



# Lineage-Specific Epigenomic and Genomic Activation of Oncogene HNF4A Promotes Gastrointestinal Adenocarcinomas

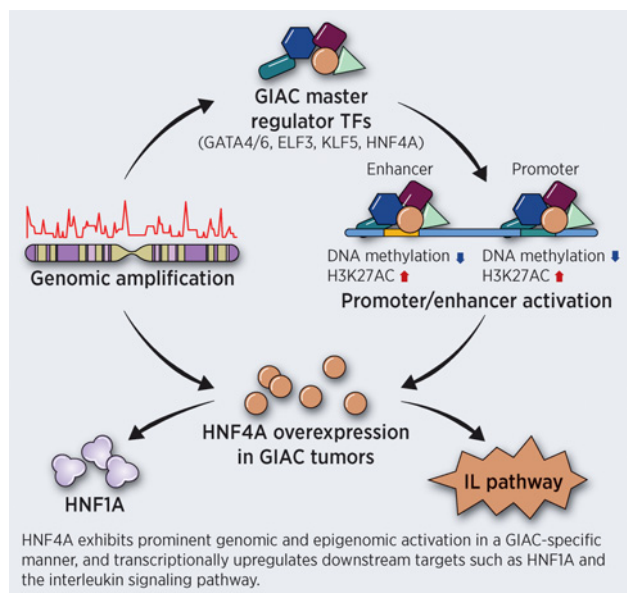
Jian Pan<sup>1,2</sup>, Tiago C. Silva<sup>3,4</sup>, Nicole Gull<sup>3</sup>, Qian Yang<sup>2,5</sup>, Jasmine T. Plummer<sup>3</sup>, Stephanie Chen<sup>3</sup>, Kenji Daigo<sup>6</sup>, Takao Hamakubo<sup>6</sup>, Sigal Gery<sup>2</sup>, Ling-Wen Ding<sup>7</sup>, Yan-Yi Jiang<sup>7</sup>, Shaoyan Hu<sup>1</sup>, Li-Yan Xu<sup>5</sup>, En-Min Li<sup>5</sup>, Yanbing Ding<sup>8</sup>, Samuel J. Klempner<sup>9,10</sup>, Simon A. Gayther<sup>3</sup>, Benjamin P. Berman<sup>3,11</sup>, H. Phillip Koefler<sup>2,7,12</sup>, and De-Chen Lin<sup>2</sup>

## ABSTRACT

Gastrointestinal adenocarcinomas (GIAC) of the tubular gastrointestinal (GI) tract including esophagus, stomach, colon, and rectum comprise most GI cancers and share a spectrum of genomic features. However, the unified epigenomic changes specific to GIAC are poorly characterized. Using 907 GIAC samples from The Cancer Genome Atlas, we applied mathematical algorithms to large-scale DNA methylome and transcriptome profiles to reconstruct transcription factor (TF) networks and identify a list of functionally hyperactive master regulator (MR) TF shared across different GIAC. The top candidate HNF4A exhibited prominent genomic and epigenomic activation in a GIAC-specific manner. A complex interplay between the HNF4A promoter and three distal enhancer elements was coordinated by GIAC-specific MRTL including ELF3, GATA4, GATA6, and KLF5. HNF4A also self-regulated its own promoter and enhancers. Functionally, HNF4A promoted cancer proliferation and survival by transcriptional activation of many downstream targets, including HNF1A and factors of interleukin signaling, in a lineage-specific manner. Overall, our study provides new insights into the GIAC-specific gene regulatory networks and identifies potential therapeutic strategies against these common cancers.

**Significance:** These findings show that GIAC-specific master regulatory transcription factors control HNF4A via three distal enhancers to promote GIAC cell proliferation and survival.

**Graphical Abstract:** <http://cancerres.aacrjournals.org/content/canres/80/13/2722/F1.large.jpg>.



## Introduction

Tubular gastrointestinal adenocarcinomas (GIAC) including esophageal (EAC), stomach (STAD), and colorectal adenocarcinomas (COAD) arise from columnar epithelium and share similar endodermal cell-of-origin. GIACs represent one of the most common and

aggressive tumor types, affecting over 3.5 million worldwide (1), and causing 1.4 million deaths (1) in the year of 2018 alone. At the genomic level, GIACs notably exhibit a spectrum of both common and unique characteristics, and can be subtyped accordingly into Epstein-Barr virus (EBV)-positive (STAD-specific), chromosomal instability (CIN),

<sup>1</sup>Department of Hematology and Oncology, Children's Hospital of Soochow University, Suzhou, China. <sup>2</sup>Department of Medicine, Cedars-Sinai Medical Center, Los Angeles, California. <sup>3</sup>Center for Bioinformatics and Functional Genomics, Cedars-Sinai Medical Center, Los Angeles, California. <sup>4</sup>Department of Genetics, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, Brazil. <sup>5</sup>Institute of Oncologic Pathology, Medical College of Shantou University, Shantou, China. <sup>6</sup>Department of Protein-protein Interaction Research, Institute for Advanced Medical Sciences, Nippon Medical School, Kawasaki, Kanagawa, Japan. <sup>7</sup>Cancer Science Institute of Singapore, National University of Singapore, Singapore. <sup>8</sup>Department of Gastroenterology, Affiliated Hospital of Yangzhou University, Yangzhou University, Jiangsu, China. <sup>9</sup>Department of Medicine, Massachusetts General Hospital Cancer Center, Boston, Massachusetts. <sup>10</sup>Harvard Medical School, Boston, Massachusetts. <sup>11</sup>Department of Developmental Biology and Cancer Research, Institute for Medical Research Israel-Canada, Hebrew University-Hadassah Medical School, Jerusalem, Israel.

<sup>12</sup>National University Cancer Institute, National University Hospital Singapore, Singapore.

**Note:** Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org/>).

J. Pan, T.C. Silva, N. Gull, Q. Yang, and J.T. Plummer contributed equally to this work.

**Corresponding Authors:** De-Chen Lin, Cedars-Sinai Medical Center, 8700 Beverly Blvd, Los Angeles, CA 90048. Phone: 3104237736; E-mail: dchlin1@gmail.com; and Benjamin P. Berman, Hebrew University-Hadassah Medical School, Jerusalem, Israel; E-mail: ben.berman@mail.huji.ac.il

Cancer Res 2020;80:2722-36

doi: 10.1158/0008-5472.CAN-20-0390

©2020 American Association for Cancer Research.

microsatellite instability (MSI), and genomically stable (GS; refs. 2–4). Some of these genomic features have important clinical implications. For example, EBV and MSI tumors showed increased sensitivity to immune checkpoint blockade (5, 6).

Large-scale genomic analyses (7–9) have established genetic landscapes of all three major GIAC types (EAC, STAD, and COAD). Very recently, The Cancer Genome Atlas (TCGA) consortium conducted a comparative analysis and characterized shared genomic features across these GI cancers (4). Compared with non-GIAC tumors, unique genomic changes in GIACs included amplification of lineage-specific transcription factors (TF; e.g., *CDX2*, *KLF5* and *GATA4/6*) and mutations and deletions of *APC* and *SOX9*. Interestingly, this same study also noted that a significant fraction of pancreatic adenocarcinomas (PAAD) exhibited a high-degree molecular similarity with STAD (particularly the CIN and GS subtypes) based on consensus clustering results. Indeed, other individual reports also found that several key oncogenes (such as *GATA6* and *KLF5*; refs. 10–12) are similarly essential for the proliferation and viability of both PAAD and other GIAC cancers.

Relative to genomic alterations, the epigenomic changes that are either shared or private in GIACs are less well-characterized. Most knowledge is derived from DNA methylation array data (Infinium HM450), which has identified the CpG island methylator phenotype (CIMP) uniquely in EBV+ and some MSI+ GIACs. Further integration with RNA profiling suggests that there conceivably exist GIAC-specific gene regulatory networks that contribute to the phenotypes and biological processes shared among GI cancers (7–9). However, such gene expression networks and particularly their enhancer regions and upstream regulators have not been well elucidated in a Pan-GIAC manner.

Genome-wide DNA methylation profiling has shown that enhancer demethylation represents the most dynamic change that occurs during neoplastic development (13–16). Our earlier work demonstrated that this form of demethylation could be used to identify cancer-specific enhancers that were strongly enriched for specific TF binding sites (TFBS; ref. 14). Recently, we have developed a novel computational algorithm, enhancer linking by methylation/expression relationships (ELMER), to identify and exploit systematically these TFBS-associated methylation changes in cancer (17, 18). Briefly, the approach of ELMER is to utilize TFBS methylation changes as key nodes within the larger gene regulatory network, and correlate with gene expression to infer both the upstream (master regulator TFs, MRTFs) and downstream (target genes) links for each TFBS. Specifically, when a group of enhancers is coordinately altered in a specific sample subset, this is often the result of an altered upstream MRTF in the gene regulatory network. To identify MRTFs involved in setting the tumor-specific DNA methylation patterns, ELMER correlates DNA methylation occurring within binding sites for specific MRTFs (inferred by specific TF binding motifs), with altered expression of the corresponding MRTF (17, 18).

Considering the molecular similarities among GIACs, we hypothesized that there exist GIAC-specific gene regulatory networks that are controlled by GIAC-specific MRTFs. Here, we addressed this hypothesis by applying ELMER to TCGA Pan-GIAC samples to explore whether different GIACs share functionally hyperactive MRTFs.

## Materials and Methods

### Antibodies and reagents

The following antibodies and reagents were used: anti-HNF4A for chromatin immunoprecipitation sequencing (ChIP-seq) and Western blotting (Abcam, ab41898), anti-HNF4A for IHC (a gift from Kenji

Daigo and Takao Hamakubo), anti-ELF3 (Santa Cruz Biotechnology, sc-376055 X), anti-KLF5 (Santa Cruz Biotechnology, sc-398470 X), anti-GATA6 (Cell Signaling Technology, 5851), anti-GATA4 (Thermo Fisher Scientific, MA5-15532), anti-HNF1A (Cell Signaling Technology, 89670), anti-FLAG (Sigma, F1804), anti-actin (Santa Cruz Biotechnology, sc-8432), anti-GAPDH (Cell Signaling Technology, 2118), anti-H3K27Ac (Abcam, ab4729), anti-mouse IgG-HRP (Jackson ImmunoResearch Laboratories, Inc., 115-035-003), anti-rabbit IgG-HRP (Jackson ImmunoResearch Laboratories, Inc., 115-035-144), FITC Annexin V Apoptosis Detection Kit (BD Biosciences, 556547), Lipofectamine RNAiMAX (Thermo Fisher Scientific, 13778150), BI-6015 (Cayman, 12032) and siRNAs were purchased from Suzhou GenePharma Co., Ltd. (Sequences are provided in Supplementary Table S1).

### Human cancer cell lines

Eso26, OACM5.1, SNU398, OACp4C, OE19 cells were grown in RPMI1640 medium, and Flo1, JH-EsoAD1, SKGT4, ASPC-1, Suit2, Colo205, and OE33 were cultured in DMEM. Eso26, OACp4C, OE19, Flo1, JH-EsoAD1, OE19, OE33, SKGT4, and OACM5.1 were purchased from the ATCC. ASPC-1, Suit2, Colo205, and SNU398 were purchased from JENNIO Biological Technology. Both media were supplemented with 10% FBS (Omega Scientific), 100 U/mL penicillin, and 100 mg/mL streptomycin. All cell lines were verified by short tandem repeat analysis in the year of 2018.

### ELMER analyses

Illumina HM450 methylation data from four different TCGA projects, COAD-READ, ESCA, PAAD, and STAD were processed with SeSAMe (19), and the matched RNA-seq (FPKM-UQ) data were downloaded from GDC (Genomic Data Commons, <https://portal.gdc.cancer.gov/>) from the harmonized database (data aligned to hg38/GRCh38). Unsupervised analysis was performed using ELMER (18) for the following pair-wise comparisons: EAC primary tumors ( $n = 78$ ) versus ESCC primary tumors ( $n = 76$ ); EAC primary tumors ( $n = 78$ ) versus EAC normal adjacent samples ( $n = 5$ ), COAD-READ primary tumors ( $n = 389$ ) versus colon normal adjacent samples ( $n = 21$ ), PAAD primary tumors ( $n = 177$ ) versus pancreas normal adjacent samples ( $n = 4$ ). Because of the small number of normal adjacent samples from STAD ( $n = 2$ ), which also lacked RNA-seq data, we could not perform comparison for STAD primary tumors. We applied the default probe filter “distal,” which uses 160,944 probes that are >2 kbp from any transcription start site as annotated by GENCODE 28. ELMER version 2.5.4 was used with the following parameters: `get.diff.meth(sig.diff) = 0.3`, `get.diff.meth(p_value) = 0.01`, `get.diff.meth(minSubgroupFrac) = 0.2`, `get.pair(Pe) = 10-3`, `get.pair(raw.pvalue) = 10-3`, and `get.pair(filter.probes) = TRUE`, `get.pair(permu.size) = 10,000`, `get.pair(minSubgroupFrac) = 0.4`, `get.enriched.motif(lower.OR) = 1.1`, `get.TFs(minSubgroupFrac) = 0.4`. For each comparison, the TF subfamilies were inferred from the TFs, which had the most significant anticorrelation scores. The candidate MRTFs were next identified within the TF subfamily with  $FDR < 0.05$  using the classification from TFClass (20). An HTML file with all necessary code to reproduce ELMER analysis and figures is available as Supplementary Data S1.

### Whole genome bisulfite sequencing and data analysis

Primary tumor tissues from 5 EAC and 12 ESCC samples were collected at Shantou Center Hospital, China. Informed written consent was obtained from all patients. This study has been approved by the ethics committee of the Medical College of Shantou University. Whole

genome bisulfite sequencing (WGBS) library preparation and sequencing were performed by Novogene, Inc. Briefly, 2 to 3  $\mu\text{g}$  DNA spiked with 26 ng lambda DNA were fragmented. Cytosine-methylated barcodes were ligated to the sonicated DNA, treated twice with bisulfite using EZ DNA Methylation-Gold™ Kit (Zymo Research). The resulting DNA fragments were PCR amplified using HiFi HotStart Uracil + ReadyMix (Kapa Biosystems). The clustering of the index-coded samples was performed on Illumina cBot Cluster Generation System according to the manufacturer's instructions. DNA libraries were sequenced on Illumina HiSeq platform with 150bp paired-end reads.

WGBS reads were aligned to the genome (build GRCh38) using BISCUIT (<https://github.com/zwdzwd/biscuit>). Duplicated reads were marked using Picard Tools (<https://broadinstitute.github.io/picard/>). Methylation rates were called using BISCUIT. CpGs with fewer than five reads of coverage were excluded from further analysis. Quality control was performed using TrimGalore by default parameter for Illumina sequencing platforms, ([https://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/)), PicardTools and MultiQC (<https://multiqc.info/>). Bisulfite nonconversion was checked using the Biscuit QC module in MultiQC (<https://github.com/ewels/MultiQC/tree/master/multiqc/modules/biscuit>).

#### ELMER-dependent motif analysis of WGBS

EAC and ESCC WGBS BED files were converted into Tag Directories using HOMER (<http://homer.ucsd.edu/homer/>) with the mCpGbed option. The annotatePeaks.pl script from HOMER was then run on all Tag Directories from each EAC and ESCC sample, with focal points defined as the list of predicted HNF4A motifs adjacent to EAC-demethylated CpGs, generated by ELMER (EAC vs. ESCC run described above). A bin size of 100bp for HOMER to generate spatial methylation plot.

#### ELMER-independent motif analysis of WGBS

The WGBS analysis was performed as described in the ELMER-dependent section above, except with different focal points. Here, we generated a set of predicted HNF4A binding sites by scanning the complete human genome using HOMER's scanMotifGenomeWide.pl at a p-value threshold of  $1\text{E}-5$ . We used the same HOCOMOCO (21) HNF4A model as used in ELMER ([http://hocomoco11.autosome.ru/motif/HNF4A\\_HUMAN.H11MO.0.A](http://hocomoco11.autosome.ru/motif/HNF4A_HUMAN.H11MO.0.A)). We then filtered only those predicted HNF4A sites that overlapped one of the 72,264 esophageal-specific ATAC-seq peaks from TCGA (22), aligning to the HNF4A motif for the spatial methylation plot. For the box plot, methylation was averaged within 20kb from each HNF4A motif.

#### Cancer Cell Line Encyclopedia data

We used dependency scores from the "Combined RNAi (Broad, Novartis, Marcotte)" field, expression values from the "Expression public 19Q2" field, and methylation values from the "Methylation (1kb upstream TSS)" field.

#### Chromatin immunoprecipitation sequencing

A total of  $1 \times 10^7$  to  $5 \times 10^7$  cells were harvested in 15 mL tubes and fixed with 2 mL of 1% paraformaldehyde for 10 minutes at room temperature, which was stopped by 2 mL of 250 mmol/L of glycine. Samples were rinsed with  $1 \times$  PBS twice and lysed twice with 1 mL of  $1 \times$  lysis/wash buffer (150 mmol/L NaCl, 0.5M EDTA pH 7.5, 1M Tris pH 7.5, 0.5% NP-40). Samples were filtered through a 29 G needle during each lysis process, and were harvested by centrifuge. Cell pellets were resuspended in lysis buffer (1% SDS, 10 mmol/L EDTA pH

8.0, 50 nmol/L Tris pH 8.0) and sonicated in a Covaris sonicator. The sonicated samples were subsequently centrifuged to remove debris and supernatants were diluted  $5 \times$  with dilution buffer (0.01% SDS, 1% Triton X-100, 1.2 mmol/L EDTA pH 8.0, 150 nmol/L NaCl). The primary antibodies were then added and incubated by rotation at  $4^\circ\text{C}$  overnight. Dynabeads Protein G beads (Life Technologies) were added the next morning and incubated by rotation for 4 hours. The beads were washed with  $1 \times$  lysis/wash buffer followed by wash in cold TE buffer. DNA molecules were reverse crosslinked, purified, and subject to library preparation and sequencing on Illumina HiSeq platform.

#### IHC staining and quantification

Tissue microarrays were purchased from US Biomax Company (Col05-118e, ES8011a, PA804a). All slides were deparaffinized and immersed before incubating with anti-HNF4A antibody. IHC staining was performed at room temperature using a PowerVision Homo-Mouse IHC Kit (ImmunoVision Technologies). Hematoxylin (Sigma) was used for counterstaining. HNF4A immunopositivity was scored as follows: 0, no staining or sporadic staining in  $<5\%$  cells; 1, weak staining in 5% to 25% cells; 2, weak staining in 26% to 50% of tumor cells; 3, strong staining in 26% to 50% cells; and 4, strong staining in  $>50\%$  cells.

#### Xenograft assays in nude mice

For shRNA-based xenograft experiments has been approved by Animal Experimental Committee of Soochow University. Twelve male 6-week-old mice were randomly separated to two groups and subcutaneously injected with  $2 \times 10^6$  Eso26 cells expressing inducible scrambled shRNA or shRNA against HNF4A. After tumor inoculation, all mice were fed with 2 mg/mL doxycycline (Abcam, ab141091) containing water. For the BI-6015 experiment, 12 male 6-week-old mice were subcutaneously injected with  $2 \times 10^6$  Eso26 cells. Either BI-6015 or vehicle control (5% DMSO+45% PEG 300+H<sub>2</sub>O) were administered three times per week by intraperitoneal injection at either 50 or 100 mg/kg. Xenograft size was measured two times per week for a total of 4 weeks. Mice were euthanized at the end of experiment and xenograft tumors were extracted for analysis. All animal experiments have been approved by Animal Experimental Committee of Soochow University.

#### Construction of expression vectors

The pBABE-puro-HNF4A expression vector was amplified on the basis of CMV-68 HNF4A (Addgene, #31092) and a 3xFLAG-tag was added via PCR. The amplified 3xFLAG-tagged HNF4A was then cloned into pBABE-puro vector (Addgene, #1764). The shRNA expression vector was designed on the basis of siRNA sequences (provided in Supplementary Table S1) and cloned into Tet-pLKO-puro vector (Addgene, #21915) and pLKO-puro vector (Addgene, #8453). To produce viral particles, recombinant viral vectors and packaging vectors were cotransfected into 293T cells. Supernatants containing viral particles were harvested and filtered through a 0.45  $\mu\text{m}$ /L filter 48 hours after transfection. GIAC cells were then infected with the virus in the presence of 10 mg/mL polybrene.

#### In vitro cell proliferation assay

A total of 3,000 to 5,000 GIAC cells were seeded into 96-well plates and cultured for indicated periods of time. For inhibitor treatment, culture medium was replaced with either DMSO- or BI-6015-containing medium after 24 hours. Cell proliferation was measured by staining of 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT).

### Luciferase reporter assay

Three enhancer elements (E1: chr20:42,838,300–42,839,699; E2: chr20:42,931,956–42,930,286; E3: chr20: 43,035,000–43,037,051; and negative control chr20:42,860,825–42,862,318) were cloned into pGL3-Promoter luciferase reporter vector (Promega). HNF4A promoter luciferase reporter vector HNF4A-P2-2200 was purchased from Addgene (31062). HNF1A promoter luciferase reporter vector HNF1A-P-1526 was cloned in house. A Renilla luciferase control vector was cotransfected for normalization. After 48 hours of transfection, the reporter activity was measured by the Dual-Luciferase Reporter Assay System (Promega, E1910).

### Preparation of 4C-seq library

4C-Seq libraries were prepared using our previous described protocols (23). In brief, Eso26 cells were single-cell suspended, and the chromatin was cross-linked with 1% formaldehyde at room temperature. Cells were lysed and DNA was digested with the restriction enzyme TaqI (R0149L, NEB). Digested DNA was next ligated using T4 DNA ligase (EL0013, Thermo Fisher Scientific), followed by removal of cross-link using Proteinase K (19133, Qiagen), yielding 3C libraries. These 3C libraries were subjected to a second enzyme digestion by Csp6i (R0639L, NEB), followed by another ligation using T4 DNA ligase. A total of 3.2 µg of the resulting 4C templates was used for a scale-up inverse nested PCR for 29 cycles. The PCR products were purified using Macherey Nagel Nucleospin Gel and PCR Purification Kit (Takara Bio). 4C sequencing libraries were made from the PCR products using ThruPLEX DNA-seq Kit (R400427; Takara Bio). The libraries were subjected to Agencourt AMPure XP Bead clean-up (A63881; Beckman Coulter) using a bead-to-DNA ratio of 1:1. Libraries were sequenced using  $1 \times 75$ bp for 2 million read depth using MiniSeq system (Illumina).

### Primer design for 4C-seq

The inverse primers were designed on the basis of a viewpoint region. DpnII restriction sites flanking the region of interest were identified and the sequence between the nearest DpnII and TaqI restriction sites were selected as the viewpoint region. On the basis of this region, 4C primers (F1-AGGAGACGGACCTTAATCAGATC, R1-AATGTGACCGCTTCCCTAAGCT) were designed with the following settings: optimal melting temperature of 60°C (57°C–62°C); GC content: 40% to 60%.

### 4C-seq data analysis

4C-Sequencing (4C-seq) data were analyzed using the R package r3CSeq (24). Briefly, for each replicate, the raw reads were aligned to the reference human genome, which was masked for the gap, repetitive, and ambiguous sequences. The masked version of the genome was downloaded from the R Bioconductor repository (BSgenome.Hsapiens.UCSC.hg19.masked). The number of mapped reads for each window were counted and normalized to obtain RPM (reads per million per window) values, which was used to perform statistical analysis. Interacted regions were plotted and visualized on UCSC Genome Browser as custom tracks. 4C-seq data have been deposited to Gene Expression Omnibus (GSE132813).

### ChIP-seq data analysis

We generated ChIP-seq data for HNF4A and GATA4 in Eso26 cells in the present work, which has been deposited to Gene Expression Omnibus (GSE132813). ChIP-seq data of GATA6, KLF5, and ELF3 in Eso26 cells, and H3K27Ac ChIP-seq data in the following cell lines were generated in-house recently (GSE132686 and GSE106563;

refs. 25, 26): Eso26, Flo1, JH-EsoAD1, KYAE1, OACM5.1, OACp4C, OE19, OE33, SKGT4, KYSE70, KYSE140, KYSE510, TT, TE5, and TE7. H3K27Ac ChIP-seq from STAD normal ( $n = 20$ ) and tumor samples ( $n = 20$ ) were from Wen Fong Ooi and colleagues (27). For data analysis, ChIP-seq reads were aligned to reference genome (UCSC hg19) using Bowtie (v1.1.2; ref. 28) or BWA (29) under default parameters. PCR duplicates were removed using Picard tools. Peaks were filtered to remove all peaks overlapping ENCODE blacklisted regions. The bigwig files were generated by bamCompare in DeepTools (v3.1.3; ref. 30).

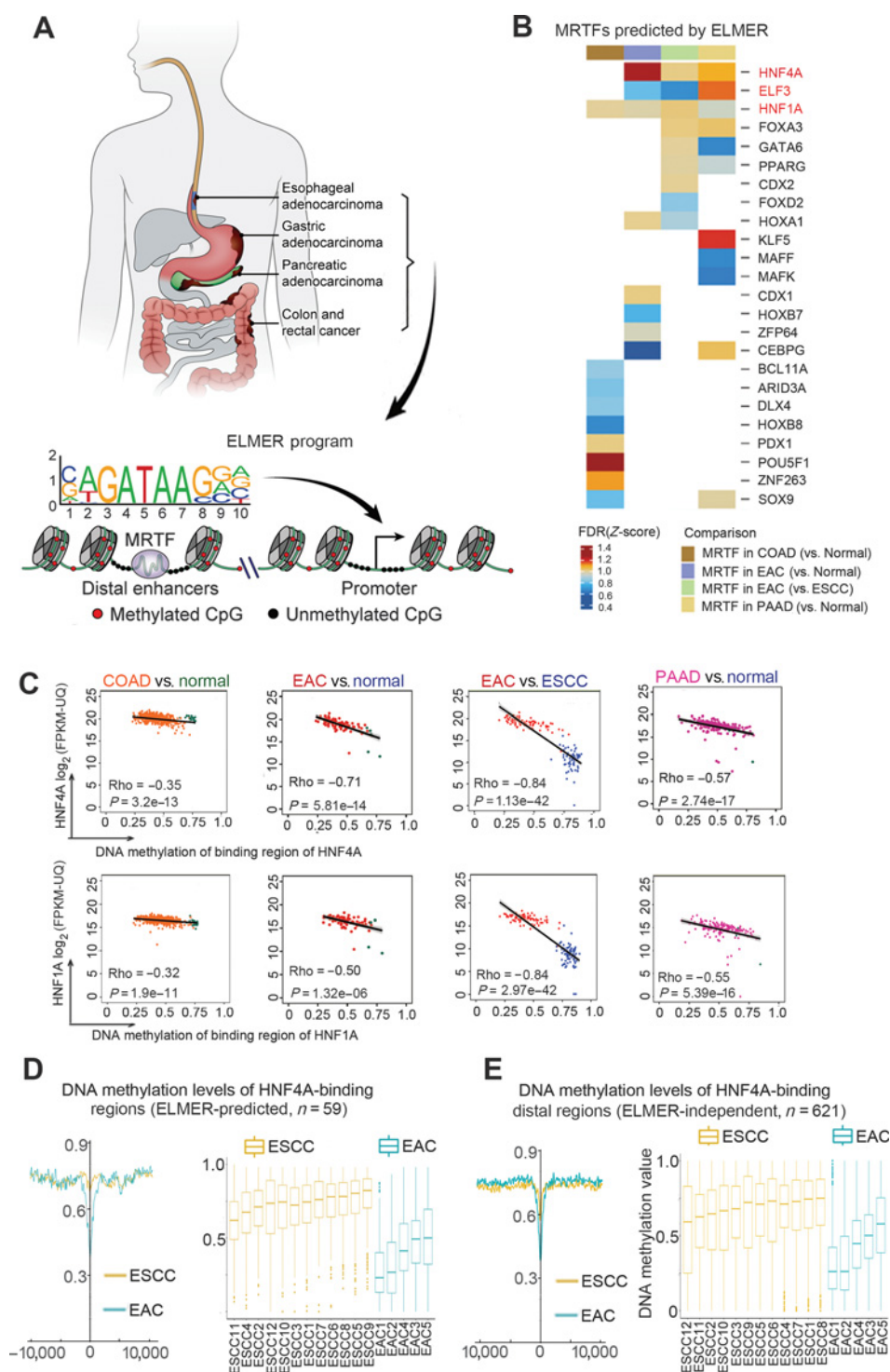
Motif enrichment analyses were performed using bedtools (v2.27.1; ref. 31) and HOMER (v4.10.3; ref. 32). Briefly, genomic regions of E1 to E3 and HNF4A promoter were divided into 557 subregion (25bp each) using makewindows function in bedtools. The resultant file containing 557 subregion were used to identify enriched motifs using HOMER with the parameters—size 200—lens 8, 10, 12.

## Results

### Identification of hyperactive MRTFs in GIAC using ELMER

To identify MRTFs with higher transcriptional activity in GIACs than adjacent normal samples, we analyzed paired transcriptome sequencing (RNA-seq) and DNA methylation array data (Infinium HM450 Beadchip) from TCGA samples, including EAC (78 tumors vs. 5 normal), STAD (338 tumors vs. 2 normal), and COAD (389 tumors vs. 21 normal). We additionally included PAAD (177 tumors vs. 4 normal) because this cancer exhibits strong molecular similarity with a significant proportion of STAD, as introduced earlier (4). The ELMER method was applied to compare tumor and normal specimens in each cancer type to identify MRTFs with higher activity in tumor samples (Fig. 1A). We further contrasted EAC vs. ESCC tumors as they are biologically distinct subtypes from the same organ, serving as ideal analytical controls to discover EAC-specific networks. As described earlier, ELMER calculates the anticorrelation score of each MRTF candidate (corrected by FDR) through correlating DNA methylation levels occurring within distal-binding sites for specific MRTFs (inferred by motif analysis), with altered expression of the corresponding MRTFs. As a result (Supplementary Table S2), ELMER identified a total of 24 MRTFs (FDR  $q$ -value  $< 0.05$ ), and notably, 10 of them were shared in at least two different comparisons (Fig. 1B). This result suggests that different GIACs indeed exhibit notable similarity in terms of their hyperactive MRTFs, which likely control similar downstream transcriptional programs. Some of these MRTFs are well-established GIAC specific oncogenes, such as *CDX2* (33, 34), *GATA6* (10, 35), *KLF5* (10, 11), and *HNF1A* (36), highlighting that ELMER is capable of identifying cancer-specific MRTFs.

Because this work was aimed to discover hyperactive MRTFs shared by different GIACs, we focused on the only three candidates (HNF4A, ELF3, and HNF1A) that were identified in at least three comparisons (Fig. 1B). Among them, HNF4A was particularly interesting because: (i) it had the lowest FDR  $q$  value among all 24 MRTFs; (ii) it has been shown as an upstream regulator of HNF1A (another top-ranked MRTF) in hepatocytes (37), which we functionally validated in different types of GIAC cells (Supplementary Fig. S1A). Specifically, knockdown of HNF4A decreased the expression of HNF1A whereas the opposite was not observed (Supplementary Figs. S1B–S1C). Moreover, HNF4A directly occupied the promoter region of HNF1A in several different GIAC cell lines (Supplementary Fig. S1A). As shown in the scatter plots from each type of GIAC (Fig. 1C), HNF4A-high tumors had



**Figure 1.** Identification of hyperactive MRTFs in GIAC using ELMER. **A**, Schematic of the study design, showing tumor anatomical locations (top) and the ELMER approach for identifying MRTFs (bottom). **B**, ELMER analysis was applied in each type of GIAC to determine the anticorrelation score of each MRTF by correlating methylation levels within distal-binding sites for specific MRTFs, with altered expression of the corresponding MRTF. A color is shown in each cell where the MRTF was identified (FDR  $q < 0.05$ ), and the color represents the  $P$ -value normalized for each GIAC (Z-score). **C**, Scatter plots of top candidate MRTFs (HNF4A and HNF1A), showing the average DNA methylation level of predicted MRTF-binding regions (x-axis) vs. the mRNA expression of the MRTF (y-axis) in each sample (dot). **D** and **E**, Plots showing average WGBS methylation levels of predicted HNF4A binding sites, either averaged across all samples of the same subtype to show spatial pattern (left) or for each sample individually (right). **D**, ELMER-predicted HNF4A binding sites. **E**, ESCA-specific ATAC-seq peaks overlapping an HNF4A binding motif.

Downloaded from <http://aacrjournals.org/cancerres/article-pdf/80/13/2722/212791628/2722.pdf> by guest on 24 May 2025

significantly lower DNA methylation in predicted HNF4A binding regions than either HNF4A-low tumors or normal tissues. The ELMER analysis was based on the Infinium HM450 design, which covers less than 5% of all enhancers (38). To validate the ELMER results, we generated WGBS data from an independent cohort of 5 EAC and 12 ESCC samples, which covered an average of 97.29% of all CpGs in the genome with at least five sequencing reads. We first

determined the DNA methylation levels of ELMER-predicted HNF4A-binding regions and confirmed that indeed they were lower in EAC than ESCC WGBS samples (Fig. 1D). To further validate the demethylation of HNF4A binding regions genome-wide, we next performed a completely independent analysis without using the ELMER results. Briefly, we analyzed methylation within ESCA-specific open chromatin regions from TCGA ATAC-seq

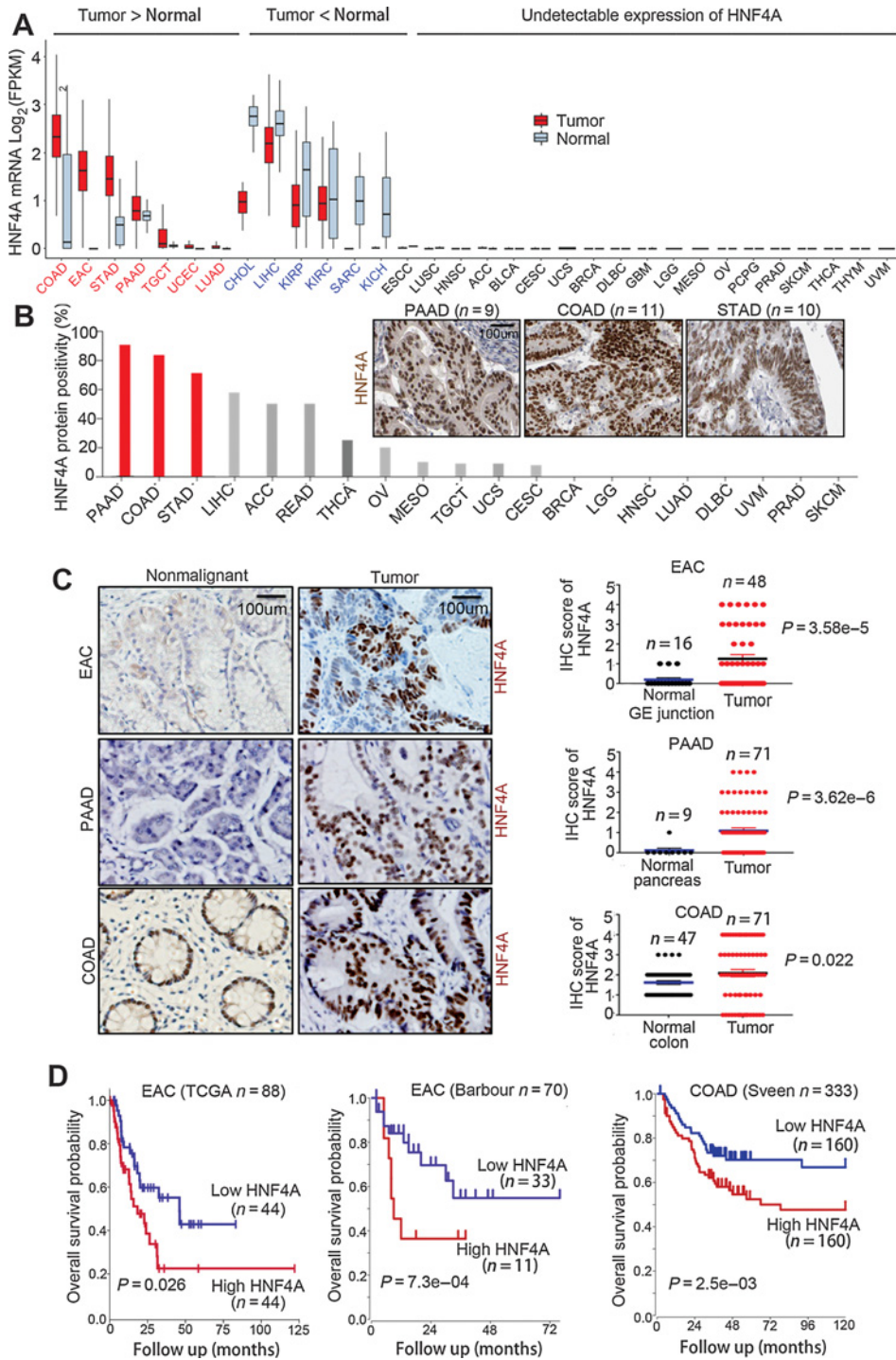
(12 ESCC and 6 EAC samples; ref. 22), which contained HNF4A binding motif sequences (see Materials and Methods). As expected, these predicted HNF4A binding sites were consistently hypomethylated in EAC samples than ESCC samples (Fig. 1E). Moreover, the DNA methylation difference was much more pronounced in distal regions than promoters (Supplementary Fig. S1E), consistent with the distal location of binding sites from HNF4A ChIP-seq in GIAC cells (Supplementary Fig. S1D).

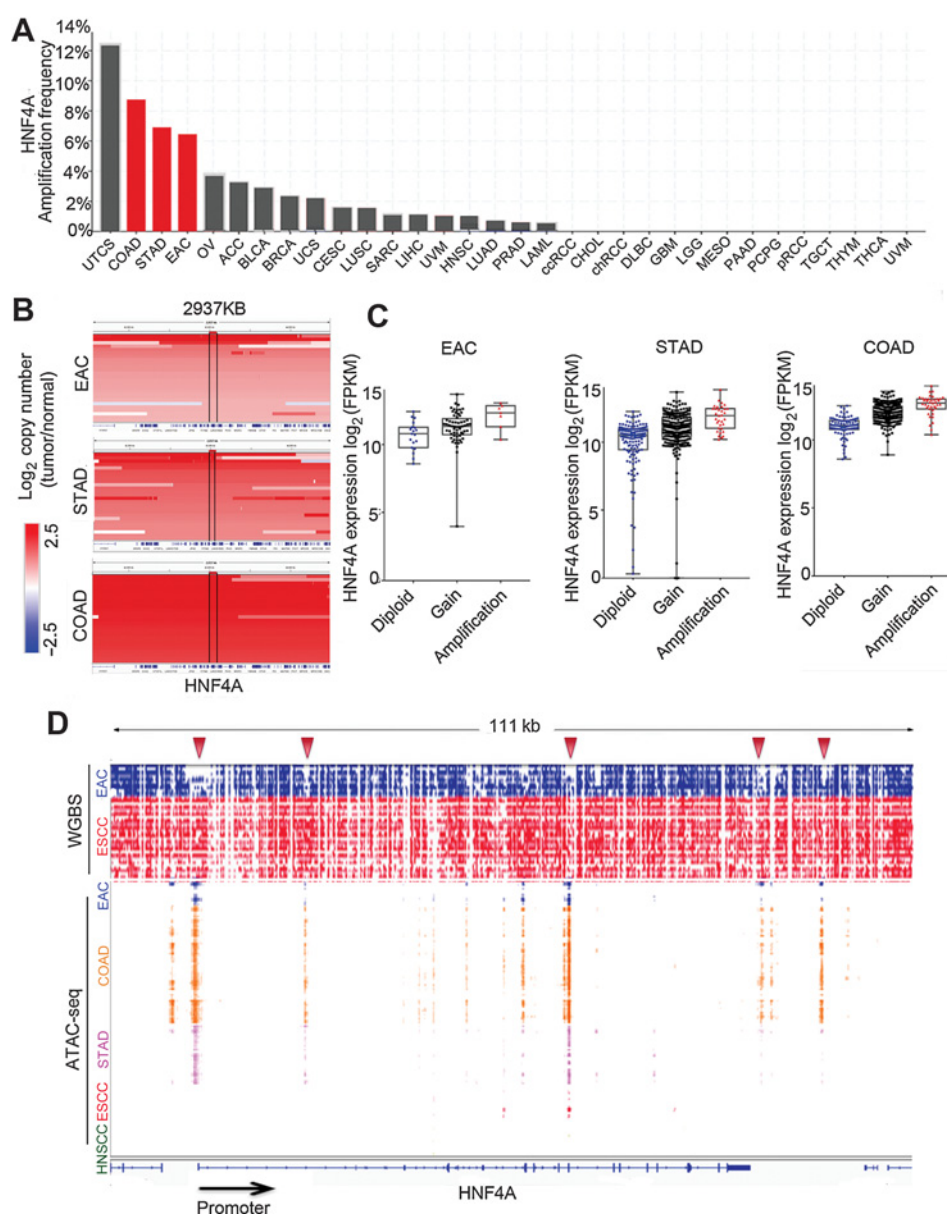
### Unique genomic and epigenomic activation of HNF4A in GIACs

The above results suggest that HNF4A is highly activated in GIACs. Indeed, Pan-Cancer RNA-seq data from TCGA showed that the expression of HNF4A was upregulated in only 7 of 32 cancer types and top four were all GIACs (Fig. 2A). Although HNF4A was expressed comparatively high in hepatocellular, cholangio, and kidney cancers, it was downregulated comparing with the correspondent nonmalignant tissues. In the remaining 19 tumor types, HNF4A was

**Figure 2.**

HNF4A is specifically and highly expressed in GIACs. **A**, Expression of HNF4A in RNA-seq data from TCGA Pan-Cancer samples. Each box represents a cancer type (TCGA abbreviations can be found at Supplementary Table S3). **B**, HNF4A protein expression was interrogated using the IHC data from Human Protein Atlas project, and GIACs are highlighted in red. Representative IHC photos are shown as inserted panels. **C**, IHC staining on three GIAC cohorts of internal samples. **D**, Overall survival analyses of HNF4A expression in both patients with EAC and COAD, from TCGA and other named studies.



**Figure 3.**

Genomic and epigenomic activation of HNF4A is specific to GIACs. **A**, HNF4A gene amplification in SNP-array data from TCGA Pan-Cancer samples, with GIACs highlighted in red. **B**, Integrative Genomics Viewer plots of HNF4A-amplified GIACs samples. The relative copy number ratio between tumor and normal samples is indicated by the color bar. **C**, mRNA level of HNF4A in GIAC samples with different HNF4A genomic status. Copy number gain refers to samples with between two and four copies of the HNF4A locus, and amplification refers to those with greater than four copies. **D**, Integrative Genomics Viewer plots showing HNF4A locus using both in-house WGBS samples (EAC = 5, ESCC = 12) and TCGA ATAC-seq samples (EAC = 6, COAD = 38, STAD = 21, ESCC = 12, HNSCC = 9). Arrows, several regions that are specifically accessible and demethylated in GIAC tumor types. The methylation value of each CpG site (from 0% to 100%) is proportional to the height of the bar.

undetectable in either nonmalignant or cancerous samples (Fig. 2A). HNF4A protein was then interrogated using the IHC results from The Human Protein Atlas (39, 40), and consistently, its protein level was most abundant in GIACs across all cancer types, displaying a prominent lineage-specific expression pattern (Fig. 2B). We next performed IHC staining on three cohorts of internal GIAC samples, and confirmed that HNF4A protein was significantly upregulated in EAC, PAAD, and COAD tumors compared with adjacent normal samples (Fig. 2C). Moreover, higher HNF4A expression was significantly associated with worse overall survival in both patients with EAC and COAD (Fig. 2D). Searching for the mechanisms underlying the GIAC-specific overexpression of HNF4A, we found that with the exception of uterine tumors, the HNF4A gene locus was specifically amplified in 7% to 9% of different GIAC types, and had low-level copy number gain in a majority of GIAC samples (Fig. 3A and B). Expectedly, HNF4A-amplified and copy number gain tumors had higher HNF4A mRNA level than nonamplified tumors (Fig. 3C).

However, we reasoned that copy number gain of HNF4A did not completely account for its mRNA upregulation, and epigenetic changes were also involved. This was because gene amplification only increased its mRNA log<sub>2</sub> FPKM level by 2 to 4 (Fig. 3C), whereas its log<sub>2</sub> FPKM was elevated by at least 10 compared with most non-GIAC samples (Fig. 2A). To explore the possibility of epigenetic upregulation, we queried TCGA ATAC-seq data (Fig. 3D) and observed that indeed the HNF4A promoter was markedly more accessible in GIAC samples than other cancer types (such as squamous tumors). Furthermore, several distal HNF4A intronic regions also exhibited higher accessibility (Fig. 3D). We aligned these ATAC-seq tracks with in-house WGBS data and found that GIAC-specific open chromatin regions were concordantly demethylated in EACs but fully methylated in ESCCs (red arrows, Fig. 3D), suggesting that these are cis-regulatory elements potentially contributing to the transcriptional activation of HNF4A in GIAC in a lineage-specific manner.

### HNF4A specifically promotes the proliferation and survival of GIAC cells

The above results demonstrate that HNF4A is specifically over-expressed in Pan-GIAC samples because of both genomic and epigenomic alterations. We thus hypothesized that this is a *bona fide* MRTF playing a functional role in GIAC biology. Indeed, HNF4A was reported to be essential for the proliferation and metabolism of STAD cells (10, 41, 42). Nevertheless, it is unclear whether HNF4A activity is functionally required for other GIAC types including EAC, PAAD, and COAD. Moreover, its biological functions and associated molecular mechanisms in GIAC remain incompletely understood. To address these questions, we initially silenced HNF4A expression in cell lines from EAC, COAD, and PAAD with high endogenous level of this MRTF. Cell proliferation and colony formation capacity were strongly and consistently suppressed in HNF4A-knockdown group compared with control cells (Fig. 4A–C). We also noted that HNF4A-depletion caused massive apoptosis in all types of GIAC cells (Fig. 4D; Supplementary Fig. S2A and S2B). We next asked whether excessive levels of HNF4A could promote GIAC proliferation in GIAC cells with low endogenous HNF4A. Indeed, its ectopic overexpression enhanced the proliferation and colony formation in EAC, COAD, and PAAD cells (Fig. 4E–G).

Considering that HNF4A is specifically activated in GIAC cancers, we next sought to test whether the functional significance of this MRTF exhibits similar GIAC specificity. Expectedly, HNF4A mRNA level was the highest in GIAC cells in a pan-cancer analysis of over 1,000 cell lines from CCLE (Cancer Cell Line Encyclopedia; Fig. 4H; ref. 43). Strikingly, HNF4A was uniquely required for the viability of GIAC cells in the unbiased high-throughput shRNA screen (Fig. 4H). Consistent with our WGBS data, the HNF4A promoter was hypomethylated and correlated with high expression uniquely in GIAC cell lines in the CCLE dataset (Fig. 4I). We next tested whether its promoter methylation lever was similarly associated with its essentiality for cell viability. Indeed, cells with lower methylation of HNF4A promoter were more dependent on HNF4A for proliferation in a GIAC-specific manner (Fig. 4J). These results together identified HNF4A as a prominent lineage-specific oncogene in GIAC tumors.

### Targeting HNF4A suppresses growth of GIAC cancer cells and xenografts

Using a doxycycline-inducible shRNA system, we tested HNF4A function *in vivo* using an EAC-derived xenograft model (Fig. 5A). Both the growth and weight of the tumor xenografts were significantly inhibited by HNF4A-silencing (Fig. 5B and C). IHC analysis showed that Ki-67 expression (a marker of cell proliferation) was down-regulated in HNF4A-knockdown tumors (Fig. 5D). We next investigated the function of HNF4A by a small-molecule inhibitor (BI-6015; ref. 44) developed against this MRTF. This antagonist binds to HNF4A with high affinity by forming hydrogen bonds with HNF4A Arg226 and Gly237, which occupies a hydrophobic pocket. This occupancy by BI-6015 inhibits the transcriptional function of HNF4A by preventing interaction with its ligand binding pocket. To test the effects of BI-6015 *in vitro*, we first conducted luciferase reporter assays using promoters from two direct HNF4A targets, *HNF1A* and *HNF4A* itself (direct binding at the *HNF1A* promoter is shown in Supplementary Fig. S1A, and autoregulation of the *HNF4A* promoter is shown below). BI-6015 strongly inhibited the reporter activities of both promoters, as well as the mRNA levels of both genes (Fig. 5E). In an *in vitro* proliferation assay, different GIAC cells from either EAC, PAAD, or COAD were significantly more sensitive to this compound than non-GIAC cells (Fig. 5F and G), and that cell sensitivity to BI-6015 (measured by IC<sub>50</sub>)

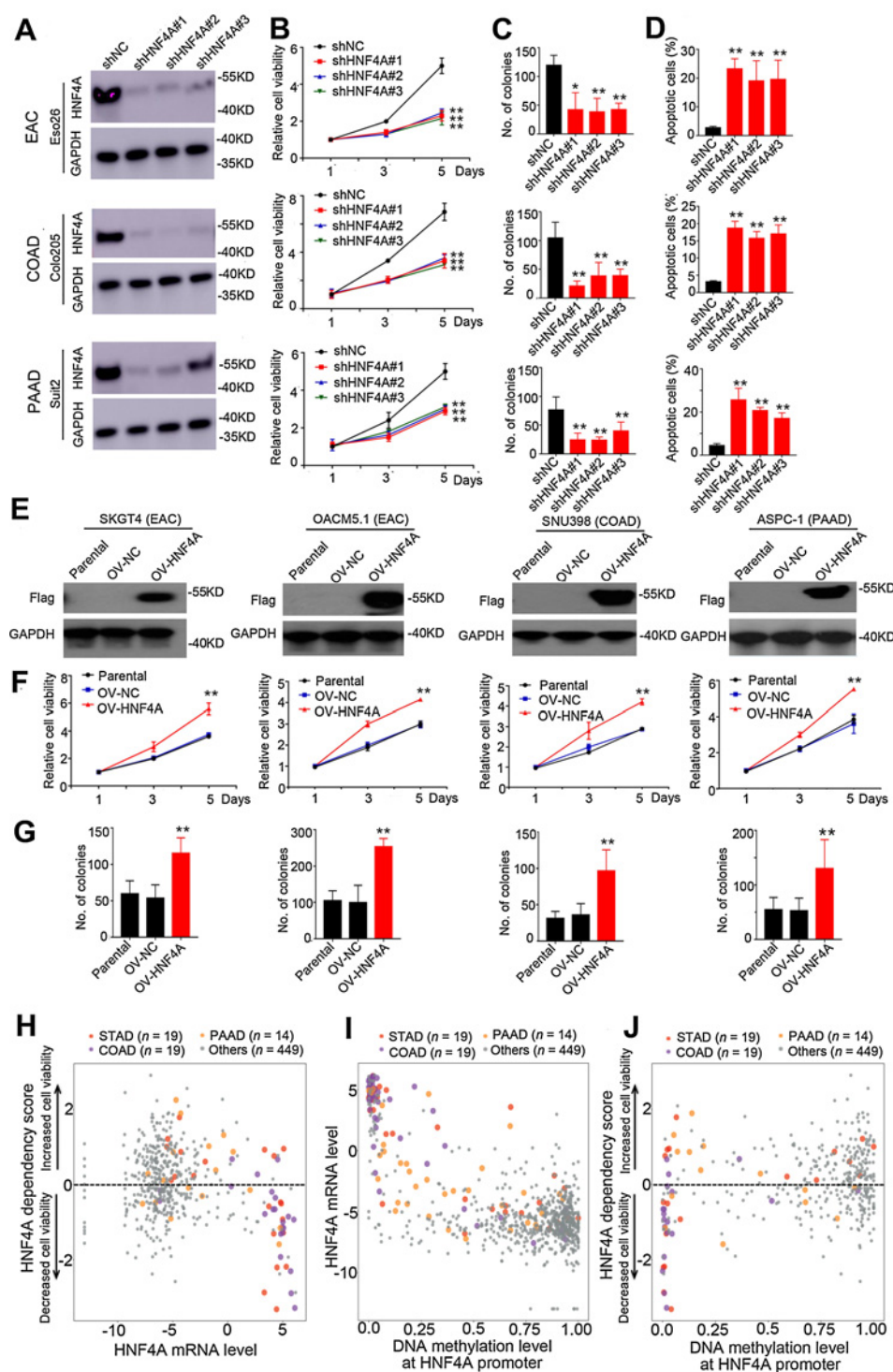
was negatively correlated with the expression of HNF4A (Fig. 5H), again confirming the on-target effect. The mouse xenograft assay showed that BI-6015 potentially inhibited the growth of GIAC tumors but did not cause systematic toxicity (Fig. 5I and J; Supplementary Figs. S3A and S3B). Together, these results obtained from genetic and chemical approaches demonstrate that HNF4A functionally promotes the proliferation and survival specifically in GIAC cancer cells.

### Upstream regulation of HNF4A transcription by GIAC MRTFs

To probe the mechanism underlying GIAC-unique epigenomic activation of HNF4A, we first performed circularized chromosome conformation capture (4C) sequencing to explore the interaction landscape of the *HNF4A* locus in EAC cells, using its promoter as the 4C-seq bait (Viewpoint). We identified that all of the significant interactions ( $q < 0.001$ ) were restricted to the 500kb window flanking the *HNF4A* promoter (Fig. 6A). Importantly, by cross-referencing H3K27Ac ChIP-seq data generated in the matched sample, we found that the majority of interacting regions had positive H3K27Ac signals (Fig. 6B), suggesting their potential regulatory function. On the basis of the  $q$  value, we identified the top 10 most significant interactions that overlapped with three H3K27Ac+ regions (referred to as E1–E3, Fig. 6C). To search for potential TFs occupying these regions, we performed motif enrichment analysis and found that top enriched motifs included the GATA family, ELF3 and HNF4A itself (Fig. 6D). This result was encouraging because GATA4/6 are known GIAC-specific MRTFs and ELF3 was the second most highly ranked MRTF identified earlier (Fig. 1B). Furthermore, the enrichment of HNF4A motif is also in keeping with the notion that self-regulation is an important and common property of many MRTFs in different cell types (45, 46). To validate this motif analysis, we performed ChIP-seq, wherein we additionally included KLF5 as our recent work identified it as an integral component of core regulatory circuitry in EAC (47). Importantly, the ChIP-seq results confirmed that the four regulatory regions (promoter region and E1–E3) were indeed occupied by GATA4, GATA6, ELF3, HNF4A, and KLF5 in different types of GIAC cell lines (Fig. 6C; Supplementary Fig. 4). Notably, both E3 and promoter regions were co-occupied by all five MRTFs (Fig. 6C), and these regulatory elements were highly concordant across different GIAC cells (Supplementary Fig. S5). To test the transcriptional activity of these regulatory elements, they were cloned individually into the luciferase reporter vector. All three enhancer elements showed robust reporter activities uniquely in GIAC cells but not in SCC cells (Fig. 6E). Moreover, silencing each of the five MRTFs inhibited the activities of promoter and E3 region. In addition, ELF3, KLF5, and HNF4A contributed to the activity of E1 and E2 (Fig. 6F). These results demonstrated strong and complex regulation of HNF4A transcription, which is GIAC-specific and cooperatively controlled by GATA4, GATA6, ELF3, KLF5, and HNF4A itself.

Importantly, individually silencing each MRTF significantly down-regulated the expression of HNF4A at both mRNA and protein levels, confirming the regulation of HNF4A by these factors (Fig. 7A and B). Given the colocalization of these five MRTFs in HNF4A promoter and enhancers, we tested if there existed direct protein–protein interactions among these upstream factors. Immunoprecipitation (IP) analysis showed that GATA4, GATA6, and KLF5 formed protein complexes in GIAC cells (Fig. 7C; Supplementary Fig. S6B). Although this protein complex did not contain HNF4A or ELF3, these two MRTFs might participate in the transcription co-operation through indirect mechanisms (indirect cooperativity, such as the “Billboard model”; ref. 48). Considering that some of these MRTFs (e.g., GATA4/6 and KLF5) have been reported to have copy number gains in GIAC





**Figure 4.**

HNF4A specifically promotes the proliferation and survival of GIAC cells. In EAC (Eso26), COAD (Colo205), and PAAD (Suit2) cell lines, HNF4A expression was silenced by three different shRNAs and followed by Western blotting assay (A), cell proliferation assay (B), colony formation assay (C), and apoptosis assays (D). In E, HNF4A was ectopically expressed and validated by Western blotting in EAC (SKGT4 and OACM5.1), COAD (SNU398), and PAAD (ASPC-1) cells. F and G, HNF4A-overexpressing cells were subjected to cell proliferation (F) and colony formation (G) assays. OV-NC, overexpression control; OV-HNF4A, overexpression of HNF4A. Data are presented as mean  $\pm$  SD. \*,  $P < 0.05$ ; \*\*,  $P < 0.01$ . H, High-throughput shRNA screen in a pan-cancer analysis of 500 cell lines from CCLE and DepMap. Scatter plot showing the HNF4A shRNA dependency score vs. the expression of HNF4A. I, Scatter plot showing the mRNA expression of HNF4A and the methylation of its promoter. J, Scatter plot showing the HNF4A dependency score vs. the methylation of HNF4A promoter. Each dot represents one cell line, and GIAC cells are highlighted.

samples (10, 11), and our present work identified that *HNF4A* was also amplified specifically in GIACs, we next analyzed the genomic status of these MRTFs in TCGA dataset. The copy number gains of *GATA4/6* and *KLF5* were confirmed in both EAC and STAD cohorts (Supplementary Fig. S6A). Notably, the amplification of these MRTFs exhibited a mutually exclusive pattern (Supplementary Fig. S6A), strongly supporting our results that these MRTFs converge on HNF4A sig-

naling, and thus redundant gain-of-function events are not required in the same GIAC samples.

#### HNF4A transcriptionally activates IL signaling pathway in GIAC cells

To investigate the downstream targets of HNF4A in GIAC cells, we performed RNA-seq upon knockdown of this MRTF in Eso26 cells.

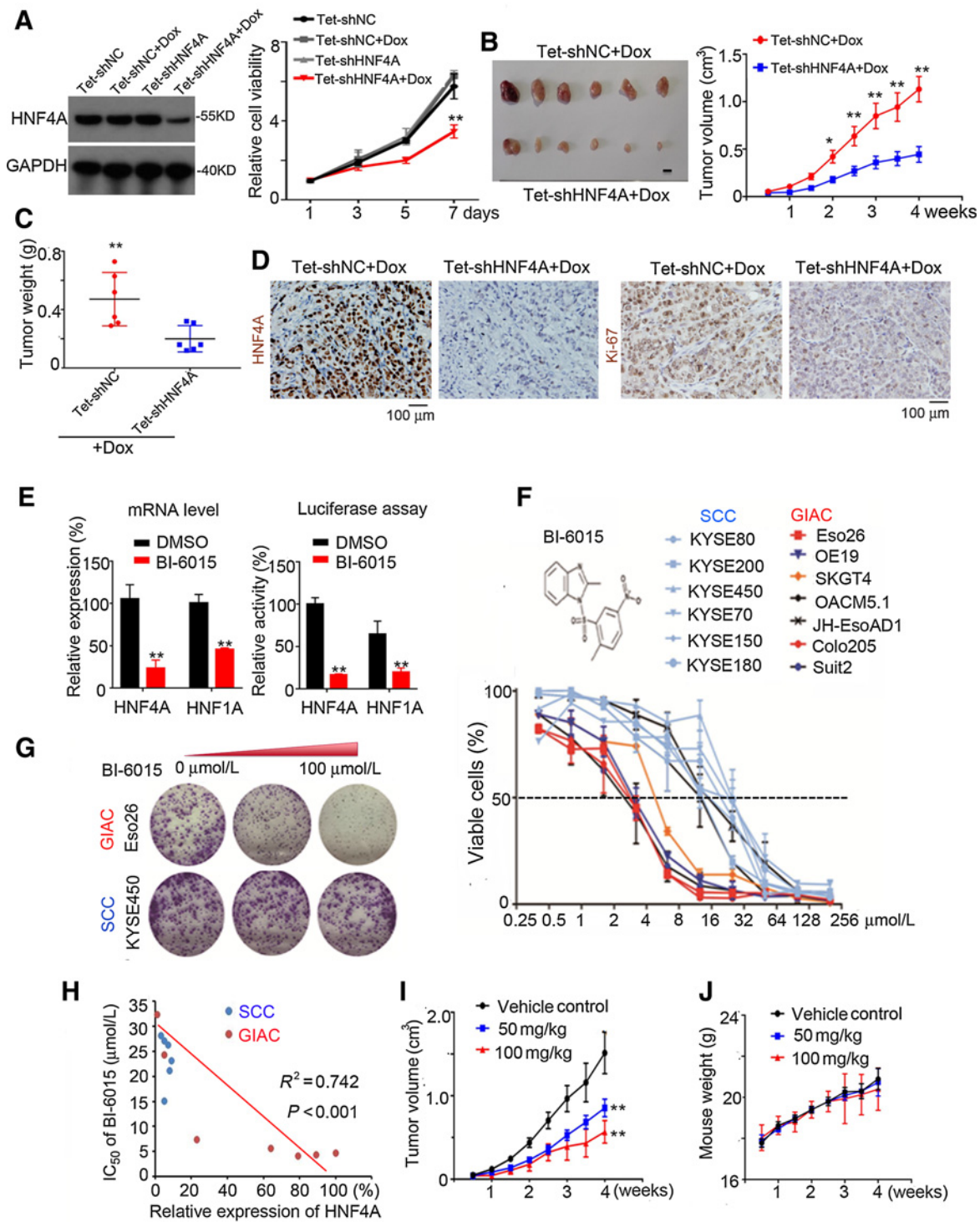
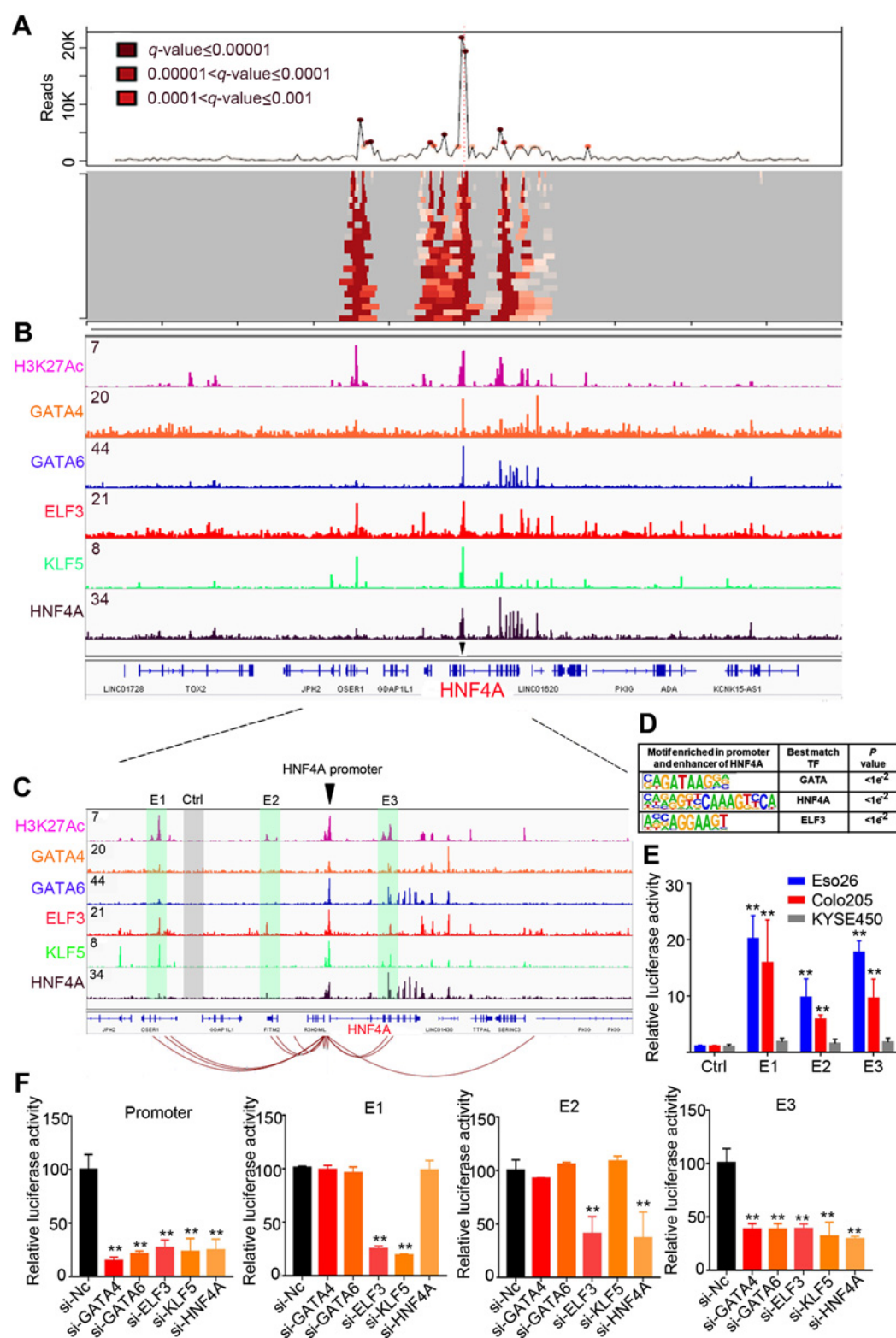


Figure 5.

Targeting HNF4A suppresses growth of GIAC cancer cells and xenografts. **A**, Western blotting assay (left) validating the knockdown of HNF4A using doxycycline (Dox)-inducible shRNA in Eso26 cells, which led to decreased cell proliferation *in vitro* (right). **B** and **C**, Tumor growth (**B**) and weight (**C**) of the xenografts were significantly inhibited by HNF4A-silencing. **D**, IHC staining of HNF4A and Ki-67 in xenograft samples. **E**, qRT-PCR analysis (left) and luciferase reporter assay (right) validating the on-target effect of BI-6015. **F**, Dose-response curves of GIAC and squamous cancer cell (SCC) lines treated with BI-6015 (72 hours). Data represent mean  $\pm$  SD of three replicates. **G**, Colony formation of Eso26 (EAC cell) and KYSE450 (SCC cell) treated with increasing dose of BI-6015. **H**, Scatter plot showing the negative correlation between cell sensitivity (measured by IC<sub>50</sub> of BI-6015) and HNF4A expression. Each dot represents one cell line. **I** and **J**, Xenograft assay showed that BI-6015 inhibited the growth of GIAC tumors (**I**) but did not affect the weight of the mice (**J**). Mean  $\pm$  SD are shown. \*,  $P < 0.05$ ; \*\*,  $P < 0.01$ .

**Figure 6.**

GIAC MRTFs occupy and regulate HNF4A promoter and enhancers. **A**, 4C-Seq assay showing the long-range interactions anchored on HNF4A promoter in Eso26 cells. Deeper red color indicates higher interaction frequency. **B**, ChIP-seq profiles for H3K27Ac and indicated MRTFs at HNF4A loci in Eso26 cells. **C**, Zoom in view of ChIP-seq signals. Connecting lines showing the most significant interactions detected by 4C. Shaded regions showing three enhancers (E1-E3) and one negative control (Ctrl) region that were separately cloned into the luciferase reporter vector. **D**, Motif enrichment analysis of promoter and E1-E3 regions. **E**, Enhancer activity measured by luciferase assays in Eso26, Colo205, and KYSE450 cells. **F**, Enhancer and promoter activity measured by luciferase assays in Eso26 cells upon silencing each of the 5 MRTFs. Mean  $\pm$  SD are shown. \*\*,  $P < 0.01$ .

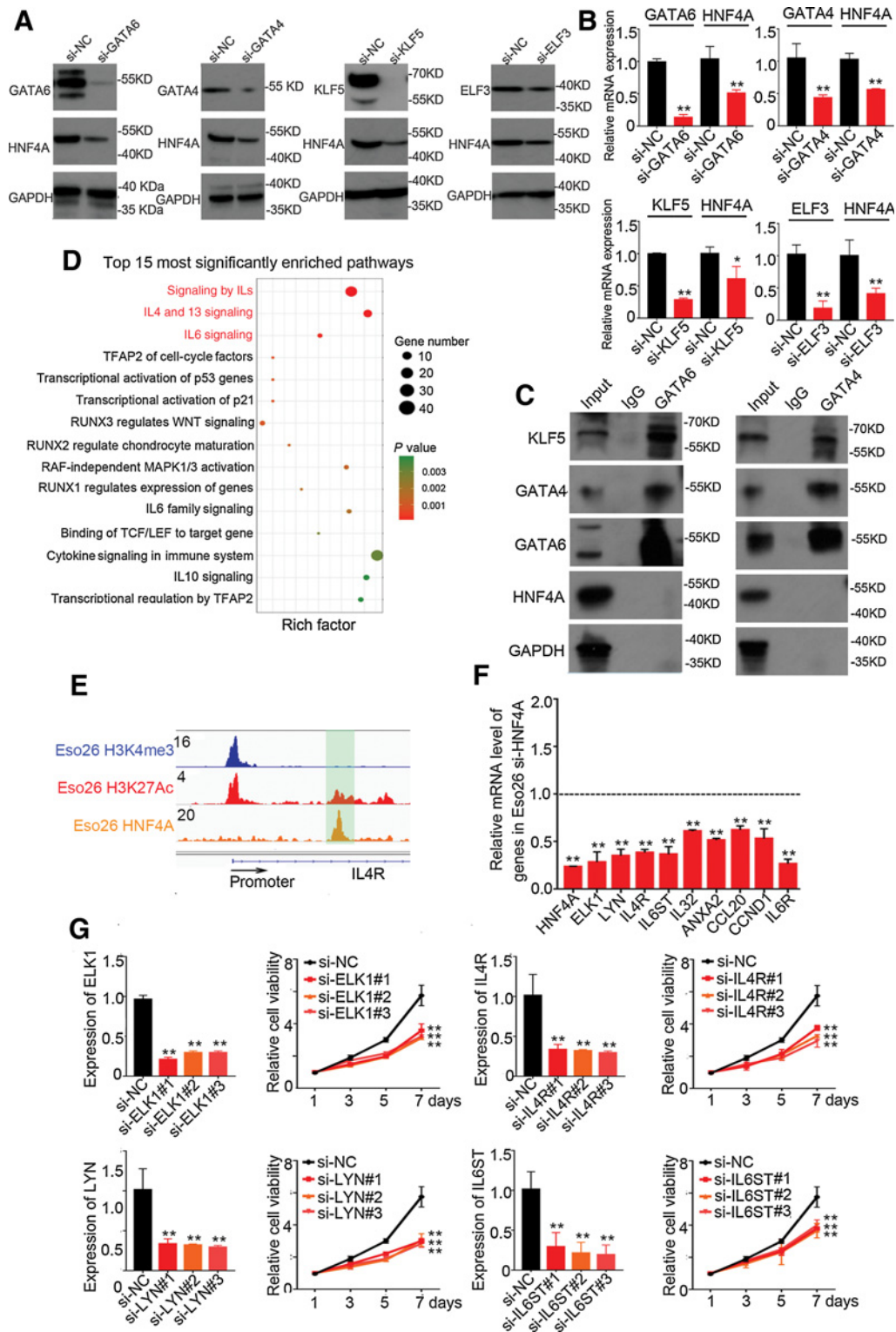


Figure 7.

Upstream and downstream regulation of HNF4A in GIAC cells. **A** and **B**, Western blotting (**A**) and qRT-PCR assay (**B**) upon each individual knockdown of indicated MRFs in Eso26 cells. **C**, Coimmunoprecipitation assay of GATA4, GATA6, and HNF4A in Eso26 cells. **D**, Pathway enrichment analysis of the differentially expressed genes upon silencing of HNF4A in Eso26 cells. **E**, HNF4A ChIP-seq showing its binding peak at the enhancer of IL4R in Eso26 (EAC) cells. **F**, qRT-PCR analysis showed that knockdown of HNF4A decreased the expression level of the HNF4A targets of interleukin signaling. **G**, IL4R, LYN, ELK1, and IL6ST were silenced by individual siRNAs in Eso26 cells, and cell proliferation was measured. Mean  $\pm$  SD are shown. \*,  $P < 0.05$ ; \*\*,  $P < 0.01$ .

Differential expression analysis of replicates identified 335 upregulated genes and 346 downregulated genes when compared with scrambled control (fold change > 2,  $P < 0.05$ ). Pathway enrichment analysis of the differential expressed genes showed that a number of cancer-related pathways were top ranked, such as IL signaling, WNT signaling, and MAPK1/3 pathway (Fig. 7D). Among these, IL signaling pathways were most notable, as they were the top 3 most significantly enriched. Specifically, a total of 21 genes of this pathway were downregulated upon knockdown of HNF4A. Integrating HNF4A ChIP-seq results found that 14 of the 21 pathway components were directly bound by HNF4A (Fig. 7E; Supplementary Table S3). The direct occupancy of HNF4A on these IL pathway genes was also highly concordant across different GIAC cell lines (Supplementary Fig. S7A). We randomly selected a number of targets and RT-PCR analysis confirmed their expression changes (Fig. 7F; Supplementary Figs. S7B and S7C). Some of these 14 direct targets (such as ELK1, LYN, IL4R, and IL6ST) have protumor properties in several cancer types (49–55) but their functions have not been explored in GIAC. We thus next tested whether they contributed to HNF4A-mediated oncogenic functions in GIAC cells. Importantly, silencing these genes significantly inhibited the proliferation of GIAC cells (Fig. 7G; Supplementary Figs. S7D–S7G). Together with our earlier result that HNF4A regulated HNF1A (another candidate GIAC MRTF, which also promoted GIAC cell proliferations, Supplementary Figs. S8A and S8B), these data demonstrate that HNF4A plays an important role in promoting GIAC proliferation and survival by transcriptionally activating many downstream targets, including HNF1A and those belonging to the IL signaling.

## Discussion

The clinical management of GIACs has advanced only modestly over the last several decades, and thus urgent needs exist to decipher the biological and pathologic basis of these malignancies to improve prevention, diagnosis, and therapy. Recently, unbiased genomic studies have strongly suggested unified genomic features among different types of GIACs, which has prognostic and therapeutic implications. For example, patients with MSI-high STAD and COAD have shown better clinical response to immune checkpoint blockade independent of anatomic tumor origin (5, 6). In addition, ERBB2 amplified and overexpressed patients with EAC, STAD, and COAD similarly benefit from anti-Her2 therapy. Thus, identification and characterization of unified genomic and/or epigenomic features across different types of GIACs may have important clinical implications. In this study, we aimed to identify such shared epigenomic characteristics across anatomically distinct GIACs. Motivated by the prominent changes in DNA methylation and gene expression in cell type- and cancer-specificity manner, we inferred gene expression networks, *cis*-regulatory elements as well as their upstream MRTFs in different types of GIACs using a novel mathematical tool (ELMER) we developed previously. This algorithm provides an unbiased systematic approach to reconstruct gene regulatory networks by integrating matched methylome and gene expression data.

We found that different types of GIACs display notable similarity with respect to their hyperactive upstream MRTFs, which in turn conceivably orchestrate similar downstream transcriptional programs. Indeed, some of the MRTFs (such as GATA4/6, KLF5, and CDX2) are known to have similar oncogenic functions in different types of GIACs. As ELMER is based on DNA methylation array data that have limited resolution and representability, we generated WGBS results in an independent cohort of EAC and ESCC samples and successfully

validated the ELMER prediction. These results highlight that our approach is capable of identifying *bona fide* MRTFs for cancer biology, and the list of candidate GIAC MRTFs warrant further investigation.

Focusing on the most significant MRTF, HNF4A, we first observed that the genomic and epigenomic activation of this factor was unique and specific in GIACs. Specifically, Pan-Cancer transcriptomic data from either TCGA or CCLE showed that the expression of HNF4A was uniquely high in GIACs comparing with other cancer types. Both internal IHC staining of different GIAC samples and Human Protein Atlas results consistently identified the same lineage-specific pattern of HNF4A protein. Moreover, higher HNF4A expression was associated with worse overall survival in different types of patients with GIAC, supporting a shared functional contribution to GIAC biology independent of anatomic origin. Genomic analyses demonstrated that copy number amplification accounted partially for the GIAC-specific overexpression of HNF4A. The other important source to drive the high HNF4A level was epigenomic activation, which was supported by ATAC-seq, WGBS, and H3K27ac ChIP-seq profiling. These epigenomic analyses identified that GIAC-specific open chromatin regions, which were concordantly demethylated and had high H3K27ac signals, likely contributed to the transcriptional activation of HNF4A in GIAC in a lineage-specific manner. Prompted by this epigenomic observation, we next characterized the underlying mechanisms. We first applied 4C-seq and identified three enhancer elements (E1–E3) interacting with the *HNF4A* promoter. Motif enrichment analysis of these regulatory regions coupled with ChIP-seq experiments identified that they were occupied by five upstream GIAC-specific MRTFs, GATA4/6, ELF3, KLF5, and HNF4A itself. Luciferase reporter assays and loss-of-function experiments of individual MRTFs confirmed that HNF4A transcription is activated by these five MRTFs through interacting with HNF4A promoter and three local enhancers. HNF4A has been shown to be regulated by GATA4, GATA6, and KLF5 in STAD cells (10), which is consistent with our unbiased motif search and epigenomic profiling. Furthermore, the self-regulatory property of HNF4A identified in this study has been observed in many TFs in different cell types, which is regarded as an important and common characteristic of key MRTFs (45, 56).

Functionally, HNF4A plays a role in normal development of the liver (57–59), kidney (60), and intestine (61, 62). In cancer biology, HNF4A has opposite roles in different cell types. For example, HNF4A is a tumor suppressor in hepatocellular cancer (63) while it was required for maintaining the proliferation of STAD cells (10, 41, 42). During the preparation of the manuscript, another study was published suggesting that HNF4A-mediated transcriptional program was more active in both Barrett's esophagus and EAC cells than normal esophageal squamous cells (64). Using analysis from unbiased shRNA library screen of over 500 cancer cell lines, we showed that HNF4A is uniquely and specifically essential for the viability of GIAC cells. We further confirmed this finding using both genetic and chemical tools in different types of GIAC cells *in vitro* and *in vivo*, highlighting HNF4A as a prominent lineage-specific oncogene in GIAC tumors. Interestingly, a previous report showed that HNF4A activity could be inhibited by metformin (41), making HNF4A an appealing and specific therapeutic target for GIAC, including early chemoprevention. We further performed ChIP-seq and transcriptomic analysis and identified Interleukin signaling as a key downstream pathway of HNF4A. Considering that agents inhibiting Interleukin signaling (such as Satralizumab, Pascolizumab, and Tralokinumab) are available clinically, these results together imply that HNF4A–IL axis might offer potential opportunity for the development of anti-GIAC therapeutic strategies.

In summary, applying ELMER to matched DNA methylation and expression profiles to reconstruct TF networks; the present work identifies a panel of hyperactive MRTFs shared by different GIAC cancers. As a top candidate, HNF4A is highlighted as a key oncogenic MRTF with prominent genomic and epigenomic activations in GIAC-specific manner. By providing mechanistic insights into the upstream and downstream regulation of HNF4A, this work significantly advances our understanding of the GIAC-specific gene regulatory networks, while providing potential therapeutic strategies against these common cancers.

### Disclosure of Potential Conflicts of Interest

S.J. Klempner is an ad-hoc advisory board member at Eli Lilly, Merck, Bristol Myers Squibb, Foundation Medicine, Inc., and Boston Biomedical; has ownership interest (including patents) in Turning Point Therapeutics; and is a unpaid consultant/advisory board relationship at NCCN and Hope for Stomach Cancer. No potential conflicts of interest were disclosed by the other authors.

### Authors' Contributions

**Conception and design:** J. Pan, B.P. Berman, H.P. Koeffler, D.-C. Lin

**Development of methodology:** J. Pan, J.T. Plummer, S. Chen, L.-W. Ding, S.A. Gayther, B.P. Berman

**Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.):** T.C. Silva, J.T. Plummer, T. Hamakubo, S.J. Klempner, J. Pan  
**Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis):** T.C. Silva, N. Gull, Q. Yang, J.T. Plummer, L.-W. Ding, S.J. Klempner, B.P. Berman, H.P. Koeffler

**Writing, review, and/or revision of the manuscript:** J. Pan, T.C. Silva, N. Gull, Q. Yang, J.T. Plummer, S. Gery, S.J. Klempner, B.P. Berman, H.P. Koeffler, D.-C. Lin

### References

- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;68:394–424.
- Arnold M, Soerjomataram I, Ferlay J, Forman D. Global incidence of oesophageal cancer by histological subtype in 2012. *Gut* 2015;64:381–7.
- Torre LA, Siegel RL, Ward EM, Jemal A. Global cancer incidence and mortality rates and trends—an update. *Cancer Epidemiol Biomarkers Prev* 2016;25:16–27.
- Liu Y, Sethi NS, Hinoue T, Schneider BG, Cherniack AD, Sanchez-Vega F, et al. Comparative molecular analysis of gastrointestinal adenocarcinomas. *Cancer Cell* 2018;33:721–35.e8. DOI: 10.1016/j.ccell.2018.03.010.
- Kim ST, Cristescu R, Bass AJ, Kim KM, Odegaard JI, Kim K, et al. Comprehensive molecular characterization of clinical responses to PD-1 inhibition in metastatic gastric cancer. *Nat Med* 2018;24:1449–58.
- Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, et al. PD-1 blockade in tumors with mismatch-repair deficiency. *N Engl J Med* 2015;372:2509–20.
- Cancer Genome Atlas N. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 2012;487:330–7.
- Cancer Genome Atlas Research N. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* 2014;513:202–9.
- Cancer Genome Atlas Research N, Analysis Working Group: Asan U, Agency BCC, Brigham, Women's H, Broad I, et al. Integrated genomic characterization of oesophageal carcinoma. *Nature* 2017;541:169–75.
- Chia NY, Deng N, Das K, Huang D, Hu L, Zhu Y, et al. Regulatory crosstalk between lineage-survival oncogenes KLF5, GATA4 and GATA6 cooperatively promotes gastric cancer development. *Gut* 2015;64:707–19.
- He P, Yang JW, Yang VW, Bialkowska AB. Kruppel-like factor 5, increased in pancreatic ductal adenocarcinoma, promotes proliferation, acinar-to-ductal metaplasia, pancreatic intraepithelial neoplasia, and tumor growth in mice. *Gastroenterology* 2018;154:1494–508.e13.
- Zhang X, Choi PS, Francis JM, Gao GF, Campbell JD, Ramachandran A, et al. Somatic superenhancer duplications and hotspot mutations lead to oncogenic activation of the KLF5 transcription factor. *Cancer Discov* 2018; 8:108–25.
- Lin CY, Erkek S, Tong Y, Yin L, Federation AJ, Zapotka M, et al. Active medulloblastoma enhancers reveal subgroup-specific cellular origins. *Nature* 2016;530:57–62.
- Berman BP, Weisenberger DJ, Aman JF, Hinoue T, Ramjan Z, Liu Y, et al. Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nat Genet* 2011;44:40–6.
- Hovestadt V, Jones DT, Picelli S, Wang W, Kool M, Northcott PA, et al. Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing. *Nature* 2014;510:537–41.
- Aran D, Sabato S, Hellman A. DNA methylation of distal regulatory sites characterizes dysregulation of cancer genes. *Genome Biol* 2013;14:R21. DOI: 10.1186/gb-2013-14-3-r21.
- Yao L, Shen H, Laird PW, Farnham PJ, Berman BP. Inferring regulatory element landscapes and transcription factor networks from cancer methylomes. *Genome Biol* 2015;16:105.
- Silva TC, Coetzee SG, Gull N, Yao L, Hazelett DJ, Noushmehr H, et al. ELMER v.2: An R/Bioconductor package to reconstruct gene regulatory networks from DNA methylation and transcriptome profiles. *Bioinformatics* 2018;35:1974–77.
- Zhou W, Triche TJ Jr., Laird PW, Shen H. SeSAMe: reducing artifactual detection of DNA methylation by Infinium BeadChips in genomic deletions. *Nucleic Acids Res* 2018;46:e123. DOI: 10.1093/nar/gky691.
- Wingender E, Schoeps T, Haubrock M, Krull M, Donitz J. TFClass: expanding the classification of human transcription factors to their mammalian orthologs. *Nucleic Acids Res* 2018;46:D343–D7. DOI: 10.1093/nar/gkx987.
- Kulakovskiy IV, Vorontsov IE, Yevshin IS, Sharipov RN, Fedorova AD, Rumynskiy EI, et al. HOCOMO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. *Nucleic Acids Res* 2018;46:D252–9. DOI: 10.1093/nar/gkx1106.
- Corces MR, Granja JM, Shams S, Louie BH, Seoane JA, Zhou W, et al. The chromatin accessibility landscape of primary human cancers. *Science* 2018;362.
- Xie JJ, Jiang YY, Jiang Y, Li CQ, Lim MC, An O, et al. Super-enhancer-driven long non-coding RNA LINC01503, regulated by TP63, is over-expressed and oncogenic in squamous cell carcinoma. *Gastroenterology* 2018;154:2137–51.e1.

**Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases):** T.C. Silva, K. Daigo, Y.-Y. Jiang, L.-Y. Xu, E.-M. Li  
**Study supervision:** J. Pan, S. Hu, B.P. Berman, H.P. Koeffler, D.-C. Lin

### Acknowledgments

This work was supported by NIH grant (1R01 CA200992) to H.P. Koeffler and the National Research Foundation Singapore under its Singapore Translational Research Investigator Award (NMRC/STaR/0021/2014 to H.P. Koeffler) and administered by the Singapore Ministry of Health's National Medical Research Council (NMRC); the NMRC Centre Grant awarded to National University Cancer Institute; the National Research Foundation Singapore and the Singapore Ministry of Education under its Research Centres of Excellence initiatives to H.P. Koeffler. D.-C. Lin is supported by the DeGregorio Family Foundation, the Price Family Foundation, Samuel Oschin Comprehensive Cancer Institute SOCCI through the Translational Pipeline Discovery Fund. He is a member of UCLA Jonsson Comprehensive Cancer Center, UCLA Molecular Biology Institute as well as UCLA Cure: Digestive Disease Research Center. B.P. Berman and T.C. Silva were supported by the Genomic Data Analysis Network of the NCI via grant U24CA210969. N. Gull was supported by institutional funding from the Cedars-Sinai Center for Bioinformatics and Functional Genomics. This research was also partly supported by National Natural Science Foundation of China (NSFC; Nos. 81570125, 81770145), Jiangsu province's science and technology support program (Social Development) project (BE2017658), and the philanthropic donations from the Melamed family. S.J. Klempner was supported by the Howard H. Hall fund for esophageal cancer research.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received February 4, 2020; revised March 24, 2020; accepted April 21, 2020; published first April 24, 2020.

24. Thongjuea S, Stadhouders R, Grosveld FG, Soler E, Lenhard B. r3Cseq: an R/Bioconductor package for the discovery of long-range genomic interactions from chromosome conformation capture and next-generation sequencing data. *Nucleic Acids Res* 2013;41:e132.
25. Xie JJ, Jiang YY, Jiang Y, Li CQ, Chee LM, An O, et al. Increased expression of the long non-coding RNA LINC01503, regulated by TP63, in squamous cell carcinoma and effects on oncogenic activities of cancer cell lines. *Gastroenterology*. 2018;154.
26. Jiang YY, Lin DC, Mayakonda A, Hazawa M, Ding LW, Chien WW, et al. Targeting super-enhancer-associated oncogenes in oesophageal squamous cell carcinoma. *Gut* 2017;66:1358–68.
27. Ooi WF, Xing M, Xu C, Yao X, Ramlee MK, Lim MC, et al. Epigenomic profiling of primary gastric adenocarcinoma reveals super-enhancer heterogeneity. *Nat Commun* 2016;7:12983.
28. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;9:357–9.
29. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:13033997; 2013.
30. Ramirez F, Dunder F, Diehl S, Gruning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* 2014;42:W187–91.
31. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010;26:841–2.
32. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 2010;38:576–89.
33. Dang LH, Chen F, Ying C, Chun SY, Knock SA, Appelman HD, et al. CDX2 has tumorigenic potential in the human colon cancer cell lines LOVO and SW48. *Oncogene* 2006;25:2264–72.
34. Hinoi T, Gesina G, Akyol A, Kuick R, Hanash S, Giordano TJ, et al. CDX2-regulated expression of iron transport protein hephaestin in intestinal and colonic epithelium. *Gastroenterology* 2005;128:946–61.
35. Martinelli P, Carrillo-de Santa Pau E, Cox T, Sainz B Jr., Dusetti N, Greenhalf W, et al. GATA6 regulates EMT and tumour dissemination, and is a marker of response to adjuvant chemotherapy in pancreatic cancer. *Gut* 2017;66:1665–76.
36. Abel EV, Goto M, Magnuson B, Abraham S, Ramanathan N, Hotaling E, et al. HNF1A is a novel oncogene that regulates human pancreatic cancer stem cell properties. *eLife* 2018;7:e33947.
37. Odom DT, Zizlsperger N, Gordon DB, Bell GW, Rinaldi NJ, Murray HL, et al. Control of pancreas and liver gene expression by HNF transcription factors. *Science* 2004;303:1378–81.
38. Zhou W, Laird PW, Shen H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Res* 2017;45:e22.
39. Uhlen M, Zhang C, Lee S, Sjostedt E, Fagerberg L, Bidkhori G, et al. A pathology atlas of the human cancer transcriptome. *Science* 2017;357:eaan2507.
40. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. *Science* 2015;347:1260419.
41. Chang HR, Nam S, Kook MC, Kim KT, Liu X, Yao H, et al. HNF4alpha is a therapeutic target that links AMPK to WNT signalling in early-stage gastric cancer. *Gut* 2016;65:19–32.
42. Xu C, Ooi WF, Qamra A, Tan J, Chua BY, Ho SWT, et al. HNF4alpha pathway mapping identifies wild-type IDH1 as a targetable metabolic node in gastric cancer. *Gut* 2020;69:231–42.
43. Ghandi M, Huang FW, Jane-Valbuena J, Kryukov GV, Lo CC, McDonald ER 3rd, et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* 2019;569:503–8.
44. Kiselyuk A, Lee SH, Farber-Katz S, Zhang M, Athavankar S, Cohen T, et al. HNF4alpha antagonists discovered by a high-throughput screen for modulators of the human insulin promoter. *Chem Biol* 2012;19:806–18.
45. Ran L, Chen Y, Sher J, Wong EWP, Murphy D, Zhang JQ, et al. FOXF1 defines the core-regulatory circuitry in gastrointestinal stromal tumor. *Cancer Discov* 2018;8:234–51.
46. Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, et al. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* 2005;122:947–56.
47. Chen L, Huang M, Plummer J, Pan J, Jiang YY, Yang Q, et al. Master transcription factors form interconnected circuitry and orchestrate transcriptional networks in oesophageal adenocarcinoma. *Gut* 2020;69:630–40.
48. Long HK, Prescott SL, Wysocka J. Ever-changing landscapes: transcriptional enhancers in development and evolution. *Cell* 2016;167:1170–87.
49. Venmar KT, Fingleton B. Lessons from immunology: IL4R directly promotes mammary tumor metastasis. *Oncoimmunology* 2014;3:e955373.
50. Venmar KT, Carter KJ, Hwang DG, Dozier EA, Fingleton B. IL4 receptor ILR4alpha regulates metastatic colonization by mammary tumors through multiple signaling pathways. *Cancer Res* 2014;74:4329–40.
51. Kawakami K, Leland P, Puri RK. Structure, function, and targeting of interleukin 4 receptors on human head and neck cancer cells. *Cancer Res* 2000;60:2981–7.
52. Thaper D, Vahid S, Nip KM, Moskalev I, Shan X, Frees S, et al. Targeting lyn regulates snail family shuttling and inhibits metastasis. *Oncogene* 2017;36:3964–75.
53. Tornillo G, Knowlson C, Kendrick H, Cooke J, Mirza H, Aurrekoetxea-Rodriguez I, et al. Dual mechanisms of LYN kinase dysregulation drive aggressive behavior in breast cancer cells. *Cell Rep* 2018;25:3674–92 e10.
54. Li ZW, Sun B, Gong T, Guo S, Zhang J, Wang J, et al. GNAI1 and GNAI3 reduce colitis-associated tumorigenesis in mice by blocking IL6 signaling and down-regulating expression of GNAI2. *Gastroenterology* 2019;156:2297–312.
55. Rao VN, Reddy ES. elk-1 domains responsible for autonomous DNA binding, SRE:SRF interaction and negative regulation of DNA binding. *Oncogene* 1992;7:2335–40.
56. Neph S, Stergachis AB, Reynolds A, Sandstrom R, Borenstein E, Stamatoyannopoulos JA. Circuitry and dynamics of human transcription factor regulatory networks. *Cell* 2012;150:1274–86.
57. DeLaForest A, Di Furio F, Jing R, Ludwig-Kubinski A, Twaroski K, Urlick A, et al. HNF4A regulates the formation of hepatic progenitor cells from human iPSC-derived endoderm by facilitating efficient recruitment of RNA pol II. *Genes (Basel)* 2018;10:21.
58. DeLaForest A, Nagaoka M, Si-Tayeb K, Noto FK, Konopka G, Battle MA, et al. HNF4A is essential for specification of hepatic progenitors from human pluripotent stem cells. *Development* 2011;138:4143–53.
59. Yahoo N, Pournasr B, Rostamzadeh J, Fathi F. Forced expression of Hnf4a induces hepatic gene activation through directed differentiation. *Biochem Biophys Res Commun* 2016;476:313–8.
60. Marable SS, Chung E, Adam M, Potter SS, Park JS. Hnf4a deletion in the mouse kidney phenocopies Fanconi renotubular syndrome. *JCI insight* 2018;3:e97497.
61. San Roman AK, Aronson BE, Krasinski SD, Shivdasani RA, Verzi MP. Transcription factors GATA4 and HNF4A control distinct aspects of intestinal homeostasis in conjunction with transcription factor CDX2. *J Biol Chem* 2015;290:1850–60.
62. Chen L, Toke NH, Luo S, Vasoya RP, Fullem RL, Parthasarathy A, et al. A reinforcing HNF4-SMAD4 feed-forward module stabilizes enterocyte identity. *Nat Genet* 2019;51:777–85.
63. Fitamant J, Kottakis F, Benhamouche S, Tian HS, Chuvin N, Parachoniak CA, et al. YAP inhibition restores hepatocyte differentiation in advanced HCC, leading to tumor regression. *Cell Rep* 2015;10:1692–707.
64. Rogerson C, Britton E, Withey S, Hanley N, Ang YS, Sharrocks AD. Identification of a primitive intestinal transcription factor network shared between esophageal adenocarcinoma and its precancerous precursor state. *Genome Res* 2019;29:723–36.