

## ARTICLE OPEN



# DNA methylation-activated full-length EMX1 facilitates metastasis through EMX1-EGFR-ERK axis in hepatocellular carcinoma

Dong-sheng Wen<sup>1,3</sup>, Li-chang Huang<sup>1,3</sup>, Xiao-yun Bu<sup>2,3</sup>, Min-ke He<sup>1</sup>, Zhi-cheng Lai<sup>1</sup>, Ze-feng Du<sup>1</sup>, Ye-xing Huang<sup>1</sup>, Anna Kan<sup>1</sup> and Ming Shi<sup>1</sup>

© The Author(s) 2023

Altered DNA methylation is a crucial epigenetic event in hepatocellular carcinoma (HCC) development and progression. Through methylation-transcriptomic analysis, we identified a set of sixty potential DNA methylation-based epigenetic genes. In this set of genes, we focused on the hypermethylation of EMX1, which is frequently observed in hepatobiliary tumors. Despite its frequent occurrence, the function of EMX1 remains largely unknown. By utilizing bisulfite-next-generation sequencing, we have detected EMX1 DNA hypermethylation on the gene body, which is positively correlated with EMX1 mRNA expression. Further analysis revealed that EMX1 mRNA terminal exon splicing in HCC generated two protein isoforms: EMX1 full length (EMX1-FL) and alternative terminal exon splicing isoform (EMX1-X1). Cellular functional assays demonstrated that gain-of-function EMX1-FL, but not EMX1-X1, induced HCC cells migration and invasion while silencing EMX1-FL inhibited HCC cells motility. This result was further validated by in vivo tumor metastasis models. Mechanistically, EMX1-FL bound to EGFR promoter, promoting EGFR transcription and activating EGFR-ERK signaling to trigger tumor metastasis. Therefore, EGFR may be a potential therapeutic target for EMX1-high expression HCC. Our work illuminated the crucial role of gene body hypermethylation-activated EMX1-FL in promoting tumorigenesis and metastasis in HCC. These findings pave the way for targeting the EMX1-EGFR axis in HCC tumorigenicity and metastasis.

*Cell Death and Disease* (2023)14:769; <https://doi.org/10.1038/s41419-023-06293-y>

## INTRODUCTION

Hepatocellular carcinoma (HCC) is considered one of the most fatal cancers and the fastest increasing malignancy [1]. Even though advanced therapies in regional or systemic have been developed [2], the 5-year survival rate for overall HCC remains low (~20%) [1]. To a great extent, the dismal outcomes can be attributed to the high incidence of intra- and extrahepatic metastasis [3]. Finding novel potential molecular targets that could decelerated HCC metastasis is therefore an urgent matter.

The landscape of genome-wide DNA methylation has proven to be a reliable biomarker for HCC epigenetic subtyping and classification [4]. Methylation analysis of tumor tissue DNA or circulating tumor DNA has been shown to be an effective clinical prognostic predictor in primary and recurrence HCC [5–8]. However, the mechanisms of methylation-based epigenetic drivers in HCC remain unclear. While many studies focused on DNA methylation on gene promoters, methylation occurred throughout the genome, including neighboring CGI shores and shelves, intergenic non-coding regions, and gene bodies [9]. Additionally, gene body methylation has been observed to be positively

correlated with gene expression, chromatin modifications and RNA alternative splicing, though its function and targeting properties are largely unknown [9–12]. A recent study revealed that ADAMTSL5 overexpression in HCC was positively linked with gene body hypermethylation and conferred tumorigenesis and drug resistance [13]. These findings suggest that more targeted therapies might be discovered by re-evaluating the relationship between gene methylation and expression regulation and by re-evaluating the mechanisms of DNA methyltransferase inhibitors (DNMTi) in HCC treatment [14]. Further investigation into HCC DNA methylation-based epigenetic drivers, particularly in the gene body regions, is therefore a promising field.

Under normal circumstances, the Empty Spiracles Homeobox genes, which include paralog gene EMX1 and EMX2, are restricted in the developmental period of the cerebral cortex [15]. EMX1 is deemed as a marker for pyramidal neurons in the cerebral cortex [16] and widely used in the *Cre-LoxP* system to target neurons [17, 18]. Downregulation of the EMX2 gene has been linked to the development of several solid tumors, such as gastric cancer, lung cancer, endometrial cancer and colorectal cancer, while restoration of

<sup>1</sup>Department of Hepatobiliary Oncology, Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Guangdong Provincial Clinical Research Center for Cancer, Guangzhou 510060, P. R. China. <sup>2</sup>Department of Colorectal Surgery, Hunan Cancer Hospital and the Affiliated Cancer Hospital of Xiangya School of Medicine, Central South University, Changsha 410013, P. R. China. <sup>3</sup>These authors contributed equally: Dong-sheng Wen, Li-chang Huang, Xiao-yun Bu. ✉email: [annakan@sysucc.org.cn](mailto:annakan@sysucc.org.cn); [shiming@sysu.edu.cn](mailto:shiming@sysu.edu.cn)

Edited by Professor Stephen Tait

Received: 12 May 2023 Revised: 1 November 2023 Accepted: 13 November 2023

Published online: 25 November 2023

EMX2 suppresses cancer progression [19–23]. However, the biological function and detailed molecular mechanism of EMX1 in cancer remain unclear. Despite a few studies focused on EMX1 DNA hypermethylation in hepatobiliary tumors recently [6, 24], its underlying mechanisms have not yet been elucidated.

Here, we have identified EMX1 as a DNA epigenetically upregulated oncogene in HCC. Our analysis revealed an unreported EMX1 3' splice variant. Specifically, while the EMX1 full length (EMX1-FL) promotes HCC migration and invasion in vitro and metastasis in vivo, this is not observed with the EMX1-X1 (an alternative terminal exon splicing isoform). Additionally, we demonstrated that EMX1-FL activated the transcription of EGFR and upregulated the EGFR-ERK signaling. As a result, this study highlights EMX1 as both the HCC epidriver marker and a potential therapeutic target.

## RESULTS

### EMX1 gene body is hypermethylated in HCC

To screen out HCC potential DNA methylation driver genes, we analyzed the TCGA-LIHC DNA methylome and transcriptome. Finally, 60 potential epidriver genes were selected (Supplementary Fig. S1A). These epidrivers mainly enriched in cell differentiation and Homeobox genes (Fig. 1A). Among the genes, six Homeobox family genes (EMX1, OTX1, HOXA10, PITX1, TLX1 and DLX5) consistently expressed higher in HCC tumors (Supplementary Fig. S1B). Four genes (OTX1, HOXA10, TLX1 and DLX5) were reported as oncogenes in multiple tumors [25–28], while PITX1 was extensively reported as tumor suppressor gene [29, 30]. Remarkably, EMX1 was recognized recently as a potential clinically available epi-marker in hepatobiliary tumors [6, 24], but with undefined mechanism and function in HCC.

Next, we assembled four additional methylation validation datasets (GSE112791, GSE113017, GSE113019 and GSE89852) along with TCGA-LIHC to further explore EMX1 aberrant methylation in HCC. We identified the most different methylation probes between tumor and adjacent normal liver tissue (ANLT) (Fig. 1B and Supplementary Fig. S1C). Surprisingly, these probes were located in the region of EMX1 CpG island of intron 1, exon 2 and a predictive alternative exon 3 (namely R2, R3 and R4 respectively), but not the promoter (R1) (the genomic positions of R1–R4 are marked in Supplementary Fig. S1D). We further conducted bisulfite sequencing PCR (BSP) in 12 paired fresh HCC tumors and ANLT on R1–R4 for confirmation (Fig. 1C). The methylation CpG sites that significantly amplified located in R2 ( $P < 0.001$ ), while R3 ( $P < 0.05$ ) and R4 ( $P < 0.01$ ) also showed comparatively higher methylation in tumor. The promoter (R1), however, showed unmethylated status in both tumor and ANLT (Fig. 1D). Methylation-specific PCR (MSP) assay in HCC tumor and ANLT further supported these findings (Supplementary Fig. S1E). We then analyzed the DNA methylation profile of matched samples including primary HCC tumor, portal vein tumor thrombus and ANLT from 20 HCC patients. Probe cg06498720 (locate in R2) showed the highest methylation in portal vein tumor thrombus, compared to both tumor and ANLT (Supplementary Fig. S1F). Taken together, EMX1 gene body hypermethylation is considered as an epidriver that may contribute to HCC tumorigenesis.

### Hypermethylation in gene body upregulated EMX1 expression in HCC

After analyzing BSP sequencing results, we noticed positive correlations between EMX1 gene body methylation and mRNA expression following R2 ( $r = 0.80$ ,  $P < 0.01$ ), R4 ( $r = 0.70$ ,  $P = 0.01$ ) and R3 ( $r = 0.52$ ,  $P = 0.08$ ) in tumor (Fig. 1E). However, correlations were not observed in R1 ( $P = 0.30$ ) and ANLT (Supplementary Fig. S1G). This result was confirmed by TCGA-LIHC and GSE77269 methylation and RNA-seq analyses (<https://mexpress.be/index.html>) and Supplementary Fig. S1H). Then, we conducted

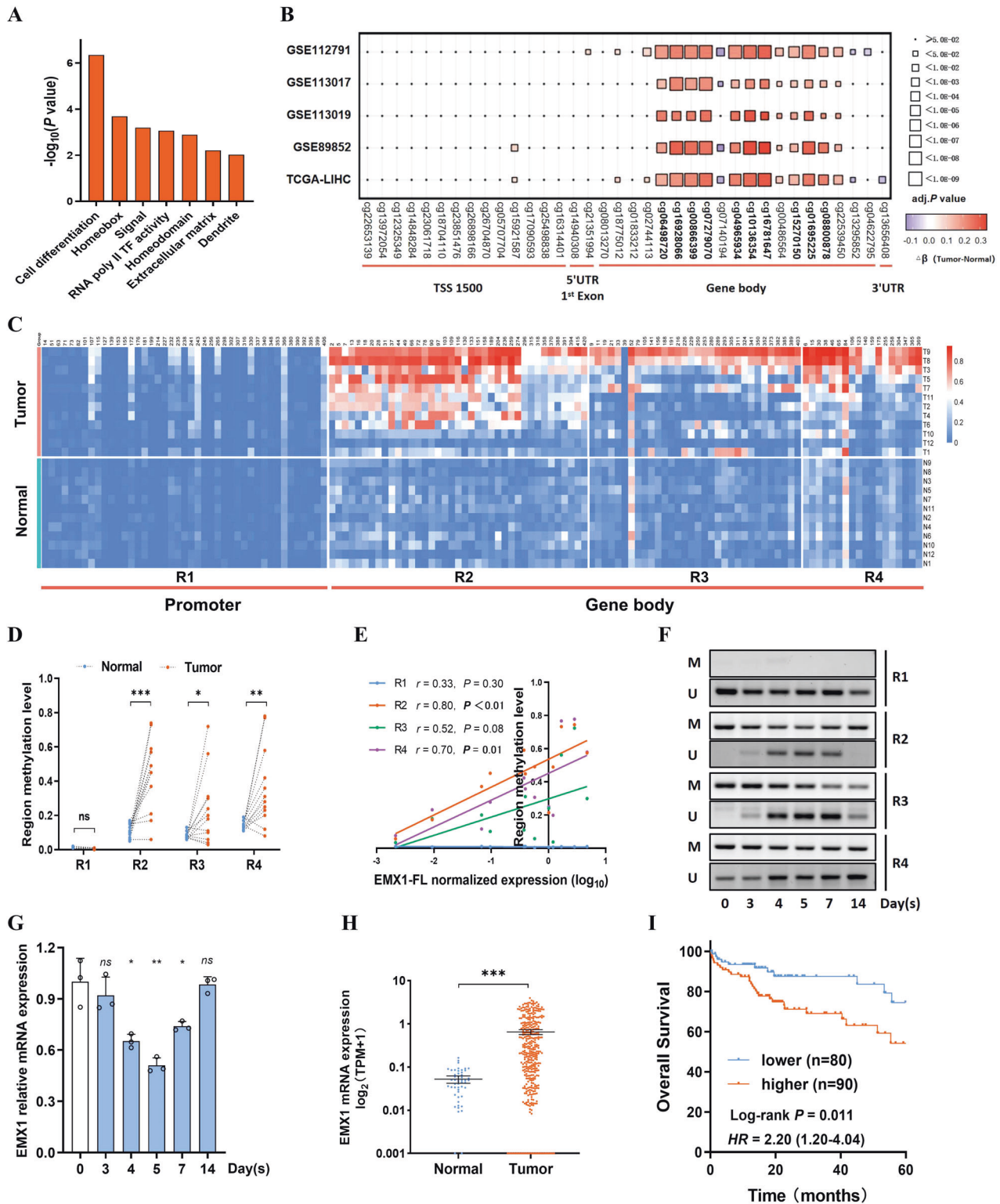
DNA demethylation and remethylation experiments by exposure (3 days) and withdrawal (within 2 weeks) of Decitabine (DNMTi) in SNU-398 cells for further verification. MSP analysis indicated that Decitabine increased EMX1 DNA demethylation at R2–R4 at day 3–7, concomitantly, EMX1 mRNA expression decreased during the first week. Conversely, DNA remethylation and EMX1 expression recovered in the withdrawal phase (Fig. 1F, G and Supplementary Fig. S1I). This was further confirmed by methylation bead chip and RNA-seq in SNU-398 and Hep-G2 treated by Guadecitabine (DNMTi) (Supplementary Fig. S1J, K) [14]. These results implied that EMX1 gene body DNA methylation at R2–R4 positively regulates EMX1 expression in HCC.

We then analyzed EMX1 expression level using patient samples. Both TCGA-LIHC cohort (Fig. 1H) and Fudan cohort (Supplementary Fig. S1L) showed that EMX1 total mRNA was significantly upregulated in HCC, while EMX2, an important paralog gene, was barely expressed (data not shown). The TCGA-LIHC iClust subgroups were extracted for further analyses. iClust1 subgroup was previously characterized as a “progenitor-like” subtype with more aggressive clinical features, such as vascular invasion tumors [4]. Interestingly, the expression of EMX1 was significantly higher in iClust1 compared to iClust2&3 (Supplementary Fig. S1M). Besides, higher expression of EMX1 indicated shorter overall survival ( $P = 0.011$ , TCGA-LIHC cohort, Fig. 1I) and recurrence-free survival ( $P = 0.011$ , SYSUCC cohort 1, Supplementary Fig. S1N) in early-stage HCC subgroup. In summary, hypermethylation of the EMX1 gene body facilitates EMX1 overexpression and potentially drives HCC tumorigenesis and progression.

### HCC exploits an EMX1 mRNA terminal exon splicing event that regulates its protein nucleus location

Western blotting was performed to examine the expression of EMX1 targeting the non-Homeodomain sequence (i.e., detecting potential EMX1 different isoforms). The results revealed that besides the prominent band (canonical EMX1-FL, NP\_004088, ~31.3 kDa), multiple bands potentially mapped to alternative isoforms were also detected, and EMX1 levels were overall higher in HCC tumor than ANLT (Fig. 2A). Immunohistochemistry (IHC) analyses indicated that EMX1 had different intracellular localization (both cytoplasm and nucleus) in HCC (Fig. 2B). Additionally, we performed IHC and RNA-seq on 30 HCC tissue samples from SYSUCC cohort 2, and verified that the total protein levels of EMX1 were consistent with the total mRNA levels ( $r = 0.78$ ,  $P < 0.0001$ , Fig. 2C and Supplementary Fig. S2A).

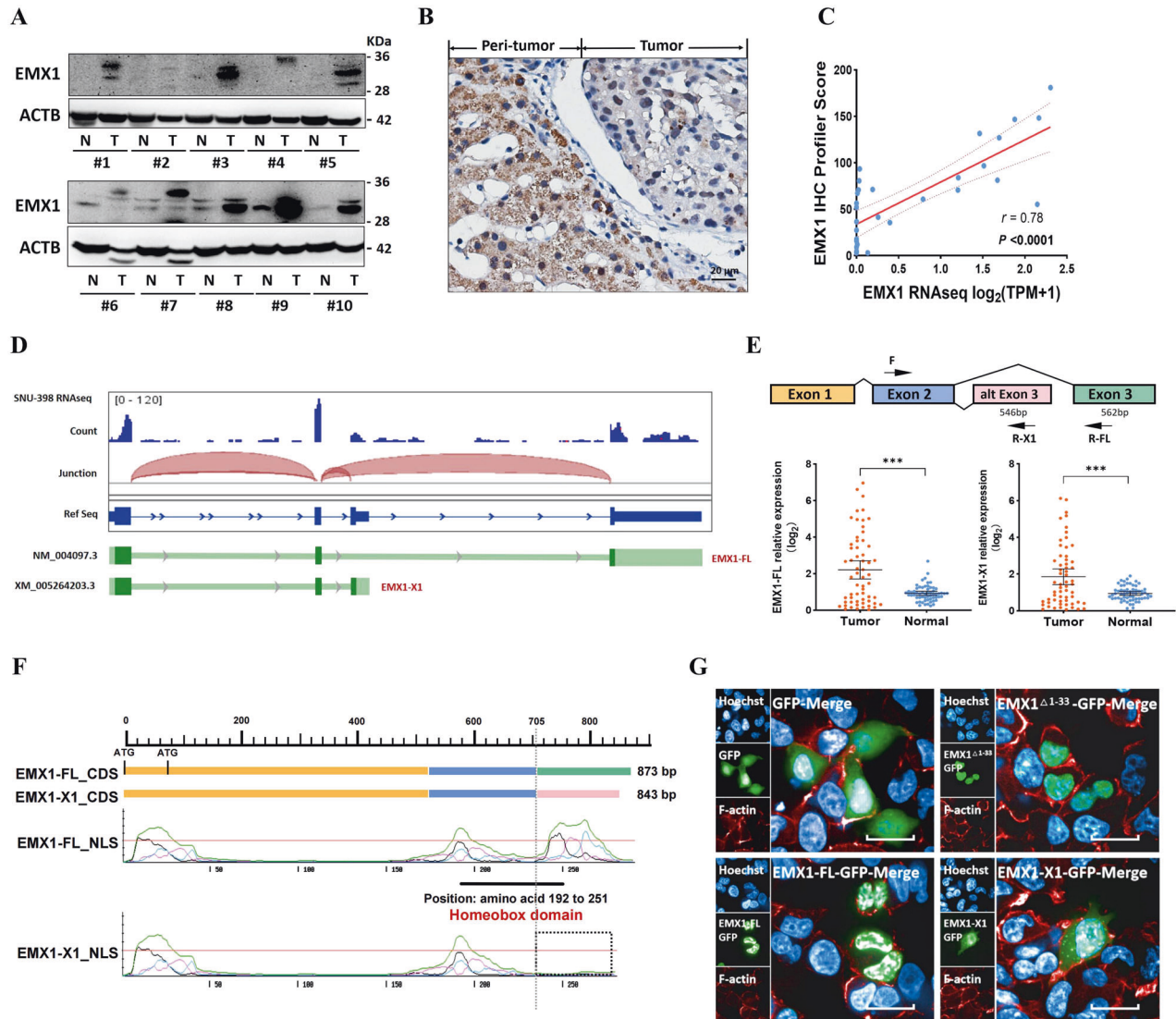
Previous reports have suggested that methylation in the gene body potentially stimulated transcription elongation and splicing outcomes [9, 11]. Upon retrieving expressed sequence tag databases (<https://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/>) [31], we noticed that region R4 is not only located in EMX1 intron 2 but can also be found on an alternative exon 3 (Supplementary Fig. S1D), indicating that EMX1 mRNA excites terminal exon splicing. Spliced alignment of RNA-seq was therefore performed in SNU-398, and the results showed a proximal alternative third exon that could substitute the third exon, altering the coding region of EMX1 protein at C-terminal (Fig. 2D). Further confirmation was done through 3'RACE sequencing in HCC tumor tissue cDNA (Supplementary Fig. S2B), where it was found that EMX1 mRNA 3' end had a consensual 562-bp-long terminal exon and a shorter alternative terminal exon (546 bp). Thus, two distinct EMX1 proteins can be generated: a canonical EMX1 full-length (EMX1-FL) isoform and an EMX1 isoform with alternative C-terminal end (EMX1-X1). Then the transcript-specific intron-spanning primers were designed for qRT-PCR (Fig. 2E upper panel and Supplementary Fig. S2C, D). The results from sixty paired samples showed that both of these two EMX1 transcripts were highly expressed in HCC tumor tissue compared to ANLT (Fig. 2E lower panel and Supplementary Fig. S2F), and displayed a high similarity of expression patterns ( $r = 0.95$ ,  $P < 0.0001$ ; Supplementary Fig. S2E).



After aligning the sequence, we found that EMX1-FL terminator exon coding part of the Homeodomain (between amino acid 192 to 251) along with the nuclear localization signals (NLS) [32] (Fig. 2F). Since annotation results suggesting that EMX1 might shuttle between nucleus and cytoplasm, we investigated whether the alternative terminator splicing (by truncating the terminal NLS) or potential alternative translation initial sites (ATG<sup>34</sup>, by truncating the

initial NLS) regulated its pivotal nuclear location. We constructed GFP-tagged control vector, EMX1-FL, EMX1-X1 and truncated EMX1<sup>Δ1-33</sup>, and transfected them into SNU-398. Immunofluorescence assays confirmed the exogenous expression and subcellular localization. EMX1-FL and EMX1<sup>Δ1-33</sup> (with entire Homeodomain) were predominantly localized in the nucleus, while EMX1-X1 particularly accumulated in the cytoplasm or vesicles (Fig. 2G).

**Fig. 1** **EMX1 is an epidriver and its gene body hypermethylation enhanced EMX1 expression in HCC.** **A** The DAVID functional annotation chart of sixty potential DNA epidriver genes. **B** The methylation sites of EMX1 were compared between tumor and ANLT in HCC methylation datasets (GSE112791, GSE113017, GSE113019, GSE89852 and TCGA-LIHC). The top ten most different probes were highlighted (cg06498720, cg16928066, cg00866399, cg07279070, cg04965934, cg10136354, cg16781647, cg15270150, cg01695225 and cg08800878; adj.  $P < 0.0001$  and  $|\beta_{\text{tumor}} - \beta_{\text{normal}}| > 0.25$ ). **C** The methylation status of individual CpG sites was quantified by BSP sequencing in HCC tumors and ANLT from R1 to R4 ( $n = 12$ ). R1 is located in the promoter region, R2–R4 are located in the gene body. R1–R4 consist of 44, 44, 42 and 23 CpG sites, respectively. **D** The overall methylation status from R1 to R4 in BSP sequencing ( $n = 12$ ). ns, no significant;  $*P < 0.05$ ;  $**P < 0.01$ ;  $***P < 0.001$ . **E** Scatterplot of the correlation between EMX1 regions methylation status and mRNA expression (normalized and  $\log_{10}$  transformed) in HCC tumor ( $n = 12$ ). **F, G** SNU-398 cells were treated with Decitabine (10  $\mu\text{M}$ ) for 3 days, and withdrawn for an extra 11 days. **F** DNA methylation status in R1–R4 was quantified using MSP (M, methylated; U, unmethylated). **G** Quantification of relative EMX1 mRNA expression using qRT-PCR. Data shown represent the means ( $\pm$ SD) of three biological replicates. ns, no significant;  $*P < 0.05$ ;  $**P < 0.01$ . **H** The EMX1 total mRNA expression was elevated in HCC tumor tissues derived from the TCGA-LIHC cohort. **I** Kaplan–Meier analysis of overall survival in early-stage HCC (AJCC Stage1,  $n = 170$ ) from TCGA-LIHC cohort based on EMX1 gene expression.

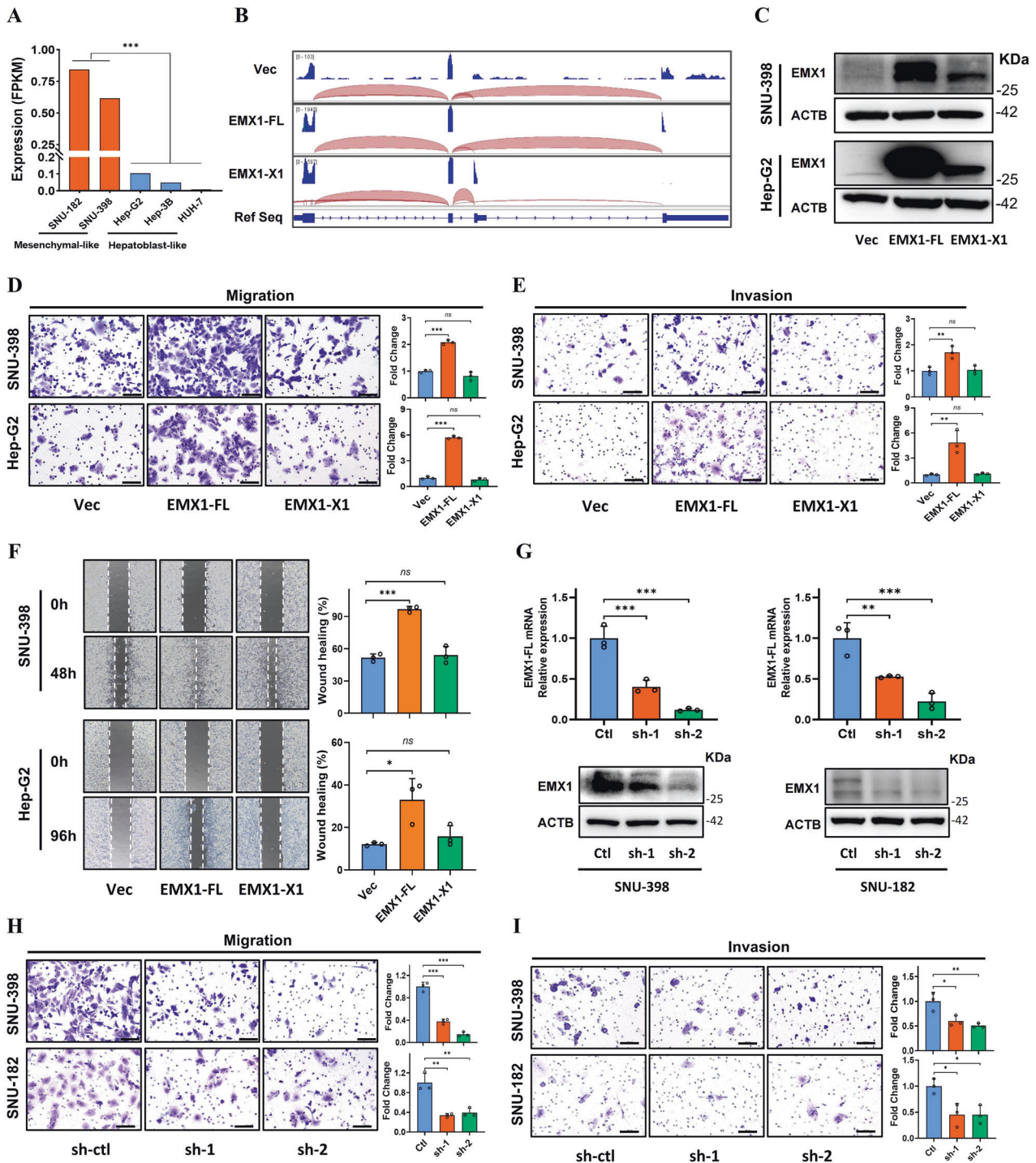


**Fig. 2** **HCC exploits an EMX1 mRNA terminal exon splicing event and regulates its nucleus location.** **A** EMX1 multiple isoforms in HCC tumor and ANLT samples detected by western blotting. **B** IHC with anti-EMX1 was performed in HCC samples. Scale bars, 20  $\mu\text{m}$ . **C** Scatterplot of the correlation between EMX1 total mRNA expression (RNA-seq) and total protein (IHC) in the SYSUCC cohort 2. **D** Generation of EMX1 isoforms by terminal exon splicing: EMX1-FL transcript consists of exons 1, 2 and canonical exons 3; EMX1-X1 transcript consists of exons 1, 2 and proximal alternative exons 3; The exons (green) were identified using SNU-398 RNA-seq and spliced alignment analysis. **E** The transcript-specific intron-spanning primers were designed (upper panel) for qRT-PCR and performed on paired HCC tumor and ANLT (SYSUCC cohort 1,  $n = 60$ ). The EMX1-FL expression was showed in lower left panel and the EMX1-X1 expression was showed in lower right panel. The mRNA expression data were shown to represent the means ( $\pm$ 95% confidence interval).  $***P < 0.001$ . **F** Graph of two EMX1 transcripts' coding sequence (CDS) and corresponding nuclear localization signal (NLS) predicted by NLStradamus. The EMX1-FL Homeobox domain is located in amino acid 192–251. **G** Subcellular localization of GFP-tagged vector, EMX1-FL, EMX1-X1 and EMX1 $\Delta^{1-33}$  (with a EMX1-FL 1–33 amino acid truncation) shown by immunofluorescence. Phalloidin was used to indicate cytoskeleton. Overlay, merged images of GFP, cytoskeleton and Hoechst.

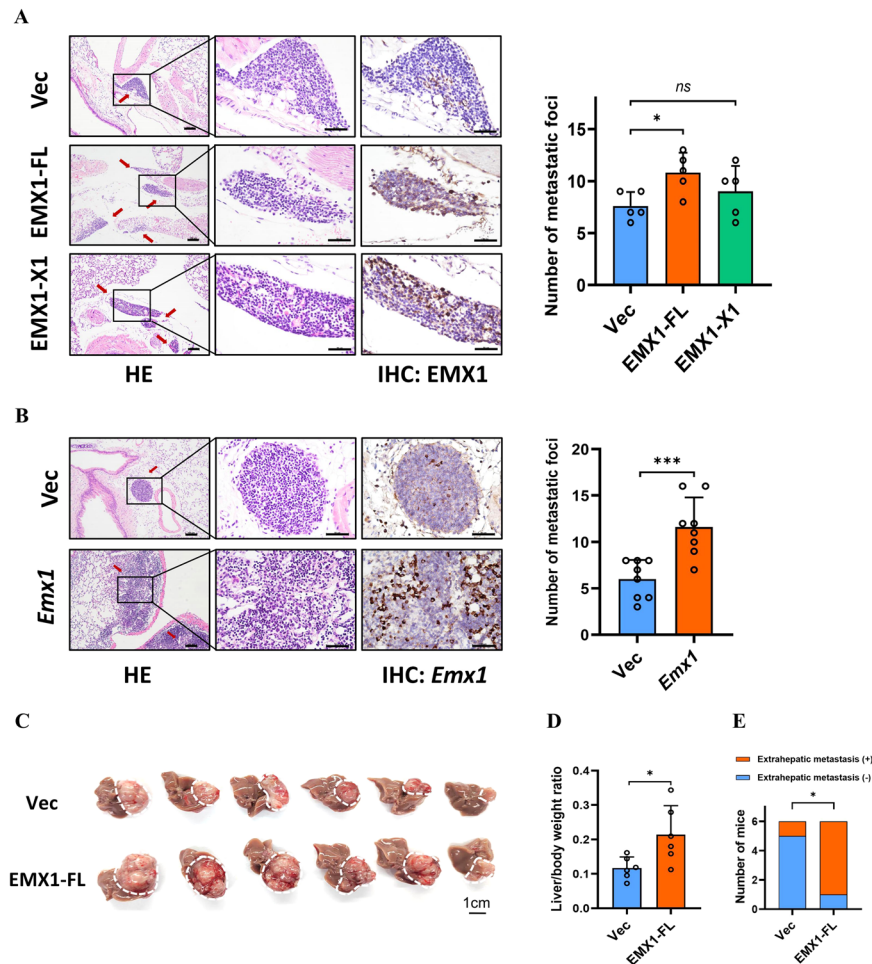
### EMX1-FL but not its alternative splicing isoform enhances HCC migration and metastatic ability in vitro and in vivo

We started out by investigating the total EMX1 mRNA expression in SNU-182, SNU-398, Hep-G2, Hep-3B and Huh-7 HCC cell lines [33]. Interestingly, compared to the hepatoblast-like cell lines

(Hep-G2, Hep-3B and Huh-7), the mesenchymal-like cell lines (SNU-182 and SNU-398) exhibited relatively higher expression of EMX1 (Fig. 3A) [34]. We then established stable EMX1 transcript-specific overexpression in two mesenchymal-like cell lines (SNU-398 and SNU-182) and two hepatoblast-like cell lines (Hep-G2 and



**Fig. 3** EMX1-FL but not its alternative splicing isoform enhances HCC migration and invasion ability in vitro. **A** HCC cell lines total EMX1 expression by RNA-seq from Liver Cancer Cell Lines Database. **B, C** Forced expression of transcript-specific EMX1 in SNU-398 and Hep-G2 and validated by RNA-seq (**B**) and western blotting (**C**). **D–F** Representative images and quantification of in vitro migration (**D**), invasion (**E**) and wound healing (**F**) assays in SNU-398 (upper panel) and Hep-G2 (lower panel) cells stably transfected with transcript-specific EMX1 or vector control. Scale bar, 100  $\mu$ m. **G** Silencing of EMX1 in SNU-398 (left panel) and SNU-182 (right panel) using shRNA lentivirus and validated by qRT-PCR (upper panel) and western blotting (lower panel). **H, I** Representative images and quantification of in vitro migration (**H**) and invasion (**I**) assays in SNU-398 (upper panel) and SNU-182 (lower panel) cells stably transfected with EMX1 shRNA or vector control. Scale bar, 100  $\mu$ m. Data shown represent the means ( $\pm$ SD) of three biological replicates. ns, no significant; \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ .



**Fig. 4** **EMX1-FL enhances HCC metastasis in vivo.** **A, B** The tail-vein injection lung metastasis model: nude mice model (**A**) and the C57BL/6 mice model (**B**). In the left panel, representative images of HE staining (scale bar, 100  $\mu$ m) and IHC anti-EMX1 staining (scale bar, 50  $\mu$ m) are shown after the tail-vein injection of **A** SNU-398 vector control, OE-EMX1-FL or OE-EMX1-X1 cells ( $n = 5$  per group), **B** Hepa1-6 vector control or OE-*Emx1* cells ( $n = 8$  per group). The arrows indicate lung metastasis foci. The right panels show quantification of the number of metastasis foci. **C–E** The orthotopic xenograft HCC mouse model. **C** Gross images with the white dotted box indicates intrahepatic tumor nodules. **D** Quantification of the liver/body weight ratio. **E** Quantification of mice with extrahepatic metastases ( $n = 6$  per group). Error bars represent the mean  $\pm$  SD. ns, no significant; \* $P < 0.05$ ; \*\*\* $P < 0.001$ .

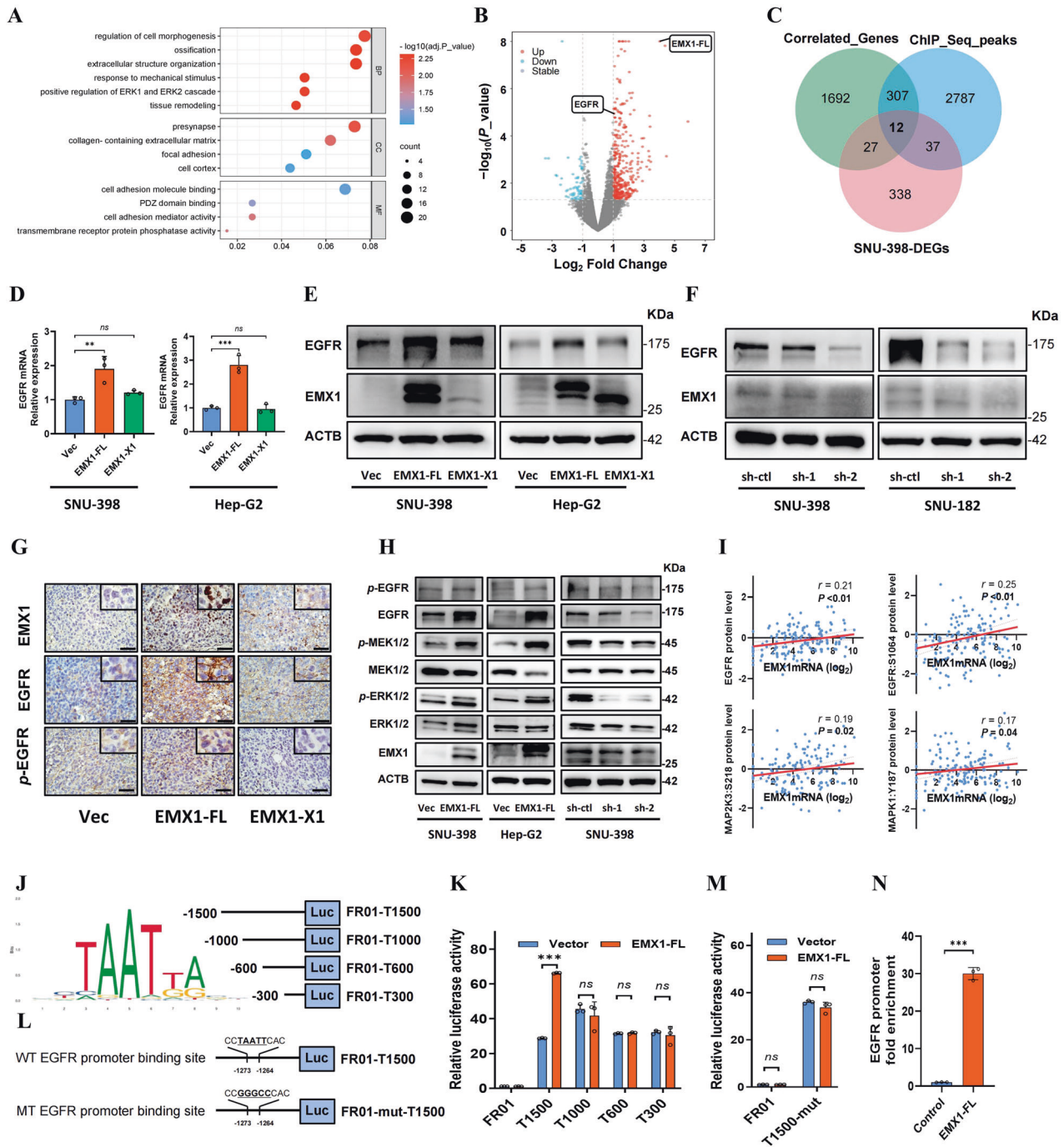
Huh-7) as well as knockdown in SNU-182 and SNU-398 to assess the functions of EMX1 in HCC. We confirmed the ectopic EMX1 isoforms expression by qRT-PCR (Supplementary Fig. S3A), RNA-seq (Fig. 3B) and western blotting (Fig. 3C). The ectopic EMX1 isoforms in SNU-398 and Hep-G2 showed similar growth patterns to the vector control in vitro (Supplementary Fig. S3B) and in vivo (the subcutaneous xenograft model showed EMX1-FL seemed to grow faster but had no statistical significance; Supplementary Fig. S3C–E), indicating that neither EMX1-FL nor EMX1-X1 contribute to proliferation of HCC cells. We next investigated the role of EMX1 in cell migration and invasion. The results showed that over-expression of EMX1-FL, but not the alternative terminal exon splicing isoform EMX1-X1, significantly promoted SNU-398, SNU-182, Hep-G2 and Huh-7 cells migration, invasion, and wound healing in vitro (Fig. 3D–F and Supplementary Fig. S3F, G). Conversely, EMX1-FL knockdown in SNU-398 and SNU-182 cell lines (Fig. 3G) showed significant weaker migration and invasion (Fig. 3H, I).

To further clarify the functions of EMX1 in HCC in vivo, we injected EMX1-FL and EMX1-X1 overexpressing SNU-398 cells into the nude mice via tail-vein. Histological examination of lung tissues revealed a significant increase in the number of lung metastasis foci in the OE-EMX1-FL group, while no statistical

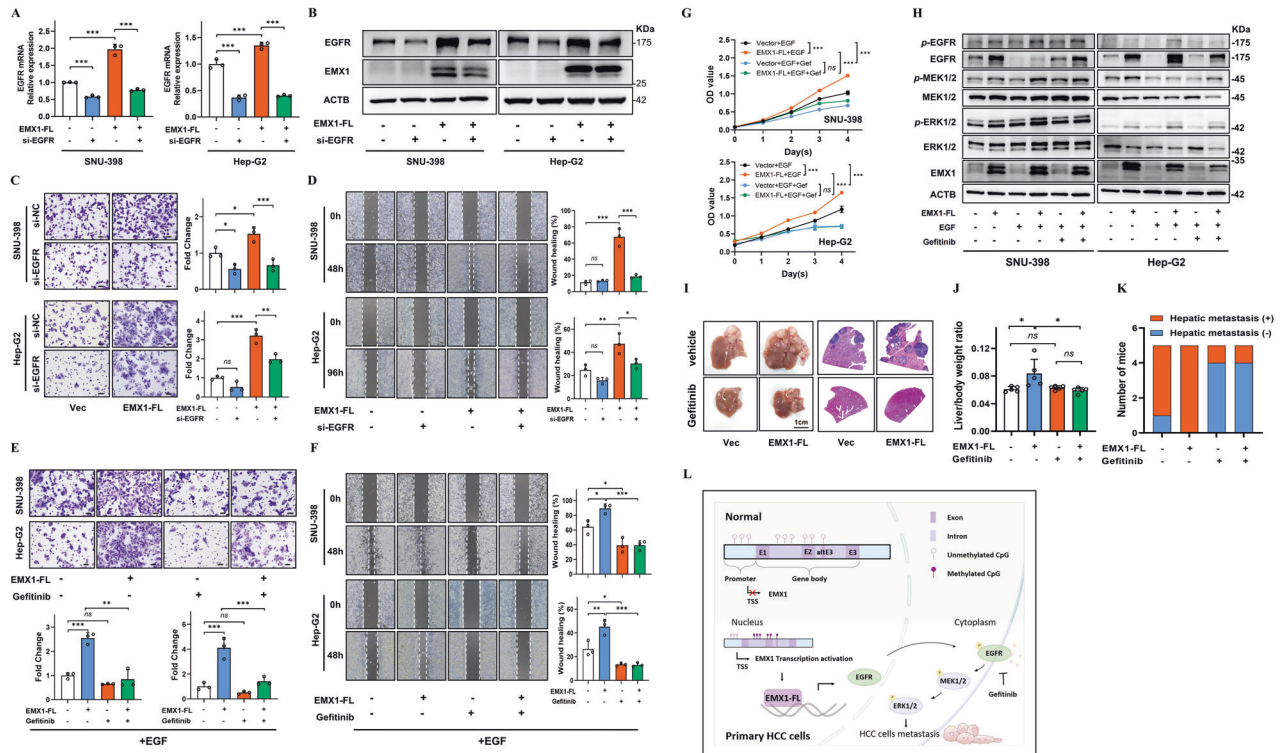
significance was observed in OE-EMX1-X1 group (Fig. 4A). For *Mus musculus*, there is only one validated transcript of *Emx1* (NM\_010131.2). Thus, we constructed an *Emx1* overexpressing Hepa1-6 cells model (Supplementary Fig. S4A). The tail-vein lung metastasis model in C57BL/6 also showed that *Emx1* over-expression increased the metastatic foci (Fig. 4B). For orthotopic xenograft implantation HCC model, OE-EMX1-FL promoted vigorously SNU-398 intrahepatic proliferation and extrahepatic metastasis (Fig. 4C–E and Supplementary Fig. S4B, C). To sum up, our findings suggest that EMX1-FL, but not EMX1-X1, promotes HCC migration, invasion and metastasis in vitro and in vivo.

#### EMX1-FL promotes EGFR-ERK signaling in HCC

EMX1-FL embeds an entire sequence-specific DNA binding Homeodomain, which is predicted to regulate gene transcription. However, the precise downstream pathway and target genes have not been identified. We performed Gene Ontology analysis upon EMX1-FL overexpressing SNU-398 and found that differentially expressed genes (DEGs) were mainly enriched in cell adhesion and migration-associated pathways (Fig. 5A). The metastasis of cancers often relies on cell-surface receptors specific to various growth factors [35]. Interestingly, we also noticed that transmembrane receptor protein phosphatase and associated ERK1/2 cascade



**Fig. 5** **EMX1-FL promotes EGFR-ERK signaling in HCC.** **A, B** The DEGs of SNU-398 EMX1-FL overexpression were enriched by Gene Ontology analysis (**A**) and shown as a volcano plot (**B**). **C** The joint analyses of SNU-398 DEGs, express-correlated genes in TCGA-LIHC and EMX1 Chip-seq peaks from a public databases (<http://cistrome.org/db/>). **D, E** Ectopic expression of EMX1-FL, but not the EMX1-X1, increased the total EGFR mRNA (**D**) and protein (**E**) levels in SNU-398 and Hep-G2. **F** EMX1 Silencing decreased the total EGFR and *p*-EGFR protein level in SNU-398 and SNU-182. **G** Ectopic expression of EMX1-FL, but not the EMX1-X1, increased the total EGFR and *p*-EGFR (Tyr1068) protein in the subcutaneous xenograft model detected by IHC. **H** EMX1-FL overexpression (SNU-398 and Hep-G2) activated EGFR-ERK signaling, while silencing (SNU-398) repressed EGFR-ERK signaling. **I** Scatterplots of the correlation between EMX1 mRNA expression and total protein of EGFR, *p*-EGFR (S1064), *p*-MAP2K3(S218) and *p*-MAPK1(Y187) in the Fudan cohort. **J** According to the Homeodomain rich frequency matrix (left panel), four truncations of EGFR promoter were constructed into the pEZX-FR01 luciferase reporter plasmids (right panel). **K** Relative promoter activity of truncations in SNU-398 luciferase reporter assays. **L** Wild type and mutant binding site on the EGFR promoter. **M** Relative activity of mutant promoter in SNU-398 luciferase reporter assays. **N** The CUT&Tag-qPCR assay measured the occupancy of EMX1-FL on the promoter of EGFR in SNU-398. Data shown represent the means ( $\pm$ SD) of three biological replicates. ns, no significant; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ .



were significantly upregulated in EMX1-FL overexpressing cells. It was also worth noting that analysis on EMX1-X1 overexpressing SNU-398 showed enrichment in pathways regulating vasculature development, angiogenesis, mesenchymal cell differentiation, stem cell development and so on (Supplementary Fig. S5A, B). Through integrating the EMX1-FL DEGs, express-correlated genes and EMX1 Chip-seq peaks in public dataset (Fig. 5B, C), we identified EGFR, which also was the hub gene in our DEG network (Supplementary Fig. S5C). We next assumed that whether EGFR is the key downstream of EMX1-FL.

We detected EGFR mRNA and protein levels and found significantly increased under EMX1-FL ectopic expression, but not in the EMX1-X1 (Fig. 5D, E). Conversely, EMX1-FL knockdown showed the opposite results in SNU-398 and SNU-182 (Fig. 5F). We further confirmed the protein levels in vivo, showing that EMX1-FL promotes EGFR and p-EGFR (Tyr1068) expression in the subcutaneous xenograft model (Fig. 5G). The expression levels of p-MEK1/2, p-ERK1/2 were also increased in EMX1-FL overexpressing SNU-398, Hep-G2 and Huh-7, while the opposite was observed in EMX1-FL knockdown SNU-398 (Fig. 5H and Supplementary Fig. S5D). Correspondingly, we observed a significant positive expression correlation between EMX1 mRNA and EGFR-ERK signaling in HCC in Fudan cohort (Fig. 5I), TCGA cohort and our SYSUCC cohort 1, but not in ANLT samples (Supplementary Fig. S5E). These results suggested that EMX1-FL plays a positive regulative role in EGFR-ERK signaling in HCC.

To identify the potential direct transcriptional regulation of EMX1 on the EGFR promoter, we generated a series of truncated

EGFR promoters according to predicted binding sites (Fig. 5J). Luciferase reporter assays suggested that the truncated promoters exhibited lower activity compared to the full-length promoter length (1.5 Kb) (Fig. 5K), indicating the region between  $-1500$  and  $-1000$  bp upstream of the transcription start site (TSS) was necessary for EMX1-FL binding to EGFR. We then generated an EGFR promoter with a mutated binding site at  $-1273$  to  $-1264$  bp upstream of the TSS (Fig. 5L). The mutation abolished EMX1-FL mediated transcriptional activity (Fig. 5M), indicating that EMX1-FL directly bonded to this site to regulate the transcription of EGFR. Further CUT&Tag sequencing (Supplementary Fig. S5F) and CUT&Tag-qPCR (Fig. 5N) also showed that EMX1-FL was remarkably enriched on the EGFR promoter. In conclusion, we demonstrated that EMX1-FL directly activates EGFR transcription and upregulates EGFR-ERK signaling.

### EMX1-FL promotes metastasis via EGFR signaling, and EGFR1 reverses the effects of overexpressing EMX1-FL

To determine whether the regulation of migration by EMX1-FL depended on EGFR signaling, we silenced EGFR in SNU-398 and Hep-G2 by using siRNA. Partial silencing of EGFR (Fig. 6A, B) successfully reversed the cell migration and wound healing capacity upon EMX1-FL overexpression (Fig. 6C, D). The transwell, wound healing and CCK8 assays verified that the addition of exogenous EGF significantly enhanced cellular migration and growth in EMX1-FL overexpressing cells, as compared to vector cells. What's more, targeting EGFR using Gefitinib (a first-generation EGFR inhibitor) offset the growth-promoting effect,



as well as the migration-promoting effect caused by the addition of exogenous EGF in EMX1-FL overexpressing cells (Fig. 6E–G). The immunoblotting revealed that Gefitinib could block the EMX1-FL promoting EGFR-ERK signaling (Fig. 6H). In the intrasplenic injection orthotopic HCC model, we observed high intrahepatic metastasis incidents in both the EMX1-FL overexpressing group (5 of 5) and the control group (4 of 5), while the EMX1-FL overexpressing group showed a higher liver/body weight ratio. These were significantly attenuated after Gefitinib treatment (Fig. 6I–K). In conclusion, our results suggested that EMX1-FL promotes HCC cells migration by activating EGFR transcription and EGFR-ERK signaling, while blocking the EGFR signals can reverse these effects and reduce HCC metastasis (Fig. 6L).

## DISCUSSION

HCC is a highly lethal malignancy with a high potential for metastasis and frequent recurrence [1–3]. Alterations in DNA methylation landscape are key epigenetic events in HCC [4]. In this study, we identified sixty potential epigenetic genes. We were especially intrigued by EMX1, a gene previously linked to neurodevelopment. We discovered that hypermethylation of the gene body of EMX1 was associated with HCC tumorigenesis and positively correlated to its expression. However, the promoter region remained largely unmethylated. Recently, Su et al. conducted a pan-cancer analysis and found a correlation between gene body hypermethylation of DNA methylation canyons and overexpression in approximately 43% of Homeobox genes, interestingly, many of which are oncogenes [36]. These findings aligned with our observation in EMX1 and were confirmed by exposing and withdrawing DNMTi in HCC cells. Our results indicated that some of the CpGs located in EMX1 gene bodies may represent functional elements under tumorigenesis circumstances, as they are associated with histone methylation modifications (such as H3K4 and H3K27), and histone (de) methylase enrichment (such as EZH2 and JARID) in HCC tissues and cell lines (Supplementary Fig. S6). However, further studies are required to uncover the molecular details of this regulation.

Transcription factors regulate gene programs, thereby controlling diverse cellular processes [37]. Alternative splicing events can generate cancer-associated transcription factors isoforms with altered activity [38]. We noticed that EMX1 mRNA exhibits terminal exon splicing, where the alternative third exon substitutes the third exon and the coding region of protein at the C-terminal, which disrupting the Homeodomain NLS. To our knowledge, this alternative splicing event has not been documented in the literature but can be retrieved from some expressed sequence tag databases [31]. The characteristics are often drastically different between splice isoforms of the same transcription factor, and these differences cannot be predicted based on nominal domain annotations or consensus sequence [37]. The advance of results indicated that EMX1-FL, but not its alternative splicing isoform EMX1-X1, is located in the nucleus and enhances HCC migration and metastatic ability in vitro and in vivo. Additionally, we found that EMX1-X1 may be involved in vasculature development and cell differentiation processes in DEGs enrichment analyses. This may be partly contributed to the different terminator in the Homeodomain. Additionally, the link between terminal exon splicing and oncogene altered activity can also be observed through a difference of 3'UTR regions by posttranscriptional regulatory mechanisms [39]. Nonetheless, further studies are required to unravel the EMX1 splicing regulators and the distinct isoforms in cancers, linking the potential relationship between gene body methylation, RNA splicing, and the roles played by EMX1 isoforms.

The human EMX1 mRNA is encoded by the 18.7-kb EMX1 gene located at 2p13.2. This is not a susceptible locus that frequently changes in copy number or mutates in HCC [4, 40]. Previously, EMX1 was considered to be a tissue-specific gene in the developing

brain [15, 16]. The potential of EMX1 in tumorigenesis was not appreciated until Jimenez-Garcia et al. discussed EMX1 and its analog (EMX2) suppressing osteosarcoma progression [41, 42]. They highlighted that EMX1 and EMX2 act as tumor suppressors by restraining the effector of the canonical WNT pathway, though the precise molecular mechanism remained unknown [42]. However, our study suggested that EMX1 could be an oncogene activated by DNA methylation in the development of HCC. Through a series of metastatic assays following EMX1 perturbation, we unexpectedly discovered that EMX1-FL promoted cell migration and invasion in vitro and metastasis in vivo in an isoform-dependent manner. And there was not significant change in WNT pathway under EMX1 perturbation in our system (data not shown). These suggest that EMX1 potentially exhibits both oncogenic and tumor-suppressive activities in a context-dependent manner.

Recent studies have pointed out that EGFR signaling contributed to the essential regulation of malignant biological behaviors of HCC cells, mainly in metastasis, recurrence, and drug resistance [43, 44]. In the present study, we identified EGFR as the crucial target of Homeodomain-dependent EMX1 signaling. This may be supported by a report that *Emx1* null mice neural stem cells reduced migratory capacity in response to serum or vascular endothelial growth factor during brain development [45]. In our in vitro experiments, we found that EMX1-FL overexpression promoted cell migration and invasion, while EMX1-X1 did not. These can be due to the different regulation of EGFR. Interestingly, both two EMX1 transcripts overexpression individually did not demonstrate a significant proliferation effect in vitro. However, in vivo models (the cell line derived orthotopic xenograft mouse model and the intrasplenic injection orthotopic model) suggested that EMX1-FL promoted cell metastasis and proliferation. These discrepancies may be due to differences in the in vivo cellular conditions. Indeed, the cell lines also showed a proliferative response to exogenous EGF. We then demonstrated that EMX1-FL's promotion of metastatic potential depend on the transcription activation of EGFR, and subsequently upregulating EGFR-ERK signaling. Blocking the EGFR signals by EGFRi (Gefitinib, etc.) can reverse these effects and reduce HCC metastasis. These findings pave the way for targeting the EMX1-EGFR axis in HCC tumorigenicity and metastasis.

There are several limitations in this study. First of all, the EMX1 gene body methylation and expression in this study mainly based on bioinformatic analyses, and whether the promotion of EMX1 was caused by gene body hypermethylation remained unclear as Decitabine is a global hypomethylation agent. The special methylation editing techniques (e.g., dCas9 mediated DNA methylation editing) may be suitable for further investigation. Secondly, we mainly focused on EMX1-FL function in promoting EMX1-EGFR-ERK signaling. However, more results need to be elucidated as to whether the alternative transcript EMX1-X1 also plays a role in HCC tumorigenesis and progression, but specific antibodies against EMX1 isoforms are currently lacking. Thirdly, the utilization of EGFRi to demonstrate the pro-metastasis role of EMX1-FL in HCC dependent on EGFR signaling is insufficient. In our ongoing research, we are currently employing gene editing techniques to specifically mutate the target sequence on the EGFR promoter that binds to EMX1-FL. Finally, targeting EMX1 inhibitor also merits further exploration.

In conclusion, our study revealed the EMX1-EGFR-ERK axis and, for the first time, shed light on the mechanism of EMX1-FL in stimulating HCC metastatic properties, and EMX1-EGFR therefore is a potential target for reducing HCC metastasis.

## MATERIALS AND METHODS

### Patients and public datasets

The study was approved by the Committee for the Ethical Review of Research of Sun Yat-sen University Cancer Center (SYSUCC). Written

informed consent was obtained following institutional guidelines and the declaration of Helsinki guidelines. The Cancer Genome Atlas of Liver Hepatocellular Carcinoma (TCGA-LIHC) methylation profile ( $n = 368$ ) and RNA-seq profile ( $n = 371$ ) were obtained from <http://www.cbioportal.org>. Another five HCC methylation gene sets (GSE112791, GSE113017, GSE113019, GSE89852, GSE77269) were collected from Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/gds>) for verification. The Fudan RNA-seq cohort ( $n = 159$ ) was obtained from Zhongshan Hospital within Fudan University (Shanghai, China) [40]. Additionally, sixty paired fresh-frozen tumors and adjacent normal liver tissue (ANLT) specimens were obtained from pathological diagnosis HCC patients who undergone surgery in SYSUCC for experimental validation (SYSUCC cohort 1). SYSUCC cohort 2 ( $n = 30$ , RNA-seq) was obtained from HCC tumor biopsies. Details of the patients' clinical characteristics and high-throughput data analyses are provided in the Supplementary Information.

### Bisulfite-next-generation sequencing PCR (BSP) and methylation-specific PCR (MSP)

Genomic DNA was extracted from 12 paired fresh HCC tumors and ANLT (from SYSUCC cohort 1) using a Ezup Column Animal Genomic DNA Purification Kit (Sangon Biotech, Shanghai, China). The purified DNA was bisulfite-converted using the EZ DNA Methylation-Gold™ Kit (Zymo-research, USA) according to the manufacturer's protocol. For BSP, the bisulfite-converted genomic DNA was amplified using specific BSP primers, and PCR products were pooled and subjected to next-generation deep sequencing to determine the methylation status of CpG sites. Sangon Biotech (Shanghai, China) performed library preparation and sequencing. BSP covered four regions of the EMX1 gene, each with different numbers of CpG sites: R1–R4 contained 44, 44, 42 and 23 CpG sites, respectively. MSP of bisulfite-converted DNA was carried out with specific methylated or unmethylated primer. Amplified PCR products were separated by 3% agarose gel electrophoresis and visualized with GelRed. Specific CpG dinucleotides analyzed within each region and the primers used in BSP and MSP were shown in Supplementary Information.

### Mouse experiments

All mouse husbandry and procedures were in accordance with protocols approved by the Institutional Care and Animal Use Committee of SYSUCC. No blinding was used during the experiment procedures.

For the subcutaneous xenograft model, SNU-398 vector control, OE-EMX1-FL or OE-EMX1-X1 ( $1 \times 10^6$  cells per mouse) were injected subcutaneously into the right posterior flanks of 6-week-old male nude mice ( $n = 5$  per group). The tumor volume was evaluated based on calliper measurements and calculated using the modified ellipsoidal formula: tumor volume =  $1/2$  length  $\times$  width<sup>2</sup>.

For the tail-vein injection lung metastasis model, SNU-398 vector control, OE-EMX1-FL or OE-EMX1-X1 ( $2 \times 10^6$  cells per mouse) were injected into 6-week-old male nude mice via the tail-vein ( $n = 5$  per group). The mice were sacrificed after 6 weeks. Hepa1-6 vector control or OE-Emx1 ( $2 \times 10^6$  cells per mouse) were injected into 6-week-old male C57BL/6 mice via the tail-vein ( $n = 8$  per group). Mice were sacrificed after 8 weeks. The murine lungs were perfused with 10 ml of PBS through the right ventricle until they were cleared of blood. The tissues were then rinsed with pre-chilled PBS and fixed in 4% paraformaldehyde overnight, and then embedded in paraffin. H&E staining was then performed on consecutive sections of the whole lung. The metastatic foci were recorded microscopically as the maximum number in one section to assess the development of pulmonary metastasis.

For the cell line derived orthotopic xenograft model, subcutaneous tumors were first established by inoculating SNU-398 vector control or OE-EMX1-FL ( $1 \times 10^6$  cells per mouse) into the right posterior flanks of nude mice. When they grew to 10 mm in diameter, tumors were harvested, and non-necrotic tissues were cut into 1 mm<sup>3</sup> pieces and then implanted into the liver left lobe of tumor-free male nude mice ( $n = 6$  per group). The mice were euthanized after 4 weeks. Liver and body weight were recorded, and the presence of extrahepatic metastasis foci was carefully documented.

For the intrasplenic injection HCC orthotopic model, SNU-398 vector control or OE-EMX1-FL ( $2 \times 10^6$  cells per mouse) were injected into 6-week-old male nude mice ( $n = 10$  per group) via intrasplenic injection. After 3 weeks, mice were randomly assigned to 5 days per week treatment with vehicle or Gefitinib (80 mg/kg, oral gavage) ( $n = 5$  per group). All mice were euthanized after 2 weeks of drug treatment and 1 week of observation. Body weight was monitored throughout the treatment. Liver

and body weight were recorded at the end of experiments, and the livers were then processed for histological examination.

### Drug treatment

Gefitinib (S1025), Decitabine (S1200) were purchased from Selleck Chemicals. For DNA methyltransferase inhibitor assay, adherent SNU-398 cells were treated with 10  $\mu$ M Decitabine at the onset day (D0). After 3 days (D3), the media were replaced with fresh drug-free media and refreshed every other day for an extra 11 days (until D14). Control cells were cultured under similar experimental conditions in the absence of drugs. For blockade of EGFR activation, cells were treated with Gefitinib (10  $\mu$ M) for 48 h before sample collection.

### Statistics

All statistical analyses in this study were performed using the GraphPad Prism 8.0 and R 4.0. Two-tailed Student's *t*-test or one-way ANOVA was applied for calculating statistical significance. The Log-rank test was used for survival analyses. Gene expression correlation was calculated through Pearson correlation analysis. The *P* values < 0.05 were considered statistically significant.

More details for other methods, including cell lines information, 3' rapid amplification of cDNA ends (3'RACE), plasmids construct, lentivirus infection, transiently silencing or overexpression, RNA extraction and quantitative real-time PCR (qRT-PCR), western blotting, immunohistochemistry (IHC) and immunofluorescence (IF), cell viability and proliferation assays, transwell assay, wound healing assay, luciferase assay, CUT&Tag assay, high-throughput data analyses, see Supplementary Information.

### Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

### DATA AVAILABILITY

All raw data have been deposited into the Research Data Deposit with the accession code RDDB2023459670 (<http://www.researchdata.org.cn>). Uncropped western blots are shown in Supplementary Information. The RNA-seq data generated in this study have been deposited in the GEO database with the accession code GSE245505. Other datasets generated during the current study are available from the corresponding author on reasonable request.

### REFERENCES

- Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. *CA Cancer J Clin.* 2022;72:7–33. <https://doi.org/10.3322/caac.21708>.
- Llovet JM, Kelley RK, Villanueva A, Singal AG, Pikarsky E, Roayaie S, et al. Hepatocellular carcinoma. *Nat Rev Dis Primers.* 2021;7:6. <https://doi.org/10.1038/s41572-020-00240-3>.
- European Association for the Study of the Liver. EASL Clinical Practice Guidelines: management of hepatocellular carcinoma. *J Hepatol.* 2018;69:182–236. <https://doi.org/10.1016/j.jhep.2018.03.019>.
- Network TCGA. Comprehensive and integrative genomic characterization of hepatocellular carcinoma. *Cell.* 2017;169:1327–41. <https://doi.org/10.1016/j.cell.2017.05.046>.
- Villanueva A, Portela A, Sayols S, Battiston C, Hoshida Y, Mendez-Gonzalez J, et al. DNA methylation-based prognosis and epidrivers in hepatocellular carcinoma. *Hepatology.* 2015;61:1945–56. <https://doi.org/10.1002/hep.27732>.
- Kisiel JB, Dukek BA, Kanipakam RVSR, Ghosh HM, Yab TC, Berger CK, et al. Hepatocellular carcinoma detection by plasma methylated DNA: discovery, phase I pilot, and phase II clinical validation. *Hepatology.* 2019;69:1180–92. <https://doi.org/10.1002/hep.30244>.
- Xu RH, Wei W, Krawczyk M, Wang W, Luo H, Flagg K, et al. Circulating tumour DNA methylation markers for diagnosis and prognosis of hepatocellular carcinoma. *Nat Mater.* 2017;16:1155–61. <https://doi.org/10.1038/nmat4997>.
- Qiu J, Peng B, Tang Y, Qian Y, Guo P, Li M, et al. CpG methylation signature predicts recurrence in early-stage hepatocellular carcinoma: results from a multicenter study. *J Clin Oncol.* 2017;35:734–42. <https://doi.org/10.1200/JCO.2016.68.2153>.
- Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet.* 2012;13:484–92. <https://doi.org/10.1038/nrg3230>.
- Yang X, Han H, De Carvalho DD, Lay FD, Jones PA, Liang G. Gene body methylation can alter gene expression and is a therapeutic target in cancer. *Cancer Cell.* 2014;26:577–90. <https://doi.org/10.1016/j.ccr.2014.07.028>.
- Shukla S, Kavak E, Gregory M, Imashimizu M, Shutinoski B, Kashlev M, et al. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature.* 2011;479:74–9. <https://doi.org/10.1038/nature10442>.

12. Hahn MA, Wu X, Li AX, Hahn T, Pfeifer GP. Relationship between gene body DNA methylation and intragenic H3K9me3 and H3K36me3 chromatin marks. *PLoS One*. 2011;6:e18844. <https://doi.org/10.1371/journal.pone.0018844>.
13. Arechederra M, Bazai SK, Abdouni A, Sequera C, Mead TJ, Richelme S, et al. ADAMTSL5 is an epigenetically activated gene underlying tumorigenesis and drug resistance in hepatocellular carcinoma. *J Hepatol*. 2021;74:893–906. <https://doi.org/10.1016/j.jhep.2020.11.008>.
14. Liu M, Zhang L, Li H, Hinoue T, Zhou W, Ohtani H, et al. Integrative epigenetic analysis reveals therapeutic targets to the DNA methyltransferase inhibitor guadecitabine (SGI-110) in hepatocellular carcinoma. *Hepatology*. 2018;68:1412–28. <https://doi.org/10.1002/hep.30091>.
15. Simeone A, Gulisano M, Acampora D, Stornaiuolo A, Rambaldi M, Boncinelli E. Two vertebrate homeobox genes related to the *Drosophila* empty spiracles gene are expressed in the embryonic cerebral cortex. *EMBO J*. 1992;11:2541–50. <https://doi.org/10.1002/j.1460-2075.1992.tb05319.x>.
16. Chan CH, Godinho LN, Thomaidou D, Tan SS, Gulisano M, Parnavelas JG. Emx1 is a marker for pyramidal neurons of the cerebral cortex. *Cereb Cortex*. 2001;11:1191–8. <https://doi.org/10.1093/cercor/11.12.1191>.
17. Guo H, Hong S, Jin XL, Chen RS, Avasthi PP, Tu YT, et al. Specificity and efficiency of Cre-mediated recombination in Emx1-Cre knock-in mice. *Biochem Biophys Res Commun*. 2000;273:661–5. <https://doi.org/10.1006/bbrc.2000.2870>.
18. Gaveriaux-Ruff C, Kieffer BL. Conditional gene targeting in the mouse nervous system: insights into brain function and diseases. *Pharmacol Ther*. 2007;113:619–34. <https://doi.org/10.1016/j.pharmthera.2006.12.003>.
19. Wang L, Jin J, Zhou Y, Tian Z, He B, Huang Y, et al. EMX2 is epigenetically silenced and suppresses epithelial-mesenchymal transition in human esophageal adenocarcinoma. *Oncol Rep*. 2019;42:2169–78. <https://doi.org/10.3892/or.2019.7284>.
20. Monnier A, Boniface R, Bouvet R, Etcheverry A, Aubry M, Avril T, et al. The expression of EMX2 lead to cell cycle arrest in glioblastoma cell line. *BMC Cancer*. 2018;18:1213. <https://doi.org/10.1186/s12885-018-5094-y>.
21. Aykut B, Ochs M, Radhakrishnan P, Brill A, Hocker H, Schwarz S, et al. EMX2 gene expression predicts liver metastasis and survival in colorectal cancer. *BMC Cancer*. 2017;17:555. <https://doi.org/10.1186/s12885-017-3556-2>.
22. Zhang Y, Cao G, Yuan QG, Li JH, Yang WB. Empty Spiracles Homeobox 2 (EMX2) inhibits the invasion and tumorigenesis in colorectal cancer cells. *Oncol Res*. 2017;25:537–44. <https://doi.org/10.3727/096504016X14756640150695>.
23. Yue D, Li H, Che J, Zhang Y, Tolani B, Mo M, et al. EMX2 is a predictive marker for adjuvant chemotherapy in lung squamous cell carcinomas. *PLoS One*. 2015;10:e132134. <https://doi.org/10.1371/journal.pone.0132134>.
24. Yang JD, Ghos H, Aboelsoud MM, Taylor WR, Yab TC, Berger CK, et al. DNA methylation markers for detection of cholangiocarcinoma: discovery, validation, and clinical testing in biliary brushings and plasma. *Hepatol Commun*. 2021;5:1448–59. <https://doi.org/10.1002/hep4.1730>.
25. Yang R, Yang L, Long M, Wu F, Shi Y, Xia R, et al. Taurodeoxycholic acid-YAP1 upregulates OTX1 in promoting gallbladder cancer malignancy through IFITM3-dependent AKT activation. *Oncogene*. 2023;42:1466–77. <https://doi.org/10.1038/s41388-023-02660-3>.
26. Shao M, Yang Q, Zhu W, Jin H, Wang J, Song J, et al. LncHOXA10 drives liver TICs self-renewal and tumorigenesis via HOXA10 transcription activation. *Mol Cancer*. 2018;17:173. <https://doi.org/10.1186/s12943-018-0921-y>.
27. De Keersmaecker K, Real PJ, Gatta GD, Palomero T, Sulis ML, Tosello V, et al. The TLX1 oncogene drives aneuploidy in T cell transformation. *Nat Med*. 2010;16:1321–7. <https://doi.org/10.1038/nm.2246>.
28. Huang Y, Yang Q, Zheng Y, Lin L, Xu X, Xu XE, et al. Activation of bivalent factor DLX5 cooperates with master regulator TP63 to promote squamous cell carcinoma. *Nucleic Acids Res*. 2021;49:9246–63. <https://doi.org/10.1093/nar/gkab679>.
29. Kolfshoten IG, van Leeuwen B, Berns K, Mullenders J, Beijersbergen RL, Bernards R, et al. A genetic screen identifies PITX1 as a suppressor of RAS activity and tumorigenicity. *Cell*. 2005;121:849–58. <https://doi.org/10.1016/j.cell.2005.04.017>.
30. Tai WT, Chen YL, Chu PY, Chen LJ, Hung MH, Shiau CW, et al. Protein tyrosine phosphatase 1B dephosphorylates PITX1 and regulates p120RasGAP in hepatocellular carcinoma. *Hepatology*. 2016;63:1528–43. <https://doi.org/10.1002/hep.28478>.
31. Thierry-Mieg D, Thierry-Mieg J. AceView: a comprehensive cDNA-supported gene and transcripts annotation. *Genome Biol*. 2006;7(Suppl 1):S11–2. <https://doi.org/10.1186/gb-2006-7-s1-s12>.
32. Nguyen BA, Pogoutse A, Provart N, Moses AM. NLStradamus: a simple Hidden Markov Model for nuclear localization signal prediction. *BMC Bioinformatics*. 2009;10:202. <https://doi.org/10.1186/1471-2105-10-202>.
33. Qiu Z, Li H, Zhang Z, Zhu Z, He S, Wang X, et al. A pharmacogenomic landscape in human liver cancers. *Cancer Cell*. 2019;36:179–93. <https://doi.org/10.1016/j.ccell.2019.07.001>.
34. Caruso S, Calatayud AL, Pilet J, La Bella T, Reikik S, Imbeaud S, et al. Analysis of liver cancer cell lines identifies agents with likely efficacy against hepatocellular carcinoma and markers of response. *Gastroenterology*. 2019;157:760–76. <https://doi.org/10.1053/j.gastro.2019.05.001>.
35. Weiss F, Lauffenburger D, Friedl P. Towards targeting of shared mechanisms of cancer metastasis and therapy resistance. *Nat Rev Cancer*. 2022;22:157–73. <https://doi.org/10.1038/s41568-021-00427-0>.
36. Su J, Huang YH, Cui X, Wang X, Zhang X, Lei Y, et al. Homeobox oncogene activation by pan-cancer DNA hypermethylation. *Genome Biol*. 2018;19:108. <https://doi.org/10.1186/s13059-018-1492-3>.
37. Joung J, Ma S, Tay T, Geiger-Schuller KR, Kirchgatterer PC, Verdine VK, et al. A transcription factor atlas of directed differentiation. *Cell*. 2023;186:209–29. <https://doi.org/10.1016/j.cell.2022.11.026>.
38. Belluti S, Rigillo G, Imbriano C. Transcription factors in cancer: when alternative splicing determines opposite cell fates. *Cells*. 2020;9:760. <https://doi.org/10.3390/cells9030760>.
39. Mayr C, Bartel DP. Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell*. 2009;138:673–84. <https://doi.org/10.1016/j.cell.2009.06.016>.
40. Gao Q, Zhu H, Dong L, Shi W, Chen R, Song Z, et al. Integrated proteogenomic characterization of HBV-related hepatocellular carcinoma. *Cell*. 2019;179:561–77. <https://doi.org/10.1016/j.cell.2019.08.052>.
41. Jimenez-Garcia MP, Lucena-Cacace A, Otero-Albiol D, Carnero A. Regulation of sarcomagenesis by the empty spiracles homeobox genes EMX1 and EMX2. *Cell Death Dis*. 2021;12:515. <https://doi.org/10.1038/s41419-021-03801-w>.
42. Jimenez-Garcia MP, Lucena-Cacace A, Otero-Albiol D, Carnero A. Empty spiracles homeobox genes EMX1 and EMX2 regulate WNT pathway activation in sarcomagenesis. *J Exp Clin Cancer Res*. 2021;40:247. <https://doi.org/10.1186/s13046-021-02048-9>.
43. Su T, Huang M, Liao J, Lin S, Yu P, Yang J, et al. Insufficient radiofrequency ablation promotes hepatocellular carcinoma metastasis through N6-methyladenosine mRNA methylation-dependent mechanism. *Hepatology*. 2021;74:1339–56. <https://doi.org/10.1002/hep.31766>.
44. Jin H, Shi Y, Lv Y, Yuan S, Ramirez C, Liefstink C, et al. EGFR activation limits the response of liver cancer to lenvatinib. *Nature*. 2021;595:730–4. <https://doi.org/10.1038/s41586-021-03741-7>.
45. Kobeissy FH, Hansen K, Neumann M, Fu S, Jin K, Liu J. Deciphering the role of Emx1 in neurogenesis: a neuroproteomics approach. *Front Mol Neurosci*. 2016;9:98. <https://doi.org/10.3389/fnmol.2016.00098>.

## ACKNOWLEDGEMENTS

These results are partly based upon public data generated by TCGA research network and GEO. We sincerely appreciate all the patients whose tissue specimens were included in our retrospective study. We thank all members of our labs for helpful discussions and comments. This work is supported by National Natural Science Foundation of China and China Postdoctoral Science Foundation.

## AUTHOR CONTRIBUTIONS

AK, MS and D-SW conceptualized the project and designed the experiments. D-SW and L-CH performed the experiments. X-YB, Z-FD and Y-XH performed high-throughput data analyses. M-KH and Z-CL contributed to clinical specimens and data collection. D-SW wrote the manuscript. AK and MS revised of the whole manuscript. All authors read and approved the final manuscript. The work reported in the paper has been performed by the authors, unless clearly specified in the text.

## FUNDING

This work was supported by National Natural Science Foundation of China (No. 82203126, No.82072610, No.82102985), and China Postdoctoral Science Foundation (No. 2021TQ0383).

## COMPETING INTERESTS

The authors declare no competing interests.

## ETHICAL APPROVAL

The study was approved by the committee for the Ethical Review of Research of SYSUCC (approved number: B2022-076-01). It was performed in accordance with the principles of the Declaration of Helsinki. Mouse experiments were performed under protocols approved by the Institutional Care and Animal Use Committee of SYSUCC (approved number: 2022-000071).

**ADDITIONAL INFORMATION**

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41419-023-06293-y>.

**Correspondence** and requests for materials should be addressed to Anna Kan or Ming Shi.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023