Finding Cell-Type Specific Glioblastoma Marker Genes

Shashank Katiyar and Jonathan Zhu

Abstract

Glioblastoma, the most common form of brain cancer, is difficult to diagnose due to uncharacteristic symptoms and the need for MRI or other advanced medical procedures. Previous work regarding the genetic differences of Glioblastoma cells has been performed, but not with respect to specific marker genes. We analyzed a single-cell RNAseq Glioblastoma dataset, with the aim of finding relevant biomarkers in the form of differentially expressed genes in cancer vs normal cells, celltype-wise. We utilized a combination of existing methods to label cell types and find marker genes based on expression levels (with Seurat) and copy number (copykat). Combined, the top produced hits for various cell types were NUDT1 for endothelial cells, CACNA1A for neurons, and MAP1B for OPCs; we examined the relationships of these genes to glioblastoma in the literature and highlight their expression in normal human brain. Initially, we hoped these biomarkers could lead to significant insights into which genes are correlated with the phenotypic differences between cancer cells and normal cells, specific to a particular cell type. While these genes showed some relation to glioblastoma and show some promise as therapeutic targets and risk assessment genes, more research is necessary to confirm their relationship. Regardless, our results and pipeline can help other researchers (looking into the biological data through different lenses) focus their analyses on these and other genes that seem to be correlated with the cancer phenotype.

Background

Glioblastoma is the most common and aggressive form of brain cancer (Ohgaki and Kliehues, 2013) whose main causes are unknown, but several risk factors have been identified, including prior radiotherapy and Li-Fraumeni syndrome (Gallego, 2015). While the majority of glioblastomas are considered primary and arise without prior indication, a small number of them are considered secondary and arise from prior astrocytomas (a type of brain tumor-specific to astrocytes). Glioblastoma symptoms can range from mood and personality changes to headaches and seizures (Aliferis and Trafalis, 2015) and so cannot be determined through symptoms alone, typically requiring MRI for full diagnosis. Furthermore, the success of glioblastoma treatments (including surgery) have low success rates and result in patient mortality, though these numbers have been increasing (Gallego, 2015).

Research has already been conducted into the genetic aspect of glioblastoma. Parsons et al. (2011) conducted a genome-wide analysis of glioblastoma DNA and identified a significant number of candidate cancer genes (CANs). Though their analysis focused on sequence differences, this study did perform some transcriptional analysis, and the scientists identified over 100 genes that were differentially expressed between glioblastoma multiforme (GBMs, the most common type of glioblastoma) and normal human brain. While these transcriptional data can be the starting point for new therapeutics, such a profiling of a potentially cancerous brain can be expensive. As such, we seek to utilize similar expression data in GBMs and normal human brains to look for potential disease biomarkers that can more easily and confidently identify a glioblastoma, similar to the marker genes for the variety of brain cells.

Dataset

We analyzed a single dataset of glioblastoma in the human brain from the 10x genomics datasets website. It is a set of 1,615 cells from a patient with glioblastoma (2k Sorted Cells from Human Glioblastoma Multiforme, 3' v3.1).

Methods

We followed the following pipeline (Fig. 1) to process our data and find marker genes for glioblastoma cancer cells. Firstly, the cells needed to be identified as cancerous or healthy; copykat is able to do this by examining cell ploidy. Cells also need to be identified for their specific type based on known and common marker genes. Following this, each cell type was analyzed for different marker genes based on gene expression (Seurat) and copy number (copykat).

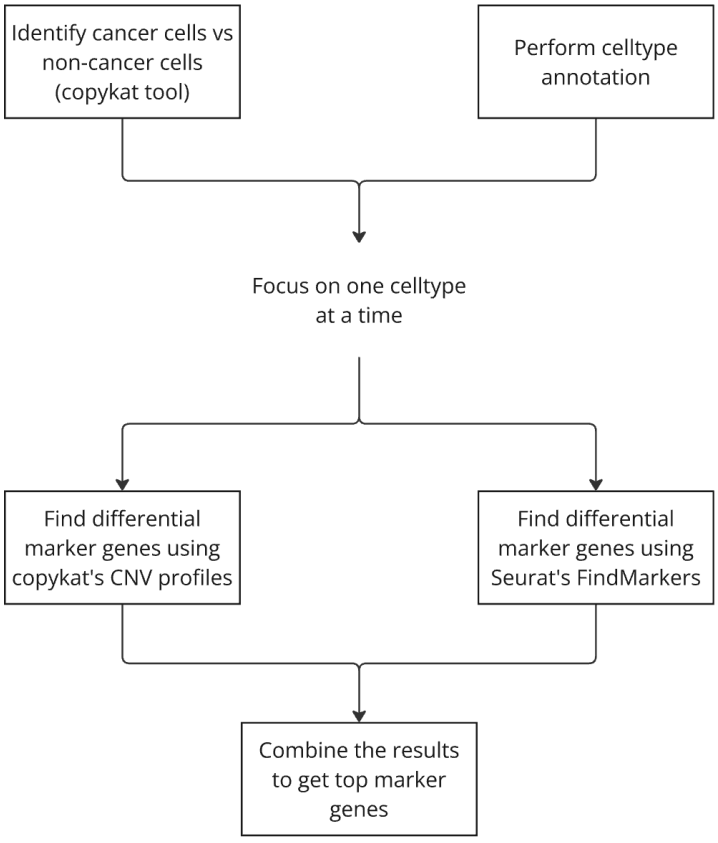

Fig.1: The pipeline of our analysis. It consisted of employing two methods to find marker genes for glioblastoma cancer cells and finding the consensus marker genes identified by both methods.

To run copykat, we just had to specify the gene-expression matrix of our data. We used the default settings to run the tool. Our output looked like as shown in Fig. 2. It shows the estimated ploidy of 200k base pair regions of the genome. When the ploidy is significantly higher than 2, that genomic area is marked red, if it is significantly lower than 2, then it is labelled blue. Once the aneuploid nature (significantly different from 2) of the genomic regions is determined, the

cells are clustered according to these labels, and eventually, we get labels for the cells as aneuploid (likely cancerous) and diploid. All this is done by the copykat tool behind the scenes and we get the labels for each cell and also the CNV profile of each genomic region. In each genomic region labelled by copykat, there lie a few genes. Hence, associated to each gene is a CNV number.

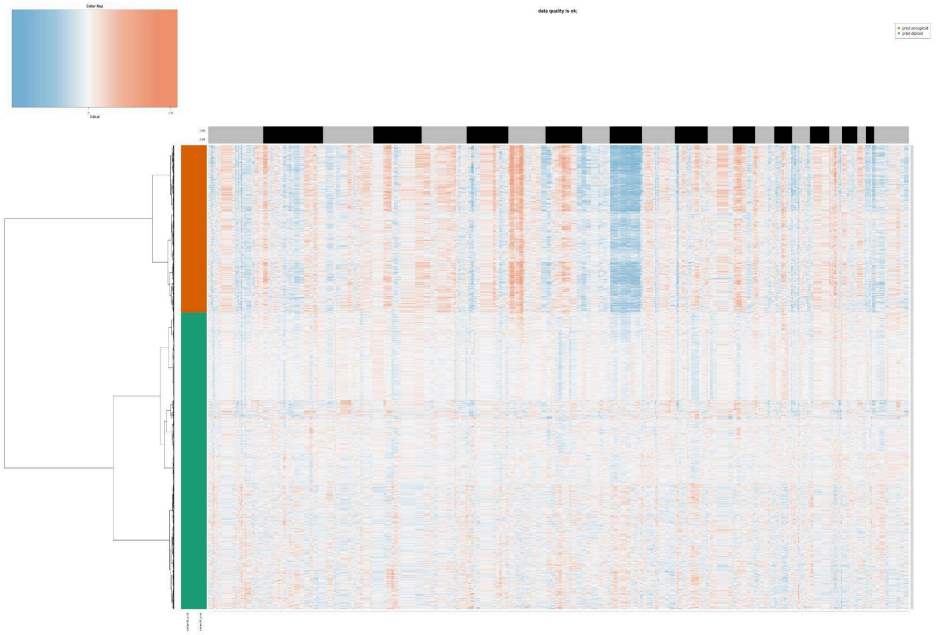

Fig. 2: The output of copykat. Each row represents a cell and each column a genomic region of size 200k. The colors correspond to the ploidy of any genomic region and the colors (red/green) on the left denote the aneuploid/diploid nature of the cell.

Parallel to this, we follow a standard single cell data celltype annotation pipeline (Fig. 3, a) to annotate our cells into common celltypes found in the human brain (Fig. 3, b). This pipeline followed the ordering of log-normalizing the data, finding the top 2000 features, scaling and running PCA for clustering, using a UMAP plot to show clusters, and examining known marker genes for manual assignment of clusters. All of these procedures were available with Seurat, and the marker genes we used for the manual celltype annotation (once the clustering was done) were obtained from the PanglaoDB resource.

Once we the cells were annotated with celltypes, we focus on the cells from one celltype at a time, and identify the marker genes for the cancer and normal cells for that celltype using the FindMarkers function from Seurat, which uses the Wilcoxon Rank Sum test to compute the likelihood of any gene coming from different distributions for the cancer and the normal cells. We took the top 100 of the most significant genes for each celltype from this. Then, we computed the top 100 marker genes for each celltype using copykat's results. To do this, we compute the ratio of average CNV scores corresponding to each gene for the cancer and normal cells separately and take their ratio. Using this ratio, we find out the genes with differential CNV numbers in cancer and normal cells. We take top 100 of these genes.

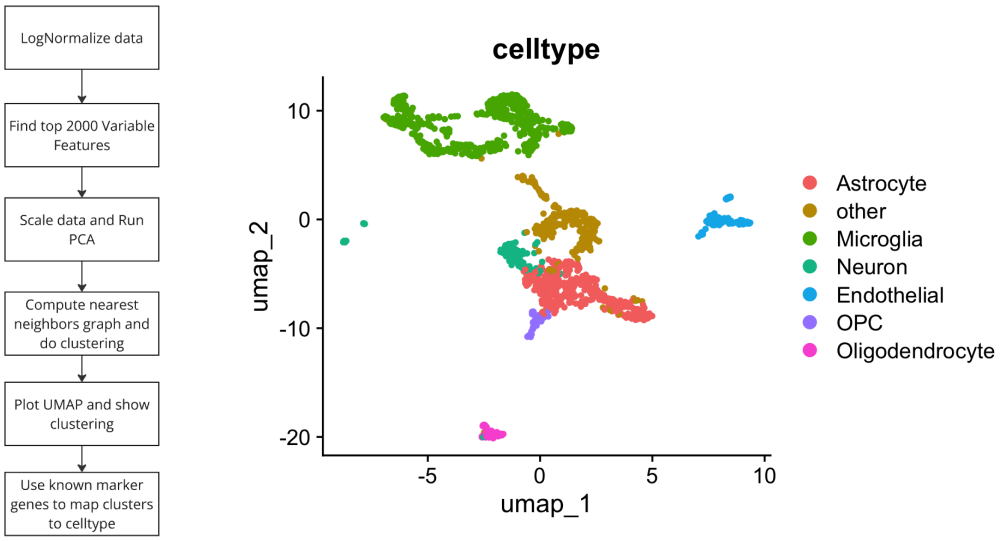

Fig. 3: Celltype annotation pipeline and results. a) The single cell celltype annotation pipeline we used. b) UMAP showing the annotations of the different cells in our dataset. The marker genes were obtained from the PanglaoDB resource.

Finally, we find the common genes in the top 100 genes' list obtained through both methods for each celltype and perform investigation on the genes, for possible link to glioblastoma.

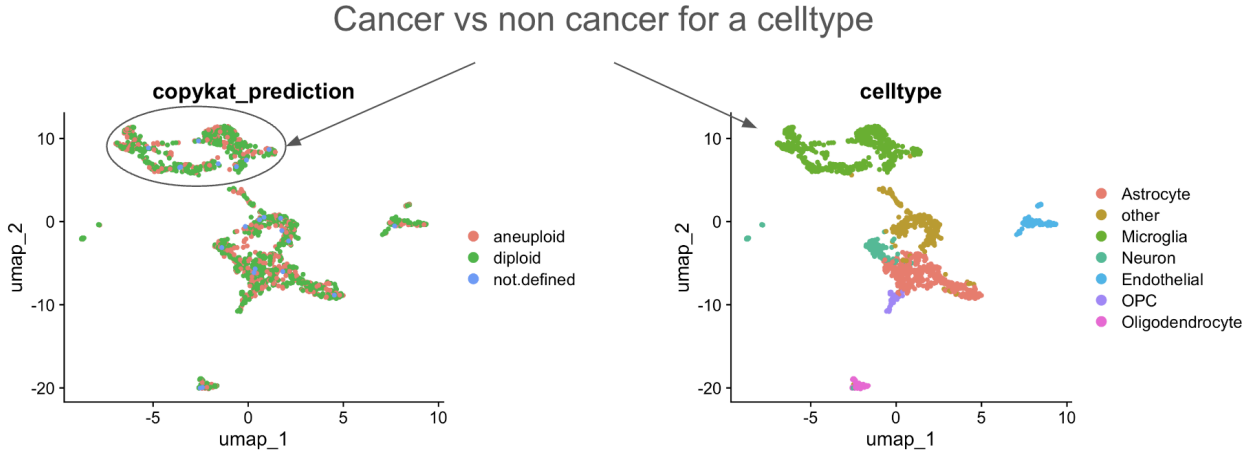

Fig. 4: Example of finding differential marker genes for cancer and normal cells for a particular celltype (Microglia).

Results

From both of our methods (Seurat and Copykat) we found common genes in the top 100 for each cell type to be the following: astrocytes, microglia, and oligodendrocytes had none, but endothelial cells had NUDT1, neurons had CACNA1A, and OPCs had MAP1B. In this section,

we present some of the literature relating these genes to glioblastoma and/or cancer cell development in general.

In endothelial cells, the gene we identified to be most common was NUDT1, commonly referred to as MTH1. This gene is largely linked to colorectal and kidney cancers, but some evidence suggests it also plays an important role in glioblastoma progression. NUDT1 acts on 8-oxo-dGTPases, which play a role in genomic stability (Bialkowski and Kasprzak, 2020); in simpler terms, this means that NUDT1 mutations have been shown to be associated with many cancers. Experimental evidence supports this; Bialkowski et al. (2009) found higher expression of NUDT1 in mouse brains when exposed to gamma radiation. Furthermore, NUDT1 has been a promising target for anticancer drug development (Taiyab et al., 2024), including with regards to glioblastoma (Pudelko et al., 2020; Tu et al., 2016). NUDT1 has been shown to also be a risk indicator for glioblastoma (Wu et al., 2023). Therefore, our identification of NUDT1 as a common glioblastoma gene is highly consistent with much research on glioblastoma genetics.

In neurons, the gene we identified to be most common was CACNA1A, a protein found in ion channels; specifically, it is found in voltage-gated calcium channels (VGCCs). While commonly associated with epilepsy, one study found this gene to be upregulated with respect to glioblastoma and cancer cell development (Hsu et al., 2021). Furthermore, some evidence indicates that its involvement in ion VGCCs make it a promising candidate for cancer therapies, but evidence is conflicting; Wang et al. (2015) agree that CACNA1A is more highly expressed in glioblastomas, but Phan et al. (2017) provide results that show a decrease in CACNA1A expression in glioblastomas. While a promising candidate for cancer risk assessments and drug therapies, more research is necessary to discern the true nature of CACNA1A.

In OPCs, the gene we found to be most common was MAP1B. Commonly associated with nervous system development, MAP1B is highly expressed in developing neurons and glial cells, so finding it in OPCs is reasonable (Villarroel-Campos and Gonzalez-Billault, 2014). MAP1B has been shown to play a role in not only glioblastoma development but also cancer drug resistance (Laks et al., 2018) despite various efforts to target it for drug development (including Chen and Chou, 2015). While the body of research on the exact links between MAP1B and glioblastoma is relatively small, our results indicating MAP1B as a potential risk assessment and/or drug target could indicate that more research is required to understand this link.

To place our gene hits in the context of known single-cell expression, we looked at these genes on the Allen Brain Map (Figure 5) and on CATLAS (Figure 6, 7) to examine normal expression of these genes and compare it with our findings. NUDT1 is shown to have low or no expression throughout all cells; this is consistent with the literature showing it to be a risk indicator for glioblastoma. CACNA1A was found to have moderate expression in all neural cells, which is consistent with its involvement in VGCCs; however, due to it already being present in neurons in some amount, it is unlikely that CACNA1A could be a risk factor or marker for glioblastoma. MAP1B was found to be present in all cells, but it was much more present in neurons compared to non-neuronal cells. Critically, however, it is present at low amounts in many OPCs, so our results indicating it is a common gene in OPCs for glioblastoma patients could mean that MAP1B is a promising risk indicator for glioblastoma, as well as a potential therapeutic target.

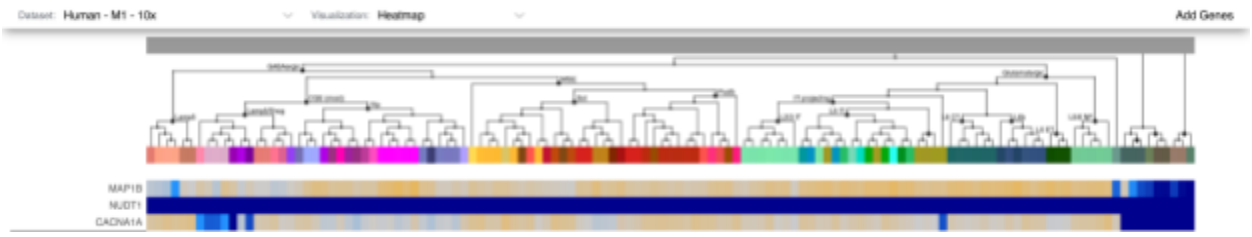

Fig. 5: Allen Brain Map heatmap visualization of the genes shown to be common in glioblastoma patients for the normal human brain. Critically, NUDT1 is not present in any cells and MAP1B is present at low levels in oligodendrocytes, contrary to what we found for glioblastoma patients.

CATLAS did not show any data for endothelial cells, so we were unable to look at NUDT1 with respect to other cells in CATLAS. For neuronal cells, we found that CACNA1A expression and openness of cis-elements varied between neurons, even those within the same family (Figure 6), consistent with CACNA1A being necessary for many neurons and also supporting that CACNA1A may not be a promising risk indicator for glioblastoma. MAP1B showed interesting results for OPCs, showing that the cis-elements of OPCs were highly active near MAP1B, especially in comparison to many other cells (Figure 7). This result is contrary to what we might expect, given that more MAP1B could indicate glioblastoma; however, due to its necessity in development, it makes sense that there would be high levels of activity for regulation.

Fig. 6: CATLAS visualization of CACNA1A in GABAergic (left) and Glutamatergic (right) neurons.

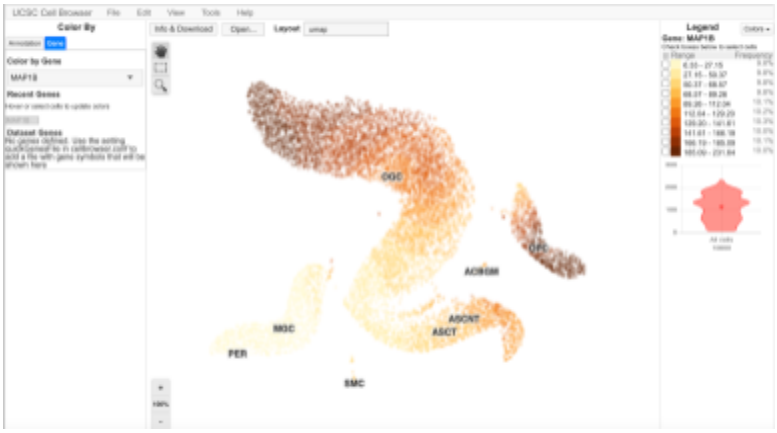

Fig. 7: CATLAS visualization of MAP1B in non-neuronal cells. Critically, OPCs are consistently labeled with high scores, possibly indicating high levels of activity.

Discussion and Conclusions

We believed that in cancer, the marker genes are likely to be specific to a celltype and not general to all celltypes. Hence, we used this opportunity to try to do that. Since copykat already gave us the CNV profiles of different genes, we thought of making use of that information too, to identify marker genes. Using our unique approach of focusing on one celltype at a time and finding marker genes for cancer cells in that celltype using two different methods (one utilizing pure gene expressions and the other utlizing estimated copy numbers for a genomic region in which a gene lies), we were able to find biologically relevant genes, which are known to be linked to glioblastoma. This shows that there is relevance in finding marker genes specific to a celltype in cancer.

One possible challenge with our approach is the possible noise in estimates of copykat, and lack of validation of copykat's results. There is some room to explore other cancer-cell identification procedures such as inferCNV, and using them in conjunction with copykat can allow us to perform some validation in the results too. Also, there is possibility of performing the analyses in a larger dataset for better quality and robustness of the results. Our dataset size being small (only ~1600 cells) did not help in deriving robust results.

Regardless, our results indicated some potential promise in the discovery of new risk indicators for glioblastoma as well as confirming the usage of current ones, those being MAP1B and NUDT1, respectively. As such, future work would likely include the experimental validation of these risk indicators, whether from surveying additional cancer patients, performing experiments on mice, or knocking down these genes *in vitro*. Given these results, we also believe that more work is necessary for studying the efficacy of therapeutically targeting MAP1B and NUDT1 in glioblastoma.

References

Alifieris, C., & Trafalis, D. T. (2015). Glioblastoma multiforme: Pathogenesis and treatment. *Pharmacology & therapeutics, 152*, 63-82.

Gallego, O. (2015). Nonsurgical treatment of recurrent glioblastoma. *Current oncology, 22*(4), 273-281.

Ohgaki, H., & Kleihues, P. (2013). The definition of primary and secondary glioblastoma. *Clinical cancer research, 19(*4), 764-772.

Parsons, D. W., Jones, S., Zhang, X., Lin, J. C. H., Leary, R. J., Angenendt, P., ... & Kinzler, K. W. (2008). An integrated genomic analysis of human glioblastoma multiforme. *Science, 321*(5897), 1807-1812.

Bialkowski K, Szpila A, Kasprzak KS. Up-regulation of 8-oxo-dGTPase activity of MTH1 protein in the brain, testes and kidneys of mice exposed to (137)Cs gamma radiation. Radiat Res. 2009 Aug;172(2):187-97. doi: 10.1667/RR1636.1. PMID: 19630523.

Bialkowski K, Kasprzak KS. A profile of 8-oxo-dGTPase activities in the NCI-60 human cancer panel: Meta-analytic insight into the regulation and role of MTH1 (NUDT1) gene expression in carcinogenesis. Free Radic Biol Med. 2020 Feb 20;148:1-21. doi: 10.1016/j.freeradbiomed.2019.12.036. Epub 2019 Dec 25. PMID: 31883466.

Chen JH, Chou CC. Acrylamide inhibits cellular differentiation of human neuroblastoma and glioblastoma cells. Food Chem Toxicol. 2015 Aug;82:27-35. doi: 10.1016/j.fct.2015.04.030. Epub 2015 May 7. PMID: 25959841.

Hsu JB, Lee TY, Cheng SJ, Lee GA, Chen YC, Le NQK, Huang SW, Kuo DP, Li YT, Chang TH, Chen CY. Identification of Differentially Expressed Genes in Different Glioblastoma Regions and Their Association with Cancer Stem Cell Development and Temozolomide Response. J Pers Med. 2021 Oct 20;11(11):1047. doi: 10.3390/jpm11111047. PMID: 34834399; PMCID: PMC8625522.

Laks DR, Oses-Prieto JA, Alvarado AG, Nakashima J, Chand S, Azzam DB, Gholkar AA, Sperry J, Ludwig K, Condro MC, Nazarian S, Cardenas A, Shih MYS, Damoiseaux R, France B, Orozco N, Visnyei K, Crisman TJ, Gao F, Torres JZ, Coppola G, Burlingame AL, Kornblum HI. A molecular cascade modulates MAP1B and confers resistance to mTOR inhibition in human glioblastoma. Neuro Oncol. 2018 May 18;20(6):764-775. doi: 10.1093/neuonc/nox215. PMID: 29136244; PMCID: PMC5961175.

Phan NN, Wang CY, Chen CF, Sun Z, Lai MD, Lin YC. Voltage-gated calcium channels: Novel targets for cancer therapy. Oncol Lett. 2017 Aug;14(2):2059-2074. doi: 10.3892/ol.2017.6457. Epub 2017 Jun 22. PMID: 28781648; PMCID: PMC5530219.

Pudelko L, Rouhi P, Sanjiv K, Gad H, Kalderén C, Höglund A, Squatrito M, Schuhmacher AJ, Edwards S, Hägerstrand D, Berglund UW, Helleday T, Bräutigam L. Glioblastoma and glioblastoma stem cells are dependent on functional MTH1. Oncotarget. 2017 Jul 20;8(49):84671-84684. doi: 10.18632/oncotarget.19404. Erratum in: Oncotarget. 2020 Feb 25;11(8):827. doi: 10.18632/oncotarget.27484. PMID: 29156675; PMCID: PMC5689565.

Taiyab A, Ashraf A, Sulaimani MN, Rathi A, Shamsi A, Hassan MI. Role of MTH1 in oxidative stress and therapeutic targeting of cancer. Redox Biol. 2024 Nov;77:103394. doi: 10.1016/j.redox.2024.103394. Epub 2024 Oct 11. PMID: 39418911; PMCID: PMC11532495.

Tu Y, Wang Z, Wang X, Yang H, Zhang P, Johnson M, Liu N, Liu H, Jin W, Zhang Y, Cui D. Birth of MTH1 as a therapeutic target for glioblastoma: MTH1 is indispensable for gliomatumorigenesis. Am J Transl Res. 2016 Jun 15;8(6):2803-11. PMID: 27398163; PMCID: PMC4931174.

Villarroel-Campos D, Gonzalez-Billault C. The MAP1B case: an old MAP that is new again. Dev Neurobiol. 2014 Oct;74(10):953-71. doi: 10.1002/dneu.22178. Epub 2014 Apr 17. PMID: 24700609.

Wang CY, Lai MD, Phan NN, Sun Z, Lin YC. Meta-Analysis of Public Microarray Datasets Reveals Voltage-Gated Calcium Gene Signatures in Clinical Cancer Patients. PLoS One. 2015 Jul 6;10(7):e0125766. doi: 10.1371/journal.pone.0125766. PMID: 26147197; PMCID: PMC4493072.

Wu Y, Liu L, Huang D, Li Z, Xu R, Cheng M, Chen L, Wang Q, You C. Uncover DNA damage and repair-related gene signature and risk score model for glioma. Ann Med. 2023 Dec;55(1):2200033. doi: 10.1080/07853890.2023.2200033. PMID: 37086071; PMCID: PMC10124985.

Oscar Franzén, Li-Ming Gan, Johan L M Björkegren, PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data, Database, Volume 2019, 2019, baz046, doi:10.1093/database/baz046