**Identification of transcriptomic hypoxia signature from tumour microenvironment by machine learning: a systematic review**

Abstract

This systematic review comprehensively analyses the application of machine learning in identifying transcriptomic hypoxia signatures (HSs) within the tumour microenvironment, a key factor in understanding tumour progression, tailoring treatment strategies, and potentially improving patient outcomes. Certain HSs have been validated to stratify patients into different risk groups and predict overall survival (OS) (Liu et al., 2020). Traditional methods for identifying HSs, including in vivo–derived analyses using hypoxia markers(Marotta et al., 2011), meta-analysis of microarray-based gene signatures(Seigneuric et al., 2007), and gene function and co-expression patterns analysis (Buffa et al., 2010), provided foundational insights. However, since 2014 (Lalonde et al., 2014), machine learning has revolutionized this field. Its ability to efficiently analyse complex, large datasets has significantly enhanced the accuracy and depth of HS identification. This review synthesizes findings from a comprehensive analysis of literature focused on using machine learning methods to identify HSs in the tumour microenvironment, adhering to strict topic relevance criteria. This review reveals that LASSO and LASSO Cox regression are predominant in this area. Besides, while they are well-established, this review highlights an untapped potential in the application of autoencoders (AEs). The insights gained from this review could pave the way for the HS identification for a broader range of tumour types and the development of novel identification methods.

Keywords: hypoxia signature, tumour, cancer, microenvironment, machine learning, prognosis, autoencoder

Introduction

A HS, a set of differentially expressed genes (DEGs) under hypoxic conditions (low oxygen level) in the tumour microenvironment, plays a pivotal role in measuring tumour hypoxia. Hypoxia, prevalent in various tumours or cancers, is linked with increased tumour aggressiveness, reduced therapy response, especially to radiotherapy (Toustrup et al., 2012), and an overall poorer clinical prognosis(Tawk et al., 2016). Despite of the importance of hypoxia in tumours, its detection remains to be a challenge. Using identified HSs as biomarkers, the intratumoural hypoxia level can be inferred by measuring the expression level of certain HSs. Moreover, HS can also be use to predict more valuable information, including response to immunotherapy (Hong et al., 2021). However, traditional HSs identification has faced significant limitations.

Machine learning, including techniques such as AE, presents a transformative approach in advancing the identification of HSs. AEs, specialized in unsupervised learning tasks like dimensionality reduction and feature learning, offer potential solutions for identifying gene signatures with increased precision. Compared to traditional methods, machine learning requires less empirical gene signature knowledge, automates the signature selection process, and efficiently handles larger, more complex gene datasets. This review is necessitated by the need to systematically examine these advanced machine learning techniques, assessing their impact and potential in refining HS identification.

This review aims to systematically analyse and synthesize all the existing literature on the application of machine learning methods for identifying HSs in the tumour microenvironment. Given the nascent nature of this research area, our scope will be broad and inclusive, covering studies across various tumour types and machine learning techniques.

To effectively collate and assess the findings from the selected papers, the narrative synthesis approach is employed. This method will allow qualitatively summarization and explanation of the diverse range of methodologies and outcomes observed in the studies. The narrative synthesis is particularly suitable given the expected heterogeneity in study designs, machine learning algorithms used, and types of tumours examined. By synthesizing this information narratively, we aim to draw comprehensive insights into the current state of machine learning applications in HS identification and highlight potential directions for future research.

Objectives

The primary objective of this review is to explore and synthesize current methodologies and findings in the field of tumour microenvironment transcriptomics, particularly focusing on the development of HSs using AEs. The research questions guiding this review are:

1. How is machine learning being utilized to develop transcriptomic HSs in the tumour microenvironment?
2. How can AEs be applied in identifying HSs and how effective in comparison with other machine learning methods?

Methodology

1. Identification of Studies

1.1 Searching Strategies

In this systematic review, we employed an "Iterative and Expert-Informed Search Strategy" to meticulously gather a comprehensive set of studies on machine learning applications in identifying HSs within tumour microenvironments. This approach involved dynamically refining search terms based on initial findings and ongoing insights.

Primarily PubMed is utilized, noted for its repository of peer-reviewed biomedical literature, ensuring access to high-quality scientific studies. Recognizing its limitations in encompassing newly published or grey literature, this review supplemented search with Google Scholar. While Google Scholar offers a broader range of literature, including grey and non-peer-reviewed articles, it may include sources with varying degrees of rigor.

Our search terms, carefully selected to cover critical aspects of our research theme, included (hypoxia), (signature), (transcriptomic), ((cancer) or (tumour)), ((micro-environment) or (TME) or (microenvironment)), and ((develop) or (autoencoders) or (AE) or (machine learning) or (deep learning)). Given the field's novelty, we imposed no date restrictions, with relevant studies dating back to 2014. This strategy yielded 90 papers from PubMed and 13 from Google Scholar, augmented by three expert-recommended papers, enriching the search's breadth and depth. For full transparency and reproducibility, the complete search methodology and iterations are detailed in Appendix A.

* 1. Inclusion/Exclusion Criteria and Screening

After the removal of duplicates, the screening of papers was conducted by the author. The initial stage involved evaluating titles and abstracts against the inclusion criteria outlined in Table 1. Papers were included in the review based on their alignment with the inclusion criteria, which encompassed topic relevance, utilization of specified technologies, and validation of results. This process resulted in the selection of 18 papers. Additionally, consultations with field experts led to the identification of three more pertinent publications, bringing the total count to 21 papers included in the review.

|  |  |
| --- | --- |
| **Inclusion Criteria** | **Reason** |
| Focus on machine learning in signatures identification | Ensures that the paper contributes to the understanding of how machine learning is used in the identification of signatures, which is central to the review's aim. |
| Focus on HSs in tumour microenvironment | Directly aligns with the review's objective to understand how HSs are identified in the context of tumour microenvironments. |
| Development of signature conducted on real gene profiles | Ensures that the study's findings are based on empirical data, enhancing the practical applicability and relevance of the research. |
| Validation of signature development | Validation of the developed signatures is crucial to ascertain their reliability and utility in real-world applications. |
|  | |
| **Exclusion Criteria** | **Reason** |
| Using Statistic Method only (e.g., only Cox regression or multivariate analysis) | Studies using statistic methods are not relevant to the specific focus on machine learning. This exclusion ensures the review concentrates on innovative, cutting-edge techniques. |
| Not focusing on HSs (e.g., immune signatures) | Studies focusing on other types of signatures are not relevant to the specific focus on HSs and thus fall outside the scope of the review. |
| Not focusing on gene signatures (e.g., HIF) | Studies not cantered on gene signatures do not contribute to the understanding of transcriptomic HSs, which is the core interest of the review. |
| Non-tumour studies (e.g., DSS-induced colitis) | The review is specifically interested in tumour-related hypoxia, so studies on other diseases are not within its scope. |

Table 1. Inclusion and exclusion criteria

1. Coding and Analysing
   1. Data extraction and Critical Appraisal

The data extraction strategy employed in this systematic review was meticulously designed based on the CHARMS checklist (Moons et al., 2014). This approach ensured that only terms relevant to the review's focus were included, with non-relevant terms being removed or adjusted appropriately. The extracted data was stored in a separate document for clarity and ease of analysis. Key data elements extracted included TUMOUR type, data source, data size, model development, performance, evaluation results, and interpretative and discussion data.

The extracted data were then rigorously assessed by the author following the Critical Appraisal guidelines provided by the PROBAST tool (Wolff et al., 2019). This appraisal aimed to evaluate the risk of bias and the applicability of the studies included in the review.

The critical appraisal process revealed a significant level of bias within the studies: approximately 40% were found to have a high risk of bias, with 30% demonstrating significant concerns regarding applicability. Notably, the most substantial risk of bias was concentrated in the domains of data and analysis assessment.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Author, Year** | **Risk of Bias** | | | | **Applicability** | | | **Overall** | |
| 1. Participants | 2. Predictors | 3. Outcome | 4. Analysis | 1. Participants | 2. Predictors | 3. Outcome | **Risk of Bias** | **Applicability** |
| Jill M. Brooks, 2019 | + | + | + | + | + | + | + | + | + |
| Edian F. Franco, 2013 | + | + | + | + | + | + | + | + | + |
| Cheng-Peng Gui, 2021 | - | + | + | ? | + | + | + | - | + |
| Zhi Liu, 2021 | + | + | ? | ? | + | + | + | ? | + |
| Emilie Lalonde, 2014 | + | + | + | + | + | + | + | + | + |
| Jia Li, 2022 | + | + | + | - | + | + | - | - | - |
| Run Shi , 2021 | + | + | + | + | - | + | + | + | - |
| Jun Shao, 2021 | - | + | + | + | + | + | + | - | + |
| Baohui Zhang, 2020 | - | + | + | - | + | + | + | - | + |
| Qiangnu Zhang, 2021 | + | + | + | + | + | ? | + | + | ? |
| Fanhong Zeng, 2021 | + | + | + | + | + | + | + | + | + |
| Brian Lane, 2022 | - | - | + | ? | + | + | - | - | - |
| Ke Wang, 2022 | + | ? | + | + | + | + | + | ? | + |
| Xiong Tian, 2022 | + | + | + | + | + | - | + | + | - |
| Dongjie Chen, 2021 | + | + | + | + | + | + | - | + | - |
| Yanhong Shou, 2021 | + | + | + | + | + | + | + | + | + |
| Xiangqian Zhang, 2023 | - | + | + | + | + | + | + | - | + |
| Chenyu Nie, 2022 | + | + | + | + | - | + | + | + | - |
| Jinman Zhong, 2021 | - | + | + | + | + | + | + | - | + |
| Xia Yang, 2021 | + | - | + | + | + | + | + | - | + |

**Table 2.** Risk of bias and applicability assessment

2.2 Synthesis

The findings of this review are reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Moher et al., 2009). Owing to considerable heterogeneity in study design, tumour types, result measurement techniques, validation methods, and data sizes, a meta-analysis was deemed unsuitable. Instead, a narrative synthesis approach was utilized to integrally interpret the diverse studies focusing on the identification of hypoxia signatures (HSs) across various tumour types using a range of machine learning methodologies.

Table 3 provides a comprehensive summary, detailing the characteristics and methods employed for HS identification in the included studies. These studies are systematically categorized based on the methodologies used, such as LASSO Cox regression, LASSO, Random Forest, K-means, Logistic Regression, Unsupervised Clustering, t-SNE, and Autoencoders. It is noted that some studies may fall into multiple categories if they employed several methods. In such instances, the function and effectiveness of each method in relation to the study are discussed separately.

|  |  |  |  |
| --- | --- | --- | --- |
| First Author | Publication Year | Methods | Tumour |
| Cheng-Peng Gui | 2021 | t-SNE and LASSO | ccRCC |
| Yifan Liu | 2020 | t-SNE and LASSO Cox Regression | Gastric Cancer |
| Zhi Liu | 2021 | LASSO | BLCA |
| Yanhong Shou | 2021 | LASSO | Melanoma |
| Jinman Zhong | 2021 | LASSO | AML |
| Xia Yang | 2021 | LASSO | Breast Cancer |
| Ke Wang | 2022 | LASSO | GBM |
| Chenyu Nie | 2022 | LASSO | Cervical Cancer |
| Xiong Tian | 2022 | LASSO | PAAD |
| Fanhong Zeng | 2021 | K-mean | HCC |
| Brian Lane | 2022 | K-mean | LUDA |
| Jun Shao | 2021 | K-mean and LASSO Cox Regression | LUAD |
| Baohui Zhang | 2020 | k-mean and LASSO Cox Regression | HCC |
| Jill M. Brooks | 2019 | Unsupervised Hierarchical Clustering | HNC |
| Jia Li | 2022 | Random Forest | Breast Cancer |
| Qiangnu Zhang | 2021 | LASSO Cox Regression | HCC |
| Run Shi | 2021 | LASSO Cox Regression | LUDA |
| Dongjie Chen | 2021 | LASSO Cox Regression | PDAC |
| Xiangqian Zhang | 2023 | LASSO Cox regression | Gastric Cancer |
| Edian F. Franco | 2013 | Autoencoder | GBM |

Table 3. Methods and Tumour Type of Selected Papers

This methodical categorization has enabled the identification of distinct patterns and variations in the functionality and efficacy of these methods for HS identification across different tumour types. Such a structured approach provides a comprehensive understanding of how each technique contributes to the broader field of tumour microenvironment analysis and hypoxia signature identification.

1. Findings

In this systematic review, we observed a predominant application of LASSO and LASSO Cox Regression methods, utilized in 40% and 35% of the studies, respectively.

3.1 t-SNE

t-SNE is a nonparametric unsupervised algorithm for dimensionality reduction. Liu et al. (Liu et al., 2020) utilized t-SNE to classify 200 hypoxia hallmark gene sets into “hypoxia-high” and “hypoxia-low” groups for gastric cancer risk stratification. Similarly, Gui et al. (Gui et al., 2021) applied t-SNE in clear cell renal cell carcinoma (ccRCC), categorizing multiomics data into 'low-hypoxia', 'moderate-hypoxia', and 'high-hypoxia' to identify hypoxia-related DEGs.

3.2 LASSO

LASSO, a linear regression variant incorporating regularization for dimensionality reduction, has been widely applied in 28.6% studies of this review. Gui et al used LASSO on the multiomics dataset obtained from the Cancer Genome Atlas (TCGA) including the 'low-hypoxia & high-immunity' and 'high-hypoxia & low-immunity' DEGs to produce a refined set of variables including the signature that most relevant for prognosis of ccRCC. Differently, Shou et al. (Shou et al., 2021) applied LASSO on 200 hypoxia DEGs identified from the log-rank test for melanoma. Applying LASSO on DEGs obtained from univariate Cox regression, Tian et al. (Tian et al., 2022) implemented LASSO in pancreatic adenocarcinoma (PAAD) to select the most relevant signatures from hypoxia-stemness-related genes, Wang et al. (Wang et al., 2022) employed LASSO for hypoxia DEGs filtration in Glioblastoma Multiforme (GBM), similar to Liu et al. (Liu et al., 2021) in bladder cancer (BLCA), Shao et al. (Shao et al., 2021) in lung adenocarcinoma (LUDA), Nie et al. (Nie et al., 2022) in cervical cancer, Zhong et al. (Zhong et al., 2021) in Acute Myeloid Leukaemia (AML) and Yang et al. (Yang et al., 2021) in triple-negative breast cancer.

3.3 K-mean

K-mean is an unsupervised clustering algorithm that partitioning a dataset into a set number of clusters. Lane et al. (Lane et al., 2022) employed K-mean on the TCGA-LUAD dataset to segregate genes into hypoxia and non-hypoxia in total two sets for DEG detection. Experimenting with a range of K values 2 to 9, Zeng et al. (Zeng et al., 2021) identified LUAD HSs, using K-mean to group genes data from TCGA based on expression similarities under hypoxic conditions, similar to Zhang et al. for hepatocellular carcinoma (HCC) (Zhang et al., 2020) and Shao et al. (Shao et al., 2021) who used data from molecular signatures database for LUAD.

3.4 Unsupervised Hierarchical Clustering

Unsupervised hierarchical clustering is a clustering algorithm used to group data points based on their similarity without pre-labelled categories. Brooks et al. (Brooks et al., 2019) focused on developing a prognostic classifier for Head and Neck Cancer (HNC), employing hierarchical clustering to subgroup a 54-gene hypoxia-immune signature derived from literature, using Spearman distance and the Ward criterion.

3.5 Random Survival Forest (RSF)

RSF is a an extension of the Random Forests algorithm that accounts for interactions and non-linear relationships between variables. Li et al. (Li et al., 2022) applied RSF to identify the most relevant DEGs for breast cancer prognosis from univariate Cox regression, focusing on Hypoxia- and Lactate Metabolism-Related Signatures.

3.6 LASSO Cox Regression

LASSO Cox regression combines LASSO penalties with Cox proportional hazards models, suitable for predicting survival times in medical research. Shi et al. (Shi et al., 2021) identified 10 HSs for stage I lung adenocarcinoma, applying LASSO Cox regression to 199 candidate genes from the weighted gene co-expression network analysis (WGCNA). Zhang et al. (Zhang et al., 2021) proposed a HS for hepatocellular carcinoma using LASSO Cox regression on TNM staging and hypoxia scores, which includes HSs. Chen et al. (Chen et al., 2021) developed a Hypoxia- and Immune-Associated Signature for Pancreatic Ductal Adenocarcinoma, employing LASSO Cox regression post-Univariate Cox regression on DEGs. In a similar approach, after univariate cox regression, Zhang et al. (Zhang et al., 2023) developed a hypoxia–glycolysis–lactylation-related gene signature using LASSO Cox regression and Liu et al. (Liu et al., 2020) developed a hypoxia-immune-related signature.

3.7 AE

AEs are unsupervised deep learning algorithms used for dimensionality reduction. Constructed from layers of simple neurons, where each layer's output feeds into the next, AEs typically exhibit a "butterfly" structure characterized by equal numbers of inputs and outputs and bottleneck hidden layers in the middle (Franco et al., 2021). Franco et al. evaluated the performance of various AEs with different regularizations in detecting GBM subtypes. Their study utilized diverse data types, including mRNA, DNA methylation, and miRNA expression from colorectal adenocarcinoma datasets. The AEs processed features extracted from these datasets, with reconstructed data subsequently applied to unsupervised clustering, including K-means, to identify GBM subtypes. In a related study focusing on tumour subtype detection, Zhang et al. (Zhang et al., 2021) employed unsupervised hierarchical clustering on pre-selected data for HCC.

1. Discussion

This systematic review notices a ambiguity in the application and description of LASSO and LASSO Cox Regression methods in the development of HSs. This ambiguity is exemplified in cases like Shou et al. (2021), where LASSO Cox Regression is claimed in the abstract, but the method section reveals the use of LASSO. Additionally, the typical progression from LASSO to Multivariate Cox Regression, for instance as seen in Yang et al.'s work (Yang et al., 2021), further blurs the functional distinction between these two methodologies. Despite these inconsistencies, LASSO and LASSO Cox Regression have emerged as the predominant techniques in HS development, underlining their central role in this research domain.

The papers reviewed collectively suggest a generalized three-step process for developing HSs in various tumour types:

Identification of Hypoxia DEGs: This initial step involves obtaining all hypoxia-related DEGs, which can be sourced through literature surveys, directly from databases, or identified using algorithms like K-mean or unsupervised Hierarchical Clustering.

Dimensionality Reduction on DEGs: Subsequently, dimensionality reduction techniques such as LASSO, LASSO Cox, or Random Forest are applied to these DEGs to pinpoint prognostic HSs.

Prognostic Model Development: The final step involves developing a prognostic model, which may take the form of a score or a more complex model incorporating additional features.

However, due to the variance in tumour types, performance measures, and processes of HS identification, conducting comparative research on the efficiency of different methods remains challenging. While the quality of each study has been rigorously evaluated in the Critical Appraisal section, a direct comparison of methodological efficacy is complex.

Moreover, our review has observed the emerging use of AEs in tasks similar to HS identification. Demonstrating their capability in related fields, AEs possess the unique ability to amalgamate the first two steps of HS identification. They accept raw gene data and reconstruct it to reduce the feature list in a single step. The processed data thus obtained represents the most relevant signatures, which can then be effectively discerned using methods like K-mean clustering. Consequently, AEs show considerable promise for future applications in HS identification.

1. Conclusion

Our findings highlight a general three-step process prevalent in HS development across studies: the identification of hypoxia DEGs, dimensionality reduction on these DEGs, and the subsequent development of prognostic models. However, the review also underscores challenges in methodological consistency and clarity, particularly in distinguishing between similar techniques like LASSO and LASSO Cox Regression.

The potential of AEs to streamline the HS identification process by simultaneously obtain DEGs from raw gene data and reducing DEG lists is particularly noteworthy. This capability not only enhances efficiency but also opens new avenues for extracting the most relevant biological signatures from vast and complex datasets.

Looking ahead, it is clear that as the field of tumour biology continues to grapple with the intricacies of the tumour microenvironment, the role of sophisticated machine learning techniques will become increasingly pivotal. Future research should focus on refining these methods, exploring their comparative efficacies, and developing more streamlined approaches to HS identification. Such advancements will be crucial in enhancing our understanding of tumour behaviour and in guiding the development of targeted therapeutic strategies.

In summary, this systematic review sheds light on the current state of HS identification, emphasizing both the achievements and challenges in this dynamic field, and sets a direction for future research that harnesses the power of advanced machine learning techniques to unravel the complexities of tumour hypoxia.

Brooks, J. M., Menezes, A. N., Ibrahim, M., Archer, L., Lal, N., Bagnall, C. J., von Zeidler, S. V., Valentine, H. R., Spruce, R. J., Batis, N., Bryant, J. L., Hartley, M., Kaul, B., Ryan, G. B., Bao, R., Khattri, A., Lee, S. P., Ogbureke, K. U. E., Middleton, G., . . . Mehanna, H. (2019). Development and Validation of a Combined Hypoxia and Immune Prognostic Classifier for Head and Neck Cancer. *Clinical Cancer Research*, *25*(17), 5315-5328. <https://doi.org/10.1158/1078-0432.CCR-18-3314>

Buffa, F. M., Harris, A. L., West, C. M., & Miller, C. J. (2010). Large meta-analysis of multiple cancers reveals a common, compact and highly prognostic hypoxia metagene [Article]. *British Journal of Cancer*, *102*(2), 428-435. <https://doi.org/10.1038/sj.bjc.6605450>

Chen, D., Huang, H., Zang, L., Gao, W., Zhu, H., & Yu, X. (2021). Development and Verification of the Hypoxia- and Immune-Associated Prognostic Signature for Pancreatic Ductal Adenocarcinoma. *Front Immunol*, *12*, 728062. <https://doi.org/10.3389/fimmu.2021.728062>

Franco, E. F., Rana, P., Cruz, A., Calderón, V. V., Azevedo, V., Ramos, R. T. J., & Ghosh, P. (2021). Performance Comparison of Deep Learning Autoencoders for Cancer Subtype Detection Using Multi-Omics Data. *Cancers*, *13*(9).

Gui, C.-P., Wei, J.-H., Chen, Y.-H., Fu, L.-M., Tang, Y.-M., Cao, J.-Z., Chen, W., & Luo, J.-H. (2021). A new thinking: extended application of genomic selection to screen multiomics data for development of novel hypoxia-immune biomarkers and target therapy of clear cell renal cell carcinoma. *Briefings in Bioinformatics*, *22*(6), bbab173. <https://doi.org/10.1093/bib/bbab173>

Hong, S., Zhang, Y., Cao, M., Lin, A., Yang, Q., Zhang, J., Luo, P., & Guo, L. (2021). Hypoxic Characteristic Genes Predict Response to Immunotherapy for Urothelial Carcinoma [Original Research]. *Frontiers in Cell and Developmental Biology*, *9*. <https://www.frontiersin.org/articles/10.3389/fcell.2021.762478>

Lalonde, E., Ishkanian, A. S., Sykes, J., Fraser, M., Ross-Adams, H., Erho, N., Dunning, M. J., Halim, S., Lamb, A. D., Moon, N. C., Zafarana, G., Warren, A. Y., Meng, X., Thoms, J., Grzadkowski, M. R., Berlin, A., Have, C. L., Ramnarine, V. R., Yao, C. Q., . . . Bristow, R. G. (2014). Tumour genomic and microenvironmental heterogeneity for integrated prediction of 5-year biochemical recurrence of prostate cancer: a retrospective cohort study. *The Lancet Oncology*, *15*(13), 1521-1532. <https://doi.org/https://doi.org/10.1016/S1470-2045(14)71021-6>

Lane, B., Khan, M. T., Choudhury, A., Salem, A., & West, C. M. L. (2022). Development and validation of a hypoxia-associated signature for lung adenocarcinoma. *Sci Rep*, *12*(1), 1290. <https://doi.org/10.1038/s41598-022-05385-7>

Li, J., Qiao, H., Wu, F., Sun, S., Feng, C., Li, C., Yan, W., Lv, W., Wu, H., Liu, M., Chen, X., Liu, X., Wang, W., Cai, Y., Zhang, Y., Zhou, Z., Zhang, Y., & Zhang, S. (2022). A novel hypoxia- and lactate metabolism-related signature to predict prognosis and immunotherapy responses for breast cancer by integrating machine learning and bioinformatic analyses [Original Research]. *Frontiers in Immunology*, *13*. <https://doi.org/10.3389/fimmu.2022.998140>

Liu, Y., Wu, J., Huang, W., Weng, S., Wang, B., Chen, Y., & Wang, H. (2020). Development and validation of a hypoxia-immune-based microenvironment gene signature for risk stratification in gastric cancer. *Journal of Translational Medicine*, *18*(1), 201. <https://doi.org/10.1186/s12967-020-02366-0>

Liu, Z., Tang, Q., Qi, T., Othmane, B., Yang, Z., Chen, J., Hu, J., & Zu, X. (2021). A Robust Hypoxia Risk Score Predicts the Clinical Outcomes and Tumor Microenvironment Immune Characters in Bladder Cancer [Original Research]. *Frontiers in Immunology*, *12*. <https://doi.org/10.3389/fimmu.2021.725223>

Marotta, D., Karar, J., Jenkins, W. T., Kumanova, M., Jenkins, K. W., Tobias, J. W., Baldwin, D., Hatzigeorgiou, A., Alexiou, P., Evans, S. M., Alarcon, R., Maity, A., Koch, C., & Koumenis, C. (2011). In vivo profiling of hypoxic gene expression in gliomas using the hypoxia marker EF5 and laser-capture microdissection [Article]. *Cancer Research*, *71*(3), 779-789. <https://doi.org/10.1158/0008-5472.CAN-10-3061>

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *Annals of Internal Medicine*, *151*(4), 264-269. <https://doi.org/10.7326/0003-4819-151-4-200908180-00135>

Moons, K. G. M., de Groot, J. A. H., Bouwmeester, W., Vergouwe, Y., Mallett, S., Altman, D. G., Reitsma, J. B., & Collins, G. S. (2014). Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies: The CHARMS Checklist. *PLOS Medicine*, *11*(10), e1001744. <https://doi.org/10.1371/journal.pmed.1001744>

Nie, C., Qin, H., & Zhang, L. (2022). Identification and validation of a prognostic signature related to hypoxic tumor microenvironment in cervical cancer. *PLOS ONE*, *17*(6), e0269462. <https://doi.org/10.1371/journal.pone.0269462>

Seigneuric, R., Starmans, M. H. W., Fung, G., Krishnapuram, B., Nuyten, D. S. A., van Erk, A., Magagnin, M. G., Rouschop, K. M., Krishnan, S., Rao, R. B., Evelo, C. T. A., Begg, A. C., Wouters, B. G., & Lambin, P. (2007). Impact of supervised gene signatures of early hypoxia on patient survival [Article]. *Radiotherapy and Oncology*, *83*(3), 374-382. <https://doi.org/10.1016/j.radonc.2007.05.002>

Shao, J., Zhang, B., Kuai, L., & Li, Q. (2021). Integrated analysis of hypoxia-associated lncRNA signature to predict prognosis and immune microenvironment of lung adenocarcinoma patients. *Bioengineered*, *12*(1), 6186-6200. <https://doi.org/10.1080/21655979.2021.1973874>

Shi, R., Bao, X., Unger, K., Sun, J., Lu, S., Manapov, F., Wang, X., Belka, C., & Li, M. (2021). Identification and validation of hypoxia-derived gene signatures to predict clinical outcomes and therapeutic responses in stage I lung adenocarcinoma patients [Research Paper]. *Theranostics*, *11*(10), 5061-5076. <https://doi.org/10.7150/thno.56202>

Shou, Y., Yang, L., Yang, Y., Zhu, X., Li, F., & Xu, J. (2021). Determination of hypoxia signature to predict prognosis and the tumor immune microenvironment in melanoma. *Mol Omics*, *17*(2), 307-316. <https://doi.org/10.1039/d0mo00159g>

Tawk, B., Schwager, C., Deffaa, O., Dyckhoff, G., Warta, R., Linge, A., Krause, M., Weichert, W., Baumann, M., Herold-Mende, C., Debus, J., & Abdollahi, A. (2016). Comparative analysis of transcriptomics based hypoxia signatures in head- and neck squamous cell carcinoma. *Radiotherapy and Oncology*, *118*(2), 350-358. <https://doi.org/https://doi.org/10.1016/j.radonc.2015.11.027>

Tian, X., Zheng, J., Mou, W., Lu, G., Chen, S., Du, J., Zheng, Y., Chen, S., Shen, B., Li, J., & Wang, N. (2022). Development and validation of a hypoxia-stemness-based prognostic signature in pancreatic adenocarcinoma. *Front Pharmacol*, *13*, 939542. <https://doi.org/10.3389/fphar.2022.939542>

Toustrup, K., Sørensen, B. S., Alsner, J., & Overgaard, J. (2012). Hypoxia Gene Expression Signatures as Prognostic and Predictive Markers in Head and Neck Radiotherapy. *Seminars in Radiation Oncology*, *22*(2), 119-127. <https://doi.org/https://doi.org/10.1016/j.semradonc.2011.12.006>

Wang, K., Lu, Y., Liu, Z., Diao, M., & Yang, L. (2022). Establishment and External Validation of a Hypoxia-Derived Gene Signature for Robustly Predicting Prognosis and Therapeutic Responses in Glioblastoma Multiforme. *Biomed Res Int*, *2022*, 7858477. <https://doi.org/10.1155/2022/7858477>

Wolff, R. F., Moons, K. G. M., Riley, R. D., Whiting, P. F., Westwood, M., Collins, G. S., Reitsma, J. B., Kleijnen, J., & Mallett, S. (2019). PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Annals of Internal Medicine*, *170*(1), 51-58. <https://doi.org/10.7326/M18-1376>

Yang, X., Weng, X., Yang, Y., Zhang, M., Xiu, Y., Peng, W., Liao, X., Xu, M., Sun, Y., & Liu, X. (2021). A combined hypoxia and immune gene signature for predicting survival and risk stratification in triple-negative breast cancer. *Aging (Albany NY)*, *13*(15), 19486-19509. <https://doi.org/10.18632/aging.203360>

Zeng, F., Zhang, Y., Han, X., Zeng, M., Gao, Y., & Weng, J. (2021). Employing hypoxia characterization to predict tumour immune microenvironment, treatment sensitivity and prognosis in hepatocellular carcinoma. *Computational and Structural Biotechnology Journal*, *19*, 2775-2789. <https://doi.org/https://doi.org/10.1016/j.csbj.2021.03.033>

Zhang, B., Tang, B., Gao, J., Li, J., Kong, L., & Qin, L. (2020). A hypoxia-related signature for clinically predicting diagnosis, prognosis and immune microenvironment of hepatocellular carcinoma patients. *J Transl Med*, *18*(1), 342. <https://doi.org/10.1186/s12967-020-02492-9>

Zhang, Q., Qiao, L., Liao, J., Liu, Q., Liu, P., & Liu, L. (2021). A novel hypoxia gene signature indicates prognosis and immune microenvironments characters in patients with hepatocellular carcinoma. *J Cell Mol Med*, *25*(8), 3772-3784. <https://doi.org/10.1111/jcmm.16249>

Zhang, X., Li, Y., & Chen, Y. (2023). Development of a Comprehensive Gene Signature Linking Hypoxia, Glycolysis, Lactylation, and Metabolomic Insights in Gastric Cancer through the Integration of Bulk and Single-Cell RNA-Seq Data. *Biomedicines*, *11*(11). <https://doi.org/10.3390/biomedicines11112948>

Zhong, J., Wu, H., Bu, X., Li, W., Cai, S., Du, M., Gao, Y., & Ping, B. (2021). Establishment of Prognosis Model in Acute Myeloid Leukemia Based on Hypoxia Microenvironment, and Exploration of Hypoxia-Related Mechanisms [Original Research]. *Frontiers in Genetics*, *12*. <https://doi.org/10.3389/fgene.2021.727392>

Appendix A.

Please contact author to access appendix due to the wordcount limitation.