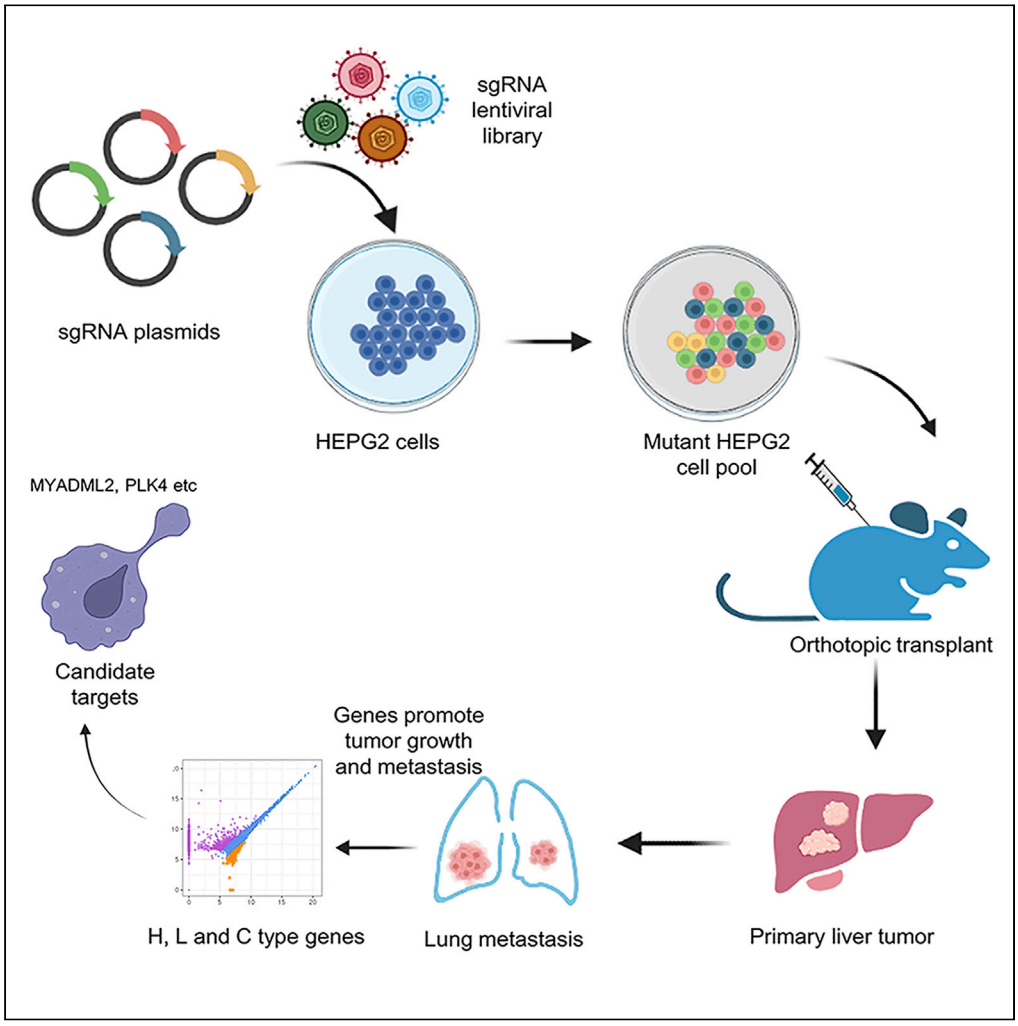
iScience

CellPress
OPEN ACCESS

Article

CRISPR activation screening in a mouse model for drivers of hepatocellular carcinoma growth and metastasis



Bei Zhang, Zhiyao Ren, Hongmei Zheng, Meilan Lin, Guobing Chen, Oscar Junhong Luo, Guodong Zhu

luojh@jnu.edu.cn (O.J.L.)
eyzhugd@scut.edu.cn (G.Z.)

Highlights

We develop CRISPRa screen for driver of HCC invasion and metastasis *in vivo*

The screen uncovered 4 novel modulator targets:XAGE1B, PLK4, LMO1 and MYADML2

High MYADML2 exhibited worse overall survival in HCC and reduced drug sensitivity

Zhang et al., iScience *26*, 106099
March 17, 2023
https://doi.org/10.1016/j.isci.2023.106099

iScience

CellPress
OPEN ACCESS

Article

CRISPR activation screening in a mouse model for drivers of hepatocellular carcinoma growth and metastasis

Bei Zhang,[1,2,8] Zhiyao Ren,[2,3,4,8] Hongmei Zheng,[5,8] Meilan Lin,[1] Guobing Chen,[6,7] Oscar Junhong Luo,[2,7,*] and Guodong Zhu[1,3,9,*]

SUMMARY

Hepatocellular carcinoma (HCC) remains a major cause of cancer-related mortality worldwide. Here we described a genome-wide screen by CRISPR activation (CRISPRa) library *in vivo* for drivers of HCC growth and metastasis. Pathological results showed the cell population formed highly metastatic tumors in lung after being mutagenized with CRISPRa. *In vitro* validation indicated overexpression of XAGE1B, PLK4, LMO1 and MYADML2 promoted cells proliferation and invasion, and the inhibition suppressed HCC progress. In addition, we reported high MYADML2 protein level exhibited worse overall survival in HCC, which increased significantly in patients over 60 years. Moreover, high MYADML2 reduced the sensitivity to chemotherapeutic drugs. Interestingly, immune cell infiltration analysis showed that the dendritic cells, macrophages, and so forth might play important role in HCC progress. In brief, we provides a roadmap for screening functional genes related to HCC invasion and metastasis *in vivo*, which may provide new potential targets for the treatment of HCC.

INTRODUCTION

Primary liver cancer is the sixth most prevalent malignancy and rank as the third leading cause of cancer-related deaths worldwide. Hepatocellular carcinoma (HCC) is most common primary liver cancer, which accounts for 90% of all cases, characterized by high mortality, recurrence, metastasis and poor prognosis.[1–3] Surgical intervention through resection or liver transplantation is the only option for cure, however, only 10-23% of patients are suitable for surgery at the time of presentation.[4–6] Unfortunately, most patients are diagnosed with advanced HCC and the 1-year survival is 15-39% due to limited treatment options.[7,8] It has been reported that the growth and metastasis of primary tumors were different yet linked processes in the development of solid tumors,[9,10] which might likely be driven by mutations of certain key genes. Knowledge of the biology and genetic basis behind tumor metastasis will assist in designing effective clinical intervention strategies in the effort of reducing the mortality of HCC.

Genetic screening is a powerful tool to analyze phenotypes and identify causal genes in various hallmarks of cancer progression. The clustered regularly interspaced short palindromic repeats (CRISPR)-associated protein 9 (Cas9) system has revolutionized the study of gene function and has had a huge impact on genetic therapy for human health.[11] Recently, most of these studies have exclusively considered loss-of-function screening through CRISPR KO library in eukaryotic cells[12,13] or mouse model.[14,15] CRISPR KO libraries were widely used to identify new biological mechanisms, such as drug resistance or cell survival signals. However, the method can only be used to examine the genes that have been expressed in the cellular system under study. CRISPR activation (CRISPRa) is a potent tool for selective transcriptional upregulation of endogenous genes and have been used to systematically identify regulators of neuronal-fate specification,[16] identify epigenetic regulators[17] and functionalize lncRNAs in drug resistance.[18] Moreover, CRISPRa is preferable to traditional overexpression techniques, such as cloned cDNA overexpression.[19]

Previously, CRISPR-Cas9 knockout library was used in HCC with lenvatinib resistance[20] and identified suppressors of liver tumor formation in mice.[21] Wang C, et al.[22] used a kinome-focused genetic screen and found CDC7 inhibition induces senescence selectively in liver cancer cells. An *in vivo* screening focused on a set of highly mutated pan-cancer tumor suppressors through AAV pools carrying a library of 278 sgRNAs.[23]

[1]Departments of Geriatrics and Oncology, Guangzhou First People's Hospital, School of Medicine, South China University of Technology, Guangzhou, Guangdong, China

[2]Department of Systems Biomedical Sciences, School of Medicine, Jinan University, Guangzhou, China

[3]Guangzhou Geriatric Hospital, Guangzhou, China

[4]Collaborative Innovation Center for Civil Affairs of Guangzhou, Guangzhou, China

[5]Department of Breast Surgery, Hubei Cancer Hospital, Tongji Medical College, Huazhong University of Science and Technology and Hubei Provincial Clinical Research Center for Breast Cancer, Wuhan, Hubei, China

[6]Guangdong-Hong Kong-Macau Great Bay Area Geroscience Joint Laboratory, Guangzhou, China

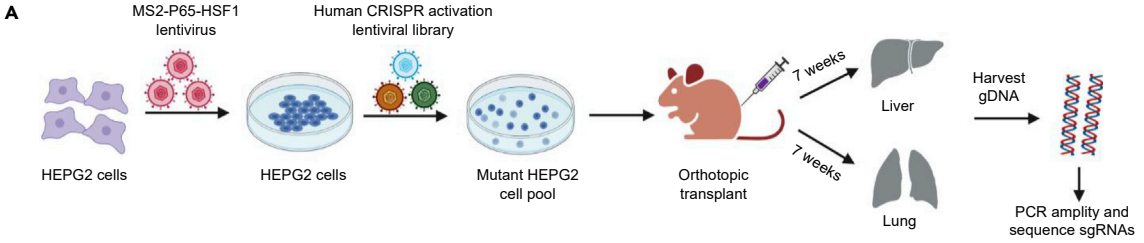
[7]Department of Microbiology and Immunology, Institute of Geriatric Immunology, School of Medicine, Jinan University, Guangzhou, China

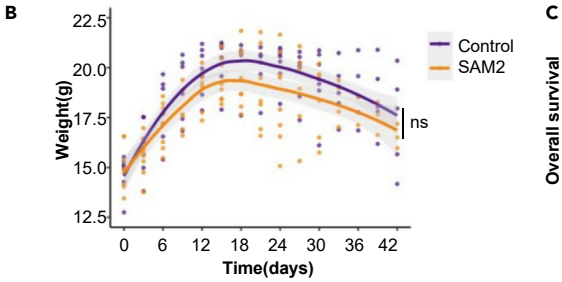
[8]These authors contributed equally

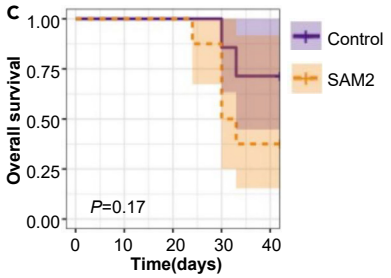
[9]Lead contact

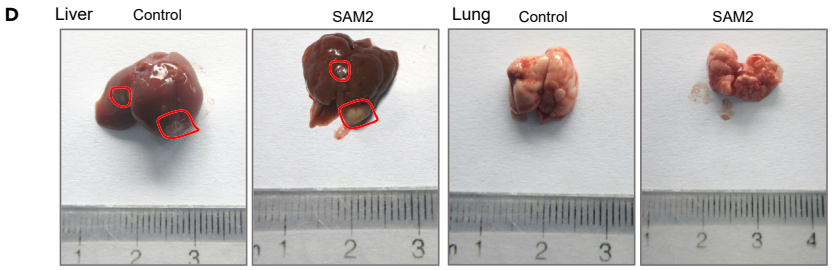
*Correspondence: luojh@jnu.edu.cn (O.J.L.), eyzhugd@scut.edu.cn (G.Z.)

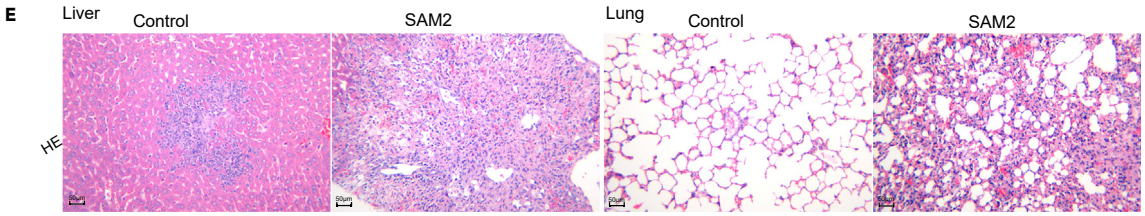
https://doi.org/10.1016/j.isci.2023.106099


A
MS2-P65-HSF1 lentivirus
Human CRISPR activation lentiviral library
HEPG2 cells
HEPG2 cells
Mutant HEPG2 cell pool
Orthotopic transplant
7 weeks
7 weeks
Liver
Lung
Harvest gDNA
PCR amplity and sequence sgRNAs


B
Weight(g)
Time(days)
Control
SAM2
ns


C
Overall survival
Time(days)
Control
SAM2
P=0.17


D
Liver
Control
SAM2
Lung
Control
SAM2


E
Liver
Control
SAM2
Lung
Control
SAM2
HE

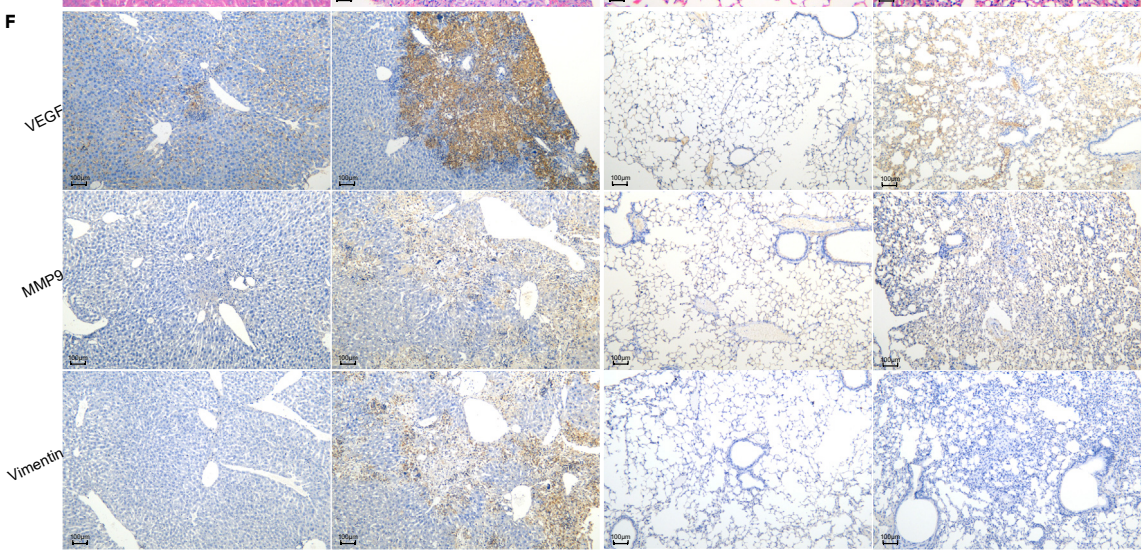

F
VEGF
MMP9
Vimentin

Figure 1. Tumor Growth and Metastasis in Transplanted HEPG2 Cells with SAM2 Library

(A) Schematic diagram for gain-of-function metastasis screening by the genome-scale CRISPR/Cas9 activation library (SAM2); (B-C) Body weight and survival curves of the mice untransduced or transduced with the SAM2 lentiviral library; (D) Visual observation of the liver and lung in mice untransduced or transduced with the SAM2 lentiviral library; (E) Representative H&E staining of primary liver tumor and lungs in Nu/Nu mice from the SAM2 lentivirus library transduction group or non-transduction group (×100); (F) Representative immunohistochemical stains of VEGF, Vimentin and MMP9 protein in primary liver tumor and lung from Nu/Nu mice that was either untransduced or transduced with the SAM2 lentiviral library (×200). Comparisons between groups were analyzed through unpaired Student's *t* test or Kaplan-Meier Log Rank test. Data were represented as means ± SD of a representative experiment. ns: No significance, compared to the control group.

However, the CRISPRa have yet been used to directly evaluate genes responsible for liver tumor growth and metastasis.

In the current study, we provided a systematic phenotypic measurement of gain-of-function screening by CRISPRa during the growth and metastasis of primary liver tumors, which was of great significance to develop a new targeted therapy for HCC. We identified candidate target XAGE1B, PLK4, LMO1 and MYADML2 that promoted cells proliferation and invasion. High expression of LMO1, MYADML2, PLK4 and XAGE1B exhibited worse overall survival in HCC. In addition, we found high MYADML2 level reduced the sensitivity to chemotherapeutic drugs. What's more, the dendritic cells, macrophages and B cells, and so forth might play an important role in promoting HCC proliferation and invasion.

RESULTS

Genome-wide clustered regularly interspaced short palindromic repeats-associated protein 9 screening *in vivo*

We performed gain-of-function metastasis screen in mouse model utilizing the human genome-scale CRISPR/Cas9 SAM2 pooled library. The HEPG2 cell line were transduced with the SAM2 pooled library in eight independent infection replicate experiments. We constructed orthotopic tumor model by transplanting 4×10^6 cells into the liver of immunocompromised BLAB/c Nu/Nu mice after *in vitro* culture for 1 week (Figure 1A). Figures 1B and 1C showed that SAM2 library transduced mice has lower body weight ($p > 0.05$) and lower survival rate ($p = 0.17$) compared with the control, but there is no significant difference. Seven weeks post-transplantation, mice were sacrificed and the metastasis of various organs were checked. Under visual observation, both SAM2 library transduced and untransduced HEPG2 cells formed tumors at the injection site (Figure 1D). According to the original survival data, 71.43% (5/7) mice were alive in control group, while 37.5% ($^3/_8$) were alive in SAM2-transduced group. In addition, the SAM2-transduced mice had 100% (3/3) positive metastases in the lung lobes and the mice in control group developed none (0/5) detectable metastases in the lungs (Figure 1E). Furthermore, the expression levels of VEGF, Vimentin and MMP9 protein were significantly higher in the liver of SAM2-transduced mice and the lung of SAM2-transduced mice had higher VEGF, and MMP9 protein levels (Figure 1F). These results suggested that SAM2 library transduction enhanced the ability of the HEPG2 cells to form metastases in the lung.

Identifies pathways affecting liver tumor growth and metastasis

Seven weeks post-transplantation, liver and lung are collected to extract DNA for PCR amplification. After amplifying the sgRNA sequence in the genome, high-throughput sequencing was used to calculate the coverage of sgRNAs in the cell. The flowchart in Figure 2A shows the data analysis process. There are 21066 and 10807 sgRNAs in the liver and lung identified, respectively, and then genes with sgRNA abundance greater than 100 (Table S1) were selected for the further analyses and 4167 sgRNAs were screened. Then, the corresponding 4167 genes were divided into three groups by comparing the differences of the normalized sgRNA abundance between lung and liver (Table S1). The H-type gene (Log2FC > 1, sgRNAs abundance in lungs greater than that in liver, 141 genes) might mediate lung colonization population and is a lung metastasis-specific gene, the C-type gene ($1 > \text{Log2FC} > -1$, 2752 genes) might promote both tumor growth and lung metastasis, the L-type gene (Log2FC < −1, sgRNAs abundance in lungs less than that in liver, 1274 genes) might mainly promote tumor growth but not metastasis (Figure 2B). We further analyzed these genes among these three groups by Kyoto encyclopedia of genes and genomes (KEGG) pathway analysis. The result showed these commonly differentially expressed sgRNAs of H type gene were enriched in the cell adhesion molecules (CAMs), Notch, and so on. The C type gene demonstrated their association with metabolic, PI3K-Akt, regulation of actin cytoskeleton, MAPK, FoxO signaling pathways, and so on. And the L type gene suggested they were involved in VEGF, tight junction, cell cycle, transcriptional misregulation in cancer pathways, and so forth (Figure 2C, Table S2). All of these pathways are closely related to tumor growth and metastasis.

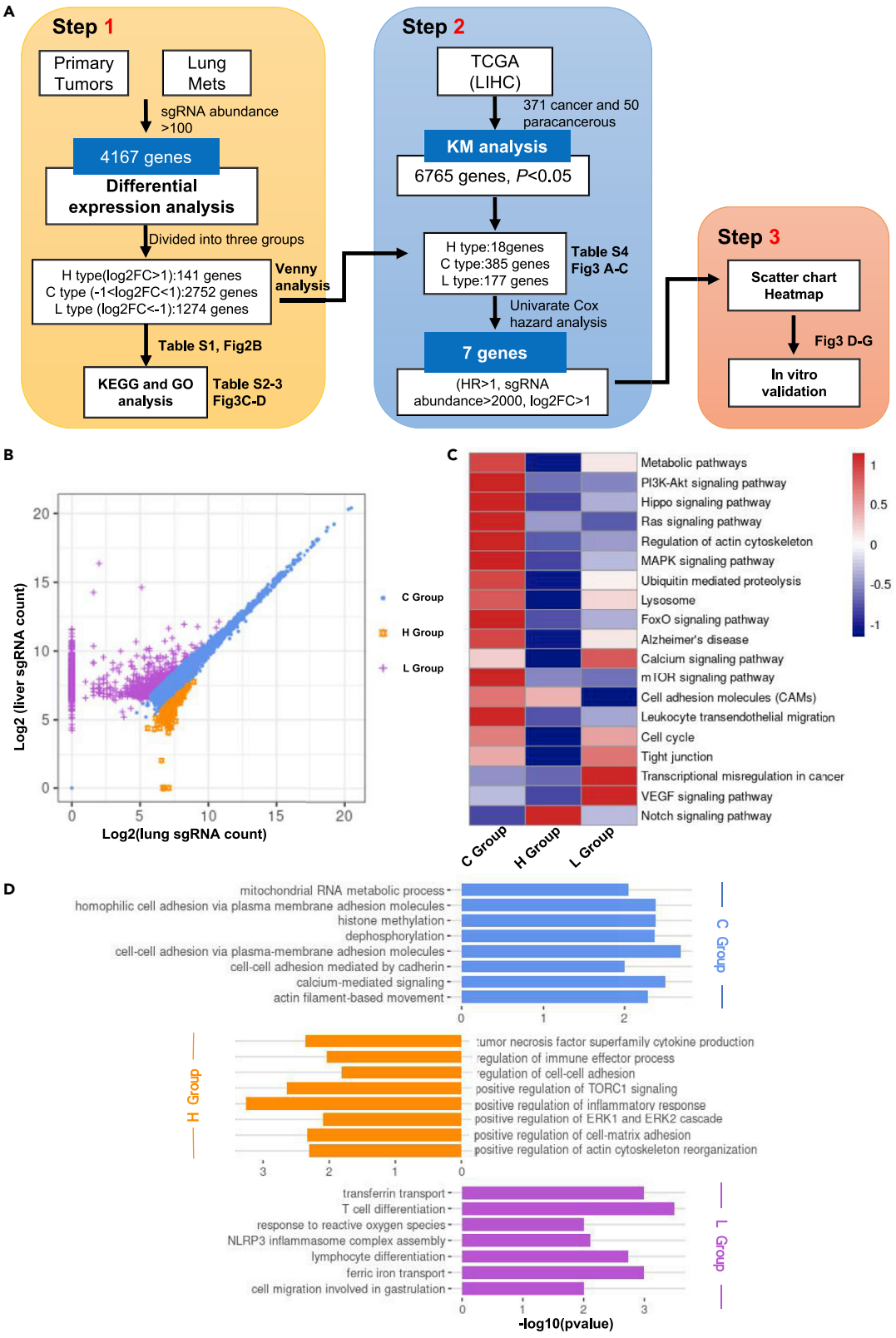

A
Step 1
Primary Tumors
Lung Mets
sgRNA abundance >100
4167 genes
Differential expression analysis
Divided into three groups
H type(log2FC>1):141 genes
C type (-1<log2FC<1):2752 genes
L type (log2FC<-1):1274 genes
Venny analysis
Table S1, Fig2B
KEGG and GO analysis
Table S2-3 Fig3C-D
Step 2
TCGA (LIHC)
371 cancer and 50 paracancerous
KM analysis
6765 genes, P<0.05
H type:18genes
C type:385 genes
L type:177 genes
Table S4 Fig3 A-C
Univarate Cox hazard analysis
7 genes
(HR>1, sgRNA abundance>2000, log2FC>1
Step 3
Scatter chart Heatmap
Fig3 D-G
In vitro validation
B
Log2 (liver sgRNA count)
Log2(lung sgRNA count)
C Group
H Group
L Group
C
Metabolic pathways
PI3K-Akt signaling pathway
Hippo signaling pathway
Ras signaling pathway
Regulation of actin cytoskeleton
MAPK signaling pathway
Ubiquitin mediated proteolysis
Lysosome
FoxO signaling pathway
Alzheimer's disease
Calcium signaling pathway
mTOR signaling pathway
Cell adhesion molecules (CAMs)
Leukocyte transendothelial migration
Cell cycle
Tight junction
Transcriptional misregulation in cancer
VEGF signaling pathway
Notch signaling pathway
C Group
H Group
L Group
D
mitochondrial RNA metabolic process
homophilic cell adhesion via plasma membrane adhesion molecules
histone methylation
dephosphorylation
cell-cell adhesion via plasma-membrane adhesion molecules
cell-cell adhesion mediated by cadherin
calcium-mediated signaling
actin filament-based movement
C Group
tumor necrosis factor superfamily cytokine production
regulation of immune effector process
regulation of cell-cell adhesion
positive regulation of TORC1 signaling
positive regulation of inflammatory response
positive regulation of ERK1 and ERK2 cascade
positive regulation of cell-matrix adhesion
positive regulation of actin cytoskeleton reorganization
H Group
transferrin transport
T cell differentiation
response to reactive oxygen species
NLRP3 inflammasome complex assembly
lymphocyte differentiation
ferric iron transport
cell migration involved in gastrulation
L Group
-log10(pvalue)

Figure 2. The genome-wide pooled sgRNA library screen identified pathways involved in liver tumor growth and metastasis
(A) The flowchart of bioinformatics analysis; (B) The 4,167 genes with sgRNA abundance greater than 100 were divided into three groups (C, H, L) by comparing the differences of the normalized sgRNA abundance between lung and liver; (C) KEGG analysis of the enriched pathways in the three groups (C, H, L); (D) GO analysis of the enriched biological process terms in the three groups (C, H, L).

GO analysis enriched the biological process terms in the three groups. The function of genes in C group was related to cell-cell adhesion mediated by cadherin or plasma-membrane adhesion molecules, calcium-mediated signaling, and so forth. Genes in H group was relevant to the biological process of positive regulation of inflammatory response, TORC1 signaling, actin cytoskeleton reorganization and ERK1 and ERK2 cascade. While, L group genes are mainly enriched in T cell differentiation, ferric iron transport, NLRP3 inflammasome complex assembly and cell migration involved in gastrulation (Figure 2D, Table S3).

Enriched genes that affect tumor growth and metastasis

We next analyzed a total of 421 samples, including 371 liver cancer samples and 50 paracancerous samples from TCGA with both molecular and clinical data. In total 6765 genes had prognostic significance ($p < 0.05$) according to the Kaplan-Meier survival analysis. AS was showed in Figure 3A, of all the 6765 genes from TCGA ($p < 0.05$), there are 385 genes intersected with C group, 177 genes intersected with L group and 18 genes intersected with H group. Thus, in total 580 genes were selected. The volcano plots showed a differential analysis of 580 genes in liver cancer, representing a comparison of 371 cancer samples and 50 paracancerous samples (Figure 3B), and 50 pairs of matched HCC and paracancerous comparisons (Figure 3C). We used three methods to identify enriched genes in primary liver tumors and lung metastases: (1) sgRNAs abundance in liver and lung greater than 2000. (2) log2FC for comparison between 371 HCC samples and 50 paracancerous samples, and 50 pairs of matched HCC and paracancerous comparisons are all greater than 1. (3) Top-ranked sgRNA in C Group. Through these methods a total of 7 genes (POLR2L, XAGE1B, HOXD1, PLK4, C17orf99, LMO1 and MYADML2) were selected. The sgRNAs abundance for each selective gene was shown (Figure 3D). XAGE1B had two significantly resistant sgRNAs. The counts of sgRNAs targeting the 7 specific genes in lung and liver, and the number of sgRNAs targeting 7 genes versus NGS read fold-changes were showed in Figures 3E and 3F. Some of the genes are well-known oncogenes, for example, LMO1, belongs to the family of LIM-only domain genes (LMOs), cooperates with MYCN to initiate neuroblastoma and contribute to metastatic disease progression.[24] and has been reported to have an oncogenic role in cancer.[25] Another oncogene, Polo-like kinase 4 (PLK4), is a serine/threonine protein kinase that regulates centriole duplication, its overexpression leads to centriole amplification and further genomic instability and tumorigenesis.[26]

Overexpression of candidate genes promote hepatocellular carcinoma cells proliferation and invasion

To validate the results from our CRISPR/Cas9 activation library screening, we next investigated the effects of the candidate genes on HCC cell phenotype. The HCC cell lines HEPG2 and HUH7 were infected with CRISPR activation lentivirus POLR2L, XAGE1B, HOXD1, PLK4, C17orf99, LMO1 and MYADML2 or the control lentivirus. Then q-PCR confirmed that all sgRNA sequences (Table 1) were able to effectively upregulated POLR2L, XAGE1B, HOXD1, PLK4, C17orf99, LMO1 and MYADML2 in HEPG2 and HUH7 cells ($p < 0.05$) (Figure S1). CCK8 assays showed that the overexpression of POLR2L, XAGE1B, HOXD1, PLK4, C17orf99, LMO1 and MYADML2 could significantly promote HCC cell proliferation ($p < 0.05$) (Figure 4A). The data showed that the gain function of POLR2L, XAGE1B, HOXD1, PLK4, C17orf99, LMO1 and MYADML2 increases survival of HCC cells.

The 7 candidate genes screened all belong to "C group" genes, implying that they could not only promote liver tumors growth, but also mediate lung colonization and promote lung metastasis. Thus, these cells with high expression of POLR2L, XAGE1B, HOXD1, PLK4, C17orf99, LMO1 and MYADML2 were then subjected to the cell invasion experiment. As expected, the number of experimental invasion HEPG2 cell by all tested genes increased significantly, which was significantly different from the control group ($p < 0.01$) (Figure 4B), and the overexpression of LMO1, MYADML2, PLK4 and XAGE1B remarkably promoted Huh7 cell invasion ($p < 0.01$) (Figure 4C). Our studies showed that under this genetic background, the gain-of-function of LMO1, MYADML2, PLK4 and XAGE1B is sufficient to accelerate the invasion of HCC cells.

Inhibition of the candidate genes suppress hepatocellular carcinoma cells proliferation and invasion

We further explored whether inhibiting LMO1, MYADML2, PLK4 and XAGE1B expression could suppress HCC cells proliferation and invasion. The siRNAs of LMO1, MYADML2, PLK4 and XAGE1B (Table 2) were

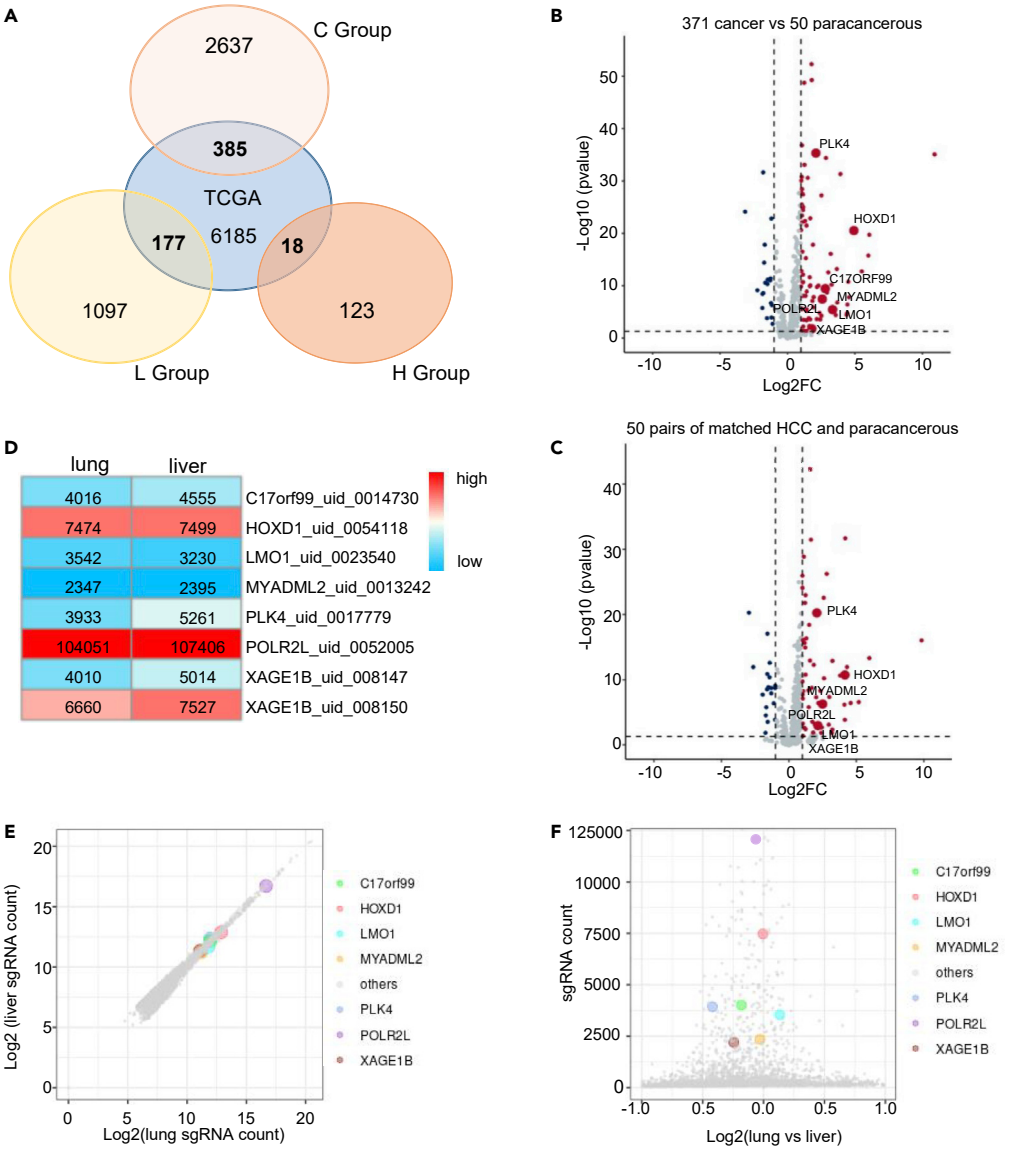


Figure 3. Enriched sgRNAs from the genome-scale CRISPR/Cas9 activation screen in liver tumor growth and metastasis

(A) The Venn diagram showed 385 C type genes, 18 H type genes and 177 L type genes intersected with 6765 prognostic genes with from TCGA ($p < 0.05$); (B) The volcano plots showed a differential analysis of the 580 genes in liver cancer, representing a comparison of 371 cancer samples and 50 paracancerous samples, the candidate 7 genes were marked; log2FC: log2FoldChange between 371 HCC and 50 paracancerous; (C) The volcano plots showed a differential analysis of 580 genes in liver cancer, representing a comparison of 50 pairs of matched HCC and paracancerous comparisons, the candidate 7 genes were marked; log2FC: log2FoldChange between50 pairs of matched HCC and paracancerous; (D) The heatmap shows the counts of resistant sgRNAs for the selected 7 genes; (E) The number of sgRNAs from the 7 candidate genes in lung and liver; (F) Fold changes the 7 candidate genes exhibiting posteriorly enriched sgRNAs.

used to transfect the HEPG2 and Huh7 cell lines. WB was used to detect the protein expression levels and selected the siRNA with the best inhibitory effect (Figure 5A). Then, CCK8 and Transwell assays were used to evaluate the cell proliferation and migration. The result showed that the down-regulation of LMO1, MYADML2, PLK4 and XAGE1B could significantly inhibit HCC cell proliferation at 24h, 48h and 72h ($p < 0.05$, $p < 0.01$, $p < 0.001$) (Figure 5B). Transwell assays clearly revealed the invasion cells in LMO1, MYADML2, PLK4 and XAGE1B-siRNA group was significantly reduced compared to siNC group in HEPG2 and Huh7 cells at 48h ($p < 0.001$) (Figures 5C and 5D).

Table 1. Primers sequences for q-PCR

Primer name	Product size	Sequences
β-actin	185 bp	F: TGGCACCAGCACAATGAA
		R: CTAAGTCATAGTCCGCCTAGAAGCA
C17orf99	250bp	F: CAAGGCACGGGAGGAAGAAATTA
		R: CCAGCAGAAGTAGGTGAGCA
LMO1	210 bp	F: TGATGGTGCTGGACAAGGAG
		R: GGTTGGCCTTGGTGTAGAGG
MYADML2	129bp	F: TCTGGCCAGTCTTCTGTTTCG
		R: ACAGGAGCAGGTTGACGTAG
XAGE1B	153 bp	F: CAGCAGCTGAAAGTCGGGAT
		R: CTTGACGCCGGAACCCAAAT
PLK4	120bp	F: AAGAGAGGTACTCACCCACAGA
		R: GATGATTGGAGAGTGCTTGATTGT
POLR2L	199bp	F: GATCATCCCTGTACGCTGCT
		R: CTTCTCCAGGGGTGCATAA
HOXD1	169bp	F: CTTGTCTCAAAGCGTCAGCC
		R: GCGAGTTTGCCTTTCTTAG

Functional enrichment in the protein level of MYADML2

In order to understand the expression profile of these candidate genes in HCC, we used the RNA seq data of Cancer Genome Atlas (TCGA), including molecular and clinical data. We evaluated the expression levels of these candidate genes in the TCGA database (371 HCC samples and 50 paracancerous samples), which provide gene expression data from for both paired non-tumor and HCC tissues. Of the identified hits, LMO1 ($p = 0.026$), MYADML2 ($p = 0.0006$), PLK4 ($p = 0.000$) and XAGE1B ($p = 0.005$) mRNA expression was selectively up-regulated in HCC (Figure 6A). The 371 HCC samples was devided into low and high expression group according to the median of LMO1, MYADML2, PLK4 and XAGE1B mRNA levels. Kaplan-Meier analysis showed that patients with HCC with high expression of LMO1, MYADML2, PLK4 and XAGE1B had poorer overall survival compared to patients with low expression of these genes ($p = 0.0182$, $p = 0.048$, $p = 0.0162$, $p = 0.0011$, $p = 0.0084$, log rank test, Figure 6B). In addition, the Cox proportional hazards model was used to evaluate whether LMO1, MYADML2, PLK4 and XAGE1B are valuable predictor for the survival of patients with HCC from the TCGA dataset. Univariate analysis indicated that these genes were significant prognostic factor: LMO1 (HR, 1.5; 95% CI,1.0-2.1, $p = 0.031$), MYADML2 (HR,1.7; 95% CI,1.2-2.4, $p = 0.0025$), PLK4 (HR, 0.66; 95% CI,0.47-0.93; $p = 0.019$), XAGE1B (HR, 1.6; 95% CI,1.1-2.4, $p = 0.011$). These results suggests that the high expression of LMO1, MYADML2, PLK4 and XAGE1B not only leads to a higher occurrence of lung metastasis probability but also affects the prognosis of HCC. Univariate Cox regression analysis of the candidate genes showed that MYADML2 were the most significant prognostic factor.

Thus, immunohistochemistry was used to detect MYADML2 protein expression in 80 HCC tissues. The result showed MYADML2 was positively expressed in HCC cells and mainly located in the cytoplasm and membrane (Figure 6C). Kaplan-Meier analysis indicated that patients with HCC with high MYADML2 protein level had worse overall survival than these with low MYADML2 protein level ($p = 0.013$, Figure 6D). Interestingly, we also found that MYADML2 protein level was significantly increased in patients with HCC over 60 years of age ($p < 0.05$) (Figure 6E). The XAGE1B mRNA level was significantly increased in patients with HCC over 60 years of age ($p < 0.0001$). While, there are no is no difference of the LMO1, MYADML2 and PLK4 mRNA level between young and old ($p > 0.05$) (Figure 6F). Besides, we used the GDSC database to predict the sensitivity of each sample in TCGA to different chemotherapeutic agents, the result showed that the IC50 of Sunitinib, Imatinib, Gefitinib and Lapatinib was higher in MYADML2 high expression group (Figure 6G).

Immune cell infiltration analysis

Proportional analysis of the 15 immune cell types in the TCGA dataset revealed that some cells were highly abundant, such as macrophages cells and naive B cells (Figures 7A and 7B). The infiltration levels of immune cells between low and high expression groups of these genes showed that high PLK4, LMO1 and MYADML2

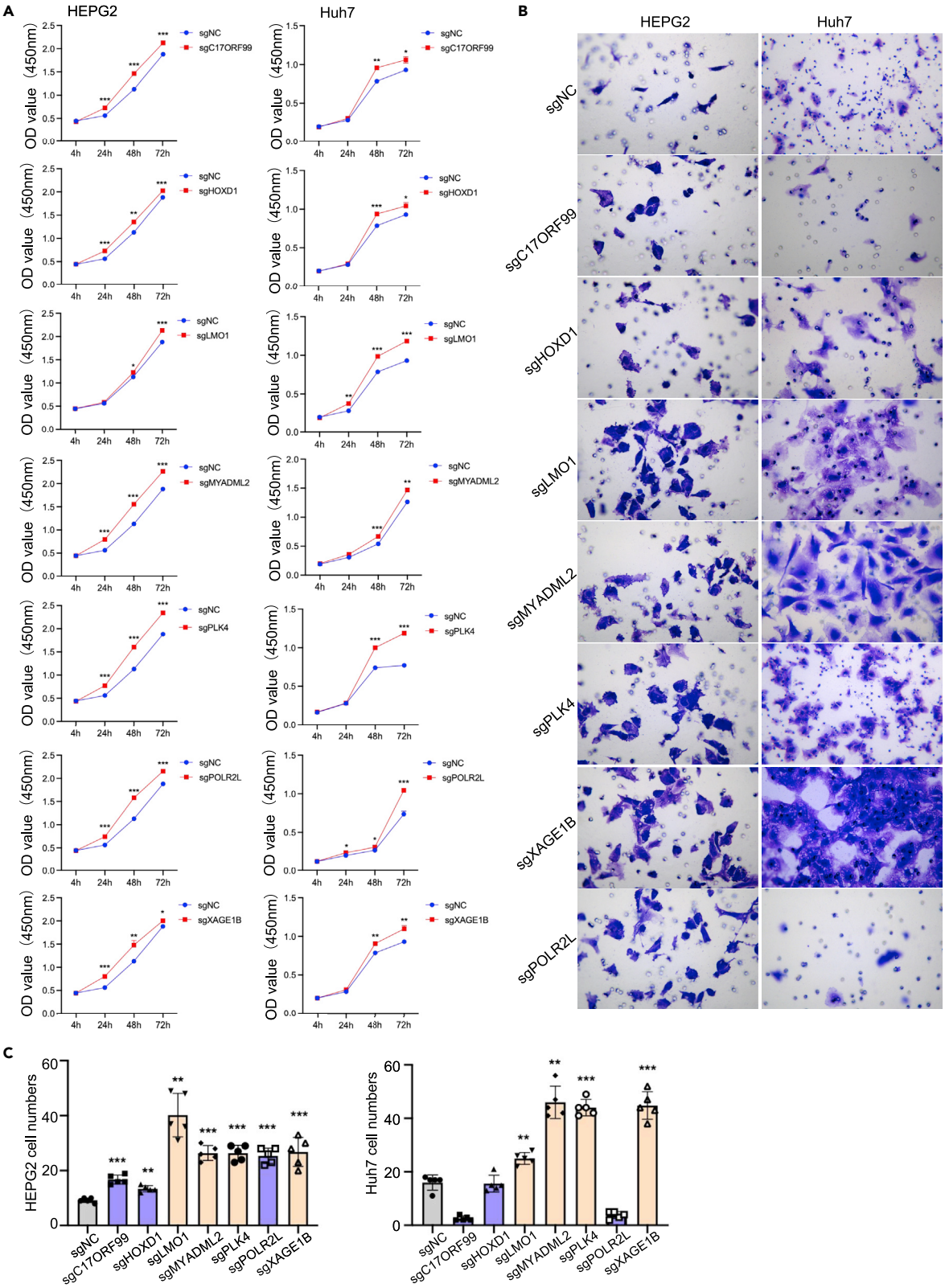

A
HEPG2
Huh7
OD value (450nm)
4h
24h
48h
72h
sgNC
sgC17ORF99
sgHOXD1
sgLMO1
sgMYADML2
sgPLK4
sgPOLR2L
sgXAGE1B
B
HEPG2
Huh7
C
HEPG2 cell numbers
Huh7 cell numbers

Figure 4. Overexpression of candidate genes promoted hepatocellular carcinoma cells proliferation and invasion

(A) CCK8 assay was used to detect cell viability at 4h, 24h, 48h and 72h for HEPG2 and HUH7 cells with individual gene overexpression.

(B-C) Transwell assays clearly revealed the invasion cells of in HEPG2 and Huh7 at 24 h after vaccination in the wells. The cell invasion numbers were counted and compared. Statistical significance was determined through one-way ANOVA. Data were represented as means ± SD of a representative experiment (of three or five independent experiments). **p < 0.01, ***p < 0.001 compared with the NC group.

level was consistent with high proportion of M0 macrophages cells ($p < 0.0001$, $p < 0.01$, $p < 0.0001$), high PLK4 level was also consistent with resting dendritic cells ($p < 0.001$). In addition, PLK4 and MYADML2 was negatively correlated with the infiltration level of monocytes ($p < 0.0001$), naive B cells ($p < 0.001$) and M2 macrophages cells ($p < 0.0001$) (Figure 7C). The result indicating that the dendritic cells, macrophages and B cells, and so forth played an important role in promoting HCC proliferation and invasion.

DISCUSSION

Age is considered to be the single most important risk factor for many chronic conditions, including most common malignancies. The incidence of common malignancies is increased exponentially with age. HCC is the main cause of cancer-related deaths worldwide. Although many scientific advances have been made, the prognosis of many patients with HCC is still poor due to intrahepatic and extrahepatic metastasis and postoperative recurrence. The pathogenesis of HCC progression and metastasis is still unclear, resulting in a lack of effective treatment for advanced HCC.[27] In this study, the orthotopic HCC model was created through directly intrahepatic injection of HEPG2 cells, and the cells underwent the clonal growth cascade to form distal metastases.[14,28] Beginning 3 weeks after injection, death was observed in tumor-bearing animals. Seven weeks after HEPG2 cell implantation, the mortality of animal untransduced with SAM2 lentiviral library is 28.57%, while 62.5% of animals transduced with the SAM2 lentiviral library died of tumor burden. The result also showed that after HEPG2 cell implantation, tumors formed in the livers of nude mice and a small white tumor mass was observed on the surface of the lobe where cells were injected. Furthermore, these tumors grew rapidly in the injected lobe and began to spread to other lobes. No metastases to other organs were observed in any of the animals tested. However, HE staining showed that the cell population formed highly metastatic tumors in lung following the induction of a genome-scale CRISPR/Cas9 activation library. In addition, compared with the untransduced group, MMP9 and VEGF protein levels were significantly elevated in transduced group. Thus, these mutations, acting in simple or complex pleiotropic ways, accelerate metastasis. In this model, the effect of mutations on metastasis strongly correlates with their abundance in primary tumors.

CRISPR/Cas9 activation library screening enables researchers to perform forward or reverse genetic screening and identify genes contributed to a specific phenotype in a genome-wide scale. In this study, we adopted the SAM2 lentiviral library screening strategy. The quality check of our library screening was conducted by accessing the read depth, sgRNA coverage in each group. The result showed the remarkably enriched genes in lung metastasis and the abundantly enriched genes in primary liver tumors overlap to a large extent. Moreover, we found sgRNA dynamics has undergone dramatically changes over the course of tumorigenesis and metastasis, reflecting the targeted genes involved in fundamental cellular processes.

After normalizing the sgRNA abundance and comparing the differences between liver and lung, the 4,167 genes were divided into three groups. The H-type gene (141 genes) can mediate lung colonization population and is a lung metastasis-specific gene, the C-type gene (2752 genes) can promote both tumor growth and lung metastasis, the L-type gene (1274 genes) can mainly promote tumor growth but not metastasis. Two methods were used to identify genes enriched in primary liver tumors and lung metastases, and finally the 7 candidate genes (POLR2L, XAGE1B, HOXD1, PLK4, C17orf99, LMO1 and MYADML2) were considered to promote liver cancer growth and lung metastasis, which had a association with metabolic, PI3K-Akt, regulation of actin cytoskeleton, MAPK signaling pathways, and so on. Several genes enriched in primary tumors and lung metastasis are related to cancer, but their role in tumor growth is still poorly understood. POLR2L encodes RNA polymerase II subunit, which is the responsible for the synthesis of mRNA polymerase in eukaryotes,[29] and was associated with cisplatin resistance in gastric cancer.[30] Ovarian cancer variant rs2072590 is associated with HOXD1 gene expression,[31] which encodes a protein with a homeobox DNA-binding domain.

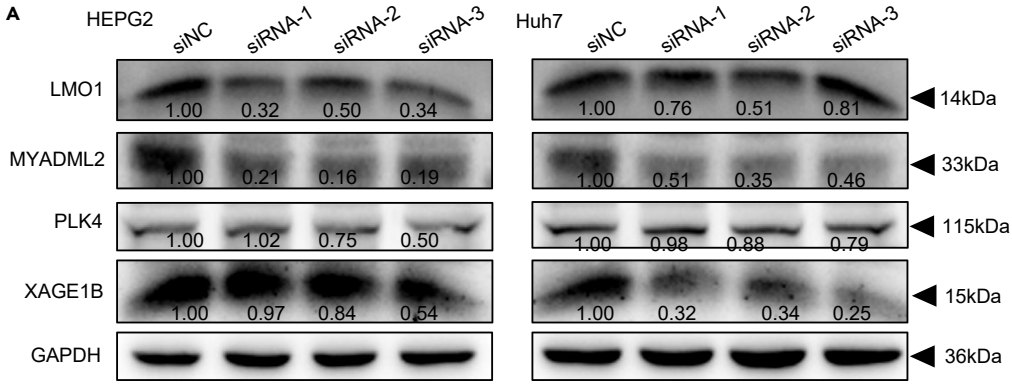
Functionally, we validated the effect of the 7 candidate genes *in vitro*, we observed the overexpression of XAGE1B, PLK4, LMO1 and MYADML2 can significantly promote the liver cancer cell proliferation and

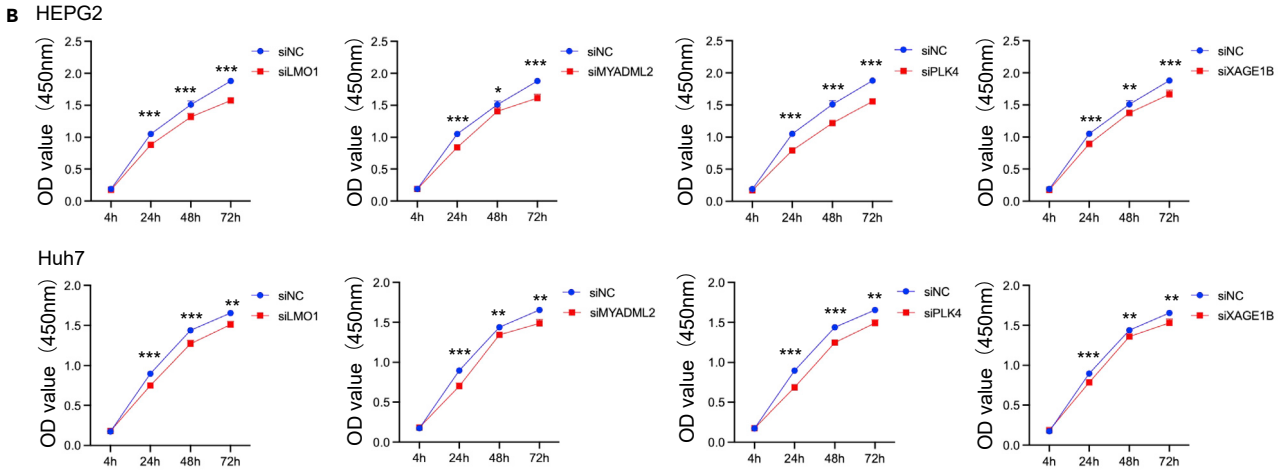
Table 2. siRNA targeting sequences

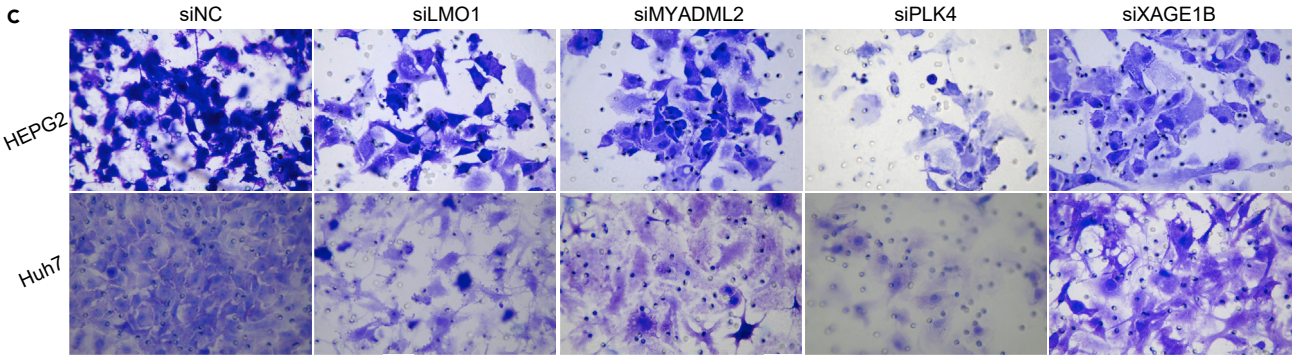
Primer name	Sequences(5′to3′)
LMO1 siRNA-1	GAGACAAAUUCUUCCUGAATT
	UUCAGGAAGAAUUUGUCUCTT
LMO1 siRNA-2	UCUUUGGCACCACAGGGAATT
	UUCCCUGUGGUGCCAAAGATT
LMO1 siRNA-3	CACCUUUGAAUCCCAAGUUTT
	AACUUGGGAUUCAAAGGUGTT
MYADML2 siRNA-1	CCAAGUACGGUGAGCCCAATT
	UUGGGCUCACCGUACUUGGTT
MYADML2 siRNA-2	GGUGGUGGCCUGUGAGUUCTT
	GAACUCACAGGCCACCACCTT
MYADML2 siRNA-3	GCUAUAUGGCCACGGUGUCTT
	GACACCGUGGCCAUAUAGCTT
PLK4 siRNA-1	ACAACUAGGUAUGGAGAAATT
	UUUCUCCAUACCUAGUUGUTT
PLK4 siRNA-2	ACAGUAAACCUGAGAUAAUTT
	AUUAUCUCAGGUUUACUGUTT
PLK4 siRNA-3	CCAAAUGGUGGUAGAGGUUTT
	AACCUCUACCACCAUUUGGTT
XAGE1B siRNA-1	GCGUCAAGGUGAAGAUAAUTT
	AUUAUCUUCACCUUGACGCTT
XAGE1B siRNA-2	CAUGGAAGGUGAUCUGCAATT
	UUGCAGAUCACCUUCCAUGTT
XAGE1B siRNA-3	GCUGAAAGUCGGGAUCCUATT
	UAGGAUCCCGACUUUCAGCTT

invasion, and inhibition of the candidate genes suppress HCC proliferation and invasion, which was consistent with the previous studies. These genes are well-known oncogenes. LMO1 cooperates with MYCN to initiate neuroblastoma and contribute to metastatic disease progression and have an oncogenic role in cancer.[24,25] Another oncogene, PLK4, is a serine/threonine protein kinase that regulates centromere duplication, whose overexpression leads to centromere amplification and further genomic instability and tumorigenesis.[26] Interestingly, PLK4 is a potential therapeutic target in many cancers.[32] While, the role of some genes enriched in primary tumors and lungs in tumor growth remains poorly understood. XAGE-1b, a gene with two independent sgRNAs, has been identified as an overexpressed surface antigen in NSCLC cells and has been shown to be immunogenic.[33] The *in vivo* screening model also shown that the gain of XAGE1B, PLK4, LMO1 and MYADML2 promoted the growth and metastasis in mouse HCC model. However, its role in HCC metastasis is largely unknown and further studies are needed to determine whether the candidate genes can serve as a therapeutic target to address HCC progression.

In our analysis of TCGA samples from patients with liver cancer, we observed XAGE1B, PLK4, LMO1 and MYADML2 mRNA levels was significantly upregulated in tumor compared with the precancerous. The further cox proportional hazard model and Kaplan-Meier survival analysis showed the expression levels of XAGE1B, PLK4, LMO1 and MYADML2 were significantly correlated with overall survivals in HCC. These results suggests that the high expression of LMO1, MYADML2, PLK4 and XAGE1B not only means a higher occurrence of lung metastases probability but also affects the prognosis of HCC. There are nearly no reports on MYADML2. Yıldız Bölükbaşı E, et al.[34] reported a homozygous deletion of the coding sequences of MYADML2 as the first disease-causing variant in the gene together with a deletion in PYCR1. Further immunohistochemical staining of the expression of MYADML2 protein in HCC TMA showed that MYADML2 protein was positively expressed in HCC tissues. Kaplan-Meier analysis indicated that patients with HCC with high MYADML2 protein level exhibited worse overall survival. Interestingly, the expression of MYADML2 protein


A
HEPG2
siNC
siRNA-1
siRNA-2
siRNA-3
Huh7
LMO1
1.00 0.32 0.50 0.34
1.00 0.76 0.51 0.81
14kDa
MYADML2
1.00 0.21 0.16 0.19
1.00 0.51 0.35 0.46
33kDa
PLK4
1.00 1.02 0.75 0.50
1.00 0.98 0.88 0.79
115kDa
XAGE1B
1.00 0.97 0.84 0.54
1.00 0.32 0.34 0.25
15kDa
GAPDH
36kDa


B
HEPG2
OD value (450nm)
siNC
siLMO1
siMYADML2
siPLK4
siXAGE1B
4h 24h 48h 72h
Huh7


C
siNC
siLMO1
siMYADML2
siPLK4
siXAGE1B
HEPG2
Huh7

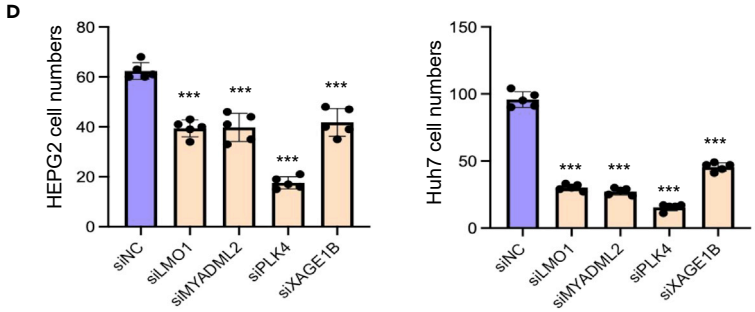

D
HEPG2 cell numbers
Huh7 cell numbers
siNC
siLMO1
siMYADML2
siPLK4
siXAGE1B

Figure 5. Inhibition of the candidate genes suppress hepatocellular carcinoma cells proliferation and invasion

(A) The siRNAs of LMO1, MYADML2, PLK4 and XAGE1B were used to transfect the HEPG2 and Huh7 cell lines. WB was used to detect the LMO1, MYADML2, PLK4 and XAGE1B protein expression levels, ImageJ was used for quantitative analysis; (B) CCK8 assay was used to detect cell viability at 4h, 24h, 48h and 72h in siNC, siLMO1, siMYADML2, siPLK4 and siXAGE1B groups of HEPG2 and Huh7 cells; (C-D) Transwell assays clearly revealed the invasion cells in siNC, siLMO1, siMYADML2, siPLK4 and siXAGE1B groups of HEPG2 and Huh7 cells at 48 h after vaccination in the wells. The cell invasion numbers were counted and compared. Statistical significance was determined through one-way ANOVA. Data were represented as means ± SD of a representative experiment (of three or five independent experiments). *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ compared with the siNC group.

was significantly increased in patients with HCC over 60 years. Moreover, the high expression of MYADML2 reduced the sensitivity to chemotherapeutic drugs, such as Sunitinib, Imatinib, Gefitinib and Lapatinib, which implying that targeting MYADML2 may be a potential strategy to increase sensitivity to chemotherapy drugs.

The tumor microenvironment (TME) is a complex cell population, including fibroblasts, epithelial cells and immune cells (i.e. neutrophils, monocytes/macrophages, dendritic cells, Treg cells, and so forth). Tumor-associated macrophages (TAMs) are the most abundant immune cells in TME, which have a large impact on tumor development and immunotherapy.[35] Immune response mediated by TAMs is critical in tumor progression and M0-like macrophage is the feature of malignancy of glioblastoma. Our result showed that high PLK4, LMO1 and MYADML2 levels was consistent with the high proportion of M0 macrophages cells, which indicated the increased expression of these genes in HCC may be closely related to immunosuppression in tumor microenvironment. Monocytes and dendritic cells are innate immune cells, which are involved in antigen presentation and induction of immune response, we found PLK4 and MYADML2 was negatively correlated with the infiltration levels of monocytes and resting dendritic cells, which may lead to reduced recognition and killing of tumor cells.

In summary, we used the CRISPR/Cas9 genome editing provides a roadmap for screening functional genes related to liver tumor invasion and metastasis *in vivo*. Our results suggested that the high expression of LMO1, MYADML2, PLK4 and XAGE1B not only meant a higher occurrence of lung metastases probability but also affected the prognosis of HCC. In addition, we found the expression of MYADML2 protein was significantly increased in patients with HCC over 60 years. What' more, high MYADML2 level in HCC reduced the sensitivity to chemotherapeutic drugs. These findings provided new potential biological targets for clinical diagnosis, treatment and prognosis evaluation of HCC.

Limitations of the study

This study focuses on identifying targets responsible for HCC cell growth and metastasis in a mouse model by the genome-wide CRISPRa library. The validation of the modulator was achieved in cellular models and HCC samples but the specific roles of the modulators in HCC would need further investigation using a animal model. Furthermore, high PLK4 and MYADML2 levels in HCC may be closely related to immunosuppression in tumor microenvironment and reduced ability to recognize and kill tumor cells, it would thus be interesting to investigate the potential mechanism.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell lines
  - Animal model
  - Tissue samples
- METHOD DETAILS
  - Lentivirus preparation of sgRNAs library
  - Lentiviral transduction
  - Genomic DNA extraction
  - sgRNA deep sequencing
  - Establishment of candidate genes overexpression HCC cells

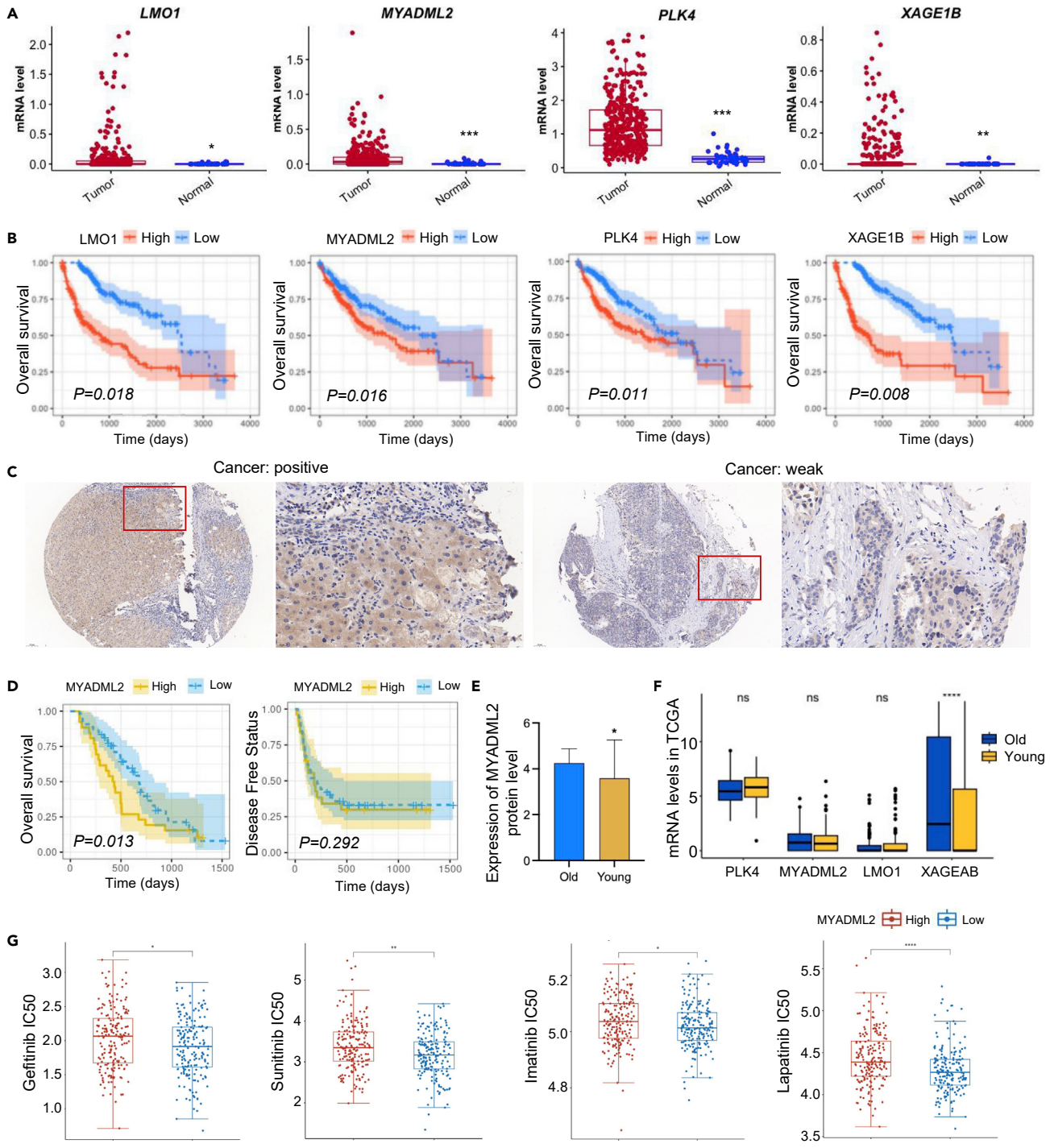


Figure 6. These candidate genes were valuable biomarker in clinical HCC samples

(A) The LIHC data from TCGA was used to analyze the LMO1, MYADML2, PLK4 and XAGE1B mRNA expression in HCC tumor (371 samples) and paracancerous (50 samples) group; (B) The HCC samples from TCGA was devided into low and high expression group according to the median of mRNA levels. Kaplan-Meier analysis revealed that high expression of LMO1, MYADML2, PLK4 and XAGE1B exhibited worse overall survival according the TCGA dataset; (C) MYADML2 protein showed positive or weak immunostaining in the cytoplasm and membrane of HCC tissues from a tissue microarray; (D) MYADML2 protein level has increased in HCC patient over 60 years (Young group: age<60 years, N = 68; Old group: age≥60 years, N = 12); (E) Overall survival and disease-free survival analysis of MYADML2 protein in HCC tissues (N = 80); (F) The HCC samples from TCGA was devided into young (age≤60) and old expression group (age>60), the differences of LMO1, MYADML2, PLK4 and XAGE1B mRNA were analyzed between the young and old group; (G) The distribution of IC50 score. The abscissa represents different groups of samples, and the ordinate represents the distribution of the IC50 score. Comparisons between groups were analyzed through unpaired Student's *t* test or Kaplan-Meier Log Rank test. Data were represented as means ± SD *p < 0.05; **p < 0.01; ***p < 0.001. ****p < 0.0001 compared with the control group.

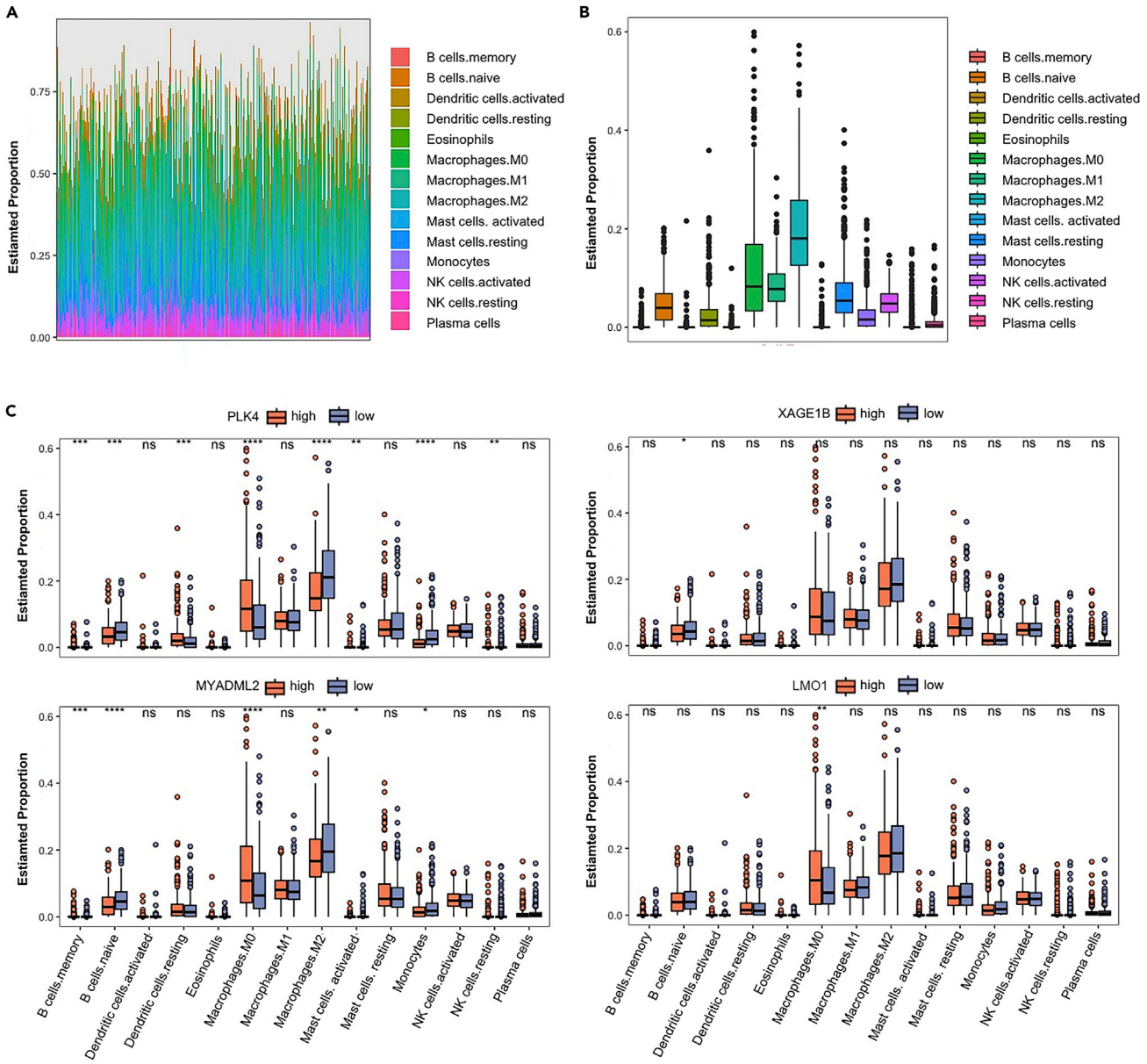


Figure 7. Analysis of immune cell composition in liver cancer

(A-B) TCGA data analyzes the proportion of 15 immune cell types in liver cancer; (C) The boxplot showed the different infiltration levels of 15 immune cell types in the high/low PLK4, MYADML2, XAGE1B and LMO1 expression groups. Comparisons between groups were analyzed through unpaired Student's *t* test. Data were represented as means ± SD of a representative experiment. *p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001, ns: No significance compared with the control group.

- siRNA transfection
- Real-time quantitative reverse transcriptase PCR
- Cell proliferation assay
- Migration assay
- Mouse tissue collection
- Immunohistochemistry
- Western blot
- TCGA data analysis
- Evaluating the sensitivity of chemotherapeutic agents
- Immune cell infiltration analysis

● QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.106099.

ACKNOWLEDGMENTS

This work was supported by Guangzhou Planned Project of Science and Technology (202102010030) and the National Key Research and Development Program of China (2021YFC2009400).

AUTHOR CONTRIBUTIONS

All the authors contributed extensively to the work presented in this article. Zhu G and Luo OJ: participated in study design and coordination, material support for obtained funding, article modification and supervised study. Zhang B performed most of the experiments, statistical analysis and drafted the article. Ren Z participated in the bioinformatics analysis. Zheng H and Lin M: participate in the clinical sample analysis. Chen G: modified the article. All authors read and approved the final article.

DECLARATION OF INTERESTS

No potential conflicts of interest were disclosed.

Received: June 10, 2022
Revised: December 20, 2022
Accepted: January 25, 2023
Published: February 2, 2023

REFERENCES

1. El-Serag, H.B. (2011). Hepatocellular carcinoma. N. Engl. J. Med. *365*, 1118–1127.
2. Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., and Bray, F. (2021). Global cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA. Cancer J. Clin. *71*, 209–249.
3. Heimbach, J.K., Kulik, L.M., Finn, R.S., Sirlin, C.B., Abecassis, M.M., Roberts, L.R., Zhu, A.X., Murad, M.H., and Marrero, J.A. (2018). AASLD guidelines for the treatment of hepatocellular carcinoma. Hepatology *67*, 358–380.
4. Zheng, Z., Liang, W., Milgrom, D.P., Zheng, Z., Schroder, P.M., Kong, N.S., Yang, C., Guo, Z., and He, X. (2014). Liver transplantation versus liver resection in the treatment of hepatocellular carcinoma: a meta analysis of observational studies. Transplantation *97*, 227–234.
5. Shah, S.A., Smith, J.K., Li, Y., Ng, S.C., Carroll, J.E., and Tseng, J.F. (2011). Underutilization of therapy for hepatocellular carcinoma in the Medicare population. Cancer *117*, 1019–1026.
6. Sonnenday, C.J., Dimick, J.B., Schulick, R.D., and Choti, M.A. (2007). Racial and geographic disparities in the utilization of surgical therapy for hepatocellular carcinoma. J. Gastrointest. Surg. *11*, 1636–1646.
7. Ayoub, W.S., Steggerda, J., Yang, J.D., Kuo, A., Sundaram, V., and Lu, S.C. (2019). Current status of hepatocellular carcinoma detection: screening strategies and novel biomarkers. Ther. Adv. Med. Oncol. *11*. 1758835919869120.
8. Altekruse, S.F., McGlynn, K.A., and Reichman, M.E. (2009). Hepatocellular carcinoma incidence, mortality, and survival trends in the United States from 1975 to 2005. J. Clin. Oncol. *27*, 1485–1491.
9. Valastyan, S., and Weinberg, R.A. (2011). Tumor metastasis: molecular insights and evolving paradigms. Cell *147*, 275–292.
10. Vanharanta, S., and Massagué, J. (2013). Origins of metastatic traits. Cancer Cell *24*, 410–421.
11. Caobi, A., Dutta, R.K., Garbinski, L.D., Esteban-Lopez, M., Ceyhan, Y., Andre, M., Manevski, M., Ojha, C.R., Lapierre, J., Tiwari, S., et al. (2020). The impact of CRISPR-Cas9 on age-related disorders: from pathology to therapy. Aging Dis. *11*, 895–915.
12. Hart, T., Chandrashekhar, M., Aregger, M., Steinhart, Z., Brown, K.R., MacLeod, G., Mis, M., Zimmermann, M., Fradet-Turcotte, A., Sun, S., et al. (2015). High-resolution CRISPR screens reveal fitness genes and genotype-specifc cancer liabilities. Cell *163*, 1515–1526.
13. Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelson, T., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G., and Zhang, F. (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. Science *343*, 84–87.
14. Chen, S., Sanjana, N.E., Zheng, K., Shalem, O., Lee, K., Shi, X., Scott, D.A., Song, J., Pan, J.Q., Weissleder, R., et al. (2015). Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. Cell *160*, 1246–1260.
15. Manguso, R.T., Pope, H.W., Zimmer, M.D., Brown, F.D., Yates, K.B., Miller, B.C., Collins, N.B., Bi, K., LaFleur, M.W., Juneja, V.R., et al. (2017). In vivo CRISPR screening identifies Ptpn2 as a cancer immunotherapy target. Nature *547*, 413–418.
16. Liu, Y., Yu, C., Daley, T.P., Wang, F., Cao, W.S., Bhate, S., Lin, X., Still, C., 2nd, Liu, H., Zhao, D., et al. (2018). CRISPR activation screens systematically identify factors that drive neuronal fate and reprogramming. Cell Stem Cell *23*, 758–771.e8.
17. Bester, A.C., Lee, J.D., Chavez, A., Lee, Y.R., Nachmani, D., Vora, S., Victor, J., Sauvageau, M., Monteleone, E., Rinn, J.L., et al. (2018). An integrated genome-wide CRISPRa approach to functionalize lncRNAs in drug resistance. Cell *173*, 649–664.e20.
18. Alda-Catalinas, C., Bredikhin, D., Hernando-Herraez, I., Santos, F., Kubinyecz, O., Eckersley-Maslin, M.A., Stegle, O., and Reik, W. (2020). A single-cell transcriptomics CRISPR-activation screen identifies epigenetic regulators of the zygotic genome activation Program. Cell Syst. *11*, 25–41.e9.
19. Joung, J., Konermann, S., Gootenberg, J.S., Abudayyeh, O.O., Platt, R.J., Brigham, M.D., Sanjana, N.E., and Zhang, F. (2017). Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. Nat. Protoc. *12*, 828–863.
20. Lu, Y., Shen, H., Huang, W., He, S., Chen, J., Zhang, D., Shen, Y., and Sun, Y. (2021). Genome-scale CRISPR-Cas9 knockout screening in hepatocellular carcinoma with lenvatinib resistance. Cell Death Discov. *7*, 359.
21. Song, C.Q., Li, Y., Mou, H., Moore, J., Park, A., Pomyen, Y., Hough, S., Kennedy, Z., Fischer, A., Yin, H., et al. (2017). Genome-wide CRISPR screen identifies regulators of

mitogen-activated protein kinase as suppressors of liver tumors in mice. Gastroenterology *152*, 1161–1173.e1.

22. Wang, C., Vegna, S., Jin, H., Benedict, B., Lieftink, C., Ramirez, C., de Oliveira, R.L., Morris, B., Gadiot, J., Wang, W., et al. (2019). Inducing and exploiting vulnerabilities for the treatment of liver cancer. Nature *574*, 268–272.

23. Wang, G., Chow, R.D., Ye, L., Guzman, C.D., Dai, X., Dong, M.B., Zhang, F., Sharp, P.A., Platt, R.J., and Chen, S. (2018). Mapping a functional cancer genome atlas of tumor suppressors in mouse liver using AAV-CRISPR-mediated direct in vivo screening. Sci. Adv. *4*, eaao5508.

24. Liu, Z., and Thiele, C.J. (2017). When LMO1 meets MYCN, neuroblastoma is metastatic. Cancer Cell *32*, 273–275.

25. Oldridge, D.A., Wood, A.C., Weichert-Leahey, N., Crimmins, I., Sussman, R., Winter, C., Diamond, M., Hart, L.S., Zhu, S., , et al.McDaniel, L.D. (2015). Genetic predisposition to neuroblastoma mediated by a LMO1 super-enhancer polymorphism. Nature *528*, 418–421.

26. Kim, D.H., Ahn, J.S., Han, H.J., Kim, H.M., Hwang, J., Lee, K.H., Cha-Molstad, H., Ryoo, I.J., Jang, J.H., Ko, S.K., et al. (2019). Cep131 overexpression promotes centrosome amplification and colon cancer progression by regulating Plk4 stability. Cell Death Dis. *10*, 570.

27. Ding, W., You, H., Dang, H., LeBlanc, F., Galicia, V., Lu, S.C., Stiles, B., and Rountree, C.B. (2010). Epithelial-to-mesenchymal transition of murine liver tumor cells promotes invasion. Hepatology *52*, 945–953.

28. Yao, X., Hu, J.F., Daniels, M., Yien, H., Lu, H., Sharan, H., Zhou, X., Zeng, Z., Li, T., Yang, Y., and Hoffman, A.R. (2003). A novel orthotopic tumor model to study growth factors and oncogenes in hepatocarcinogenesis. Clin. Cancer Res. *9*, 2719–2726.

29. Acker, J., Murroni, O., Mattei, M.G., Kedinger, C., and Vigneron, M. (1996). The gene (POLR2L) encoding the hRPB7.6 subunit of human RNA polymerase. Genomics *32*, 86–90.

30. Zhou, D., Li, X., Zhao, H., Sun, B., Liu, A., Han, X., Cui, Z., and Yuan, L. (2018). Combining multi-dimensional data to identify a key signature (gene and miRNA) of cisplatin-resistant gastric cancer. J. Cell. Biochem. *119*, 6997–7008.

31. Guo, L., Peng, Y., Sun, L., Han, X., Xu, J., and Mao, D. (2017). Ovarian cancer variant rs2072590 is associated with HOXD1 and HOXD3 gene expression. Oncotarget *8*, 103410–103414.

32. Zhao, Y., and Wang, X. (2019). PLK4: a promising target for cancer therapy. J. Cancer Res. Clin. Oncol. *145*, 2413–2422.

33. Tarek, M.M., Shafei, A.E., Ali, M.A., and Mansour, M.M. (2018). Computational prediction of vaccine potential epitopes and 3-dimensional structure of XAGE-1b for non-small cell lung cancer immunotherapy. Biomed. J. *41*, 118–128.

34. Yıldız Bölükbaşı, E., Shabbir, R.M.K., Malik, S., and Tolun, A. (2021). Homozygous deletion of MYADML2 in cranial asymmetry, reduced bone maturation, multiple dislocations. lumbar lordosis, and prominent clavicles. J. Hum. Genet. *66*, 171–179.

35. Bai, R., Li, Y., Jian, L., Yang, Y., Zhao, L., and Wei, M. (2022). The hypoxia-driven crosstalk between tumor and tumor-associated macrophages:mechanisms and clinical treatment strategies. Mol. Cancer *21*, 177.

36. Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq-a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-MYADML2 antibody	Bioss	Cat# bs-19119R
Anti-Vimentin antibody	Servicebio	Cat# GB11192
Anti-MMP9 antibody	Bioss	Cat# bs-4593R
Anti-VEGFa antibody	Abcam	Cat# ab1316; RRID: AB_2997338
Anti-MYADML2 antibody	Novus Biologicals	Cat# NBP2-85342
Anti-PLK4 antibody	Abclonal	Cat# A9863
Anti-LMO1 antibody	Abclonal	Cat# A7561
Anti-XAGE1B antibody	Novus Biologicals	Cat# NB100-262SS
Biological samples		
Human HCC specimens	Wellbiotech	Cat# LVC1607
Chemicals, peptides, and recombinant proteins		
Blasticidin	MDBio	Cat# D0120601
Hieff TMqPCR SYBR Green Master Mix (LoW Rox Plus)	YEASEN	Cat# 11202ES08
Fetal bovine serum	Gibco	Cat# 1027-106
MEM	Gibco	Cat# 31985070
DMEM	HyClone	Cat# SH30022-01B
Critical commercial assays		
HiPure Tissue DNA Mini Kit	Magen	Cat# D3125-02
HiPure Gel Pure DNA Mini Kit	Magen	Cat# D2111-03
High-purity total RNA rapid extraction kit (spin column type)	BIOTEKE	Cat# RP1201
CCK8 kits	Beyotime	Cat# C0039
Universal Virus Concentration Kit	Beyotime	Cat# C2901M
Experimental models: Cell lines		
HEK293T	Procell	RRID:CVCL_0063
HEPG2	Procell	Cat# CL-0103; RRID:CVCL_C5RP
HUH7	Procell	Cat# CL-0120; RRID:CVCL_U443
Experimental models: Organisms/strains		
Balb/c nude	Experimental Animal Center of Sun Yat-sen University	SCXK (Yue) 2016-0029
Oligonucleotides		
siRNA targeting sequence (Table 2)	This paper	N/A
primer sequences for PCR (Table 1)	This paper	N/A
sgRNA targeting sequence (Table S5)	This paper	N/A
Recombinant DNA		
CRISPR/Cas9 SAM pooled library plasmids	Addgene	Cat#1000000078; RRID:Addgene_164896
MS2-P65-HSF1 activator plasmid	Addgene	Cat# 89308; RRID:Addgene_92120
pCMV-VSV-G plasmid	GENEROL BIOL	N/A
pMDLg pRRE plasmid	GENEROL BIOL	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
pRSV-Rev plasmid	GENEROL BIOL	N/A
sgC17orf99	GENEROL BIOL	N/A
sgHOXD1	GENEROL BIOL	N/A
sgLMO1	GENEROL BIOL	N/A
sgMYADML2	GENEROL BIOL	N/A
sgPLK4	GENEROL BIOL	N/A
sgPOLR2L	GENEROL BIOL	N/A
sgXAGE1B	GENEROL BIOL	N/A
siMYADML2	GENEROL BIOL	N/A
siPLK4	GENEROL BIOL	N/A
siLMO1	GENEROL BIOL	N/A
siXAGE1B	GENEROL BIOL	N/A
Software and algorithms		
Prism version 8	Graphpad	https://www.graphpad.com/
SPSS Statistics 26	IBM	https://www.ibm.com/
DESeq2 v1.30.1	R v4.0.4	https://github.com/mikelove/DESeq2
survival v3.2.10	R v4.0.4	https://github.com/therneau/survival
survminer v0.4.9	R v4.0.4	https://github.com/kassambara/survminer/issues
ggplot2 v3.3.3	R v4.0.4	https://github.com/tidyverse/ggplot2/issues

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfifilled by the corresponding author, Professor.Guodong Zhu (eyzhugd@scut.edu.cn).

Materials availability

This study did not generate new unique reagents. Primers and siRNA sequences used were provided in Tables 1 and 2, and available upon request to the corresponding author.

Data and code availability

- The data reported in this paper will be shared upon request to the lead corresponding author (eyzhugd@scut.edu.cn).
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell lines

The human liver cancer cell lines HEPG2 and HUH7 were obtained from Procell Life Science & Technology Co.,Ltd (Wu Han, China) and was cultured in Minimum Eagle's medium (MEM, Gibco) or Dulbecco's modified Eagle's medium (DMEM, HyClone), with 10% fetal bovine serum (FBS, Gibco) and 0.1% Penicillin/Streptomycin (P/S, TBD, PS2004HY) in a humidified incubator with 5% CO2 at 37°C.

Animal model

The purpose is to establishment of an *in situ* liver cancer mouse model. In the study, sixteen 4-6-week-old male BLAB/c nu/nu *Mus musculus*, weighed 15 to 18g, purchased from the Experimental Animal Center of Sun Yat-sen University (the certificate number SCXK (Yue) 2016-0029) were utilized for the animal experiments described. The study was approved by the Institutional Review Board of Guangzhou Forevergen Biosciences (IACUC-AEWC-F1902026). All mice were housed in groups, given free access to standard rodent diet and water and were randomly assigned to experimental groups. For orthotopic tumor model, BLAB/c nu/nu mice were directly injected with 4×10^6 of lentiSAM2 pooled library-transduced HEPG2 cells through intrahepatic. After 7 weeks post-transplantation, mice were sacrificed and liver, lung organs were dissected for examination of metastases.

Tissue samples

In the study, 80 HCC tissues samples was used to detect MYADML2 protein level by immunohistochemistry.The clinical and pathological details of these patients are shown in Table S4.

METHOD DETAILS

Lentivirus preparation of sgRNAs library

HEK293T cells were seeded in a 15 cm dish with DMEM medium the day prior to transfection. And then, co-transfected with CRISPR/Cas9 SAM2 pooled library plasmids (Addgene, #1000000078) or MS2-P65-HSF1 activator plasmids (Addgene plasmid #89308), pCMV-VSV-G, pMDLg pRRE and pRSV-Rev (3:1:1:1). After 6 h, the medium was replaced with fresh DMEM medium. Viral particles were harvested 48 h and 72h later after transfection. The Universal Virus Concentration Kit (Beyotime, C2901M) was used to enrich the concentrated virus.

Lentiviral transduction

HEPG2 cells were transfected with pooled CRISPR/Cas9 SAM2 human lentiviral library which contains 70,290 unique sgRNA sequences targeting 23,430 human genes at a low MOI (~0.5). Followed selected with blasticidin (MDBio, D0120601) for 15 days to generate a mutant cell pool. Polymerase chain reaction (PCR) was applied to identify the successful transfection of lentiCRISPRa vector in HEPG2 cells. CRISPRa-F: TCTTGTGGAAAGGACGAAACACCG and CRISPRa-R: CTCCTTTCAAGACCTAGGATC were selected for PCR amplification region (209 bp) from lentiCRISPRa vector, followed by electrophoresis. The thermocycling parameters of PCR were 95°C for 60 s, 20 cycles of (95°C for 10 s, 56°C for 10 s, 72°C for 30 s) and 72°C for 1 min.

Genomic DNA extraction

Genomic DNA from 200 mg tissues (primary tumors and lungs) was extracted using the HiPure Tissue DNA Mini Kit (Magen). The gDNA concentration was measured using a Nanodrop (Thermo Scientific). The sgRNA library readout was performed using two steps of polymerase chain reaction (PCR) and the first PCR includes enough genomic DNA to preserve full library complexity.[14] For PCR#1, a region containing sgRNA cassette was amplified using primers specific to the sgRNA-expression vector. The primer of PCR#1: CRISPRa-F (TCTTGTGGAAAGGACGAAACACCG) and CRISPRa-R (CTCCTTTCAAGACCTAGGATC). The thermocycling parameters were: 95°C for 60 s, 20 cycles of (95°C for 10 s, 60°C for 10 s, 72°C for 30 s) and 72°C for 1 min. In the study, 3 mg gDNA were used in each PCR#1 reaction and100 mg of gDNA was used per sample. The second PCR was used to add appropriate sequencing adapters to the first PCR products. After the PCR products were electrophoresed, HiPure Gel Pure DNA Mini Kit was used for Gel Extraction.

sgRNA deep sequencing

DNA fragments were size selected using agarose gel and sequenced using a 209 bp run on the HiSeq2500 (Illumina Inc., San Diego, CA, USA). The counts of sgRNA were summarized using htseq.[36]

Establishment of candidate genes overexpression HCC cells

The candidate genes C17orf99, HOXD1, LMO1, MYADML2, PLK4, POLR2L, and XAGE1B were selected for validation.The sgRNA plasmids of the candidate genes were synthesized (GENEROL BIOL) and used for HCC cells transfection.The sgRNA sequence was shown in Table S5. Lentivirus was produced in 293T cells

and 9 μg sgRNA plasmid, 3 μg pCMV-VSVG plasmid, 3 μg pMDLg pRRE plasmid and 3 μg pRSV-Rev plasmid were used. The HEPG2 and Huh7 cell was transfected with the sgRNA lentivirus with the MOI 10 and 5, respectively. Then, the cells were selected with blasticidin (MDBio, D0120601).

siRNA transfection

HEPG2 and Huh7 cells were seeded in 6-well plates and transient transfection with LMO1, MYADML2, PLK4 and XAGE1B siRNA and control siRNA when the cells had reached 60% confluency. Briefly, 6 μL siRNA was pre-mixed with 100 μL Opti-MEM (Gibco). Then, 8 μL Lipo8000 transfection reagent (Beyotime, C0533) was added into 100 Opti-MEM and dropped into the plasmid mixture. The complex were incubated for 20 min at room temperature and then added to HEPG2 and Huh7 cell culture.Transfection media was removed 5 h later and replaced with DMEM or MEM complete medium. The siRNA were obtained from GENERAL BIOL (Chuzhou, China). The siRNA sequences were shown in Table 2.

Real-time quantitative reverse transcriptase PCR

The C17orf99, HOXD1, LMO1, MYADML2, PLK4, POLR2L, XAGE1B mRNA levels were detected by real-time quantitative reverse transcriptase PCR (q-PCR). Briefly, total RNA was extracted using the total RNA Rapid Extraction Kit (BIOTEKE, RP1201) according to the manufacturer's instructions. The SYBR Green ® Real-time PCR Master Mix (#QPK-201, Toyobo Co, Ltd, Osaka) was used for qRT-PCR assays. The data were analyzed on an Applied Biosystems 7900 Real-Time PCR System. The target genes were normalized to the mean β-actin expression.The PCR primers used were shown in Table 1.

Cell proliferation assay

We detected the overexpression of C17orf99, HOXD1, LMO1, MYADML2, PLK4, POLR2L, XAGE1B on the proliferation of liver cancer cells by CCK8 assays (Beyotime, Shanghai, China). Briefly, 5,000 cells were inoculated to 96-well plates and the OD values were detected at 4h, 24h, 48h and 72h according to the manufacturer's instructions.

Migration assay

For transwell (Corning) migration assay, the cells were resuspended in serum-free media and added to the upper compartment of the chamber, and 600 mL of medium (containing 20% FBS) was added to the lower chamber. The cells were incubated in a humidified environment with 95% air and 5% CO2 at 37 °C and allowed to migrate for 24 h. After removal of the non-migrated cells, the cells that had migrated through the filter were stained with crystal violet, photographed, and counted using ImageJ (National Institutes of Health, Bethesda, MD, USA).

Mouse tissue collection

Primary tumors and lung were dissected manually. Homogenized tissues were used for DNA/RNA extractions using standard molecular biology protocols. Tissues for histology were then fixed in 4% formaldehyde overnight, embedded in paraffin, and sectioned at 6 mm with a microtomeas. Slices were subjected to H&E staining and an optical microscope was used for overall histological assessment.

Immunohistochemistry

The MYADML2 (bs-19119R, BIOSS, 1:400) protein expression in the HCC tissues was detected by immunohistochemistry. Human HCC specimens were scored in a semi-quantitative manner due to the heterogeneity of the staining of MYADML2 proteins. The percentage was grouped as follows: 0–25% = 1, 25–50% = 2, 50–75% = 3 and 75–100% = 4. The staining intensity was categoried as follows: none = 0, weak = 1, moderate = 2, and strong = 3. A final immunoreactivity scores (IRS) was obtained for each case by products of the percentages and the intensity scores. The case information including clinical and follow-up data was shown in Table S1. Vimentin (GB11192, Servicebio, 1:400), MMP9 (bs-4593R, BIOSS, 1:200) and VEGFa (ab1316, abcam, 1:1000) protein levels in liver and lung tissue were detected by immunohistochemistry.

Western blot

Cells were lysed with RIPA buffer (Beyotime, P0013B) for total lysate extraction. Then, 10 μg proteins was used for electrophoresis with 10% SDS-PAGE and then transferred to a polyvinylidene difluoride membrane. The membrane was first blocked by 5% BSA for 1.5 h at room temperature, followed by incubation at 4 °C overnight with the following primary antibodies: Anti-MYADML2 antibody (Novus Biologicals,

NBP2-85342), Anti-PLK4 antibody (Abclonal, A9863), Anti-LMO1 antibody (Abclonal, A7561), Anti-XAGE1B antibody (Novus Biologicals, NB100-262SS) and GAPDH (Proteintech, HRP-60004). After washing three times with TBST, the membrane was incubated with antimouse and antirabbit horseradish peroxidaseconjugated secondary antibodies (boster, BA1054 and BA1051) for 2 h at room temperature.

TCGA data analysis

For liver cancer types (424 samples) in TCGA (hepatocellular carcinoma, LIHC), mRNA-seq and clinical data were retrieved from UCSC Xena (https://xenabrowser.net/datapages/) and convert FPKM data to TPM with the formula: exp (log(fpkm**2-1) - log(sum(fpkm**2-1)) + log(1e6)). Finally, 365 patients with mRNA expression data and prognosis information were divided into high expression group and low expression group according to median, and then Kaplan-Meier survival analysis and univarate Cox hazard analysis were performed. p value less than 0.05 was considered as prognostic.

Evaluating the sensitivity of chemotherapeutic agents

To examine the sensitivity of each sample in TCGA to different chemotherapeutic agents the Genomics of Drug Sensitivity in Cancer (GDSC, https://www.cancerrxgene.org/) database were used. The half-maximal inhibitory concentration (IC50) of chemotherapy drugs for the comparison of the high- and low-MYADML2 groups in training cohort was implemented by R package "pRRophetic"(3.6.1).

Immune cell infiltration analysis

Cell-type Identification by Estimating Relative Subsets of RNA Transcripts (CIBERSORT,http://cibersort.stanford.edu) was used to characterize the infiltration of 22 kinds of immune cell types with the RNA expression profile of each patient A. We obtained the abundance ratio matrix of 15 immune cell types with the criteria of $p < 0.05$. We also compared the different infiltration levels of 15 immune cell types in low and high XAGE1B, PLK4, LMO1 and MYADML2 expression by "ggplot2" packages.

QUANTIFICATION AND STATISTICAL ANALYSIS

Our data were expressed as means $\pm$ SD Analyses were performed using IBM SPSS Statistics for Windows, version 26.0 (IBM Corp., Armonk, NY, USA). Comparisons between groups were analyzed through one-way ANOVA or unpaired Student's *t* test. P-values less than 0.05 were considered statistically significant.