JOURNAL OF
EMERGING INVESTIGATORS

Article

Citrate and lactate drive glioblastoma progression via activation of tumor-associated macrophages

Nikila Swaminathan[1], Venkata Naga Goutham Davuluri[2]

[1] Allen High School, Allen, Texas

[2] Department of Pathology and Immunology, Baylor College of Medicine, Houston, Texas

SUMMARY

Glioblastoma multiforme (GBM), the most aggressive primary brain tumor, presents a poor prognosis with a high resistance to current therapies. GBMs are highly associated with the tumor microenvironment (TME), which has become central to cancer research. Within the TME, tumor-associated macrophages (TAMs) can adopt pro-inflammatory or anti-inflammatory states, playing a significant role in tumor progression. We investigated the metabolic interaction between GBM cells and TAMs, specifically focusing on how citrate and lactate contribute to GBM advancement. We hypothesized that citrate and lactate drive GBM progression by modulating TAM polarization, promoting an anti-inflammatory state that supports tumor growth. Through gene expression analysis of GBM tumors and TAMs, alongside survival analysis of specific markers, we found a strong correlation between poor prognosis and high levels of anti-inflammatory-like TAM markers, such as cluster of differentiation 163 (CD163) and histone lysine demethylases (KDMs). Citrate and lactate were further identified as critical metabolites that were correlated with changes in α-ketoglutarate (α-KG) production and promoted anti-inflammatory macrophage polarization. Our findings underscore the potential of targeting citrate and lactate metabolism to enhance immunotherapeutic strategies in GBM. Future studies should prioritize biomarkers of citrate and lactate metabolism and explore the mechanisms by which these metabolites drive TAM polarization and GBM progression.

INTRODUCTION

Glioblastoma multiforme (GBM) represents approximately 15% of all brain tumors, marking it as one of the most deadly and aggressive cancers affecting the brain and spinal cord (1). Each year, around 12,000 new cases are diagnosed in the United States, and despite advancements in cancer treatments, the median survival time remains around 15 months (2). One critical factor contributing to GBM's aggressive behavior is its interaction with the tumor microenvironment (TME), a dynamic and complex niche that facilitates tumor growth and progression.

A significant component of the TME in GBM is tumor-associated macrophages (TAMs), which play a pivotal role in promoting tumor growth, angiogenesis, and metastasis (3). TAMs exist primarily in two phenotypes: pro-inflammatory and anti-inflammatory macrophages. Pro-inflammatory macrophages, or classically activated macrophages, contribute to immune responses by inhibiting cell proliferation and causing tissue damage, while anti-inflammatory macrophages, or alternatively activated macrophages, promote tissue repair and cell proliferation (4). In the TME, TAMs are typically polarized toward the anti-inflammatory-like phenotype, fostering an immunosuppressive environment that supports tumor progression by secreting anti-inflammatory cytokines and inhibiting effector cell function (5).

Metabolic alterations within the TME are significantly associated with changes in cancer behavior, particularly through metabolites like lactate and citrate (6, 7). Metabolic interactions between TAMs and cancer cells have been shown to be associated with the establishment of immunosuppressive conditions that facilitate tumor progression (8). The Warburg effect, a well-documented metabolic feature of cancer cells, plays a central role in this metabolic reprogramming. Through the Warburg effect, cancer cells convert glucose to lactate even in the presence of oxygen, enabling rapid energy production and contributing to hypoxic and acidic conditions in the TME (9). By predominantly relying on glycolysis over oxidative phosphorylation, tumor cells consume substantial amounts of glucose, which leads to resource depletion in the TME and potentially contributes to hypoxia (10, 11).

Under hypoxic conditions, hypoxia-inducible factor 1-alpha (*HIF1A*) stabilizes and activates glycolytic genes, increasing glycolysis rates and leading to lactate production (12). HIF1A also diverts pyruvate from the tricarboxylic acid (TCA) cycle toward lactate production via lactate dehydrogenase A (LDHA), sustaining glycolysis by regenerating cellular nicotinamide adenine dinucleotide (NAD+) and maintaining pH balance through lactate export via monocarboxylate transporters (MCTs) (13). As lactate accumulates in the TME, it creates a cycle of lactic acidosis and nutrient depletion, which may suppress immune cell function and potentially facilitate anti-inflammatory-like TAM polarization (14).

The presence of lactate within the TME exerts multiple effects on TAMs that are associated with the polarization of TAMs toward the anti-inflammatory-like state. Lactate not only serves as an energy source but also acts as a signaling molecule by interacting with specific macrophage receptors that promote anti-inflammatory polarization. By inhibiting pro-inflammatory pathways, lactate reinforces the anti-inflammatory phenotype, thereby supporting tissue repair and immunosuppression within the TME, both of which favor tumor growth (15).

As glycolysis intensifies in tumor cells due to HIF1A activation, their demand for glucose creates competition for resources in the TME (16). High glucose levels are associated with pro-inflammatory macrophage polarization, which requires

https://doi.org/10.59720/24-258

glycolysis to meet energy demands (17). However, glucose depletion in the TME hinders pro-inflammatory macrophage functions and shifts macrophages toward the anti-inflammatory phenotype, further reinforcing an anti-inflammatory environment that favors tumor growth (18). Tumor cells actively deplete glucose in the TME as the disease progresses, while simultaneously producing increased lactate levels, which may be associated with the recruitment and polarization of monocytes into an immunosuppressive anti-inflammatory-like state (13).

Moreover, as a result of this metabolic shift, key intermediates from the TCA cycle, such as α-KG and citrate, accumulate in the hypoxic TME due to reduced pyruvate availability for the cycle (19, 20). These metabolites are exported from tumor cells into the TME, where they are taken up by TAMs and utilized as carbon sources for biosynthetic processes that support anti-inflammatory polarization. Excess citrate is imported into TAMs via specific transporters and metabolized to produce acetyl-CoA, which fuels anti-inflammatory functions (21, 22). Similarly, extracellular α-KG can be taken up by tumor associated macrophages (TAMs) via Solute Carrier Family 13 (SLC13) transporters, promoting immune modulation and contributing to the anti-inflammatory macrophage polarization (23, 24). Through these mechanisms, α-KG helps regulate the balance between pro-inflammatory and anti-inflammatory macrophage polarization, potentially reducing inflammation and favoring tumor progression (25).

Recent *in vitro* studies in GBM cell lines reveal that elevated lactate levels are associated with increased GBM cell migration and invasion, which may contribute to tumor progression and metastasis (26). Patient-derived GBM cells exposed to varying citrate concentrations have shown alterations in signaling pathways, such as the phosphatidylinositol 3-kinase (PI3K) / protein kinase B (Akt) pathway, providing evidence that metabolic modulation may be a promising therapeutic target (27). These findings highlight the impact of lactate and citrate on GBM at the cellular level and underscore the potential of targeting these metabolites for GBM therapy.
In this study, we hypothesized that lactate and citrate metabolism within the TME of GBM played a critical role in tumor progression by polarizing TAMs to the anti-inflammatory-like phenotype, thereby promoting tumor growth. Specifically, we propose that elevated lactate and citrate levels may drive an anti-inflammatory-like macrophage phenotype, contributing to GBM cell survival, invasion, and immune evasion. Our results indicated a significant association between anti-inflammatory TAM markers and poor survival in GBM patients. Moreover, our findings suggest that the upregulation of specific KDM genes and the presence of lactate in the TME are linked to the immunosuppressive microenvironment in GBM. Additionally, we observed that lactate uptake via MCT1 and MCT2 may influence TAM polarization, while α-KG played a crucial role in regulating macrophage polarization. By focusing on these metabolic interactions within the TME, we aim to uncover novel therapeutic targets for GBM treatment, with the potential to disrupt the metabolic pathways supporting tumor growth and immune evasion. This study lays the groundwork for future investigations into targeting metabolic alterations as a strategy to treat GBM and improve patient survival.

RESULTS

Gene expression analysis in GBM tumor dataset

We analyzed the GBM tumor dataset from the Cancer Genome Atlas (TCGA), which included five healthy controls and 156 primary GBM cancer samples. Our analysis focused on glycolysis and TCA cycle genes, along with *HIF1A* due to its role in regulating these processes. Compared to healthy controls, we saw an upregulation of *LDHA*, *HIF1A*, and *IDH1*, and downregulation of phosphofructokinase platelet (*PFKP*) and citrate synthase (*CIT*) (**Figure 1**). *LDHA* expression increased by 7.94% (p-value = 5.56e-3), *PFKP* was downregulated by 15.63% (p-value = 1.86e-4), and *CIT1* showed a 22.41% decrease (p-value = 1.36e-3). *IDH1* was upregulated by 25.56% (p-value = 1.47e-4), and *HIF1A* showed a 15.50% increase (p-value = 2.34e-4), suggesting significant metabolic reprogramming in GBM. These findings indicated that enhanced glycolysis and TCA cycle dysregulation, driven in part by *HIF1A*, supported tumor survival and growth. Targeting these pathways may provide potential therapeutic strategies for GBM treatment.

Further analysis, incorporating datasets from the Genotype-Tissue Expression (GTEx) portal, revealed positive correlations between the anti-inflammatory-like TAM markers *CD163* and *MSR1* with key metabolic genes involved in the TCA cycle (*IDH1*) and lactate metabolism (*HIF1A*, *MCT4*) (**Figure 2**). Statistical analysis was performed using Pearson correlation. Moreover, analysis of *IDH1* mutation status in 156 samples revealed that 15% of GBM samples harbored *IDH1* mutations, exhibiting a distinct metabolic phenotype characterized by reduced glycolysis and altered TCA cycle activity. This highlighted the metabolic heterogeneity of GBM and the importance of considering *IDH1* mutations in understanding tumor metabolism (28).

TAM marker analysis on GBM patient prognosis

Building on the gene expression analysis, we explored the impact of TAM markers on patient prognosis in GBM. Specifically, we focused on CD163, CD204 or macrophage scavenger receptor 1 (MSR1), and CD206 or MRC1 (mannose receptor C type 1), markers associated with anti-inflammatory-polarized macrophages (29). Survival analysis revealed that elevated expression of CD163, MSR1, and CD206 was correlated with significantly poorer prognosis in GBM patients, highlighting the potential role of TAMs, particularly the anti-inflammatory-polarized subset, in GBM development and patient outcomes (**Figure 3**).

Gene set enrichment analysis in GBM-TAM

We performed gene set enrichment analysis (GSEA) using the TCGA expression data to investigate gene expression patterns within GBM-associated TAMs. We saw upregulation of several *KDM* and *MCT* genes within GBM TAMs (**Table 1**). Specifically, *KDM4B*, *KDM5B*, *KDM5C*, *KDM6B*, and *KDM7A* were upregulated, along with *MCT1* and *MCT2*, which are transporters of lactate. These findings suggested that these genes played roles in the uptake of lactate into TAMs and may have been involved in metabolic reprogramming.

Patient survival analysis

Our analysis extended to exploring the associations between gene expression, TAM markers, and patient survival. Several combinations of *KDM* and *CD163* exhibited signifi-

https://doi.org/10.59720/24-258

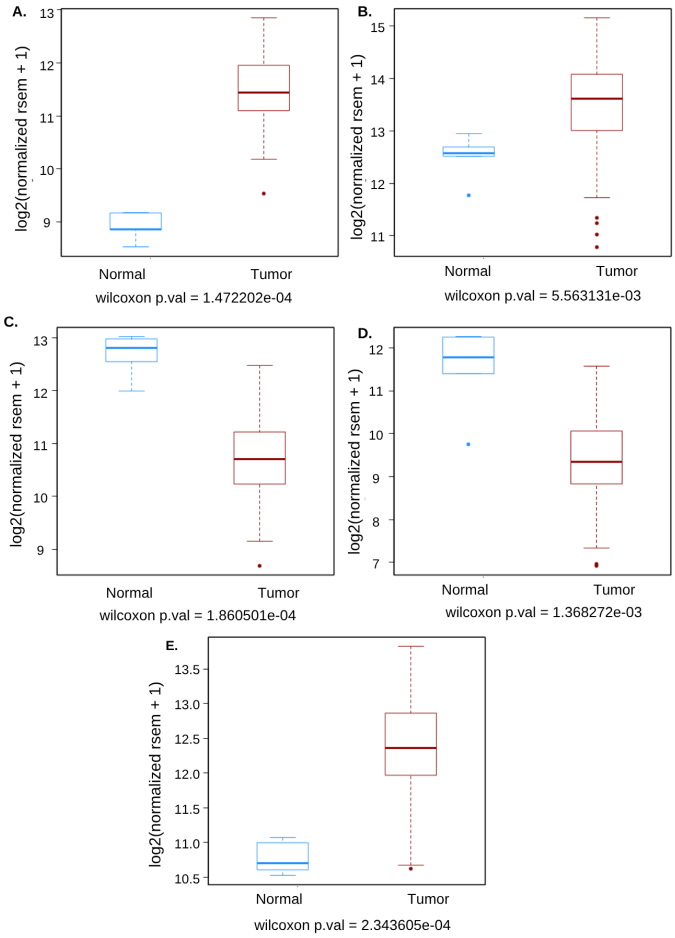

Figure 1: Gene expression profiles of glycolysis and TCA genes in the GBM cancer dataset. Gene expression profiles of **A)** *IDH1,* **B)** *LDHA,* **C)** *PFKP,* **D)** *CIT1,* and **E)** *HIF1A* showing log2(x+1) transformed RSEM normalized counts of genes in the GBM cancer dataset. p = **A)** 1.47 e-4, **B)** 5.56 e-3, **C)** 1.86 e-4, **D)** 1.36 e-3, and **E)** 2.34 e-4. Tumor samples (n=150) and normal samples (n=5) were included in the analysis.

cant effects on survival (**Figure 4A - D**). For example, the combination of *KDM5B* and *CD163* was found to correlate with poorer survival outcomes (p-value = 1.956e-6, **Figure 4B**). This suggests that the interaction between this specific KDM and CD163 may have contributed to an immunosuppressive tumor microenvironment, promoting tumor progression. Additionally, combined high expression levels of CD206, ARG1, and CD163 were significantly associated with poorer survival outcomes (p-value = 0.03, **Figure 5**), further supporting the potential prognostic value of these markers.

Hypothetical model of GBM TAM interaction

Based on the findings of this study, we constructed a hypothetical model of the interaction between TAMs and tumor cells in GBM using BioRender (**Figure 6**). This model suggests that tumors secrete lactate through HIF1A and α-KG through the Krebs cycle, which are then absorbed by TAMs. Additionally, citrate is metabolized directly in TAMs, inducing higher α-KG. The activation of KDMs is controlled by α-KG, leading to anti-inflammatory polarization. This model highlights the key metabolic interactions within the TME that influence TAM polarization and contribute to GBM progression, providing insights into potential therapeutic targets for GBM treatment.

DISCUSSION

GBM tumors exhibit an immunosuppressive TME, which often results in poor responses to immunotherapies. Crosstalk between the various components of the TME can alter tumor cell metabolism, including extracellular metabolites (8). IDH1, an enzyme in the TCA cycle, converts isocitrate to α-KG, and overexpression of *IDH1* has been observed in GBM, supporting tumor growth and resistance (30). We observed upregulation of *LDHA* and *IDH1* in GBM tumor cells, leading to an excess production of lactate and α-KG, respectively (**Figure 1**). These metabolites are then released into the TME, reinforcing the dysregulated metabolic environment. Additionally, HIF1A plays a critical role in promoting lactate secretion under hypoxic conditions. Our findings showed significantly high expression of *HIF1A* in GBM patients (p-value = 2.34e-4), suggesting that *HIF1A* contributes to metabolic reprogramming in the TME (31). Notably, *CIT1* was downregulated in tumor cells, which suggests suppression of the TCA cycle and a shift toward anaerobic respiration, resulting in increased lactate production.

Anti-inflammatory-like TAMs, characterized by markers such as CD163, MSR1, and MRC1, are typically associated with tissue repair and wound healing. However, in GBM,

https://doi.org/10.59720/24-258

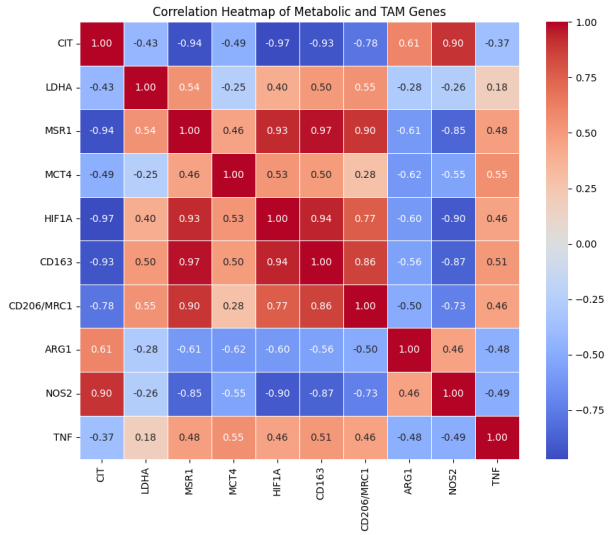

Figure 2: Correlation heatmap of correlations between TAM markers and metabolic genes in the tumor microenvironment. The heatmap was generated from gene expression data of tumor-associated macrophage (TAM) markers (*CD163, NOS2, MSR1*) and metabolic genes (*HIF1A, IDH1, MCT4, CIT*) in the GBM tumor microenvironment. It shows significant correlations between these TAM markers and key metabolic genes involved in the tricarboxylic acid (TCA) cycle and lactate metabolism. Notably, anti-inflammatory-like TAM markers, such as *CD163* and *MSR1*, exhibit strong positive correlations with metabolic genes like *HIF1A, IDH1,* and *MCT4.* Conversely, pro-inflammatory-like TAM marker **NOS2** shows positive correlations with *CIT*, a key TCA cycle gene. Statistical analyses were performed using Pearson correlation to compute the relationships between metabolic and TAM genes. Significant correlations had a value greater than 0.7.

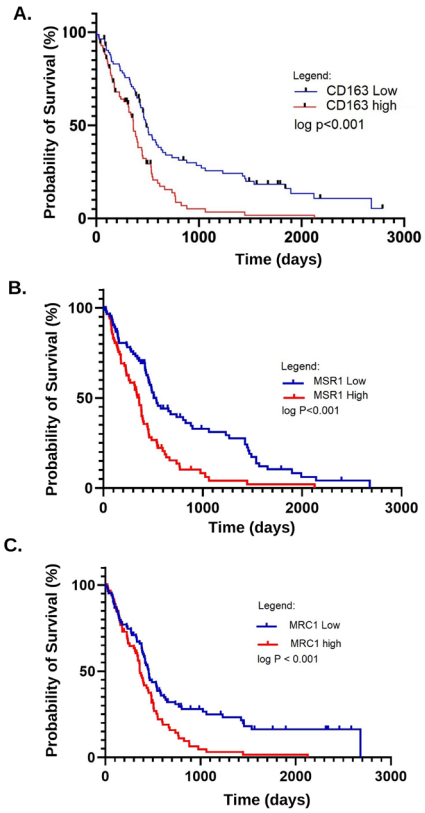

Figure 3: Survival analysis of GBM TAM markers. Kaplan-Meier survival curves for GBM patients stratified by expression levels of **A)** *CD163,* **B)** *MSR1,* and **C)** *MRC1* in GBM TAMs. Statistical significance was assessed using the log-rank test ($p < 0.001$)

these markers are upregulated, correlating with poor prognosis (**Figure 3**). This dual role of anti-inflammatory macrophages—promoting tumor progression while suppressing immune responses—makes them integral to the tumor's metabolic reprogramming (32). Lactate uptake by TAMs has been shown to enhance their immunosuppressive phenotype, inhibiting immune surveillance and facilitating tumor progression (26). The uptake of lactate by TAMs via MCT1 and MCT2 transporters correlates with increased anti-inflammatory polarization (**Figure 2**), exacerbating the immunosuppressive environment and contributing to poor clinical outcomes in GBM patients. Elevated levels of lactate in the TME are known to activate anti-inflammatory polarization in TAMs, which is reflected in the observed increased expression of lactate transporters, such as MCT1 and MCT2, in GBM TAMs (33). These findings underscore the prognostic value of these TAM markers and pave the way for further investigations into the intricate interplay between TAMs and GBM.

Our study highlighted a significant association between anti-inflammatory-like TAM markers and poor survival in GBM patients. However, it is important to note that these relationships are observational, and causality cannot be established without functional studies and in vivo models. Further research is necessary to investigate the exact role of TAMs in GBM development and their influence on tumor progression. *KDM*s (lysine demethylases), such as *KDM5C*, *KDM5B*, *KDM7A,* and *KDM6B*, are highly expressed in GBM TAMs, correlating with poor survival outcomes (**Figure 5**). These results suggest that *KDMs* may regulate the immune microenvironment and contribute to immune evasion in GBM, potentially reinforcing the immunosuppressive TME. α-KG, a byproduct of the TCA cycle, serves as a cofactor for KDMs, which are essential for demethylating histones and activating anti-inflammatory-related genes (34). α-KG may promote anti-inflammatory polarization by enhancing prolyl hydroxylase domain (PHD) activity, which inhibits pro-inflammatory polarization and enhances anti-inflammatory activation via the Jumonji domain-containing protein D3 (α-KG-Jmjd3) pathway (35).

Our findings align with previous research showing that high expression of *LDHA*, *IDH1*, and other key metabolic regulators contribute to the altered metabolic phenotype of GBM, which may be linked to pro-inflammatory activation of TAMs and tumor progression (36). The polarization of TAMs to an anti-inflammatory phenotype, influenced by markers like CD206 (MRC1) and ARG1, plays a critical role in tumor progression (37, 38).

Additionally, lactate has been shown to activate anti-inflammatory polarization in TAMs, with transporters like MCT1 facilitating lactate uptake and supporting immunosuppressive environments in GBM (13, 39, 40). MCT1 and MCT2, which are involved in lactate transport, were found to be upregulated in GBM TAMs, further supporting the hypothesis that lactate contributes to anti-inflammatory polarization in these macrophages.

Building on these findings, we propose a model of the

https://doi.org/10.59720/24-258

Gene	Average enrichment score	Occurrences
KDM4B	0.775	4
KDM5B	0.833	8
KDM5C	0.588	1
KDM6B	0.77	8
KDM7A	0.7288	10
SLC16A1/MCT1	1.0	21
SLC16A7/MCT2	0.75	17

Table 1: Average enrichment scores of KDM and MCT genes. Results of GSEA for key KDM and MCT genes in GBM TAMs using the C7 immunological signature gene set. The table includes the average enrichment scores (ES) and the number of occurrences for selected genes involved in immune modulation and metabolic reprogramming. Occurrences represent how many times a given gene appears in the top-ranking gene sets across different samples. Enrichment Score (ES) indicates how strongly a gene is associated with a particular biological process or pathway. A high ES suggests the gene plays a key role in that process. The p-value for the analysis is 0.009, indicating that the results are statistically significant.

interaction between TAMs and tumor cells in GBM (**Figure 6**). The model suggests that tumor cells secrete lactate via HIF1A and α-KG through the Krebs cycle, which are subsequently absorbed by TAMs. Additionally, citrate is metabolized in TAMs, inducing higher levels of α-KG. This metabolic shift activates KDMs, leading to anti-inflammatory polarization. The model illustrates the crucial metabolic interactions within the TME that shape TAM polarization, contributing to GBM progression and highlighting potential therapeutic targets.

Therapeutically, targeting metabolic pathways involved in glycolysis and the TCA cycle could offer promising strategies for GBM treatment. Inhibiting *LDHA*, *IDH1*, and *HIF1A* could potentially reduce lactate and α-KG levels, impairing tumor metabolism and disrupting the immunosuppressive environment (32). Additionally, targeting *MCT1* and *MCT2* to block lactate uptake by TAMs may reduce their immunosuppressive phenotype and improve the efficacy of immunotherapies. Further studies are required to explore these therapeutic avenues and validate their clinical potential in GBM.

Dietary interventions that influence citrate, lactate, or α-KG levels in the TME could provide complementary therapeutic strategies. These interventions might help reprogram TAMs and enhance the overall response to treatment. TAMs play a critical role in GBM development, and their metabolic activities are significantly influenced by lactate. Therapeutic strategies aimed at reprogramming TAM metabolism—such as inhibiting lactate uptake via MCT1 and MCT2 inhibitors or targeting glycolytic enzymes—could potentially reduce the immunosuppressive phenotype of TAMs, improving the efficacy of immunotherapies. Furthermore, therapies that modulate α-KG production or its interactions with *KDMs* may influence TAM polarization, enhancing the therapeutic response in GBM.

Despite the promising insights from this study based on publicly available data from the TCGA and Genomic Data Commons (GDC), there are limitations. The absence of *in vitro* and *in vivo* experiments means our conclusions are primarily based on gene expression data, which may not fully capture direct functional interactions. Therefore, further experimental studies are needed to validate these findings and clarify the mechanisms underlying TAM polarization and metabolic reprogramming in GBM. Additionally, while our study

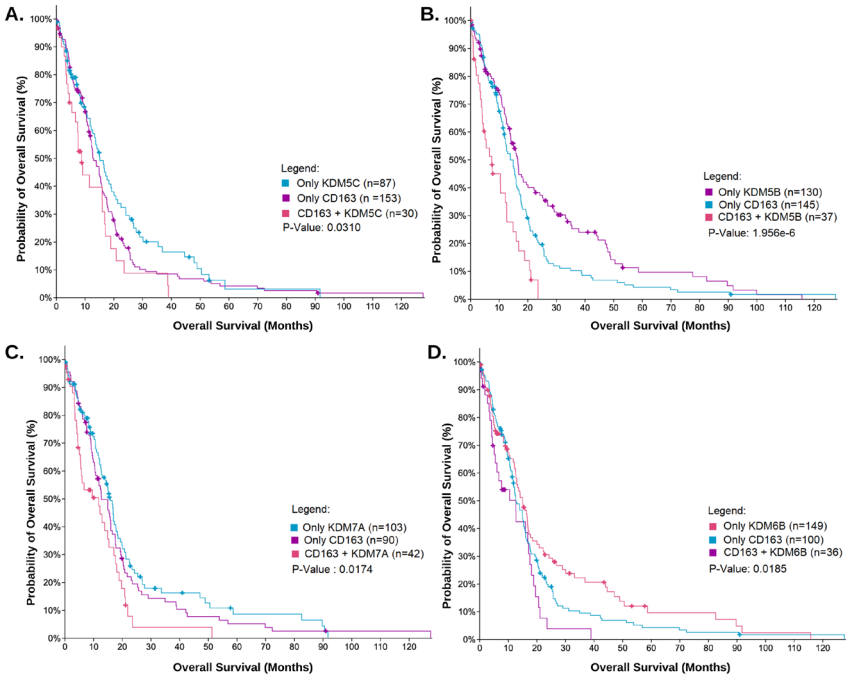

Figure 4: Survival analysis of individual and combined GBM tumor markers. Kaplan-Meier survival curves for GBM patients stratified by the expression levels of individual and combined markers: **A)** *CD163* & *KDM5C,* **B)** *CD163* & *KDM5B,* **C)** *CD163* & *KDM7A,* and **D)** *CD163* & *KDM6B***.** Statistical significance was assessed using the log-rank test, with p = **A)** 0.031, **B)** 1.956e-6, **C)** 0.0174, **D)** 0.085.

https://doi.org/10.59720/24-258

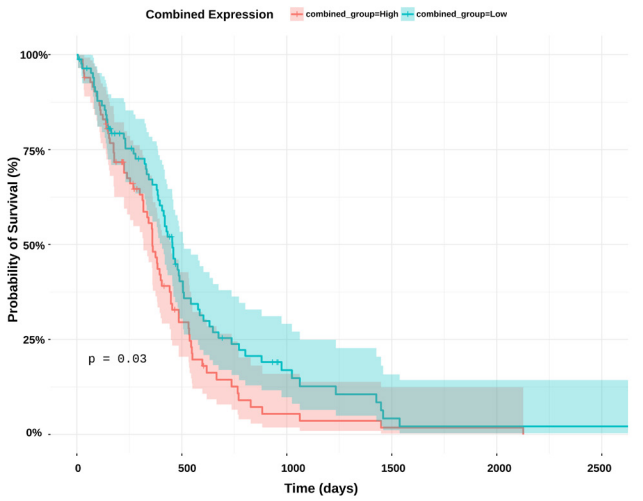

Figure 5: Survival analysis of combined anti-inflammatory TAM markers. Kaplan-Meier survival curves for GBM patients were stratified by expression levels of combined markers CD206, ARG1, and CD163. Statistical significance was assessed using the log-rank test, with a p-value of 0.03.

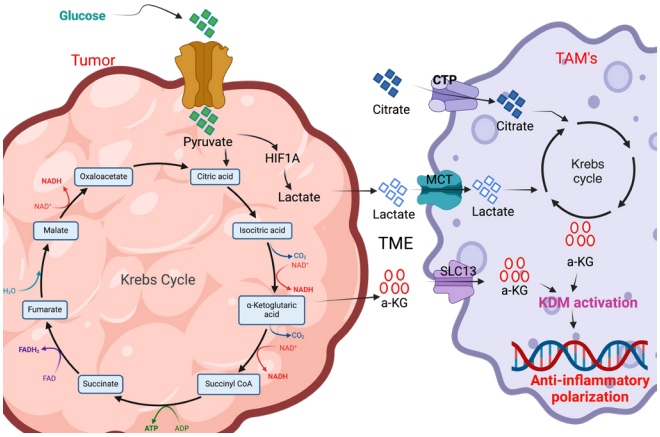

Figure 6: Hypothetical model of GBM TAM regulation. Model illustrating the metabolic crosstalk between GBM tumor cells and TAMs within the TME, emphasizing the roles of lysine demethylases (KDMs) and monocarboxylate transporters (MCTs) in metabolic reprogramming and immune evasion. Under hypoxic conditions driven by HIF1A, GBM tumor cells secrete lactate and α-KG. Lactate, exported via MCTs, contributes to an acidic TME, promoting immune evasion. α-KG, a byproduct of TCA cycle activity, serves as a key regulator of epigenetic and metabolic processes. TAMs absorb lactate via MCTs, driving metabolic shifts that polarize them toward the anti-inflammatory phenotype, supporting an immunosuppressive TME and tumor progression. Tumor-derived citrate is also metabolized within TAMs, producing additional α-KG that activates KDMs, further reinforcing anti-inflammatory polarization and immune suppression. Additionally, GBM tumor cells overexpress IDH1, an enzyme in the TCA cycle that converts isocitrate to α-KG. This overexpression represents a critical metabolic adaptation, enabling tumor growth and resistance to stress. The high levels of lactate and α-KG within the TME contribute to immunosuppression, reducing the efficacy of immunotherapies and promoting tumor survival. The model highlights potential therapeutic targets, such as inhibitors of KDMs, MCTs, and IDH1, to disrupt the metabolic interactions between tumor cells and TAMs. The figure was created using BioRender.com.

highlights the potential of targeting specific metabolic pathways and markers, further research is needed to identify the best strategies for therapeutic interventions and to determine their safety and efficacy in clinical settings.

In conclusion, our research underscores the critical roles of citrate, lactate, and α-KG in driving anti-inflammatory polarization in TAMs within the GBM TME. Targeting metabolic pathways, such as *LDHA*, *IDH1*, *HIF1A*, and *MCTs*, offers novel therapeutic strategies for GBM treatment. Future research should focus on identifying biomarkers related to citrate and lactate metabolism, validating these findings through preclinical and clinical studies, and exploring multimodal treatment strategies to improve patient survival.

MATERIALS AND METHODS

Our experimentation process involved a series of steps: Gene expression analysis, survival analysis, gene set enrichment analysis, and prognostic evaluation (**Figure 7**).

Gene expression analysis of the GBM tumor

We analyzed the expression dispersion of glycolysis and citric acid cycle genes in GBM tumor cells using the online software tool, Wanderer (41). The analysis included a dataset of 161 RNA-Seq samples (5 healthy controls and 156 GBM patient samples, 80% of which were from patients undergoing treatment) sourced from the TCGA database (42). The data were obtained from Illumina HiSeq RNA sequencing. Additionally, we used the R-script (CorrelateGenes_SurvivalAnalysis.R) in R Studio to identify correlations between TAM and metabolic genes in the GBM tumor dataset. Gene expression values for glycolysis, the TCA cycle, and *HIF1A* were visualized using RNA-Seq by Expectation-Maximization (RSEM) formatted data. Furthermore, we calculated the changes in log2(x+1) transformed (normalized by applying the base-2 logarithm after adding 1 to each value) RSEM normalized counts of cancer-related glycolysis and citric acid cycle genes between average healthy controls and GBM cancer samples.

Survival analysis of GBM TAM markers

In the GDC portal, we filtered data using “TCGA-GBM”, resulting in 523 cases, followed by the selection of “Gene Expression Quantification” as the data type, narrowing the cases to 169 (Project Id IS TCGA-GBM AND Data Type IS Gene Expression Quantification). We extracted the patient data (transcriptome profiling) with detailed follow-up and survival information from the downloaded clinical data from these 169 cases. We additionally utilized RNA-Seq read count data from 773 GTEx normal brain samples, including data from the cortex, frontal cortex, and anterior cingulate cortex (43). All the downloaded transcriptome profiling files were combined into a single csv file with only tpm-unstranded data by running the R-script (mergeGETables.R) in R Studio. Then we mapped the filename and submitter ID to correlate the clinical data and gene expression table. (e.g., a4e0b059 and TCGA-02-0047). We used GraphPad Prism to plot a Kaplan-Meier survival curve of the selected GBM TAM marker or gene with a p-value < 0.001.

https://doi.org/10.59720/24-258

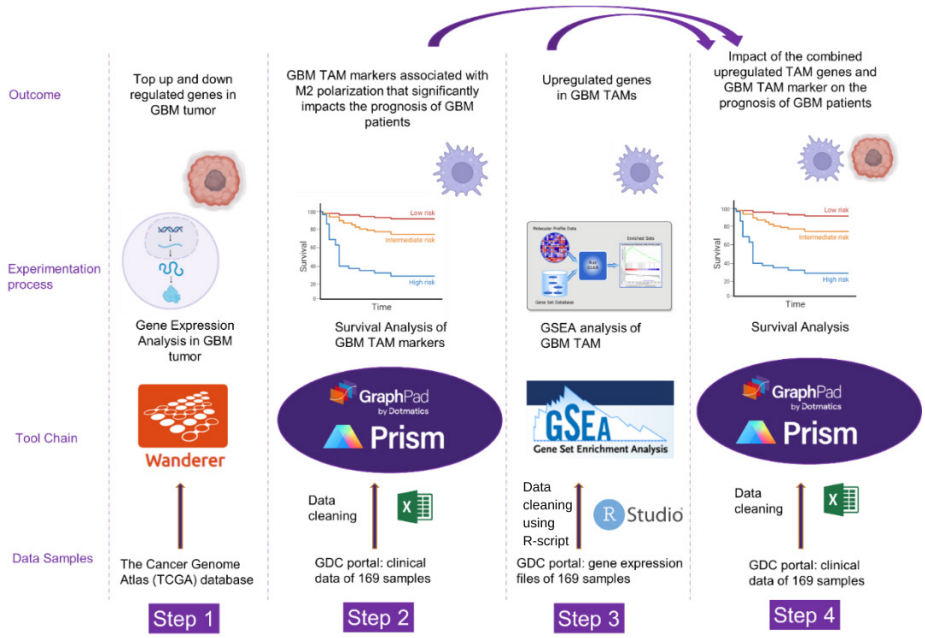

Figure 7: Stepwise experimentation process for GBM TAM analysis. Experimental approach used to identify key findings in GBM TAM analysis. The process includes four major steps: 1) gene expression analysis of glycolysis and citric acid cycle genes in GBM tumor cells, 2) survival analysis of GBM TAM markers, 3) gene set enrichment analysis in TAMs, and 4) prognostic evaluation of combined GBM TAM markers. Key findings from each step, such as gene expression differences, survival correlations, and functional enrichment, are summarized.

Gene set enrichment analysis

An R script was used to prepare the necessary data for analysis. GSEA was performed to identify whether certain gene sets are overrepresented at the top or bottom of a ranked list of genes within the GBM-TAM context. Gene expression data for all genes in all samples, formatted as TPM-unstranded, were combined into a CSV file. A subset of relevant genes was extracted from this dataset and formatted into a gene cluster text (GCT) file (44). The GCT file, along with a phenotype file, was then input into the GSEA software for analysis.

For the analysis, TAMs in GBM were defined based on the expression of established TAM marker genes, such as CD163 and MRC1. To ensure the analysis focused on macrophage-specific gene expression, TAM-specific profiles were extracted by filtering the dataset to include only samples expressing these marker genes. This approach mitigated the issue of analyzing bulk tumor data, which includes gene expression from all cells in the tumor microenvironment (TME). Using this TAM-specific subset, we conducted gene set enrichment analysis to identify overrepresented genes from the *KDM* and *MCT* classes associated with TAMs in GBM.

Survival analysis of combined GBM TAM markers

We investigated the prognostic value of tumor-associated genes and markers of GBM TAMs, as well as their combined occurrence, by plotting survival curves for the selected markers or genes with a p-value < 0.001 using GraphPad Prism. Additionally, we used the R-script (CorrelateGenes_SurvivalAnalysis.R) in R Studio to generate a survival graph for the combined *CD206, CD163,* and *ARG1* in the GBM tumor dataset.

ACKNOWLEDGMENTS

We acknowledge Swaminathan Arunachalam and Bhavanthi Thirumanam Raghukumar (Nikila's parents) for their support in getting the tools access, setting up the environment for the experiments, and reviewing the research paper. In addition, we would like to thank Dr. Jagan Lakshmanan, Nithyasree Murugan, Sasikala Anbarasan, and Sushmita Halder for reviewing and commenting on this research paper.

Received: March 25, 2024
Accepted: December 19, 2024
Published: June 28, 2025

REFERENCES

1. "Glioblastoma Multiforme Translational Center of Excellence - Abramson Cancer Center, Penn Medicine." Penn Medicine - Abramson Cancer Center, www.pennmedicine.org/cancer/cancer-research/translating-research-to-practice/gbm-tce. Accessed 16 July 2024.
2. Glioblastoma Multiforme Translational Center of Excellence - Abramson Cancer Center, Penn Medicine." Penn Medicine - Abramson Cancer Center
3. Louis, David N, et al. "The 2016 World Health Organization Classification of Tumors of the Central Nervous System: A Summary." *Acta Neuropathologica*, vol. 131, no. 6, 9 May 2016, pp. 803–20, https://doi.org/10.1007/s00401-016-1545-1.
4. Viola, Antonella, et al. "The Metabolic Signature of Macrophage Responses." *Frontiers in Immunology,* vol. 10, 3 July 2019, pp. 1462, https://doi.org/10.3389/fimmu.2019.01462.
5. Liu, Yan-Cun, et al. "Macrophage Polarization in Inflammatory Diseases." *International Journal of Biological Sciences*, vol. 10, no. 5, 1 May 2014, pp. 520–29, https://

doi.org/10.7150/ijbs.8879.
6. Pan, Yueyun, et al. "Tumor-Associated Macrophages in Tumor Immunity." *Frontiers in Immunology*, vol. 11, 3 Dec. 2020, https://doi.org/10.3389/fimmu.2020.583084.
7. Martínez-Reyes, Inmaculada, and Navdeep S. Chandel. "Cancer Metabolism: Looking Forward." *Nature Reviews Cancer*, vol. 21, no. 10, 16 July 2021, pp. 669–80, https://doi.org/10.1038/s41568-021-00378-6.
8. Xie Sun-Zhe, et al. "The Critical Function of Metabolic Reprogramming in Cancer Metastasis." *Aging and Cancer*, vol. 3, no. 1, 23 Feb. 2022, pp. 20–43, https://doi.org/10.1002/aac2.12044.
9. Chen, Degao, et al. "Metabolic Regulatory Crosstalk between Tumor Microenvironment and Tumor-Associated Macrophages." *Theranostics*, vol. 11, no. 3, 2021, pp. 1016–30, https://doi.org/10.7150/thno.51777.
10. Vander Heiden, M. G, et al. "Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation." *Science*, vol. 324, no. 5930, 21 May 2009, pp. 1029–33, https://doi.org/10.1126/science.1160809.
11. Zhuang, Yan. "Hypoxia Signaling in Cancer: Implications for Therapeutic Interventions." *MedComm*, vol. 4, no. 1, 23 Jan. 2023, p. e203. https://doi.org/10.1002/mco2.203.
12. Li, Yue, et al. "Hypoxia and the Tumor Microenvironment." *Technology in Cancer Research & Treatment*, vol. 20, 1 Jan. 2021, https://doi.org/10.1177/15330338211036304.
13. Li, Xiaolu, et al. "Lactate Metabolism in Human Health and Disease." *Signal Transduction and Targeted Therapy*, vol. 7, no. 1, 1 Sept. 2022, pp. 1–22, https://doi.org/10.1038/s41392-022-01151-3.
14. Noe, Jordan T, et al. "Lactate Supports a Metabolic-Epigenetic Link in Macrophage Polarization." *Science Advances*, vol. 7, no. 46, 12 Nov. 2021, https://doi.org/10.1126/sciadv.abi8602.
15. Thapa, Bikash, and Keunwook Lee. "Metabolic Influence on Macrophage Polarization and Pathogenesis." *BMB Reports*, vol. 52, no. 6, 30 June 2019, pp. 360–72, https://doi.org/10.5483/bmbrep.2019.52.6.140.
16. Tao, Hongxia, et al. "Unveiling the Veil of Lactate in Tumor-Associated Macrophages: A Successful Strategy for Immunometabolic Therapy." *Frontiers in Immunology*, vol. 14, 26 July 2023, https://doi.org/10.3389/fimmu.2023.1208870
17. Taylor, Cormac T., and Carsten C. Scholz. "The Effect of HIF on Metabolism and Immunity." *Nature Reviews Nephrology,* vol. 18, no. 9, 20 June 2022, pp. 573–87, https://doi.org/10.1038/s41581-022-00587-8.
18. Torres-Castro, Israel, et al. "Human Monocytes and Macrophages Undergo M1-Type Inflammatory Polarization in Response to High Levels of Glucose." *Immunology Letters,* vol. 176, 1 Aug. 2016, pp. 81–9, https://doi.org/10.1016/j.imlet.2016.06.001.
19. Lin, Jieshan, et al. "High Glucose Contributes to the Polarization of Peritoneal Macrophages to the M2 Phenotype in Vivo and in Vitro." Molecular Medicine Reports, vol. 22, no. 1, May 2020, pp. 127–34, https://doi.org/10.3892/mmr.2020.11130.
20. Zheng, Jie. "Energy Metabolism of Cancer: Glycolysis versus Oxidative Phosphorylation (Review)." *Oncology Letters*, vol. 4, no. 6, 1 Dec. 2012, pp. 1151–7, https://doi.org/10.3892/ol.2012.928.
21. Icard, Philippe, et al. "Understanding the Central Role of Citrate in the Metabolism of Cancer Cells and Tumors: An Update." *International Journal of Molecular Sciences*, vol. 22, no. 12, 19 June 2021, p. 6587, https://doi.org/10.3390/ijms22126587.
22. Schiliro, Chelsea, and Bonnie L. Firestein. "Mechanisms of Metabolic Reprogramming in Cancer Cells Supporting Enhanced Growth and Proliferation." *Cells*, vol. 10, no. 5, 29 Apr. 2021, p. 1056, https://doi.org/10.3390/cells10051056.
23. Sun, Jiakang, et al. "Mitochondrial and Plasma Membrane Citrate Transporters: Discovery of Selective Inhibitors and Application to Structure/Function Analysis." *Molecular and Cellular Pharmacology*, vol. 2, no. 3, 2010, pp. 101–10, www.ncbi.nlm.nih.gov/pmc/articles/PMC2913483/.
24. Yoo, Hee Chan, et al. "Glutamine Reliance in Cell Metabolism." *Experimental & Molecular Medicine*, vol. 52, no. 9, 1 Sept. 2020, pp. 1496–516, https://doi.org/10.1038/s12276-020-00504-8.
25. Mehla, Kamiya, and Pankaj K. Singh. "Metabolic Regulation of Macrophage Polarization in Cancer." *Trends in Cancer*, vol. 5, no. 12, Nov. 2019, pp. 822–34, https://doi.org/10.1016/j.trecan.2019.10.007.
26. Boutilier, Ava J., and Sherine F. Elsawa. "Macrophage Polarization States in the Tumor Microenvironment." *International Journal of Molecular Sciences*, vol. 22, no. 13, 29 June 2021, p. 6995, https://doi.org/10.3390/ijms22136995.
27. Wang, Zi-Hao et al. "Lactate in the tumour microenvironment: From immune modulation to therapy." *EBioMedicine*, vol. 73, 2021, https://doi.org/10.1016/j.ebiom.2021.103627.
28. Icard, Philippe, et al. "Why May Citrate Sodium Significantly Increase the Effectiveness of Transarterial Chemoembolization in Hepatocellular Carcinoma?" *Drug Resistance Updates*, vol. 59, Dec. 2021, p. 100790, https://doi.org/10.1016/j.drup.2021.100790.
29. Solomou, G., Finch, A., Asghar, A., & Bardella, C. (2023). Mutant IDH in Gliomas: Role in Cancer and Treatment Options. *Cancers*, vol. 15, no. 11, p. 2883, https://doi.org/10.3390/cancers15112883.
30. Rakina, M., Larionova, I., & Kzhyshkowska, J. (2024). Macrophage diversity in human cancers: New insight provided by single-cell resolution and spatial context. *Heliyon*, vol. 10, no. 7, https://doi.org/10.1016/j.heliyon.2024.e28332.
31. Calvert, Andrea E., et al. "Cancer-Associated IDH1 Promotes Growth and Resistance to Targeted Therapies in the Absence of Mutation." *Cell Reports*, vol. 19, no. 9, May 2017, pp. 1858–73, https://doi.org/10.1016/j.celrep.2017.05.014.
32. Lum, J. J., et al. "The Transcription Factor HIF-1 Plays a Critical Role in the Growth Factor- Dependent Regulation of Both Aerobic and Anaerobic Glycolysis." *Genes & Development*, vol. 21, no. 9, 16 Apr. 2007, pp. 1037–49, https://doi.org/10.1101/gad.1529107.
33. Feng, Yawei et al. "The Role of TAMs in Tumor Microenvironment and New Research Progress." *Stem Cells International*, vol. 2022, 15 Aug. 2022, pp. 1–11, https://doi.org/10.1155/2022/5775696.
34. Tao, Hongxia et al. "Unveiling the veil of lactate in tumor-associated macrophages: a successful strategy for immunometabolic therapy." *Frontiers in immunology,* vol. 14, 26

https://doi.org/10.59720/24-258

Jul. 2023, https://doi.org/10.3389/fimmu.2023.1208870.
35. Tran, Thai Q., et al. "Molecular Pathways: Metabolic Control of Histone Methylation and Gene Expression in Cancer." *Clinical Cancer Research*, vol. 23, no. 15, 1 Aug. 2017, pp. 4004–9, https://doi.org/10.1158/1078-0432.ccr-16-2506.
36. Liu, Shaojuan, et al. "The Regulatory Role of α-Ketoglutarate Metabolism in Macrophages." *Mediators of Inflammation*, vol. 2021, 29 Mar. 2021, pp. 1–7, https://doi.org/10.1155/2021/5577577.
37. Viola, Antonella et al. "The Metabolic Signature of Macrophage Responses." *Frontiers in Immunology*, vol. 10, 3 July 2019, https://doi.org/10.3389/fimmu.2019.01462.
38. Chen, Yibing, et al. "Tumor-Associated Macrophages: An Accomplice in Solid Tumor Progression." *Journal of Biomedical Science*, vol. 26, no. 1, 20 Oct. 2019, p. 78, https://doi.org/10.1186/s12929-019-0568-z.
39. de Groot, Amber E., and Kenneth J. Pienta. "Epigenetic Control of Macrophage Polarization: Implications for Targeting Tumor-Associated Macrophages." *Oncotarget*, vol. 9, no. 29, 7 Feb. 2018, pp. 20908-27, https://doi.org/10.18632/oncotarget.24556.
40. Morris, Marilyn E, and Melanie A Felmlee. "Overview of the Proton-Coupled MCT (SLC16A) Family of Transporters: Characterization, Function and Role in the Transport of the Drug of Abuse γ-Hydroxybutyric Acid." *The AAPS Journal,* vol. 10, 2, 1 June 2008, pp. 311-21, https://doi.org/10.1208/s12248-008-9035-6.
41. Miranda-Gonçalves, Vera, et al. "Hypoxia-Mediated Upregulation of MCT1 Expression Supports the Glycolytic Phenotype of Glioblastomas." *Oncotarget*, vol. 7, no. 29, 16 June 2016, pp. 46335–53, https://doi.org/10.18632/oncotarget.10114.
42. «Wanderer.» MapLab, www.maplab.imppc.org/wanderer/. Accessed 16 July 2024.
43. «Cancer Genomics." National Cancer Institute, www.cancergenome.nih.gov. Accessed 16 July 2024.
44. "GTEx portal" Broad Institute of MIT and Harvard, www.gtexportal.org/home/datasets. Accessed 6 December 2024.
45. «GSEA User Guide.» Broad Institute, www.gsea-msigdb.org/gsea/doc/GSEAUserGuideFrame.html. Accessed 16 July 2024.

https://doi.org/10.59720/24-258

APPENDIX

mergeGETables.R

```
# This R program takes individual gene expression sample files downloaded from TCGA as
# input and creates a consolidated file with each sample as one column and different genes as
# rows.

#### first few list of the .tsv file downloaded for each sample from TCGA database will be like below ####
#gene_id   gene_name   gene_type   unstranded  stranded_first  stranded_second tpm_unstranded  fpkm_unstranded fpkm_uq_
unstranded
#N_unmapped         5456436 5456436 5456436
#N_multimapping          3426782 3426782 3426782
#N_noFeature           3860513 31085642    30811764
#N_ambiguous           5775384 1332069 1355930
#ENSG00000000003.15 TSPAN6  protein_coding  4167    2114    2053    67.1411 18.1950 18.5714
#ENSG00000000005.6  TNMD    protein_coding  1   0   1   0.0495  0.0134  0.0137

# Loading necessary library
library(dplyr)
# Set the working directory to the location where all the TCGA files are stored
setwd("/Users/xyz/GBM/Data/TCGA")
# Iterate through the list of sample files in the directory
for (filename in list.files()) {
  # If no dataset exists yet, create the dataset
  if (!exists("dataset")) {
    # Read the file and create a dataset
    dataset <- read.delim(filename, stringsAsFactors = FALSE)
    # Remove the first 4 irrelevant rows
    dataset <- dataset %>% filter(!row_number() %in% c(1, 2, 3, 4))
    # Select only relevant columns (gene_id, gene_name, tpm_unstranded)
    dataset <- dataset[c(1, 2, 7)]
    # Create a new column name based on the first part of the filename
    col_name <- sapply(strsplit(filename, split='-', fixed=TRUE), function(x) (x[1]))
    colnames(dataset)[3] <- col_name
  } else {
    # If the dataset exists, process the new file and append to the dataset
    temporary <- read.delim(filename, stringsAsFactors = FALSE)
    temporary <- temporary %>% filter(!row_number() %in% c(1, 2, 3, 4))
    # Select only relevant columns (gene_id, tpm_unstranded)
    temporary <- temporary[c(1, 7)]
    # Create a new column name based on the first part of the filename
    col_name <- sapply(strsplit(filename, split='-', fixed=TRUE), function(x) (x[1]))
    colnames(temporary)[2] <- col_name
    # Join the new data with the existing dataset by gene_id
    dataset <- full_join(dataset, temporary, by = "gene_id")
    # Delete the temporary dataset to free up memory
    rm(temporary)
  }
}
# Write the consolidated dataset into an output CSV file
write.csv(dataset, file="/Users/xyz/GBM/Data/combine.csv", row.names = FALSE)
#Sample output file's first 2 lines
#gene_id   gene_name   03ddea6f    051dc36f    075029c6    08dce278    0a80d60d    0ad4aa09    0c6b79d0    0ce1eb02    0d3fac43
0db34205    0f89ce0f    1069403e    1076483a    113df549    139fe07b    13da7123    1435785d    1483c347    1871e575    19f6db33
1a61593f    1f027bee    209797f2    210daf0d    23f3b0c8    24397ba5    251de1f4    2856d609    2acc7210    2c25aa49    2e33127c
32291119    325f91c9    328b149f    ...
#ENSG00000000003.15 TSPAN6  5.8836  111.9005    54.7729 34.3045 67.1411 114.9393    52.4703 60.4303 44.6427 55.9218 93.3216
66.0662 70.8772 21.6593 71.9519 76.9722 65.7113 81.7969 82.0863 56.7117 105.1259    106.8657    61.4441 81.9417 66.2069 44.7954
146.2413    NA  70.3354 40.8241 46.4405 129.5973    128.4459    49.6363 58.244  79.3682 65.4239 85.0811 92.3314 …
```

https://doi.org/10.59720/24-258

CorrelateGenes_SurvivalAnalysis.R

The code is a comprehensive analysis pipeline that performs the following steps:

- **Package Installation:** The necessary R packages (dplyr, tibble, tidyverse, googledrive, DESeq2, pheatmap, ggplot2, survival, survminer) are installed if not already present. These packages support data manipulation, visualization, and survival analysis.
- **Google Drive Authentication:** The script authenticates access to Google Drive to download raw count data files for normal and tumor samples (normal_raw_counts.csv and tumor_raw_counts.csv).
- **Data Cleaning and Merging:** The raw count data for normal and tumor samples is loaded, unnecessary columns are removed, and column names are standardized. The datasets are then merged by gene_id, with missing values replaced by zeros. The merged dataset is saved with gene IDs as row names.
- **Gene Selection:** A list of genes related to tumor-associated macrophages (TAM) and metabolic pathways is defined. These genes are extracted from the merged dataset for further analysis.
- **Correlation Analysis:** A Spearman correlation matrix is computed for the selected genes, and the results are visualized in a heatmap using the pheatmap package.
- **Survival Analysis:** Metadata for tumor samples is read and filtered to include only “Alive” or “Dead” statuses. Survival time and status columns are created, and data is aligned with the expression data. Expression levels of immune-related markers (CD206, ARG1, CD163) are extracted, and a combined expression value is calculated.
- **Kaplan-Meier Survival Curve:** The combined expression of the selected markers is categorized into “High” and “Low” groups based on the median value. A Kaplan-Meier survival curve is generated to evaluate the relationship between the expression of these markers and patient survival. The plot is saved as a PNG file.

```
# Install necessary packages if not already installed
# This section ensures that all required libraries are installed and loaded
install.packages(“dplyr”)
install.packages(“tibble”)
install.packages(“tidyverse”)
install.packages(“googledrive”)

# Install Bioconductor packages for DESeq2 and visualization tools
if (!requireNamespace(“BiocManager”, quietly = TRUE)) install.packages(“BiocManager”)
BiocManager::install(“DESeq2”)
BiocManager::install(“pheatmap”)
BiocManager::install(“ggplot2”)

# Install required libraries (if not already installed)
if (!requireNamespace(“survival”, quietly = TRUE)) install.packages(“survival”)
if (!requireNamespace(“survminer”, quietly = TRUE)) install.packages(“survminer”)

# Load required libraries
library(dplyr)
library(tibble)
library(tidyverse)
library(googledrive)
library(DESeq2)
library(ggplot2)
library(pheatmap)
library(survival)
library(survminer)

drive_auth()  # Authenticate Google Drive access

# --- Download Raw Count Data ---
# Download the cleaned and aligned normal (GTEX) and tumor raw counts (TCGA)
file <- drive_get(“normal_raw_counts.csv”)
drive_download(file, path = “normal_raw_counts.csv”, overwrite = TRUE)

file <- drive_get(“tumor_raw_counts.csv”)
drive_download(file, path = “tumor_raw_counts.csv”, overwrite = TRUE)

# Load and clean raw counts data
normal_counts <- read.csv(“normal_raw_counts.csv”)
tumor_counts <- read.csv(“tumor_raw_counts.csv”)

# Standardize column names and remove unnecessary columns
normal_counts <- normal_counts[, !(colnames(normal_counts) %in% “Description”)]
tumor_counts <- tumor_counts[, !(colnames(tumor_counts) %in% “Description”)]
colnames(normal_counts)[colnames(normal_counts) == “Name”] <- “gene_id”

# Preview cleaned datasets
cat(“Normal counts columns after cleaning:\n”)
```

https://doi.org/10.59720/24-258

```
print(colnames(normal_counts))
cat("Tumor counts columns after cleaning:\n")
print(colnames(tumor_counts))

# --- Merge Datasets ---
# Identify and summarize common and unique genes
common_genes <- intersect(tumor_counts$gene_id, normal_counts$gene_id)
tumor_only_genes <- setdiff(tumor_counts$gene_id, normal_counts$gene_id)
normal_only_genes <- setdiff(normal_counts$gene_id, tumor_counts$gene_id)

cat("Number of common genes:", length(common_genes), "\n")
cat("Number of genes only in tumor_counts:", length(tumor_only_genes), "\n")
cat("Number of genes only in normal_counts:", length(normal_only_genes), "\n")

# Merge data based on 'gene_id'
combined_counts <- merge(tumor_counts, normal_counts, by = "gene_id", all = FALSE)
combined_counts[is.na(combined_counts)] <- 0
rownames(combined_counts) <- combined_counts$gene_id
combined_counts <- combined_counts[,-1]

# Verify dimensions
cat("Number of rows and columns in the combined dataset:\n")
cat(nrow(combined_counts), "rows\n")
cat(ncol(combined_counts), "columns\n")

# Save row information for reference
row_info <- data.frame(row_number = 1:nrow(combined_counts), gene_id = rownames(combined_counts))
write.table(row_info, file = "Gene_id-sequence.csv", sep = ",", quote = FALSE, row.names = FALSE, col.names = TRUE)
cat("Row numbers and gene IDs written to Gene_id-sequence.csv\n")

# --- Gene Selection and Analysis ---
# Define TAM-related and metabolic genes of interest
tam_genes <- c("ENSG00000177575.13","ENSG00000118520.15","ENSG00000007171.18",
            "ENSG00000232810.4","ENSG00000260314.3","ENSG00000180425.11",
            "ENSG00000134595.9","ENSG00000038945.15")

metabolic_genes <- c("ENSG00000100644.17", "ENSG00000122966.17",
                  "ENSG00000134333.14", "ENSG00000138413.14",
                  "ENSG00000168679.18")

selected_genes <- c(tam_genes, metabolic_genes)
selected_counts <- combined_counts[rownames(combined_counts) %in% selected_genes, ]

# Calculate Spearman correlation and visualize as a heatmap
correlation_matrix <- cor(t(selected_counts), method = "spearman")
pheatmap(correlation_matrix,
       cluster_rows = TRUE,

       cluster_cols = TRUE,

         main = "Correlation between TAM-related and Metabolic Genes")

cat("Correlation analysis and heatmap generation completed.\n")
#############Survival Analysis############################
# The following code includes survival analysis based on the combined expression of immune-related markers
# (CD206, ARG1, CD163) in TNBC tumor samples. The Kaplan-Meier survival curve and statistical analysis will
# help evaluate the correlation between the expression levels of these markers and patient survival.

# Read the normalized counts CSV of the tumor and set the first column as row names (this is the output of the previous R script)
tumor_normalized_counts <- read.csv("combine.csv", row.names = 1)

# Load metadata
metadata_surv <- read.csv("metadata_tumorsamples.csv")  # Replace with your file name

# Filter rows to include only "Alive" or "Dead"; remove "not reported" status
metadata_surv <- metadata_surv[metadata_surv$vital_status %in% c("Alive", "Dead"), ]

# Create a survival time column
metadata_surv$time <- ifelse(metadata_surv$vital_status == "Alive",
                     metadata_surv$days_to_last_follow_up,
```

https://doi.org/10.59720/24-258

```
                    metadata_surv$days_to_death)

metadata_surv$time <- as.numeric(as.character(metadata_surv$time))

# Create a survival status column (1 = deceased, 0 = alive)
metadata_surv$status <- ifelse(metadata_surv$vital_status == "Alive", 0, 1)

# Check the metadata
head(metadata_surv[, c("Case.ID", "SampleID", "time", "status", "age_at_diagnosis")])

# Remove the 'X' prefix from column names if they start with 'X' followed by digits
colnames(tumor_normalized_counts) <- gsub("^X(\\d+)", "\\1", colnames(tumor_normalized_counts))

# Confirm that the column names are updated correctly
print(head(colnames(tumor_normalized_counts)))

cat("Columns Names of the Tumor normalized counts:", colnames(tumor_normalized_counts), "\n")

# Align metadata_surv with tumor_normalized_counts
metadata_surv <- metadata_surv[metadata_surv$SampleID %in% colnames(tumor_normalized_counts), ]
tumor_normalized_counts <- tumor_normalized_counts[, colnames(tumor_normalized_counts) %in% metadata_surv$SampleID]

# Reorder metadata_surv to match tumor_normalized_counts
metadata_surv <- metadata_surv[match(colnames(tumor_normalized_counts), metadata_surv$SampleID), ]

# Confirm the filtered dataset
cat("Number of columns after alignment:", ncol(tumor_normalized_counts), "\n")
cat("Number of samples in metadata after alignment:", nrow(metadata_surv), "\n")

# Check the first few entries of each
head(metadata_surv$SampleID)
head(colnames(tumor_normalized_counts))

# Check lengths
length(metadata_surv$SampleID)
length(colnames(tumor_normalized_counts))

anyNA(metadata_surv$SampleID)  # Should return FALSE
anyNA(colnames(tumor_normalized_counts))  # Should return FALSE

setdiff(metadata_surv$SampleID, colnames(tumor_normalized_counts))  # IDs in metadata but not in counts
setdiff(colnames(tumor_normalized_counts), metadata_surv$SampleID)  # IDs in counts but not in metadata

all(metadata_surv$SampleID == colnames(tumor_normalized_counts))  # Should return TRUE
print(metadata_surv$SampleID)

# Check alignment after filtering
cat("Metadata rows: ", nrow(metadata_surv), "\n")
cat("Expression columns: ", ncol(tumor_normalized_counts), "\n")

if (!all(metadata_surv$SampleID == colnames(tumor_normalized_counts))) {
  stop("Mismatch between metadata and tumor_normalized_counts!")
}

# Now, create the expression columns in 'metadata_surv'
metadata_surv$CD206_expression <- as.numeric(tumor_normalized_counts["ENSG00000260314.3", ]) # CD206
metadata_surv$ARG1_expression <- as.numeric(tumor_normalized_counts["ENSG00000118520.15", ]) # ARG1
metadata_surv$CD163_expression <- as.numeric(tumor_normalized_counts["ENSG00000177575.13", ]) # CD163

# Combine the expression values for the markers (ensure to exclude NAs)
metadata_surv$combined_expression <- rowMeans(metadata_surv[, c("CD206_expression", "ARG1_expression", "CD163_expression")],
na.rm = TRUE)

# Stratify combined expression into High/Low groups based on the median value
metadata_surv$combined_group <- ifelse(metadata_surv$combined_expression > median(metadata_surv$combined_expression, na.rm =
TRUE), "High", "Low")


# Create survival object
surv_obj <- Surv(time = metadata_surv$time, event = metadata_surv$status)
```

https://doi.org/10.59720/24-258

```
# Fit Kaplan-Meier model using the combined expression groups
(High vs Low)
surv_fit_combined <- survfit(surv_obj ~ combined_group, data =

metadata_surv)


# Plot Kaplan-Meier curve for combined expression values
km_plot <- ggsurvplot(
  surv_fit_combined,
  data = metadata_surv,
  pval = TRUE,
  conf.int = TRUE,
  risk.table = TRUE,
  title = "Survival Analysis for Combined Expression of CD206,
ARG1, and CD163",
  legend.title = "Combined Expression",
  xlab = "Time (days)",
  ylab = "Survival Probability (%)",
  ggtheme = theme_minimal()
)

# Adjust y-axis to display percentages
km_plot$plot <- km_plot$plot +
  scale_y_continuous(
    labels = scales::percent_format(accuracy = 1)  # Convert to
percentages with 1 decimal place
  )

# Print the Kaplan-Meier plot
print(km_plot)
# Save the plot
ggsave("combined_expression_survival_curve.png",
       plot = km_plot$plot,
       dpi = 600,
       width = 10,
       height = 8,
       units = "in",
       bg = "white")
```