npj Precision Oncology

https://doi.org/10.1038/s41698-025-01213-z

Article in Press

Machine learning-driven comprehensive profiling of tumor heterogeneity and sialylation in hepatocellular carcinoma

Received: 15 September 2025

Accepted: 19 November 2025

Cite this article as: Tang, K., Han, L., Li, J. *et al.* Machine learning-driven comprehensive profiling of tumor heterogeneity and sialylation in hepatocellular carcinoma. *npj Precis. Onc.* (2025). https://doi.org/10.1038/s41698-025-01213-z

Kaiqiang Tang, Lu Han, Junlin Li & Kang Li

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

ARTICLE IN PRESS

© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

ARTICLE IN PRESS

Machine Learning-Driven Comprehensive Profiling of Tumor Heterogeneity and Sialylation in Hepatocellular Carcinoma

Kaiqiang Tang[1, †], Lu Han[2, †], Junlin Li[3], Kang Li[1*]

[1] Department of Radiology, Chongqing General Hospital, Chongqing, China

[2] Comprehensive Ward, Affiliated Hospital of Guizhou Medical University, Guiyang, Guizhou Province, China; Department of Infectious Diseases, The Affiliated Hospital of Guizhou Medical University, Guiyang550000, Guizhou, China

[3] Department of radiology, The 13th People's Hospital of Chongqing, No. 16, Huangjueping Railway New Village, Jiulongpo District, Chongqing400050, China

†These authors contributed equally: Kaiqiang Tang, Lu Han.

***Corresponding Authors:**

Kang Li, MD, E-mail: lkrmyydoctor@126.com; Department of Radiology, Chongqing General Hospital, Chongqing University, Xingguang District, Chongqing, 401147, China; Tel: 023-63390062; Fax: 023-63390062.

ARTICLE IN PRESS

Abstract

Hepatocellular carcinoma (HCC) exhibits profound cellular heterogeneity, the understanding of which is critical for improving prognosis and therapy. Using single-cell RNA sequencing of 32,247 cells from human HCC samples, we characterized the tumor ecosystem and identified five malignant hepatocyte subpopulations with distinct molecular profiles and stage-specific enrichment. Among these, the S100A6$^+$ C1 and S100A9$^+$ C4 subpopulations were predominantly associated with advanced tumors and actively remodeled the tumor microenvironment through enhanced signaling pathways such as MDK and MIF. We further identified PGAM2 as a key transcriptional regulator in early-stage tumors, whose activity correlated with sialylation—a process linked to immune evasion. Based on these findings, we developed a prognostic model integrating PGAM2 and sialylation-related genes, which robustly stratified patients into high- and low-risk groups with significantly different survival outcomes, immune contextures, and predicted

ARTICLE IN PRESS

therapeutic responses. Functional experiments validated AGRN, a component of the signature, as a functional driver of HCC proliferation and invasion. Collectively, our results decode the cellular and molecular heterogeneity of HCC, provide a clinically relevant prognostic tool, and highlight potential targets for further investigation.

KEYWORDS: Hepatocellular carcinoma, Tumor heterogeneity, Sialylation, PGAM2, Prognostic model.

Introduction

Liver cancer is a leading cause of cancer-related mortality worldwide, with over 760,000 new cases and 870,000 deaths reported in 2022[1]. Among its various subtypes, hepatocellular carcinoma (HCC) represents the predominant form, accounting for approximately 80% of all cases[2]. Despite advancements in diagnostic methods, the prognosis for HCC remains poor, with a 5-year survival rate below 20%[3]. This is partly due to the diagnostic and therapeutic challenges posed by the tumor heterogeneity[4, 5]. The complex molecular and histological variability within and between tumors complicates early detection and effective treatment[6], underscoring the need for a deeper understanding of its underlying mechanisms.

Tumor heterogeneity in HCC manifests at multiple levels, including genetic, epigenetic, and phenotypic diversity[7, 8]. This variability is driven by factors such as genetic mutations, tumor microenvironment interactions, and the dynamic behavior of cancer stem cells (CSCs)[9, 10]. The presence of distinct tumor subpopulations with differing molecular profiles complicates treatment, as certain cell populations may exhibit resistance to conventional therapies[11]. Additionally, the process of lineage transition, where cancer cells evolve to adopt drug-resistant or metastasis-prone phenotypes, further exacerbates treatment failure and tumor recurrence[12]. Understanding these mechanisms is crucial for developing more precise therapeutic strategies.

Sialylation, a glycosylation modification that adds sialic acid residues to proteins and lipids,

ARTICLE IN PRESS

plays a critical role in regulating tumor progression[13]. In HCC, aberrant sialylation influences several key processes, including immune evasion, metastasis, and resistance to therapy[14-16]. Elevated expression of sialyltransferases, such as St3gal3, has been linked to poor prognosis and aggressive tumor phenotypes[17]. Sialylated glycans interact with Siglec receptors on immune cells, promoting immune suppression and creating an immunosuppressive tumor microenvironment (TME)[15, 18]. These modifications not only contribute to immune evasion but also affect tumor cell signaling pathways, thus influencing tumor survival and progression. Targeting sialylation presents a promising therapeutic approach to overcome these barriers.

The complexity and high-dimensional nature of single-cell and bulk transcriptomic data pose a significant challenge to deciphering the cellular and molecular underpinnings of tumor heterogeneity. Traditional analytical methods often fall short in capturing the nonlinear relationships and intricate patterns within such data. To this end, machine learning approaches have emerged as powerful tools for unbiased exploration and integration of multi-omics data[19]. In the context of HCC, machine learning algorithms can deconvolute the tumor microenvironment, reconstruct cellular trajectories, and infer cell-cell communication networks, thereby revealing previously unappreciated layers of heterogeneity[20].

ARTICLE IN PRESS

This study aims to explore how tumor heterogeneity and sialylation modifications regulate HCC progression and influence treatment outcomes. By integrating these machine learning-powered insights, we aim to develop a prognostic model that integrates tumor molecular profiles and sialylation-related signatures, which could improve patient stratification and guide therapeutic decision-making.

Results

Comprehensive Single-Cell Profiling Reveals Key Cellular Heterogeneity in HCC

ARTICLE IN PRESS

To explore the cellular diversity within HCC, we performed single-cell RNA sequencing on a cohort of 32,247 cells. Through rigorous quality control and dimensionality reduction, we identified eight major cell types based on well-established marker genes: hepatocytes, fibroblasts, endothelial cells, T/NK cells, B cells, plasma cells, myeloid cells, and MAST cells (Figure 1A-B). These classifications were further validated by analyzing the cell cycle and distribution of cell types across early and advanced tumor stages, hepatocytes from advanced stage tumors showed a significant reduction in proportion compared to early-stage tumors, highlighting the dynamic changes in tumor cell composition as the disease progresses (Figure 1C-D). Marker expression for each cell type was subsequently visualized (Figure 1E-F), ensuring accurate classification. Hepatocytes were characterized by high ALB expression, fibroblasts by high DCN expression, endothelial cells by high PECAM1 expression, T/NK cells by high CD3D and NKG7 expression, B cells by high CD79A expression, plasma cells by high JCHAIN expression, myeloid cells by high LYZ expression, and mast cells by high KIT expression. Notably, we also observed distinct distributions of cell cycle scores (G2M, S) and RNA features (nFeatureRNA and nCountRNA) across the different cell types (Figure 1G-H), providing further insight into the functional state of these cells. Hepatocytes exhibited the highest number of features and counts, indicating their active participation in tumor progression and cellular heterogeneity within the tumor microenvironment. KEGG pathway enrichment analysis highlighted critical biological processes associated with each cell type. Hepatocytes, in particular, showed significant enrichment for pathways related to "Chemical carcinogenesis - reactive

ARTICLE IN PRESS

oxygen species," suggesting a key role of oxidative stress in driving liver cancer progression (Figure 1H).

Identification of distinct malignant subpopulations with stage-specific molecular profiles

To further analyze the heterogeneity of hepatocytes, we performed secondary dimensionality reduction and clustering, identifying five distinct hepatocyte subpopulations (Supplementary Figure 1A). When cross-referenced with the TNM staging system, we observed that cluster 4 was predominantly composed of stage IV hepatocytes, underscoring its relevance in advanced tumor progression (Supplementary Figure 1B). We assessed CNV in hepatocytes using infercnv, with fibroblasts and endothelial cells serving as reference populations. In terms of CNV scores, cluster 4 exhibited the highest scores, consistent with its association with stage IV tumors, which also showed the highest CNV scores (Supplementary Figure 1C-D). Based on the gene expression profiles，We identified five distinct subpopulations of malignant cells, labeled as C0 (CYP2D6+ malignant), C1 (S100A6+ malignant), C2 (UBE2C+ malignant), C3 (HULC+ malignant), and C4 (S100A9+ malignant) (Figure 2A-B). Subpopulations C0, C2, and C3 were predominantly present in early-stage tumors, while C1 and C4 were enriched in advanced-stage tumors, with C4 almost exclusively observed in advanced samples (Figure 2B-E). These findings were further supported by the Ro/e heatmap analysis, which confirmed C1 and C4 as the dominant populations in advanced tumors (Figure 2F). Differential marker expression was

ARTICLE IN PRESS

visualized, revealing key markers associated with each subpopulation，For example, C4 cells exhibited high expression of S100A8, S100A9, SLPI, S100P, and CCL26, highlighting their distinct molecular profile (Figure 2G-H). CNV scores were significantly higher in C4 compared to other subpopulations (Figure 2I-J), supporting its role as a prominent malignant cell type. Additionally, GO enrichment analysis revealed that C0 was enriched for metabolic pathways, C1 for ribosomal functions, C2 for nuclear-related processes, C3 for mitochondrial energy metabolism, and C4 for protein translation (Figure 2K-M). Notably, pathway analysis for metabolic processes revealed that, compared to the other four subpopulations, C4 exhibited significantly enhanced oxidative phosphorylation scores, suggesting a heightened metabolic activity associated with its malignant phenotype (Figure 2N).

Tumor cell differentiation trajectories define early and late-stage progression in HCC

To investigate the differentiation trajectories of tumor cells, we performed monocle2 trajectory analysis, which revealed distinct differentiation paths. C1 was primarily located at the early stages of the trajectory, while C0, C2, and C3 were situated in the later stages. Interestingly, C4 formed a separate branch, indicating a unique differentiation pathway (Figure 3A-E). Slingshot trajectory analysis further defined two main differentiation paths, with C4 occupying an intermediate position in the trajectory. Combining the CytoTRACE scores with its intermediate location in the differentiation path, it is evident that C4 represents a pivotal population in driving disease progression

ARTICLE IN PRESS

(Figure 3G-I). Marker gene expression along pseudotime demonstrated distinct functional changes across cell populations, providing further insights into the dynamic changes in gene regulation as tumor cells differentiate (Figure 3J). Genes correlated with pseudotime were identified, which allowed us to investigate the transcriptional dynamics underlying tumor progression (Figure 3K).

IMalignant cell communication networks highlight key signaling interactions driving tumor progression

Cell-cell communication analysis was performed using CellChat, revealing significant interactions between malignant tumor subpopulations (C1 and C4) and other cell types, particularly endothelial and myeloid cells (Figure 4A-C). Upon comparing the incoming and outgoing communication patterns, we found that tumor cells exhibited significant activity in several signaling pathways, including MIF, MIK, SPP1, and PARs, suggesting their critical role in tumor microenvironment modulation (Figure 4D). We identified key signaling pathways involved in these interactions, with a focus on the MK pathway, which was highly active in both C1 and C4 (Figure 4E-F). C1 cells exhibited increased release of MK signaling molecules, influencing endothelial and fibroblast cells (Figure 4E). Additionally, the MIF-CD74/CXCR4 signaling axis was strongly active between C1 malignant cells and myeloid cells (Figure 4G), highlighting a potential immune evasion mechanism. Expression of key ligand genes in these pathways, including MIF and MDK, was significantly elevated in C1 cells (Figure 4H), further supporting the importance of these interactions in tumor progression. The MDK-

ARTICLE IN PRESS

NCL communication network was also visualized, suggesting potential therapeutic targets (Figure 4I).

PGAM2 acts as a pivotal regulator in the transcriptional landscape of HCC

We identified key transcription factors regulating tumor progression by analyzing the CSI matrix and performing clustering. Three major transcriptional modules (M1, M2, and M3) were identified (Figure 5A–C). By integrating transcription factor activity scores (AUC values), we found that module M1 exhibited high activity across C0, C1, C2, and C3 populations, whereas module M3 showed pronounced activation specifically in C1 and C4 cells (Figure 5D–E). These findings indicate distinct transcriptional regulatory programs underlying different malignant subpopulations. At the individual transcription factor level, XBP1 activity was predominant in C0, C2, and C3 subpopulations, while PGAM2 was most active in C1 cells (Figure 5F). Furthermore, XBP1 exhibited the highest activity in early-stage tumors, whereas PGAM2 and ELF3 were markedly activated in advanced-stage tumors, reflecting a stage-dependent shift in transcriptional regulation (Figure 5G). Across all five malignant cell subpopulations, PGAM2 showed the highest AUC score in C1, ELF3 exhibited strong activity in both C1 and C4, and XBP1 maintained consistently high activity across C0–C3 (Figure 5H–I). Collectively, these results suggest that PGAM2, ELF3, and XBP1 function as key transcriptional regulators orchestrating stage-specific tumor progression and malignant phenotype transitions.

ARTICLE IN PRESS

We investigated the relationship between PGAM2 expression and sialylation (Supplementary Figure 2A-B), a post-translational modification often associated with cancer progression. Using ssGSEA, we calculated sialylation scores and found that C1 cells, which are in the early stage of tumor differentiation, exhibited significantly higher sialylation scores compared to late-stage cells (C2, C3) (Supplementary Figure 2C). A positive correlation between PGAM2 expression and sialylation was confirmed, suggesting that PGAM2 may regulate sialylation pathways, which are critical for tumor cell metastasis and immune evasion (Supplementary Figure 2D).

Development of a prognostic model based on PGAM2 and sialylation-related genes in HCC

To further evaluate the clinical significance of PGAM2 and its associated sialylation-related pathways, we constructed a prognostic model integrating genes linked to both PGAM2 expression and sialylation activity. Candidate genes were first identified through univariate Cox regression analysis, yielding 12 prognosis-related genes, among which ST6GALNAC4, SIGLECL1, GALNTL5, B4GALT5, and AGRN were identified as prominent risk factors associated with poor survival outcomes (Figure 6A). Subsequently, LASSO regression and multivariate Cox analysis refined the model to nine key prognostic genes, with GALNTL5, B4GALT5, and SIGLECL1 emerging as the most statistically significant contributors to patient prognosis (Figure 6B–C). The derived risk score model demonstrated strong predictive capacity. Kaplan–Meier survival analysis revealed that patients in the high-risk group exhibited significantly

ARTICLE IN PRESS

shorter overall survival compared to those in the low-risk group (Figure 6D–F), indicating the clinical relevance of the gene signature. PCA further showed a clear separation between high- and low-risk groups, reflecting distinct transcriptional profiles underlying the two risk categories (Figure 6G). In addition, time-dependent ROC curve analysis showed that the model achieved robust predictive accuracy, with AUC values exceeding 0.7 for 1-, 3-, and 5-year survival (Figure 6H). Together, these findings demonstrate that the PGAM2- and sialylation-associated gene signature provides an effective tool for prognostic stratification in HCC, highlighting its potential utility for risk assessment and clinical decision-making (Figure 6I).

High-risk and low-risk subgroups exhibit distinct functional signatures linked to tumor progression

Differential expression analysis between the high-risk and low-risk groups revealed a distinct transcriptional landscape that underpins their divergent clinical outcomes. Several genes, including CA9, PAEP, EPO, and SLC10A1, were markedly upregulated in the high-risk subgroup (Figure 7A–B), suggesting their potential involvement in promoting tumor aggressiveness and metabolic reprogramming. To further elucidate the biological processes associated with these differential expression patterns, KEGG pathway enrichment analysis demonstrated a significant enrichment of the IL-17 signaling pathway (Figure 7C), a well-established mediator of tumor-associated inflammation and immune modulation. In parallel, GO enrichment analysis revealed that high-risk group–associated genes were predominantly linked to cell cycle–related

ARTICLE IN PRESS

processes, including mitotic nuclear division and chromatid segregation (Figure 7D), indicating that enhanced proliferative activity is a hallmark of tumors with poor prognosis. Additionally, GSEA analysis identified key pathways related to cell chemotaxis and sphingolipid metabolism that were upregulated in high-risk tumors, providing further insights into the molecular mechanisms driving tumor progression (Figure 7E-F).

Immune microenvironment, mutation patterns, and drug sensitivity further define high-risk tumor characteristics

Immune infiltration analysis revealed a higher proportion of Tregs in the high-risk group, along with decreased infiltration of anti-tumor immune cells (Figure 8A-B). Correlation analysis between immune cell populations and risk scores identified significant associations with macrophages and Tregs (Figure 8C-D). Additionally, the TIDE algorithm predicted poorer responses to immunotherapy in the high-risk group (Figure 8E). Further comprehensive assessment of immune infiltration levels in HCC using the ESTIMATE, CIBERSORT, and xCell algorithms consistently demonstrated an immunosuppressive microenvironment in the high-risk group (Figure 8F). Mutation frequency analysis and drug sensitivity predictions further emphasized the clinical relevance of the prognostic model, with high-risk patients exhibiting poorer responses to common therapies (Figure 8G-I).

Functional Validation of AGRN and Its Role in Liver Cancer Progression

ARTICLE IN PRESS

The role of AGRN as a prognostic marker in HCC was further explored through experimental validation. Comparison of mRNA expression levels between HCC tumor and adjacent normal tissues revealed significantly higher AGRN expression in the tumors (Figure 9A). Among several HCC cell lines, Hep 3B and HuH-6 exhibited the highest AGRN levels, and were selected for subsequent experiments (Figure 9B). To investigate the functional role of AGRN, siRNA targeting AGRN was transfected into HCT116 and Hep 3B cells, resulting in a substantial reduction in AGRN expression (Figure 9C). Cell proliferation, measured by the CCK-8 assay, showed a significant decrease in growth rates in AGRN-depleted cells, suggesting that AGRN promotes HCC cell proliferation (Figures 9D–E). Flow cytometry-based apoptosis assays revealed a marked increase in apoptosis in AGRN-knockdown Hep 3B cells compared to controls (Figures 9F–G). Further functional assays using transwell chambers demonstrated that AGRN knockdown significantly reduced cell migration and invasion, supporting the notion that AGRN enhances the aggressive behavior of HCC cells (Figures 9H–I).

ARTICLE IN PRESS

Discussion

Our single-cell transcriptomic analysis of HCC delineates five distinct malignant cell subpopulations with stage-specific dominance. The advanced-stage associated C1 and C4 subpopulations engage in specific pro-tumorigenic crosstalk, such as MDK and MIF signaling, within the tumor microenvironment. Building on this, we established a potent prognostic model centered on PGAM2 and sialylation-related genes, which

ARTICLE IN PRESS

effectively stratifies patients and reflects an immunosuppressive phenotype in high-risk cases. Furthermore, functional validation confirmed AGRN, a key model component, as a bona fide oncoprotein driving HCC proliferation and invasion.

Our findings reinforce the concept of cellular heterogeneity in HCC, central to understanding its aggressive nature and therapeutic resistance. Previous studies have suggested that hepatocellular carcinoma, like many cancers, is driven by a subpopulation of tumor-initiating cells (TICs) with stem-like properties and immune evasion capabilities[21, 22]. Our trajectory analysis revealed distinct differentiation pathways for tumor cells, suggesting that different subpopulations—particularly C1 may drive early and late-stage progression, respectively. This is consistent with the concept of early dissemination and high neoantigen intratumor heterogeneity observed in metastases, which complicates immune recognition and treatment efficacy[23].

Our results underscore the cellular diversity within HCC, where distinct tumor subpopulations contribute to various stages of tumor progression. The C1 subpopulation, predominantly found in early-stage tumors, exhibited elevated expression of S100A11 and increased secretion of MDK—findings consistent with previous literature linking MDK to tumor progression and metastasis[24, 25]. Specifically, circMDK, derived from the MDK gene, has been identified as an oncogenic circRNA that activates the PI3K/AKT/mTOR pathway, promoting tumor cell proliferation, migration, and invasion[26]. This observation aligns with our finding that C1 cells secrete

ARTICLE IN PRESS

higher levels of MDK, suggesting a potential link between PGAM2-driven metabolic changes and MDK signaling in HCC progression.

Additionally, the elevated expression of S100A11 in C1 tumor cells is consistent with prior studies highlighting its role in inflammation and tumor progression[27]. S100A11 is often deregulated early in hepatocellular carcinoma, particularly in the context of steatosis, and is involved in multiple stages of cancer development[28]. Our findings suggest that S100A11, in conjunction with PGAM2 and MDK, may contribute to the aggressive nature of early-stage tumors, making it a potential therapeutic target for early intervention.

Beyond the early-stage PGAM2-driven dynamics, our study highlights the C4 subpopulation ($S100A9^{+}$ malignant) as a metabolic engine propelling advanced HCC. Characterized by the highest CNV burden and near-exclusive enrichment in late-stage tumors, C4 cells undergo profound metabolic rewiring, transitioning from glycolysis to a predominant reliance on oxidative phosphorylation (OXPHOS). Importantly, this OXPHOS surge is not merely a metabolic adaptation but is intimately coupled to elevated reactive oxygen species (ROS) production, positioning ROS as a central signaling mediator and driver of malignancy. This OXPHOS–ROS axis establishes a self-reinforcing loop that fuels tumor aggressiveness. The enhanced OXPHOS satisfies the substantial energy and biosynthetic demands of rapidly proliferating and invasive cells[29], while the concomitant ROS burst functions as a potent intracellular

ARTICLE IN PRESS

signal. Elevated ROS can stabilize hypoxia-inducible factors (HIFs) even under normoxic conditions, reinforcing a pro-tumorigenic transcriptional program[30, 31]. Concurrently, ROS promotes degradation of key tumor suppressors such as E-cadherin—via mechanisms including RNF25-mediated ubiquitination—thereby facilitating EMT and metastatic potential[32].

The marked enrichment of S100A9 in C4 provides a plausible upstream mechanism for this vicious cycle. S100A9, often induced under TACE-associated hypoxia, can scaffold PGAM5, enhancing its stability and promoting mitochondrial fission[33]. The resulting fragmented mitochondria are predisposed to ROS generation, locking C4 cells in a state of sustained OXPHOS-driven oxidative stress. Moreover, this ROS-enriched environment functions as a selective pressure favoring cancer stem-like properties[34], activating stress-response pathways such as Nrf2 and triggering mitophagy[35], processes that are closely linked to stemness maintenance and therapy resistance in HCC, as observed in ADAR1-enriched LCSCs[36]. Positioned at a critical branch point in the differentiation trajectory, C4 cells leverage their OXPHOS–ROS core not only to meet energetic demands but also to orchestrate a pro-metastatic signaling network, sustain stem-like plasticity, and resist therapeutic pressures. Together, these features underscore C4 as a pivotal driver of late-stage HCC malignancy.

Sialylation modifications have long been implicated in cancer progression, particularly

ARTICLE IN PRESS

in immune evasion and metastasis. Our study found a significant correlation between PGAM2 expression and increased sialylation in early-stage tumor cells, supporting the idea that sialylation is critical for metastasis and immune evasion. The HBV X protein has been reported to regulate sialyl Lewis A synthesis, suggesting that viral proteins modulate glycosylation events to enhance metastatic potential[37]. Interestingly, our study also revealed that higher sialylation scores were associated with early-stage C1 cells, which further supports the hypothesis that sialylation is not only a feature of late-stage disease but also a key factor in early-stage tumor progression. These findings contribute to the growing body of evidence linking sialylation to immune escape in liver cancer.

Beyond sialylation, we observed significant enrichment in sphingolipid metabolism pathways in high-risk tumor subgroups. This is in line with literature reports linking sphingolipid metabolism in tumor-associated macrophages to resistance to immunotherapy in hepatocellular carcinoma[38]. Our study's finding that high-risk tumors exhibit altered sphingolipid metabolism suggests that metabolic reprogramming in the TME plays a crucial role in conferring resistance to treatment, particularly immunotherapy. This highlights the potential for targeting metabolic pathways such as sphingolipid metabolism to improve therapeutic outcomes in HCC.

PGAM2, identified as a key regulator of early-stage HCC, exhibited the highest expression in C1 tumor cells, which are associated with early tumor stages. This is

ARTICLE IN PRESS

consistent with existing literature, which implicates PGAM2 in promoting glycolysis—a hallmark of cancer metabolism[39]. Our results further support the notion that PGAM2 plays a pivotal role in metabolic reprogramming during tumor progression, particularly in early-stage HCC, where high glycolytic activity is essential for tumor growth and survival[40]. Although our study did not directly investigate the regulatory mechanisms of PGAM2, existing literature suggests that PGAM2 activity is regulated by post-translational modifications, such as acetylation at lysine 100[41]. The acetylation status of PGAM2 affects its enzymatic activity, and enzymes like SIRT2 can regulate this process[39]. Oxidative stress has been shown to promote PGAM2 activation, thereby enhancing glycolysis in tumor cells. This suggests that PGAM2 may function as a metabolic switch, responding to the tumor microenvironment to support cancer cell proliferation. Furthermore, PGAM2's involvement in glycolysis, nucleotide biosynthesis, and redox homeostasis highlights its broad role in cancer metabolism.

ARTICLE IN PRESS

Our functional validation firmly establishes AGRN as a potent oncoprotein in HCC, driving proliferation, invasion, and suppressing apoptosis. Although studies on AGRN in liver cancer remain scarce, the convergence of evidence from other malignancies underscores its potential as a central orchestrator of HCC malignancy, extending beyond a prognostic marker. Mechanistically, AGRN appears to operate through two interconnected axes: tumor-intrinsic signaling and tumor-immune modulation. Intrinsically, the pronounced pro-invasive phenotype associated with AGRN expression aligns with its capacity to activate the NOTCH1 pathway, a well-established

ARTICLE IN PRESS

driver of EMT and invasion in HCC[42]. This positions the AGRN–NOTCH1 axis as a plausible mechanistic core underlying the aggressive traits observed in our high-risk subgroup. Extrinsically, the immunosuppressive microenvironment characterized by Treg enrichment and predicted poor response to immunotherapy in high-risk patients may be partially explained by AGRN-mediated immune modulation. Evidence from breast cancer indicates that AGRN can scaffold interactions stabilizing PD-L1, and its knockdown enhances $CD8^+$ T cell infiltration[43]. By analogy, elevated AGRN in our high-risk HCC cohort may actively sculpt an immune-cold microenvironment, contributing to therapy resistance. Taken together, AGRN emerges not as a passive component but as a nexus protein, potentially coordinating a dual assault on tumor progression: promoting aggressive tumor cell behavior via pathways such as NOTCH1, while simultaneously suppressing anti-tumor immunity. This integrated model, grounded in our functional data and informed by cross-cancer insights, elevates AGRN from a gene signature component to a compelling therapeutic candidate, offering a strategy to simultaneously curb tumor growth and restore immune surveillance in HCC.

While this study provides valuable insights into the cellular heterogeneity of HCC, there are several limitations. Although the sample size of 32,247 cells is substantial, it may not fully capture the diversity of rare tumor or immune cell subpopulations. Future studies with larger cohorts could provide a more comprehensive view of the tumor microenvironment. Additionally, while PGAM2 was identified as a key regulator, functional validation using techniques such as CRISPR-Cas9 is necessary to confirm

ARTICLE IN PRESS

its role in tumor progression and therapy response. Lastly, while our prognostic model showed strong predictive power, it requires validation in independent cohorts to ensure its clinical applicability.

In conclusion, our work delineates an association between metabolic reprogramming, immune evasion, and clinical prognosis in HCC. Our data suggest that PGAM2 may serve as a key node, connecting glycolytic flux to sialylation, which is correlated with an immunosuppressive microenvironment and aggressive tumor behavior. The prognostic model we developed, rooted in these observations, effectively stratified patients in our cohort and offers a candidate tool for outcome prediction. Future work focused on validating these targets in preclinical models is essential to determine the translational potential of these insights for HCC patients.

Materials and Methods

Acquisition and Processing of Transcriptomic Data

RNA expression data for HCC and corresponding clinical information were obtained from the Xena database (https://xenabrowser.net/) for a cohort of 368 patients. The transcriptomic data were pre-processed and converted to transcripts per million (TPM) format. Log2 transformation was applied for downstream analyses. A validation cohort was used to assess the stability and accuracy of the predictive model. Gene expression data were normalized and subsequently analyzed for model development and validation.

ARTICLE IN PRESS

Acquisition and Processing of Single-Cell RNA Sequencing Data

Single-cell RNA sequencing (scRNA-seq) data was downloaded from the GEO database (GSE149614), containing 10 HCC tumor samples (4 early-stage and 6 advanced-stage). The original data were generated by aligning sequencing reads to the human reference genome GRCh38 (hg38) using the Cell Ranger pipeline. Data analysis was performed using R software (version 4.1.3) and the Seurat package (version 4.0.0). Initial data preprocessing included quality control of cells based on mitochondrial gene content (less than 20%), blood cell contamination (less than 3%), and total UMI count (200–40,000) and gene count (200–7,000). Normalization was performed using the NormalizeData function, and highly variable genes were selected using FindVariableFeatures (top 2000 genes). Cell cycle effects were regressed out with the parameters vars.to.regress = c ("S.Score", "G2M.Score"). Batch effect correction was performed using the Harmony algorithm. Dimensionality reduction and clustering were conducted using UMAP (RunUMAP) and Louvain clustering (FindClusters), respectively. Differential gene expression between clusters was performed using FindAllMarkers, with thresholds of p-value < 0.05, log2 fold change > 0.25, and expression detected in >10% of cells.

Acquisition of Sialylation-Related Genes

Sialylation-related genes were curated from the Molecular Signatures Database (MSigDB), as previously described[44]. The gene set encompasses key enzymes

ARTICLE IN PRESS

involved in sialylation metabolism, including sialyltransferases, transporters, and neuraminidases. his comprehensive set was used for downstream analyses, including the calculation of the sialylation signature score.

Cell Annotation and Classification

Cell type annotations were performed using established markers for various cell types, as follows: Hepatocytes: EPCAM, KRT18, KRT19, ALB; Fibroblasts: DCN, THY1, COL1A1, COL1A2; Endothelial cells: PECAM1, CLDN5, FLT1, RAMP2; T cells: CD3D, CD3E, CD3G, TRAC; NK cells: NKG7, GNLY, NCAM1, KLRD1; B cells: CD79A, IGHM, IGHG3, IGHA2; Plasma cells: JCHAIN; Myeloid cells: LYZ, MARCO, CD68, FCGR3A; Mast cells: KIT, MS4A2, GATA2. These markers were used for cell classification, and visualizations were created to ensure accurate cell type annotations.

CNV Analysis of Single-Cell Data

The InferCNV package was used to analyze CNV across the tumor cell subpopulations[45]. Fibroblasts and endothelial cells were selected as reference cell populations to help detect CNVs in tumor cells. The CNV scores were calculated using default settings in InferCNV and were visualized on a heatmap to identify malignant subpopulations exhibiting copy number alterations. A threshold of 0.3 was set for CNV significance to call a region as a gain or loss.

Pseudotime Analysis of Tumor Cell Differentiation

ARTICLE IN PRESS

Pseudotime analysis was performed using the Monocle2 package (version 2.18.0)[46]. The DDRTree algorithm was employed for dimensionality reduction, and the default parameters were used for trajectory analysis. This method allowed us to map the differentiation trajectories of tumor cells, providing insights into tumor progression from early to late stages.

Transcription Factor Analysis in HCC Tumor Cells

The SCENIC package (version 1.1.2) was used to identify key transcription factors regulating gene expression in tumor cell subpopulations[47]. The analysis utilized the RcisTarget and GRNBoost motif databases for motif discovery. Transcription factor activity scores for each cell type were computed using the AUCell function in SCENIC. These scores allowed us to identify differentially active transcription factors across tumor cell subpopulations.

Cell-Cell Communication Analysis

The CellChat package (version 1.1.2) was used to investigate potential cell-cell communication within the tumor microenvironment[48]. A normalized gene expression matrix was imported into CellChat, and communication interactions were analyzed using the functions identifyOverExpressedGenes, identifyOverExpressedInteractions, and ProjectData. Cell-cell communication networks were constructed using the aggregateNet function, and interactions between tumor cells and immune or stromal cells were identified.

ARTICLE IN PRESS

Calculation of Sialylation Signature Scores

Sialylation-related genes were used to calculate a sialylation signature score for each cell using the ssGSEA method implemented in the GSVA package (version 1.34.0)[49]. The sialylation signature score was calculated for each tumor cell, and its correlation with tumor progression and PGAM2 expression was assessed.

Immune Infiltration Analysis

Immune infiltration levels in high-risk and low-risk tumor groups were evaluated using three different algorithms: ESTIMATE, CIBERSORT, and xCell[50]. The IOBR package (version 1.0.0) was used for immune cell deconvolution, and the immune infiltration profiles of each group were visualized using heatmaps and bar plots[51]. The results were used to explore the relationship between immune cell composition and tumor progression.

ARTICLE IN PRESS

Differential Gene Expression and Pathway Enrichment Analysis

Differential gene expression between high-risk and low-risk groups was performed using the limma package (version 3.46.0)[52]. The threshold for differential expression was set to |log2FoldChange| > 1.5 and adjusted p-value < 0.05. Pathway enrichment analysis was performed using the clusterProfiler package (version 4.0) for KEGG and GO databases[53]. Enriched pathways were visualized using ggplot2 (version 3.3.5).

ARTICLE IN PRESS

Genomic Variation Landscape Analysis Between Risk Groups

The genomic variation landscape was compared between high-risk and low-risk groups using the "maftools" package (version 2.8.0) in R[54]. This package was used to examine mutational burden and create mutation waterfall plots. The mutational profiles of the two risk groups were compared to identify significant differences in mutation frequencies.

Development of a Prognostic Model Based on PGAM2 and Sialylation-Related Genes

A prognostic model was developed using a set of 116 genes, including PGAM2 and sialylation-related genes, identified through single-factor Cox analysis[55]. The model was constructed using the LASSO (least absolute shrinkage and selection operator) regression combined with Cox regression (using glmnet package, version 4.1). The accuracy of the prognostic model was evaluated using the timeROC package, with the calculation of AUC values for 1-year, 3-year, and 5-year survival predictions.

Clinical Sample Collection and Preparation

Tumor and adjacent normal tissues were collected from 15 liver metastasis patients undergoing surgery at Chongqing General Hospital between May 2019 and April 2024. Normal tissues were obtained at least 3 cm from the tumor margins. Immediately after excision, the samples were snap-frozen in liquid nitrogen and stored at −80°C to preserve RNA integrity for subsequent analysis. Ethical approval was obtained from

ARTICLE IN PRESS

the Institutional Review Board (IRB) of Chongqing General Hospital, and informed consent was acquired from all participants.

RNA Extraction and Quantitative Real-Time PCR (qRT-PCR)

Total RNA was extracted from tissue and cell samples using TRIzol reagent (Invitrogen, USA) according to the manufacturer's protocol. RNA quality was assessed using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, USA), with only samples having an A260/A280 ratio between 1.8 and 2.0 included. cDNA synthesis was carried out using 1 μg of RNA per sample with the PrimeScript RT Kit (Takara, Japan). AGRN expression was quantified by qRT-PCR using SYBR Premix Ex Taq (Takara, Japan) on the QuantStudio 5 PCR system (Applied Biosystems, USA). The following cycling conditions were used: 95°C for 30 seconds, followed by 40 cycles of 95°C for 5 seconds and 60°C for 30 seconds. Relative gene expression was calculated using the 2^−ΔΔCt method, normalized to GAPDH. All experiments were performed in triplicate.

ARTICLE IN PRESS

Cultivation and Validation of Cell Lines

Cell lines used in this study were authenticated via short tandem repeat (STR) profiling and routinely tested for mycoplasma contamination. Cells were cultured in Dulbecco's Modified Eagle Medium (DMEM; Gibco, USA) supplemented with 10% fetal bovine serum (FBS; Gibco, USA) and 1% penicillin-streptomycin (Gibco, USA) at 37°C in a 5% CO_2 incubator. Upon reaching ~80% confluence, RNA was extracted and AGRN

ARTICLE IN PRESS

expression was evaluated by qRT-PCR as described for tissue samples. All experiments were conducted in triplicate.

Procedure for siRNA Transfection

Small interfering RNAs (siRNAs) targeting AGRN and control siRNAs were synthesized by TsingKe Biotechnology and dissolved in nuclease-free water at a concentration of 10 μM. HCC cell lines (Hep 3B and HuH-6) were seeded in 6-well plates (2×10^5 cells/well) and incubated overnight. For transfection, siRNAs (50 nM) were mixed with 5 μL of Lipofectamine 3000 (Invitrogen, USA) in Opti-MEM (Gibco, USA) and incubated for 15 minutes at room temperature before adding to the cells. After 6 hours, the medium was replaced with fresh DMEM containing 10% FBS. Cells were harvested 48 hours post-transfection for RNA extraction and AGRN expression analysis by qRT-PCR. AGRN expression was reduced by over 70% compared to controls ($p < 0.01$, Student's t-test). Each experiment was performed in triplicate.

ARTICLE IN PRESS

Cell Proliferation Evaluation Using CCK-8 Assay

Cell proliferation was assessed using the Cell Counting Kit-8 (CCK-8; Dojindo, Japan). Cells were seeded in 96-well plates (3,000 cells/well) and transfected with siRNAs as described earlier. Proliferation was measured at 24, 48, 72, and 96 hours post-transfection by adding 10 μL of CCK-8 reagent to each well and incubating for 2 hours at 37°C. Absorbance at 450 nm was measured using a Synergy H1 microplate reader (BioTek, USA). Data were presented as fold-change in absorbance relative to baseline

ARTICLE IN PRESS

(0-hour) measurements, with five replicates per condition.

Measurement of Cell Apoptosis Using Flow Cytometry

Apoptosis was analyzed 48 hours post-transfection using the Annexin V-FITC/PI apoptosis detection kit (BD Biosciences, USA). Cells were harvested, washed twice with PBS, and stained according to the manufacturer's protocol. Flow cytometric analysis was conducted on a BD FACSVerse instrument to quantify early apoptotic (Annexin V+/PI−) and late apoptotic (Annexin V+/PI+) cells. Data were analyzed using FlowJo software (version 10.0).

Evaluation of Cell Migration and Invasion

Cell migration was assessed using Transwell chambers (Corning, USA) with 8 μm pore size membranes. Cells (5×10^4) in serum-free medium were seeded into the upper chamber, while the lower chamber contained medium with 10% FBS as a chemoattractant. After 24 hours of incubation, non-migrated cells on the upper surface were removed, and migrated cells on the lower surface were fixed with 4% paraformaldehyde and stained with 0.1% crystal violet. Cells were counted from five random visual fields per chamber. For the invasion assay, Transwell membranes were pre-coated with Matrigel (BD Biosciences, USA) before cell seeding. Migration and invasion assays were repeated in triplicate.

Statistical Analysis

ARTICLE IN PRESS

All statistical analyses and visualizations were performed using R (version 4.1.3). Pearson's correlation coefficient was used to assess correlations between continuous variables. Chi-square tests were applied to compare categorical variables, and Wilcoxon rank-sum tests were used for continuous variables. The survminer package was used to determine the optimal cutoff value for survival analysis. Cox regression and Kaplan-Meier survival analysis were performed using the survival package (version 3.2-13).

Acknowledgements

None.

Data availability statement

The datasets supporting the findings of this study are available in public repositories. The bulk RNA-seq data and corresponding clinical information for the HCC cohort were sourced from the TCGA-HCC project via the UCSC Xena database (https://xenabrowser.net/). The single-cell RNA-seq data analyzed in this study are available in the GEO database under accession code GSE149614. All other data generated or analyzed during this study are included in this published article and its supplementary information files.

Code Availability Statement

The code used for data analysis, including single-cell RNA sequencing processing, cell type identification, and prognostic model construction, is available upon request from

ARTICLE IN PRESS

the corresponding author.

Competing Interests

No conflicts of interest to disclose.

Author contributions

Conception and design of the research: Kaiqiang Tang, Lu Han, Kang Li. Experiment: Kaiqiang Tang, Lu Han, Junlin Li. Analysis and interpretation of data: Kaiqiang Tang. Statistical analysis: Kaiqiang Tang, Lu Han, Junlin Li. Drafting the manuscript: Kaiqiang Tang, Lu Han. Revision of manuscript for important intellectual content: Kang Li. All authors read and approved the final manuscript.

ARTICLE IN PRESS

Funding

None.

References

1. Bray, F.; Laversanne, M.; Sung, H.; Ferlay, J.; Siegel, R. L.; Soerjomataram, I.; Jemal, A., Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* **2024,** *74* (3), 229-263.

2. Rumgay, H.; Ferlay, J.; de Martel, C.; Georges, D.; Ibrahim, A. S.; Zheng, R.; Wei, W.; Lemmens, V.; Soerjomataram, I., Global, regional and national burden of primary liver cancer by subtype. *Eur J Cancer* **2022,** *161*, 108-118.

3. Cai, K.; Fu, W.; Wang, Z.; Yang, X.; Liu, H.; Ji, Z., Optimizing Prognostic Predictions in Liver Cancer with Machine Learning and Survival Analysis. *Entropy (Basel)* **2024,** *26* (9).

4. Bao, M.; Wu, A., Understanding the heterogeneity in liver hepatocellular carcinoma with a special focus on malignant cell through single-cell analysis. *Discov Oncol* **2024,** *15* (1), 241.

5. Yang, J. D.; Heimbach, J. K., New advances in the diagnosis and management of hepatocellular carcinoma. *BMJ* **2020,** *371*, m3544.

6. Cordani, M.; Dando, I.; Ambrosini, G.; Gonzalez-Menendez, P., Signaling, cancer cell plasticity, and intratumor heterogeneity. *Cell Commun Signal* **2024,** *22* (1), 255.

7. Lin, D. C.; Mayakonda, A.; Dinh, H. Q.; Huang, P.; Lin, L.; Liu, X.; Ding, L. W.; Wang, J.; Berman, B. P.; Song, E. W.; Yin, D.; Koeffler, H. P., Genomic and Epigenomic Heterogeneity of Hepatocellular Carcinoma. *Cancer Res* **2017,** *77* (9), 2255-2265.

8. Su, X.; Zhao, L.; Shi, Y.; Zhang, R.; Long, Q.; Bai, S.; Luo, Q.; Lin, Y.; Zou, X.; Ghazanfar, S.; Tao, K.; Yang, G.; Wang, L.; He, K. Y.; Cui, X.; He, J.; Wu, J. X.; Han, B.; Yan, B.; Deng, B.; Wang, N.; Li, X.; Yang, P.; Hou, S.; Sun, J.; Yang, J. Y. H.; Chen, J.; Han, Z. G., Clonal evolution in liver cancer at single-cell and single-variant resolution. *J Hematol Oncol* **2021,** *14* (1), 22.

9. Liang, R.; Hong, W.; Zhang, Y.; Ma, D.; Li, J.; Shi, Y.; Luo, Q.; Du, S.; Song, G., Deep dissection of stemness-related hierarchies in hepatocellular carcinoma. *J Transl Med* **2023,** *21* (1), 631.

10. Sanchez-Martin, A.; Sanchon-Sanchez, P.; Romero, M. R.; Marin, J. J. G.; Briz, O., Impact of tumor suppressor genes inactivation on the multidrug resistance phenotype of hepatocellular carcinoma cells. *Biomed Pharmacother* **2023,** *165*, 115209.

11. Huang, T.; Song, X.; Xu, D.; Tiek, D.; Goenka, A.; Wu, B.; Sastry, N.; Hu, B.; Cheng, S. Y., Stem cell programs in cancer initiation, progression, and therapy resistance. *Theranostics* **2020,** *10* (19), 8721-8743.

12. Moorman, A.; Benitez, E. K.; Cambulli, F.; Jiang, Q.; Mahmoud, A.; Lumish, M.; Hartner, S.; Balkaran, S.; Bermeo, J.; Asawa, S.; Firat, C.; Saxena, A.; Wu, F.; Luthra, A.; Burdziak, C.; Xie, Y.; Sgambati, V.; Luckett, K.; Li, Y.; Yi, Z.; Masilionis, I.; Soares, K.; Pappou, E.; Yaeger, R.; Kingham, T. P.; Jarnagin, W.; Paty, P. B.; Weiser, M. R.; Mazutis, L.; D'Angelica, M.; Shia, J.; Garcia-Aguilar, J.; Nawy, T.; Hollmann, T. J.; Chaligne, R.; Sanchez-Vega, F.; Sharma, R.; Pe'er, D.; Ganesh, K., Progressive plasticity during colorectal cancer metastasis. *Nature* **2025,** *637* (8047), 947-954.

13. Li, F.; Ding, J., Sialylation is involved in cell fate decision during development, reprogramming and cancer progression. *Protein Cell* **2019,** *10* (8), 550-565.

14. Dashzeveg, N. K.; Jia, Y.; Zhang, Y.; Gerratana, L.; Patel, P.; Shajahan, A.; Dandar, T.; Ramos, E. K.; Almubarak, H. F.; Adorno-Cruz, V.; Taftaf, R.; Schuster, E. J.; Scholten, D.; Sokolowski, M. T.; Reduzzi, C.; El-Shennawy, L.; Hoffmann, A. D.; Manai, M.; Zhang, Q.; D'Amico, P.; Azadi, P.; Colley, K. J.; Platanias, L. C.; Shah, A. N.; Gradishar, W. J.; Cristofanilli, M.; Muller, W. A.; Cobb, B. A.; Liu, H., Dynamic Glycoprotein Hyposialylation Promotes Chemotherapy Evasion and Metastatic Seeding of Quiescent Circulating Tumor Cell Clusters in Breast Cancer. *Cancer Discov* **2023,** *13* (9), 2050-2071.

15. Wieboldt, R.; Sandholzer, M.; Carlini, E.; Lin, C. W.; Borsch, A.; Zingg, A.; Lardinois, D.; Herzig, P.; Don, L.; Zippelius, A.; Laubli, H.; Mantuano, N. R., Engagement of sialylated glycans with Siglec receptors on suppressive myeloid cells inhibits anticancer immunity via CCL2. *Cell Mol Immunol* **2024,** *21* (5), 495-509.

16. Yang, M.; Shi, D.; Lyu, J.; Pan, Y.; Lyv, Y.; Chen, X.; Ouyang, Y.; Liu, Y.; Li, Y.; Song, L., Supplementing sialic acid analogs overcomes radiotherapy resistance in triple-negative breast cancer by exacerbating ER stress. *Redox Biol* **2025,** *85*, 103712.

17. Cao, K.; Zhang, G.; Yang, M.; Wang, Y.; He, M.; Zhang, C.; Huang, Y.; Lu, J.; Liu, H., Attenuation of Sialylation Augments Antitumor Immunity and Improves Response to

ARTICLE IN PRESS

ARTICLE IN PRESS

Immunotherapy in Ovarian Cancer. *Cancer Res* **2023,** *83* (13), 2171-2186.
18. Jame-Chenarboo, Z.; Gray, T. E.; Macauley, M. S., Advances in understanding and exploiting Siglec-glycan interactions. *Curr Opin Chem Biol* **2024,** *80*, 102454.
19. Gao, Q.; Yang, L.; Lu, M.; Jin, R.; Ye, H.; Ma, T., The artificial intelligence and machine learning in lung cancer immunotherapy. *J Hematol Oncol* **2023,** *16* (1), 55.
20. Gao, J.; Zhang, M.; Chen, Q.; Ye, K.; Wu, J.; Wang, T.; Zhang, P.; Feng, G., Integrating machine learning and molecular docking to decipher the molecular network of aflatoxin B1-induced hepatocellular carcinoma. *Int J Surg* **2025,** *111* (7), 4539-4549.
21. Yang, C.; Geng, H.; Yang, X.; Ji, S.; Liu, Z.; Feng, H.; Li, Q.; Zhang, T.; Zhang, S.; Ma, X.; Zhu, C.; Xu, N.; Xia, Y.; Li, Y.; Wang, H.; Yu, C.; Du, S.; Miao, B.; Xu, L.; Wang, H.; Cao, Y.; Li, B.; Zhu, L.; Tang, X.; Zhang, H.; Zhu, C.; Huang, Z.; Leng, C.; Hu, H.; Chen, X.; Yuan, S.; Jin, G.; Bernards, R.; Sun, C.; Zheng, Q.; Qin, W.; Gao, Q.; Wang, C., Targeting the immune privilege of tumor-initiating cells to enhance cancer immunotherapy. *Cancer Cell* **2024,** *42* (12), 2064-2081 e19.
22. Shan, Q.; Yin, L.; Zhan, Q.; Yu, J.; Pan, S.; Zhuo, J.; Zhou, W.; Bao, J.; Zhang, L.; Hong, J.; Xiang, J.; Que, Q.; Chen, K.; Xu, S.; Wang, J.; Zhu, Y.; He, B.; Wu, J.; Xie, H.; Zheng, S.; Feng, T.; Ling, S.; Xu, X., The p-MYH9/USP22/HIF-1alpha axis promotes lenvatinib resistance and cancer stemness in hepatocellular carcinoma. *Signal Transduct Target Ther* **2024,** *9* (1), 249.
23. Nguyen, P. H. D.; Ma, S.; Phua, C. Z. J.; Kaya, N. A.; Lai, H. L. H.; Lim, C. J.; Lim, J. Q.; Wasser, M.; Lai, L.; Tam, W. L.; Lim, T. K. H.; Wan, W. K.; Loh, T.; Leow, W. Q.; Pang, Y. H.; Chan, C. Y.; Lee, S. Y.; Cheow, P. C.; Toh, H. C.; Ginhoux, F.; Iyer, S.; Kow, A. W. C.; Young Dan, Y.; Chung, A.; Bonney, G. K.; Goh, B. K. P.; Albani, S.; Chow, P. K. H.; Zhai, W.; Chew, V., Intratumoural immune heterogeneity as a hallmark of tumour evolution and progression in hepatocellular carcinoma. *Nat Commun* **2021,** *12* (1), 227.
24. Filippou, P. S.; Karagiannis, G. S.; Constantinidou, A., Midkine (MDK) growth factor: a key player in cancer progression and a promising therapeutic target. *Oncogene* **2020,** *39* (10), 2040-2054.
25. Karaman, S.; Alitalo, K., Midkine and Melanoma Metastasis: A Malevolent Mix. *Dev Cell* **2017,** *42* (3), 205-207.
26. Du, A.; Li, S.; Zhou, Y.; Disoma, C.; Liao, Y.; Zhang, Y.; Chen, Z.; Yang, Q.; Liu, P.; Liu, S.; Dong, Z.; Razzaq, A.; Tao, S.; Chen, X.; Liu, Y.; Xu, L.; Zhang, Q.; Li, S.; Peng, J.; Xia, Z., M6A-mediated upregulation of circMDK promotes tumorigenesis and acts as a nanotherapeutic target in hepatocellular carcinoma. *Mol Cancer* **2022,** *21* (1), 109.
27. Comitre-Mariano, B.; Segura-Collar, B.; Vellila-Alonso, G.; Contreras, R.; Henandez-Lain, A.; Valiente, M.; Sepulveda, J. M.; Marcus, S. G.; Garcia-Posadas, G.; Jimenez-Roldan, L.; Perez-Nunez, A.; Gargini, R., S100A proteins show a spatial distribution of inflammation associated with the glioblastoma microenvironment architecture. *Theranostics* **2025,** *15* (2), 726-744.
28. Sobolewski, C.; Abegg, D.; Berthou, F.; Dolicka, D.; Calo, N.; Sempoux, C.; Fournier, M.; Maeder, C.; Ay, A. S.; Clavien, P. A.; Humar, B.; Dufour, J. F.; Adibekian, A.; Foti, M., S100A11/ANXA2 belongs to a tumour suppressor/oncogene network deregulated early with steatosis and involved in inflammation and hepatocellular carcinoma development. *Gut* **2020,** *69* (10), 1841-1854.

29. Qiu, X.; Li, Y.; Zhang, Z., Crosstalk between oxidative phosphorylation and immune escape in cancer: a new concept of therapeutic targets selection. *Cell Oncol (Dordr)* **2023,** *46* (4), 847-865.

30. Alva, R.; Wiebe, J. E.; Stuart, J. A., Revisiting reactive oxygen species production in hypoxia. *Pflugers Arch* **2024,** *476* (9), 1423-1444.

31. Zhang, J.; Yao, M.; Xia, S.; Zeng, F.; Liu, Q., Systematic and comprehensive insights into HIF-1 stabilization under normoxic conditions: implications for cellular adaptation and therapeutic strategies in cancer. *Cell Mol Biol Lett* **2025,** *30* (1), 2.

32. Huang, Z.; Zhou, L.; Duan, J.; Qin, S.; Jiang, J.; Chen, H.; Wang, K.; Liu, R.; Yuan, M.; Tang, X.; Nice, E. C.; Wei, Y.; Zhang, W.; Huang, C., Oxidative Stress Promotes Liver Cancer Metastasis via RNF25-Mediated E-Cadherin Protein Degradation. *Adv Sci (Weinh)* **2024,** *11* (13), e2306929.

33. Zhong, C.; Niu, Y.; Liu, W.; Yuan, Y.; Li, K.; Shi, Y.; Qiu, Z.; Li, K.; Lin, Z.; Huang, Z.; Zuo, D.; Yang, Z.; Liao, Y.; Zhang, Y.; Wang, C.; Qiu, J.; He, W.; Yuan, Y.; Li, B., S100A9 Derived from Chemoembolization-Induced Hypoxia Governs Mitochondrial Function in Hepatocellular Carcinoma Progression. *Adv Sci (Weinh)* **2022,** *9* (30), e2202206.

34. Zhao, H.; Wu, S.; Li, H.; Duan, Q.; Zhang, Z.; Shen, Q.; Wang, C.; Yin, T., ROS/KRAS/AMPK Signaling Contributes to Gemcitabine-Induced Stem-like Cell Properties in Pancreatic Cancer. *Mol Ther Oncolytics* **2019,** *14*, 299-312.

35. Xu, Y.; Hu, S.; Chen, R.; Xu, S.; Yu, G.; Ji, L., Interplay between Nrf2 and ROS in regulating epithelial-mesenchymal transition: implications for cancer metastasis and therapy. *Mol Biol Rep* **2025,** *52* (1), 628.

36. Luo, J.; Gong, L.; Yang, Y.; Zhang, Y.; Liu, Q.; Bai, L.; Fang, X.; Zhang, B.; Huang, J.; Liu, M.; Liu, B.; Tang, Y.; Wong, C. N.; Huang, J.; Liu, S.; Li, S.; Ding, T.; Man, K.; Lee, V. H.; Li, Y.; Ma, S.; Guan, X. Y., Enhanced mitophagy driven by ADAR1-GLI1 editing supports the self-renewal of cancer stem cells in HCC. *Hepatology* **2024,** *79* (1), 61-78.

37. Chung, T. W.; Kim, S. J.; Choi, H. J.; Song, K. H.; Jin, U. H.; Yu, D. Y.; Seong, J. K.; Kim, J. G.; Kim, K. J.; Ko, J. H.; Ha, K. T.; Lee, Y. C.; Kim, C. H., Hepatitis B virus X protein specially regulates the sialyl lewis a synthesis among glycosylation events for metastasis. *Mol Cancer* **2014,** *13*, 222.

38. Zhang, X.; Lao, M.; Sun, K.; Yang, H.; He, L.; Liu, X.; Liu, L.; Zhang, S.; Guo, C.; Wang, S.; Shi, J.; Zhang, X.; Xu, D.; Lu, X.; Bai, X.; Liang, T., Sphingolipid synthesis in tumor-associated macrophages confers immunotherapy resistance in hepatocellular carcinoma. *Sci Adv* **2025,** *11* (21), eadv0558.

39. Wang, Z.; Guo, Y.; Hu, K.; He, T.; Qin, T.; Zhang, L.; Xu, F.; Xu, Y.; Cheng, M.; Zhang, J.; Zhao, Q., Hepatocellular carcinoma cells downregulate PGAM2 via SIRT2-mediated deacetylation modification to enhance aerobic glycolysis. *NPJ Precis Oncol* **2025,** *9* (1), 143.

40. Zhou, D.; Duan, Z.; Li, Z.; Ge, F.; Wei, R.; Kong, L., The significance of glycolysis in tumor progression and its relationship with the tumor microenvironment. *Front Pharmacol* **2022,** *13*, 1091779.

41. Xu, Y.; Li, F.; Lv, L.; Li, T.; Zhou, X.; Deng, C. X.; Guan, K. L.; Lei, Q. Y.; Xiong, Y., Oxidative stress activates SIRT2 to deacetylate and stimulate phosphoglycerate mutase. *Cancer Res* **2014,** *74* (13), 3630-42.

42. Zhang, H.; Liang, J.; Lu, T.; Li, M.; Shan, G.; Bi, G.; Zhao, M.; Jin, X.; Wang, Q.;

ARTICLE IN PRESS

ARTICLE IN PRESS

Chen, Z.; Zhan, C., AGRN promotes lung adenocarcinoma progression by activating Notch signaling pathway and acts as a therapeutic target. *Pharmacol Res* **2023,** *194*, 106819.

43. Tao, J.; Shen, L.; Zhuang, M.; Zhai, C.; Zeng, H.; Mao, Y.; Liu, X., Suppression of AGRN enhances CD8+ T cell recruitment and inhibits breast cancer progression. *FASEB J* **2024,** *38* (7), e23582.

44. Zhou, M.; Lv, S.; Hou, Y.; Zhang, R.; Wang, W.; Yan, Z.; Li, T.; Gan, W.; Zeng, Z.; Zhang, F.; Yang, M., Characterization of sialylation-related long noncoding RNAs to develop a novel signature for predicting prognosis, immune landscape, and chemotherapy response in colorectal cancer. *Front Immunol* **2022,** *13*, 994874.

45. Su, Z.; He, Y.; You, L.; Zhang, G.; Chen, J.; Liu, Z., Coupled scRNA-seq and Bulk-seq reveal the role of HMMR in hepatocellular carcinoma. *Front Immunol* **2024,** *15*, 1363834.

46. Smelik, M.; Diaz-Roncero Gonzalez, D.; An, X.; Heer, R.; Henningsohn, L.; Li, X.; Wang, H.; Zhao, Y.; Benson, M., Combining Spatial Transcriptomics, Pseudotime, and Machine Learning Enables Discovery of Biomarkers for Prostate Cancer. *Cancer Res* **2025,** *85* (13), 2514-2526.

47. Aibar, S.; Gonzalez-Blas, C. B.; Moerman, T.; Huynh-Thu, V. A.; Imrichova, H.; Hulselmans, G.; Rambow, F.; Marine, J. C.; Geurts, P.; Aerts, J.; van den Oord, J.; Atak, Z. K.; Wouters, J.; Aerts, S., SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* **2017,** *14* (11), 1083-1086.

48. Jin, S.; Plikus, M. V.; Nie, Q., CellChat for systematic analysis of cell-cell communication from single-cell transcriptomics. *Nat Protoc* **2025,** *20* (1), 180-219.

49. Ferreira, M. R.; Santos, G. A.; Biagi, C. A.; Silva Junior, W. A.; Zambuzzi, W. F., GSVA score reveals molecular signatures from transcriptomes for biomaterials comparison. *J Biomed Mater Res A* **2021,** *109* (6), 1004-1014.

50. Liu, Y.; Jiang, C.; Xu, C.; Gu, L., Systematic analysis of integrated bioinformatics to identify upregulated THBS2 expression in colorectal cancer cells inhibiting tumour immunity through the HIF1A/Lactic Acid/GPR132 pathway. *Cancer Cell Int* **2023,** *23* (1), 253.

51. Zeng, D.; Ye, Z.; Shen, R.; Yu, G.; Wu, J.; Xiong, Y.; Zhou, R.; Qiu, W.; Huang, N.; Sun, L.; Li, X.; Bin, J.; Liao, Y.; Shi, M.; Liao, W., IOBR: Multi-Omics Immuno-Oncology Biological Research to Decode Tumor Microenvironment and Signatures. *Front Immunol* **2021,** *12*, 687975.

52. Ritchie, M. E.; Phipson, B.; Wu, D.; Hu, Y.; Law, C. W.; Shi, W.; Smyth, G. K., limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **2015,** *43* (7), e47.

53. Wu, T.; Hu, E.; Xu, S.; Chen, M.; Guo, P.; Dai, Z.; Feng, T.; Zhou, L.; Tang, W.; Zhan, L.; Fu, X.; Liu, S.; Bo, X.; Yu, G., clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)* **2021,** *2* (3), 100141.

54. Guo, Z.; Feng, H.; Jia, W.; Zhang, T.; Chen, M.; Xu, Z.; Zheng, F.; Huang, C.; Li, Z.; Liu, K.; Zhang, T.; Zhao, R.; Zhang, W.; Cheng, X., Tertiary lymphoid structures signature predicts prognosis and clinical benefits from neoadjuvant chemotherapy and PD-1 blockade in colorectal adenocarcinoma. *Int J Surg* **2025,** *111* (8), 5088-5104.

55. Ding, Y.; Zhao, Z.; Cai, H.; Zhou, Y.; Chen, H.; Bai, Y.; Liu, Z.; Liu, S.; Zhou, W., Single-cell sequencing analysis related to sphingolipid metabolism guides immunotherapy and prognosis of skin cutaneous melanoma. *Front Immunol* **2023,** *14*, 1304466.

ARTICLE IN PRESS

Figure 1. Single-cell Expression Atlas of HCC

(A-B) UMAP plots showing the clustering of single cells from HCC tumors into eight distinct cell types: hepatocytes, fibroblasts, endothelial cells, T/NK cells, B cells, plasma cells, myeloid cells, and MAST cells.

(C) Proportion of different cell types across the G1, S, and G2M phases of the cell cycle.

(D) Proportion of cell types across early and advanced tumor stages.

(E-F) Expression of cell type-specific markers in each cluster shown in bubble plots and UMAP plots.

(G) Distribution of cell cycle scores and RNA metrics across different cell types.

(H) KEGG enrichment analysis for cell type-specific markers.

ARTICLE IN PRESS

Figure 2. Tumor Cell Analysis and Malignant Subpopulation Identification

(A-C) UMAP plots of tumor cells showing five distinct malignant subpopulations.

(D-E) Proportion of tumor cells from different stages (early vs advanced) within each malignant cluster.

(F) Ro/e heatmap showing the association between stages and tumor subpopulations.

(G) Bubble plots of marker expression across malignant cell clusters.

(H) Volcano plots showing differential gene expression between tumor subpopulations.

(I-J) Comparison of CNV scores between different subpopulations.

(K-M) GO enrichment analysis of differentially expressed genes in malignant

ARTICLE IN PRESS

subpopulations. (L) Shared enrichment results across all tumor subpopulations. (M) Unique enrichment results for each tumor subpopulation.

(N) ssGSEA scores for KEGG metabolic pathways in malignant subpopulations.

Figure 3. Trajectory Analysis of Tumor Cell Subpopulations

(A) Monocle2 trajectory plot showing the progression of tumor cells along a pseudotime axis.

(B-C) Distribution of different tumor cell subpopulations across the pseudotime axis in Monocle2, with each subpopulation colored accordingly.

(D) Monocle2 trajectory plot colored by state.

(E) Proportion of cell subpopulations within each state along the pseudotime trajectory.

(F) CytoTRACE analysis of tumor cell differentiation potential.

(G-I) Slingshot trajectory analysis, showing two distinct differentiation pathways colored by tumor stage, cell type, and state.

(J) Expression of top markers in each subpopulation across pseudotime.

(K) Heatmap of genes correlated with pseudotime.

Figure 4. Cell-Cell Communication Analysis in HCC

(A-C) Cell-cell communication network maps for tumor cells, showing interactions between C1 (malignant) and C4 cells with other cell types. Line thickness represents the number of interactions, and weight indicates the strength of communication.

(D) Bubble plots of incoming and outgoing signaling pathways for each cell type.

ARTICLE IN PRESS

(E-F) MK signaling pathway communication heatmap.

(G) Bubble plot showing the communication from C1 malignant cells to other cell types.

(H) Bubble plot showing the expression of key ligand genes in different cell types.

(I) Cell-cell communication network between MDK and NCL receptors.

Figure 5. Transcription Factor Activity in Tumor Cell Subpopulations

(A) Heatmap of transcription factor activity in tumor cell subpopulations, based on CSI matrix and clustering of regulatory modules (M1, M2, M3).

(B) UMAP plot showing the distribution of five distinct tumor subpopulations.

(C) UMAP plot showing the distribution of early and advanced stage tumor cells.

(D-E) AUC scores for transcription factors in each cluster.

(F) Scatter plot showing RSS for different transcription factors in the five tumor subpopulations.

(G) Scatter plot showing RSS for different transcription factors in early and advanced tumor subpopulations.

(H-I) Visualization of four transcription factors (PGAM2, ELF3, XBP1, JUNB) using AUC bar plots and UMAP plots.

Figure 6. Prognostic Model Based on PGAM2-Regulated Genes and Sialylation

(A) Forest plot showing the results of univariate Cox regression analysis for 12 genes associated with survival, selected from the PGAM2-related transcription factor TARGET gene chain (importance > 10) and sialylation-related genes.

ARTICLE IN PRESS

(B) LASSO-Cox regression analysis to select prognostic genes, leading to the identification of a risk model based on 9 genes.

(C) Multivariate Cox regression analysis and forest plot showing the hazard ratios for each prognostic gene.

(D) Bar plot displaying the Coef of the 9 prognostic genes used to calculate risk scores.

(E-F) Risk score vs survival time scatter plot and heatmap showing the expression of prognostic genes across different risk groups.

(G) PCA plot showing dimensionality reduction of prognostic genes, visualizing the separation between high- and low-risk groups.

(H) Time-dependent ROC curves for 1, 3, and 5 years, demonstrating the predictive accuracy of the risk model.

(I) Kaplan-Meier survival curve comparing high-risk and low-risk groups.

Figure 7. Functional Analysis of High- and Low-Risk Groups

(A) Heatmap of the top 30 differentially expressed genes between high-risk and low-risk groups, sorted by logFC.

(B) Volcano plot showing the differential gene expression analysis of high- and low-risk groups.

(C-D) KEGG and GO enrichment analysis of upregulated genes in the high-risk group.

(E) Heatmap of pathway scores from msigdb.

(F) GSEA results for five key biological pathways.

ARTICLE IN PRESS

Figure 8. Immune Infiltration, Mutation Landscape, and Drug Sensitivity

(A-B) Bar and box plots showing immune cell composition predicted by the CIBERSORT algorithm in high-risk and low-risk groups.

(C) Correlation between immune cell infiltration and risk scores.

(D) Heatmap showing the correlation between immune cell infiltration and prognostic genes.

(E) TIDE analysis showing significantly higher scores in the high-risk group.

(F) Heatmap of immune infiltration levels in HCC based on ESTIMATE, CIBERSORT, and xCell algorithms.

(G) Mutation waterfall plot comparing the top 20 mutations in high- and low-risk groups.

(H) Correlation analysis between immune checkpoint genes and prognostic genes.

(I) Box plots comparing drug sensitivity between high-risk and low-risk groups for 12 chemotherapy agents.

Figure 9. Expression and functional analysis of AGRN in HCC.

(A) Relative AGRN mRNA expression in HCC tumors and adjacent normal tissues.

(B) AGRN mRNA levels in LO2 and various HCC cell lines.

(C) qRT-PCR validation of AGRN knockdown efficiency in Hep 3B and HuH-6 cells post-siRNA transfection.

(D–E) CCK-8 assays evaluating cell proliferation over 4 days in control and AGRN knockdown groups (n = 5 per group).

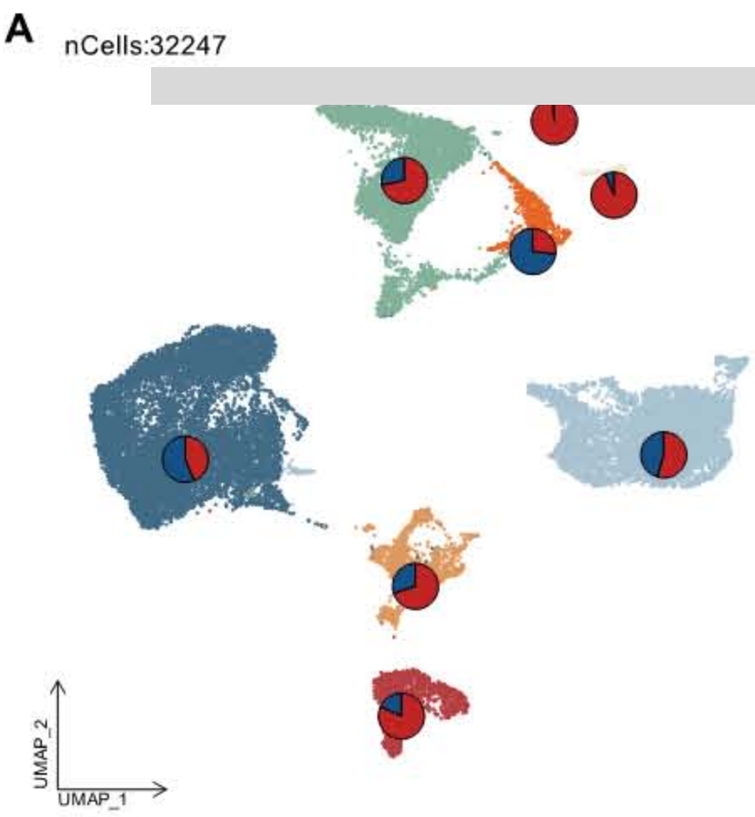
(F–G) Flow cytometric analysis of apoptosis in AGRN-silenced Hep 3B cells.

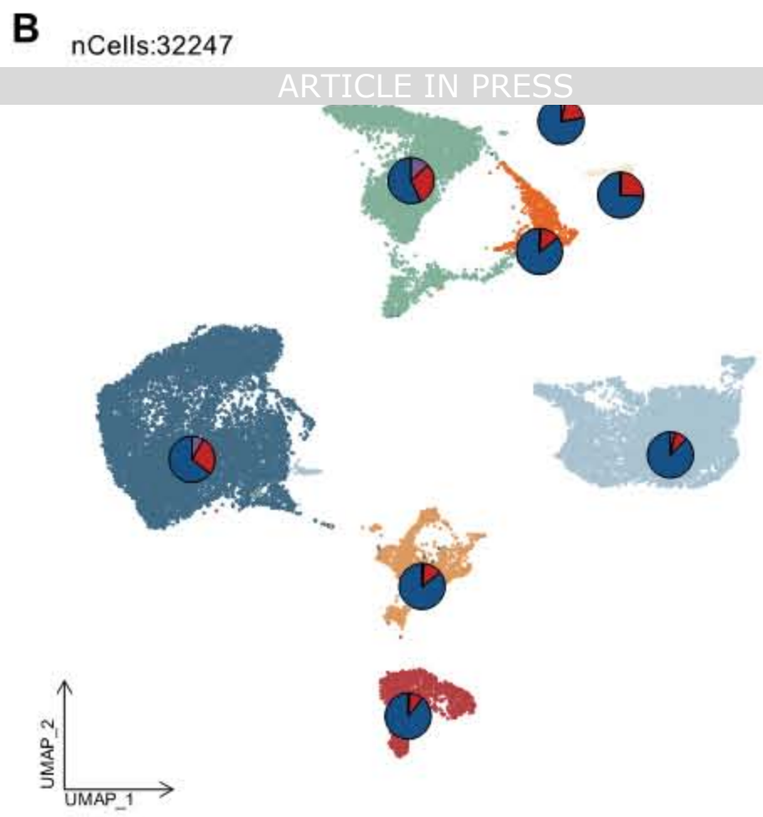
ARTICLE IN PRESS

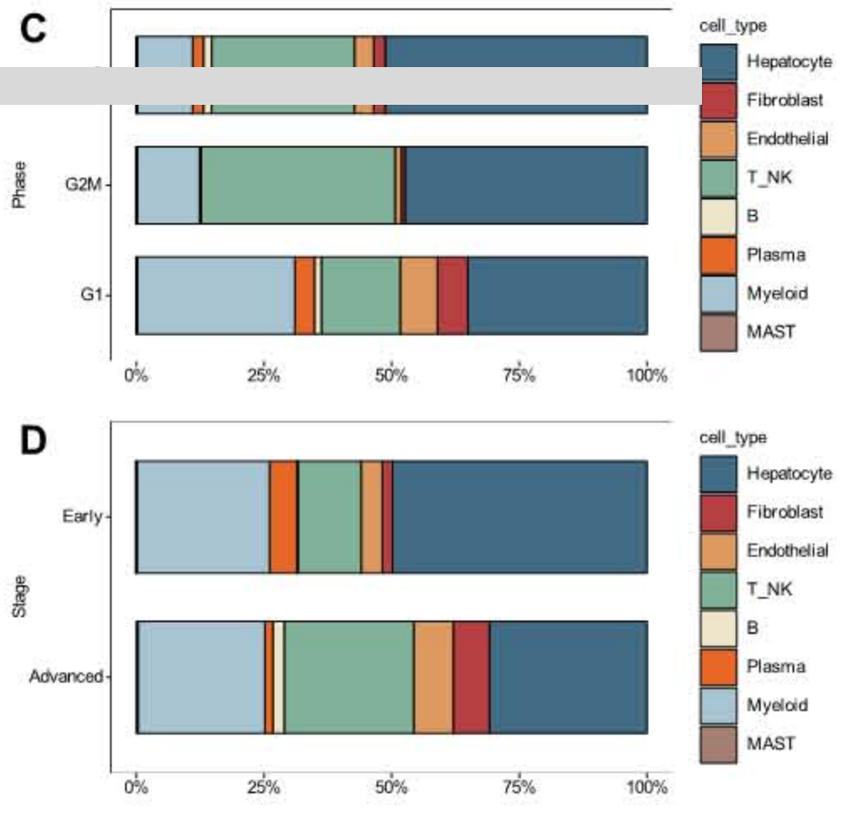
(H–I) Transwell migration and invasion assays in Hep 3B cells with AGRN knockdown.

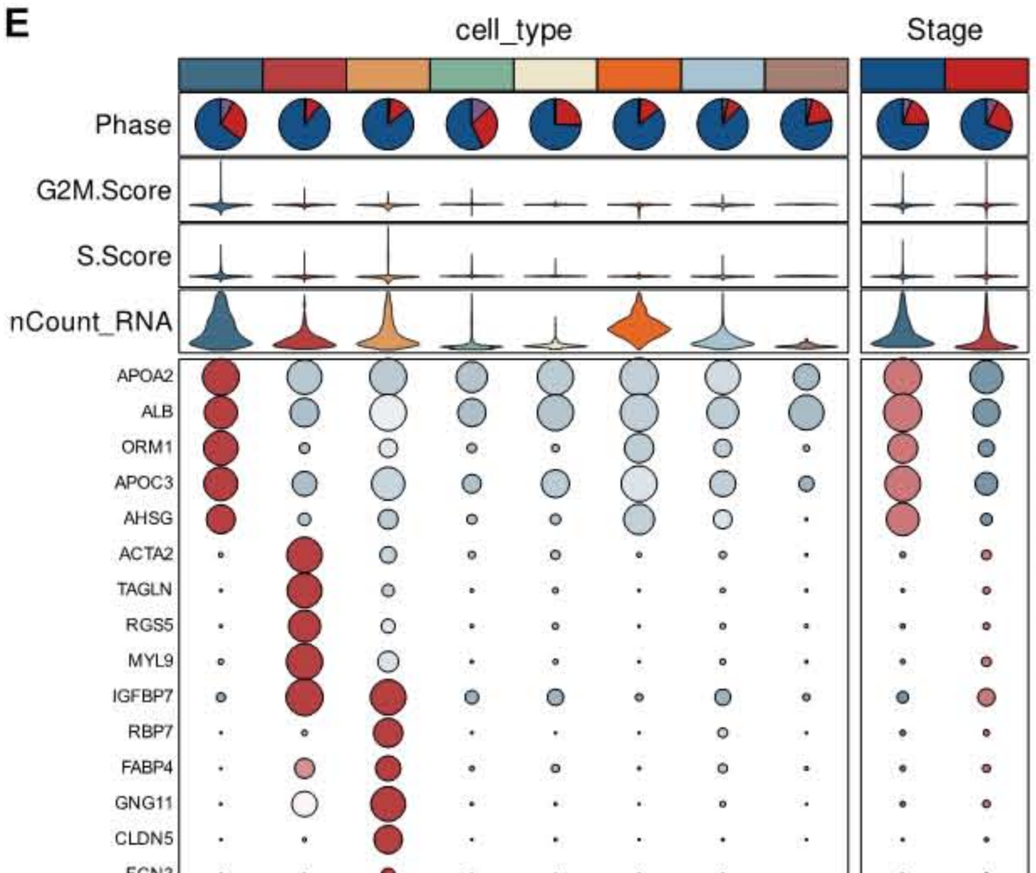
ARTICLE IN PRESS

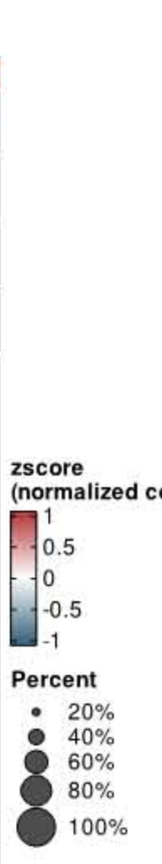
ARTICLE IN PRESS

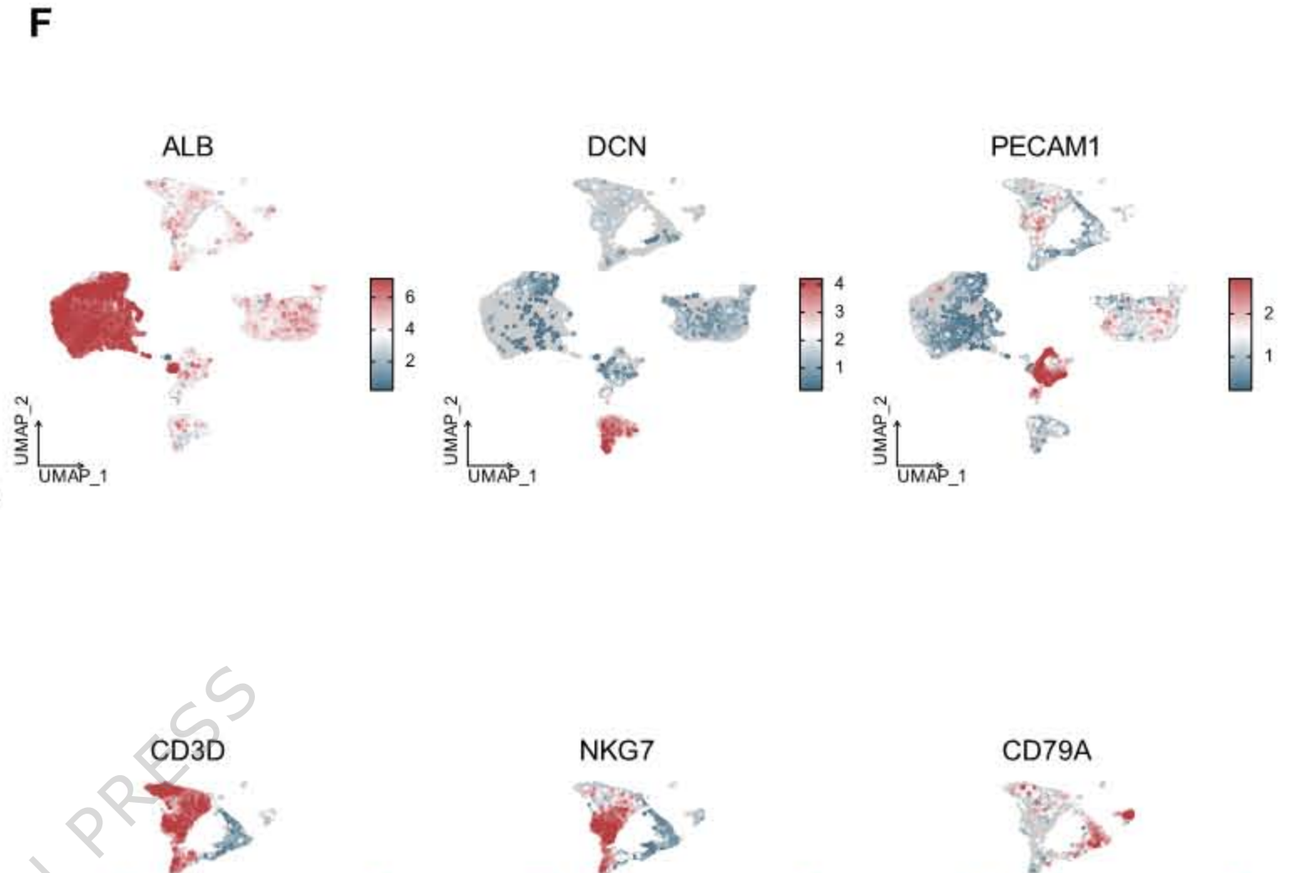


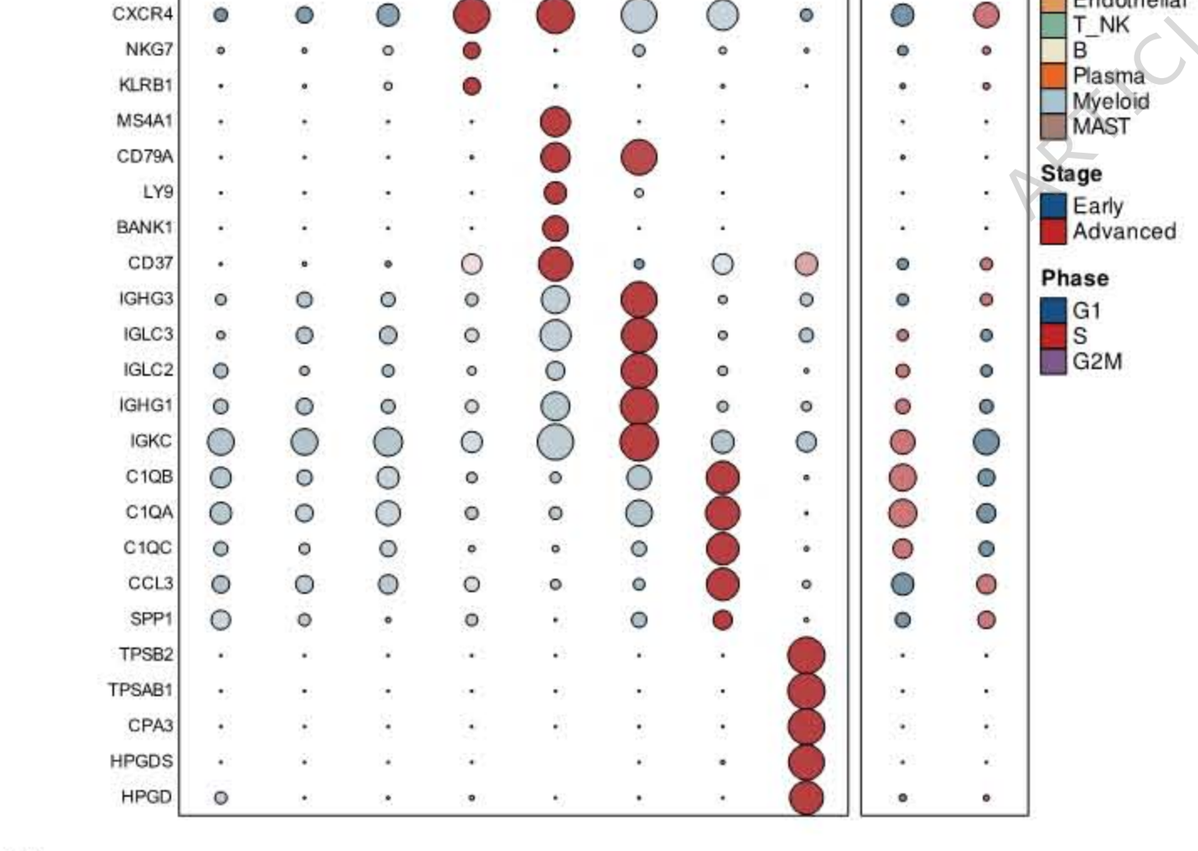


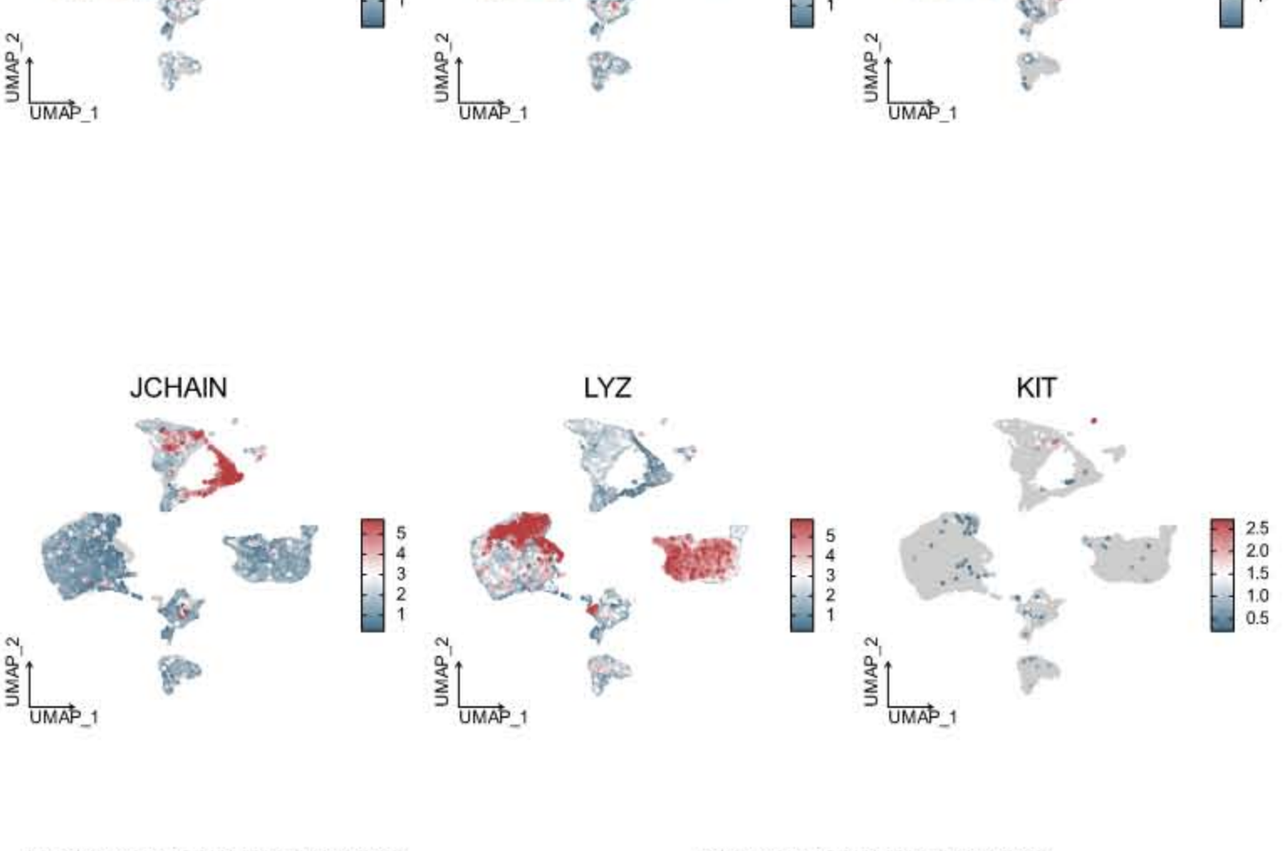




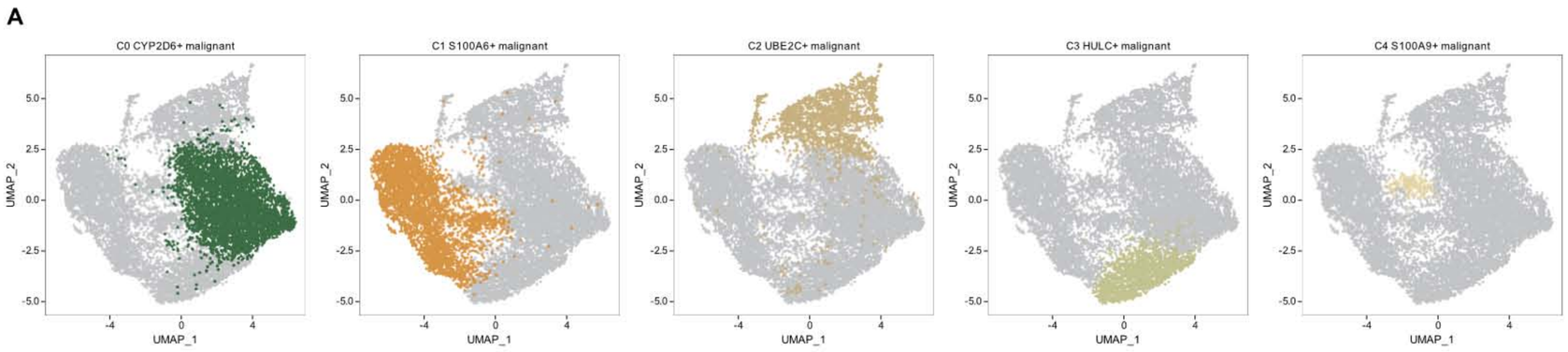


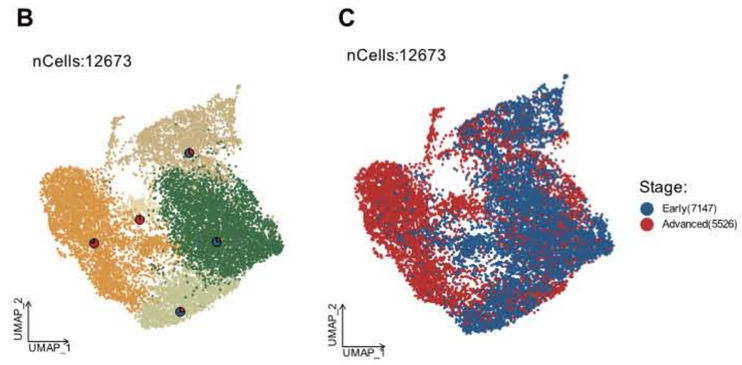


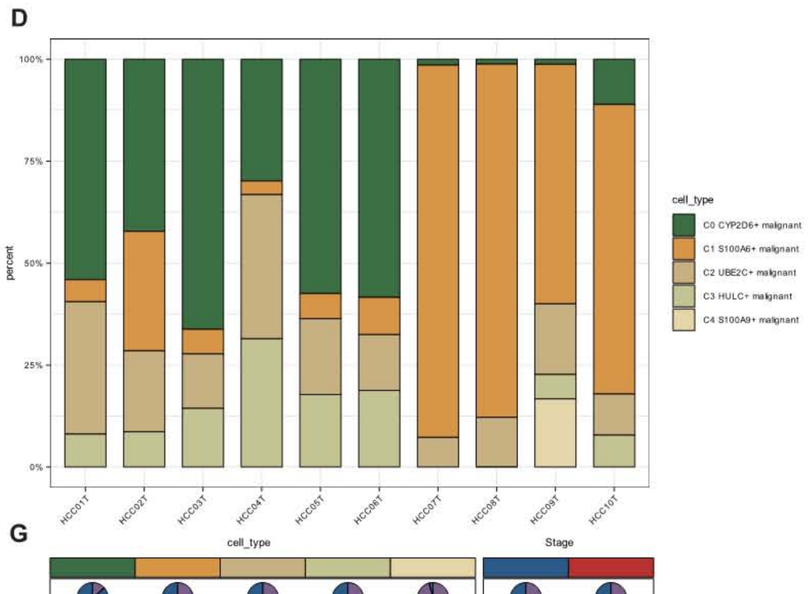


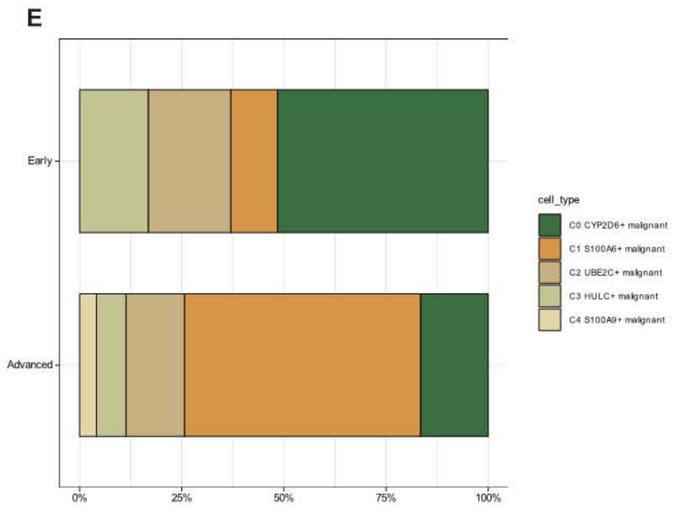


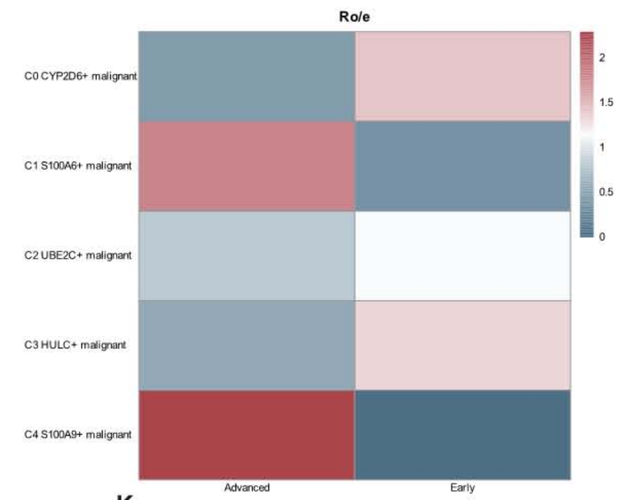
ARTICLE IN PRESS



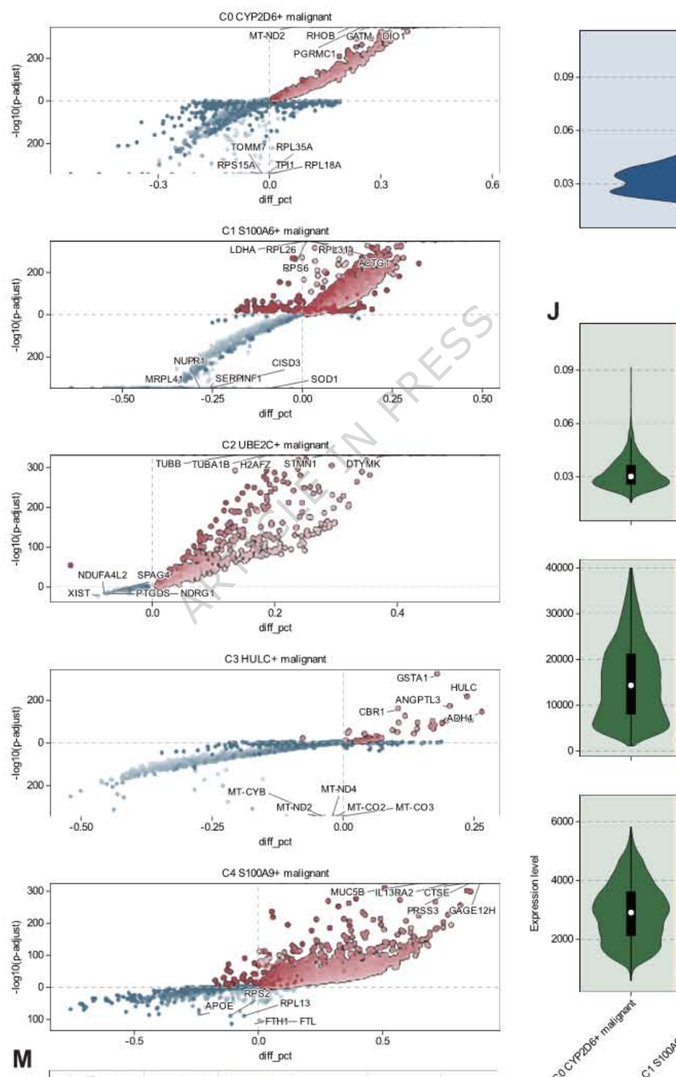


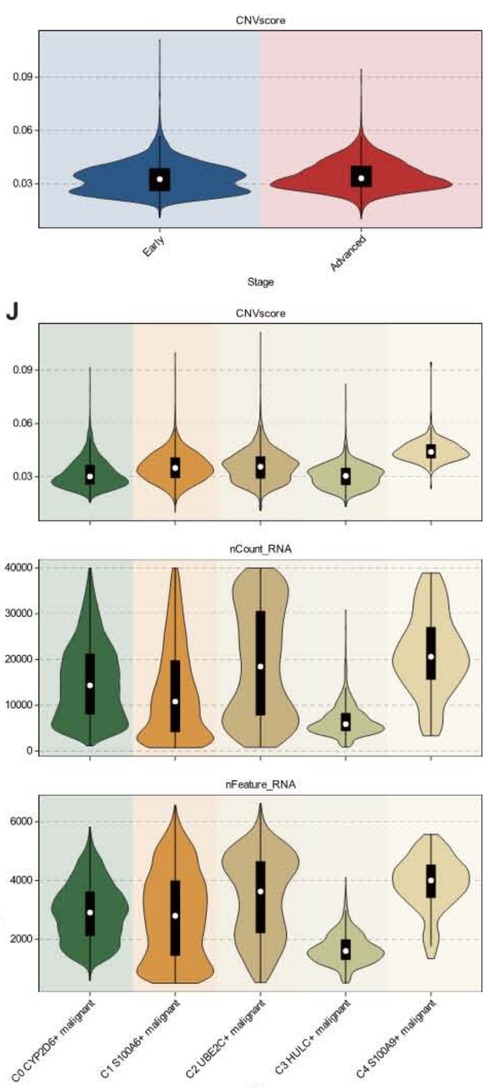


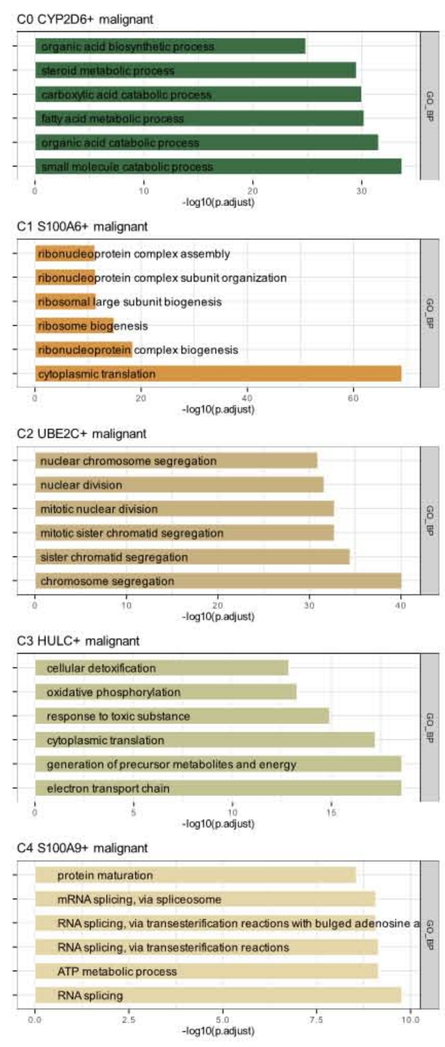


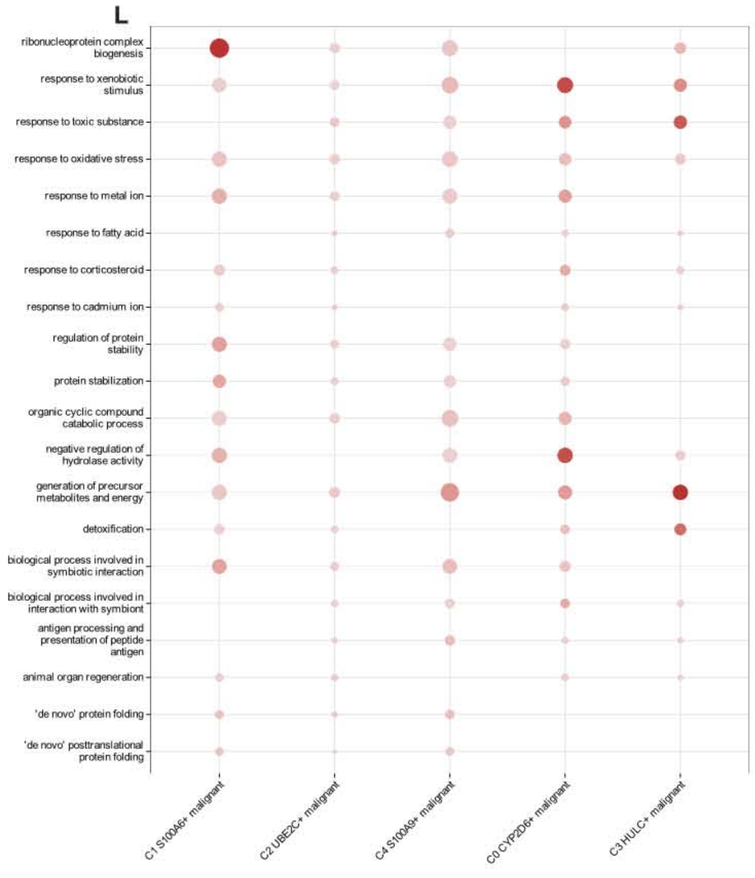


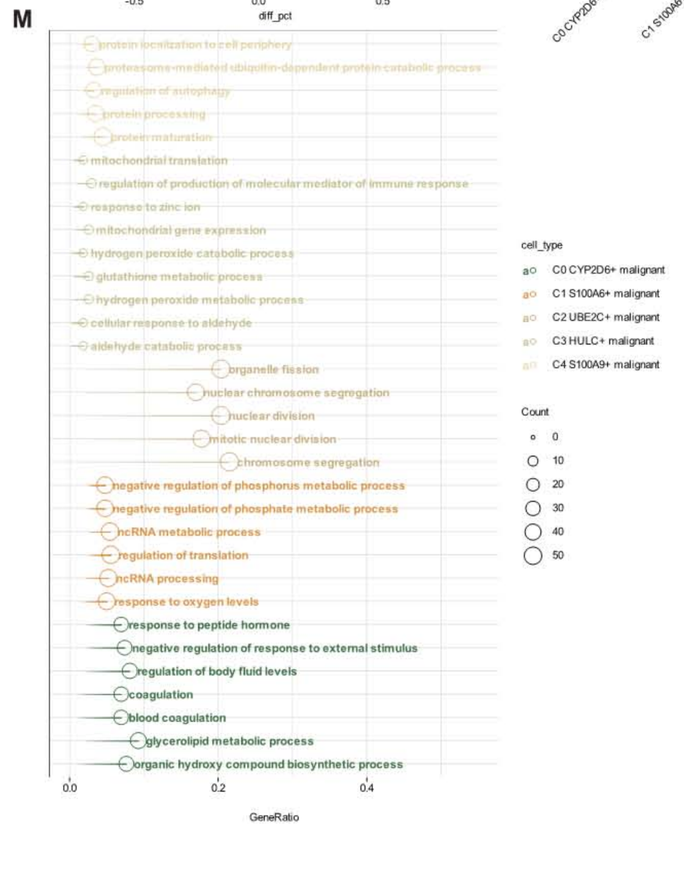


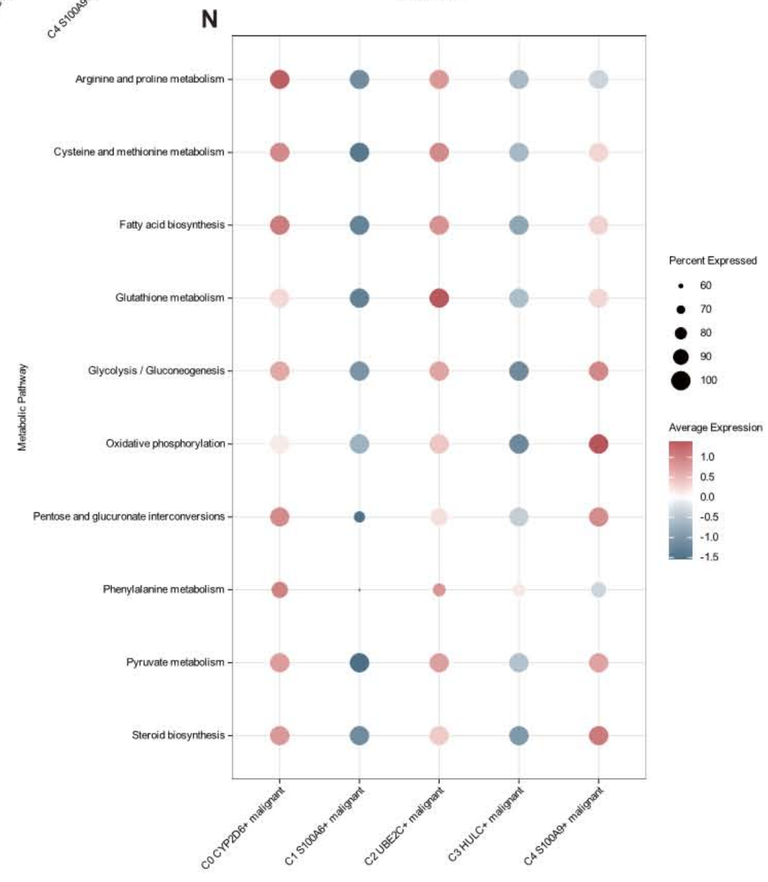




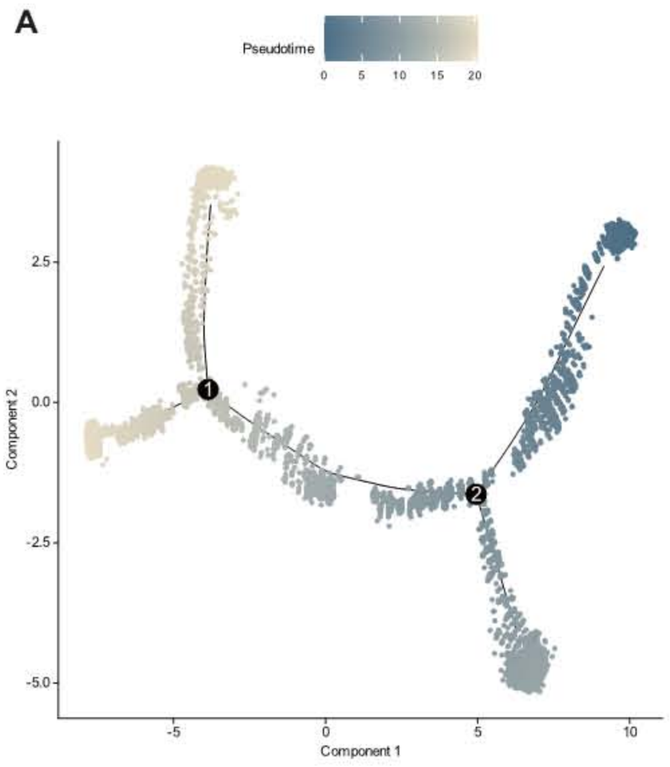


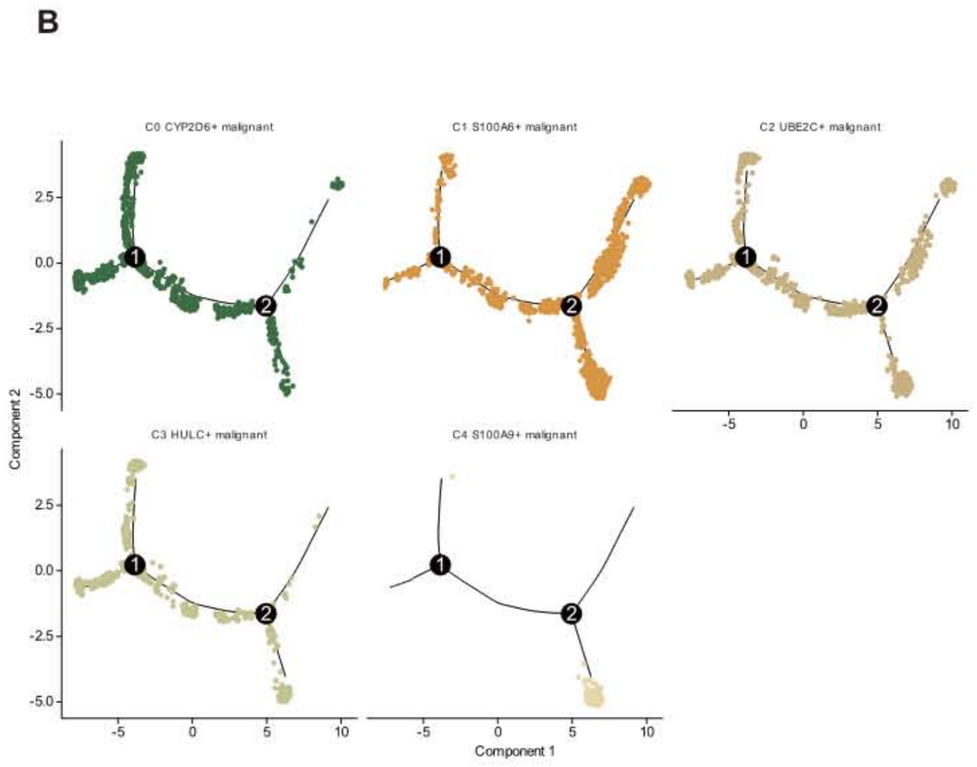


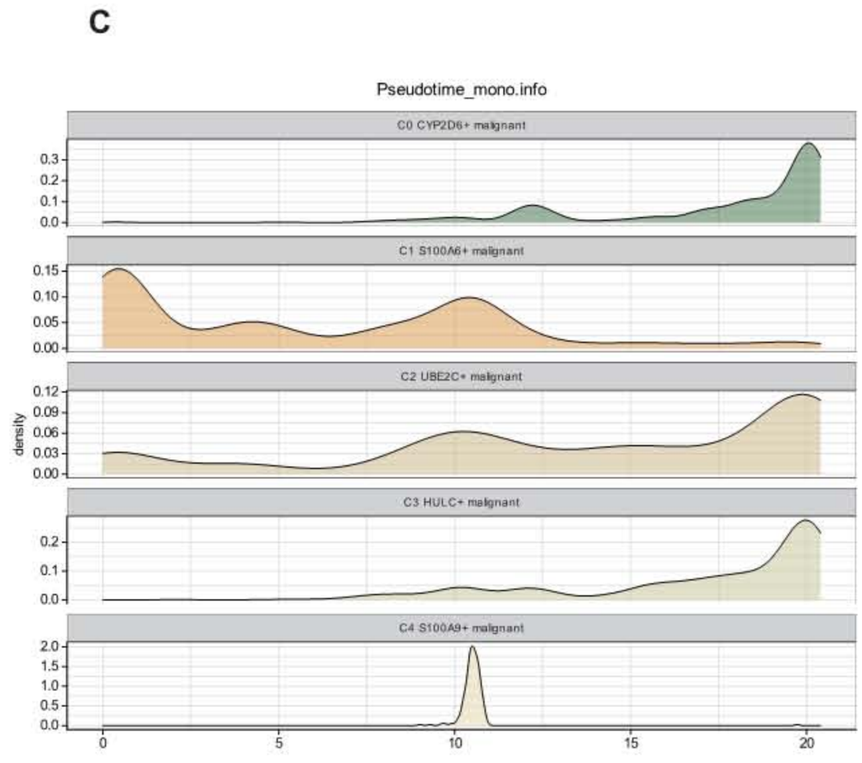


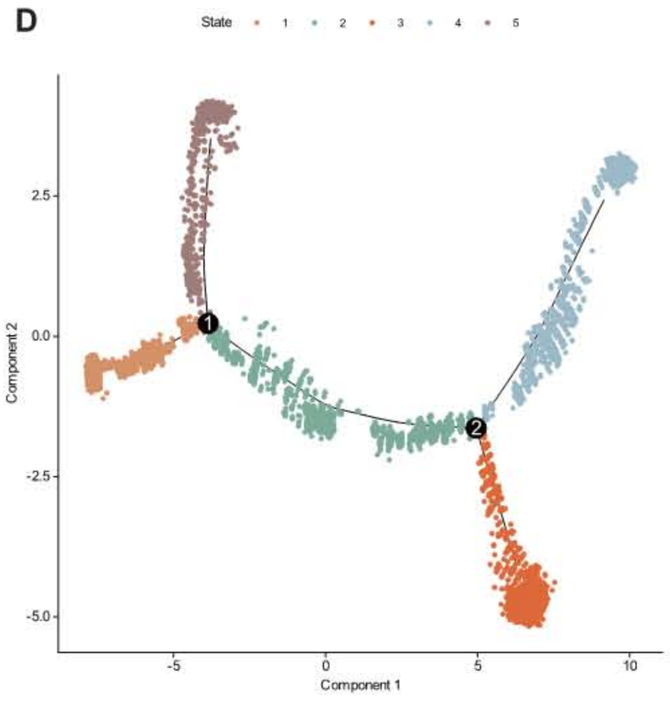


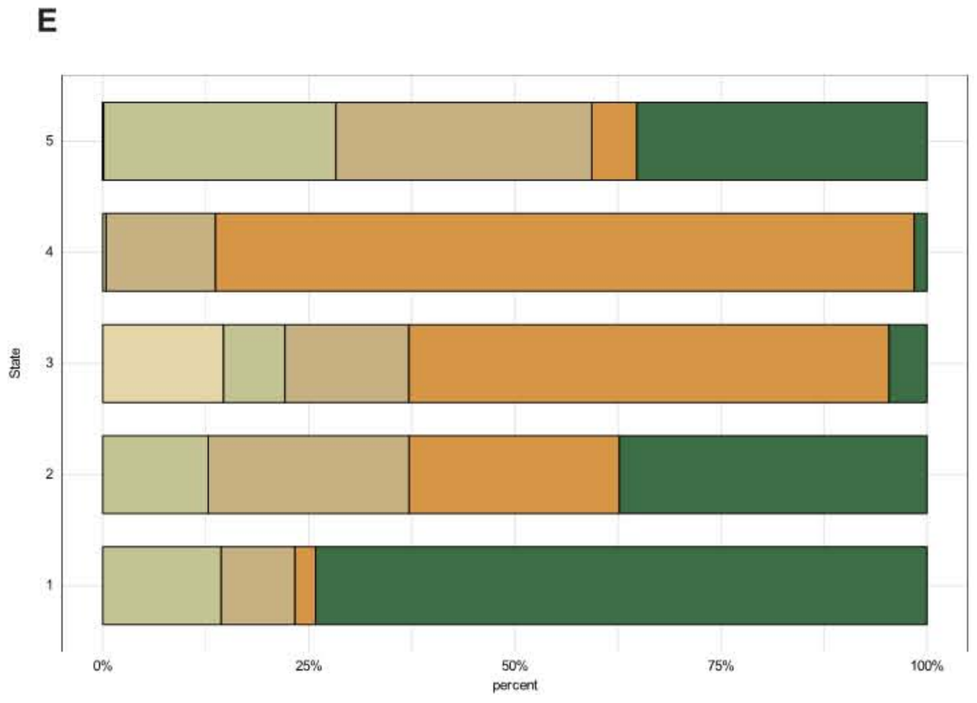
ARTICLE IN PRESS

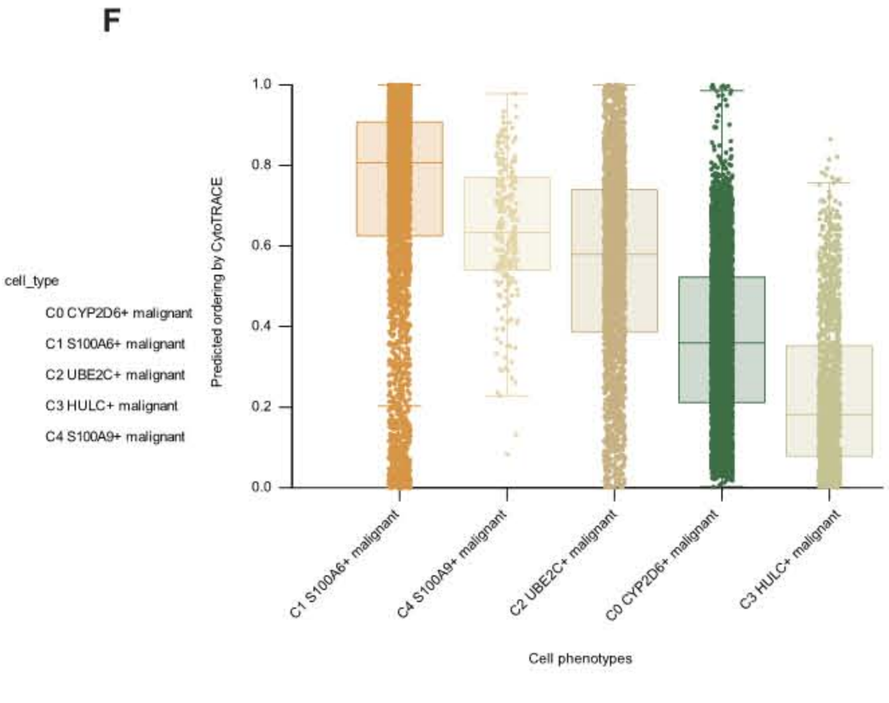


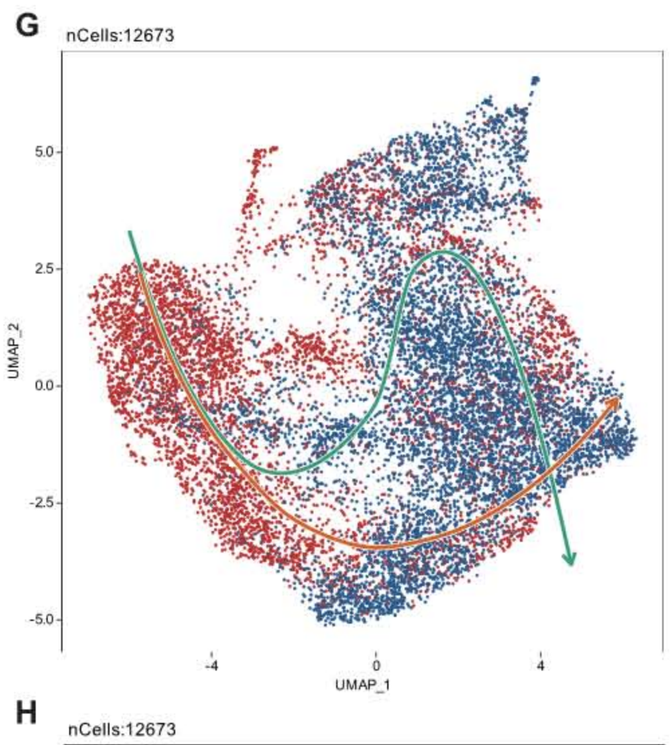




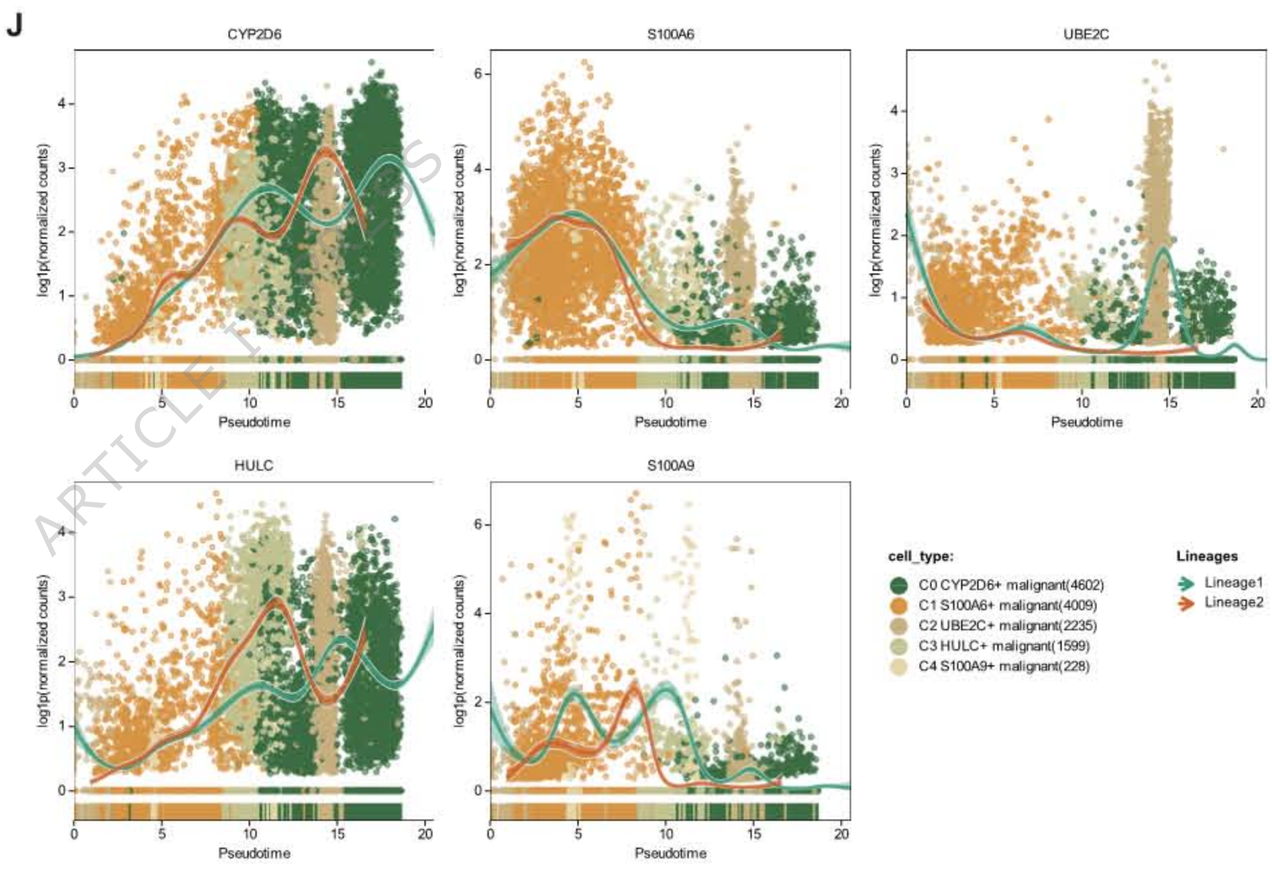


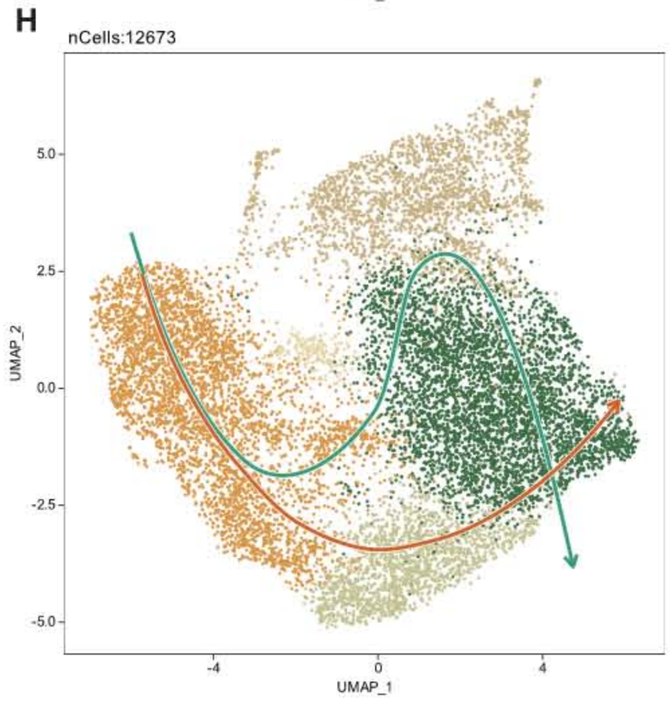


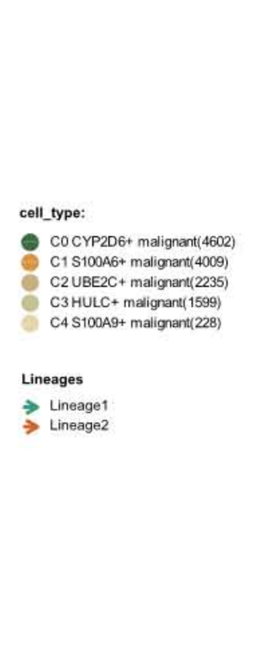


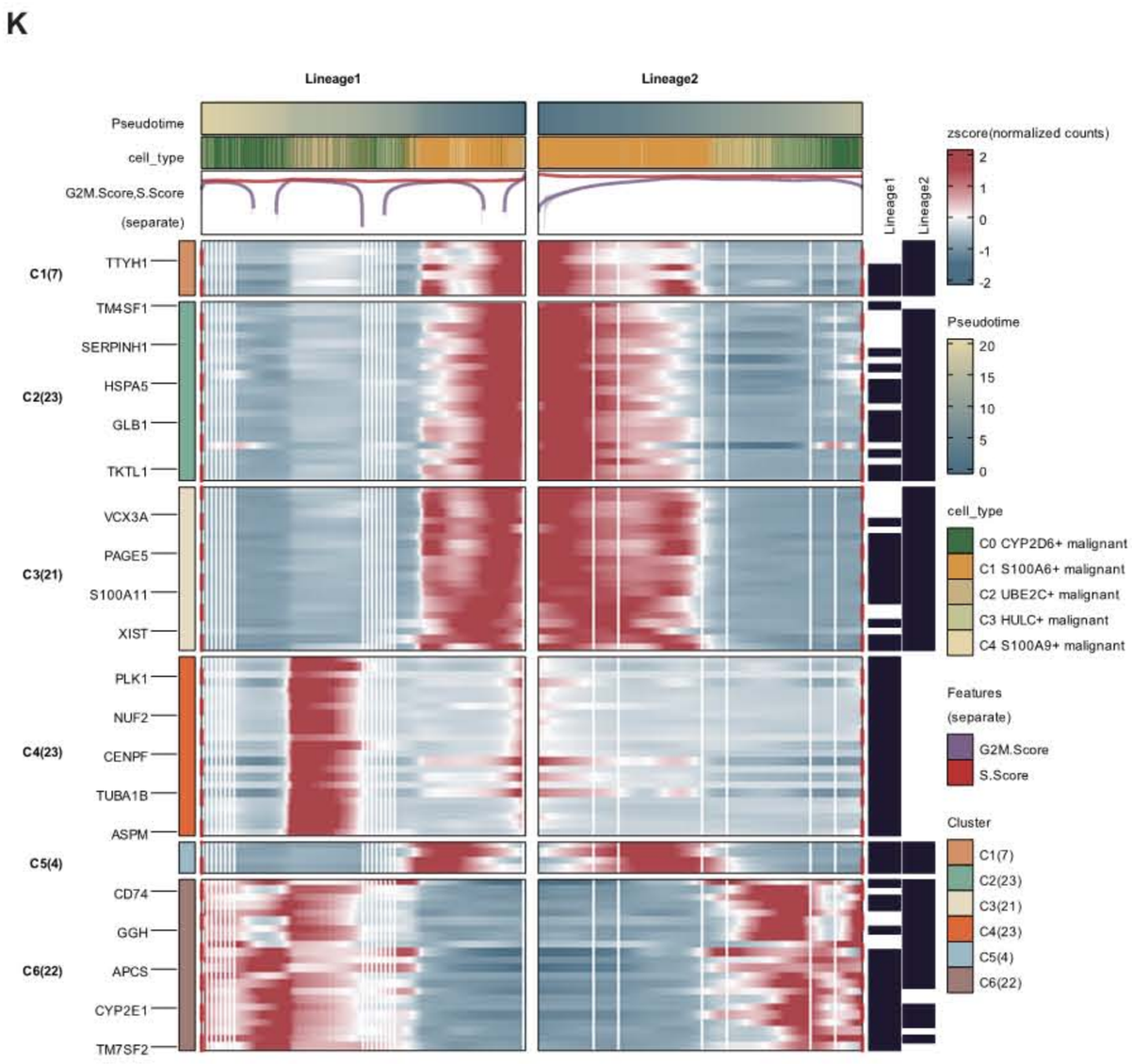


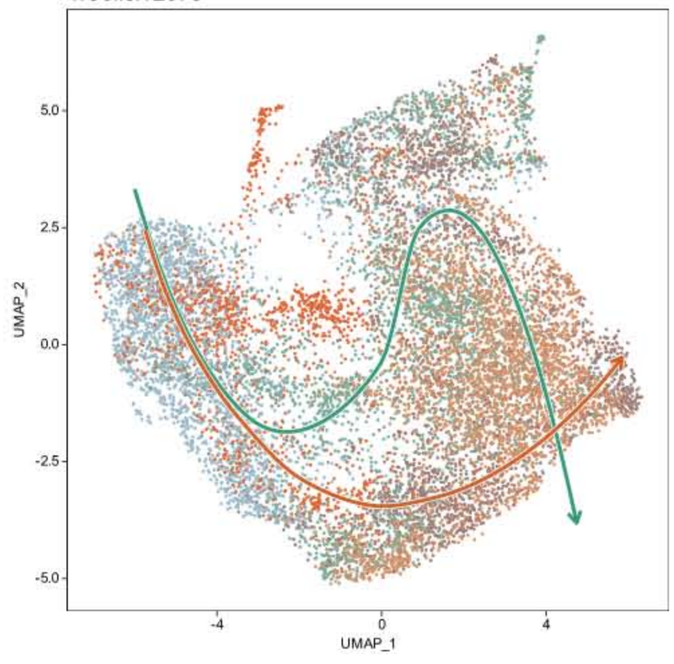


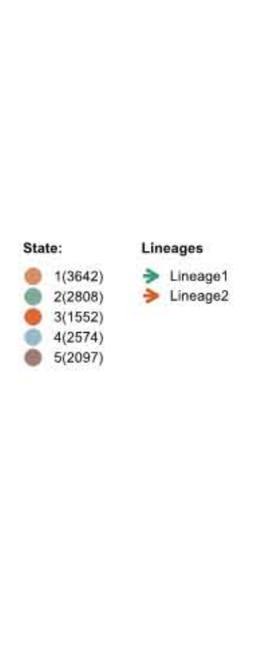






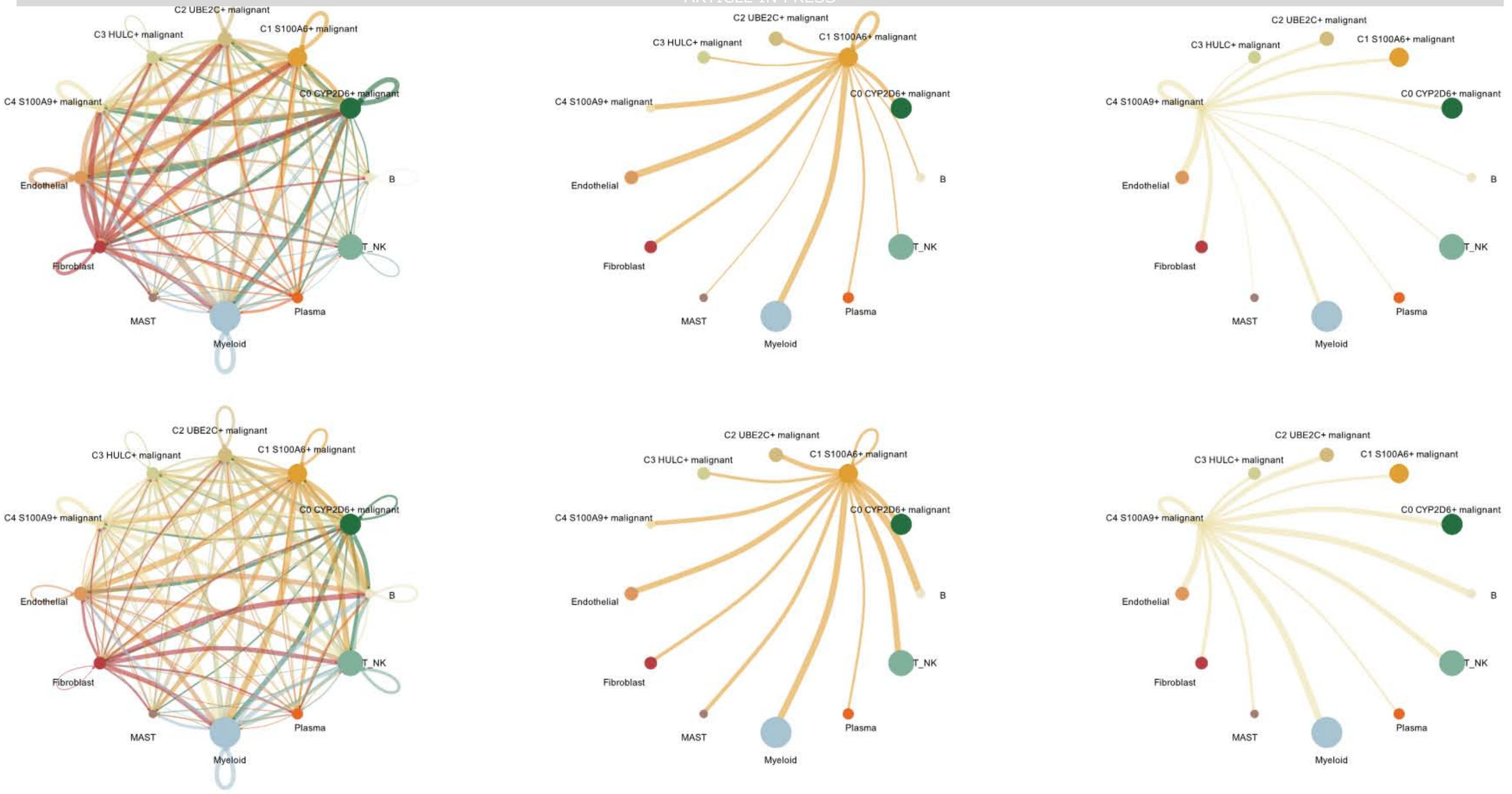




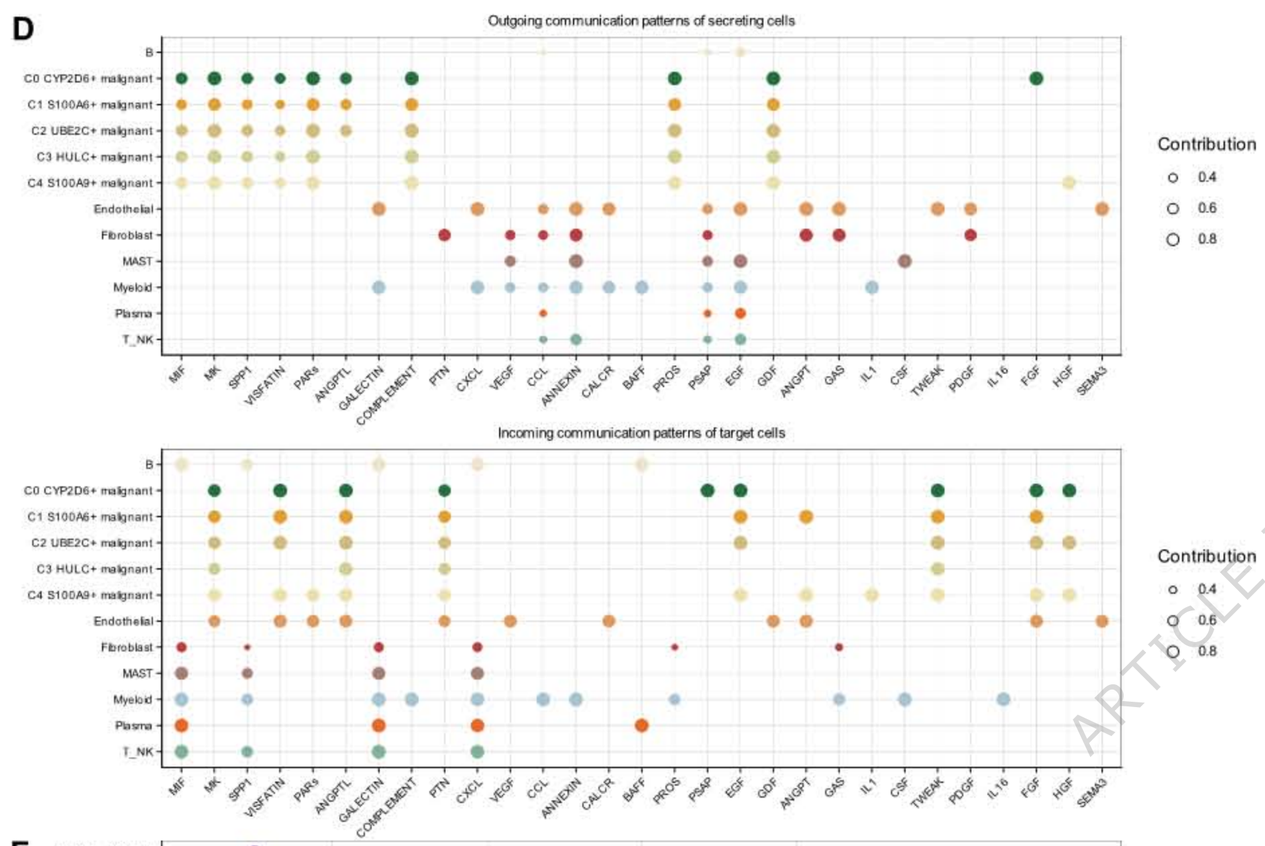


ARTICLE IN PRESS

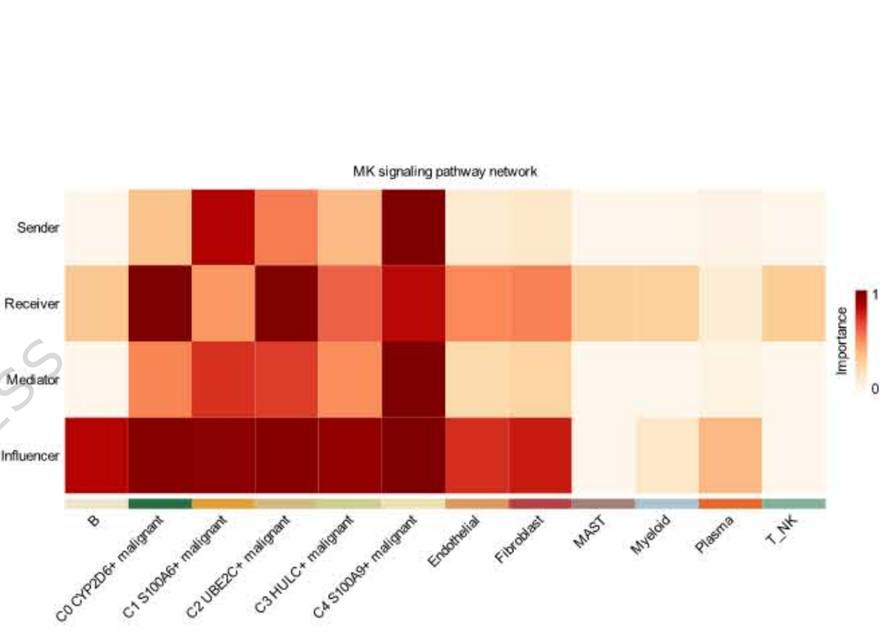
A



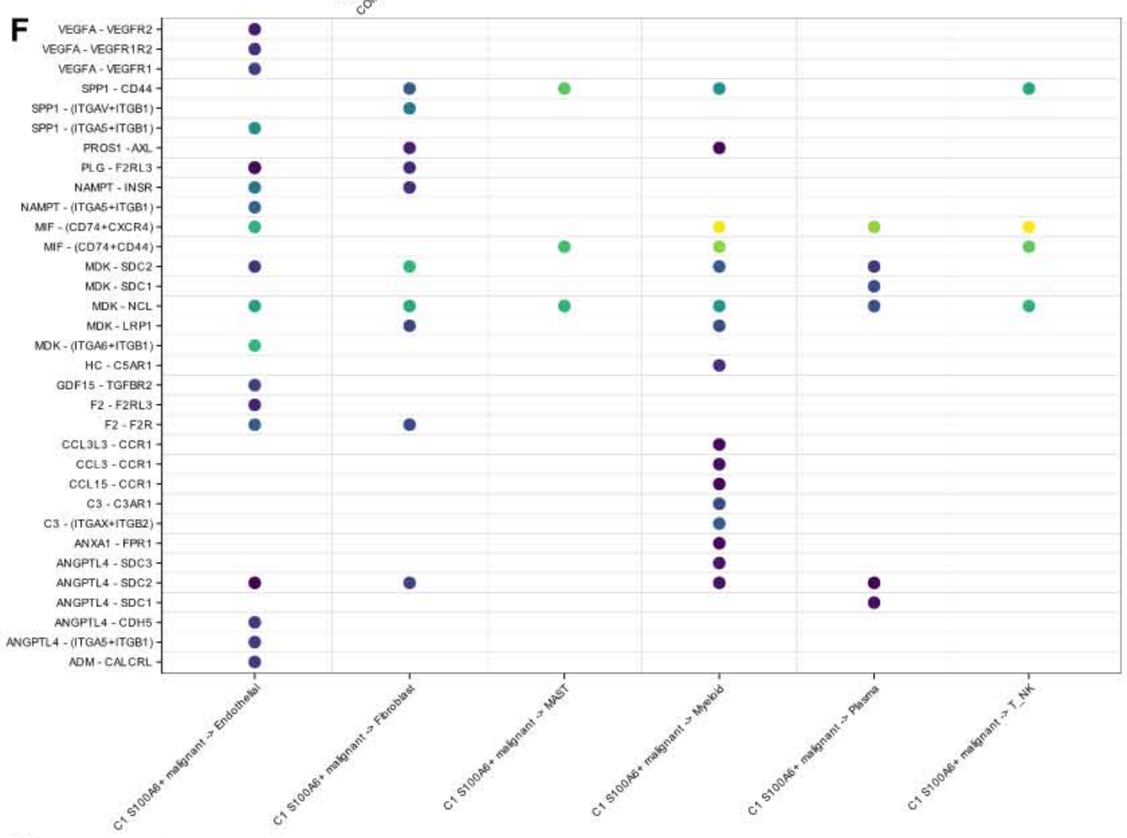
D



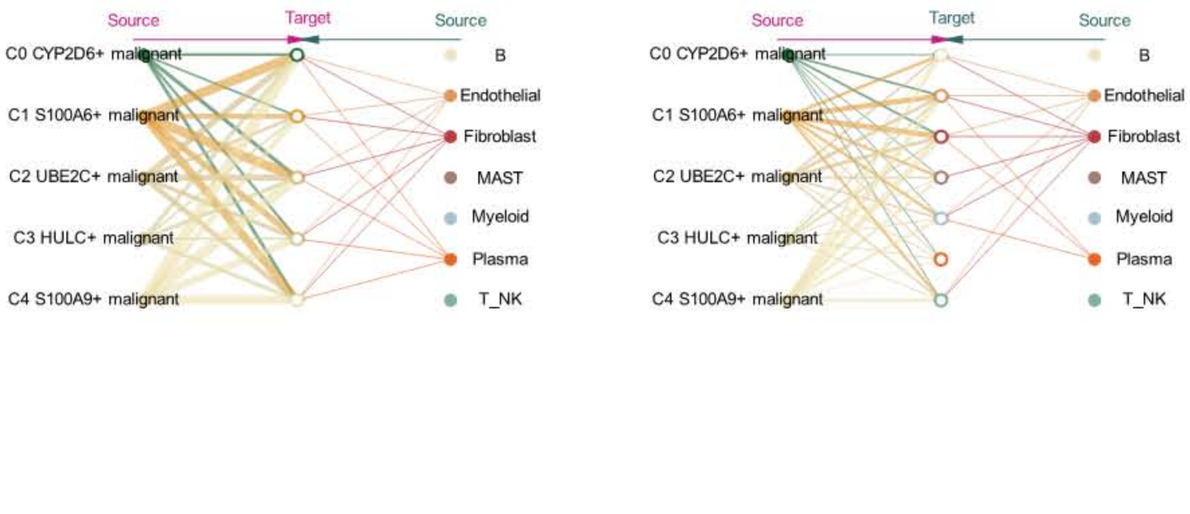
E



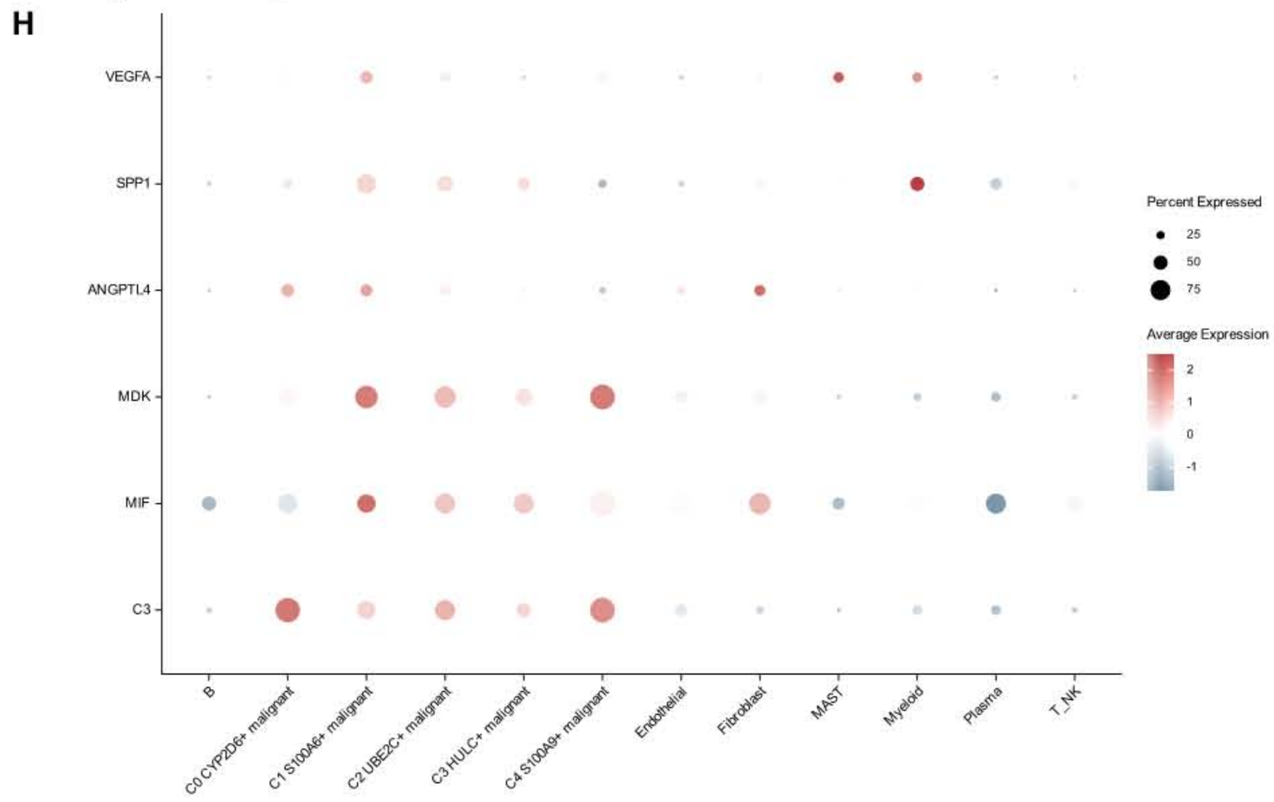
F



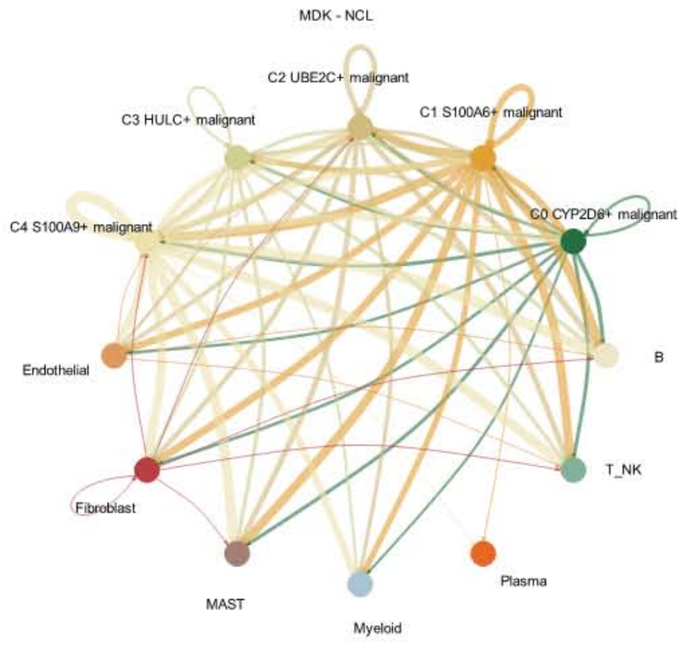
G



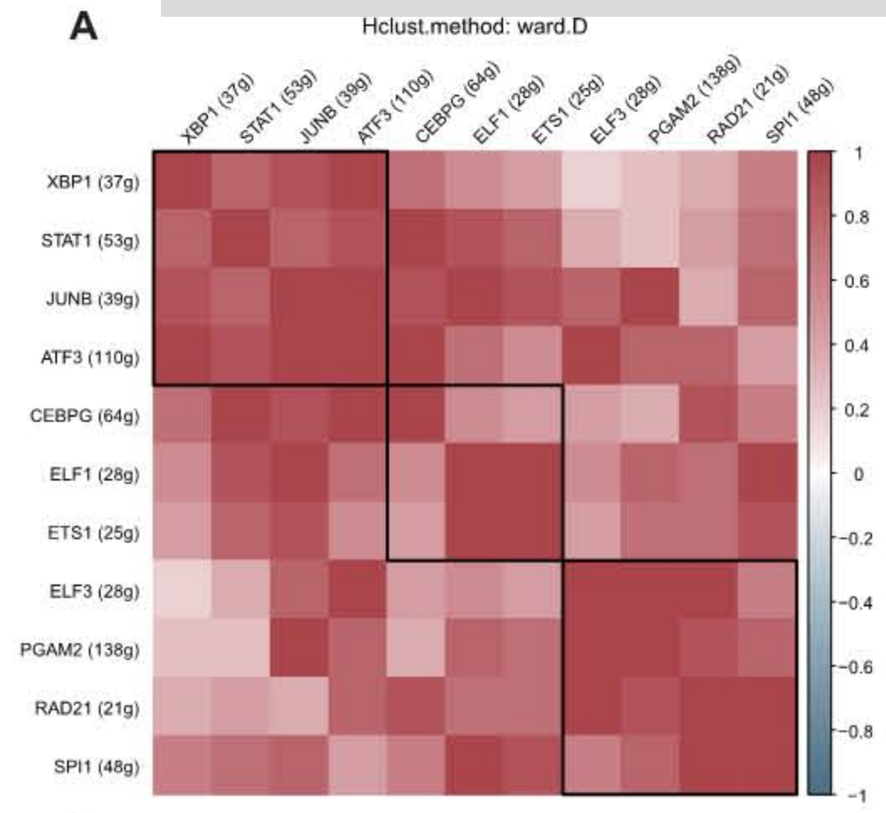
H

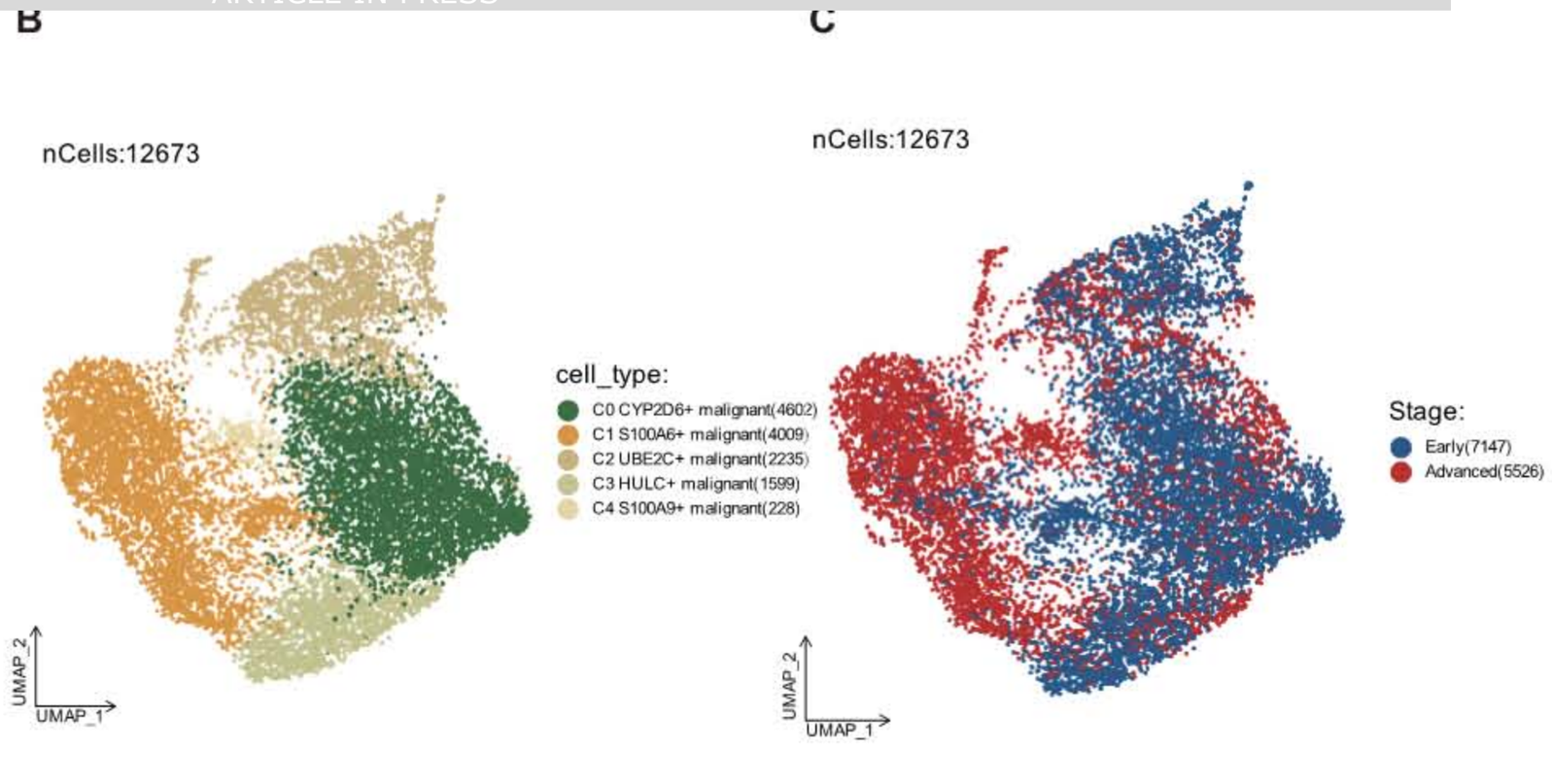


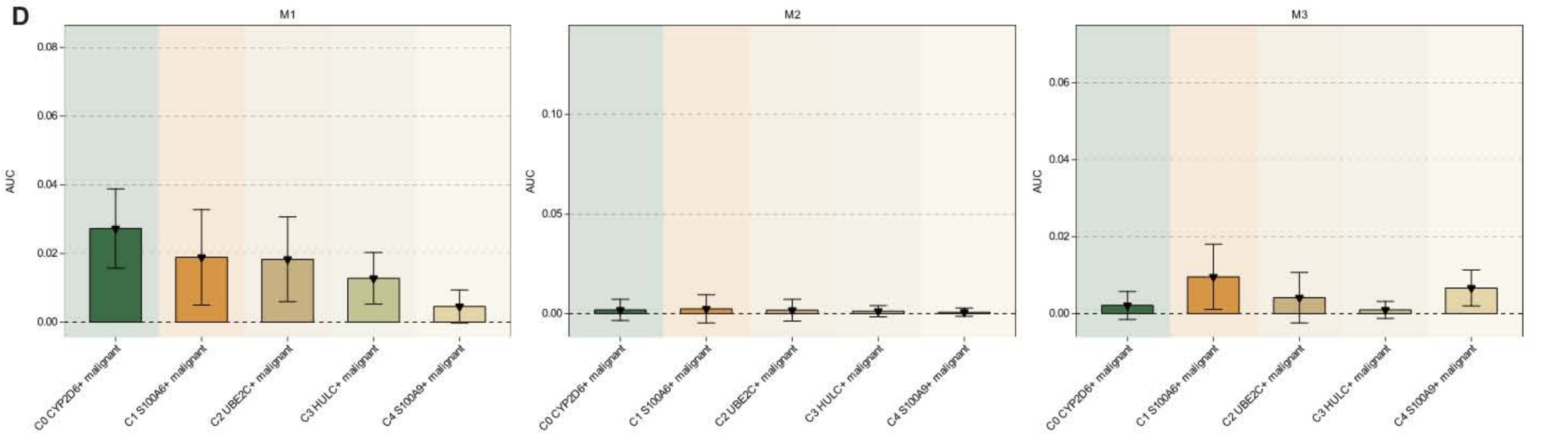
I

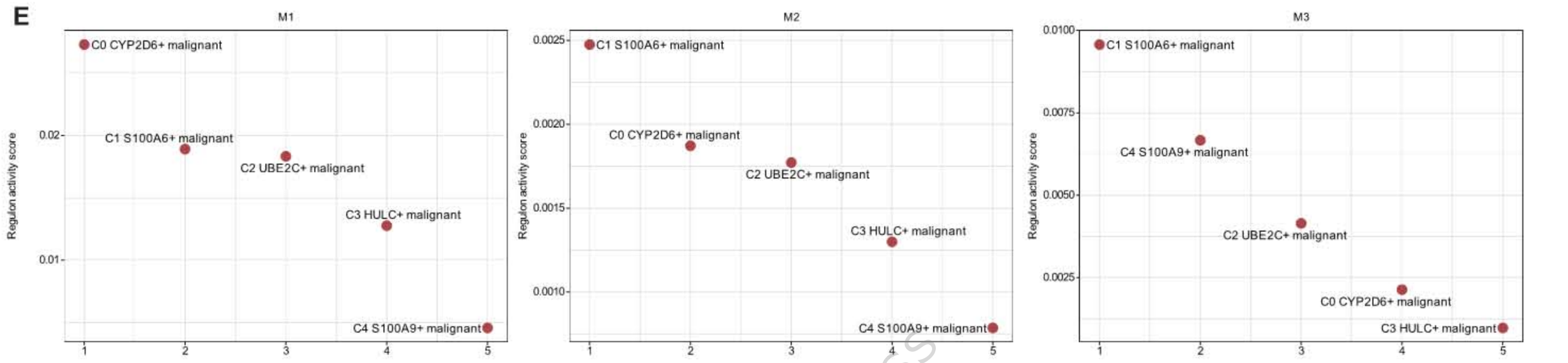


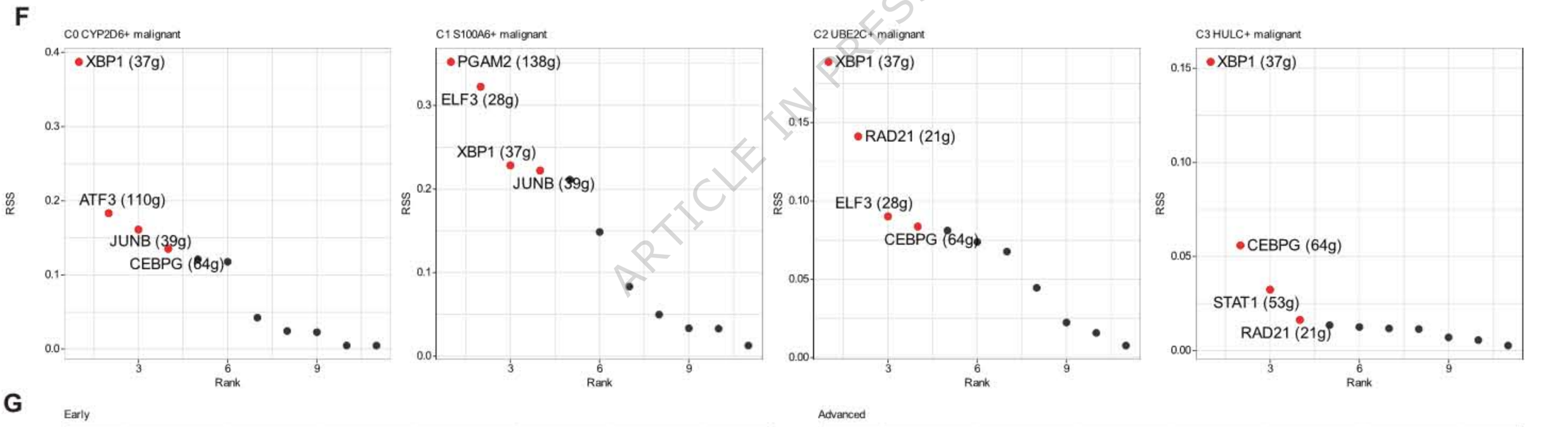
ARTICLE IN PRESS

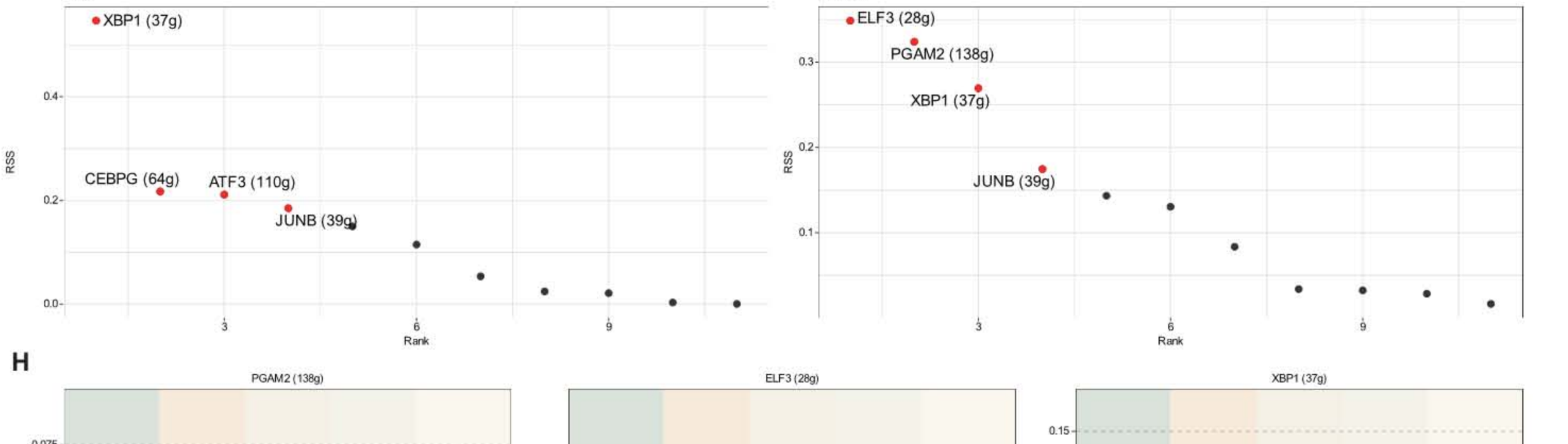


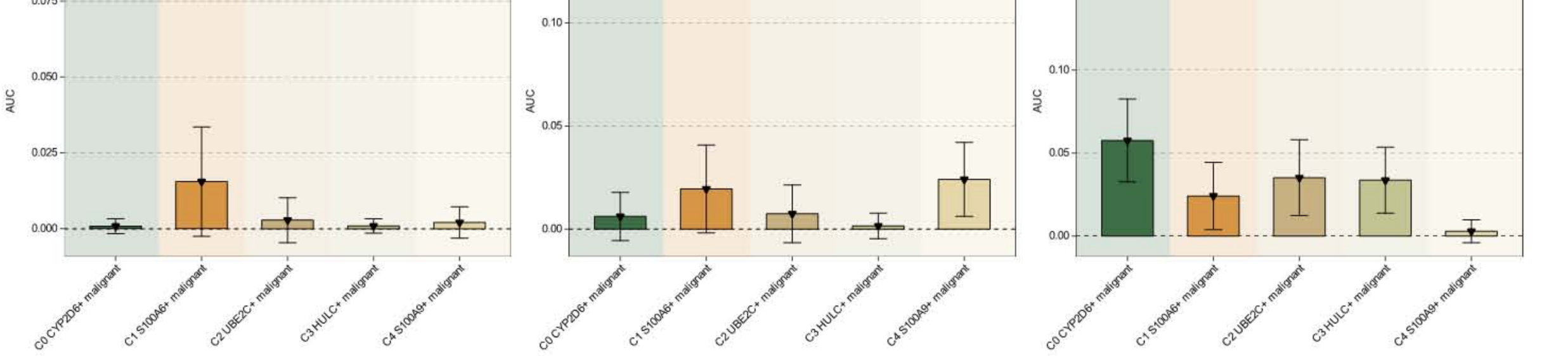


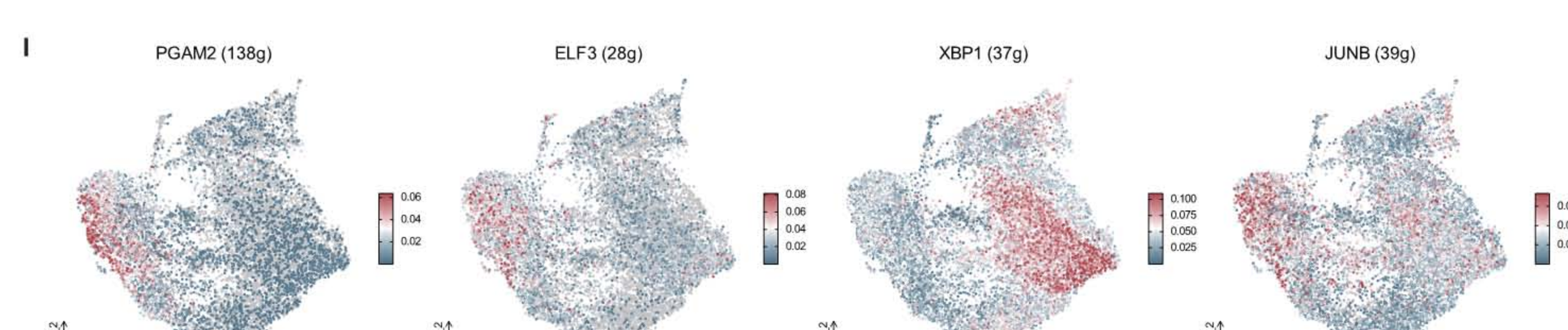




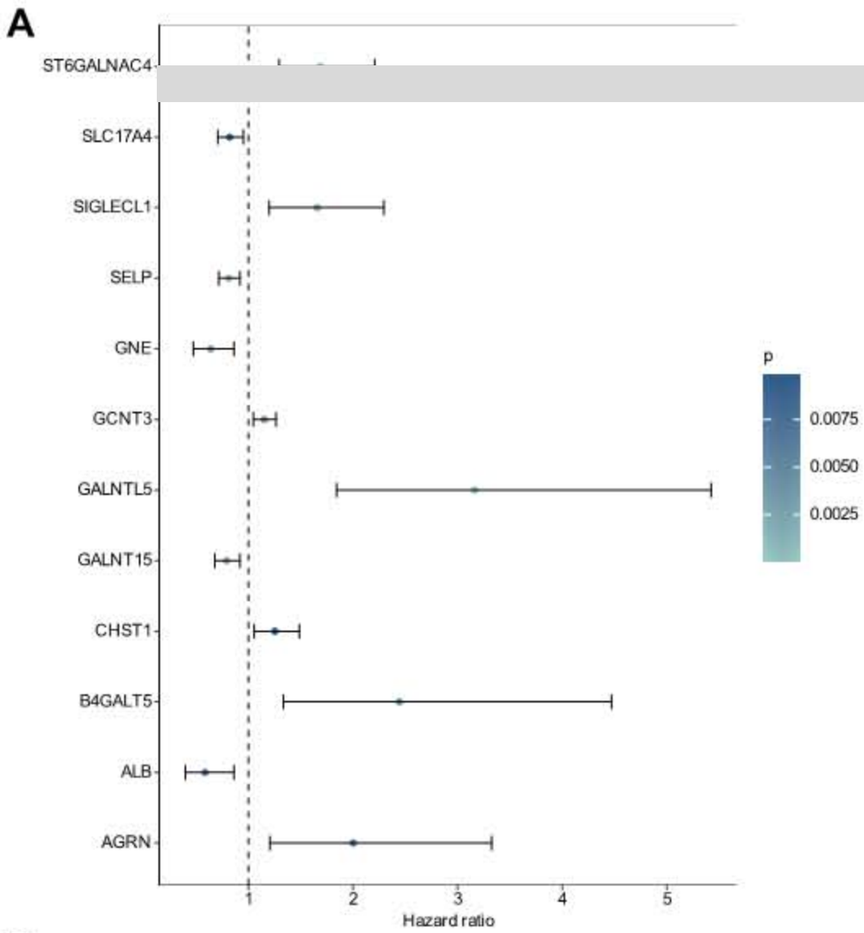


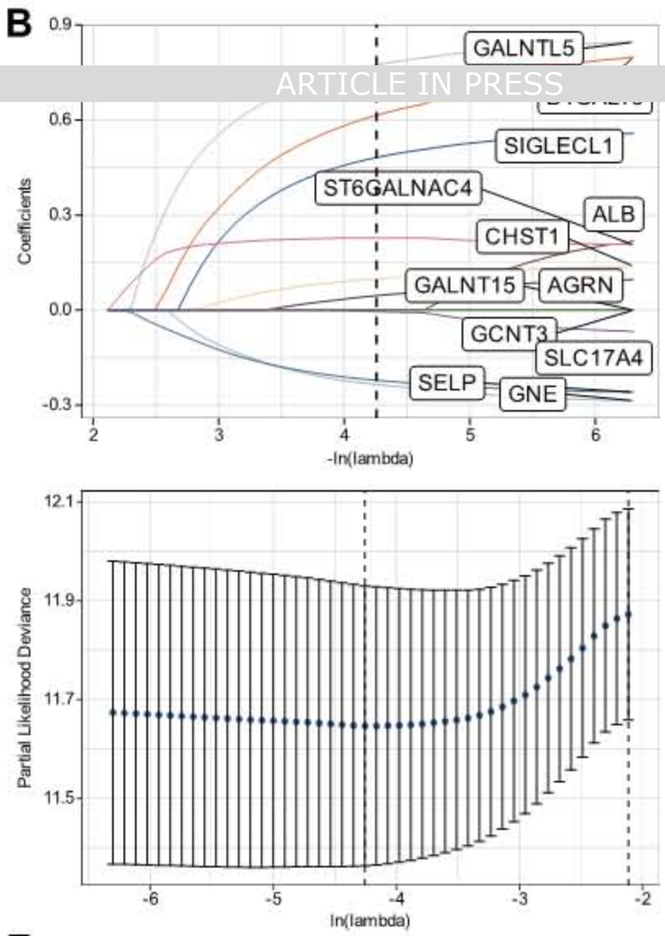


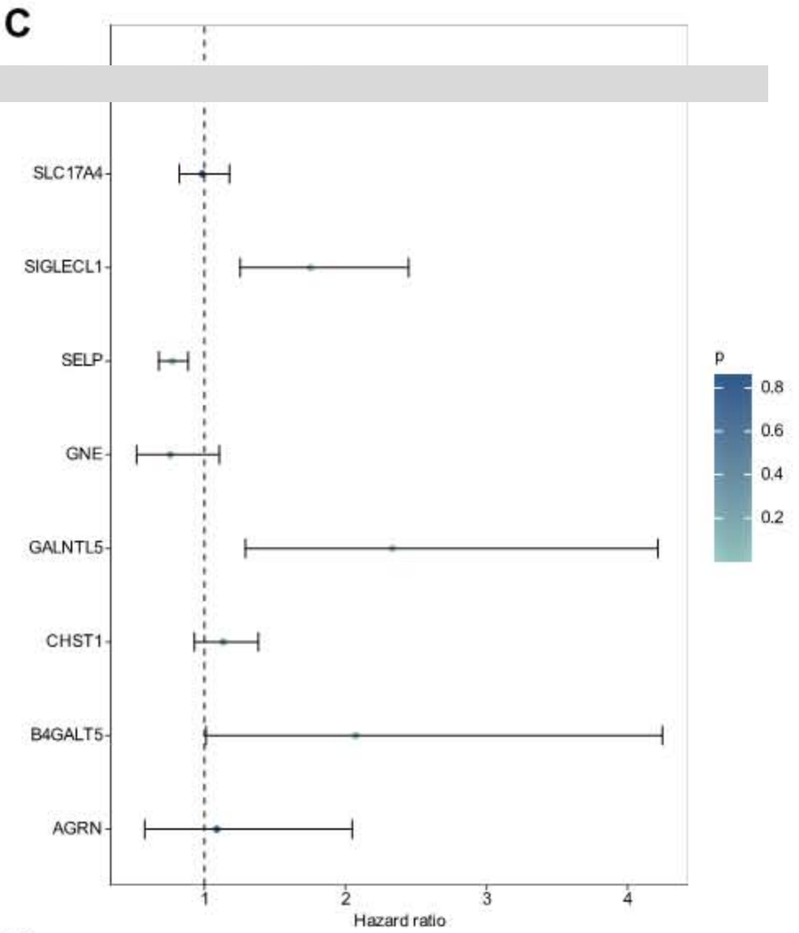


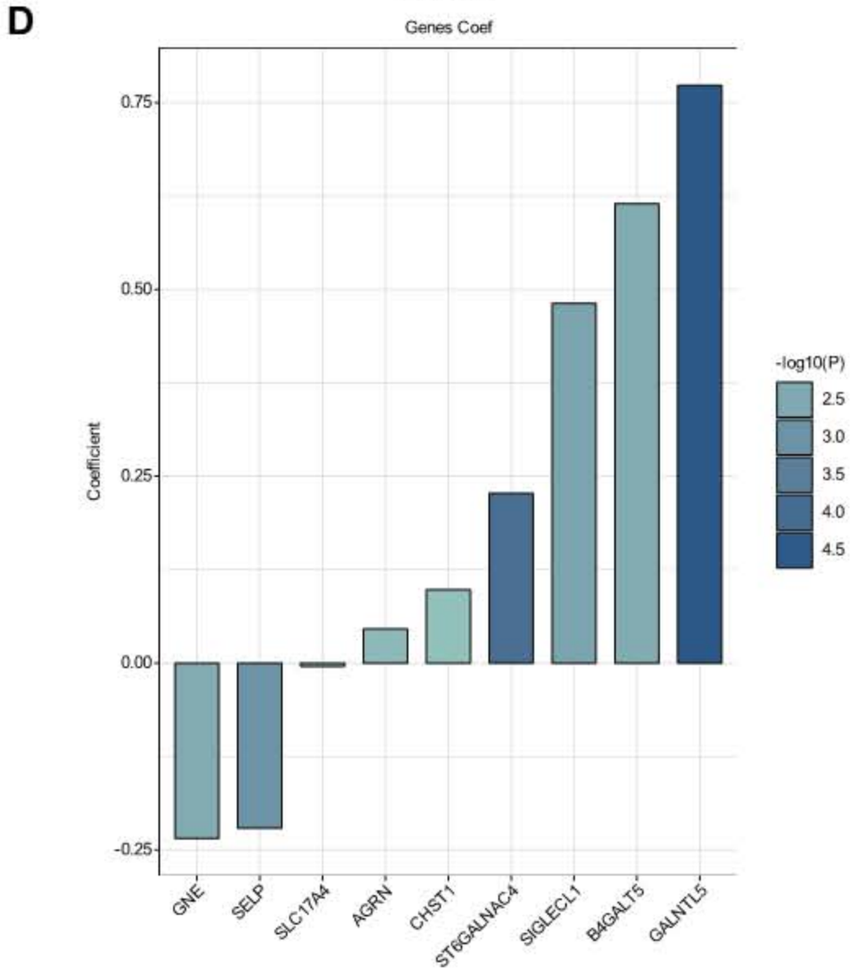


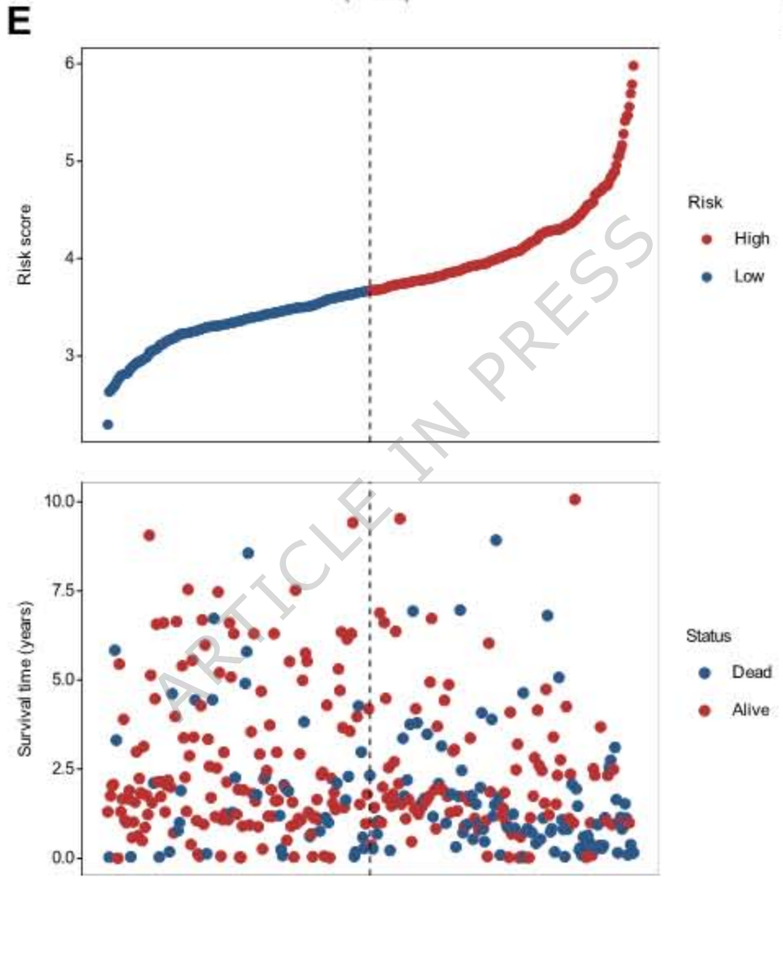
ARTICLE IN PRESS

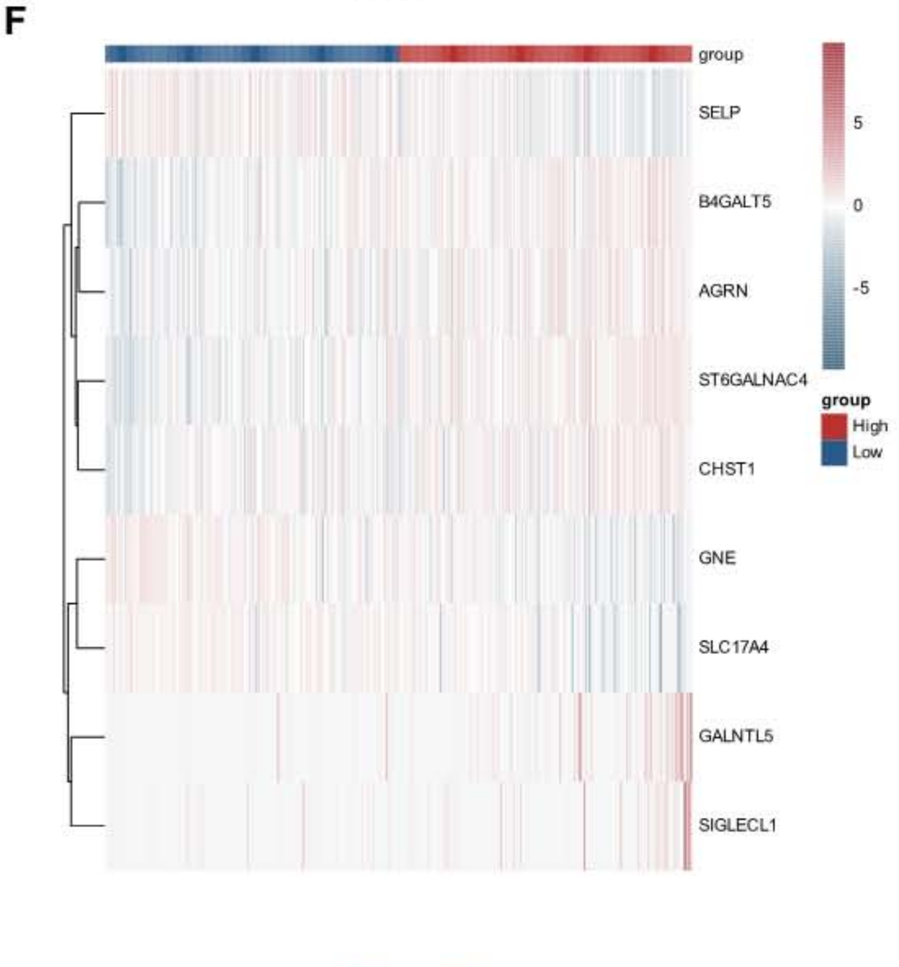


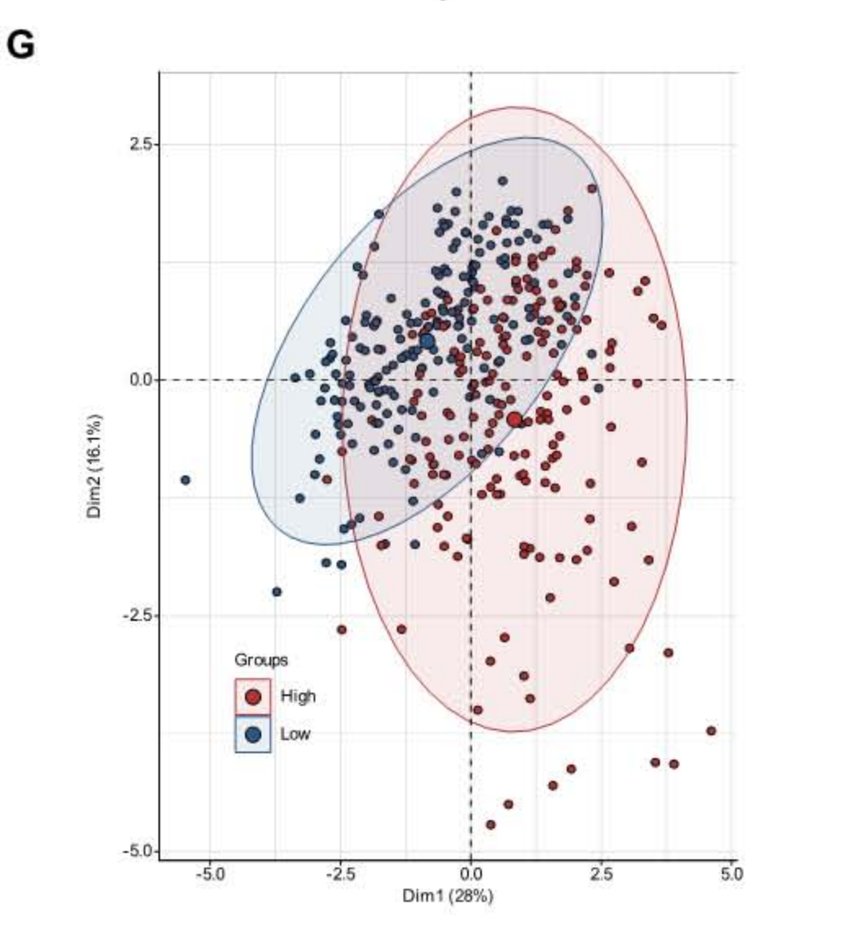


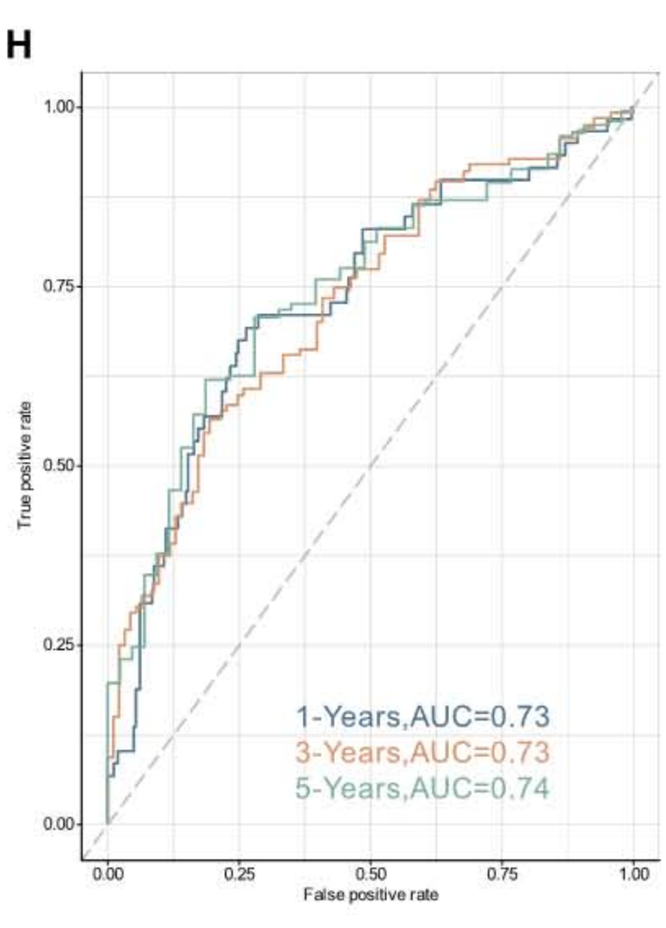


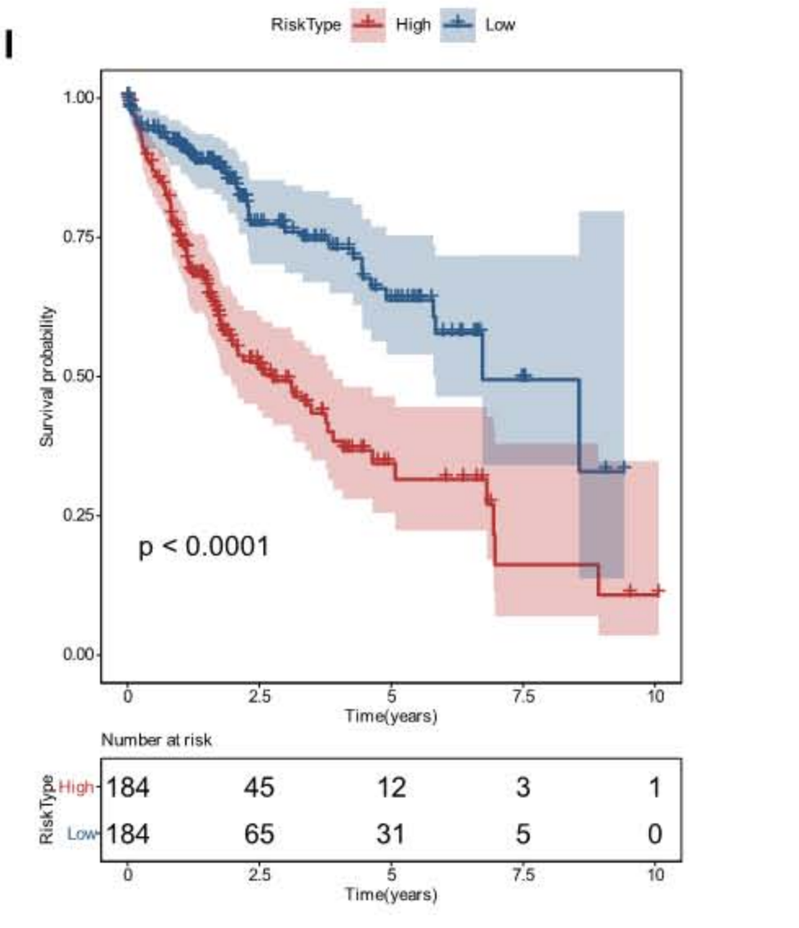




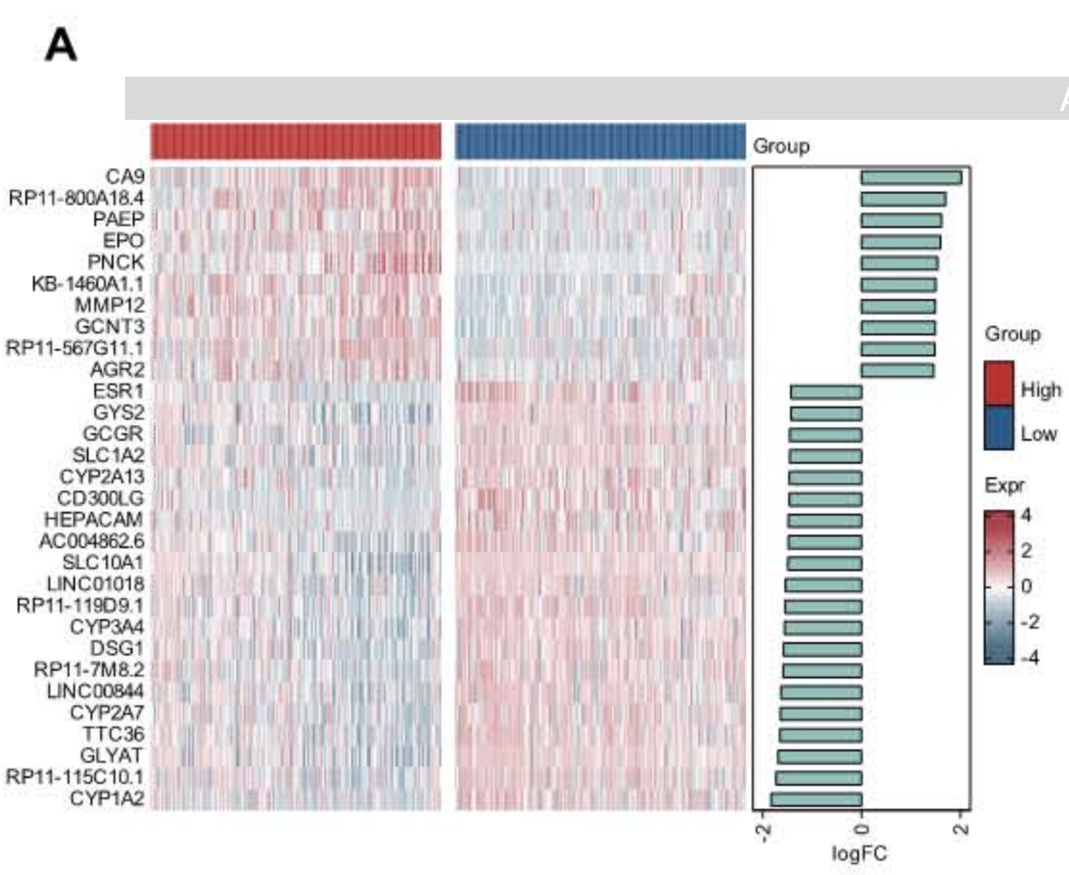


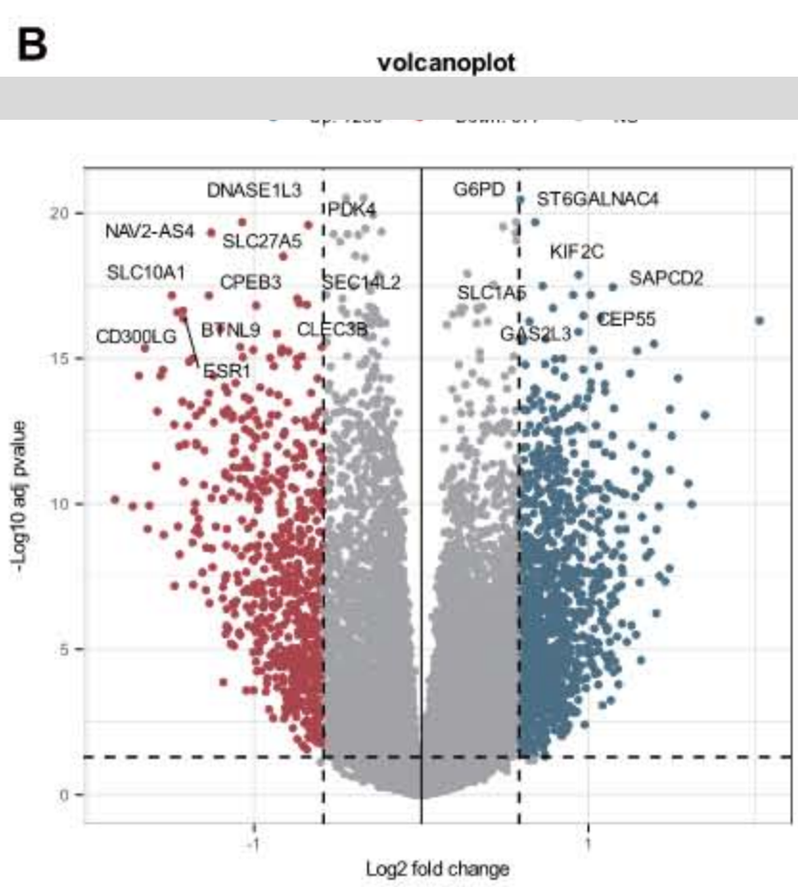


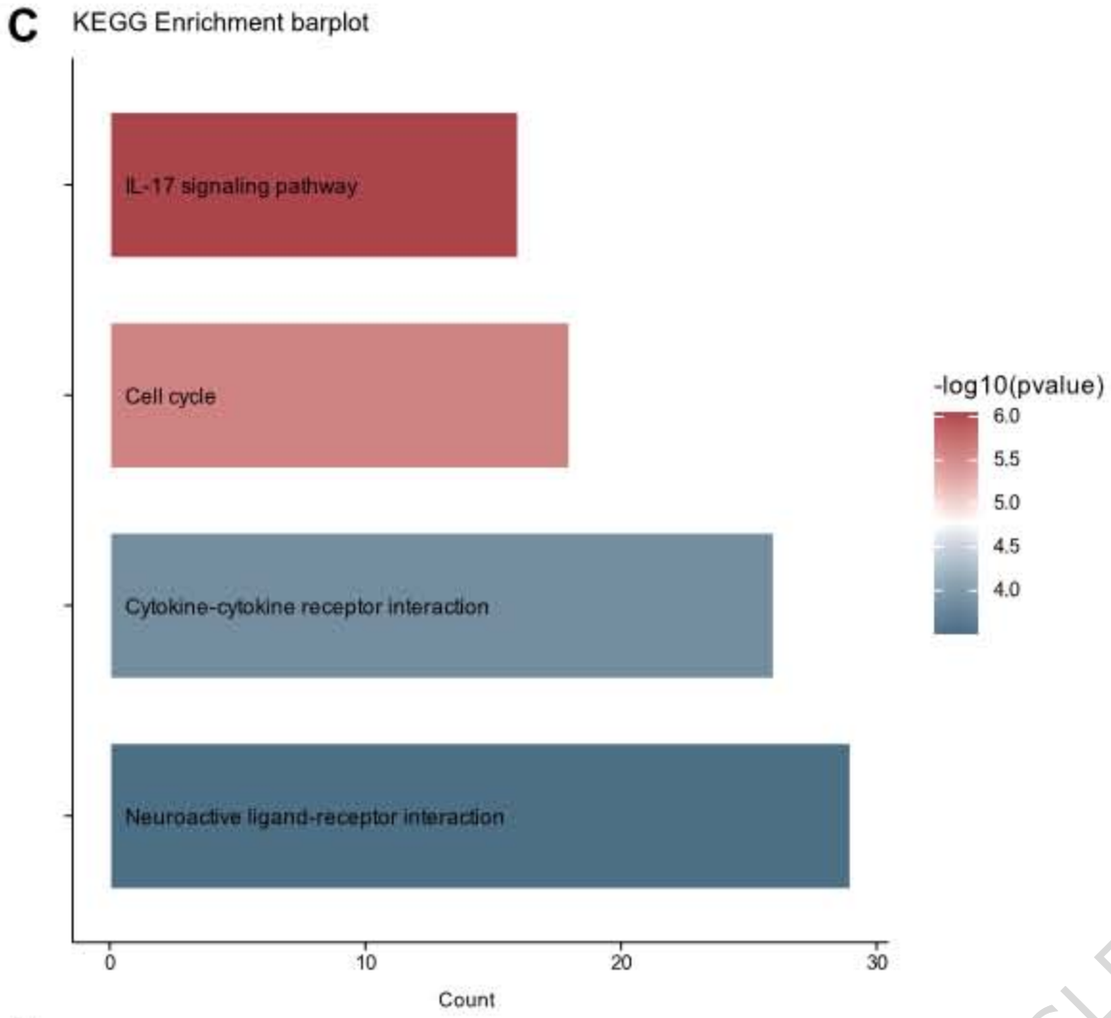


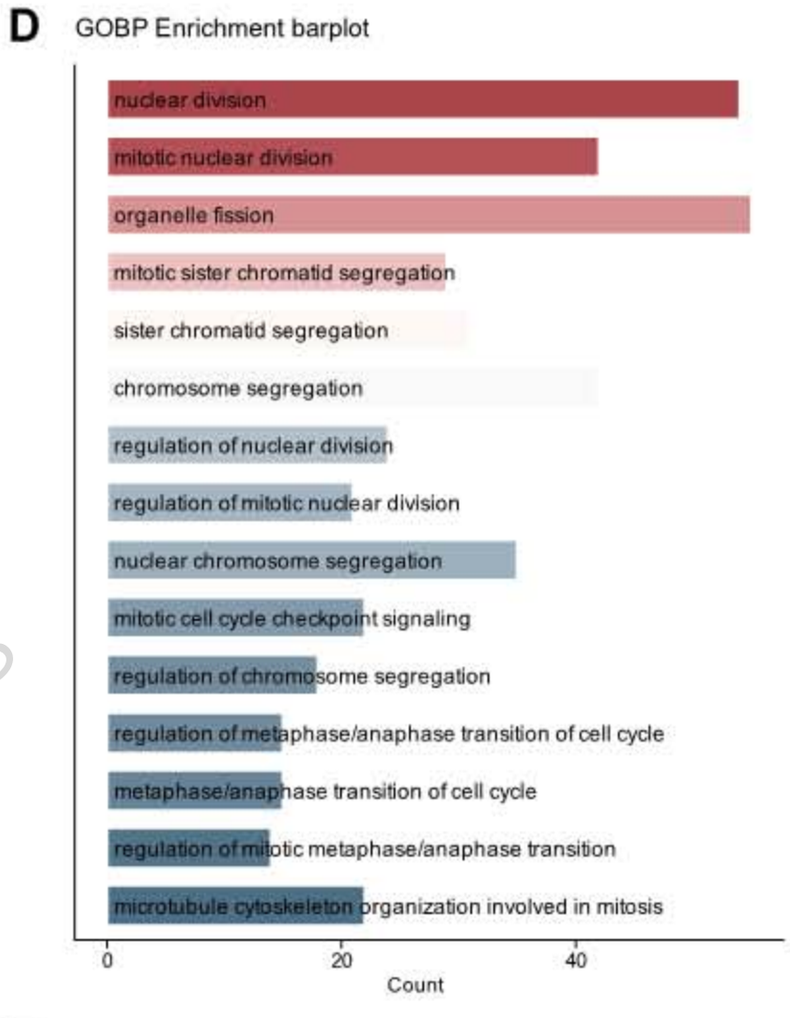


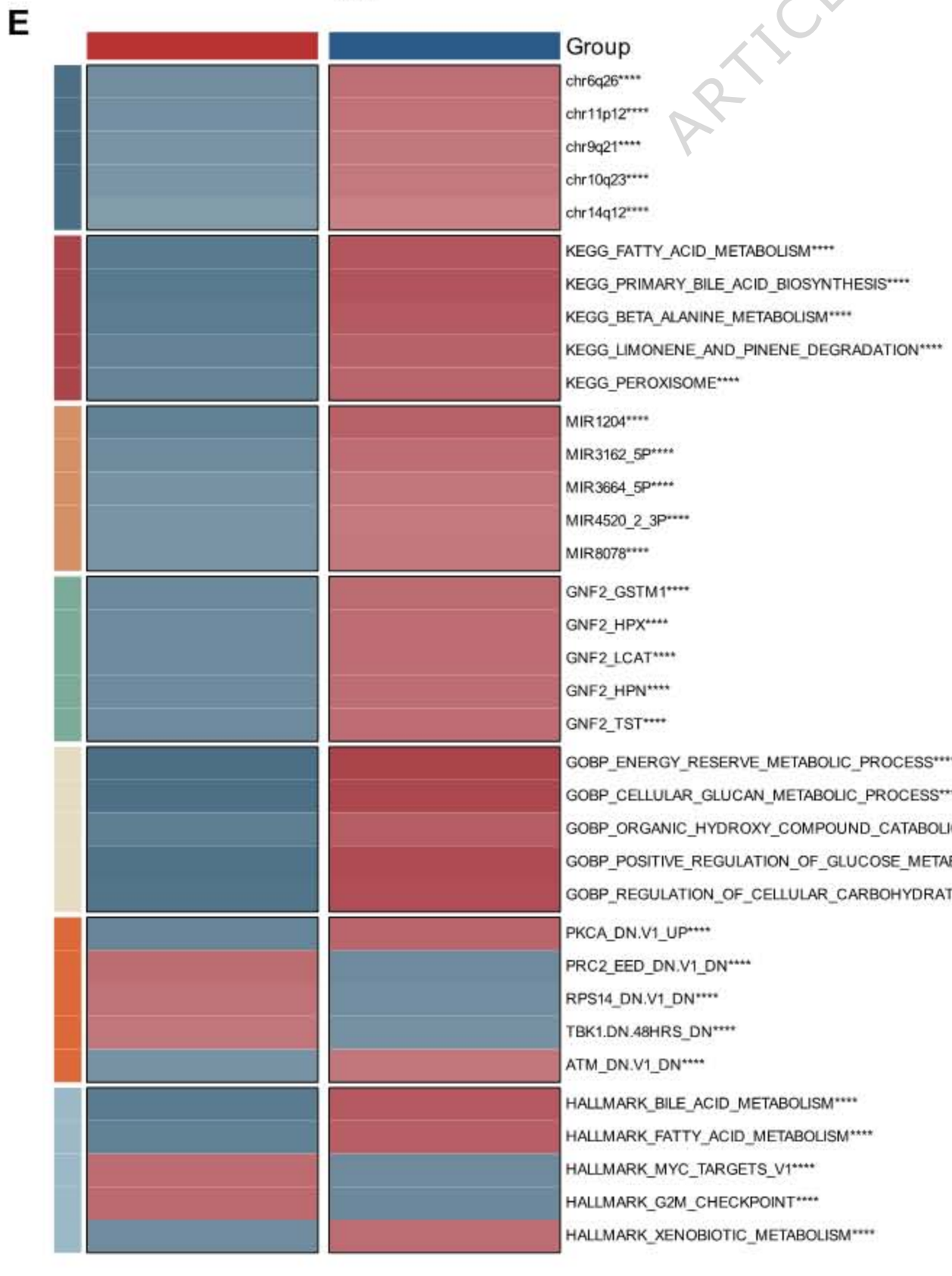
ARTICLE IN PRESS

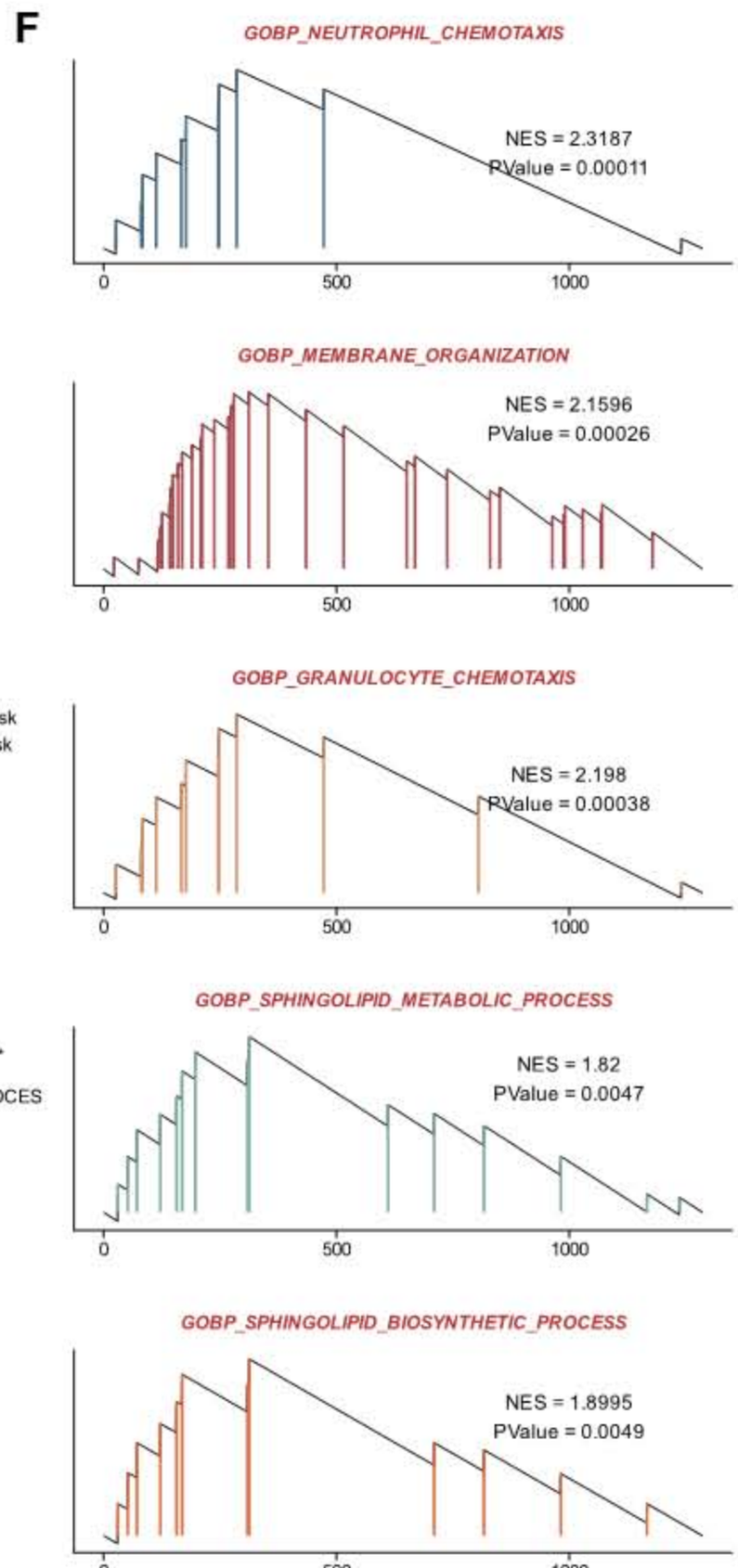






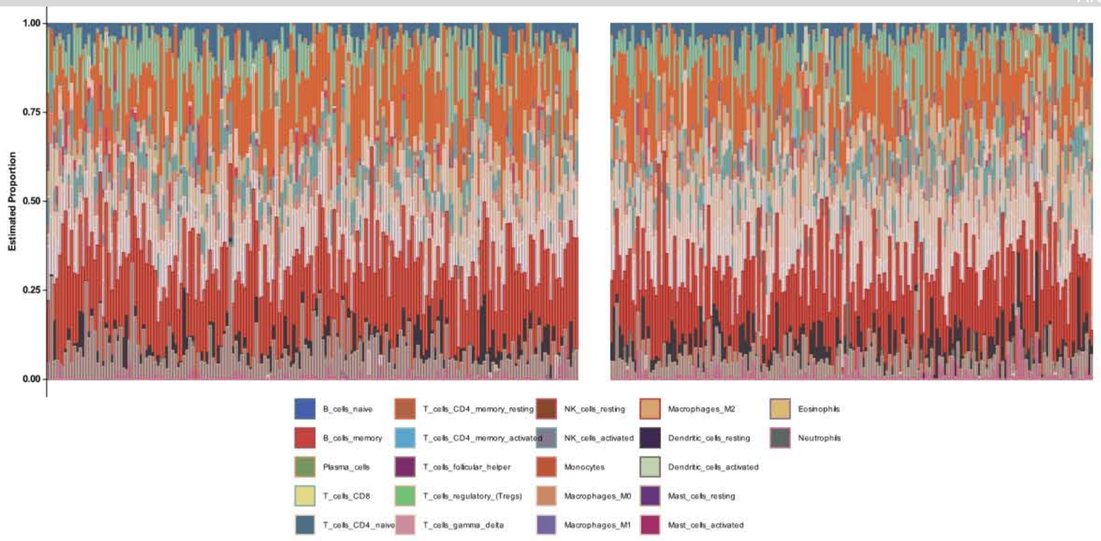




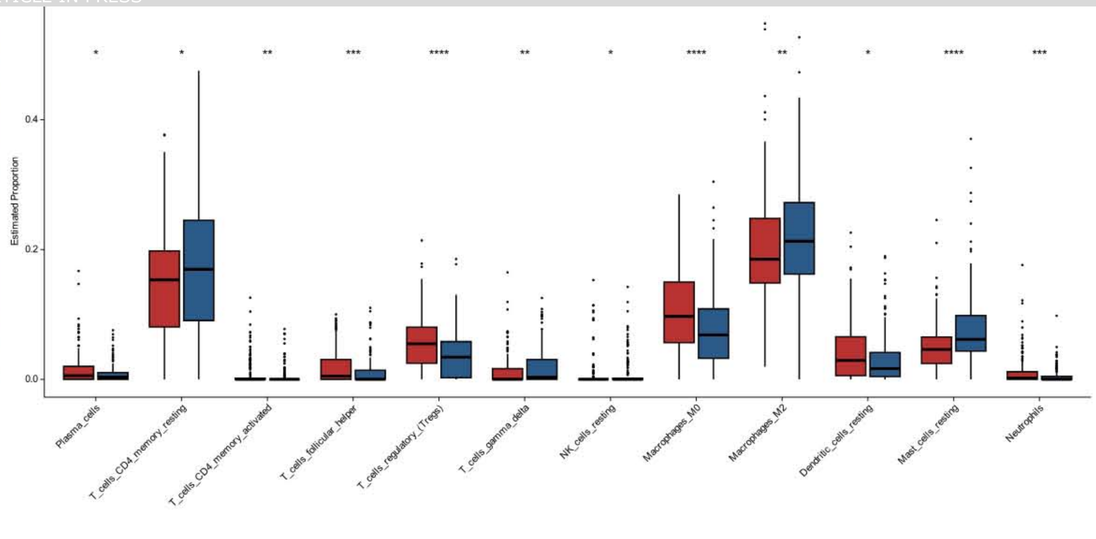


ARTICLE IN PRESS

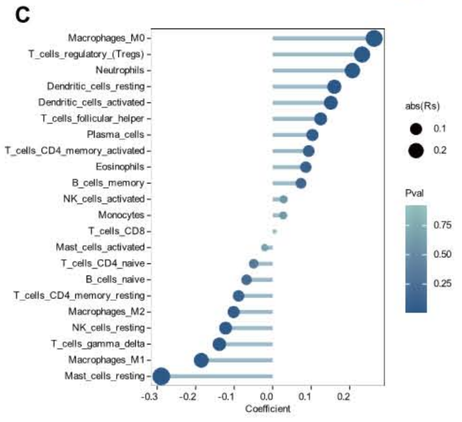
A



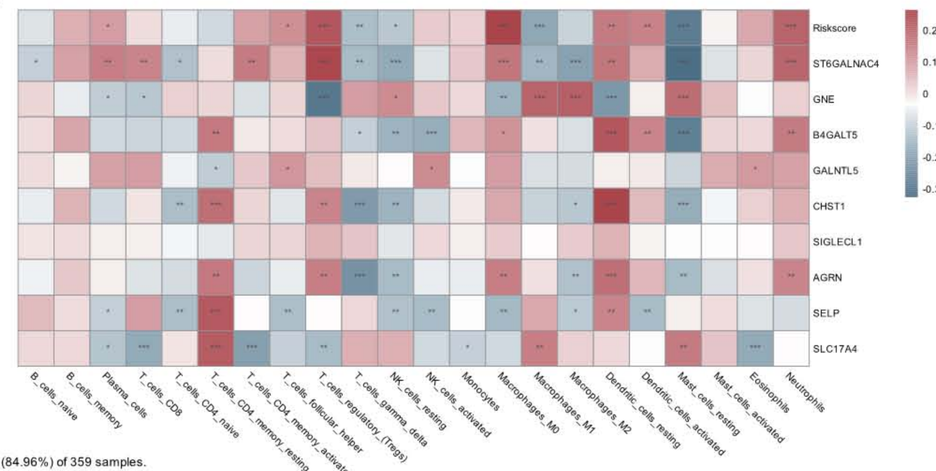
B



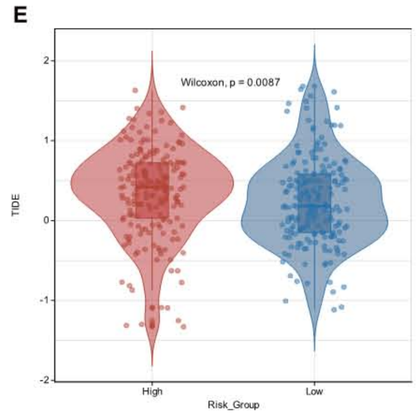
C



D

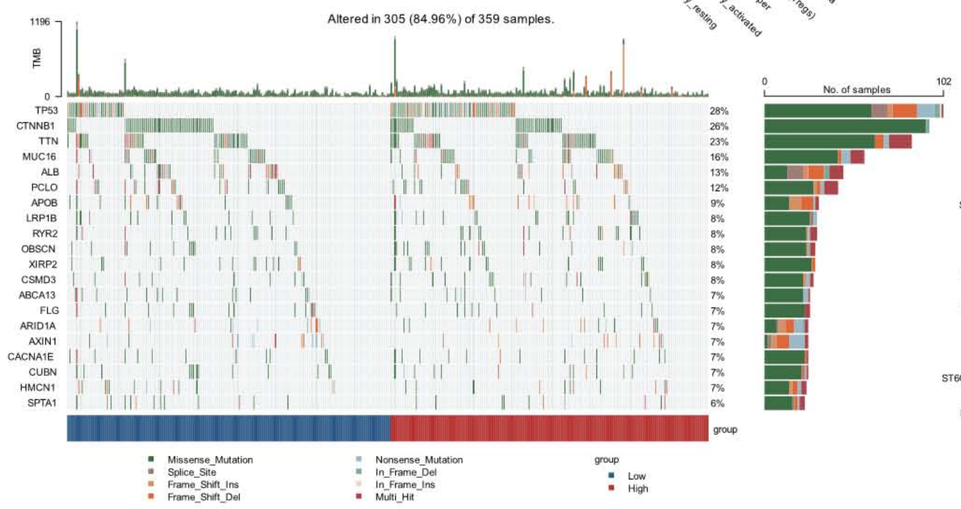


E

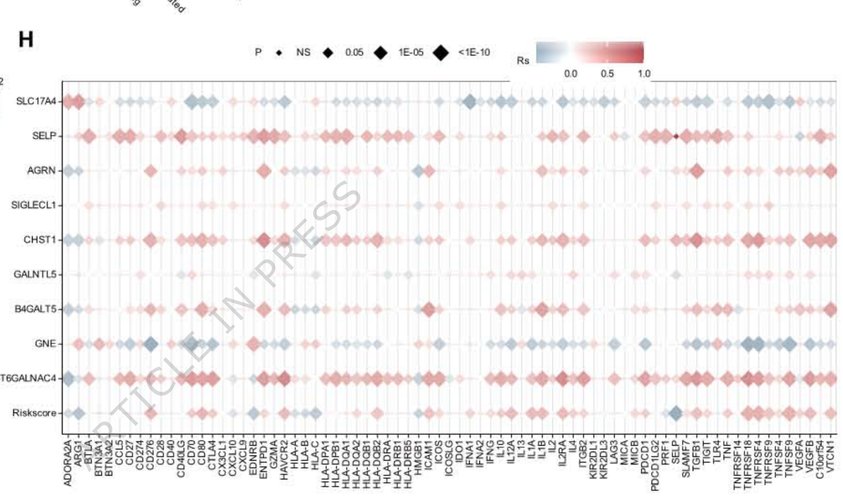


F

G

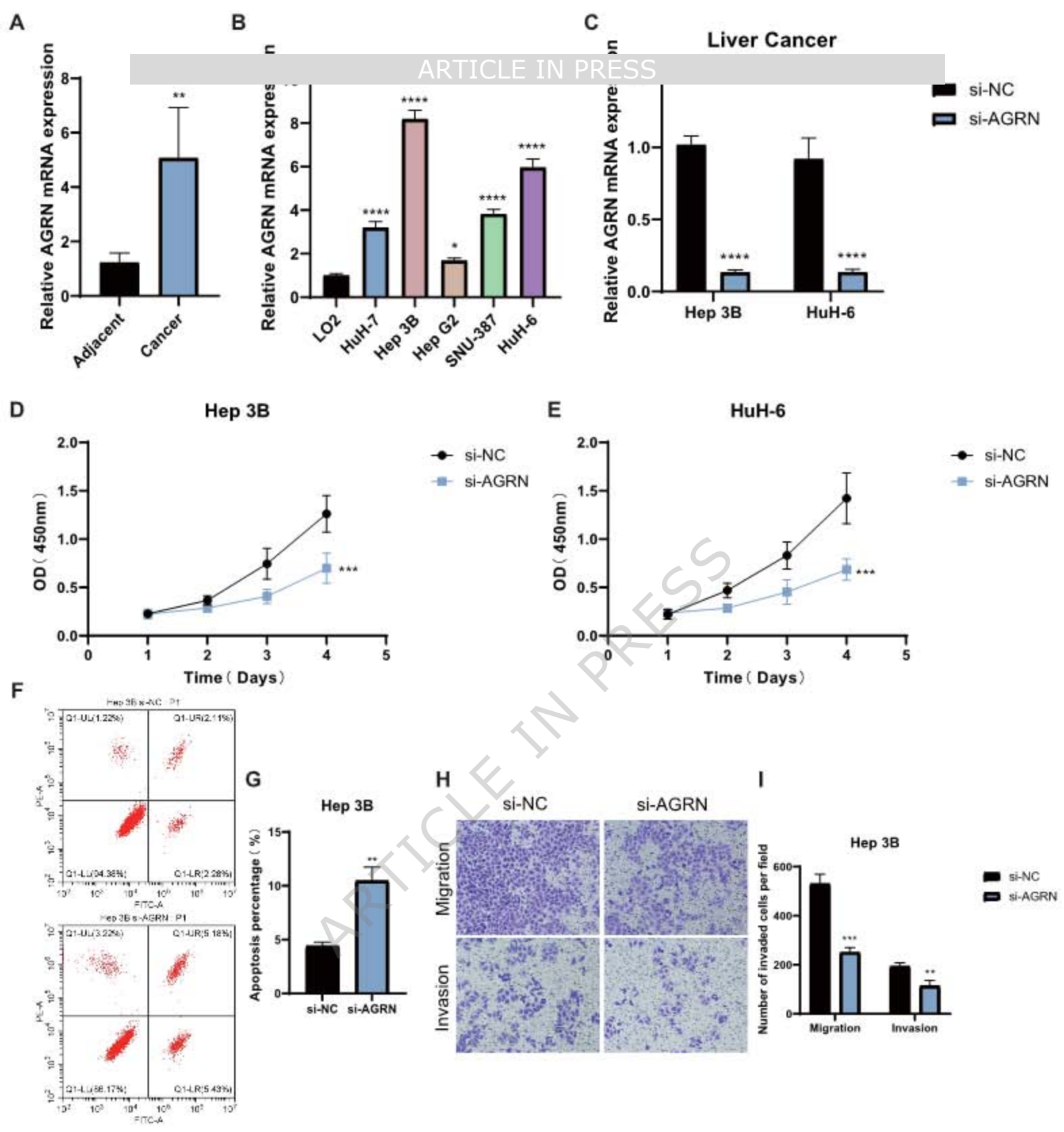


H



I

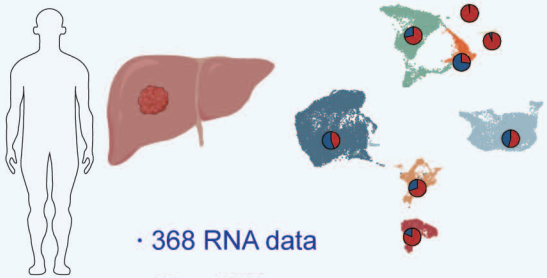
ARTICLE IN PRESS

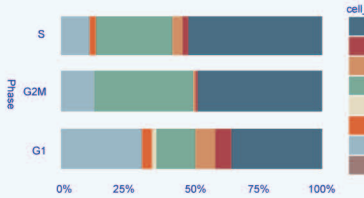


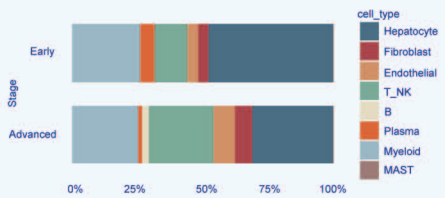
Machine Learning-Driven Comprehensive Profiling of Tumor Heterogeneity and Sialylation in Hepatocellular Carcinoma

ARTICLE IN PRESS

A. Cellular heterogeneity







· 368 RNA data

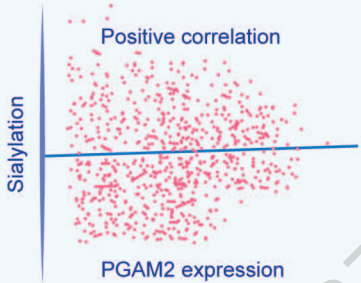
· 10 scRNA-seq

C0(CYP2D6$^+$), C2(UBE2C$^+$), C3(HULC$^+$): early stage tumor

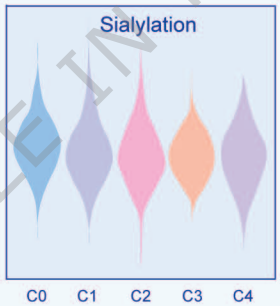
C1(S100A6$^+$), C4(S100A9$^+$): advanced stage tumor

B. Key mechanisms

The regulatory role of PGAM2



The role of sialylation



Intercellular communication networks

C1(S100A6$^+$)

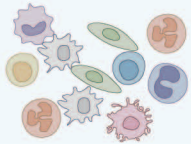
C4(S100A9$^+$)



Endothelium



Myeloid cells



C. Prognostic model

