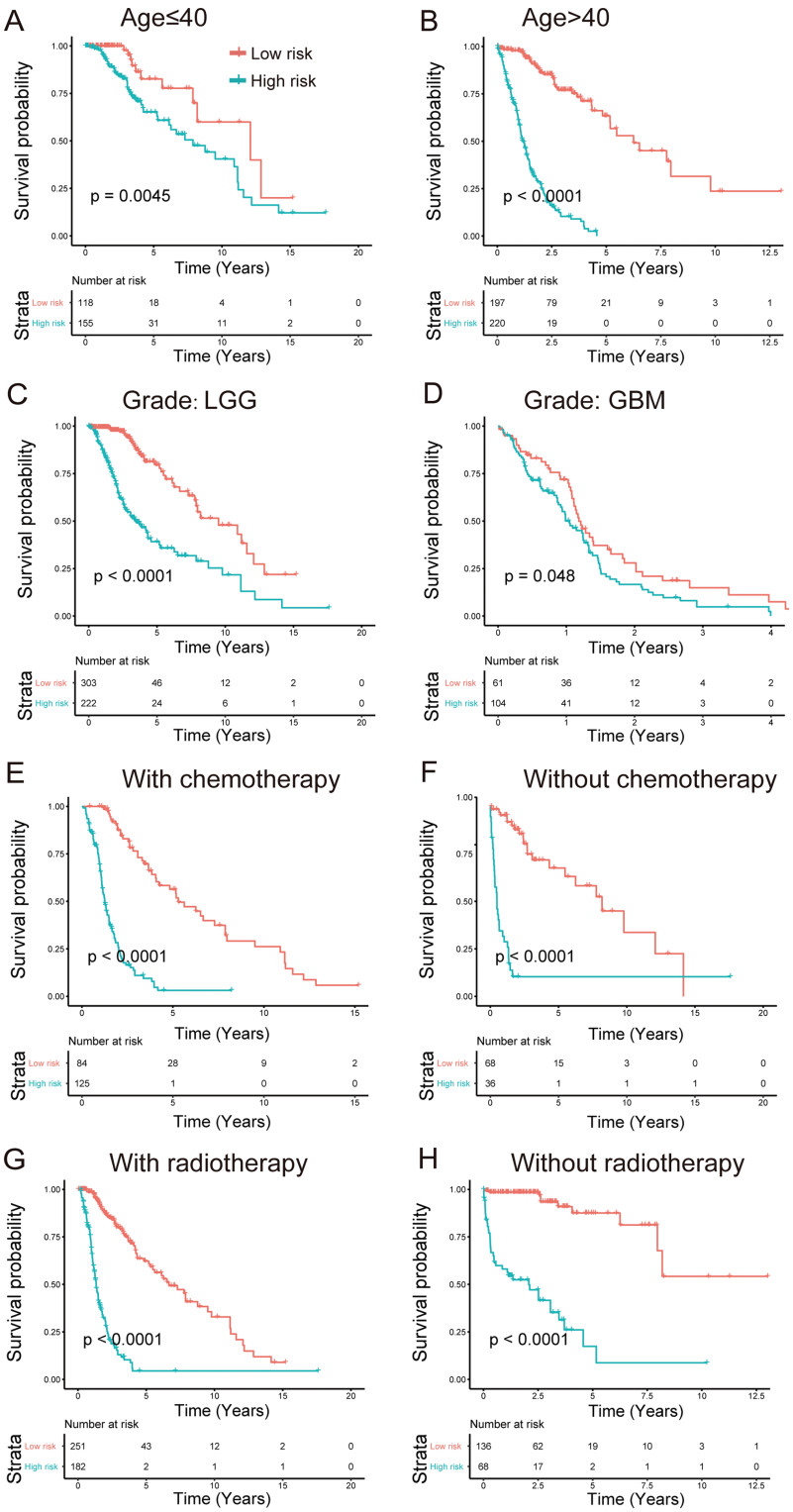

Figure 6. Prognostic significance of the MNB in the stratified cohorts of TCGA data set. (**A–D**) Prognostic significance of the TAM-related gene signature in different cohorts stratified by age (age ≤ 40 or age > 40, panels A and B, respectively) or grade (LGG and GBM, panels C and D, respectively). (**E–H**) The MNB signature retained prognostic significance for glioma patients treated with or without pharmaceutical therapy (E and F) and radiotherapy (G and H). Optimal cut-off values were used to determine high-risk and low-risk groups in each stratified cohort, and the statistical significance of the difference between two K–M survival curves was assessed using the log-rank test. See also Supplementary Figure S4 for the analysis results in the stratified cohorts of CGGA data set.

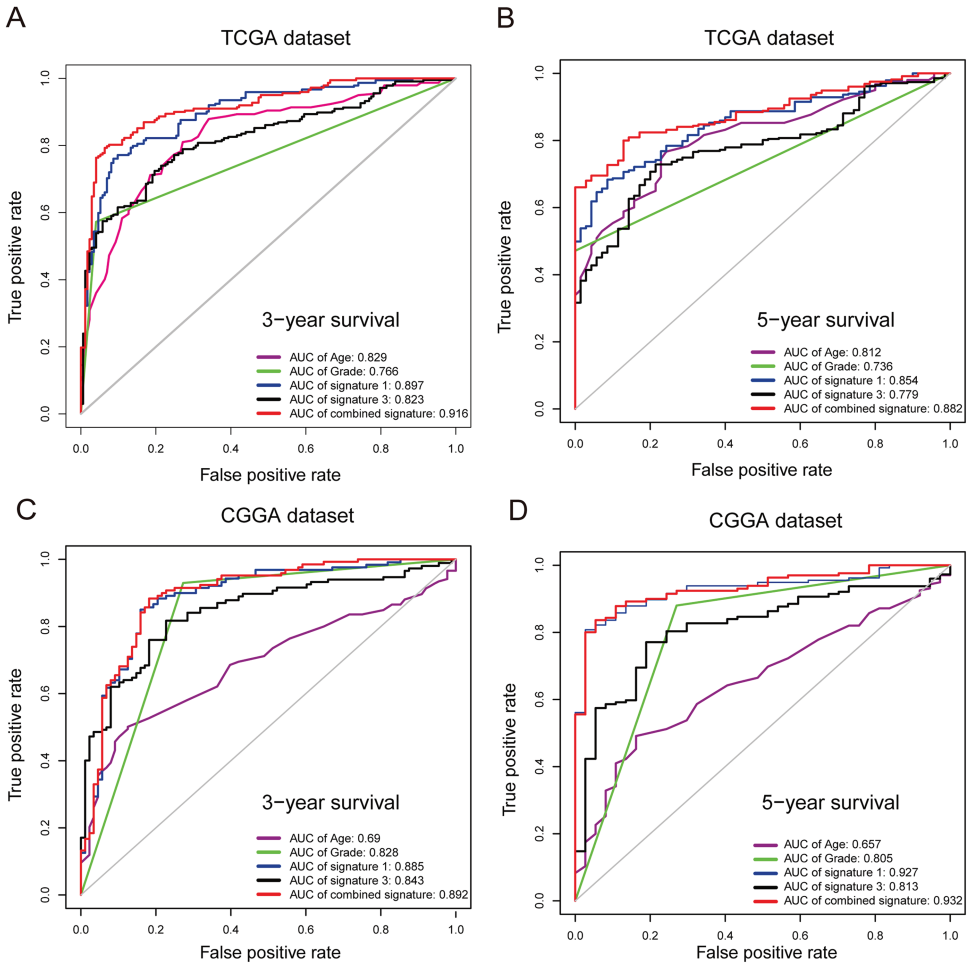

Figure 7. Time-dependent ROC curves comparing the prognostic accuracy of the MNB with clinicopathologic risk factors and other existing gene signatures or their combination. Only significant signatures in the multivariate Cox regression analysis, i.e. MNB and immune-related gene signature (Table 2; Supplementary Figure S3), were used for analysis and comparison. (**A** and **B**) Comparisons of the prognostic accuracy by age, grade (LGG or GBM), MNB, an immune-related gene signature (i.e. Cheng *et al*. signature) and the combined signature using the TCGA data set with respect to 3 year survival (A) or 5 year survival (B). (**C** and **D**) Comparisons of the prognostic accuracies with respect to 3 year survival (C) or 5 year survival (D) using the CGGA data set.

Therefore, the MNB was verified to be more accurate and robust than LASSO.

We further validated the prognostic value of the MNB on a large external data set ($N = 899$) involving both gene expression values and clinical information of glioma patients from GSE55918 [61] that was well aggregated from 16 previous microarray data sets and was used for external validation. The MNB had good prognostic significance (Figure 9A) and accuracy (Figure 9B) on this data set. Furthermore, we performed robustness tests of the MNB on 100 random sub-data sets using a bootstrapping approach as described above. The mean AUC values with 95% confidence intervals for 3 year and 5 year survival predictions were 0.755 (0.745–0.764) (Figure 9C) and 0.747 (0.747–0.756) (Figure 9D), respectively. These results demonstrated that the MNB is an independent and robust prognostic signature.

Mechanistic simulation illuminated dynamic interplay between multilayer signals in the MNB

Next, we sought to investigate the intrinsic interplay between intercellular and intracellular pathways in the TAM-mediated EGFR module to understand the role of the MNB in regulating cellular function. The TAM-mediated EGFR feedback-crosstalk signaling module is depicted in Figure 10A, where L1 and L2 represent TAMs-secreted ligands for EGFR and ERBB2, respectively, and α and β represent the crosstalk strength between ESR2 and ELK1, respectively. Generally, in mammalian cells, biochemical reactions in signaling transduction (including ligand–receptor binding, protein phosphorylation and TF activation, at a time scale of seconds) is relatively faster than the stochastic switching of TF-regulated genes (at a time scale of minutes) [62] (Figure 10B). Therefore, a multiple temporal scale model is required to describe the multilayer signaling dynamics. As such, we developed a multiscale model that couples ODEs for signaling transduction at short time scales and stochastic processes for TF-gene interactions at long time scales, simulated using a Gillespie-ODE hybrid algorithm (see the Methods section).

We investigated the combinatorial effects of extracellular ligand concentrations and intracellular crosstalk strength on the signaling dynamics. Figure 10C shows the mean values of EGFR activation at steady state (evaluated in 100 min) from 1000 simulations, in which a dramatic switch was observed as α increased above -0.12 when L1 was greater than 0.5. Meanwhile, the noise of EGFR activation increased singularly near

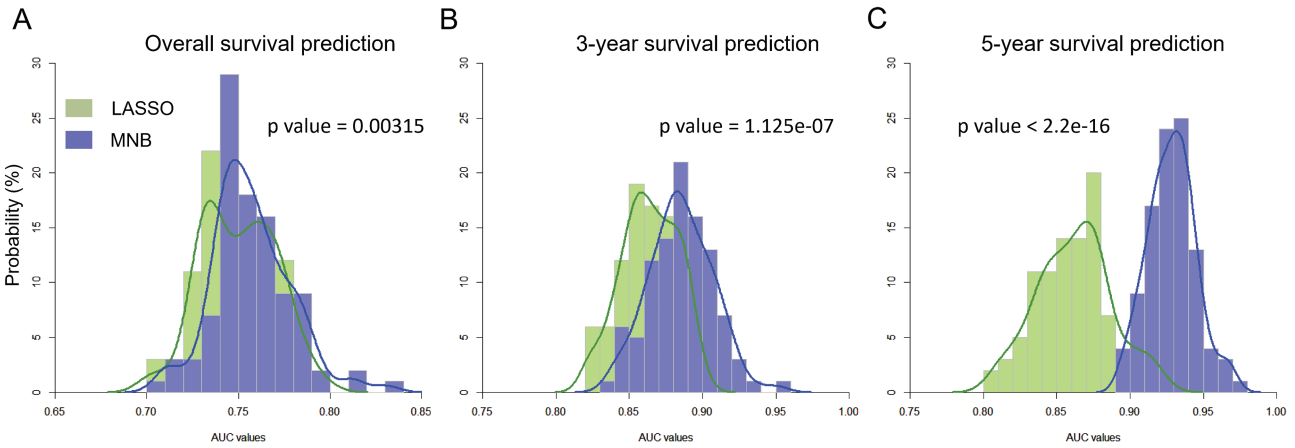

Figure 8. Robustness tests of the MNB compared with LASSO. We generated 100 random data sets by randomly taking 60% of the samples from the validation set. The AUC values of ROC with respect to overall survival (**A**), 3 year survival (**B**) and 5 year survival (**C**) were computed. K-S test P-values were computed to assess the significance of the difference between the probability distributions of AUC values of MNB and LASSO. MNB showed good robustness and better accuracy than LASSO.

$\alpha = -0.12$ (Figure 10D). These results suggest that the increase in crosstalk strength α might induce a critical transition of the signaling dynamics and cellular states. Around the critical value of $\alpha = -1.2$, there were two distinct populations of EGFR activation trajectories (Figure 10E), resulting in a bimodal steady-state distribution of EGFR activation. The stochastic dynamics of ERBB2 activation resulted in, whereas, a unimodal steady-state distribution (Figure 10F). The difference between EGFR and ERBB2 distributions verified the role of multilayered feedback-crosstalk loops in decoding ligand specificity of stochastic cell signaling [63, 64].

In addition, Supplementary Figure S6 shows the stochastic dynamics of ON–OFF switching of genes *egfr* and *erbb2* (Supplementary Figure S6A and B), the total levels of EGFR and ERBB2 proteins (Supplementary Figure S6C and D) and the activities of TFs ESR and ELK1 (Supplementary Figure S6G–H). The red lines represent means, and the brown lines represent the first and third quartiles. In the above simulations, $\alpha = -1.2$, $\beta = 1$ and $L1 = L2 = 1$. Other parameters are listed in Supplementary Table S4. The examination of the effects of β on the stochastic activation of EGFR (Supplementary Figure S7) indicated that the above results were robust with respect to the inverse crosstalk from ESR2 to ELK1. These results demonstrated that both the extracellular signals and the intracellular circuits influence the dynamics of signaling transduction and gene expression. Moreover, the model analysis revealed dynamic and nonlinear interplay between multilayer signals, highlighting the importance of the biological integration of multilayered signals in decoding ligand specificity, determining cellular fate and driving heterogeneous response kinetics.

Discussion and Conclusion

The network-based approaches to modeling cancer genomic or transcriptomic data have been developed in the past few years to understand cancer mechanisms and to predict cancer metastasis, prognosis, drug resistance and drug targets for clinical purposes [66–68]. For example, Edwin Wang's group proposed and developed cancer hallmark networks approach for individual tumors [65]. They developed a cancer hallmark gene module-based network algorithm to identify highly robust cancer biomarkers [69]. The core cancer hallmark network could also be used to effectively predict breast cancer subtype-specific drug targets [70]. These studies provided encouraging examples of using network-based prediction approaches in cancers. The success of such methods could be attributed to the inherent capability of biomolecular networks in representing underlying mechanisms and principles of tumorigenesis, cancer development and drug response.

The systematic reconstruction and modeling of intercellular signaling pathways and intracellular gene regulatory networks are important for understanding microenvironment-mediated gene expression and cell–cell interactions. The technology of scRNA-seq provides a new tool with which to accomplish this challenging task. In this study, we proposed an scRNA-seq data-based pipeline to reconstructing and modeling the multilayer signaling networks that consist of pathways, from intercellular ligand–receptor interactions to intracellular TF activation and gene switching. To illustrate our method, we used scRNA-seq transcriptomic data from glioma patients. We built a multilayer intercellular/intracellular signaling network to investigate tumor cell–microenvironment interactions. The reconstructed network not only correctly predicted some experimentally validated TAM-mediated pathways in gliomas but also discovered some new pathways that exist between cancer cells and TAMs.

We acknowledge that some assumptions and simplifications have been made for the scRNA-seq transcriptome-based reconstruction of the multilayer signaling network. Both intercellular and intracellular signaling transduction pathways involve complex posttranslational modifications of proteins, such as phosphorylation, ubiquitination and acetylation [71]. However, due to difficulties in obtaining high-throughput single-cell proteomic and secretome proteomic data, the use of scRNA-seq to approximate protein activity in cells is a good alternative given that highly expressed mRNA levels correlate well with increased protein activity because the ratio of protein to mRNA level is conserved between tissues for every protein [72]. In this way, the false-positive prediction rate would be reduced, although a variety of elements in the natural signaling network might be missed. In addition, we did not consider direct physical contact between cells in our network reconstruction, which is also important for the interaction of physically adjacent cells. In the future, we will integrate this type of cell interaction mode [21] into the multilayer signaling network to build a more comprehensive interaction map between cells.

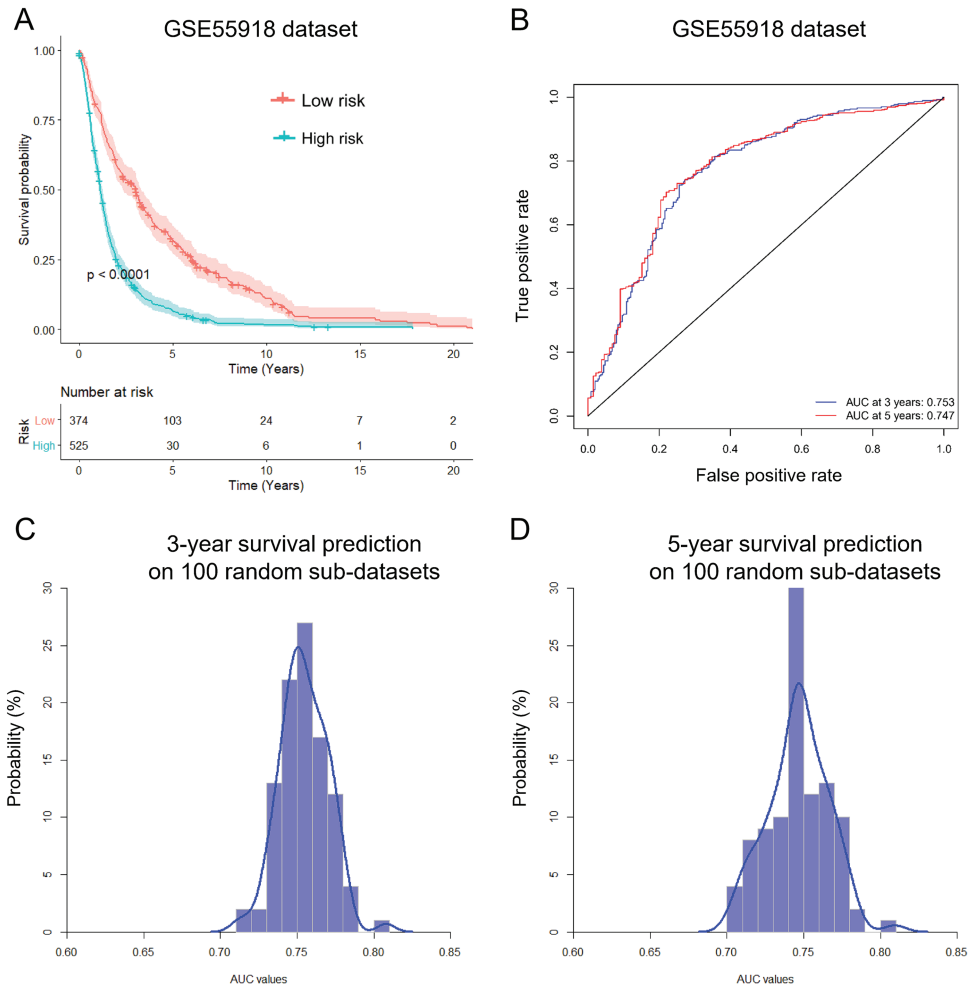

Figure 9. Validation of the prognostic value of the MNB on more data sets. An integrated data set ($N = 899$) involving both gene expression values and clinical information of glioma patients from GSE55918 that was merged from 16 previous microarray data sets was used for external validation. (**A**) K-M survival curves for patients in the high-risk group (blue) and in the low-risk group (red) predicted by the MNB signature, with log-rank test P-value less than 0.0001. (**B**) Prognostic accuracy of the MNB evaluated by the AUC of the time-dependent ROC with respect to 3 year and 5 year survival of glioma patients in the whole external validation set. The MNB showed good prognostic accuracy. (**C** and **D**) Robustness tests of the MNB using a bootstrapping approach by generating 100 random sub-data sets each randomly taking 60% of the samples from the external validation set. The mean AUC values with 95% confidence intervals for 3 year survival prediction (C) and 5 year survival prediction (D) were 0.755 (0.745–0.764) and 0.747 (0.737–0.756), respectively.

Despite these flaws in our approach, the reconstructed network verified some known pathways, such as the EGF pathway, TGF pathway and IGF1 pathway [13, 47], and discovered some novel crosstalk and feedback loops, as motioned above (Figure 2G). The proposed method enriched the downstream analysis of scRNA-seq data and represents a new approach to study cell–cell interactions that could be applied to many biological questions, such as cancer drug resistance [67]. For example, many studies have revealed that the interaction between immune cells and tumor cells influences tumor progression and drug response [43–45]. The IGF1-IGF1R pathway between TAMs and glioma cells, as revealed in this study, has been found to play important roles in glioma resistance to CSF1R inhibition therapy [8]. Our previous study [46] developed a spatio-temporal model to study the TAM-mediated drug resistance in glioma immunotherapy based on the IGF1-IGF1R pathway. In future work, we will use our developed scRNA-seq transcriptome-based multilayered signaling network approach to study intercellular and intracellular mechanisms that underlie microenvironment-mediated drug resistance.

More importantly, based on the multilayer signaling mechanism, we identified a novel multilayer network-based biomarker, namely MNB, for predicting cancer prognosis, which was proven to be of prognostic and predictive value in the clinical data sets of glioma patients. Although EGFR plays important roles in glioma development [73], the studies on prognostic significance of EGFR expression in glioma have not been consistent [74]. Some studies have shown that EGFR expression is not associated with the overall survival of secondary GBM patients [75], and it was not an independent predictor of overall survival in a cohort of glioblastoma patients [76].

We demonstrated that the EGFR-mediated MNB possessed better prognostic accuracy and robustness when compared to the conventional gene signatures (Figures 4, 5 and 8; Supplementary Figure S8) and other predictive methods such as LASSO Cox PH model (Figure 8) and random forests method for survival prediction (Supplementary Figure S9). We interpret this as follows. First, the interaction among the EGFR feedback-crosstalk module is dynamic and nonlinear as revealed by our hybrid multiscale model (Figure 10). Therefore, a combination of genes involved in the multilayered EGFR module could

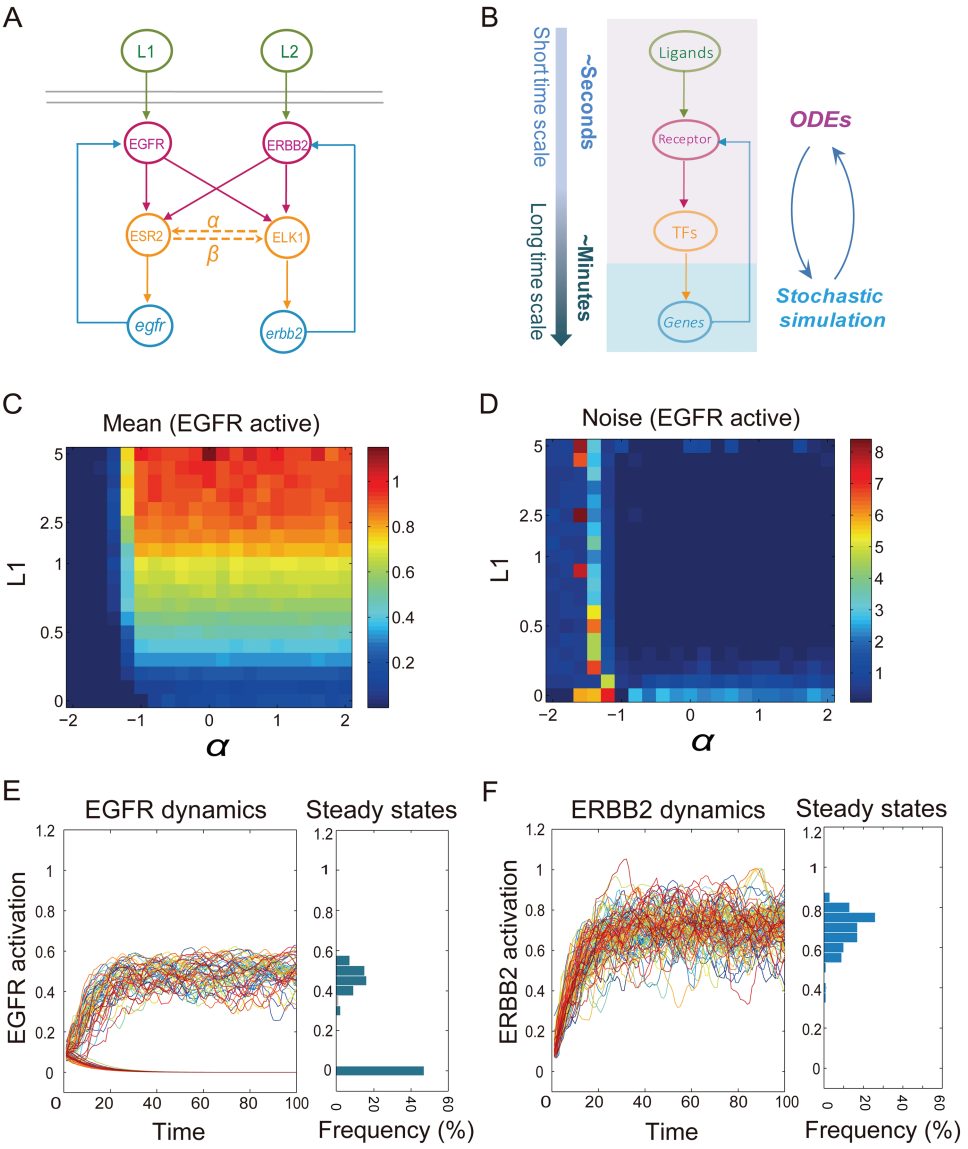

Figure 10. Multiscale modeling of the TAM-mediated EGFR feedback-crosstalk module depicts dynamic interplay between multilayer signals in the MNB. **(A)** TAM-mediated EGFR feedback-crosstalk module. Under the stimulation of the respective ligands secreted from TAMs, receptors EGFR and ERBB2 can both activate TFs ESR2 and ELK1 through crosstalk pathways, which regulate the gene ON–OFF switches of *egfr* and *erbb2* and, in turn, the corresponding protein synthesis. L1 and L2 represent the TAM-secreted ligands for EGFR and ERBB2, respectively. α and β represent crosstalk strengths between ESR2 and ELK1. **(B)** Schematic depiction of multiscale modeling. The signaling transduction from extracellular ligands to intracellular receptors and TFs activation at short time scales was modeled using ODEs, while the gene transcription at long time scales was described using stochastic processes. A hybrid simulation algorithm was designed to simulate the multiscale model. **(C)** Mean of EGFR activation at steady states under different combinations of L1 and α values. A dramatic switch was observed as α increased above −0.12. **(D)** Noise of EGFR activation at steady states under different combinations of L1 and α values. The noise increased singularly near $\alpha = -0.12$. **(E** and **F)** Stochastic dynamics of activated EGFR and ERBB2. A bimodal distribution of EGFR activation and a unimodal distribution of ERBB2 activation at steady states were observed. $\alpha = -1.2$, $\beta = 1$, $L1 = L2 = 1$.

represent the signaling regulation and cellular status more accurately and comprehensively [77], thus having more predictive power than single genes. Second, and more importantly, the MNB takes into account the mechanism of cell–cell interactions, in particular the tumor cell–TAMs interaction pathways, for predicting cancer progression. Compared to the conventional biomarkers that focus on tumor cells, the MNB could mechanistically connect microenvironmental cells (e.g. immune cells) to tumor cells through intercellular pathways (ligand–receptor interactions) and further to intracellular gene transcription (TF-gene interactions). Therefore, the MNB provides a mechanism-based approach to identify functional signaling modules that has predictive power for cancer prognosis.

Although the existing immune-related gene signature [34] was generated by screening all immune cell-related genes, it lacks clear correspondence to the specific cell types, thus falls short on mechanistic interpretation. Our MNB signature only considered TAMs while produced better predictive ability to the immune-related gene signature (Supplementary Figure S8), highlighting the dominant roles of TAMs in glioma progression and drug resistance. In the future work, we will construct multiple cell type-mediated multilayer networks, which would expectedly identify a more accurate prognostic signature.

The addition of the interaction terms between the genes in the MNB could be expected to improve the predictive power of the prognostic signature, however, which would also result in more complexity of the model and unfairness of comparison with other existing signatures that was based on linear models of gene expression features. So in the current study we did not explicitly consider the interaction terms in the model. In the future work we will integrate the interaction terms in the model to improve our approach.

In recent years, efforts have been made to develop more accurate proportional hazards models using advanced computational methods, such as random survival forests (RSF) [39], neural network Cox models (Cox-nnet) [78] and deep learning models (DeepSurv) [79]. These models usually produce a large amount of genes as features for survival prediction. For example, we compared the MNB and RSF-based signature (i.e. rfsrc signature; Supplementary Figure S9). Even though rfsrc signature employed a large amount of genes (up to 259 genes), its prognostic performance was not significantly better than that of the MNB. In particular, a robustness test using a bootstrapping approach showed that the MNB was much more accurate than the rfsrc signature with respect to 5 year survival prediction. Therefore, the MNB did not only possess good predictive ability but also had less model complexity and better mechanism interpretation.

In this study, we used scRNA-seq of IDH-mutant astrocytoma samples to construct the multilayer network. We assumed that the expression levels and the regulation strengths of genes in the multilayer network might be quantitatively different between different glioma subtypes (e.g. astrocytoma, oligodendrocytoma and GBM), but the structures of the multilayer network between these two types were qualitatively comparable except slight disparity due to different mutation status of some genes (e.g. IDH). So the multilayer network constructed from the astrocytoma samples should be comparable to that of GBM. Ideally, if the GBM samples are available and adopted also for the construction of the multilayer network, the resulting MNB signature could be expected to further improve the prognostic accuracy of glioma patients especially GBM patients.

The previous study [19] has demonstrated that the analysis of differentially expressed genes is largely independent of different scRNA-seq platforms (e.g. the SMART-seq2 and the 10X genomics platform; Supplementary Figure S5C therein). On the other hand, although different scRNA-seq platforms may detect gene expression with different levels of amplification noise [80], the Seurat based on DM employed in our pipeline could filter gene expression noise. Therefore the analysis of the differential expression of genes between different cellular clusters in our pipeline should be relatively robust with respect to different scRNA-seq platforms. Since the construction of the network was mainly based on the differentially expressed genes, so we assumed that different scRNA-seq platforms have slight influence on the construction of the network and thus the identification of biomarker.

In summary, we proposed a novel scRNA-seq transcriptome-based multilayer network approach to identify prognostic and predictive signatures of cancer patients. The MNB was shown to be of good prognostic value in glioma patients compared with other existing signatures and computational methods. The proposed approach shed lights on facilitating the identification of signatures for the clinical diagnosis, prognosis or treatment of cancers.

Key Points

- The technology of scRNA-seq provides an unprecedented opportunity for dissecting the interplay between the cancer cells and the associated microenvironment.
- We developed a novel scRNA-seq data-based approach to reconstruct a multilayer signaling network that contains pathways from intercellular ligand–receptor interactions, intracellular TFs and their target genes.
- A novel MNB was proposed to predict survival outcome and therapeutic response of glioma patients.
- The MNB showed good prognostic and predictive power and outperformed other related gene signatures, evaluated with glioma data sets.
- The MNB connects cancer cells and tumor-associated microenvironment, which would facilitate the identification of more effective signatures for the clinical prognosis and treatment of cancers.

Supplementary Data

Supplementary data are available online at https://academic.oup.com/bib.

Funding

National Natural Science Foundation of China (11871070 and 61503419 to X.S.); Guangdong Nature Science Foundation (2016A030313234 to X.S.); Guangdong Provincial Key Laboratory of Computational Science (2018003 to X.S.); National Key Research Project of China (91530320 to T.Z.); 973 Project of China (2014CB964703 to T.Z.). The funding body has no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

References

1. Pires-daSilva A, Sommer RJ. The evolution of signalling pathways in animal development. *Nat Rev Genet* 2003;**4**:39–49.
2. Boulanger CA, Mack DL, Booth BW, Smith GH. Interaction with the mammary microenvironment redirects spermatogenic cell fate in vivo. *Proc Natl Acad Sci U S A* 2007;**104**:3871–6.
3. Livesey FJ, Cepko CL. Vertebrate neural cell-fate determination: lessons from the retina. *Nat Rev Neurosci* 2001;**2**:109–18.
4. Ammeux N, Housden BE, Georgiadis A, *et al*. Mapping signaling pathway cross-talk in *Drosophila* cells. *Proc Natl Acad Sci U S A* 2016;**113**:9940.
5. Heasley LE. Autocrine and paracrine signaling through neuropeptide receptors in human cancer. *Oncogene* 2001;**20**:1563.
6. Smalley KSM, Brafford PA, Herlyn M. Selective evolutionary pressure from the tissue microenvironment drives tumor progression. *Semin Cancer Biol* 2005;**15**:451–9.
7. Hotamisligil GS. Inflammation and metabolic disorders. *Nature* 2006;**444**:860–7.
8. Quail DF, Bowman RL, Akkari L, *et al*. The tumor microenvironment underlies acquired resistance to CSF-1R inhibition in gliomas. *Science* 2016;**352**:aad3018.

9. Obenauf AC, Zou Y, Ji AL, *et al*. Therapy-induced tumour secretomes promote resistance and tumour progression. *Nature* 2015;**520**:368.
10. Straussman R, Morikawa T, Shee K, *et al*. Tumour microenvironment elicits innate resistance to RAF inhibitors through HGF secretion. *Nature* 2012;**487**:500–4.
11. Klemm F, Joyce JA. Microenvironmental regulation of therapeutic response in cancer. *Trends Cell Biol* 2015;**25**:198–213.
12. Somasundaram R, Zhang G, Fukunaga-Kalabis M, *et al*. Tumor-associated B-cells induce tumor heterogeneity and therapy resistance. *Nat Commun* 2017;**8**:607.
13. Quail DF, Joyce JA. The microenvironmental landscape of brain tumors. *Cancer Cell* 2017;**31**:326.
14. Kerkar SP, Restifo NP. Cellular constituents of immune escape within the tumor microenvironment. *Cancer Res* 2012;**72**:3125.
15. Hanahan D, Coussens LM. Accessories to the crime: functions of cells recruited to the tumor microenvironment. *Cancer Cell* 2012;**21**:309.
16. Sun X, Su J, Bao J, *et al*. Cytokine combination therapy prediction for bone remodeling in tissue engineering based on the intracellular signaling pathway. *Biomaterials* 2012;**33**:8265–76.
17. Patel AP, Tirosh I, Trombetta JJ, *et al*. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* 2014;**344**:1396.
18. Tirosh I, Venteicher AS, Hebert C, *et al*. Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. *Nature* 2016;**539**:309–13.
19. Venteicher AS, Tirosh I, Hebert C, *et al*. Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq. *Science* 2017;**355**:1391–402.
20. Lee JK, Wang J, Sa JK, *et al*. Spatiotemporal genomic architecture informs precision oncology in glioblastoma. *Nat Genet* 2017;**49**:594–9.
21. Boisset J-C, Vivié J, Grün D, *et al*. Mapping the physical network of cellular interactions. *Nat Methods* 2018;**15**:547–53.
22. Zhou JX, Taramelli R, Pedrini E, *et al*. Extracting intercellular signaling network of cancer tissues using ligand–receptor expression patterns from whole-tumor and single-cell transcriptomes. *Sci Rep* 2017;**7**:8815.
23. Ramilowski JA, Goldberg T, Harshbarger J, *et al*. A draft network of ligand–receptor-mediated multicellular signalling in human. *Nat Commun* 2015;**6**:7866.
24. Skelly DA, Squiers GT, McLellan MA, *et al*. Single-cell transcriptional profiling reveals cellular diversity and intercommunication in the mouse heart. *Cell Rep* 2018;**22**:600.
25. Coifman RR, Lafon S, Lee AB, *et al*. Geometric diffusions as a tool for harmonic analysis and structure definition of data: diffusion maps. *Proc Natl Acad Sci U S A* 2005;**102**:7426–31.
26. Ocone A, Haghverdi L, Mueller NS, Theis FJ. Reconstructing gene regulatory dynamics from high-dimensional single-cell snapshot data. *Bioinformatics* 2015;**31**:i89–96.
27. Jarvis RA, Patrick EA. Clustering using a similarity measure based on shared near neighbors. *IEEE Trans Comput* 2006;**C-22**:1025–34.
28. Satija R, Farrell JA, Gennert D, *et al*. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* 2015;**33**:495.
29. Butler A, Hoffman P, Smibert P, *et al*. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* 2018;**36**:411.
30. Choi H, Sheng J, Gao D, *et al*. Transcriptome analysis of individual stromal cell populations identifies stroma-tumor crosstalk in mouse lung cancer model. *Cell Rep* 2015;**10**:1187–201.
31. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 2010;**33**:1–22.
32. Cox DR, Oakes D. *Analysis of Survival Data*. London: Chapman & Hall, 1984.
33. Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* 2015;**56**:337–44.
34. Cheng W, Ren X, Zhang C, *et al*. Bioinformatic profiling identifies an immune-related risk signature for glioblastoma. *Neurology* 2016;**86**:2226–34.
35. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;**44**:837–45.
36. Etienne MC, Formento JL, Lebrun-Frenay C, *et al*. Epidermal growth factor receptor and labeling index are independent prognostic factors in glial tumor outcome. *Clin Cancer Res* 1998;**4**:2383–90.
37. Li J, Liang R, Song C, *et al*. Prognostic significance of epidermal growth factor receptor expression in glioma patients. *Onco Targets Ther* 2018;**11**:731–42.
38. Simon N, Friedman J, Hastie T, Tibshirani R. Regularization paths for Cox's proportional hazards model via coordinate descent. *J Stat Softw* 2011;**39**:1–13.
39. Taylor JMG. Random survival forests. *J Thorac Oncol* 2008;**2**:841–60.
40. Gordon S, Taylor PR. Monocyte and macrophage heterogeneity. *Nat Rev Immunol* 2005;**5**:953.
41. Oldham MC, Konopka G, Iwamoto K, *et al*. Functional organization of the transcriptome in human brain. *Nat Neurosci* 2008;**11**:1271.
42. Baumann N, Pham-Dinh D. Biology of oligodendrocyte and myelin in the mammalian central nervous system. *Physiol Rev* 2001;**81**:871–927.
43. Hambardzumyan D, Gutmann DH, Kettenmann H. The role of microglia and macrophages in glioma maintenance and progression. *Nat Neurosci* 2016;**19**:20.
44. Pyonteck SM, Akkari L, Schuhmacher AJ, *et al*. CSF-1R inhibition alters macrophage polarization and blocks glioma progression. *Nat Med* 2013;**19**:1264.
45. Hussain SF, Yang D, Suki D, *et al*. The role of human glioma-infiltrating microglia/macrophages in mediating antitumor immune responses. *Neuro Oncol* 2006;**8**:261–79.
46. Zheng Y, Bao J, Zhao Q, *et al*. A spatio-temporal model of macrophage-mediated drug resistance in glioma immunotherapy. *Mol Cancer Ther* 2018;**17**:814–24.
47. Quail DF, Bowman RL, Akkari L, *et al*. The tumor microenvironment underlies acquired resistance to CSF1R inhibition in gliomas. *Science* 2016;**352**:aad3018.
48. Bonavia R, Inda MM, Vandenberg S, *et al*. EGFRvIII promotes glioma angiogenesis and growth through the NF-κB, interleukin-8 pathway. *Oncogene* 2012;**31**:4054–66.
49. Gondi CS, Dinh DH, Klopfenstein JD, *et al*. MMP-2 downregulation mediates differential regulation of cell death via ErbB-2 in glioma xenografts. *Int J Oncol* 2009;**35**:257–63.
50. Sareddy GR, Li X, Liu J, *et al*. Selective estrogen receptor β agonist LY500307 as a novel therapeutic agent for glioblastoma. *Sci Rep* 2016;**6**:24185.
51. Wang Z, Yuan H, Sun C, *et al*. GATA2 promotes glioma progression through EGFR/ERK/Elk-1 pathway. *Med Oncol* 2015;**32**:87.

52. Lv P, Wang W, Cao Z, *et al*. Fsk and IBMX inhibit proliferation and proapoptotic of glioma stem cells via activation of cAMP signaling pathway. *J Cell Biochem* 2019;**120**: 321–31.
53. Xu Q, Chiao P, Sun Y. Amphiregulin in cancer: new insights for translational medicine. *Trends Cancer* 2016;**2**:111.
54. Moraes LA, Kar S, Foo SL, *et al*. Annexin-A1 enhances breast cancer growth and migration by promoting alternative macrophage polarization in the tumour microenvironment. *Sci Rep* 2017;**7**:17925.
55. Coniglio SJ, Eugenin E, Dobrenis K, *et al*. Microglial stimulation of glioblastoma invasion involves epidermal growth factor receptor (EGFR) and colony stimulating factor 1 receptor (CSF-1R) signaling. *Mol Med* 2012;**18**: 519–27.
56. Wang C, Wang L, Su B, *et al*. Serine protease inhibitor Kazal type 1 promotes epithelial–mesenchymal transition through EGFR signaling pathway in prostate cancer. *Prostate* 2014;**74**:689–701.
57. Kikuchi T, Joki T, Akasaki Y, *et al*. Induction of antitumor immunity using intercellular adhesion molecule 1 (ICAM-1) transfection in mouse glioma cells. *Cancer Lett* 1999;**142**:201.
58. Zhang Y, Qiao HX, Zhou YT, *et al*. Fibrinogen-like-protein 1 promotes the invasion and metastasis of gastric cancer and is associated with poor prognosis. *Mol Med Rep* 2018;**18**:1465–72.
59. Leone P, Shin EC, Perosa F, *et al*. MHC class I antigen processing and presenting machinery: organization, function, and defects in tumor cells. *J Natl Cancer Inst* 2013;**105**:1172–87.
60. Angelopoulou E, Piperi C. Emerging role of plexins signaling in glioma progression and therapy. *Cancer Lett* 2018;**414**:81–7.
61. Lee S, Piccolo SR, Allen-Brady K. Robust meta-analysis shows that glioma transcriptional subtyping complements traditional approaches. *Cell Oncol (Dordr)* 2014;**37**:317–29.
62. Shamir M, Baron Y, Phillips R, Milo R. SnapShot: timescales in cell biology. *Cell* 2016;**164**:1302–1302.e1301.
63. Komarova NL, Zou X, Nie Q, Bardwell L. A theoretical framework for specificity in cell signaling. *Mol Syst Biol* 2005;**1**:2005.0023.
64. Avraham R, Yarden Y. Feedback regulation of EGFR signalling: decision making by early and delayed loops. *Nat Rev Mol Cell Biol* 2011;**12**:104–17.
65. Wang E (ed). *Cancer Systems Biology*. Boca Raton, FL: CRC Press, 2010.
66. Wang E, Zaman N, Mcgee S, *et al*. Predictive genomics: a cancer hallmark network framework for predicting tumor clinical phenotypes using genome sequencing data. *Semin Cancer Biol* 2015;**30**:4–12.
67. Sun X, Hu B. Mathematical modeling and computational prediction of cancer drug resistance. *Brief Bioinform* 2018;**19**:1382–99.
68. Vaske CJ, Benz SC, Sanborn JZ, *et al*. Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. *Bioinformatics* 2010;**26**: i237–45.
69. Li J, Lenferink AEG, Deng Y, *et al*. Corrigendum: identification of high-quality cancer prognostic markers and metastasis network modules. *Nat Commun* 2010;**1**:34.
70. Zaman N, Li L, Jaramillo ML, *et al*. Signaling network assessment of mutations and copy number variations predict breast cancer subtype-specific drug targets. *Cell Rep* 2013;**5**:216–23.
71. Deribe YL, Pawson T, Dikic I. Post-translational modifications in signal integration. *Nat Struct Mol Biol* 2010;**17**:666–72.
72. Wilhelm M, Schlegl J, Hahne H, *et al*. Mass-spectrometry-based draft of the human proteome. *Nature* 2014; **509**:582.
73. Huang PH, Xu AM, White FM. Oncogenic EGFR signaling networks in glioma. *Sci Signal* 2009;**2**:re6.
74. Munirathnam V, Philip A, R B, *et al*. Analysis of EGFRvIII and EGFR overexpression in glioma and its prognostic significance. *Ann Oncol* 2016;**27** suppl_6:vi103–13.
75. Bao ZS, Chen HM, Yang MY, *et al*. RNA-seq of 272 gliomas revealed a novel, recurrent PTPRZ1-MET fusion transcript in secondary glioblastomas. *Genome Res* 2014;**24**:1765–73.
76. Heimberger AB, Dima S, Yang D, *et al*. The natural history of EGFR and EGFRvIII in glioblastoma patients. *J Transl Med* 2005;**3**:38.
77. Normanno N, De Luca A, Bianco C, *et al* . Epidermal growth factor receptor (EGFR) signaling in cancer. *Gene* 2006;**366**:2–16.
78. Ching T, Zhu X, Garmire LX. Cox-nnet: an artificial neural network method for prognosis prediction of high-throughput omics data. *PLoS Comput Biol* 2018; **14**:e1006076.
79. Katzman JL, Shaham U, Cloninger A, *et al*. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC Med Res Methodol* 2016;**18**:24.
80. Ziegenhain C, Vieth B, Parekh S, *et al*. Comparative analysis of single-cell RNA sequencing methods. *Mol Cell* 2017;**65**: 631–643.e634.