Large-scale bulk and single-cell RNA sequencing combined with machine learning reveals glioblastoma-associated neutrophil heterogeneity and establishes a *VEGFA*⁺ neutrophil prognostic model

Yufan Yang[1,3†], Ziyuan Liu[2,3†], Zhongliang Wang[4†], Xiang Fu[1,3], Zhiyong Li[7], Jianlong Li[5,6*], Zhongyuan Xu[1,3*] and Bohong Cen[1,3*]

Abstract

Background   Neutrophils play a key role in the tumor microenvironment (TME); however, their functions in glioblastoma (GBM) are overlooked and insufficiently studied. A detailed analysis of GBM-associated neutrophil (GBMAN) subpopulations may offer new insights and opportunities for GBM immunotherapy.

Methods   We analyzed single-cell RNA sequencing (scRNA-seq) data from 127 isocitrate dehydrogenase (IDH) wild-type GBM samples to characterize the GBMAN subgroups, emphasizing developmental trajectories, cellular communication, and transcriptional networks. We implemented 117 machine learning combinations to develop a novel risk model and compared its performance to existing glioma models. Furthermore, we assessed the biological and molecular features of the GBMAN subgroups in patients.

Results   From integrated large-scale scRNA-seq data (498,747 cells), we identified 5,032 neutrophils and classified them into four distinct subtypes. *VEGFA*⁺GBMAN exhibited reduced inflammatory response characteristics and a tendency to interact with stromal cells. Furthermore, these subpopulations exhibited significant differences in transcriptional regulation. We also developed a risk model termed the "*VEGFA*⁺neutrophil-related signature" (VNRS) using machine learning methods. The VNRS model showed higher accuracy than previously published risk models

†Yufan Yang, Ziyuan Liu and Zhongliang Wang contributed equally to this work.

*Correspondence:
Jianlong Li
jianlongyx@163.com
Zhongyuan Xu
nflcyljd@smu.edu.cn
Bohong Cen
cenbohong22@i.smu.edu.cn

Full list of author information is available at the end of the article

and was an independent prognostic factor. Additionally, we observed significant differences in immunotherapy responses, TME interactions, and chemotherapy efficacy between high-risk and low-risk VNRS score groups.

Conclusion Our study highlights the critical role of neutrophils in the TME of GBM, allowing for a better understanding of the composition and characteristics of GBMAN. The developed VNRS model serves as an effective tool for evaluating the risk and guiding clinical treatment strategies for GBM.

Clinical trial number Not applicable.

Graphical abstract
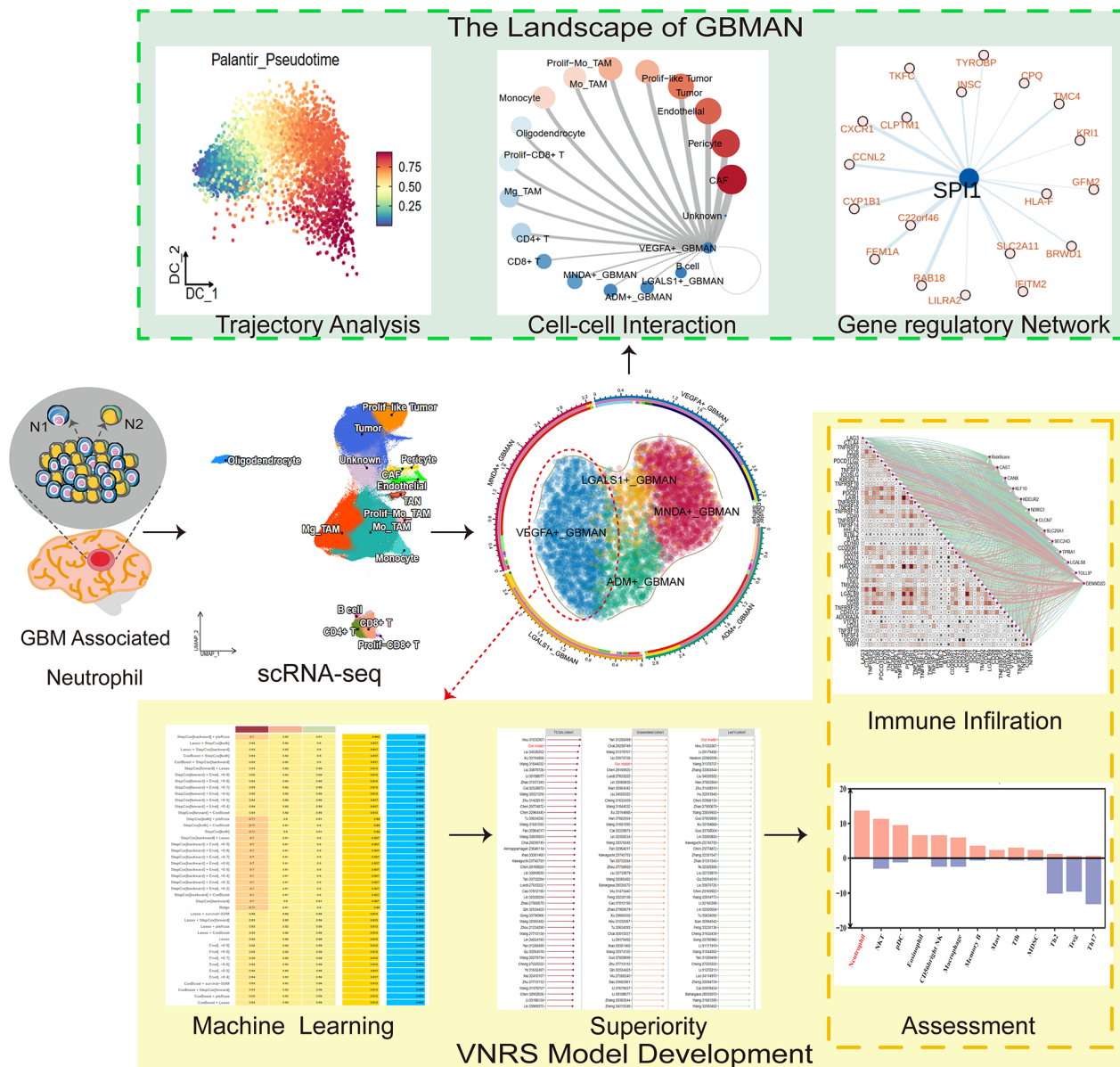
Highlights

- Large-scale scRNA-seq of GBM identified four neutrophil subtypes and characterized their molecular landscape.
- Trajectory analysis identified the developmental process of *MNDA*$^+$GBMAN into *VEGFA*$^+$GBMAN.
- *VEGFA*$^+$GBMAN exhibits significant immunosuppressive and hypoxic characteristics.
- A VNRS prognostic model based on *VEGFA*$^+$GBMAN was proposed using 117 machine learning combinations.
- The VNRS model demonstrates significant superiority and can accurately predict patient characteristics.

Keywords Glioblastoma, Single-cell RNA sequencing, Neutrophil, Machine learning, Heterogeneity

Introduction

The TME of GBM is a complex ecosystem comprising not only neoplastic cells but also a diverse array of non-tumor elements, including immune and stromal cells [1, 2]. Among these, neutrophils, the most abundant immune cells in the human circulatory system, serve as the primary line of defense against microbial infections and are rapidly recruited to sites of tissue injury and inflammation [3]. In the context of GBM, IDH-wildtype GBM represents the most common and aggressive subtype of astrocytic tumors, accounting for over 95% of all GBM cases [4]. This subtype is characterized by a higher degree of immune cell infiltration and a more pronounced immunosuppressive microenvironment compared to IDH-mutant GBM. Furthermore, the function of myeloid cells, including neutrophils, may be differentially modulated by tumor necrosis and inflammation in IDH-wildtype GBM, which constitutes a key distinguishing feature from its IDH-mutant counterpart [5].

As key components of the TME, neutrophils are widely distributed throughout the circulatory system, and those in the brain parenchyma may primarily originate from the skull and adjacent spinal bone marrow [6]. In the circulatory system, neutrophils constitute more than half of all white blood cells, whereas in the TME of GBM, their number is only about one-seventh of that of monocytes. Additionally, neutrophils are characterized by a relatively short lifespan (19 h to 5.4 days) and a lack of proliferative capacity, distinguishing them from other immune cells [7]. This limited lifespan and function have contributed to a perception of neutrophils as specialized cells with a restricted range of activities in immune defense. Consequently, research on the functional diversity of neutrophils is less extensive than that on other myeloid cells [8, 9]. These factors lead to an underestimation of the role of neutrophils in GBM, resulting in a greater research focus on macrophages, which are more numerous in the TME of GBM, while neutrophils receive comparatively less attention.

As research into gliomas advances, the critical role of neutrophils in glioma biology becomes increasingly evident. Few studies demonstrated that neutrophils undergo phenotypic changes upon interaction with glioma cells, acquiring a range of functions, including contributions to glioma progression and immune suppression [10, 11]. In the early stages, neutrophils inhibit tumor cells through antibody-dependent cellular cytotoxicity. However, with prolonged contact, their antitumor effects diminish, while pro-tumor activities intensify. Additionally, neutrophils interact with other cells via various factors and receptors, further driving glioma proliferation, invasion and angiogenesis [12].

In GBM, the precise functions and molecular basis underlying the heterogeneity of specific neutrophil subpopulations remain diverse and challenging to elucidate [13]. ScRNA-seq is a powerful tool for investigating immune cell heterogeneity [14]. Advances in single-cell technologies now enable the accurate identification of neutrophil subtypes with distinct features during GBM progression. This approach aids in uncovering dynamic changes in gene expression within these subtypes, elucidating the molecular mechanisms that drive tumorigenesis, and identifying potential diagnostic and therapeutic targets. However, it is important to note that neutrophil is a fragile cell type, prone to loss during tissue dissociation. Moreover, neutrophils express a limited number of genes, and the expression levels of key genes are often low, complicating the analysis of their subtypes and gene expression profiles. Additionally, the high cost of single-cell sequencing technologies presents a significant barrier to their broader application in clinical research on neutrophils. Nevertheless, subtyping neutrophils in GBM patients to assess prognosis and inform treatment decisions is feasible. Developing a simple and effective method for characterizing the "neutrophil signature" in GBM patients is essential. As bioinformatics advances, machine learning is increasingly used in clinical risk modeling [15]. However, the efficacy of these multi-gene expression signatures may be difficult to validate and implement effectively due to limitations in single machine learning approach and statistical methods.

Based on this foundation, we integrated an unprecedented large-scale scRNA-seq dataset to characterize GBMAN. Through trajectory analysis and differentiation potential calculations, we identified key developmental features of GBMAN. Furthermore, we examined the dynamic interactions between neutrophils and other cell types. Subsequently, transcriptional regulatory network analysis was employed to uncover the epigenetic landscape of GBMAN. Lastly, given the diagnostic importance of GBMAN, we developed a novel risk prediction model leveraging advanced machine learning techniques. In conclusion, our study offers enhanced insights into the landscape of GBMAN and provides a predictive framework that supports clinical diagnosis and improves treatment outcomes for GBM patients.

Materials and methods

Data sources

In this study, we analyzed scRNA-seq data from 127 samples of 66 IDH wild-type GBM patients collected from nine datasets. The inclusion criteria required samples to be either primary or recurrent and to include tumor cells, immune cells, and stromal cells. The datasets used were EGAS00001004656 [16], EGAS00001005300 [17], GSE138794 [18], GSE141383 [19], GSE162631 [20], GSE173278 [21], GSE182109 [22], GSE223063 [23], and GSE235676 [24], sourced from the Gene Expression

Omnibus (GEO) and European Genome-phenome Archive (EGA) databases (Table S1). These datasets were publicly available or accessed with formal authorization. We also obtained bulk RNA-seq and microarray data for GBM and normal samples from the GEO, the Cancer Genome Atlas (TCGA), and the Genotype-Tissue Expression (GTEx) databases [25, 26]. The criteria for inclusion required tumor sample sizes exceeding 150. The datasets included in this analysis were TCGA-GBM ($n = 155$), GSE13041 ($n = 191$) [27], and GSE16011 ($n = 155$) [28].

Single-cell RNA sequencing data process and integration

To ensure the accuracy of the results, we unified the reference genome for all scRNA-seq data to hg38 genome. All expression matrices, except for EGAS00001005300, were generated using hg38 genome. To align the raw data of EGAS00001005300 with the hg38 version of the 10X reference genome (v2.0.0), we employed the Cellranger (v3.0.2) pipeline. All data were preprocessed using the Scanpy (v1.9.3) package [29]. First, low-quality cells were filtered out based on a cutoff threshold of total RNA features fewer than 200 and mitochondrial RNA exceeding 30%. Subsequently, following quality control and preliminary normalization, we performed batch effect correction using Batch Balanced k-Nearest Neighbors (BBKNN) to align multiple samples while preserving biological variance [30, 31]. The implementation was carried out through Scanpy's sc.external.pp.bbknn function using the following key parameters: (1)

principal component input (n_pcs = 50) from precomputed principal component analysis (PCA), (2) per-batch neighbor connections (neighbors_within_batch = 3), (3) approximate nearest neighbor detection (approx = True) accelerated by Annoy library with n_trees = 10, and (4) Euclidean distance metric (metric = "euclidean"). The UMAP algorithm was used to construct neighborhood graphs and perform dimensionality reduction on the gene expression profiles of each cell, focusing on the 1,500 most variable genes [32]. Finally, clustering analysis was performed on the modified neighborhood graph using the Leiden community detection algorithm (resolution = 0.6) [33].

In the identification of subgroups within GBMAN, 1500 highly variable genes were selected, and the cluster resolution was set to 1. To avoid artifacts in expression due to accidental noise and dissociation, a total of 1514 genes associated with mitochondria (50 genes), heat shock proteins (178 genes), ribosomes (1253 genes), and dissociation (33 genes) were excluded [34].

Cell annotations

The h5ad data produced by Scanpy was transformed into a Seurat object [35]. Based on the clustering results, we determined the cell types of each cell group by analyzing the expression of well-established marker genes [35, 36]. Ultimately, apart from the unknown cell, a total of 15 cell types were annotated for further analysis.

Assessment of GBMAN subsets purity

We applied the default parameter settings of the ROGUE package to assess GBMAN heterogeneity, following the recommended procedures [37]. The Ratio of Global Unshifted Entropy (ROGUE) index ranges from 0 to 1, where a value of 1 indicates a completely pure subtype, and 0 represents the highest level of population heterogeneity.

Trajectory analysis

To investigate the dynamic developmental trajectories of neutrophil subsets, we mapped them onto a diffusion map and conducted trajectory analyses using the Cyto-TRACE2 and Palantir Python packages. CytoTRACE2 (v1.0.0) facilitates the inference of cell differentiation order by quantifying the similarity of gene expression profiles between individual cells [38]. In addition, Palantir (v1.3.3) was employed to further analyze the trajectory characteristics of distinct subsets and assess their differentiation potential [39]. Finally, by fitting differentiation potential curves and calculating their first and second derivatives, we identified the pseudotime points marking the rapid evolution of these subsets.

Immune response enrichment analysis

We employed immune response enrichment analysis to evaluate the correlation between gene expression and cytokine response in GBMAN subgroups [40]. The rows of the input gene expression matrix represent individual genes, while the columns correspond to cell IDs. Using cytokine response data specific to neutrophils from the immune dictionary database, we extracted the relevant gene expression matrix. Subsequently, we computed the similarity between the gene expression profiles of various neutrophil subtypes and a reference neutrophil sample. we applied the Wilcoxon rank-sum test to assess the differential response intensity of each neutrophil subgroup under various cytokine treatments and calculated the corresponding enrichment scores.

Cell communication analysis

To study the roles of different neutrophil subtypes in regulating other cells within the GBM microenvironment, we used CellChat package (v1.6.1) to infer, analyze, and visualize receptor-ligand signaling pathways between various cell types, with all parameters set to default [41]. Additionally, we utilized the scMLnet package (v0.1.0), including signals with a p-value less than 0.05 and a log2FC greater than 1 to generate a multi-layer signaling network [42]. This detailed the effects of each neutrophil

subtype on other cell types and elucidated the ligand-receptor-transcription factor (TF)-target gene cascade.

Gene regulatory network analysis
An analysis of the gene regulatory network was conducted employing the single-cell regulatory network inference and clustering (SCENIC) methodology, utilizing the pySCENIC (v0.10.0) implementation [43]. The workflow of pySCENIC was conducted with default parameters, taking the raw count matrix of all neutrophils as input [44]. The igraph package was employed to visualize the gene regulatory networks of differential regulatory pathways between primary and recurrent events [45]. The pathway enrichment analysis between the primary and recurrent groups was performed using the GSEA function in the clusterProfiler package [46].

Development of VNRS model using univariate Cox and 117 machine learning combinations
Differentially expressed genes in *VEGFA*⁺neutrophil subpopulations were identified using the FindAllMarkers function within the Seurat package, employing the Wilcoxon rank-sum test as the statistical method. Signature markers were defined as genes achieving an average log2 fold change (avg_log2FC) greater than 1 combined with an adjusted p-value below 0.05. Prognosis-associated genes were then identified through univariate Cox regression analysis using the coxph function from the survival R package, with a significance threshold set at $p < 0.1$. To develop a robust predictive model, we integrated ten machine learning algorithms: stepwise Cox, CoxBoost, Lasso, Ridge, Elastic Net (Enet, $\alpha = 0.1–0.9$), survival-SVMs, SuperPC, Generalized Boosted Regression Models (GBMs), partial least squares Cox (plsRcox), and Random Survival Forest (RSF). A total of 117 algorithm combinations were tested under a leave-one-out cross-validation (LOOCV) framework using the mime R package [47]. Model performance was evaluated by consistency index (C-index) across validation datasets. The TCGA-GBM dataset served as the training set, while external validation cohorts from Gravendeel et al. and Lee Y et al. were used for independent verification. The final model selection prioritized the highest average C-index across validation sets. Patients were stratified into high- and low-risk groups based on the median risk score derived from the prognostic model. Kaplan-Meier survival curves were generated using the survfit function from the Survival (v3.5-5) and Survminer (v0.4.9) packages to assess intergroup survival differences [48]. The risk indices of each gene in the risk model formula available in Table S2.

Immune evaluation and prediction of the response to immunotherapy
We investigated the differences in immune cell activity and proportions between VNRS high-risk and low-risk groups. The ESTIMATE algorithm was applied to evaluate both immune cell proportions and tumor purity [49]. Additionally, single-sample gene set enrichment analysis (ssGSEA) from the GSVA (v1.48.2) [50] was employed to analyze the variations in immune cell activity and immune pathways between these groups. Immunotherapy with checkpoint inhibitors is an increasingly critical strategy for treating GBM. Furthermore, we examined the correlation between the model genes and 46 immune checkpoint inhibitor (ICI) genes using correlation analysis.

Immune-cell interaction enrichment analysis
TimiGP identified immune cell interactions in VNRS high- and low-risk groups [51]. TimiGP is a robust computational method designed to infer intercellular interactions in the TME, demonstrating the association between the relative abundance of immune cells and prognosis, and it can indicate which immune cells or cell interactions are beneficial to patient outcomes. The 28 cell types reported by Charoentong et al. were used to infer cell-cell interactions in high- and low-risk groups [52].

Drug sensitive analysis
Sensitivity scores for drugs were acquired from the Genomics of Drug Sensitivity in Cancer (GDSC) database [53] through the utilization of the oncoPredict package [54]. GDSC2 dose-response curves were normalized, and batch effects were mitigated. Gene expression profiles of the studied cell lines were log2-transformed and harmonized. Tenfold cross-validation optimized model robustness, with λ automatically selected via minimum cross-validated error. A comprehensive analysis was conducted on 198 anticancer compounds from the GDSC2 dataset. Unlike GDSC1, the GDSC2 dataset integrates more contemporary sequencing data and experimental outcomes.

Statistical analysis
Statistical analyses and graphical representations were carried out by R (v4.3.1), Python (v3.10.9), and GraphPad Prism (v8.3). Survival curves were analyzed using the Kaplan-Meier method with the log-rank test for significance testing, while the Pearson correlation coefficient was employed to determine the strength of linear associations. For comparisons across multiple groups, one-way ANOVA was applied. Significance levels in the graphical data are denoted as follows: $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, $****P < 0.0001$, with "ns" indicating a lack of statistical significance.

Results

Composition of GBMAN's four main subpopulations

Neutrophils are among the less infiltrated immune cells in GBM. To comprehensively understand the landscape of GBMAN, we collected scRNA-seq data from primary and recurrent IDH wild-type GBM samples across nine datasets (Fig S1A-C). To ensure accuracy, we included only samples that contained both tumor cells and microenvironment cells, with the reference genome standardized through upstream analysis. Rigorous quality control was conducted (Fig S1D), and the BBKNN algorithm was employed to eliminate batch effects [30, 31]. As a result, we integrated a large single-cell transcriptome dataset comprising 498,747 cells from 127 samples of 66 patients (Fig. 1A).

After dimensionality reduction and clustering, we annotated the cell clusters with detailed marker information, identifying 15 distinct cell types, including
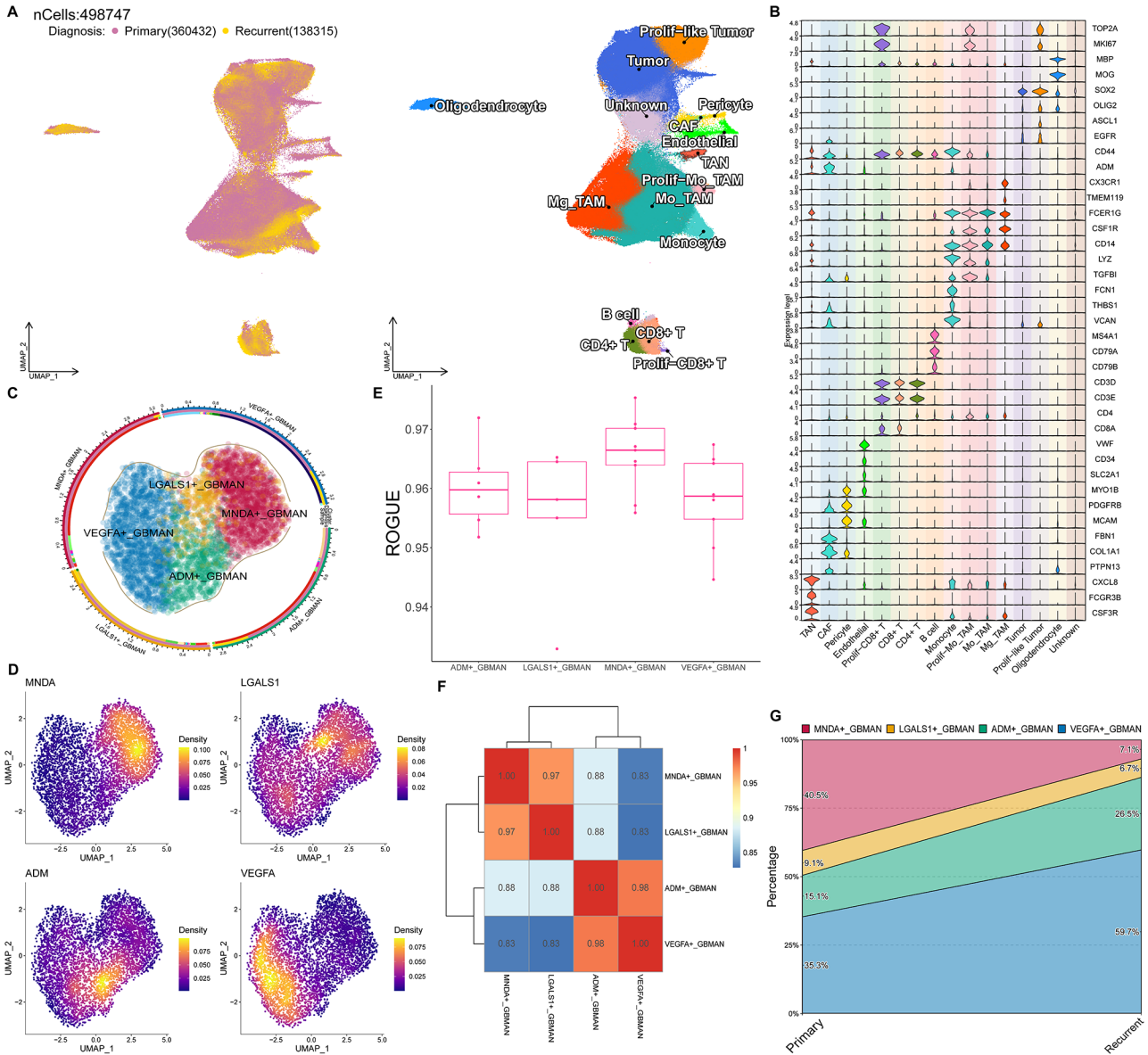

Fig. 1 Composition of GBMAN's four main subpopulations. (**A**) Integration of 127 human GBM scRNA-seq data across 9 individual datasets using BBKNN. A total of 498,747 single-cell transcriptomes were processed and integrated to reduce batch effects and ensure robust clustering. UMAP was employed to visualize the high-dimensional data in a two-dimensional space. (**B**) Violin plot illustrating marker gene expression for each major cell cluster. The plot highlights the expression levels of key markers across identified cell clusters. The width of each violin represents the distribution density of gene expression. (**C**) UMAP visualization of four distinct subgroups of neutrophils. Each subgroup is color-coded for clarity. (**D**) The box plot shows the ROGUE scores of the four subgroups. (**E**) FeaturePlot showing the spatial distribution of four marker genes across the UMAP embedding. (**F**) Heatmap depicting the correlation between the four neutrophil subgroups. Hierarchical clustering was performed to assess the similarity of gene expression profiles among subgroups, with colors representing correlation coefficients. (**G**) The proportion changes of the four neutrophil subtypes in primary and recurrent cases
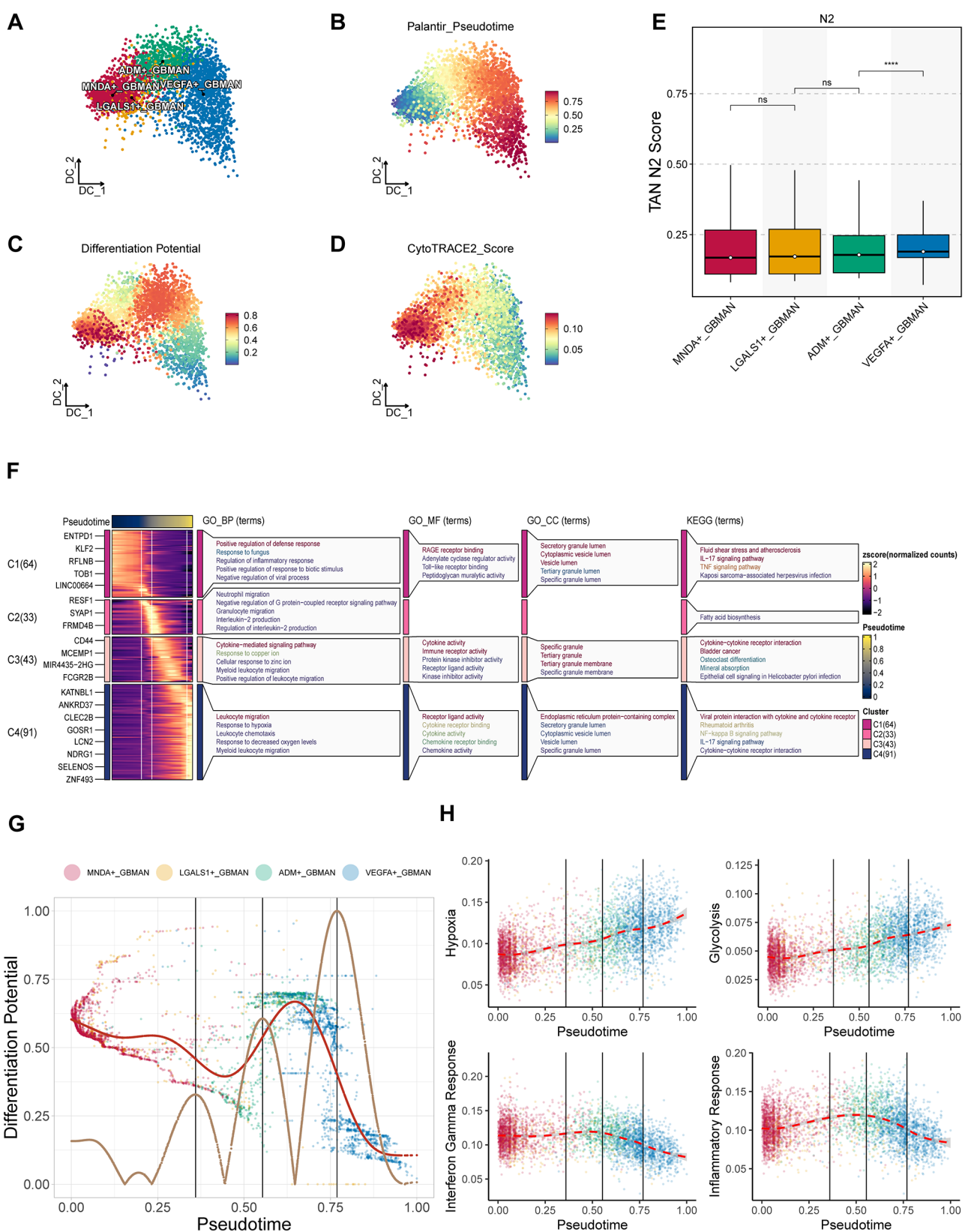
Fig. 2 (See legend on next page.)

(See figure on previous page.)

Fig. 2 Developmental Trajectory Characteristics of GBMAN. (**A**) Diffusion map visualization of the distribution of four neutrophil subsets. The diffusion map captures the underlying structure of the data, revealing the relative positions and relationships of the four neutrophil subsets. (**B-D**) Feature plots respectively display the pseudotime, differentiation potential, and stemness score. (**E**) Boxplot shows a comparison of N2 scores among the four subgroups. (**F**) Gene trajectories were clustered based on expression patterns over pseudotime, and functional enrichment analyses were performed for Gene Ontology Biological Processes (GO_BP), Molecular Functions (GO_MF), Cellular Components (GO_CC), and KEGG pathways. (**G**) Scatterplot depicting the correlation between pseudotime (developmental progression) and differentiation potential for each neutrophil subset. The plot reveals how the differentiation potential changes as cells progress along the pseudotime trajectory. (**H**) Scatterplots showing the association between pseudotime and key functional states, including hypoxia, glycolysis, interferon-γ response, and immune response. These plots illustrate how cellular functional states evolve along the developmental continuum

tumor-associated neutrophil (TAN), cancer-associated fibroblast (CAF),

pericyte, endothelial, prolif-*CD8*$^+$ T, *CD8*$^+$ T, *CD4*$^+$ T, B cell, monocyte, prolif-like Mo_TAM, Mo_TAM, Mg_TAM, tumor, prolif-like tumor, and oligodendrocyte (Fig. 1B) [35, 36]. Cells without specific marker expression were classified as unknown cell. We then isolated TANs and performed secondary dimensionality reduction and clustering, identifying four GBMAN subtypes: *MNDA*$^+$GBMAN, *LGALS1*$^+$GBMAN, *ADM*$^+$GBMAN, and *VEGFA*$^+$GBMAN (Fig. 1C). Specific marker expression for each subtype is illustrated in Fig. 1D.

To verify the validity of these subgroups, we assessed the purity of each GBMAN subtype and found that ROGUE values exceeded 0.9 for all four subtypes (Fig. 1E), indicating high purity and stability [37]. Correlation analysis revealed that *MNDA*$^+$GBMAN and *LGALS1*$^+$GBMAN shared similar characteristics, as did *ADM*$^+$GBMAN and *VEGFA*$^+$GBMAN (Fig. 1F). Finally, a comparison of subgroup proportions showed that the proportion of *MNDA*$^+$GBMAN significantly decreased in recurrent patients, while *VEGFA*$^+$GBMAN increased markedly (Fig. 1G). These findings suggested that four subtypes of neutrophils are present in the GBM microenvironment, with *VEGFA*$^+$GBMAN becoming more prominent in recurrent samples.

Developmental trajectory characteristics of GBMAN

TANs, despite their short lifespan, are continually influenced by various factors within the immune microenvironment, leading to dynamic shifts between pro-tumor and anti-tumor activities during their development [55]. To precisely map the evolutionary process of GBMANs, we analyzed primary TANs using the diffusion map method to investigate their developmental trajectories. This technique arranges cells based on transition probabilities, allowing for better preservation of differentiation pathways [56, 57]. Additionally, through the application of the Palantir algorithm, we found that neutrophils gradually develop from *MNDA*$^+$GBMAN to *VEGFA*$^+$GBMAN (Fig. 2A-B). Notably, we identified that *ADM*$^+$GBMAN also exhibits significant differentiation potential, suggesting a shift in neutrophil differentiation at this stage (Fig. 2C). Finally, using the Cytotrace2 algorithm, we confirmed that the stemness score of

MNDA$^+$GBMAN progressively decreases as it differentiates into *VEGFA*$^+$GBMAN (Fig. 2D), a finding further corroborated by the VECTOR algorithm (Fig. S2A).

TANs can adopt one of two phenotypes: N1 or N2. The N2 phenotype is regarded as immunosuppressive and plays a role in promoting tumor growth [58]. Using the ssGSEA algorithm, we observed that *VEGFA*$^+$GBMAN had the higher N2 score (Fig. 2E). To elucidate the functional dynamics, we clustered genes with distinct expression patterns into four stages and performed GO (BP/CC/MF) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses (Fig. 2F). These analyses revealed that neutrophils initially participate in defense and immune responses, with the TNF pathway activated to exert cytotoxic effects. Over time, cell migration becomes prominent, driven by ligand activity, and eventually, leukocyte migration features emerge, exhibiting clear hypoxic characteristics.

We then performed first derivative fitting to derive a curve showing how differentiation potential changes over pseudo-time. A second derivative analysis identified three critical pseudo-time points of significant changes in neutrophil differentiation potential (Fig. 2G). These points correspond to: (1) the initial phase of rapid differentiation, (2) changes in differentiation induced by specific factors, and (3) a second phase of accelerated differentiation. These findings suggested that neutrophil reprogramming is pivotal in the *ADM*$^+$GBMAN stage. Finally, we scored each cell using hallmark gene sets via ssGSEA and fitted the data to track changes (Fig. 2H). We observed that, under increasing hypoxic and glycolytic conditions, the interferon-gamma and inflammatory responses initially rise until the *ADM*$^+$GBMAN stage, after which they decline rapidly. This indicated that GBMANs initially contribute to pro-inflammatory and anti-tumor responses, but as hypoxia and glycolysis intensify, they are ultimately reprogrammed into immunosuppressive *VEGFA*$^+$GBMAN.

Intercellular communication and cytokine-mediated transformation between the GBMAN subpopulation and other cells

Next, we aimed to elucidate the interactions between GBMAN and other cells within the GBM microenvironment. Recent advancements, such as the development of

an immune dictionary, have highlighted the pleiotropic effects of cytokines, offering a framework for deducing the roles of specific cytokines and the networks of cell-to-cell communication in immune responses [40]. Utilizing neutrophil immune factor response data, we conducted matrix similarity analysis and found that ADM^+GBMAN formation is predominantly stimulated by factors such as IL18 and IL36a, while $VEGFA^+$GBMAN is primarily triggered by IL1a, IL1b, IL18, and TNFa (Fig. 3A).

Next, using CellChat analysis, we identified intercellular communication patterns and compared communication intensities across different groups (Fig. 3B). The results indicated that all subgroups had stronger interactions with stromal cells (e.g., CAF and endothelial cells), followed by tumor cells and macrophages. Moreover, comparisons between primary and recurrent samples revealed that signaling intensity is stronger in primary samples, regardless of whether the subgroups functioned as ligand or receptor providers (Fig. S3A-D).

During the differentiation from $MNDA^+$GBMAN to $VEGFA^+$GBMAN, we observed that when TANs acted as source cells, the SELL-CD34/PODXL axis exhibited a decreasing effect on endothelial cells, while the VEGFA-VEGFR1/VEGFR2 axis progressively strengthened. Additionally, the HBEGF-EGFR axis showed increasing activation in tumor cells. In other myeloid and lymphoid cells, the MIF-(CD74 + CXCR4) and MIF-(CD74 + CD44) axes were particularly prominent in $VEGFA^+$GBMAN. When TANs acted as recipient cells, several axes from multiple cell sources, such as SPP1/FN1/COL1A1-CD44 and ANXA1-FPR1/FPR2, were significantly prominent in ADM^+GBMAN (Fig. 3C) (Fig S3E), suggesting these receptor-ligand pairs may be critical in driving the transformation into $VEGFA^+$GBMAN and promoting immunosuppressive functions.

Given the strong immunosuppressive nature of $VEGFA^+$GBMAN, we further investigated its impact on other cells by integrating intercellular pathways and intracellular subnetworks (Ligand-Receptor-TF-Target Gene) (Fig. 3D-F) (Fig. S3F-G). This analysis, which incorporated cell type-specific gene expression, prior network information, and statistical inference, revealed that $VEGFA^+$GBMAN predominantly influences ITGB1 through ligands such as VEGFA, SPP1, and PLAU. This interaction affects TFs like HIF1A and TWIST1, leading to the activation of target genes such as SERPINE1, thereby promoting ECM remodeling and angiogenesis in stromal cells and facilitating immune evasion by tumor cells. Additionally, ligands such as VEGFA and HBEGF, interacting with EGFR, activate TFs like STAT1 and ETV4, leading to the expression of target genes such as CCND2, which promotes tumor cell growth and development.

Comprehensive map of the gene regulatory networks in GBMAN subtypes

Epigenetic regulation plays a critical role in determining cell phenotypes and their potential for reprogramming [59]. To explore the epigenetic characteristics of different GBMAN subtypes, we first evaluated their genomic stability. Our findings revealed a gradual decrease in genomic instability from $MNDA^+$GBMAN to $VEGFA^+$GBMAN (Fig. 4A), suggesting an increasing suppression of the anti-tumor immune response and a reduced likelihood of functional alterations due to external interventions. Identifying key gene regulatory networks (GRNs) that induce and maintain these cell states may offer novel strategies for reprogramming GBMANs. We employed SCENIC analysis and performed PCA on the TF activity matrix (Fig. 4B), which reveals a clear distinction between $MNDA^+$GBMAN and $VEGFA^+$GBMAN. This indicated a progressive differentiation in transcriptional regulatory features among the four GBMAN subtypes. Further analysis identified specific TFs for each subtype (Fig. 4C). In the primary group, $MNDA^+$GBMAN is associated with CREB5 and FOXP1, $LGALS1^+$GBMAN with BHLHE41 and CEBPE, ADM^+GBMAN with SRF and STAT4, and $VEGFA^+$GBMAN with KLF10 and SREBF1. In the recurrent group, $MNDA^+$GBMAN is associated with ZNF217 and KLF7, $LGALS1^+$GBMAN with RUNX3 and POLE4, ADM^+GBMAN with NFAT5 and EGR2, and $VEGFA^+$GBMAN with ZNF432 and ATF2.

To assess the impact of GRNs on pathway changes in GBMAN subtypes across primary and recurrent stages, we conducted GSEA analysis on the differentially expressed genes of each GBMAN subtype. We then analyzed the enriched genes based on the TF-target gene set (Fig. 4D-G). Our analysis revealed that ZBTB14 upregulates the KRAS signaling pathway in primary $MNDA^+$GBMAN by regulating genes such as DIRC3. Additionally, STAT1 downregulates the interferon-α response pathway in recurrent $LGALS1^+$GBMAN through the regulation of genes like SLC39A7, and also downregulates the MYC target gene V1 pathway in recurrent ADM^+GBMAN by regulating TAP1. SPI1 was found to downregulate the complement pathway in recurrent ADM^+GBMAN by targeting genes such as CXCR1 and CCNL2. These results demonstrated marked disparities in gene regulatory networks between primary and recurrent GBMAN, as well as notable variations among the four subgroups, with each network presiding over unique functions.

Development and evaluation of VNRS prediction model

Given the short overall survival (OS) of GBM and its cold tumor characteristics, developing effective prognostic strategies has become a central focus of GBM
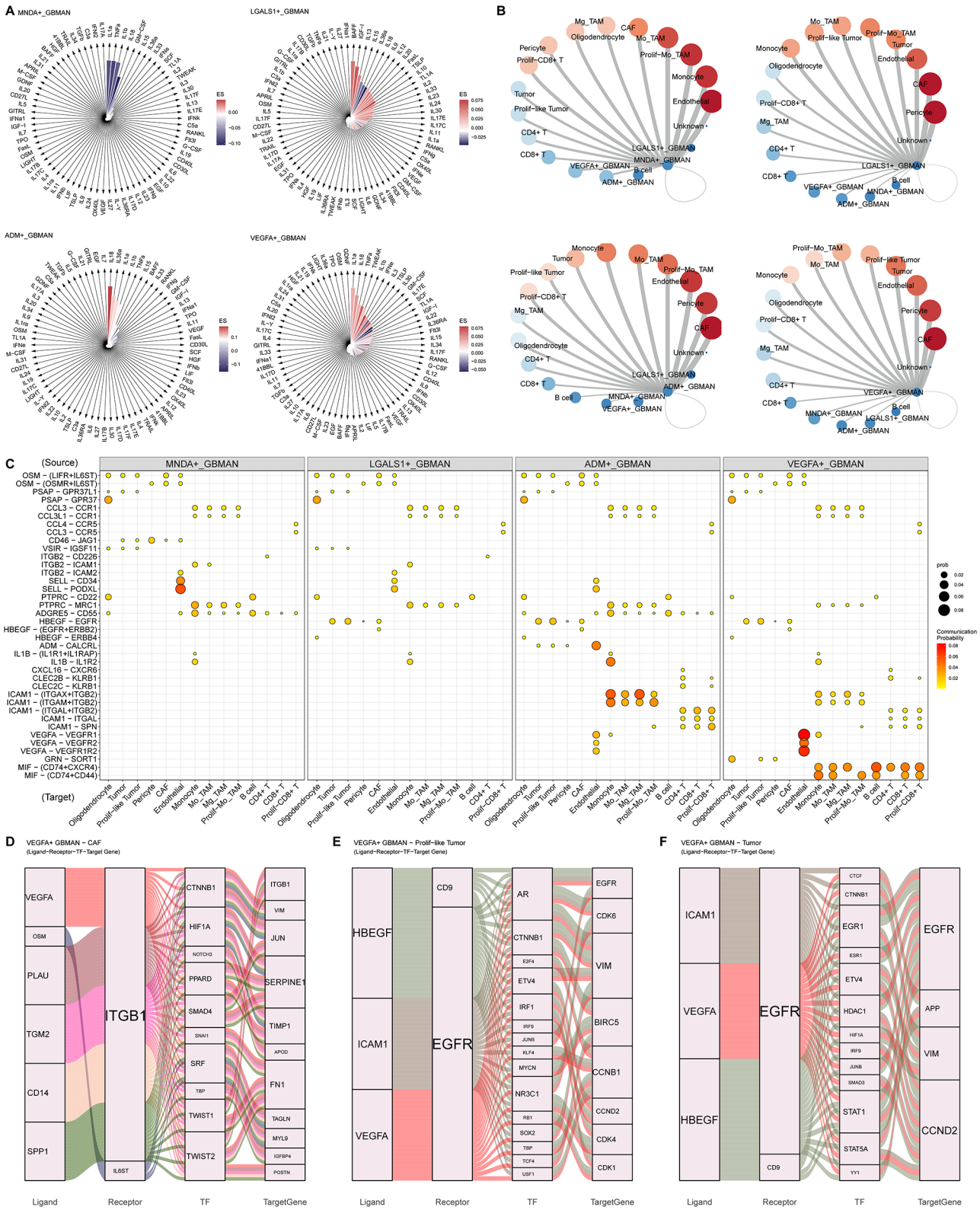
Fig. 3 Intercellular communication and cytokine-mediated transformation between the GBMAN subpopulation and other cells. (**A**) Enrichment results of immune factors in neutrophil subpopulations. (**B**) The shell plot illustrates the communication strength of each neutrophil subgroup with other types of cells. (**C**) The dot plot displays specific ligand-receptor pairs that mediate interactions between each neutrophil subset (source) and other cell types (target). (**D-F**) The Sankey diagram shows the Ligand-Receptor-TF-TargetGene relationships for the three cells with the strongest interaction intensity with *VEGFA*+GBMAN.
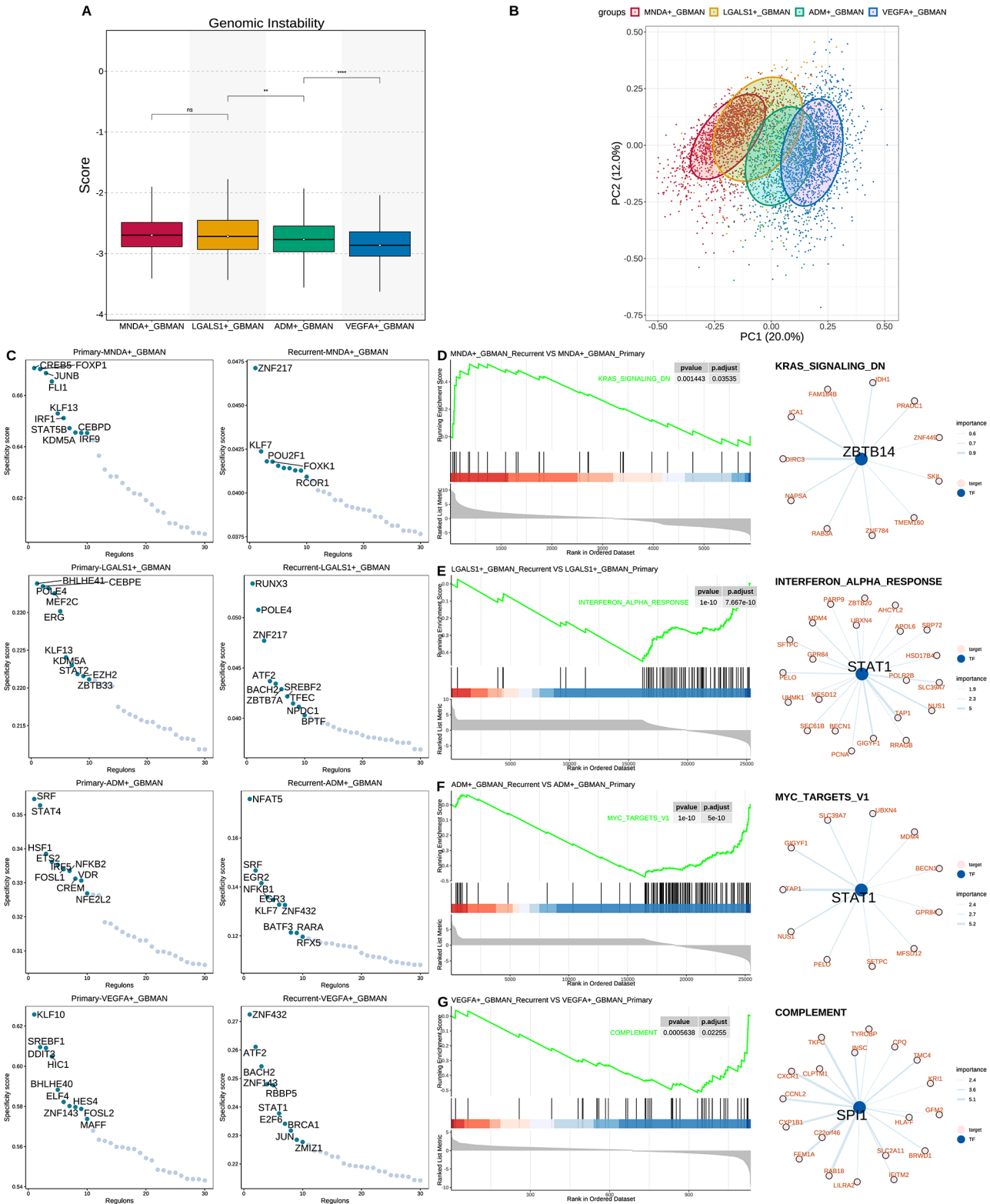
Fig. 4 Comprehensive map of the gene regulatory networks in GBMAN subtypes. (**A**) The boxplot compares the genomic instability scores across different subgroups, reflecting the extent of genomic alterations in four GBMAN subpopulations. (**B**) PCA plot visualizing the activity patterns of TFs across various neutrophil subpopulations. The plot highlights distinct clusters based on TF activity, revealing key regulatory differences between subgroups and their potential roles in functional diversity. (**C**) Specific TFs of various neutrophil subpopulations in the primary and recurrent groups. (**D-G**) The left figure shows the GSEA enrichment analysis of differences between different neutrophil subpopulations in primary and recurrent cases, while the right figure presents the gene regulatory network of each pathway

research [27]. Building on the previously identified *VEGFA*+GBMAN immunosuppressive feature, this study seeks to construct a robust predictive model, termed the "*VEGFA*+Neutrophil-Related Signature" (VNRS). Using specific marker genes from *VEGFA*+GBMAN, we identified genes associated with prognosis. Through univariate

Cox regression analysis based on TCGA-GBM samples, we identified 67 genes linked to hazard ratio (HR) and gene expression, as shown in Fig. 5A.

To develop a consistent prognostic model, we additionally gathered two large cohort datasets as validation sets and implemented a machine learning-based integration
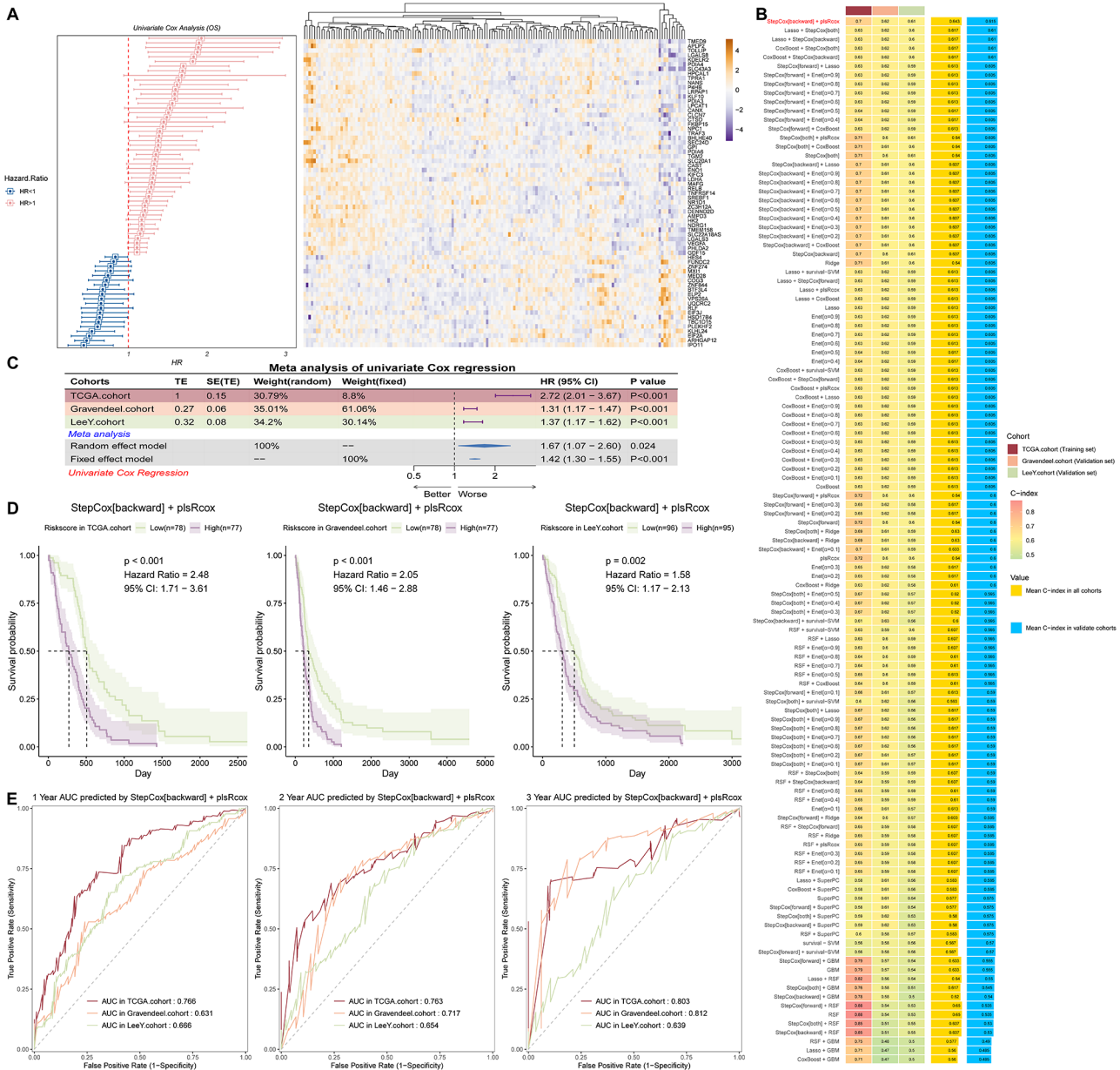

Fig. 5 Development and evaluation of VNRS prediction model. (**A**) The figure shows the univariate analysis results of differentially expressed genes in the *VEGFA*+GBMAN subgroup. The right figure shows the expression of each gene in the TCGA-GBM cohort. (**B**) Heatmap summarizing the construction of 117 predictive models and their corresponding concordance index (C-index) across three large datasets. The C-index quantifies the predictive accuracy of each model, with higher values indicating better performance in discriminating between outcomes. (**C**) Table displaying the results of univariate regression analysis for the predictive model in three independent cohorts. (**D**) Kaplan-Meier curves comparing OS between high- and low-risk groups as defined by the predictive model. The curves are shown for both the training set (used to develop the model) and validation sets (used to test its generalizability), demonstrating the model's ability to stratify patients based on risk. (**E**) ROC curves evaluating the predictive performance of the model at 1-, 2-, and 3-year time points in the three large cohorts. The AUC values are reported, reflecting the model's accuracy in predicting survival outcomes at different time intervals

framework. Specifically, we applied the "Leave-One-Out Cross-Validation" (LOOCV) approach to fit 117 predictive models and calculated the C-index for each model in both the training and validation sets, as illustrated in Fig. 5B. The optimal model, combining StepCox (backward) and plsRcox, achieved the highest average C-index of 0.643. Next, we conducted a meta-analysis of univariate Cox regression across the three large cohorts using the optimal model (Fig. 5C). The HR values were 2.72, 1.31, and 1.37, respectively, with all p-values being less than 0.001. In the meta-analysis, both the random effects and fixed effects models yielded HRs greater than 1, with p-values under 0.05. Subsequently, we classified samples into high-risk and low-risk groups based on the median risk score calculated by the VNRS model and performed a log-rank survival analysis (Fig. 5D). The findings revealed that in all three large cohorts, the high-risk group was significantly correlated with poor prognosis, with statistical significance. Accordingly, this model was designated the VNRS model. Finally, we assessed the 1-year, 2-year, and 3-year ROC curves for the model across the three large GBM cohorts (Fig. 5E). In the TCGA-GBM cohort, the area under the curve (AUC) values were 0.766, 0.763, and 0.803, respectively; in the Gravendeel cohort, the AUC values were 0.631, 0.717, and 0.812, respectively; and in the LeeY cohort, the AUC values were 0.666, 0.654, and 0.639, respectively. These results demonstrated that the VNRS model exhibits strong predictive power and robustness.

Validation of the superiority of the VNRS model compared to published predictive models

To evaluate the superiority of the VNRS model, we compiled 95 previously published glioma-related prediction models for a comprehensive comparison. Risk scores were calculated for these 95 models across three large cohorts, and corresponding HR values were derived (Fig. 6A). The results revealed that only the VNRS model consistently showed a significant association with poor prognosis in all cohorts, with HR values exceeding 1 and p-values less than 0.01, outperforming other published models. Additionally, when comparing the C-index, the VNRS model exhibited superior performance in all three cohorts, ranking second, fifth, and first, respectively (Fig. 6B). Similarly, the 1-year, 2-year, and 3-year AUC values for the VNRS model were consistently among the highest in nearly all cohorts, frequently achieving first place (Fig. 6C-E). These findings suggested that the VNRS model demonstrates great predictive accuracy and generalizability.

Immune infiltration and checkpoint activation in high VNRS risk GBM

Given the important role of the tumor immune microenvironment in GBM, we next evaluated the ability of the VNRS model to predict tumor immune infiltration. According to the ESTIMATE algorithm, higher stromal scores, immune scores and ESTIMATE scores, along with lower tumor purity scores, indicated that tumors with a high VNRS risk score had a higher degree of immune cell infiltration (Fig. 7A-D). This finding suggested that GBM with high VNRS scores has lower purity, which may lead to poorer prognosis in this subgroup of patients. The overall level of immune infiltration is insufficient to reveal specific cellular infiltration differences between the high and low-risk groups, and the relationship between them remains unclear, prompting us to conduct further analyses. By summarizing previously published tumor-related immune cell gene sets gathered by Immuno-Oncology Biological Research [60], we inferred the levels of immune cell infiltration between the high and low-risk groups (Fig. 7E). In the high-risk group of TCGA-GBM, only CD8 T cells and T helper cells showed decreased infiltration levels, while immunosuppressive myeloid-derived macrophages, neutrophils, and regulatory T cells were significantly increased, further confirming the role of the VNRS model in predicting tumorigenesis and immunosuppression.

Subsequently, we validated our conclusions using the pan-cancer metagenes for 28 immune cell subpopulations summarized by Charoentong et al., and found that macrophages and neutrophils had higher infiltration levels in the high-risk group, while the activity of chemokine receptors was also higher (Fig. 7F). Interestingly, the activity of immune checkpoints was also elevated. This led us to explore the differences in the efficacy of immune checkpoint therapy between the high and low-risk groups. Correlation analysis showed a strong association between the expression of model genes and the expression of 46 immune checkpoint genes (Fig. 7G). The high-risk group was associated with upregulated expression of immune checkpoint molecules such as CD276, PDCD1, and LGALS9 (Fig. 7H), suggesting that patients with higher VNRS risk scores might be more sensitive to related ICI therapies.

Robust prognostic assessment and therapeutic insights via the VNRS model

Since the VNRS model is based on the immunosuppressive properties of *VEGFA*+GBMAN, we sought to determine whether it accurately evaluates the negative impact of neutrophils in the samples. Using the timigp method [51], we identified cell types associated with favorable or unfavorable prognosis in both high-risk and low-risk groups. Our findings indicated that neutrophil is a
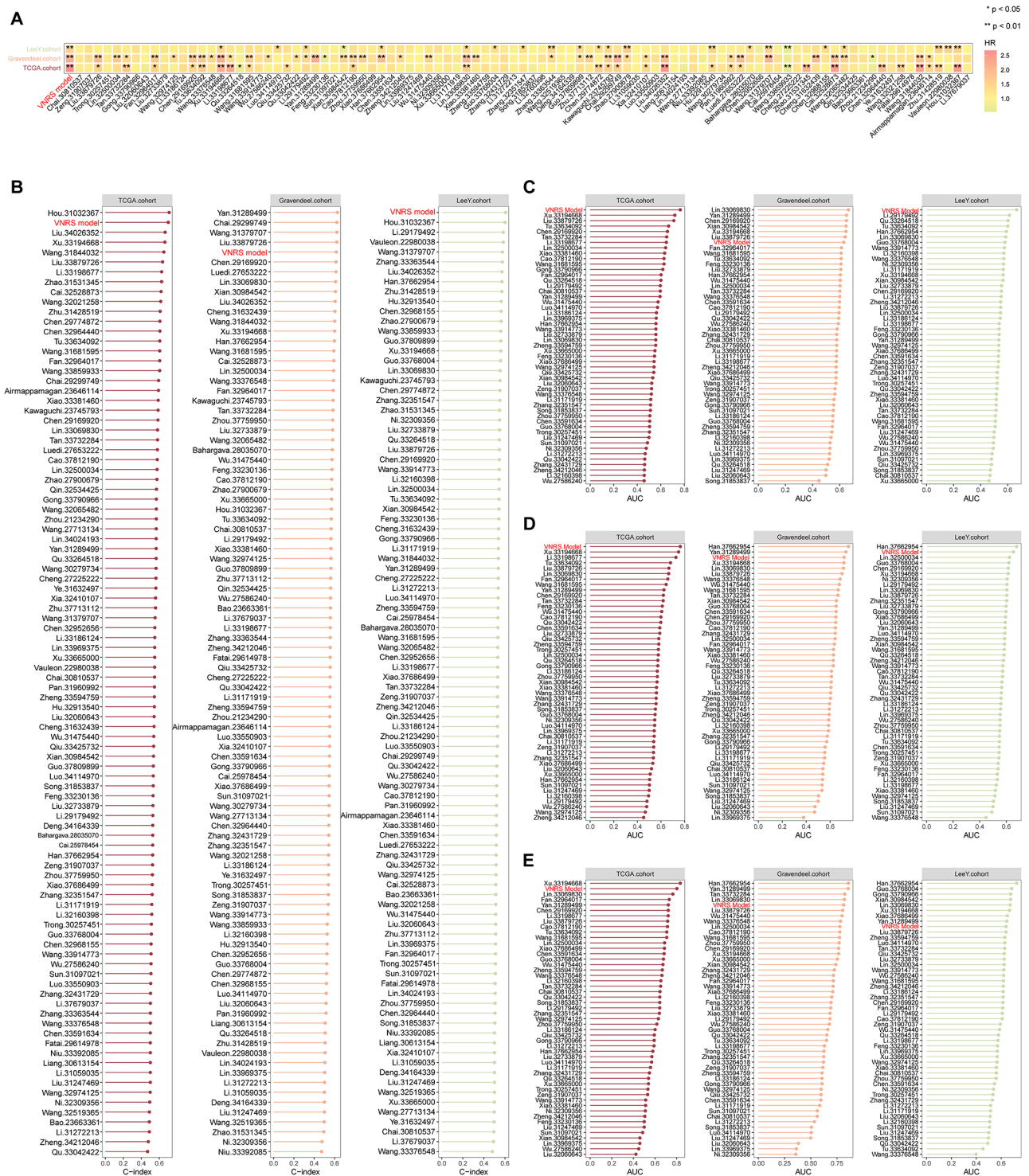
Fig. 6 Validation of the superiority of the VNRS model compared to published predictive models. (**A**) Heatmap comparing the HR of the proposed model with those of previously published models related to LGG and GBM. HR values represent the strength of association between the model's risk scores and patient outcomes, with higher HR values indicating stronger prognostic performance. (**B**) The results of the model's C-index are compared with previously published models related to LGG and GBM. The C-index quantifies the model's predictive accuracy, with higher values (closer to 1) indicating better discrimination between outcomes. (**C-E**) The results of the model's 1/2/3-year AUC values are compared with previously published models related to LGG and GBM
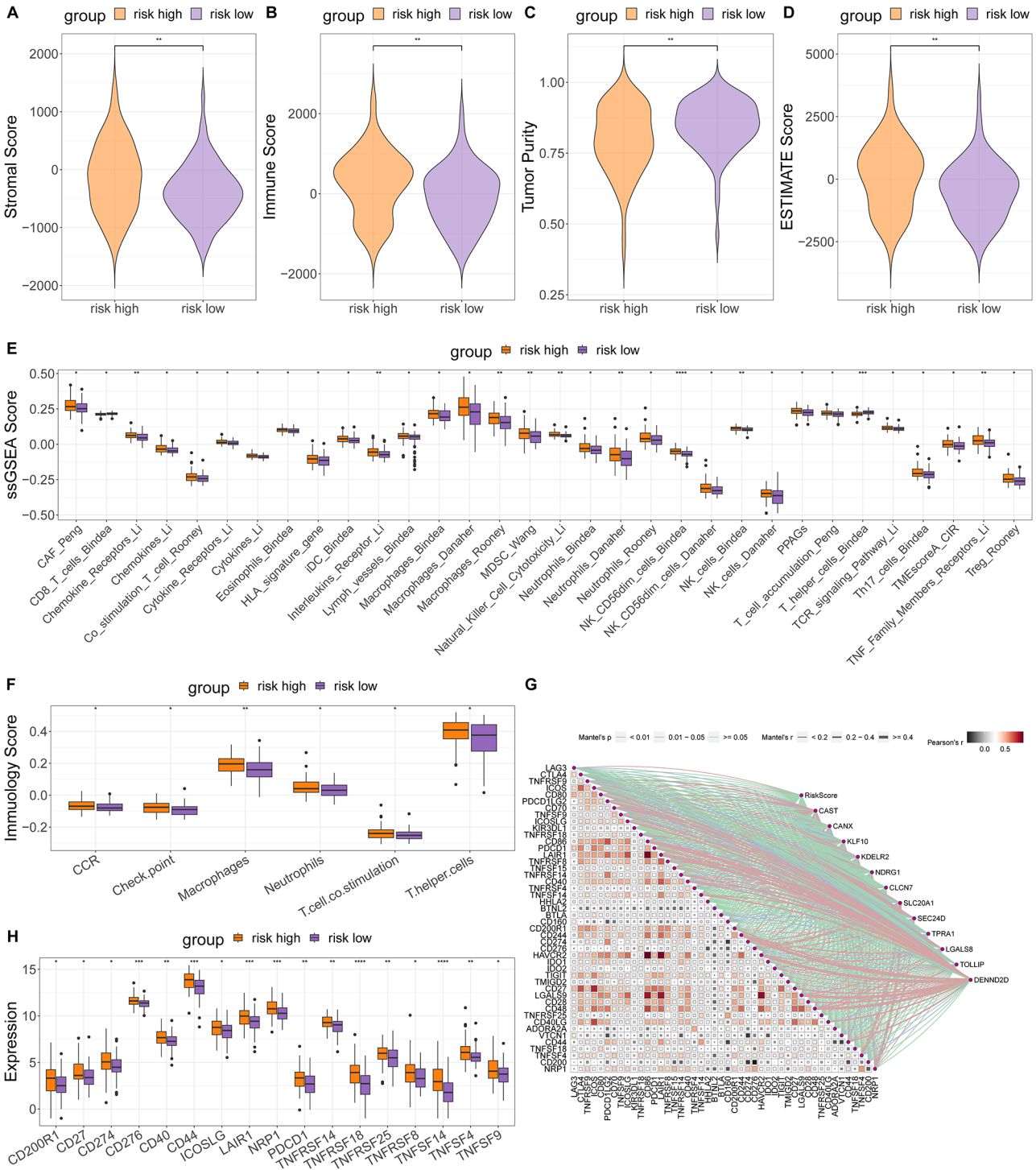
Fig. 7 Immune infiltration and checkpoint activation in high VNRS risk GBM. (**A-D**) Violin plots comparing the Stromal Score, Immune Score, Tumor Purity, and ESTIMATE Score between the high- and low-risk groups. These scores reflect the composition of the TME. (**E**) Boxplot displaying the ssGSEA scores for immune cell gene sets derived from previously published studies. These scores quantify the relative abundance of specific immune cell populations in the high- and low-risk groups, revealing differences in immune cell infiltration and activation. (**F**) Boxplot showing the scores of pan-cancer immune metagenes, which represent consensus immune-related gene signatures across various cancer types. (**G**) Correlation matrix illustrating the associations between the expression levels of ICI genes in the high- and low-risk groups. This analysis identifies potential co-expression patterns and interactions among these genes, which may influence immune evasion or response to immunotherapy. (**H**) Boxplot comparing the expression levels of ICI genes between the high- and low-risk groups. This analysis revealed differences in the expression of key immune checkpoint molecules, providing insights into the immune landscape and potential therapeutic vulnerabilities

significant contributor to poor prognosis in the high-risk group (Fig. 8A). Interestingly, in the low-risk group, neutrophils were identified as the primary beneficial cell type (Fig. 8B). Cell interaction enrichment analysis revealed no beneficial neutrophil-related interactions in the high-risk group (Fig. 8C). However, in the low-risk group, neutrophils demonstrated beneficial interactions with Th1 cells, T$\gamma\delta$ cells, and CD4 memory T cells (Fig. 8D). These findings suggested that the VNRS model effectively assesses neutrophil immune function in GBM patients. We next examined metabolic pattern differences between the high- and low-risk groups (Fig. 8E). In the high-risk group, gluconeogenesis, glycolysis, and the pentose phosphate pathway were significantly upregulated, providing energy for cancer cell proliferation and promoting tumor growth and survival. These glucose metabolism pathways also supplied alternative sugar sources to meet the high energy and biosynthetic demands of tumor cells. Further analysis indicated significant upregulation of pathways related to structural molecule degradation in the high-risk group, likely reflecting enhanced extracellular matrix remodeling, which may contribute to tumor invasion and metastasis. Enzyme-mediated drug metabolism pathways were also upregulated in this group, suggesting stronger detoxification capacities and drug resistance, which may reduce treatment efficacy. Additionally, we observed pronounced hypoxic characteristics in the high-risk group, accompanied by increased production and secretion of extracellular vesicles and exosomes (Fig. 8F). This suggested enhanced intercellular communication under hypoxic conditions, microenvironment remodeling, and augmented immune evasion by tumor cells. Exosomes carrying drug-resistant molecules further enhanced the resistance of tumor cells to treatment.

Furthermore, we performed a drug sensitivity analysis using the oncoPredict platform with 198 drugs from the GDSC2 dataset. The results demonstrated that Ribociclib, PF.4,708,671, WIKI4, ZM447439, and WZ4003 exhibited significantly enhanced sensitivity in the high-risk group compared to other agents, suggesting their potential as targeted therapeutics (Fig. 8G) (Fig. S4A). Furthermore, correlational analysis revealed a statistically significant inverse relationship ($p < 0.05$) between the half maximal inhibitory concentration (IC50) values of these compounds and VNRS risk scores - the lower IC50 values strongly correlated with elevated risk scores (Fig. S4B). This negative association further corroborated the heightened therapeutic susceptibility of high-risk group patients to these pharmacological agents. At the same time, this also indicated the strong resistance of the high-risk group to chemotherapy.

The validation of VNRS model genes expression at RNA and protein levels

Finally, we validated the differential expression of genes in GBM and low grade glioma (LGG) using TCGA and GTEx datasets. The results revealed that CAST, CANX, KLF10, KDELR2, NDRG1, CLCN7, SLC20A1, TPRA1, SEC20D and DENND2D were highly expressed in tumor samples, while LGALS8 and TOLLIP were expressed at lower levels in tumor samples (Fig. 9A-H) (Fig. S4C-F). Immunohistochemistry of high grade glioma (HGG) and LGG samples was performed from the Human Protein Atlas (HPA), the protein expression data of these genes confirmed the above results (Fig. 9I-L) (Fig. S5A-G).

Discussion

Neutrophil functions and phenotypes in cancer are highly complex and diverse. Increasing evidence highlights the pivotal role of neutrophils in the pathogenesis of GBM, influencing tumor initiation, local progression, and metastasis [61]. Further investigation into neutrophil diversity within the TME can provide valuable insights into TANs in GBM. Targeting neutrophil recruitment, reprogramming, depletion, neutrophil extracellular trap inhibition, neutrophil engineering-based drug development, and enhancing drug efficacy hold significant therapeutic potential for GBM. Advances in single-cell analysis may identify molecular markers on neutrophils, revealing their heterogeneity in circulation and the TME [34]. However, studies specifically exploring neutrophil heterogeneity in GBM remain limited. Our research aimed to address this gap, with the potential to enhance immunotherapy outcomes in GBM patients.

Our study revealed transcriptomic differences among neutrophils, classifying GBMAN into four subgroups marked by MNDA, LGALS1, ADM, and VEGFA. MNDA, a cell surface protein, is expressed in normal and abnormal hematopoietic cells and plays a crucial role in normal hematopoiesis. Its stable expression supports hematopoietic stem cell self-renewal and secretion, regulating immune cell activity and promoting interactions between immune cells and tumor cells through receptors such as PD-1. LGALS1, a glycan-binding protein, modulates intracellular and extracellular processes, promoting cell survival. Previous studie indicated that knocking down LGALS1 can inhibit M2 macrophages and MDSCs within the GBM microenvironment [62]. ADM, a peptide hormone, acts as a local paracrine and autocrine mediator, facilitating angiogenesis, cell proliferation, and anti-inflammatory activities. VEGFA is a critical biomarker widely used in diagnosing and treating tumors, cardiovascular diseases, and other conditions. By comparing the differentiation potential of each subgroup, we propose that neutrophils initially support pro-inflammatory activities, a function later co-opted by tumor-secreted
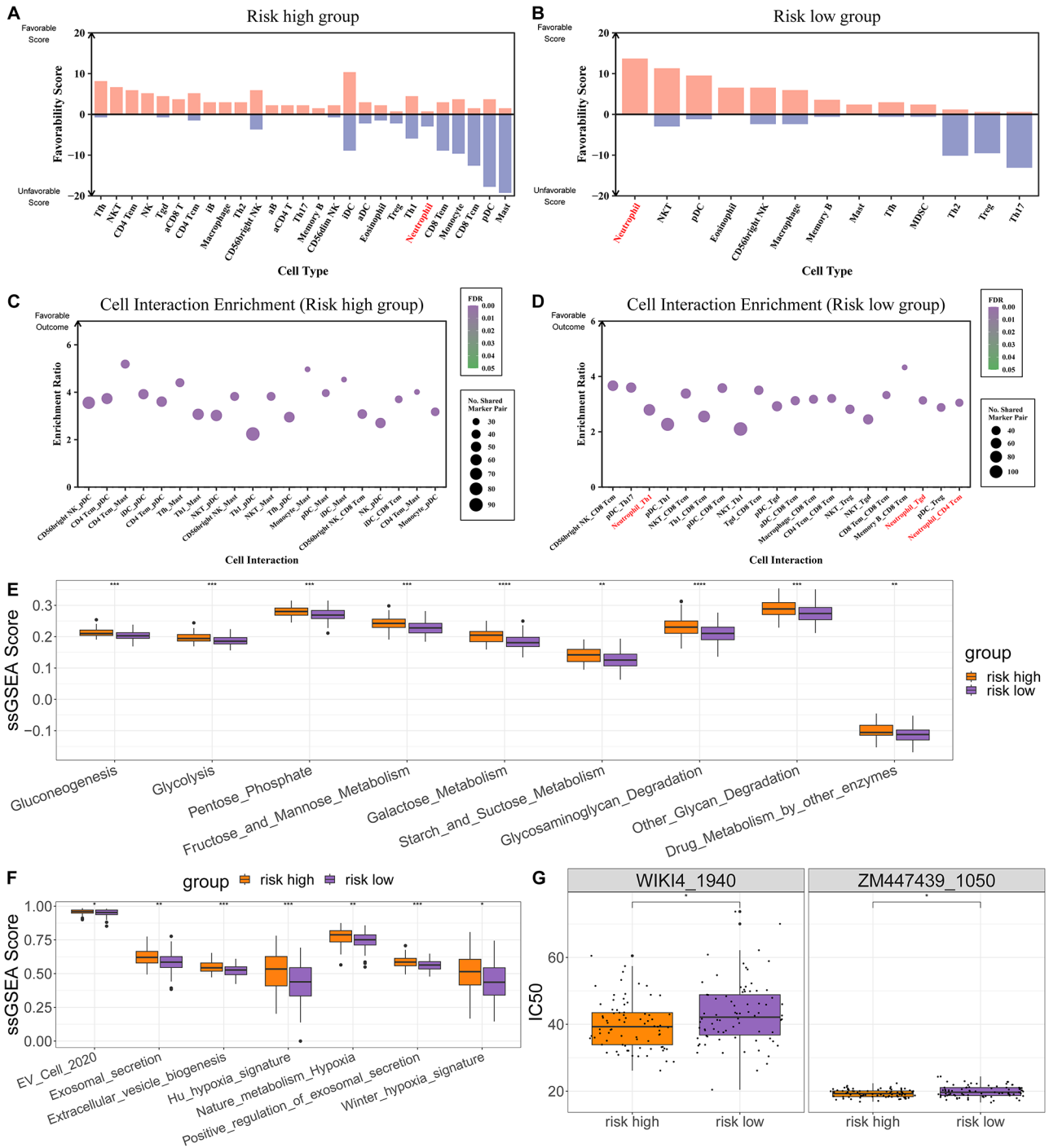
Fig. 8 Robust prognostic assessment and therapeutic insights via the VNRS model. (**A-B**) The barplot illustrates the favorable prognostic factors for the high-risk and low-risk score groups. These factors highlight key immune cells associated with better clinical outcomes in each group. (**C-D**) The dot plot displays the cell-type pairs that are beneficial for prognosis based on cell-cell interaction analysis in the high-risk and low-risk score groups. (**E**) Boxplot showing the metabolic differences between the high-risk and low-risk score groups. This analysis identifies key metabolic pathways or metabolites that are dysregulated in each group, providing insights into the metabolic reprogramming associated with risk stratification. (**F**) Boxplot comparing the expression levels of tumor-associated gene signatures between the high-risk and low-risk score groups. These signatures represent key biological processes or pathways that are differentially activated in each group, offering potential mechanistic explanations for their prognostic differences. (**G**) OncoPredict drug sensitivity analysis results show the IC50 of drugs (WIKI4 and ZM447439) between the high-risk score group and the low-risk score group
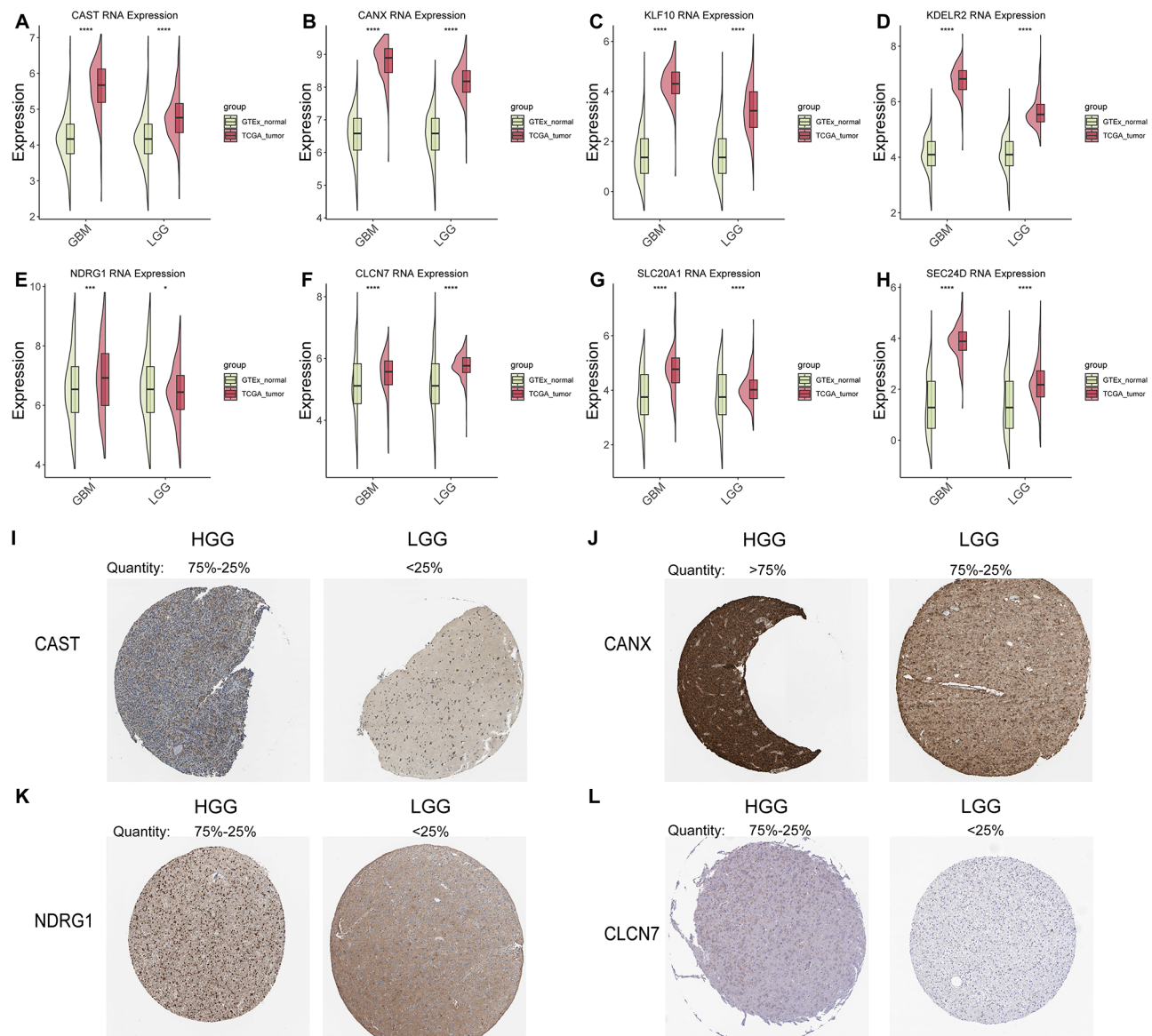
Fig. 9 The validation of VNRS model genes expression at RNA and protein levels. (**A-H**) Differences in RNA levels of risk model genes between tumor samples and normal samples in the TCGA and GTEx databases. (**I-L**) Differences in protein levels of VNRS model genes between HGG samples and LGG samples

cytokines to promote abnormal tumor growth. This transition likely occurs during the *LGALS1*+GBMAN and *ADM*+GBMAN stages, marking a critical point of neutrophil reprogramming. Moreover, a study on neutrophils in pancreatic cancer revealed that high VEGFA expression at stage T3 significantly enhances intratumoral angiogenesis, correlating with poor prognosis [12]. Our research identifies the development of *MNDA*+GBMAN into *VEGFA*+GBMAN, which is associated with increased hypoxia and glycolysis. *VEGFA*+GBMAN exhibits the lowest inflammatory response and the strongest interaction with stromal and tumor cells, suggesting a role in guiding angiogenesis and providing essential nutrients and oxygen to stromal and tumor cells, particularly

within glycolytic-hypoxic niches. Previous studies have linked human cancer resistance to anti-angiogenic therapies with neutrophil infiltration, and neutrophil depletion has been shown to reduce tumor angiogenesis and growth [13]. Glycolytic by-products promote tumor immune evasion and growth, with glycolytic activity in TANs significantly increased. Our research further explores and identifies *VEGFA*+GBMAN in GBM, highlighting its role in inhibiting anti-tumor immunity, promoting angiogenesis, and enhancing tumor cell survival, invasiveness, and metastasis through extracellular matrix degradation.

Each tumor type's unique biological and genetic characteristics may influence biomarker expression. The low

diagnostic rate in the early stages and high resistance to treatment present significant challenges for GBM [63]. Despite advances in molecular biology and immunology, large-scale studies applying machine learning models to predict patient prognosis and drug efficacy based on GBMAN remain limited. To address this, we employed univariate Cox regression analysis and machine learning-based integration procedures to identify prognostic genes from *VEGFA*+GBMAN. A prognostic model was constructed using 12 key genes. Among these, CAST inhibits calpain, a protein that regulates the WNT/β-catenin pathway [64]. Overexpression of KLF10 reduces extracellular matrix degradation, supports cell proliferation, and may extend neutrophil lifespan [65]. Abnormal KDELR2 expression predicts poor prognosis in breast cancer, with the HDAC3-KDELR2 axis accelerating cancer progression [66]. NDRG1 enhances VEGFA-induced corneal angiogenesis [67], while SLC20A1, a phosphate transporter protein, regulates apoptosis and insulin signaling [68]. These findings highlight the prognostic potential of VNRS model genes in enhancing GBM immunity. While this investigation provides valuable insights, several limitations warrant consideration. First, the study's exclusive reliance on RNA expression data constrained our ability to assess critical biological dimensions including genetic variants and post-transcriptional regulatory mechanisms. Second, though enhanced by multifaceted algorithmic validation and robust cohort analyses, the intrinsic heterogeneity of TME - where dynamic cellular interactions and spatial biology modulate immune responses - presents persistent interpretative challenges. These methodological safeguards strengthen confidence in our key findings, yet future work integrating multi-omics approaches will be essential to address these constraints systematically. Future refinement of the model incorporating these factors could improve accuracy. Despite these limitations, our results confirm the feasibility of the methods used and their potential for clinical application. The neutrophil landscape outlined in this study lays the foundation for future research.

Our findings suggested that recruited neutrophils, regardless of their initial phenotype, adapt to the tumor environment, converging toward a unified evolutionary trajectory. This process ensures that long-lived, pro-angiogenic *VEGFA*+GBMAN continuously supply resources to support tumor growth. Our research revealed how neutrophils, despite being short-lived effector cells, adjust their functions to meet the demands of the tissue environment. Through extensive bioinformatics analysis and machine learning algorithms, we developed a robust signature for the "*VEGFA*+neutrophil-related signature" in GBM patients. The VNRS model shows promise as a tool for optimizing treatment strategies and monitoring plans for GBM patients. This study provides a new perspective on neutrophil roles in GBM and proposes new strategies for treatment and prognosis. Future research may explore therapeutic interventions targeting specific neutrophil subtypes in GBM.

Conclusion

In summary, this study utilized large-scale scRNA-seq to identify different GBMAN subpopulations, analyzing their developmental trajectories, communication patterns, and regulatory networks. It particularly focused on the role of *VEGFA*+GBMAN in promoting an immunosuppressive TME, and mapped out the blueprint of GBMANs. The VNRS model was subsequently developed using machine learning techniques, resulting in a novel prognostic model. This model demonstrates comparatively favorable predictive performance within existing glioma risk models. The VNRS not only serves as an independent prognostic factor for OS but also proves valuable in assessing immunotherapy response and chemotherapy efficacy in GBM patients. Our findings provide critical insights into the role of neutrophils within the GBM microenvironment and highlight the previously underappreciated importance of GBMAN. We uncovered significant transcriptional differences between these subpopulations, indicating their potential functional diversity within the TME. The VNRS model represents a promising tool for guiding personalized treatment and improving patient outcomes in GBM, while also paving the way for new immunotherapy strategies.

Abbreviations

GBM	glioblastoma
IDH	isocitrate dehydrogenase
scRNA-seq	single-cell RNA sequencing
UMAP	uniform manifold approximation and projection
PCA	principal component analysis
TME	tumor microenvironment
Mg_TAM	microglia-derived macrophages
Mo_TAM	monocyte-derived macrophages
TAN	tumor associated neutrophil
CAF	cancer-associated fibroblast
MDSC	myeloid-derived suppressor cell
ES	enrichment score
GBMAN	glioblastoma associated neutrophil
GO	gene ontology
log2FC	log2 fold change
BP	biological processes
MF	molecular functions
CC	cellular components
KEGG	kyoto encyclopedia of genes and genomes
BBKNN	batch balanced k-nearest neighbor
SCENIC	single-cell regulatory network inference and clustering
GSEA	gene set enrichment analysis
ssGSEA	single-sample gene set enrichment analysis
GDSC	genomics of drug sensitivity in cancer
VNRS	VEGFA+neutrophil-related signature
OS	overall survival
HR	hazard ratio
AUC	area under the curve
ICI	immune checkpoint inhibitor
IC50	half maximal inhibitory concentration
ROGUE	Ratio of Global Unshifted Entropy

Yang *et al. Biology Direct*   (2025) 20:45

Page 20 of 22

C-index   consistency index

Supplementary Information

Acknowledgements

Author contributions

Funding

Data availability
The data used to support the findings of this study are available either online or from the corresponding author upon request.

Declarations

Consent for publication
All the authors give the consent for the publication of identifiable details, which can include the text, figures and other materials in this manuscript.

Competing interests
The authors declare no competing interests.

Author details
[1]Clinical Pharmacy Center, Nanfang Hospital, Southern Medical University, Guangzhou, Guangdong 510515, China
[2]Department of Pharmacy, Zhujiang Hospital, Southern Medical University, Guangzhou, Guangdong 510282, China
[3]National Medical Products Administration Key Laboratory for Research and Evaluation of Drug Metabolism & Guangdong Provincial Key Laboratory of New Drug Screening & Guangdong-Hongkong- Macao, Joint Laboratory for New Drug Screening, School of Pharmaceutical Sciences, Southern Medical University, Guangzhou, Guangdong 510515, China
[4]Department of Pathology, Zhujiang Hospital, Southern Medical University, Guangzhou, Guangdong 510282, China
[5]Department of Orthopedic Surgery, Nanfang Hospital, Southern Medical University, Guangzhou, Guangdong 510515, China
[6]Departments of Pediatrics, Weill Cornell Medicine, New York, NY, USA
[7]Department of Neurosurgery, Nanfang Hospital, Southern Medical University, Guangzhou, Guangdong 510515, China

References
1. Aaron C, Tan DM, Ashley, Giselle Y, López M, Malinzak HS. Friedman, and Mustafa Khasraw. Management of glioblastoma: state of the Art and future directions. Cancer J Clin. 2020;70(4):299–312.
2. Dapash M, Hou D, Castro B, Lee-Chang C, Maciej S. Lesniak Interplay between Glioblastoma its Microenvironment Cells. 2021;10(9):2257.
3. Xie X, Shi Q, Wu P, Zhang X, Kambara H, Su J, Yu H, Park S-Y, Guo R, Ren Q, Zhang S, Xu Y, Silberstein LE, Cheng T, Ma F, Li C, Hongbo R. Luo. Single-cell transcriptome profiling reveals neutrophil heterogeneity in homeostasis and infection. Nat Immunol. 2020;21(9):1119–33.
4. Gabriel Alzial O, Renoult François, Paris C, Gratas A, Clavreul, Pecqueur C. Wild-type isocitrate dehydrogenase under the spotlight in glioblastoma. Oncogene. 2022;41(5):613–21.
5. Pravesh Gupta M, Dang S, Oberai S, Migliozzi R, Trivedi G, Kumar M, Peshoff N, Milam A, Ahmed K, Bojja, Tuan M, Tran J, Gumin CK-M, Huse J, Cox K, Li J, Shehwana H, Sameer A, Sheth R, Saxon. Sun Baohua, Brittany Parker Kerrigan, Atul Maheshwari, Edwin Roger Parra Cuentas, Nicholas E Navin, amy B Heimberger, Frederick F Lang, Antonio Iavarone, Karen Clise-Dwyer, Linghua Wang, and Krishna P Bhat. Immune landscape of isocitrate dehydrogenase-stratified primary and recurrent human gliomas. Neurooncology. 2024;26(12):2239–55.
6. Jianbo Wang Y, Jia N, Wang X, Zhang B, Tan G, Zhang, Cheng Y. The clinical significance of tumor-infiltrating neutrophils and neutrophil-to-CD8 + lymphocyte ratio in patients with resectable esophageal squamous cell carcinoma. J Translational Med. 2014;12(1):7.
7. Janesh Pillay Iden, Braber N, Vrisekoop LM, Kwast, Rob J, de Boer JoséAM, Borghans. Kiki Tesselaar, and Leo Koenderman. In vivo labeling with 2H2O reveals a human neutrophil lifespan of 5.4 days. Blood. 2010;116(4):625–7.
8. Florian Klemm RR, Maas RL, Bowman M, Kornete K, Soukup S, Nassiri J-P, Brouland CA, Iacobuzio-Donahue C, Brennan V, Tabar PH, Gutin, Roy T, Daniel ME. Hegi, and Johanna A. Joyce. Interrogation of the microenvironmental landscape in brain tumors reveals Disease-Specific alterations of immune cells. Cell. 2020;181(7):1643–60e17.
9. Ekaterina Friebel K, Kapolou S, Unger NicolásG, Núñez S, Utz EJ, Rushing L, Regli M, Weller M, Greter. Sonia Tugues, marian Christoph Neidert, and Burkhard becher. Single-cell mapping of human brain cancer reveals tumor-specific instruction of tissue-invading leukocytes. Cell. 2020;181(7):1626–e164220.
10. Ji Liang Y, Piao L, Holmes GN, Fuller V, Henry N, Tiao, De John F. Groot. Neutrophils promote the malignant glioma phenotype through S100A4. Clin Cancer Res. 2014;20(1):187–198.
11. Prerna Magod I, Mastandrea L, Rousso-Noori L, Agemy G, Shapira. Noam Shomron, and Dinorah Friedmann-Morvinski. Exploring the longitudinal glioma microenvironment landscape uncovers reprogrammed pro-tumorigenic neutrophils in the bone marrow. Cell Rep. 2021;36(5):109480.
12. Melissa SF, Ng I, Kwok L, Tan C, Shi D, Cerezo-Wallis Y, Tan K, Leong GF, Calvo K, Yang Y, Zhang J, Jin KH, Liong D, Wu R, He D, Liu YC, Teh M, Li J, Chen Y, Liu L, Liu. Jingjing Qi, Yingbin Liu, Lingxi Jiang, Baiyong Shen, Hui Cheng, Tao Cheng, Veronique Angeli, Ankur Sharma, Yuin-han Loh, Hong Liang Tey, Shu Zhen Chong, Matteo Iannacone, Renato Ostuni, Andrés Hidalgo, Florent Ginhoux, and Lai Guan Ng. Deterministic reprogramming of neutrophils within tumors. Science. 2024;383(6679):eadf6493.
13. Merav E, Shaul, Zvi G, Fridlender. Tumour-associated neutrophils in patients with cancer. Nat Rev Clin Oncol. 2019;16(10):601–20.
14. Chu X, Zhang Y, Cheng S. Heterogeneity of tumor-infiltrating myeloid cells in era of single-cell genomics. Chin J Cancer Res = Chung-Kuo Yen Cheng Yen Chiu. 2022;34(6):543–53.
15. Shen H-Y, Xu J-L, Zhu Z, Xu H-P, Liang M-X, Xu D, Chen W-Q, Tang J-H, Zheng Fang, and, Zhang J. Integration of bioinformatics and machine learning strategies identifies APM-related gene signatures to predict clinical outcomes and therapeutic responses for breast cancer patients. Volume 45. New York, N.Y.): Neoplasia; 2023. p. 100942.

16. Laura M, Richards, Owen KN, Whitley G, MacLeod, Florence MG, Cavalli FJ, Coutinho JE, Jaramillo N, Svergun M, Riverin DC, Croucher M, Kushida K, Yu P, Guilhamon N, Rastegar M, Ahmadi JK, Bhatti DA, Bozek N, Li L, Lee C, Che E, Luis NI, Park Z, Xu T, Ketela RA, Moore MA, Marra J, Spears MD, Cusimano PB, Dirks GD. Bader, and Trevor J. Pugh. Gradient of Developmental and Injury Response transcriptional states defines functional vulnerabilities underpinning glioblastoma heterogeneity. Nature Cancer. 2021;2(2):157–173.

17. Kevin C, Johnson KJ, Anderson ET, Courtois AD, Gujar FP, Barthel, Frederick S, Varn D, Luo M, Seignon E, Yi H, Kim MRH, Estecio D, Zhao M, Tang NE, Navin PC, De Hamer W, Bulsara K, Samuels ML. Sunit Das, Paul Robson, and Roel G. W. Verhaak. Single-cell multimodal glioma analyses identify epigenetic regulators of cellular plasticity and environmental stress response. Nat Genet. 2021;53(10):1456–1468.

18. Lin Wang H, Babikir Sören, Müller G, Yagnik K, Shamardani F, Catalan G, Kohanbash B, Alvarado ED, Lullo A, Kriegstein S, Shah H, Wadhwa SM, Chang JJ, Phillips MK. Aghi, and Aaron A. Diaz. The phenotypes of proliferating glioblastoma cells reside on a single axis of variation. Cancer Discov. 2019;9(12):1708–19.

19. Andrew X, Chen RD, Gartrell J, Zhao PS, Upadhyayula W, Zhao J, Yuan HE, Minns A, Dovas JN, Bruce A, Lasorella A, Iavarone P, Canoll PA. Sims, and Raul Rabadan. Single-cell characterization of macrophages in glioblastoma reveals MARCO as a mesenchymal pro-tumor marker. Genome Med. 2021;13(1):88.

20. Xie Y, He L, Lugano R, Zhang Y, Cao H, He Q, Chao M, Liu B, Cao Q, Wang J, Yang J, Yaqin H, Han L, Zhang Y, Huang H, Uhrbom L, Betsholtz C, Wang L. Anna Dimberg, and Lei Zhang. Key molecular alterations in endothelial cells in human glioblastoma uncovered through single-cell RNA sequencing. JCI Insight. 2021;6(15):e150861.

21. Véronique G, LeBlanc DL, Trinh S, Aslanpour M, Hughes D, Livingstone D, Jin BY, Ahn MD, Blough JG, Cairncross JA, Chan JJP, Kelly, Marco A. Marra. Single-cell landscapes of primary glioblastomas and matched explants and cell lines show variable retention of inter- and intratumor heterogeneity. Cancer Cell. 2022;40(4):379–92e9.

22. Nourhan Abdelfattah P, Kumar C, Wang J-S, Leu WF, Flynn R, Gao DS, Baskin K, Pichumani OB, Ijare SL, Wood, Suzanne Z, Powell DL, Haviland, Brittany C, Parker Kerrigan FF, Lang SS, Prabhu KM, Huntoon W, Jiang, Betty YS, Kim. Joshy George, and Kyuson Yun. Single-cell analysis of human glioma and immune cells identifies S100A4 as an immunotherapy target. Nat Commun. 2022;13(1):767.

23. Saritha Krishna A, Choudhury MB, Keough K, Seo L, Ni S, Kakaizada A, Lee A, Aabedi G, Popova B, Lipkin C, Cao CN, Gonzales R, Sudharshan A, Eglady-ous N, Almeida Y, Zhang AM, Molinaro HS, Venkatesh, Andy GS, Daniel K, Shamardani J, Hyer EF, Chang A, Findlay JJ, Phillips S, Nagarajan DR. Raleigh, David Brang, Michelle Monje, and Shawn L. Hervey-Jumper. Glioblastoma remodelling of human neural circuits decreases survival. Nature. 2023;617(7961):599–607.

24. Mei Y, Wang X, Zhang J, Liu D, He J, Huang C, Liao J, Wang Y, Feng Y, Li H, Liu X, Chen L, Yi W, Chen X, Bai H-M, Wang X, Li Y, Wang L, Liang Z, Xianwen Ren, Li Qiu, Yuan Hui, Qingling Zhang, Qibin Leng, Jun Chen, and, Jia G. Siglec-9 acts as an immune-checkpoint molecule on macrophages in glioblastoma, restricting T-cell priming and immunotherapy response. Nature Cancer. 2023.

25. GTEx Consortium. The Genotype-Tissue expression (GTEx) project. Nat Genet. 2013;45(6):580–5.

26. Katarzyna, Tomczak. Patrycja Czerwińska, and Maciej Wiznerowicz. The cancer genome atlas (TCGA): an immeasurable source of knowledge. Contemp Oncol (Poznan Poland). 2015;19(1A):A68–77.

27. Lee Y, Scheck AC, Timothy F, Cloughesy A, Lai J, Dong HK, Farooqi LM, Liau S, Horvath PS, Mischel, Stanley F. Nelson. Gene expression analysis of glioblastomas identifies the major molecular basis for the prognostic benefit of younger age. BMC Med Genom. 2008;1:52.

28. Lonneke AM, Gravendeel MCM, Kouwenhoven O, Gevaert JJ, de Rooi AP, Stubbs JE, Duijm A, Daemen FE, Bleeker, Linda BC, Bralten NK, De Kloosterhof PHC, Eilers PJ, van der Spek JM, Kros, Peter AE. Sillevis Smitt, Martin J. van den Bent, and Pim J. French. Intrinsic gene expression profiles of gliomas are a better predictor of survival than histology. Cancer Res. 2009;69(23):9065–9072.

29. Wolf FA, Angerer P, Fabian J. Theis. SCANPY: Large-scale single-cell gene expression data analysis. Genome Biol. 2018;19(1):1–5.

30. Krzysztof Polański MD, Young Z, Miao KB, Meyer SA. Teichmann, and Jong-Eun park. BBKNN: fast batch alignment of single cell transcriptomes. Bioinf (Oxford England). 2020;36(3):964–5.

31. Tang F, Li J, Qi L, Liu D, Bo Y, Qin S, Miao Y, Yu K, Hou W, Li J, Peng J, Tian Z. Linnan Zhu, Hui Peng, Dongfang Wang, and Zemin Zhang. A pan-cancer single-cell panorama of human natural killer cells. Cell. 2023;186(19):4235–e425120.

32. Etienne Becht L, McInnes J, Healy C-A, Dutertre, Immanuel WH, Kwok. Lai Guan Ng, florent Ginhoux, and Evan W. Newell. Dimensionality reduction for visualizing single-cell data using UMAP. Nat Biotechnol. 2019;37(1):38–44.

33. Traag VA, Waltman L. Eck. From Louvain to Leiden: guaranteeing well-connected communities. Sci Rep. 2019;9(1):5233. van.

34. Xue R, Zhang Q, Cao Q, Kong R, Xiang X, Liu H, Feng M, Wang F, Cheng J, Li Z, Zhan Q, Deng M, Zhu J. Zemin Zhang, and Ning Zhang. Liver tumour immune microenvironment subtypes and neutrophil heterogeneity. Nature. 2022;612(7938):141–7.

35. Yuhan Hao S, Hao E, Andersen-Nissen WM, Mauck S, Zheng A, Butler MJ, Lee AJ, Wilk C, Darby M, Zager P, Hoffman M, Stoeckius E, Papalexi EP, Mimitou J, Jain A, Srivastava T, Stuart LM, Fleming B, Yeung AJ, Rogers JM, McElrath CA. Blish, Raphael Gottardo, Peter Smibert, and Rahul Satija. Integrated analysis of multimodal single-cell data. Cell. 2021;184(13):3573–e358729.

36. Yufan Yang Z, Liu Y, Wei S, He A, Gu Z, Li J, Li Z, Xu, Cen B. Single-cell multi-omics analysis reveals candidate therapeutic drugs and key transcription factor specifically for the mesenchymal subtype of glioblastoma. Cell Bioscience. 2024;14(1):151.

37. Baolin Liu C, Li Z, Li D, Wang X, Ren, Zhang Z. An entropy-based metric for assessing the purity of single cell populations. Nat Commun. 2020;11(1):3155.

38. Minji Kang JJA, Armenteros GS, Gulati R, Gleyzer S, Avagyan EL, Brown W, Zhang A, Usmani N, Earland Z, Wu J, Zou RC, Fields DY, Chen AA. Chaudhuri, and Aaron M. Newman. Mapping single-cell developmental potential in health and disease with interpretable deep learning. 2024.

39. Xiaojie Qiu A, Hill J, Packer D, Lin Y-A, Ma, Trapnell C. Single-cell mRNA quantification and differential analysis with census. Nat Methods. 2017;14(3):309–15.

40. Cui A, Huang T, Li S, Ma A, Pérez JL, Sander C, Keskin DB, Wu CJ. Ernest Fraenkel, and Nir Hacohen. Dictionary of immune responses to cytokines at single-cell resolution. Nature. 2024;625(7994):377–84.

41. Jin S, Christian F, Guerrero-Juarez L, Zhang I, Chang R, Ramos C-H, Kuan P, Myung MV. Plikus, and Qing Nie. Inference and analysis of cell-cell communication using cellchat. Nat Commun. 2021;12(1):1088.

42. Cheng J, Zhang J, Wu Z, Sun X. Inferring microenvironmental regulation of gene expression from single-cell RNA sequencing data using ScMLnet with an application to COVID-19. Brief Bioinform. 2021;22(2):988–1005.

43. Sara Aibar CB, González-Blas T, Moerman H, Imrichova G, Hulselmans F, Rambow J-C, Marine. Pierre Geurts, Jan Aerts, Joost van den Oord, Zeynep Kalender Atak, Jasper Wouters, and Stein Aerts. SCENIC: Single-cell regulatory network inference and clustering. Nat Methods. 2017;14(11):1083–1086.

44. Van de Bram C, Flerin K, Davie MD, Waegeneer G, Hulselmans S, Aibar R, Seurinck W, Saelens R, Cannoodt Q, Rouchon. Toni Verbeiren, dries de Maeyer, joke reumers, Yvan Saeys, and Stein aerts. A scalable SCENIC workflow for single-cell gene regulatory network analysis. Nat Protoc. 2020;15(7):2247–76.

45. Gábor Csárdi. and T. Nepusz. The igraph software package for complex network research. 2006.

46. Tianzhi Wu E, Hu S, Xu M, Chen P, Guo Z, Dai T, Feng L, Zhou W, Tang. Li Zhan, Xiaocong Fu, Shanshan Liu, Xiaochen Bo, and Guangchuang Yu. cluster-Profiler 4.0: A universal enrichment tool for interpreting omics data. Innov. 2021;2(3):100141.

47. Hongwei Liu W, Zhang Y, Zhang AA, Adegboro DO, Fasoranti L, Dai Z, Pan H, Liu Y, Xiong W, Li K, Peng. Siyi Wanggou, and Xuejun li. Mime: A flexible machine-learning framework to construct and visualize models for clinical characteristics prediction and feature selection. Comput Struct Biotechnol J. 2024;23:2798–810.

48. Terry M, Therneau, Patricia M. Grambsch. model.ng survival data: extending the Cox model. Statistics for biology and health. New York, NY: Springer New York; 2000.

49. Kosuke Yoshihara M, Shahmoradgoli E, Martínez R, Vegesna H, Kim W, Torres-Garcia V, Treviño H, Shen PW, Laird, Douglas A, Levine SL, Carter GB. Mills, and Roel G. W. Verhaak. Inferring tumour purity and stromal and immune cell admixture from expression data. Nature Commun. 2013;4(1):2612.

50. Sonja Hänzelmann R, Castelo, Guinney J. Gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics. 2013;14:7.

51. Li C, Zhang B, Schaafsma E, Reuben A, Wang L, Turk MJ, Zhang J, Cheng C. TimiGP: inferring cell-cell interactions and prognostic associations in the tumor immune microenvironment through gene pairs. Cell Rep Med. 2023;4(7):101121.

52. Pornpimol Charoentong F, Finotello M, Angelova C, Mayer M, Efremova D, Rieder H, Hackl, Trajanoski Z. Pan-cancer Immunogenomic analyses reveal

Genotype-Immunophenotype relationships and predictors of response to checkpoint Blockade. Cell Rep. 2017;18(1):248–62.

53. Wanjuan Yang J, Soares P, Greninger EJ, Edelman H, Lightfoot S, Forbes N, Bindal D, Beare JA, Smith IR, Thompson S, Ramaswamy P, Andrew Futreal DA, Haber MR, Stratton. Cyril Benes, Ultan McDermott, and Mathew J. Garnett. Genomics of drug sensitivity in cancer (GDSC): A resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res. 2013;41(Database issue):D955–961.

54. Danielle Maeser RF, Gruener. OncoPredict: an R package for predicting in vivo or cancer patient drug response and biomarkers from cell line screening data. Brief Bioinform. 2021;22(6):bbab260.

55. Antuamwine BB, Bosnjakovic R, Hofmann-Vega F, Wang X, Theodosiou T. Ioannis Iliopoulos, and Sven Brandau. N1 versus N2 and PMN-MDSC: A critical appraisal of current concepts on tumor-associated neutrophils and new directions for human oncology. Immunol Rev. 2023;314(1):250–79.

56. Laleh Haghverdi F, Buettner, Fabian J. Theis. Diffusion maps for high-dimensional single-cell analysis of differentiation data. Bioinf (Oxford England). 2015;31(18):2989–98.

57. Coifman RR, Lafon S, Lee AB, Maggioni M, Nadler B, Warner F. Zucker. Geometric diffusions as a tool for harmonic analysis and structure definition of data: diffusion maps. Proc Natl Acad Sci USA. 2005;102(21):7426–31.

58. Zvi G, Fridlender J, Kim SS, Kapoor V, Cheng G, Ling L, Scott Worthen G, Steven M. Albelda. Polarization of tumor-associated neutrophil phenotype by TGF-beta: N1 versus N2 TAN. Cancer Cell. 2009;16(3):183–94.

59. Ronan Chaligne F, Gaiti D, Silverbush JS, Schiffman HR, Weisman L, Kluegel S, Gritsch SD, Deochand LNG, Castro AR, Richman ML, Suvà, Dan A. Landau. Epigenetic encoding, heritability and plasticity of glioma transcriptional cell states. Nature Genet. 2021;53(10):1469–1479.

60. Dongqiang Zeng Z, Ye R, Shen G, Yu J, Wu Y, Xiong R, Zhou W, Qiu N, Huang L, Sun X, Li J, Bin. Yulin Liao, min Shi, and Wangjun Liao. Volume 12. IOBR: Multi-Omics Immuno-Oncology Biological Research to Decode Tumor Microenvironment and Signatures. Frontiers in Immunology; 2021.

61. Zhang L, Yao J, Wei Y, Zhou Z, Li P, Qu J, Badu-Nkansah A, Yuan X, Huang Y-W, Fukumura K, Mao X, Chang W-C, Saunus J, Lakhani S, Huse JT. Mien-Chie hung, and Dihua Yu. Blocking immunosuppressive neutrophils deters pY696-EZH2–driven brain metastases. Sci Transl Med. 2020;12(545):eaaz5387.

62. Chen Q, Han B, Meng X, Duan C, Yang C, Wu Z, Magafurov D, Zhao S, Safin S. Chuanlu Jiang, and Jinquan Cai. Immunogenomic analysis reveals LGALS1 contributes to the immune heterogeneity and immunosuppression in glioma. Int J Cancer. 2019;145(2):517–30.

63. Wei Wu JL, Klockow M, Zhang F, Lafortune E, Chang L, Jin Y, Wu. Daldrup-Link. Glioblastoma multiforme (GBM): an overview of current therapies and mechanisms of resistance. Pharmacol Res. 2021;171:105780.

64. Yang K-T, Yen C-C, Chang R, Wang J-T, Jin-Shuen, Chen. CAST as a potential oncogene, identified by machine search, in gastric cancer infiltrated with macrophages and associated with Lgr5. Biomolecules. 2022;12(5).

65. Yuce K and Ozkan AI. The kruppel-like factor (KLF) family, diseases, and physiological events. Gene. 2024;895:148027.

66. Wei H, Ma W, Lu X, Liu H, Lin K, Wang Y, Ye Z, Sun L, Huang Z, Pan T, Zhou Z, Cheng EY, Zhang H. Ping Gao, and Xiuying Zhong. KDELR2 promotes breast cancer proliferation via HDAC3-mediated cell cycle progression. Cancer Commun (London England). 2021;41(9):904–20.

67. Kosuke Watari T, Shibata H, Fujita A, Shinoda Y, Murakami H, Abe A, Kawahara H, Ito J, Akiba S, Yoshida. Michihiko Kuwano, and Mayumi Ono. NDRG1 activates VEGF-a-induced angiogenesis through PLCγ1/ERK signaling in mouse vascular endothelial cells. Commun Biology. 2020;3(1):1–14.

68. Xu H, Zhang A, Fang C, Zhu Q, Wang W, Liu Y, Zhang Z, Wang X, Yuan L, Xu Y. Anwen Shao, and Meiqing Lou. SLC11A1 as a stratification indicator for immunotherapy or chemotherapy in patients with glioma. Front Immunol. 2022;13:980378.

Publisher's note