



**HAL**  
open science

# Uncovering the Genetic Basis of Glioblastoma Heterogeneity through Multimodal Analysis of Whole Slide Images and RNA Sequencing Data

Ahmad Berjaoui, Louis Roussel, Eduardo Hugo Sanchez, Elizabeth Cohen-Jonathan Moyal

► **To cite this version:**

Ahmad Berjaoui, Louis Roussel, Eduardo Hugo Sanchez, Elizabeth Cohen-Jonathan Moyal. Uncovering the Genetic Basis of Glioblastoma Heterogeneity through Multimodal Analysis of Whole Slide Images and RNA Sequencing Data. 2024. hal-04746042v2

**HAL Id: hal-04746042**

**<https://hal.science/hal-04746042v2>**

Preprint submitted on 9 Jan 2025

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

# Uncovering the Genetic Basis of Glioblastoma Heterogeneity through Multimodal Analysis of Whole Slide Images and RNA Sequencing Data

Ahmad Berjaoui<sup>a,c,1,\*</sup>, Eduardo Hugo Sanchez<sup>a,1</sup>, Louis Roussel<sup>a</sup>, Elizabeth Cohen-Jonathan Moyal<sup>b,c</sup>

<sup>a</sup>*IRT Saint Exupéry, Toulouse, France*

<sup>b</sup>*Oncopole Claudius Regaud, Toulouse, France*

<sup>c</sup>*INSERM UMR 1037 Cancer Research Center of Toulouse (CRCT), University Paul Sabatier Toulouse III, Toulouse, France*

---

## Abstract

Glioblastoma is a highly aggressive form of brain cancer characterized by rapid progression and poor prognosis. Despite advances in treatment, the underlying genetic mechanisms driving this aggressiveness remain poorly understood. In this study, we employed multimodal deep learning approaches to investigate glioblastoma heterogeneity using joint image/RNA-seq analysis. Our results reveal novel genes associated with glioblastoma. By leveraging a combination of whole-slide images and RNA-seq, as well as introducing novel methods to encode RNA-seq data, we identified specific genetic profiles that may explain different patterns of glioblastoma progression. These findings provide new insights into the genetic mechanisms underlying glioblastoma heterogeneity and highlight potential targets for therapeutic intervention.

*Keywords:* Glioblastoma, Deep learning, RNA-seq, Personalized medicine

---

## 1. Introduction

Glioblastoma (GB) is the most aggressive primary brain tumor and is not curable [1] despite of treatment associating surgery when possible followed

---

\*Corresponding author

<sup>1</sup>Equal contribution

by radio-chemotherapy [2] and more recently Tumor treating Fields [3] leading to a median overall survival (OS) of 20.9 months and a progression free survival (PFS) around 7 months. Despite being a minor population of cancer cells, the cancer stem cells that are identified in glioblastoma (GSCs) are thought to be the major driving force behind glioblastoma biological heterogeneity and are likely to explain the high rates of glioblastoma recurrence. In the STEMRI clinical trial aiming to study GB heterogeneity and the enrichment of GSC in certain areas defined by multimodal MRI (NCT01872221) [4] different GSC sub-populations extracted from tumor samples obtained by multimodal MRI guided surgery were xenografted into mice brain to study their invasion patterns as well as their aggressiveness. RNA-seq on each tumor bulk samples was also performed. The observed differences in mice survival according to the GSC implanted confirm the heterogeneous nature of tumor cells lineage.

In this study, we set out to determine potential genetic markers associated with glioblastoma aggressiveness using multimodal deep learning. Our results reveal genetic targets already identified in medical literature but also highlight new potential targets. We leverage the extensive recent developments in multimodal data analysis, namely image and text, and adapt these techniques to whole slide images (WSI) and RNA-seq data. Training uses public data from The Cancer Genome Atlas (TCGA) [5] but also WSI and RNA-seq data from the STEMRI trial. Furthermore, we introduce a new RNA-seq encoding technique where genes are grouped based on biological pathways prior to encoding and show better performance in comparison to mere position based grouping.

Overall, our results can be used to test new GB treatment strategies. Our main contributions can be summarized as follows:

- i Identify genetic profiles leading to unique GB patterns.
- ii Novel method to encode RNA-seq data for use in deep learning models.
- iii Novel method to combine WSI and RNA-seq data for use in deep learning models.

This paper is structured as follows. Section 2 details the developed algorithms, the datasets and training strategies. Section 3 showcases our experimental results followed by our main findings.

## 2. Material and methods

### 2.1. Datasets

Two datasets are used to train our models. The TCGA public dataset is used to pre-train RNA-seq encoders. This dataset contains around 10k RNA-Seq gene expression samples from multiple sites (breast, brain, prostate, bladder, etc.).

51 RNA-seq samples from the STEMRI trial complete this data. These correspond to 16 patients and bulk tumor RNA from metabolically heterogeneous regions identified by spectral MRI, as explained in [4]. This corresponds to 51 different tumor cell lineages. To insure data homogeneity, only genes common to both STEMRI and TCGA were kept. Genes expressions with relatively low variance were also removed and all data was then normalized. Tumor cells from these 51 different lineages were xenografted onto mice brains. Cells from the same lineage were used onto more than one mouse whenever possible, as culture was not always successful. This led to a total of 116 mice. Whole slide images of mice brain slices were used at x10 magnification factor (fig.2) after each mouse’s death. These images are in average composed of 16k x 21k RGB pixels and human tumor cells are highlighted using specific coloration.

Observation data contains patient OS and PFS and mouse survival time in days.

### 2.2. RNA-seq encoding

RNA-seq data is a vector representing protein expressions for roughly 19K protein coding genes. Moreover, attention-based encoders are currently state-of-the-art for vision [6] and language [7]. Combining these two observations, we leveraged the Protein-to-Protein interaction (PPI) graph [8] to regroup protein scores before computing attention scores. The PPI graph is a directed graph and [9] propose a clustering algorithm that considers the directed nature of PPI.

The reader can refer to [9] for more detail about the directed Louvain algorithm but in brief, the aim is to maximize *modularity* [10]  $Q_d$  which in the case of a directed graph, can be defined by:

$$Q_d = \frac{1}{m} \sum_{i,j} [A_{ij} - \frac{d_i^{in} d_j^{out}}{m}] \delta(c_i, c_j) \quad (1)$$

$m$  is the total number of edges.  $A_{ij}$  indicates the presence of an edge between nodes  $i$  and  $j$ .  $d_i^{in}$  (*resp.*  $d_i^{out}$ ) is the in-degree (*resp.* out-degree) of node  $i$ . The degree is the number of incoming (*resp.* outgoing) edges.  $\delta(c_i, c_j)$  indicates whether nodes  $i$  and  $j$  are in the same cluster. Connected edges with low degrees contribute significantly to modularity when they are in the same cluster as this is much less likely than a connection between edges of high degrees, regardless of cluster assignment.

Applying this algorithm to PPI leads to 11 clusters. Clusters with significantly less nodes were then grouped together, leading to 7 final clusters with an average of 2905 genes per cluster (ranging between 1514 and 3182).

Genes in each cluster constitute a sub-vector that is projected to a common size. The result is 7 sub-vectors of the same size, each as a separate token as input to a masked-autoencoder [11]. During training, one token is randomly masked and the decoder uses the *cls* token and the remaining tokens to reconstruct it, as illustrated in fig.1. This results in a first loss term:

$$\mathcal{L}_{RNA}^r = \frac{1}{N} \sum_{i=1}^N \|\mathbf{x}_i - \hat{\mathbf{x}}_i\|^2 \quad (2)$$

where  $N$  is the batch size,  $\mathbf{x}_i$  represents RNA-seq vector  $i$  and  $\hat{\mathbf{x}}_i$  its reconstruction. In our experiments, we compared this encoding strategy to a similar strategy but where genes were clustered according to their chromosome. The second loss term  $\mathcal{L}_{RNA}^c$  is obtained using pairwise contrastive learning [12]: a positive pair is obtained from two RNA sub-vectors corresponding to the same tumor cell lineage, and a negative pair is obtained using two sub-vectors corresponding to two different tumor cell lineages. This forces the encoder to learn lineage specific features. Given a pair of representations  $\mathbf{z}_i$  and  $\mathbf{z}'_i$  with batch size  $N$ , temperature  $\tau$ , the loss term  $l_i$  is therefore:

$$l_i = -y(i, i') \log \frac{\exp(\mathbf{z}_i^T \mathbf{z}'_i / \tau)}{\sum_{j=1}^N \exp(\mathbf{z}_i^T \mathbf{z}_j / \tau)} \quad (3)$$

$y(i, i') = 1$  for a matching pair and 0 for negative pairs (i.e. different cell lineages). Hence,

$$\mathcal{L}_{RNA}^c = \sum_{i=1}^N l_i \quad (4)$$

The final loss for RNA-seq encoding is:

$$\mathcal{L}_{RNA} = \mathcal{L}_{RNA}^r + \mathcal{L}_{RNA}^c \quad (5)$$

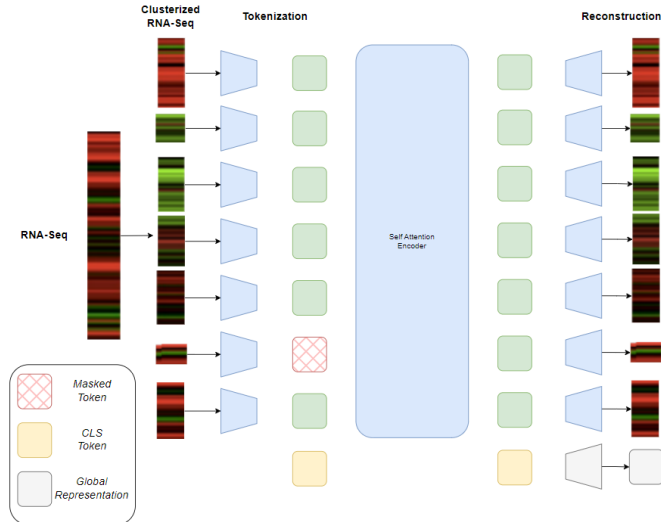


Figure 1: RNA-seq encoding. The original vector is reorganized according to directed graph clustering from the PPI knowledge graph. Sub-vectors are projected to a common embedding size. A token is randomly masked. The decoder reconstructs the RNA-seq vector using the remaining tokens and the *cls* token.

### 2.3. WSI encoding

All images are split into sequential, non-overlapping patches of 256x256 RGB pixels, which results in roughly 5K patches per image and a total of 100K patches for the whole dataset. All patches are converted to hue, saturation, luminance, space (HSL). HSL conversion makes it easier to distinguish non-tumor and tumor cells (colored in brown). Patches with a majority of tumor cells can hence be easily extracted by computing an overall pixel score, by counting pixels that respect a given hue interval, regardless of lighting conditions. Patches can then be sorted based on that score, which correlates with their tumor cells content. Only those with a brown colored pixel ratio exceeding 20% are kept for training.

Training a masked autoencoder for tumor patches would require a very high computational cost due to the size of the WSI dataset. We therefore rely solely on contrastive learning as our experiments have proved that it was enough to achieve very good performance. A 256x256 patch is split into smaller 16x16 sub-patches and we use a ViT [6] transformer encoder to obtain a global representation of the patch, using the *cls* token. This representation serves then as an anchor in a triplet loss[13]: the anchor ( $\mathbf{z}_i$ ) is matched with

the representation of another patch of the same tumor cell lineage ( $\mathbf{z}_+$ ) and it is contrasted with the representation of a patch from another tumor cell lineage ( $\mathbf{z}_-$ ).

$$\mathcal{L}_{WSI}^c = \frac{1}{N} \sum_{i=1}^N \max(\|\mathbf{z}_i - \mathbf{z}_+\|^2 - \|\mathbf{z}_i - \mathbf{z}_-\|^2 + d, 0) \quad (6)$$

where  $d$  is the margin and  $N$  is the batch size.

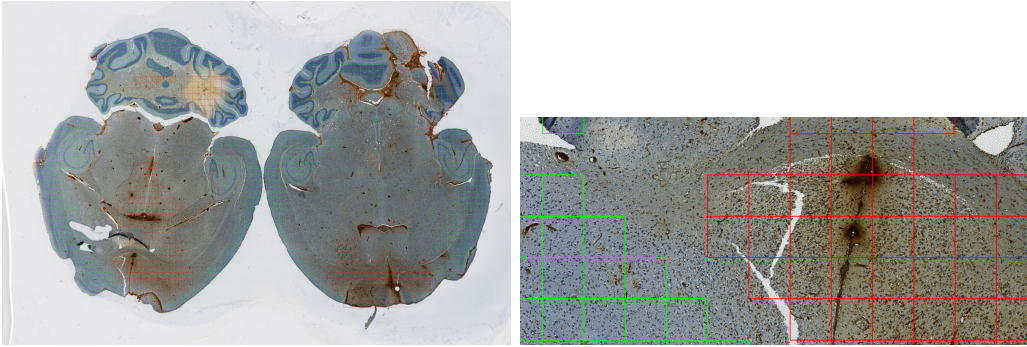


Figure 2: (Left) Example of a WSI of a mouse brain slice. (Right) Zoom on the upper-right part of the left brain slice. Red square patches have a majority of tumor cells whereas green square patches have a majority of non-tumor cells.

#### 2.4. Multimodal training

Both the RNA and the WSI encoders are pre-trained using eq.5 and eq.6. In order to consider both modalities, we draw inspiration from ALBEF[14]. A multimodal contrastive loss  $\mathcal{L}_{MM}^c$  is used to align the modalities' representations before a cross-attention encoder. For a given RNA vector and a 256x256 WSI patch, let  $\mathbf{u}_k$  and  $\mathbf{u}'_k$  be the *cls* representations at the output of their respective encoders, after linear projection to a common embedding size. Order is irrelevant as the following equations treat  $\mathbf{u}_k$  and  $\mathbf{u}'_k$  symmetrically. We use a similar pairwise loss as in eq.4, considering a pair of RNA/WSI representations  $\mathbf{u}_k$  and  $\mathbf{u}'_k$ , a temperature  $\tau_c$  and batch size  $N$ :

$$l_k = -y(k, k') \log \frac{\exp(\mathbf{u}_k^T \mathbf{u}'_k / \tau_c)}{\sum_{j=1}^N \exp(\mathbf{u}_k^T \mathbf{u}_j / \tau_c)} \quad (7)$$

$y(k, k') = 1$  for a matching pair and 0 for negative pairs.

$$\mathcal{L}_{MM}^c = \sum_{k=1}^N l_k \quad (8)$$

As is done in [14], a cross-attention model is used to obtain a joint RNA/WSI

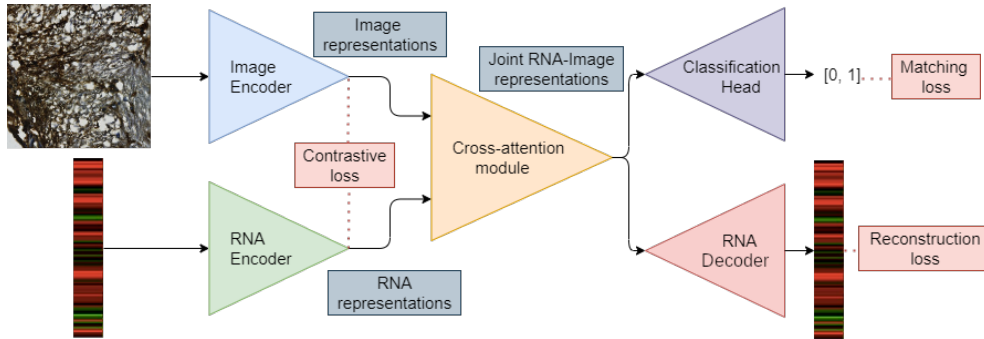


Figure 3: Multimodal model. A pairwise contrastive loss aligns RNA and WSI representations. A cross-attention module is used to obtain a joint representation. The latter is used for a classification head that matches data from the same tumor cells lineage and a RNA decoder reconstructs the original RNA vector.

representation from the unimodal representations. The cross-attention model is trained along with a classification head and a RNA decoder head. The classification head is trained using a weakly supervised loss  $\mathcal{L}_{MM}^m$  where a label of 1 indicates samples from the same tumor cells lineage and 0 for mismatching samples. The RNA decoder is trained using a reconstruction loss  $\mathcal{L}_{MM}^r$  similar to Eq.2.

### 2.5. Evaluation

Training the multimodal model leads to aligned WSI and RNA-seq representations. Given a WSI patch not used during training, it becomes possible to retrieve the closest RNA-seq vector amongst all the 51 RNA-seq vectors by measuring its cosine similarity, as shown in fig.4

The main evaluation criteria is therefore matching accuracy, computed over a subset of WSI patches (from a total of 100K patches) not used during training but that are drawn evenly amongst corresponding cell lineages and patients. Let  $\mathcal{I} = \{1..51\}$  be the set of all the 51 RNA-seq samples,  $\mathbf{z}_{WSI}$  be the output of the image encoder for a given WSI patch, and  $\mathbf{u}_i$  be the output



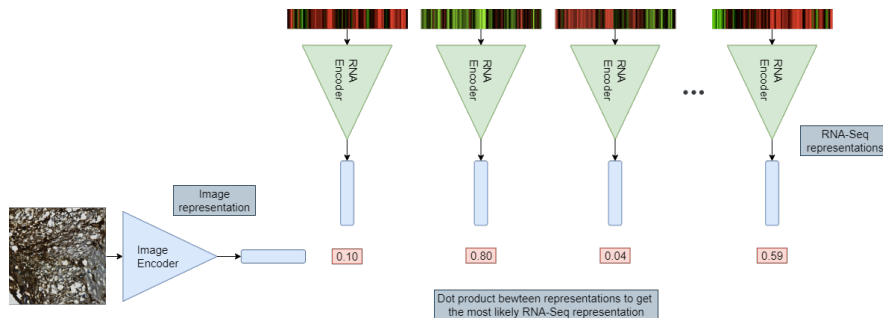


Figure 4: RNA-seq retrieval using a WSI patch.

of the RNA encoder for sample  $i$  ( $i \in \mathcal{I}$ ). The predicted sample  $\hat{p}$  is given by eq.9

$$\hat{p} = \arg \max_{i \in \mathcal{I}} \frac{\mathbf{z}_{WSI}^T \mathbf{u}_i}{\|\mathbf{z}_{WSI}\| \cdot \|\mathbf{u}_i\|} \quad (9)$$

### 3. Experiments

*RNA-encoding.* After pre-training the RNA-seq encoder depicted in §2.2 using TCGA data, we plotted the t-SNE two dimensional projections for both TCGA data and STEMRI data (fig.5). Nearly all STEMRI data points lie very close to TCGA data points corresponding to brain tumors.

*RNA retrieval.* Table.1 lists RNA-seq retrieval accuracy considering tumor cell lineage and patient of origin. The best results are obtained using pathway based clustering as explained in §2.2, and using only a matching loss (without a RNA decoder and a reconstruction loss). Results show a near perfect accuracy for both patient and cell lineage matching tasks. This indicates that our model has learned unique genetic features that can be matched against unique cellular microscopic patterns.

Model	Accuracy (patient)	Accuracy (lineage)
Random initialization	0.050	0.006
$\mathcal{L}_{MM}^r + \mathcal{L}_{MM}^c + \mathcal{L}_{MM}^m$ (frozen RNA-Image encoders)	0.29	0.199
$\mathcal{L}_{MM}^r + \mathcal{L}_{MM}^c + \mathcal{L}_{MM}^m$ (all modules are trainable)	0.946	0.941
$\mathcal{L}_{MM}^c + \mathcal{L}_{MM}^m$ (all modules are trainable/chromosome gene groups)	0.890	0.913
$\mathcal{L}_{MM}^c + \mathcal{L}_{MM}^m$ (all modules are trainable/pathways gene groups)	<b>0.970</b>	<b>0.968</b>

Table 1: Lineage and patient retrieval results. The table lists WSI patch/RNA-seq matching accuracy by considering cell lineage or patient of origin, in several configurations.

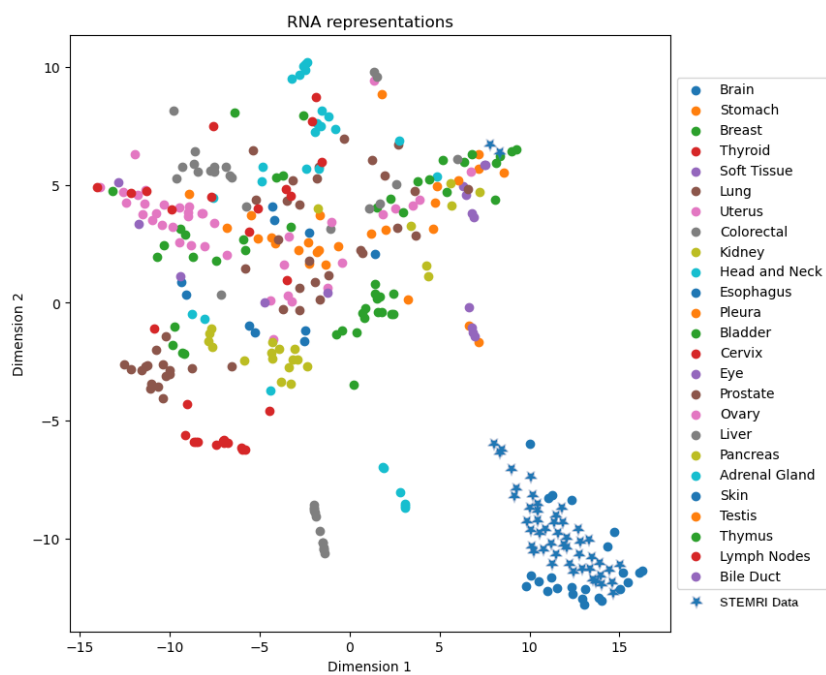


Figure 5: 2D t-SNE projections of RNA-seq representations of primary tumor locations, combining public TCGA data (circles) and STEMRI data (stars). The figure clearly showcases a strong similarity between public brain RNA-seq data and STEMRI data.

*Genetic analysis.* Grad-CAM [15] is a common technique to analyse deep neural networks’ response. We used grad-CAM to determine genetic positions that have the greatest contribution in the WSI/RNA-seq matching process. For each WSI in the validation set, we sort gene expressions according to their importance as computed using grad-CAM. Keeping the 15 most occurring gene expressions leads to the results in table.2. The highlighted

	Gene	Occurrence	Symbol
0	ENSG00000198964.14	2344	SGMS1
1	ENSG00000120008.16	2151	WDR11
2	ENSG00000111670.16	1759	GNPTAB
3	ENSG00000078098.14	1669	FAP
4	ENSG00000143147.14	1639	GPR161
5	ENSG00000070061.16	1463	ELP1
6	ENSG00000139624.14	1350	CERS5
7	ENSG00000077254.14	1347	USP33
8	ENSG00000138614.15	1249	INTS14
9	ENSG00000090054.15	1202	SPTLC1
10	ENSG00000183036.11	1187	PCP4
11	ENSG00000130717.13	1177	UCK1
12	ENSG00000184164.15	1157	CRELD2
13	ENSG00000125633.11	1138	CCDC93
14	ENSG00000128829.12	1112	EIF2AK4

Table 2: Key genes in the WSI/RNA-seq matching process. Highlighted genes have been found in glioblastoma literature.

genes have already been identified in literature as playing an important role in the occurrence or development of glioblastoma [16, 17, 18, 19, 20] (most of them playing a role in metabolism, micro-environment and invasion), other brain cancer types [21] and cancer in general [22, 23, 24, 25, 26, 27, 28].

#### 4. Conclusion

In conclusion, our study demonstrates the effectiveness of multimodal deep learning approaches in identifying genetic profiles that explain different glioblastoma patterns. By leveraging joint image/RNA-seq analysis and introducing novel methods to encode RNA-seq data, we have shed new light on the heterogeneous nature of this aggressive brain tumor. Our findings

not only confirm existing medical literature but also highlight new potential targets for therapeutic intervention. These results have significant implications for the development of personalized medicine strategies for glioblastoma patients and underscore the importance of continued research into the application of AI algorithms in cancer biology.

### **Acknowledgement**

The results shown here are in part based upon data generated by the TCGA Research Network: <https://www.cancer.gov/tcga>. The authors would like to thank Caroline Delmas<sup>2</sup> for WSI images acquisition.

### **Declaration of competing interest**

Elizabeth Moyal-Cohen Jonathan served as board expert for Novocure and received research grants from Astra Zeneca, Incyte, Bayer and Novocure. Other authors declare that they have no known competing financial interests or personal relationships that could have interfered with the study being conducted.

### **References**

- [1] O. G. Taylor, J. S. Brzozowski, K. A. Skelding, Glioblastoma Multiforme: An Overview of Emerging Therapeutic Targets, *Frontiers in Oncology* 9 (2019) 963. doi:10.3389/fonc.2019.00963.
- [2] R. Stupp, M. Weller, K. Belanger, U. Bogdahn, S. K. Ludwin, D. Lacombe, R. O. Mirimanoff, Radiotherapy plus Concomitant and Adjuvant Temozolomide for Glioblastoma, *The New England Journal of Medicine* (2005) 10.
- [3] O. Rominiyi, A. Vanderlinden, S. J. Clenton, C. Bridgewater, Y. Al-Tamimi, S. J. Collis, Tumour treating fields therapy for glioblastoma: current advances and future directions, *British journal of cancer* 124 (4) (2021) 697–709.

---

<sup>2</sup>INSERM UMR 1037 Cancer Research Center of Toulouse (CRCT), University Paul Sabatier Toulouse III

- [4] A. Lemarié, V. Lubrano, C. Delmas, A. Lusque, J.-P. Cerapio, M. Perrier, A. Siegfried, F. Arnauduc, Y. Nicaise, P. Dahan, et al., The stemri trial: Magnetic resonance spectroscopy imaging can define tumor areas enriched in glioblastoma stem-like cells, *Science Advances* 9 (44) (2023) eadi0114.
- [5] F. S. Collins, A. D. Barker, Mapping the cancer genome, *Scientific American* 296 (3) (2007) 50–57.
- [6] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al., An image is worth 16x16 words: Transformers for image recognition at scale, arXiv preprint arXiv:2010.11929 (2020).
- [7] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever, et al., Language models are unsupervised multitask learners, *OpenAI blog* 1 (8) (2019) 9.
- [8] P. Chandak, K. Huang, M. Zitnik, Building a knowledge graph to enable precision medicine, *Scientific Data* 10 (1) (2023) 67.  
URL <https://doi.org/10.1038/s41597-023-01960-3>
- [9] N. Dugué, A. Perez, Direction matters in complex networks: A theoretical and applied study for greedy modularity optimization, *Physica A: Statistical Mechanics and its Applications* 603 (2022) 127798.
- [10] M. E. Newman, The structure and function of complex networks, *SIAM review* 45 (2) (2003) 167–256.
- [11] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, R. Girshick, Masked autoencoders are scalable vision learners, in: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022*, pp. 16000–16009.
- [12] T. Chen, S. Kornblith, M. Norouzi, G. Hinton, A simple framework for contrastive learning of visual representations, in: *International conference on machine learning, PMLR, 2020*, pp. 1597–1607.
- [13] X. Dong, J. Shen, Triplet loss in siamese network for object tracking, in: *Proceedings of the European Conference on Computer Vision (ECCV), 2018*, pp. 459–474.

- [14] J. Li, R. Selvaraju, A. Gotmare, S. Joty, C. Xiong, S. C. H. Hoi, Align before fuse: Vision and language representation learning with momentum distillation, *Advances in neural information processing systems* 34 (2021) 9694–9705.
- [15] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-cam: visual explanations from deep networks via gradient-based localization, *International journal of computer vision* 128 (2020) 336–359.
- [16] C. C. Hawkins, T. Ali, S. Ramanadham, A. B. Hjelmeland, Sphingolipid metabolism in glioblastoma and metastatic brain tumors: a review of sphingomyelinases and sphingosine-1-phosphate, *Biomolecules* 10 (10) (2020) 1357.
- [17] O. B. Chernova, A. Hunyadi, E. Malaj, H. Pan, C. Crooks, B. Roe, J. K. Cowell, A novel member of the wd-repeat gene family, *wdr11*, maps to the 10q26 region and is disrupted by a chromosome translocation in human glioblastoma cells, *Oncogene* 20 (38) (2001) 5378–5392.
- [18] H. Xu, Y. Chen, C. Tan, T. Xu, Y. Yan, R. Qin, Q. Huang, C. Lu, C. Liang, Y. Lu, et al., High expression of *wdr1* in primary glioblastoma is associated with poor prognosis, *American Journal of Translational Research* 8 (2) (2016) 1253.
- [19] P. Busek, E. Balaziová, I. Matrasová, M. Hilser, R. Tomas, M. Syrucek, Z. Zemanová, E. Krepela, J. Belacek, A. Sedo, Fibroblast activation protein alpha is expressed by transformed and stromal cells and is associated with mesenchymal features in glioblastoma, *Tumor Biology* 37 (2016) 13961–13971.
- [20] A. Zhang, Z. Huang, W. Tao, K. Zhai, Q. Wu, J. N. Rich, W. Zhou, S. Bao, *Usp33* deubiquitinates and stabilizes *hif-2alpha* to promote hypoxia response in glioma stem cells, *The EMBO Journal* 41 (7) (2022) e109187.
- [21] M. Kojic, N. E. Abbassi, T.-Y. Lin, A. Jones, E. L. Wakeling, E. Clement, V. Nakou, M. Singleton, D. Dobosz, M. Kaliakatsos, et al., A novel *elp1* mutation impairs the function of the elongator complex

and causes a severe neurodevelopmental phenotype, *Journal of Human Genetics* 68 (7) (2023) 445–453.

- [22] M. E. Feigin, B. Xue, M. C. Hammell, S. K. Muthuswamy, G-protein-coupled receptor *gpr161* is overexpressed in breast cancer and is a promoter of cell proliferation and invasion, *Proceedings of the National Academy of Sciences* 111 (11) (2014) 4191–4196.
- [23] S. Wang, J. Yang, W. Huang, Z. Yu, Y. Mao, Y. Feng, J. Chen, Identification of *cers5* as a molecular biomarker in pan-cancer through multiple omics integrative analysis, *Cellular Signalling* 116 (2024) 111054.
- [24] X. Liu, J. Xu, B. Shen, J. Xu, J. Jiang, *Usp33* promotes pancreatic cancer malignant phenotype through the regulation of *tgfbr2/tgf $\beta$*  signaling pathway, *Cell Death & Disease* 14 (6) (2023) 362.
- [25] W.-K. Zhu, W.-H. Xu, J. Wang, Y.-Q. Huang, M. Abudurexiti, Y.-Y. Qu, Y.-P. Zhu, H.-L. Zhang, D.-W. Ye, Decreased *sptlc1* expression predicts worse outcomes in ccrc patients, *Journal of cellular biochemistry* 121 (2) (2020) 1552–1562.
- [26] T. Hamada, M. Souda, T. Yoshimura, S. Sasaguri, K. Hatanaka, T. Tasaki, T. Yoshioka, Y. Ohi, S. Yamada, M. Tsutsui, et al., Anti-apoptotic effects of *pcp4/pep19* in human breast cancer cell lines: a novel oncotarget, *Oncotarget* 5 (15) (2014) 6076.
- [27] G. Aji, F. Li, J. Chen, F. Leng, K. Hu, Z. Cheng, Y. Luo, X. Xu, J. Zhang, Z. Lu, Upregulation of *pcp4* in human aldosterone-producing adenomas fosters human adrenocortical tumor cell growth via akt and ampk pathway, *International Journal of Clinical and Experimental Pathology* 11 (3) (2018) 1197.
- [28] U. Jariwala, J. Prescott, L. Jia, A. Barski, S. Pregizer, J. P. Cogan, A. Arasheben, W. D. Tilley, H. I. Scher, W. L. Gerald, et al., Identification of novel androgen receptor target genes in prostate cancer, *Molecular cancer* 6 (2007) 1–15.