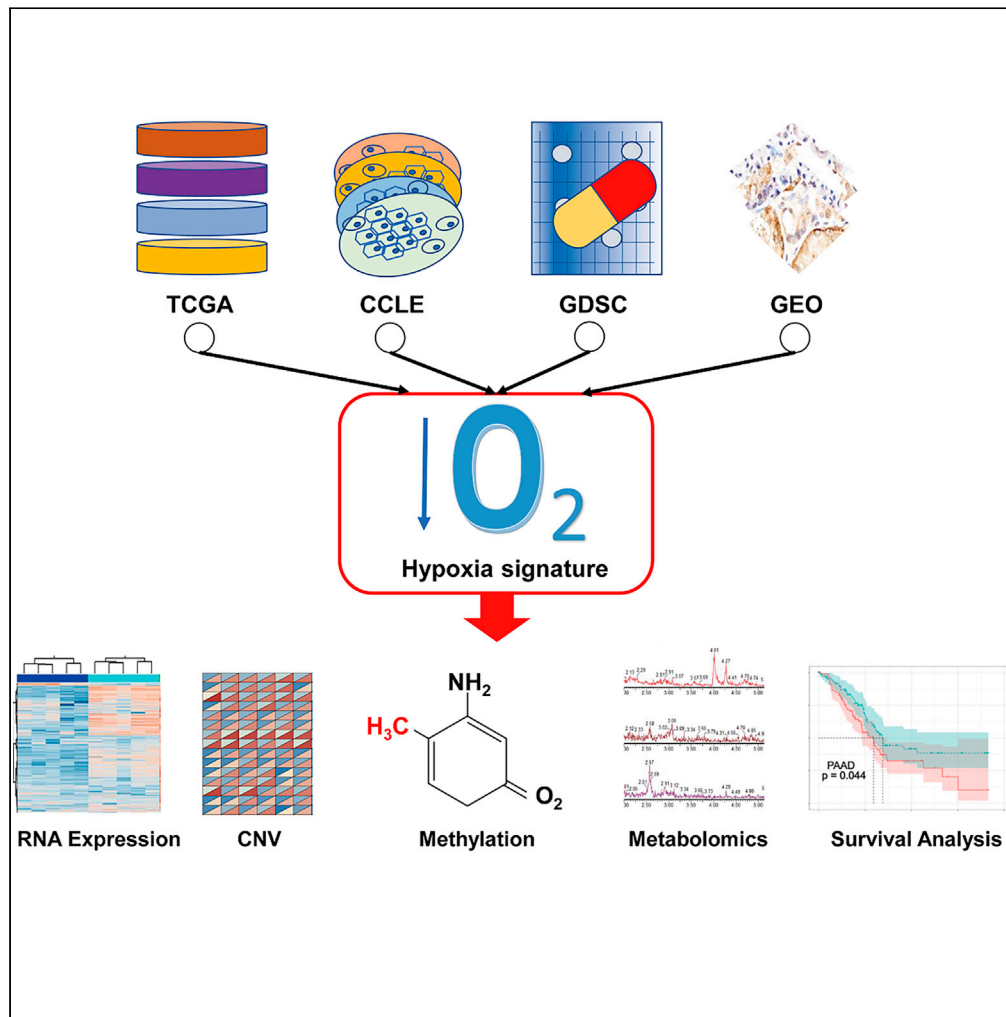# iScience

**Article**

# Integrative Analysis of Hypoxia-Associated Signature in Pan-Cancer



Qian Zhang, Rui Huang, Hanqing Hu, ..., Zheng Liu, Jiaying Li, Guiyu Wang

zhangqian1171@126.com (Q.Z.)
guiywang@163.com (G.W.)

**HIGHLIGHTS**

Majority of the 15 genes of the hypoxia signature are upregulated in cancers

The signature is closely related to cancer hallmarks and metabolism pathways

CNV is frequent, whereas mutation frequency is low across different cancer types

The 15 genes of the signature are risk factors of overall survival

CellPress

### Article

# Integrative Analysis of Hypoxia-Associated Signature in Pan-Cancer

Qian Zhang,[1,3,]* Rui Huang,[1] Hanqing Hu,[1] Lei Yu,[1] Qingchao Tang,[1] Yangbao Tao,[1] Zheng Liu,[2] Jiaying Li,[1] and Guiyu Wang[1,]*

## SUMMARY

**Hypoxia is serving crucial roles in cancers. This study aims to comprehensively analyze the molecular features and clinical relevance of a well-defined hypoxia-associated signature in pan-cancer using multi-omics data. Data were acquired from TCGA, CCLE, GDSC, and GEO. RNA expression pattern, copy number variation (CNV), methylation, and mutation of the signature were analyzed. The majority of the 15 genes were upregulated in cancer tissues compared with normal tissue, and RNA expression was negatively associated with methylation level. CNV occurred in almost all the cancers, whereas mutation frequency was low across different cancer types. The signature was also closely related to cancer hallmarks and cancer-related metabolism pathways. NDRG1 was upregulated in kidney cancer tissues as indicated by immunohistochemistry. Besides, most of the 15 genes were risk factors for patients' overall survival. Our results provide a valuable resource that will guide both mechanistic and therapeutic analyses of the hypoxia signature in cancers.**

## INTRODUCTION

Cancer is a major burden worldwide and is also a leading cause of death (Siegel et al., 2019). It is a multi-step disease and characterized by complex biological features. Hypoxia, which is defined as reduced oxygen levels, is one of the hallmarks of malignant tumors (Giaccia and Schipani, 2010). Most of the tumor cells have insufficient blood supply due to rapidly increasing cell proliferation and increasing tumor mass, resulting in a hypoxic tumor microenvironment (Foster et al., 2014), and hypoxia has a crucial role in shaping the behavior of cancers. It has been reported that hypoxia could cause tumor cells to acquire more aggressive phenotypes. Hypoxia can also induce epithelial-mesenchymal transition (EMT) (Muz et al., 2015), which is closely related to tumor migration and invasion (Mittal, 2018). Moreover, hypoxia promotes self-renewal capability and suppresses the differentiation of stem cells in certain tumors (Li et al., 2009). Besides, hypoxia is associated with abnormal vasculature (Carmeliet and Jain, 2011), metastasis, and radio-chemotherapy resistance (Muz et al., 2015).

Some approaches, including both invasive and noninvasive ones, have been developed to measure tumor oxygen levels in tumors (Vaupel et al., 2007). However, these procedures have limited applications. Therefore a robust and clinically applicable hypoxia gene signature has been introduced (Buffa et al., 2010). The signature is composed of 15 genes with known function: VEGF (Vascular endothelial growth factor A), PGAM1 (Phosphoglycerate mutase 1), ENO1 (Enolase 1), LDHA (Lactate dehydrogenase A), TPI1 (Triosephosphate isomerase 1), P4HA1 (Prolyl 4-hydroxylase, a-polypeptide I), MRPS17 (Mitochondrial ribosomal protein S17), ADM (Adrenomedullin), NDRG1 (N-myc downstream regulated 1), TUBB6 (Tubulin, b6), ALDOA (Aldolase A, fructose-bisphosphate), MIF (Macrophage migration inhibitory factor), SLC2A1 (Solute carrier family 2, member 1), CDKN3 (Cyclin-dependent kinase inhibitor 3), and ACOT7 (Acyl-CoA thioesterase 7). The signature is the best performer, proved by recent extensive researches that have evaluated and validated the power of different hypoxia signatures (Ye et al., 2019).

Despite the presence of studies about the hypoxia-associated signature, there is still a lack of researches about the expression, copy number variation (CNV), methylation, or mutations of these hypoxia-related genes in tumors. In this study, we aimed to comprehensively identify the molecular features and clinical relevance of these genes across a wide variety of cancer types using multi-omics data.

[1]Department of Colorectal Surgery, the Second Affiliated Hospital of Harbin Medical University, Harbin, Heilongjiang Province 150086, China

[2]Department of Colorectal Surgery, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Science and Peking Union Medical College, Beijing 100000, China

[3]Lead Contact

*Correspondence:
zhangqian1171@126.com (Q.Z.),
guiywang@163.com (G.W.)
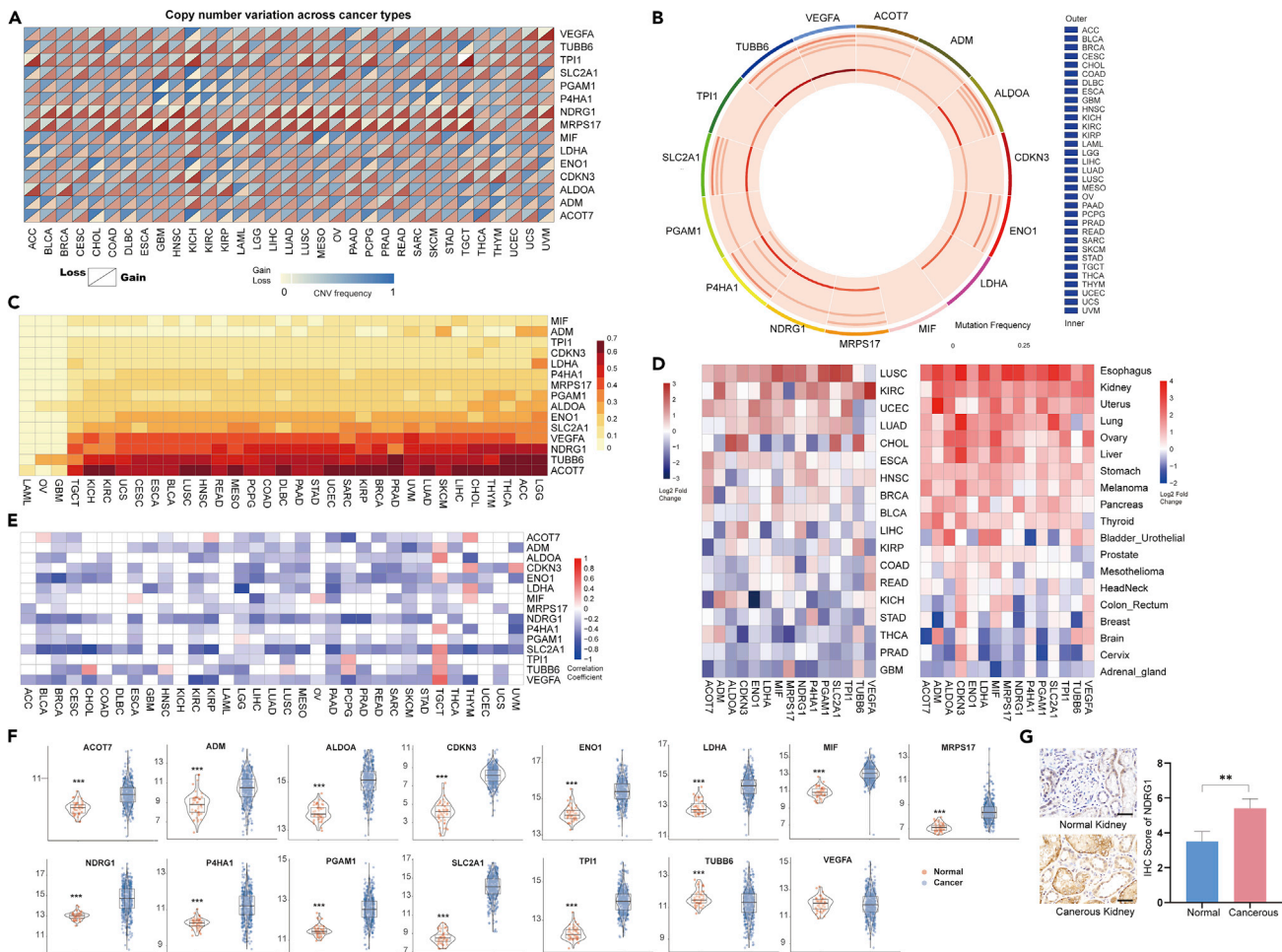https://doi.org/10.1016/j.isci.2020.101460

**Figure 1. Genomic Alterations of Hypoxia-Associated Signature in Different Cancer Tissues**

(A) CNV of the hypoxia-associated signature in cancers. The upper half of each rectangle exhibits the loss frequency, and the bottom half exhibits the gain frequency.

(B) Mutation frequency of the signature among different cancer types.

(C) Heatmap of the methylation level of the 15 genes in 33 cancer types.

(D) Log2 fold change of the genes in cancer tissues versus normal tissues. Both TCGA (left) and GEO (right) data were used.

(E) Correlation plot of RNA expression and methylation level of each gene in each cancer type. The white grid means p > 0.05

(F and G) (F) Violin plot of each gene in normal tissues and cancer tissues in LIHC. (G) IHC result of NDRG1 expression in kidney cancer tissues.

Data were represented as mean ± SEM. **p < 0.01; ***p < 0.001.

## RESULTS

### Genomic Alterations of Hypoxia-Associated Signature in Different Cancer Tissues

A common and compact hypoxia metagene has been identified and selected (Buffa et al., 2010). We analyzed the molecular landscape of the 15-gene hypoxia signature in various cancers to provide basic information for future studies. We analyzed CNV, methylation, and mutation data from TCGA, CCLE, GDSC, and GEO (Database: TCGA, CCLE, GDSC. and GEO). As shown in Figure 1A, CNV was observed in all the hypoxia-related genes, among which NDRG1 and MRPS17 had the highest gain frequency across the 33 cancer types. The average mutation frequency of the 15 genes was low, ranging from 0 to 0.25 with an average value of 0.01 (Figure 1B).

The methylation data from TCGA was also analyzed. The beta value was used to quantify methylation status. Methylation was observed in most of the 15 genes with the beta value range being 0–0.7 (Figure 1C). ACOT7, TUBB6, and NDRG1 were the most highly methylated genes as shown in Figure 1C. As gene methylation is closely correlated with mRNA expression, gene expression data from TCGA and GEO were then

analyzed. Gene expression of normal tissues and cancer tissues was compared. Interestingly, both TCGA and GEO data showed that lung, kidney, and uterus cancer tissues had general higher hypoxia-associated gene expression compared with normal tissues (Figure 1D). Linear models were also used to adjust for tumor purity, ploidy, age, sex, and race (Table S1). To assess the relationship between methylation and gene expression, the correlation was analyzed, and it indicated that there was a reverse relationship between RNA expression and methylation level in most of the 33 cancer types (Figure 1E). In particular, RNA expression data from LUSC were further analyzed, and it showed that 14 of the 15 genes were significantly upregulated in cancer tissues compared with normal tissue with p < 0.001 (Figure 1F). As NDRG1 was significantly upregulated in kidney cancer tissues in both GEO and TCGA data, we validated the results in tissues by immunohistochemistry (IHC). It showed that NDRG1 was significantly higher in kidney cancer tissues (Figure 1G).

### Genomic Alterations of Hypoxia-Associated Signature in Different Cancer Cell Lines

To further understand the molecular landscape of this 15-gene signature (Buffa et al., 2010), we also analyzed the data of cancer cell lines from CCLE and GDSC. Data of 715 cell lines of 24 cancer types in CCLE and 988 cell lines of 30 cancer types in GDSC were identified. Unexpectedly, CNV frequency of most cell lines in CCLE was quite low ranging from 0 to 0.65, whereas CNV frequency in GDSC was high on average ranging from 0 to 1 (Figures 2A and 2B).

UCEC exhibited the highest mutation frequency among all the cancer types, which is consistent with the fact that UCEC has a higher global mutation burden. The mutation frequency of the genes in CCLE was similar to that of GDSC, both ranging from 0 to less than 0.5 (Figures 2C and 2D).

Hypoxia is a potent microenvironmental factor promoting metastatic progression (Rankin and Giaccia, 2016). To evaluate and validate the potential role of the 15-gene signature in metastasis, we further assessed the expression pattern of the signature in metastatic and primary cancer cell lines. The results revealed that the expression of these genes was on an average equal to or higher in metastatic cell lines in most cancer types (Figures 2E and 2F), indicating a potential role of these genes in cancer metastasis.

Unexpectedly, there was some inconsistency between CCLE and GDSC databases, which could be due to the use of different array platforms, experimental artifacts, measurement error, or divergence of the cell lines (Haibe-Kains et al., 2013). Extra attention should be paid for future researchers when using these databases.

### Hypoxia-Associated Signature and Cancer Pathways

We hypothesized that a variety of pathways interacted with the hypoxia-related signature. Moreover, genes do not function in isolation (Li et al., 2019); we thus first investigated the expression correlation among the hypoxia gene signature. To our expectations, the majority of these genes were positively related to each other (Figure 3A). In particular, TPI1 and ENO1 were most closely correlated. GSVA analysis was adopted to assess the involvement of the signature in cancer-related hallmarks. Cancer hallmark gene sets from MSigDB collections, which summarize and represent specific well-defined biological states or processes (Subramanian et al., 2005), were applied in GSVA analysis. We then examined the correlation between the expression of individual genes of the signature and the activity of cancer hallmark-related pathways. We found that the expression of the signature was correlated with the activation (correlation coefficient > 0, p < 0.01) or inhibition (correlation coefficient < 0, p < 0.01) of multiple oncogenic pathways (Figure 3B). E2F target pathway, MYC target pathway, glycolysis pathway, hypoxia pathway, and unfolded protein response pathway were the most frequently and highly activated pathways, whereas myogenesis pathway, hedgehog signaling pathway, heme metabolism pathway, bile acid (BA) metabolism pathway, and Kras signaling pathway were the most highly inhibited pathways (Figure 3B). Besides, the activated pathways greatly outnumber the inhibited pathways (Figure 3C), indicating a cancer promotion role of these genes.

To further study the role of hypoxia-related genes in the regulation of metabolites, metabolomics data were retrieved from CCLE. The correlativity between metabolomics and transcriptomics data of the 15-gene signature were then analyzed. RNA-metabolite pairs with |R|≥0.6 and p < 0.05 were shown in Figure 3D. We further investigated the metabolic pathways that were engaged in the network. It showed that about 30 metabolism pathways were included in the network. Interestingly, most of these pathways were identified as common cancer-related pathways like BA biosynthesis, biopterin metabolism, and lysine metabolism (Table S2).
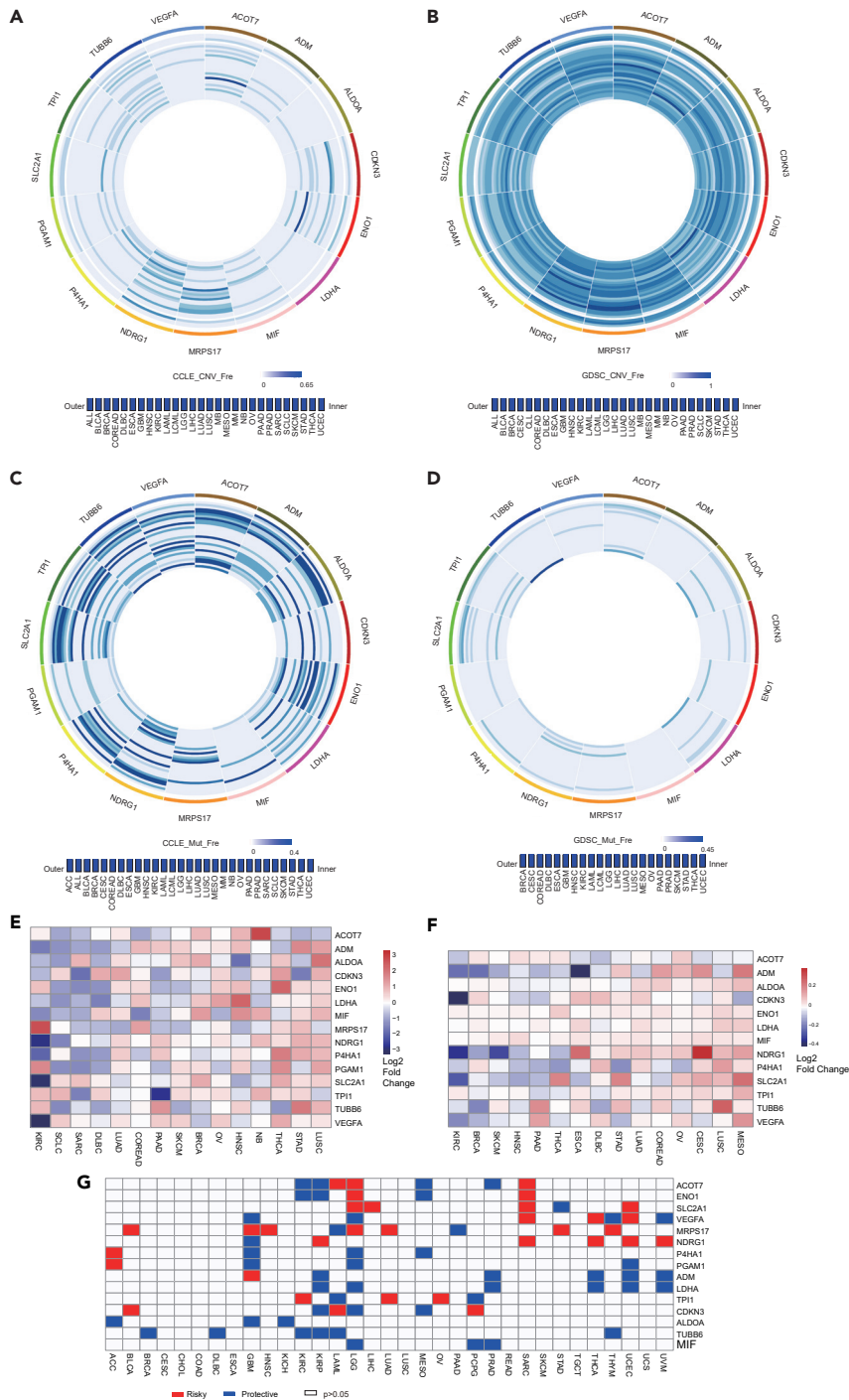
**Figure 2. Genomic Alterations of Hypoxia-Associated Signature in Different Cancer Cell Lines**

(A–G) CNV frequency of the 15 genes in cell lines of CCLE (A) and GDSC (B). Mutation frequency of the 15 genes in cell lines of CCLE (C) and GDSC (D). Log2 fold change of genes' expression in metastatic cell lines versus primary cell lines in CCLE (E) and GDSC (F). Heatmap of the correlation between CNV of the signature and patient's OS. Red shows that a higher copy number was related to a worse OS, and blue shows a link with a better OS (G).
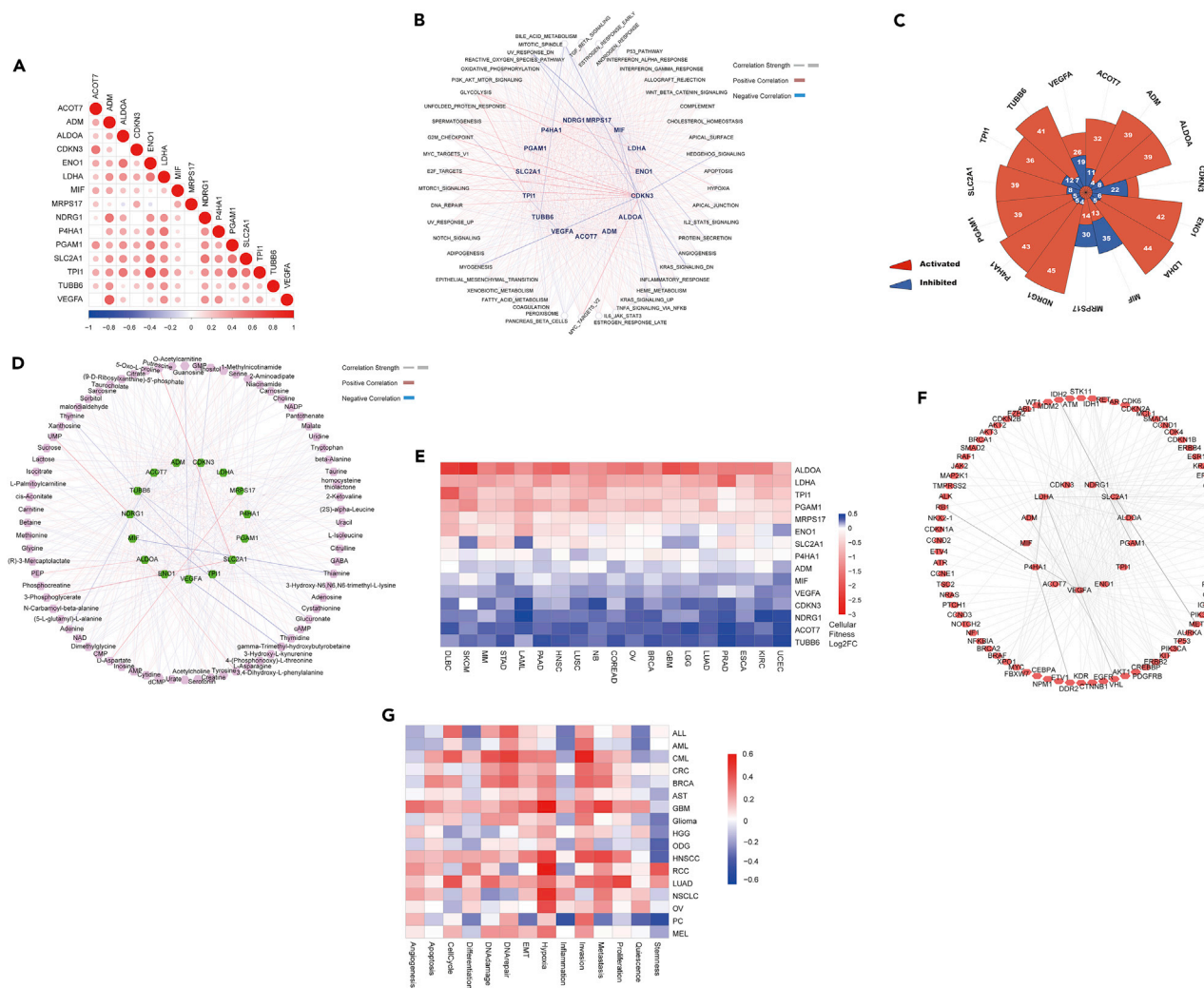
**Figure 3. Hypoxia-Associated Signature and Cancer Pathways**

(A) RNA correlation plots of the 15 hypoxia-associated genes.

(B) Network of cancer hallmarks and the hypoxia-associated signature.

(C) Number of activated and inhibited cancer hallmarks of each gene. The red sector represents activated hallmarks, and the blue sector represents inhibited hallmarks.

(D) Network of metabolites and the hypoxia-associated signature.

(E) Log2 fold change of cellular fitness after knocking out each gene in cells.

(F) Correlation between clinically actionable genes and hypoxia-associated genes.

(G) Average correlations between the gene of interest and functional states in different cancers.

Hypoxia promotes cancer malignancy, so the 15-gene hypoxia signature is supposed to contribute to cancer cells' progression. We adopted cellular fitness analysis, which could predict the functions of genes in various biological processes (Kurata and Lin, 2018; Perez et al., 2017), to validate the roles of the 15 genes in normoxic cancer cells. Cell fitness data from Sanger genome-wide CRISPR essentialities were used in our study to evaluate the function of these genes in cells. These essentiality profiles provide a measure of the loss or gain of cellular fitness resulting from knocking out a gene from targeted disruption via a single guide RNA. There was a dramatic decrease in cellular fitness after knocking out ALDOA, LDHA, TPI1, and PGAM1 in most cancer types, whereas only a slight increase in cellular fitness after knocking out TUBB6, ACOT7, NDRG1, CDKN3, VEGFA, and MIF (Figure 3E). These data suggested that the 15-gene signature mainly acts as cancer promoters. In addition, clinically actionable genes were collected and downloaded from a previously published article (Van Allen et al., 2014) to study the potential clinical and medical use of the 15-gene hypoxia signature. Interactions between these genes with the hypoxia-associated signature
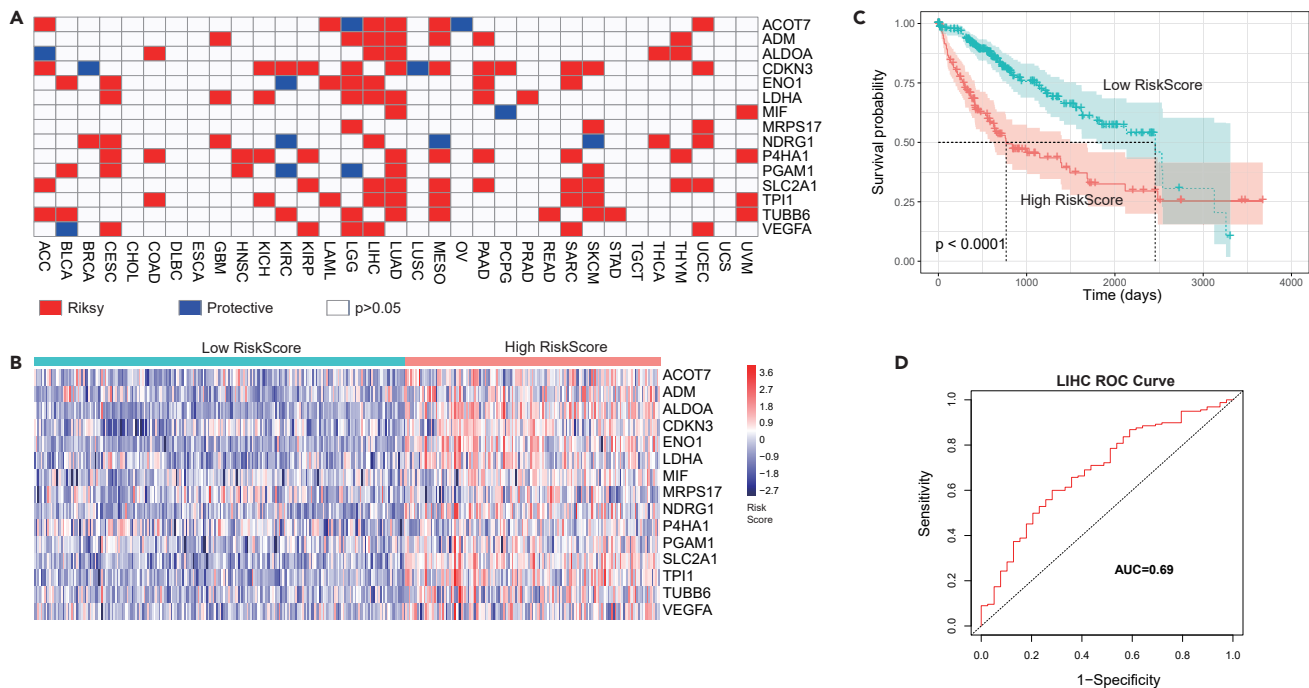
**Figure 4. Prognostic Role of the Hypoxia-Associated Signature in Cancers**

(A) Heatmap of the relationship between the expression of the signature and OS. Red shows that a higher expression was related to a worse OS, and blue shows a link with a better OS.

(B) Risk score value distribution in patients with LIHC.

(C) Survival plot of patients with LIHC according to the risk score.

(D) Receiver operating chanracteristic curve of risk score prediction model.

were evaluated, and 80 of 121 genes were found to be associated with the signature (Figure 3F), which indicated that the hypoxia signature might interact with these genes and be used as drug targets. Moreover, correlations between the gene list and functional states in different cancers were analyzed. As shown in Figure 3G, most of the genes were positively related to malignant behaviors, especially hypoxia, invasion, and EMT.

## Prognostic Role of the Hypoxia-Associated Signature in Cancers

The hypoxia-associated signature was playing important roles in cancers based on the aforementioned findings. Therefore we speculated that the signature would also affect clinical outcomes. Cox analysis was adopted to evaluate the effect of the 15 genes on overall survival (OS) of patients. The results demonstrated that the majority of these genes were risk factors for OS in various cancer types, especially in liver hepatocellular carcinoma (LIHC) and lung adenocarcinoma (LUAD) (Figure 4A). Therefore, LIHC cases, as an example, were divided into two groups according to the median risk score (Risk Score = $\sum_{i=1}^{n} Coef(i) * x(i)$). Based on the risk score pattern of the signature, two subgroups of patients with LIHC cancer were identified (Figure 4B). As shown in Figure 4C, cases of a low risk score had a much better OS than those of a high risk score. The data demonstrated that the signature is powerful to predict OS at least for LIHC. Time-dependent receiver operating characteristic analysis was then used to evaluate the accuracy of the signature-based prognostic prediction model in LIHC. Area under the curve = 0.69 showed that the model had excellent predictive value (Figure 4D). The role of P4HA1 in the prognosis of cancers was also analyzed, and it indicated that P4HA1 was a risk factor in more than one-third of all cancers (Figure S1).

## DISCUSSION

In our study, we comprehensively analyzed multi-omics data of hypoxia-associated signature. Metabolomics, transcriptomics, epigenetics, and proteomics data are integrated to investigate the role of the hypoxia-related signature in cancers in an extensive manner. Multi-omics data enable us to predict novel functional interactions between molecular mediators at multiple levels. Also, these data have the potential

to uncover crucial biological observations into hallmarks and pathways that would otherwise not be obvious through single-omics studies. Hypoxia has been paid wide attention recently. Ye et al. classified tumor samples into hypoxia score-high and score-low groups based on the same hypoxia signature and identified molecular alterations that correlated with drug responses to anticancer drugs (Ye et al., 2019). Also, Bhandari et al. quantified hypoxia in various cancer types and found elevated hypoxia associated with increased mutational load (Bhandari et al., 2020), as well as observed widespread hypoxia-associated dysregulation of microRNAs (Bhandari et al., 2019). However, molecular description of the 15-gene hypoxia signature in cancer tissues and cell lines is still lacking. Our study performed a comprehensive analysis of the hypoxia signature, aiming to provide a useful resource for future related researches.

CNVs are nearly ubiquitous in cancer (Heitzer et al., 2016; Zack et al., 2013) and have a great effect on cancer genome as a key type of genomic variations (Heitzer et al., 2016). Genes' CNV landscape is varied across different cancer types, and specific CNVs are related to cancer outcomes (Nibourel et al., 2017). In our study, we found that NDRG1 and MRPS17 are characterized by high gain frequency across the 33 cancer types. Further analysis showed that the copy number gains of NDRG1 and MRPS17 are prone to be risk factors in cancers (Figure 2G). Therefore, CNVs of the two genes might be used to predict prognosis in some cancers.

Not up to expectations, the general mutation frequency of these genes is quite low with a maximum frequency of 0.25, indicating that mutation is not the main alteration of these genes. Methylation is one of the common epigenetic mechanisms that involved in cell proliferation, autophagy, differentiation, and cell cycle in cancer (Pan et al., 2018). We also looked at the methylation level of the 15 genes across all the 33 cancer types. The methylation level of the signature is similar in 30 of the 33 cancer types, demonstrating the homogeneity of the signature in cancers. DNA methylation in the gene body is supposed to increase gene expression, whereas promoters' methylation is negatively related to gene expression (Yang et al., 2014). The negative correlation between RNA expression and methylation status for the majority of the genes reveals that promoter methylation is common for these genes, and methylation inhibitors might be used based on the presumed therapeutic target, particularly in the case of NDRG1, TUBB6, and ACOT7.

Data from TCGA and GEO showed alike RNA profiling in normal tissue and cancer tissue, proving the reliability of our analysis. Also, RNA data from cell lines invalidated the metastasis promoter roles for most of these genes. These data are consistent with previous studies (Bhandari et al., 2019, 2020; Ye et al., 2019) and further proved the roles of hypoxia in carcinogenesis and cancer development process. Interestingly, the RNA expression patterns of COAD are much the same as those of READ, indicating the general histological homogeneity of colon and rectum. On the other hand, the patterns are almost opposite between KIRC and KIRP, both of which are from renal tissues. It illustrates the discrepancy of clear cell cancer and papillary cancer of the kidney.

The majority of these 15 genes are upregulated in cancer tissues compared with normal tissue, implying that these genes could be positively correlated with each other. Data from RNA correlation analysis consistently validated the hypothesis. In particular, TPI1 and ENO1 are the most closely related. Both TPI1 and ENO1 are glycolytic enzymes, and studies have reported that TPI1 and ENO1 are overexpressed in various types of cancers, along with other glycolytic enzyme members like ALDOA and LDHA (Fedorova et al., 2019; Hamaguchi et al., 2008; Migneco et al., 2010). TPI1 and ENO1 would also be functionally correlated. Therefore, future studies about hypoxia or glycolysis could incorporate TPI1 and ENO1 into existing knowledge. Moreover, GSVA analysis also found that these 15 genes are more likely to be oncogenes as the number of activated pathways is much larger than that of inhibited pathways. Furthermore, a lot of important cancer hallmarks, such as apoptosis, p53 pathway, and transforming growth factor-beta signaling, interact with the signature. It is not surprising that the signature promotes cancer, but we listed the activated pathways in our results aiming to provide some useful clues for future studies. To our expectation, hypoxia is significantly related to the majority of these 15 genes. Of interest, there are some overlaps between the signature-related cancer hallmarks and metabolomics pathways among which are glycolysis and BA biosynthesis. Aerobic glycolysis has been closely linked with tumorigenesis for a long time ever since the "Warburg effect" was proposed (Warburg, 1956). Yet, the underlying mechanistic details regarding the causes and effect of the metabolic phenotype remained unclear (Ganapathy-Kanniappan and Geschwind, 2013). Our study might provide some clues to subsequent studies. BA biosynthesis has been serving crucial roles in multiple cancers and a large number of pathways linking BAs to cancer have been identified including oxidative stress with DNA damage and genomic instability, apoptosis, epigenetic factors, activation of nuclear receptors and metabolic and cellular homeostasis, and interactions with and changes of gut microbiota (Di Ciaula et al., 2017). Studies have showed that BA could be linked to glycolysis. BA

interruption by BA sequestrants is able to affect Carbohydrate-response element-binding protein (ChREBP) transactivity, subsequently influencing aerobic glycolysis through p53 inhibition (Iizuka, 2017). Also, BAs could potently modulate microbiome (Tian et al., 2020), and microbial communities provide molecules and metabolites, which are essential for glycolysis (Belizario et al., 2018). It has been suggested that hypoxia could repress BA synthesis through inhibition of CYP7A1 expression (Moon et al., 2016). Moreover, hypoxia could negatively regulate BA homeostasis by regulating the biosynthesis, conjugation, secretion, and absorption of BAs via Farnesoid X receptor (FXR) (Fujino et al., 2009). Our results were consistent with those of the aforementioned studies that most of the genes from the 15-gene signature were negatively associated with BA metabolism pathway. Tyrosine, a crucial component of tyrosine metabolism and biopterin metabolism pathways, is widely involved in multiple cancer pathophysiologies. We also found that some genes of the signature were correlated with tyrosine metabolism. Previous studies showed that hypoxia could decrease tyrosine by directly or indirectly upregulating tyrosine hydroxylase (Kumar et al., 2003; Lim et al., 2015). Therefore, tyrosine might be used to prevent hypoxia-induced effects. In fact, it has been proved that tyrosine administration counters hypoxia-induced decrements in learning and memory (Shukitt-Hale et al., 1996). Future researches are warranted to develop drugs targeted on tyrosine to treat cancers.

Cellular fitness has been applied in many fields to uncover new oncogenes (Kurata and Lin, 2018) and study molecular structures (Perez et al., 2017). In our study, cellular fitness was used to measure the potential function of these genes in cells. The results also indicate that most of these genes are essential to maintain normal cancer cellular fitness. It is worth noting that the cellular fitness here was performed in cancer cells, which were cultured under normoxia. Multiple hypoxia-mediated gene regulatory networks will be initiated and produced under lasting or temporary hypoxic situations and finally lead to significant change of cellular functions and behaviors (Chen et al., 2020). Subsequently, irreversible processes develop, which could result in pathophysiological disorder consequences. Therefore, the cells that are cultured in normoxia cannot completely reflect the the functions of the signature in hypoxia. However, these data could at least provide and validate some information about roles of the 15 genes. Also, cells from CCLE and GDSC were cultured under normoxia, therefore data from these cells are not closely related to hypoxia in primary cancers. However, we aimed to provide a molecular landscape of the signature in cancer cells, which could aid future researchers to select suitable hypoxia cell models.

A previous study has found that hypoxia status could affect most clinically actionable genes and experimentally validated the predicted effects of hypoxia on the response to several drugs in cultured cells (Ye et al., 2019). In our study, we further evaluated the correlation of the 15-gene signature and 121 clinically actionable genes in a direct way and also showed that the hypoxia signature is closely related to the clinically actionable genes. Both the studies warrant that hypoxia-targeted therapy is a promising cancer treatment, probably as a component of combination therapy targeting clinically actionable genes.

Prognosis prediction of cancers is crucial, as it can aid the subsequent clinical management of patients. Despite the fact that remarkable improvement has been achieved in cancer research globally over the past several decades, it is still a big challenge to predict patients' prognosis. We also attempted to evaluate the role of the signature in cancer prognosis. The findings demonstrate that the genes are risk factors in about 65% gene-cancer pairs, which accords with the cancer promoter role of these genes in cancers. The risk score based on expression and hazard ratio could greatly separate LIHC cases, and the risk score is powerful to predict the prognosis of LIHC. This evidence warrants further study of hypoxia-associated signature in cancers. It could be also used as an effective prognosis prediction model.

There may be some possible limitations in this study. Data from multiple sources are difficult to be processed and analyzed under a unifying framework given the diversity of methodological work in this area (Wu et al., 2019). It may lead to minor inconsistencies between our study and others'. Besides, some data are collected from retrospective studies in which some critical parameters are not recorded, and remarkable biases could influence selection of controls. Therefore, further prospective clinical studies are needed to validate our findings.

In conclusion, we systematically analyzed the landscape of genetic, epigenetic, and metabolomics alterations and biological and clinical relevance of hypoxia-associated signature in cancers. It helps to better understand the dysregulation of hypoxia in cancer. The findings from our study can be readily applied to further studies and are highly promising for personalized treatment regimens and clinical management.

## Limitations of the Study

Cells that were used to analyze the profile of the hypoxia signature were cultured in normoxia. This would not perfectly reflect the situation of cells that are cultured in hypoxia.

## Resource Availability

### Lead Contact

Qian Zhang, zhangqian1171@126.com.

### Materials Availability

The study did not generate new unique reagents.

### Data and Code Availability

The datasets and code used and/or analyzed during the current study are available from the corresponding author on reasonable request.

## METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

## SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2020.101460.

## AUTHOR CONTRIBUTIONS

Q.Z. and G.W. designed the work. Q.Z. and R.H. analyzed the data. Q.T. interpreted the patient data. L.Y. and Y.T. performed the acquisition of data. Q.Z. and Z.L. drafted the work. J.L. was a major contributor in writing the manuscript. All authors read and approved the final manuscript.

## DECLARATION OF INTERESTS

The authors declare that they have no competing interests.

Ethical approval and consent to participate: All data were downloaded from public databases, and the study was approved by the IRB of the Second Affiliated Hospital of Harbin Medical University.

## REFERENCES

Belizario, J.E., Faintuch, J., and Garay-Malpartida, M. (2018). Gut microbiome dysbiosis and immunometabolism: new frontiers for treatment of metabolic diseases. Mediators Inflamm. 2018, 2037838.

Bhandari, V., Hoey, C., Liu, L.Y., Lalonde, E., Ray, J., Livingstone, J., Lesurf, R., Shiah, Y.J., Vujcic, T., Huang, X., et al. (2019). Molecular landmarks of tumor hypoxia across cancer types. Nat. Genet. 51, 308–318.

Bhandari, V., Li, C.H., Bristow, R.G., Boutros, P.C., and Consortium, P. (2020). Divergent mutational processes distinguish hypoxic and normoxic tumours. Nat. Commun. 11, 737.

Buffa, F.M., Harris, A.L., West, C.M., and Miller, C.J. (2010). Large meta-analysis of multiple cancers reveals a common, compact and highly prognostic hypoxia metagene. Br. J. Cancer 102, 428–435.

Carmeliet, P., and Jain, R.K. (2011). Molecular mechanisms and clinical applications of angiogenesis. Nature 473, 298–307.

Chen, P.S., Chiu, W.T., Hsu, P.L., Lin, S.C., Peng, I.C., Wang, C.Y., and Tsai, S.J. (2020). Pathophysiological implications of hypoxia in human diseases. J. Biomed. Sci. 27, 63.

Di Ciaula, A., Wang, D.Q., Molina-Molina, E., Lunardi Baccetto, R., Calamita, G., Palmieri, V.O., and Portincasa, P. (2017). Bile acids and cancer: direct and environmental-dependent effects. Ann. Hepatol. 16 (Suppl 1), S87–S105.

Fedorova, M.S., Krasnov, G.S., Lukyanova, E.N., Zaretsky, A.R., Dmitriev, A.A., Melnikova, N.V., Moskalev, A.A., Kharitonov, S.L., Pudova, E.A.,

Guvatova, Z.G., et al. (2019). The CIMP-high phenotype is associated with energy metabolism alterations in colon adenocarcinoma. BMC Med. Genet. *20*, 52.

Foster, J.G., Wong, S.C., and Sharp, T.V. (2014). The hypoxic tumor microenvironment: driving the tumorigenesis of non-small-cell lung cancer. Future Oncol. *10*, 2659–2674.

Fujino, T., Murakami, K., Ozawa, I., Minegishi, Y., Kashimura, R., Akita, T., Saitou, S., Atsumi, T., Sato, T., Ando, K., et al. (2009). Hypoxia downregulates farnesoid X receptor via a hypoxia-inducible factor-independent but p38 mitogen-activated protein kinase-dependent pathway. FEBS J. *276*, 1319–1332.

Ganapathy-Kanniappan, S., and Geschwind, J.F. (2013). Tumor glycolysis as a target for cancer therapy: progress and prospects. Mol. Cancer *12*, 152.

Giaccia, A.J., and Schipani, E. (2010). Role of carcinoma-associated fibroblasts and hypoxia in tumor progression. Curr. Top. Microbiol. Immunol. *345*, 31–45.

Haibe-Kains, B., El-Hachem, N., Birkbak, N.J., Jin, A.C., Beck, A.H., Aerts, H.J., and Quackenbush, J. (2013). Inconsistency in large pharmacogenomic studies. Nature *504*, 389–393.

Hamaguchi, T., Iizuka, N., Tsunedomi, R., Hamamoto, Y., Miyamoto, T., Iida, M., Tokuhisa, Y., Sakamoto, K., Takashima, M., Tamesa, T., et al. (2008). Glycolysis module activated by hypoxia-inducible factor 1alpha is related to the aggressive phenotype of hepatocellular carcinoma. Int. J. Oncol. *33*, 725–731.

Heitzer, E., Ulz, P., Geigl, J.B., and Speicher, M.R. (2016). Non-invasive detection of genome-wide somatic copy number alterations by liquid biopsies. Mol. Oncol. *10*, 494–502.

Iizuka, K. (2017). The transcription factor carbohydrate-response element-binding protein (ChREBP): a possible link between metabolic disease and cancer. Biochim. Biophys. Acta Mol. Basis Dis. *1863*, 474–485.

Kumar, G.K., Kim, D.K., Lee, M.S., Ramachandran, R., and Prabhakar, N.R. (2003). Activation of tyrosine hydroxylase by intermittent hypoxia: involvement of serine phosphorylation. J. Appl. Physiol. *95*, 536–544.

Kurata, J.S., and Lin, R.J. (2018). MicroRNA-focused CRISPR-Cas9 library screen reveals fitness-associated miRNAs. RNA *24*, 966–981.

Li, Y., Xiao, J., Bai, J., Tian, Y., Qu, Y., Chen, X., Wang, Q., Li, X., Zhang, Y., and Xu, J. (2019). Molecular characterization and clinical relevance of m(6)A regulators across 33 cancer types. Mol. Cancer *18*, 137.

Li, Z., Bao, S., Wu, Q., Wang, H., Eyler, C., Sathornsumetee, S., Shi, Q., Cao, Y., Lathia, J., McLendon, R.E., et al. (2009). Hypoxia-inducible factors regulate tumorigenic capacity of glioma stem cells. Cancer Cell *15*, 501–513.

Lim, J., Kim, H.I., Bang, Y., Seol, W., Choi, H.S., and Choi, H.J. (2015). Hypoxia-inducible factor-1alpha upregulates tyrosine hydroxylase and dopamine transporter by nuclear receptor ERRgamma in SH-SY5Y cells. Neuroreport *26*, 380–386.

Migneco, G., Whitaker-Menezes, D., Chiavarina, B., Castello-Cros, R., Pavlides, S., Pestell, R.G., Fatatis, A., Flomenberg, N., Tsirigos, A., Howell, A., et al. (2010). Glycolytic cancer associated fibroblasts promote breast cancer tumor growth, without a measurable increase in angiogenesis: evidence for stromal-epithelial metabolic coupling. Cell Cycle *9*, 2412–2422.

Mittal, V. (2018). Epithelial mesenchymal transition in tumor metastasis. Annu. Rev. Pathol. *13*, 395–412.

Moon, Y., Park, B., and Park, H. (2016). Hypoxic repression of CYP7A1 through a HIF-1alpha- and SHP-independent mechanism. BMB Rep. *49*, 173–178.

Muz, B., de la Puente, P., Azab, F., and Azab, A.K. (2015). The role of hypoxia in cancer progression, angiogenesis, metastasis, and resistance to therapy. Hypoxia (Auckl) *3*, 83–92.

Nibourel, O., Guihard, S., Roumier, C., Pottier, N., Terre, C., Paquet, A., Peyrouze, P., Geffroy, S., Quentin, S., Alberdi, A., et al. (2017). Copy-number analysis identified new prognostic marker in acute myeloid leukemia. Leukemia *31*, 555–564.

Pan, Y., Liu, G., Zhou, F., Su, B., and Li, Y. (2018). DNA methylation profiles in cancer diagnosis and therapeutics. Clin. Exp. Med. *18*, 1–14.

Perez, A.M., Gomez, M.M., Kalvapalle, P., O'Brien-Gilbert, E., Bennett, M.R., and Shamoo, Y. (2017). Using cellular fitness to map the structure and function of a major facilitator superfamily effluxer. Mol. Syst. Biol. *13*, 964.

Rankin, E.B., and Giaccia, A.J. (2016). Hypoxic control of metastasis. Science *352*, 175–180.

Shukitt-Hale, B., Stillman, M.J., and Lieberman, H.R. (1996). Tyrosine administration prevents hypoxia-induced decrements in learning and memory. Physiol. Behav. *59*, 867–871.

Siegel, R.L., Miller, K.D., and Jemal, A. (2019). Cancer statistics, 2019. CA Cancer J. Clin. *69*, 7–34.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U S A *102*, 15545–15550.

Tian, Y., Gui, W., Koo, I., Smith, P.B., Allman, E.L., Nichols, R.G., Rimal, B., Cai, J., Liu, Q., and Patterson, A.D. (2020). The microbiome modulating activity of bile acids. Gut Microbes *11*, 979–996.

Van Allen, E.M., Wagle, N., Stojanov, P., Perrin, D.L., Cibulskis, K., Marlow, S., Jane-Valbuena, J., Friedrich, D.C., Kryukov, G., Carter, S.L., et al. (2014). Whole-exome sequencing and clinical interpretation of formalin-fixed, paraffin-embedded tumor samples to guide precision cancer medicine. Nat. Med. *20*, 682–688.

Vaupel, P., Hockel, M., and Mayer, A. (2007). Detection and characterization of tumor hypoxia using pO2 histography. Antioxid. Redox Signal. *9*, 1221–1235.

Warburg, O. (1956). On the origin of cancer cells. Science *123*, 309–314.

Wu, C., Zhou, F., Ren, J., Li, X., Jiang, Y., and Ma, S. (2019). A selective review of multi-level omics data integration using variable selection. High Throughput *8*, https://doi.org/10.3390/ht8010004.

Yang, X., Han, H., De Carvalho, D.D., Lay, F.D., Jones, P.A., and Liang, G. (2014). Gene body methylation can alter gene expression and is a therapeutic target in cancer. Cancer Cell *26*, 577–590.

Ye, Y., Hu, Q., Chen, H., Liang, K., Yuan, Y., Xiang, Y., Ruan, H., Zhang, Z., Song, A., Zhang, H., et al. (2019). Characterization of hypoxia-associated molecular features to aid hypoxia-targeted therapy. Nat. Metab. *1*, 431–444.

Zack, T.I., Schumacher, S.E., Carter, S.L., Cherniack, A.D., Saksena, G., Tabak, B., Lawrence, M.S., Zhsng, C.Z., Wala, J., Mermel, C.H., et al. (2013). Pan-cancer patterns of somatic copy number alteration. Nat. Genet. *45*, 1134–1140.

**Supplemental Information**

**Integrative Analysis of Hypoxia-Associated**

**Signature in Pan-Cancer**

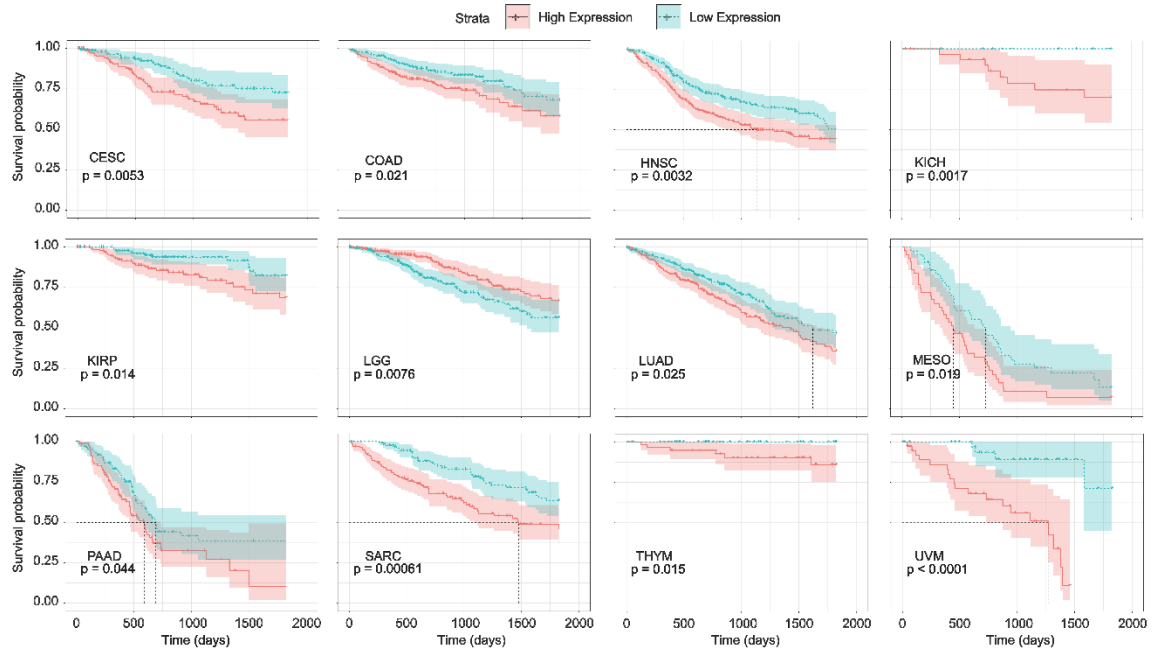Qian Zhang, Rui Huang, Hanqing Hu, Lei Yu, Qingchao Tang, Yangbao Tao, Zheng Liu, Jiaying Li, and Guiyu Wang

**Fig.S1 The role of P4HA1 in 5-year overall survival in cancers Related to Figure 1.** Cases were divided into High Expresison group and Low Expression groups depending on median P4HA1 RNA expression in each cancer type.

**Table S2 Metabolites and pathways in cancer cell lines, Related to Figure 3**

| Metabolites | Pathways |
| --- | --- |
| Taurocholate | Bile acid biosynthesis |
| Tyrosine | Tyrosine metabolism\|Biopterin metabolism |
| Sorbitol | Galactose metabolism\|Fructose and mannose metabolism |
| Lactose | Galactose metabolism |
| Sucrose | Galactose metabolism |
| Acetylcholine | Glycerophospholipid metabolism |
| Choline | Glycerophospholipid metabolism\|Glycine, serine, alanine and threonine metabolism |
| Serine | Methionine and cysteine metabolism\|Vitamin B9 (folate) metabolism\|Glycerophospholipid metabolism\|Glycosphingolipid metabolism\|Glycine, serine, alanine and threonine metabolism |
| Dimethylglycine | Glycine, serine, alanine and threonine metabolism |
| (5-L-glutamyl)-L-alanine | Glycine, serine, alanine and threonine metabolism |
| Phosphocreatine | Glycine, serine, alanine and threonine metabolism |
| Betaine | Glycine, serine, alanine and threonine metabolism |
| Sarcosine | Glycine, serine, alanine and threonine metabolism |
| 4-(Phosphonooxy)-L-threonine | Glycine, serine, alanine and threonine metabolism |
| Creatine | Glycine, serine, alanine and threonine metabolism |
| 3-Phosphoglycerate | Glycine, serine, alanine and threonine metabolism\|Glycolysis and Gluconeogenesis |
| PEP | TCA cycle\|Glycolysis and Gluconeogenesis |
| Malate | TCA cycle\|Glycolysis and Gluconeogenesis |
| Carnosine | Histidine metabolism |
| 2-Aminoadipate 6-semialdehyde | Lysine metabolism |
| 3-Hydroxy-N6,N6,N6-trimethyl-L-lysine | Lysine metabolism |
| Carnitine | Lysine metabolism\|Urea cycle and metabolism of arginine, proline, glutamate, aspartate and asparagine |
| gamma-Trimethyl-hydroxybutyrobetaine | Lysine metabolism\|Urea cycle and metabolism of arginine, proline, glutamate, aspartate and asparagine |
| (R)-3-Mercaptolactate | Methionine and cysteine metabolism |
| homocysteine thiolactone | Methionine and cysteine metabolism |
| Cystathionine | Methionine and cysteine metabolism |
| Taurine | Bile acid biosynthesis\|Methionine and cysteine metabolism |
| Methionine | Methionine and cysteine metabolism\|Glycine, serine, alanine and threonine metabolism |
| Inositol | Phosphatidylinositol phosphate metabolism |
| Glucuronate | Phosphatidylinositol phosphate metabolism |
| malondialdehyde | Prostaglandin formation from arachidonate |
| Urate | Purine metabolism |
| Inosine | Purine metabolism |
| AMP | Purine metabolism |
| Adenine | Purine metabolism |
| Xanthosine | Purine metabolism |
| (9-D-Ribosylxanthine)-5'-phosphate | Purine metabolism |
| Guanosine | Purine metabolism |
| GMP | Purine metabolism |
| Adenosine | Purine metabolism |

| | |
|---|---|
| cAMP | Purine metabolism |
| dCMP | Pyrimidine metabolism |
| Cytidine | Pyrimidine metabolism |
| N-Carbamoyl-beta-alanine | Pyrimidine metabolism |
| UMP | Pyrimidine metabolism |
| Thymine | Pyrimidine metabolism |
| Uridine | Pyrimidine metabolism |
| Uracil | Pyrimidine metabolism |
| Thymidine | Pyrimidine metabolism |
| CMP | Pyrimidine metabolism\|Glycosphingolipid biosynthesis - ganglioseries |
| L-Palmitoylcarnitine | Saturated fatty acids beta-oxidation |
| cis-Aconitate | TCA cycle |
| Isocitrate | TCA cycle |
| Citrate | TCA cycle |
| Tryptophan | Tryptophan metabolism |
| 3-Hydroxy-L-kynurenine | Tryptophan metabolism |
| Serotonin | Tryptophan metabolism |
| 3,4-Dihydroxy-L-phenylalanine | Tyrosine metabolism |
| D-Aspartate | Urea cycle and metabolism of arginine, proline, glutamate, aspartate and asparagine |
| 5-Oxo-L-proline | Urea cycle and metabolism of arginine, proline, glutamate, aspartate and asparagine |
| Putrescine | Urea cycle and metabolism of arginine, proline, glutamate, aspartate and asparagine |
| O-Acetylcarnitine | Urea cycle and metabolism of arginine, proline, glutamate, aspartate and asparagine |
| Citrulline | Urea cycle and metabolism of arginine, proline, glutamate, aspartate and asparagine |
| GABA | Urea cycle and metabolism of arginine, proline, glutamate, aspartate and asparagine |
| L-Asparagine | Urea cycle and metabolism of arginine, proline, glutamate, aspartate and asparagine |
| beta-Alanine | Histidine metabolism\|Urea cycle and metabolism of arginine, proline, glutamate, aspartate and asparagine\|Pyrimidine metabolism |
| 2-Ketovaline | Valine, leucine and isoleucine degradation |
| (2S)-alpha-Leucine | Valine, leucine and isoleucine degradation |
| L-Isoleucine | Valine, leucine and isoleucine degradation |
| Thiamine | Vitamin B1 (thiamin) metabolism |
| NAD | Vitamin B3 (nicotinate and nicotinamide) metabolism |
| 1-Methylnicotinamide | Vitamin B3 (nicotinate and nicotinamide) metabolism |
| Niacinamide | Vitamin B3 (nicotinate and nicotinamide) metabolism |
| NADP | Vitamin B3 (nicotinate and nicotinamide) metabolism |
| Pantothenate | Vitamin B5 - CoA biosynthesis from pantothenate |
| Glycine | Urea cycle and metabolism of arginine, proline, glutamate, aspartate and asparagine\|Bile acid biosynthesis\|Porphyrin metabolism\|Leukotriene metabolism\|Glycine, serine, alanine and threonine metabolism\|Vitamin B9 (folate) metabolism\|Lysine metabolism |

**Transparent Methods**

**The Cancer Genome Atlas (TCGA) data acquisition and analysis**

Gene expression, methylation, CNV and clinical data of 33 different types of cancer were downloaded from Broad GDAC Firehose (https://gdac.broadinstitute.org/). For genes with multiple methylation probes (usually from the Infinium arrays), we only include methylation data from the probe with the strongest negative correlation between the methylation signal and the gene's expression. The beta value was used to denote the methylation level. The cancer types and tissue information were as follows: ACC, Adrenocortical carcinoma; BLCA, Bladder Urothelial Carcinoma; BRCA, Breast invasive carcinoma; CESC, Cervical squamous cell carcinoma, and endocervical adenocarcinoma; CHOL, Cholangiocarcinoma; COAD, Colon adenocarcinoma; DLBC, Lymphoid Neoplasm Diffuse Large B-cell Lymphoma; ESCA, Esophageal carcinoma; GBM, Glioblastoma multiforme; HNSC, Head and Neck squamous cell carcinoma; KICH, Kidney Chromophobe; KIRC, Kidney renal clear cell carcinoma; KIRP, Kidney renal papillary cell carcinoma; LAML, Acute Myeloid Leukemia; LCML, Chronic Myelogenous Leukemia; LGG, Brain Lower Grade Glioma; LIHC, Liver hepatocellular carcinoma; LUAD, Lung adenocarcinoma; LUSC, Lung squamous cell carcinoma; MESO, Mesothelioma; OV, Ovarian serous cystadenocarcinoma; PAAD, Pancreatic adenocarcinoma; PCPG, Pheochromocytoma, and Paraganglioma; PRAD, Prostate adenocarcinoma; READ, Rectum adenocarcinoma; SARC, Sarcoma; SKCM, Skin Cutaneous Melanoma; STAD, Stomach adenocarcinoma; TGCT, Testicular Germ Cell Tumors; THCA, Thyroid carcinoma; THYM, Thymoma; UCEC, Uterine Corpus Endometrial Carcinoma; UCS, Uterine Carcinosarcoma; UVM, Uveal Melanoma.

Wilcox's rank-sum test was used to determine differentially expressed genes between cancer and normal tissues. Benjamini-Hochberg procedure was applied to adjust p values. Genes with adjusted p-values <0.05 and at least log2 fold change>1.5 in expression were considered to be differentially expressed in each cancer type. Linear models were also used to adjust for tumor purity, ploidy, age, sex and race. Methylation levels were also analyzed between cancer and normal tissues to identify potential methylation markers. Correlation between methylation level and the gene expression level was analyzed with Pearson Correlation analysis.

**Survival analysis of hypoxia-correlated signature**

To analyze the role of the hypoxia-related signature, survival analysis was performed. The log-rank test was used to examine whether CNV, methylation, and RNA expression were associated with overall survival (OS). The survival package in R program (https://cran.r-project.org/web/packages/survival/index.html) was adopted to perform survival-related analysis. The p-values <0.05 were considered as significant. The risk score of each case was calculated based on the formula: Risk Score=$\sum_{i=1}^{n} Coef(i) * x(i)$, where Coef(i) is the Hazard Ratio of each selected gene analyzed by Cox analysis, and x(i) is the FPKM (fragments per kilobase of exon model per million reads mapped) value of each selected gene (Chai et al., 2019).

**Multi-omics data for cell lines and cancers**

To comprehensively understand the effect of the 15-gene hypoxia signature, metabolite and genomic data of cancer cell lines were also retrieved and analyzed. Multi-omics data, including metabolomics, CNV, methylation, mutation and gene expression, were downloaded from the Broad Institute Cancer Cell Line Encyclopedia (CCLE) and the Genomics of Drug Sensitivity in Cancer database (GDSC) for cell lines. Cancer types of these cell lines were defined according to the annotation of the cells. The unique cancer types in CCLE and GDSC were as follows: ALL, Acute lymphocytic leukemia; CLL, Chronic lymphocytic leukemia; MB, Myeloid blasts; MM, Multiple myeloma; NB, Neuroblastoma; SCLC, Small cell lung cancer. We characterized the mutation and CNV frequency of each gene in each cancer type as the proportion of cell lines with mutations and CNV alterations respectively. Pearson Correlation was used to identify the gene expression and metabolites intensity in different cell lines. Metabolite-gene pairs with correlation coefficient more than 0.6 and adjusted p-value of less than 0.05 were plotted.

**Gene Expression Omnibus data acquisition and analysis**

Gene Expression Omnibus (GEO) was also used to get RNA expression data of the genes in the signature to further verify the molecular profiling in different cancer types. Expression data of about 22,000 samples standing for 19 tissues were acquired manually. To reduce the potential bias result from different platforms, only data generated from the Affymetrix Human Genome U133 (HG-U133) Plus 2.0 Array were applied (Li et al., 2019). Combat method was adopted to normalize and adjust the batch effect (Leek et al., 2012).

**Pathways analysis in different cancer types**

We also explored the potential pathways that are involved in the hypoxia signature. Gene Set Variation Analysis (GSVA) was used to characterize pathways from gene expression datasets (Hanzelmann et al., 2013). Pearson Correlation between the expression of genes of hypoxia-associated signatures and pathway activity was calculated to assess the relationship between the genes and pathways. The gene-pathway pairs with the absolute value of the correlation coefficient of more than 0.4 and adjusted p-value<0.05 were determined as correlated pairs.

**Immunohistochemistry (IHC)**

Standard methods were used for staining de-paraffinized and rehydrated tissue sections from the patient samples (4 normal tissues and 5 cancerous tissues from 9 patients). The histology types were defined by three pathologists independently. Anti-NDRG1 was from Wanleibio (WL03071). Each sample was assigned a score based on the proportion and the intensity of positive cells (For the propor-tion, 0≤10%, 1=10%-25%, 2=26%-50%, 3=51%-75%,4≥75%; For the intensity, 0=negative, 1=weak, 2=moderate, 3=strong).

**Correlation among hypoxia-associated genes**

Pearson Correlation method was adopted to assess the relationships among these 15 genes based on RNA expression. Besides, STRING (https://string-db.org/) database was used to further identify the protein-protein interactions (PPI) among the genes. PPI between these genes and 121 medically actionable genes which were downloaded from a previous study (https://software.broadinstitute. org/cancer/cga/target) (Van Allen et al., 2014) were visualized with Cytoscape.

**Roles of hypoxia-related signature in cell fitness**

Cell fitness data was used to evaluate the function of individual gene of the hypoxia signature in cancer cells. CRISPR KO Data were downloaded from the Cell Model Passports (https://cellmodelpassports.sanger.ac.uk/downloads). These essentiality profiles provide a measure of the

loss or gain of cellular fitness resulting from knocking-out a gene from targeted disruption via a single guide

RNA (sgRNA).