Genomic landscapes reveal post-transcriptional modifier disruption in cholangiocarcinoma

Yaodong Zhang, Zijian Ma, Changxian Li, Cheng Wang, Wangjie Jiang, Jiang Chang, Sheng Han, Zefa Lu, Zicheng Shao, Yirui Wang, Hongwei Wang, Chenyu Jiao, Dong Wang, Xiaofeng Wu, Hongbing Shen, Xuehao Wang, Zhibin Hu, Xiangcheng Li

.

Table of contents

Supplementary materials and methods

Clinical Sample Collection

All primary cholangiocarcinoma and matched adjacent normal samples were obtained from the resected specimens of patients with pCCA or iCCA between 2010 and 2017 in The First Affiliated Hospital of Nanjing Medical University (NMU). All samples were immediately frozen in liquid nitrogen and stored at -80℃ until DNA extraction. The use of clinical samples was approved by the Ethics Committee of The First Affiliated Hospital of Nanjing Medical University. Written informed patient consent was obtained in accordance with regional regulation. The data of their clinicopathological features were anonymized and were shown in **Supplementary Table 1**. All tumor samples were confirmed by pathologists that there was a minimum tumor cellularity of 70% in all CCA specimens following histopathological review of H&E slides.

Next-generation sequencing

DNA extracted from frozen tumor specimens and adjacent normal tissues using QIAmp DNA mini kit (Qiagen) and quality were determined using Picogreen (Invitrogen) and further visually inspected by agarose gel electrophoresis. Library construction and whole-exome capture of genomic DNA were performed using the Roche NimbleGen SeqCap EZ Exome SR platform V3. The captured DNA was sequenced on an Illumina HiSeq X10 sequencing system, with 150-bp paired-end sequencing. The average sequence coverage was 91.9× and 97.6× for the pCCA and iCCA tumor samples and 89.7× and 94.6× for pCCA and iCCA adjacent normal tissues for whole-exome sequencing.

Sequencing alignment and detection of somatic variants

The quality score distribution of the reads was acquired with the FastQC package (http://www.bioinformatics.babraham.ac.uk/projects/fastqc).Burrows-Wheeler Aligner (BWA-MEM v0.7.15-r1140) was used to map the read sequences to the human reference genome (GRCh37) with the default parameters[1], and duplicates were marked and discarded using Picard (v1.70) (http://broadinstitute.github.io/picard). Then, the reads were subjected to local realignment and recalibration using the Genome Analysis Toolkit (GATK)(v4). For the 102 patients from Zou et al's study[2], we downloaded the raw FASTQ files from Short Read Archive (SRA) under the accession code SRP045202 and applied the same alignment pipeline mentioned above.

Somatic substitutions and indels were detected using the MuTect2 mode in GATK (v4) on the GRCh37 genome build following the best practices for somatic SNV/indel calling (https://software.broadinstitute.org/gatk/best-practices/). Briefly, the algorithms compared the tumor with the matched normal sample to exclude germline variants. Somatic mutations were excluded (1) if they were found in a panel of normal controls assembled from matched normal tissues, (2) if they were located in the segmental duplication region marked by the UCSC browser (http://genome.ucsc.edu/), or (3) if they were found in the 1000 Genomes Project (the Phase III integrated variant set release, across 2,504 samples) with the same mutation direction. Mutations (SNVs/indels) were annotated with the local versions of Oncotator (1.8.0)[3] according to GENCODE v19. In addition, the annotated somatic mutation R298H of *METTL14* were further confirmed by Sanger sequencing, the forward primers of *METTL14* used in this study are 5'- CGAGGTAGGTAGACCACTTG-3', and reverse primer is 5'-TTCCAAATAGATGAAGGCGT-3'.

For integration analysis, we applied the somatic mutation data of 239 patients with cancer of biliary duct, including 44 pCCAs and 135 iCCAs, from BTCA-JP, International Cancer Genome Consortium (ICGC) project (https://icgc.org/icgc/cgp/91/420/1012366).

Mutational burden and signature analysis

The mutation counts were provided in **Supplementary Table 2**. The mutation burden was defined as the number of somatic mutations per mega base (/MB) in the callable regions, which was considered as the region of mapped genome with depth of at least 10x.

We converted all substitutions into a matrix (M) composed of 96 features comprising mutation counts for each mutation type (C > A, C > G, C > T, T > A, T > C and T > G) using each possible 5′ and 3′ context for all samples. R packages SignatureEstimation[4] was taken to estimate the proportion of 30 signatures from COSMIC (https://cancer.sanger.ac.uk/cosmic/signatures_v2). The signatures, percentage of which were more than one in pCCA or iCCA patients, were included in further analysis. (**Supplementary Table 4**).

Identification of CCA Driver Genes and Comparison Between the iCCA and pCCA

The IntOGen platform[5] and MutSig2CV (v3.11)[6] was used to identify SMGs among the somatic mutation data of all the patients from three databases. The IntOGen pipeline included two algorithms (OncodriveCLUST[7] and OncodriveFM[8]) that were designed

to find genes with highly clustered mutations and non-randomly distributed functional mutations, respectively. MutSig2CV was used to find genes with a higher mutation rate than the calculated background mutation rate. Multiple testing correction (Benjamini–Hochberg FDR) was performed separately and genes with q values ≤ 0.1 in any algorithm were reported as potential driver genes also mentioned as mut-drivers in this study (**Supplementary Table 5**). Then we divided the patients into two subgroups according to the CCA subtype, and comparison was conducted between mut-driver genes. Subtype specific mut-drivers was defined as genes with log2 scaled mutation rate ratio of which were > 1 between iCCA and pCCA. The specific Mutations were considered recurrent when occurred in (1) two or more NMU patients or (2) at least one NMU patient and other patient(s) from ICGC and/or Zou et al's dataset.

Somatic Copy Number Estimation

ICGC samples were not included in this copy number analysis due to the inaccessibility of the origin FASTQ and/or BAM files of the peripheral blood from those subjects. The GATK best practices for somatic copy number alterations (CNAs) in exomes were used to detect CNAs from the whole-exome sequencing data (https://software.broadinstitute.org/gatk/best-practices/). The somatic copy number was estimated by ReCapSeg, which is implemented as part of GATK (v4). Briefly, the read counts for each of the exome targets were divided by the total number of reads to generate proportional coverage. A panel of normal (PON) controls was built using proportional coverage from normal samples. Each of the tumor samples was compared with the PON, after which tangent normalization was applied. Circular binary segmentation (CBS) was then applied to segment the normalized coverage profiles. Sex chromosomes (X and Y) were excluded from this analysis. CNAB was considered as

the proportion of genome altered (amplified and/or deleted) divided by the length of the genome that all segments covered and the copy number segments with absolute value of log2 scaled copy number ratio more than 0.2 was defined as altered segments (>0.2 for amplification and <-0.2 as deletion)[9].

Highly Amplified/Deleted Regions Identification

Copy number segments of patients (67 from NMU and 102 from Zou et al.'s studies) were used as input for GISTIC2[10] to identify significantly amplified/deleted regions with the default parameters. A default q value threshold ($q < 0.25$) was used to define identified frequently amplified/deleted regions. Cancer related genes, curated by COSMIC[11], located at frequently CNA focal regions were defined as potential CNA-driver genes.

Pathway Enrichment Analysis

Pathway enrichment analysis was performed on TCGA ten pan-cancer pathways listed in Sanchez-Vega et al' study[12] using a Fisher's exact test based on the hypergeometric distribution[13]. Briefly, it determines whether the fraction of genes of interest in the pathway is higher compared to the fraction of genes in the background.

Annotation of Genomic Alterations upon Clinical Actionability

The annotation was performed with OncoKB[14]. The actionable level was classified as level 1-4. Alterations with level 1 and 2 are Food and Drug Administration (FDA)-recognized or considered standard care biomarkers predictive of response to FDA-approved drugs in specific disease settings. Level 3 alterations are believed as predictive of response based on promising clinical data to targeted agents being tested

in clinical trials whereas level 4 was the prediction of response on compelling biological evidence to targeted agents being tested in clinical trials. Clinically actionable sample frequency was calculated as the number of patients with at least one actionable alteration divided by all individuals counts in each subgroup.

Cell Culture and siRNA Transfection

CCA cell lines RBE and HCCC9810 were selected to establish the stable *METTL14*^{R298H} and *METTL14*^{wt} cells. Cells were maintained in DMEM medium (Gibco) with 10 % fetal bovine serum (Biological Industries), penicillin/streptomycin 100 units/mL at 37℃ with 5% CO^2. All the siRNAs were ordered from GenePharma. Sequences for siRNA of MACF1 are: MACF1-homo-2920, 5'-CCUUAUCUCUUGGAACUAUTT -3'; MACF1-homo-8572, 5'-GCAGAAAGCUCAGAAAUAUTT -3'. Transfection was achieved by using Lipofectamine 3000 (Invitrogen) for the siRNA following manufacturer's protocols.

Name	Supplier	Cat no.	Authentication test method
RBE	Cell Bank of the Chinese Academy of Science	TCHu179	Cell Bank of the Chinese Academy of Science
HCCC9810	Cell Bank of the Chinese Academy of Science	TCHu 17	Cell Bank of the Chinese Academy of Science

Establishment of Stable *METTL14*^{R298H} and *METTL14*^{wt} Cells and Functional Assays

To produce recombinant lentiviruses, vectors encoding empty control, *METTL14*^{R298H} and *METTL14*^{wt} gene were designed from GeneChem (Shanghai, China)[15, 16]. The lentivirus transfection assay was performed in 2×10^5 cells using 5μl/ml polybrene according to the manufacturer's protocol. Puromycin (10 μg/ml) was added for 7 days to select stable cells. Migration assays were performed in Transwells (Corning Inc., 8.0-μm pore size) and invasion assay was conducted using BD BioCoat Matrigel invasion

chambers according to the manufacturer's instructions. Briefly, 2×10^4 cells in 300 μl serum-free medium were loaded into the upper chambers. Then, 500 μl medium supplemented with 20% fetal bovine serum was loaded into the lower chamber. Cells on the underside of the membrane were stained with crystal violet and counted under a microscope in three random fields after 48 hours. For wound healing assay, cells were seeded in 6-well plates at 80–90% confluency in serum-free medium. A linear wound was performed in constant-diameter stripes using a sterile 200-μl pipette tip. After incubation for 0, 24, 48 hours, photographs were taken and wound closures were evaluated and calculated to estimate the capacity of migration.

The cell proliferation was detected by EdU (5-ethynyl-2 0 -deoxy uridine) assay using Cell-Light EdU DNA Cell Proliferation Kit (RiboBio, Shanghai, China). were seeded in each well of 6-well plates. Nucleic acids in all cells were stained with DAPI Dye. The Edu pulse chase incorporation and cell proliferation rate were applied according to the manufacturer's instructions. For colony forming assay, the transfected cells were re-plated and incubated for an additional 14 days at 37°C to allow colony formation. Colonies immobilization were maintained with -20°C methyl alcohol for 30 minutes and were stained with 0.5% crystal violet and counted. Cell proliferation rates were subsequently assessed using the cell counting kit-8 (CCK-8) (Dojindo, Tokyo, Japan) according to the manufacturer's instructions.

RNA N6-Methyladenosine (m^6A) Quantification

Total RNA was isolated using TRIzol (Invitrogen) according to the manufacturer's instructions and RNA quality was analyzed by NanoDrop. The EpiQuik™ m^6A RNA Methylation Quantification Kit (Colorimetric) was used to assess the m^6A content in total RNA. According to the manufacturer's instructions, 200 ng RNAs were an optimal

amount on assay wells. M^6A standard control was added into the assay wells at different concentrations to determine the standard curve and then capture antibody, detection antibody, enhancer solution, and developer solution were added to assay wells respectively in a suitable diluted concentration. The m^6A levels were quantified colorimetrically by the absorbance of each well at a wavelength of 450 nm, and then calculations were performed based on the standard curve.

RNA m^6A immunoprecipitation Assay and m^6A Sequencing

For m^6A sequencing, total RNA was extracted using Trizol reagent (Invitrogen) following the manufacturer's procedure. The total RNA quality and quantity were analysis of Bioanalyzer 2100 and RNA 6000 Nano LabChip Kit (Agilent) with RIN number >7.0. Approximately more than 200 ug of total RNA was subjected to isolate Poly (A) mRNA with poly-T oligo attached magnetic beads (Invitrogen). Following purification, the poly(A) mRNA fractions are fragmented into ~100-nt-long oligonucleotides using divalent cations under elevated temperature. Then the cleaved RNA fragments were subjected to incubated for 2h at 4℃ with m^6A-specific antibody (No. 202003, Synaptic Systems) in IP buffer (50 mM Tris-HCl, 750 mM NaCl and 0.5% Igepal CA-630) supplemented with BSA. The mixture was then incubated with protein-A beads and eluted with elution buffer (1×IP buffer and 6.7mM m^6A). Eluted RNA was precipitated by 75% ethanol. Eluted m^6A-containing fragments (IP) and untreated input control fragments are converted to final cDNA library in accordance with a strand-specific library preparation by dUTP method. The average insert size for the paired-end libraries was ~100±50 bp. And then we performed the paired-end 2×150bp sequencing on an Illumina Novaseq™ 6000 platforms at the LC-BIO Bio-tech Ltd (Hangzhou, China) following the vendor's recommended protocol. Enrichment of m^6A containing

mRNA was analyzed by quantitative reverse-transcription polymerase chain reaction (qRT-PCR) using m^{6}A RNA immunoprecipitation (MeRIP) method. In this assay, total RNA was chemically fragmented into 100 nucleotides or smaller RNA fragments, and m^{6}A monoclonal antibody was used for magnetic immunoprecipitation. The above methods follow the standard protocol of the Magna MeRIP m^{6}A Assay Kit (Merck Millipore).

m^{6}A Sequencing Analysis

Reads were aligned to the reference genome (GRCh37) with HISAT2[17]. The m^{6}A modification peaks were identified using R package exomePeak2[18] and a cutoff for controlling FDR < 0.05 was used to obtain high-confident m^{6}A modification peaks. Then, we retrieved the raw read counts from both input and IP RNA-Seq bam files by Genomic Alignments and performed differential methylation analysis by DESeq2 and fitted the model as below:

Normalized counts ~ Design (e.g., *METTL14*^{wt} vs control) + Experiments (IP/control) + Design × Experiments.

Candidate differential m^{6}A peaks was selected with (1) Multiple testing correction *P* value <= 0.05 and (2) log2 Fold Change > 1 between *METTL14*^{wt} and control cells or log2 Fold Change < 0 between *METTL14*^{R298H} and *METTL14*^{wt} cells.

Measurement of RNA lifetime

RBE cells were seeded in 6-well plates at 80% confluency. After 24 h, actinomycin D (MedChemExpress) was added to 1mg/ml at 6 h, 3 h, and 0 h before trypsinization and collection. The total RNA was purified according to the procedure mentioned in this study. The degradation rate of RNA (k) was estimated by plotting Nt/N0 against

time follow the standard equation mentioned by Liu et al[19].

Western Blotting and Immunohistochemistry Staining

For protein expression analysis, cells and tissues were lysed using Radioimmunoprecipitation (RIPA) lysis buffer (KeyGen Biotech) containing protease and phosphatase inhibitors (Roche) to obtain the total protein, and protein concentration were estimated using the BCA protein estimation assay (Thermo Scientific). Nuclear extracts were obtained using the NE-PER Nuclear and Cytoplasmic Extraction Reagents (Pierce Biotechnology, Thermo). Equal amounts of total protein were separated by SDS-PAGE and subsequently transferred to nitrocellulose membrane. Antibodies used were: anti-METTL14 (Novus), anti-GAPDH (Abcam), anti-Lamin B1 (Proteintech), anti-β-catenin (Abcam), anti-E-cadherin (Cell Signaling Technology), anti-N-cadherin (Cell Signaling Technology), anti-α-Tubulin (Cell Signaling Technology), anti-PCNA (Abcam), anti-Cyclin D1 (Cell Signaling Technology).

Immunohistochemical staining for METTL14 was performed on the cholangiocarcinoma (including iCCA and pCCA) tissue array using anti-METTL14 (Novus). After 4% paraformaldehyde fixation, the sections were deparaffinized. Antigens of the slides were retrieved by heating for 30 min in citrate buffer, pH 6.0. The slides were labeled with primary antibody in a blocking solution at 4°C overnight, following by counterstained with hematoxylin. Light microscopy (Nikon) was used to acquire the images, and NIS-Elements v4.0 software was used to quantify the staining (Nikon). All antibodies used in this study were listed in the table.

Name	Supplier	Cat no.	Clone no.
METTL14	Novus	NBP1-81392	Polyclonal
GAPDH	Abcam	ab181602	monoclonal EPR16891
Lamin B1	Proteintech	12987-1-AP	Polyclonal
β-catenin	Abcam	ab32572	monoclonal [E247]
E-cadherin	Cell Signaling Technology	#3195	24E10

N-cadherin	Cell Signaling Technology	#13116	D4R1H
α-Tubulin	Cell Signaling Technology	#2125	11H10
PCNA	Abcam	ab92552	EPR3821
Cyclin D1	Cell Signaling Technology	#2922	Polyclonal
MACF1	Abcam	ab221989	Polyclonal

RNA Isolation and Quantitative Real-time PCR

Total RNAs were isolated from specimens or cell lines using TRIzol (Invitrogen) following the manufacturer's protocol. RNA quality was analyzed by NanoDrop. The high capacity cDNA transcription kit (Vazyme) qRT-PCR was performed by the Thermal Cycler Dice Detection System using SYBR Green PCR Mix (Vazyme). The forward primers of *METTL14* used in this study are 5'- AGTGCCGACAGCATT GGTG', and reverse primer is 5'- GGAGCAGAGGTATCATAGGAAGC-3'. The forward primers of *MACF1* is 5'- CGGAGTGAGCGATCTACAGG -3', and reverse primer is 5'-TCATCAGCGACTCTGACCACA -3'. All data were normalized to the housekeeping gene GAPDH.

Evaluation of immunostaining

Hematoxylin and eosin (H&E) staining was performed routinely. Tissue microarrays (TMA) containing 114 cases of CCA specimens and corresponding adjacent normal tissues, which were obtained from The First Affiliate Hospital of Nanjing Medical University. Scoring was conducted based on the percentage of positive-staining cells: 0-5% scored 0, 6-35% scored 1, 36-70% scored 2, and more than 70% scored 3; and staining intensity: no staining scored 0, weakly staining scored 1, moderately staining scored 2 and strongly staining scored 3. The final score was calculated using the percentage score × staining intensity score as follows: "-" for a score of 0-1, "+" for a score of 2-3, "++" for a score of 4-6 and "+++" for a score of >6. Low expression was defined as CCA specimens' score < corresponding adjacent normal tissues' score, and

high expression was defined as CCA specimens' score ≥ corresponding adjacent normal tissues' score. These scores were determined independently by two senior pathologists in a blinded manner. Specifically, positive staining in the hepatic tissue was excluded for scoring.

Immunofluorescence Analysis

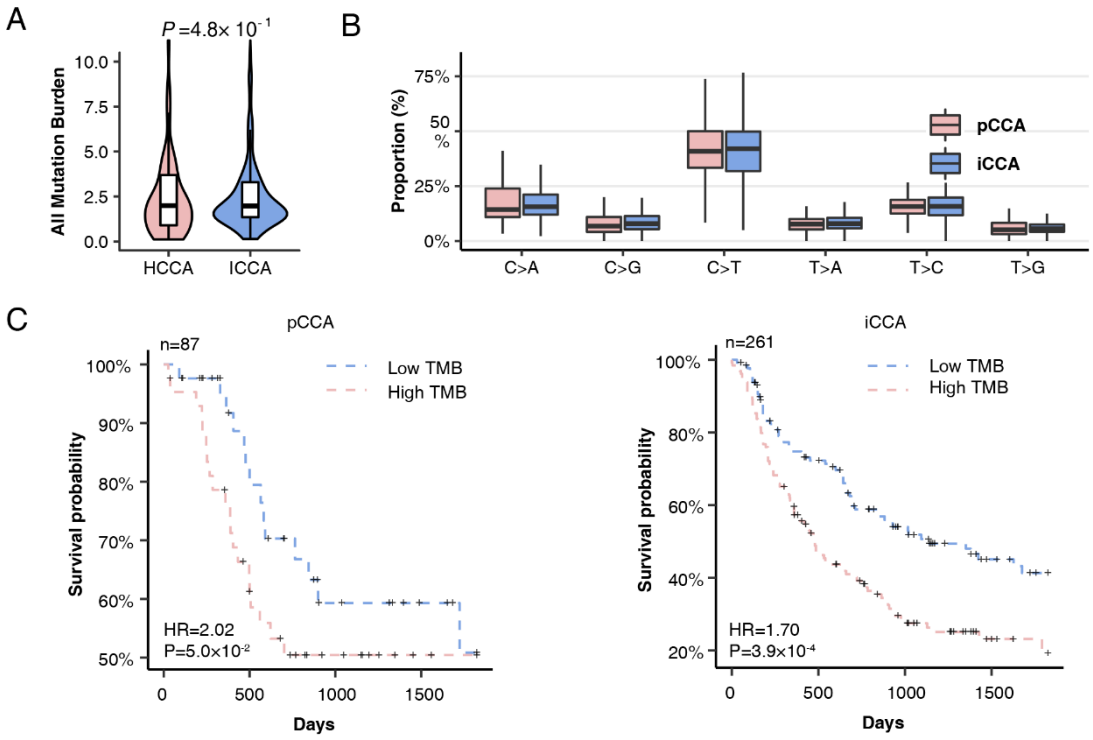
Cells were washed with phosphate-buffered saline, fixed in 4% paraformaldehyde for 10 min and permeabilized with 0.25% Triton X-100 in phosphate-buffered saline for 5 min, followed by 1 h incubation with primary antibodies, *METTL14* (Novus), *MACF1* (Abcam) and then incubation with IgG (Santa Cruz). The coverslips were counterstained with 4,6-diamidino-2-phenylindole (Invitrogen) and imaged with a confocal laser-scanning microscope (Olympus FV1000, Olympus)

Statistical Analysis and Figures

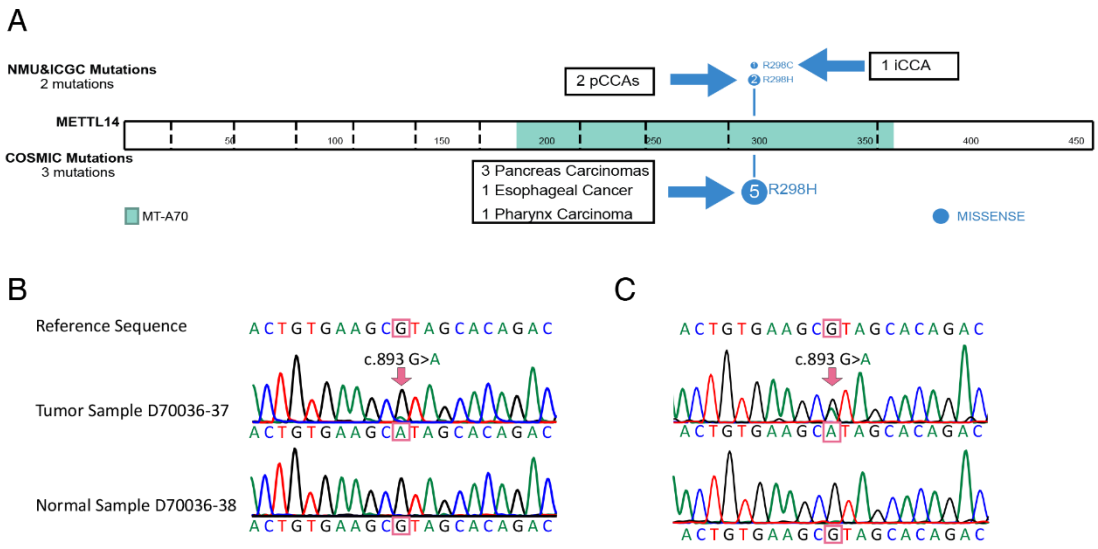
The median was used if multiple samples from the same tissues were sequenced. All statistical tests were performed using a Wilcoxon rank-sum test for continuous data. Fisher's exact test was used to assess differences in the count data. Multiple testing corrections were performed where necessary using the Benjamini-Hochberg method. All reported *P* values were two-sided. Mutational lolliplots were generated by ProteinPaint[20]. Other figures were generated using the, R with package ggplot2 [Wickham H (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4,] and package RColorBrewer [https://cran.r-project.org/web/packages/RColorBrewer/index.html].

Reference

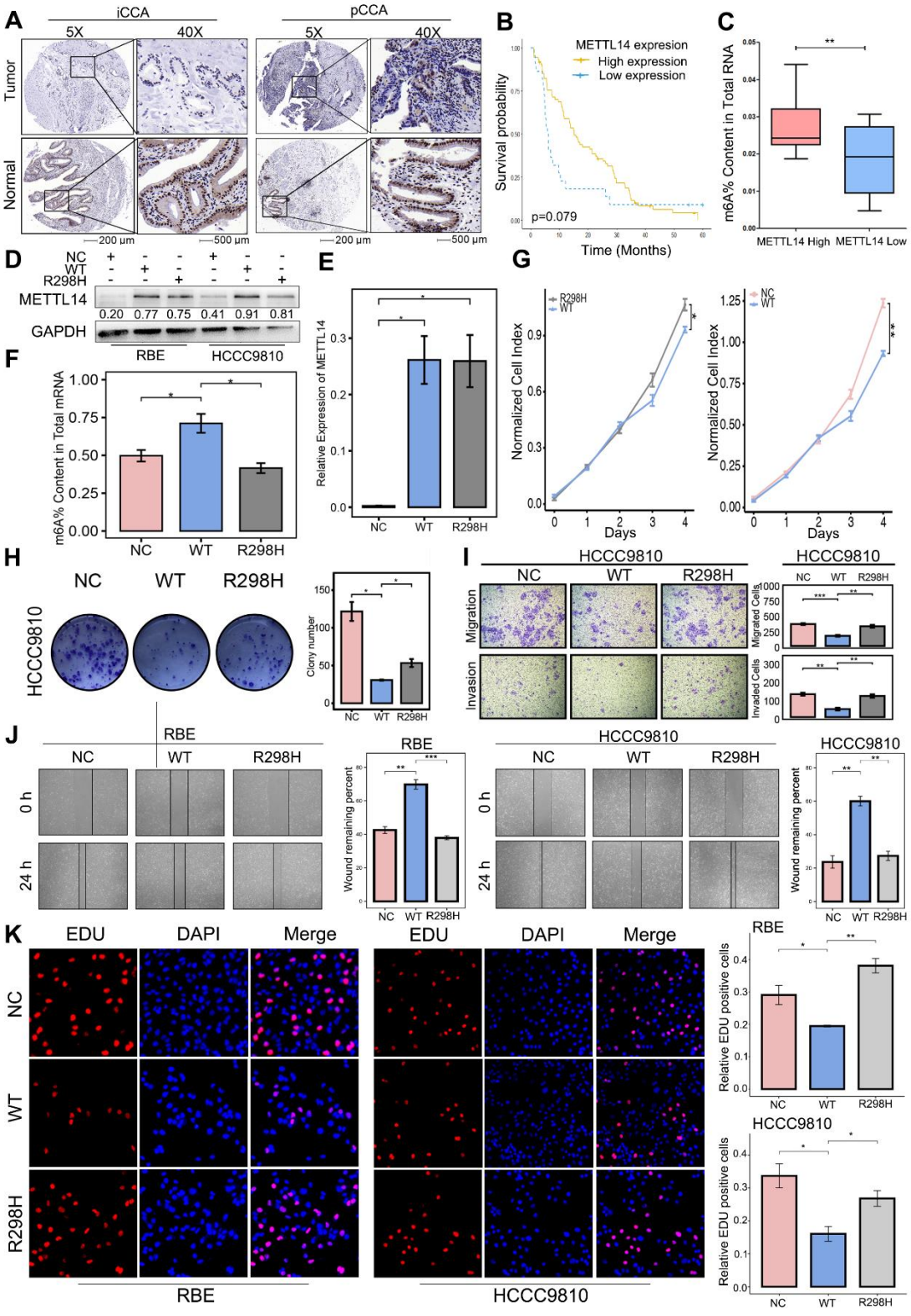
1. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
2. Zou, S. *et al.* Mutational landscape of intrahepatic cholangiocarcinoma. *Nature communications* **5**, 5696 (2014).
3. Ramos, A.H. *et al.* Oncotator: cancer variant annotation tool. *Human mutation* **36**, E2423-2429 (2015).
4. Huang, X., Wojtowicz, D. & Przytycka, T.M. Detecting presence of mutational signatures in cancer with confidence. *Bioinformatics* **34**, 330-337 (2018).
5. Gonzalez-Perez, A. *et al.* IntOGen-mutations identifies cancer drivers across tumor types. *Nature methods* **10**, 1081-1082 (2013).
6. Lawrence, M.S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214-218 (2013).
7. Tamborero, D., Gonzalez-Perez, A. & Lopez-Bigas, N. OncodriveCLUST: exploiting the positional clustering of somatic mutations to identify cancer genes. *Bioinformatics* **29**, 2238-2244 (2013).
8. Gonzalez-Perez, A. & Lopez-Bigas, N. Functional impact bias reveals cancer drivers. *Nucleic acids research* **40**, e169 (2012).
9. Xi, R. *et al.* Copy number variation detection in whole-genome sequencing data using the Bayesian information criterion. *Proceedings of the National Academy of Sciences of the United States of America* **108**, E1128-1136 (2011).
10. Mermel, C.H. *et al.* GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome biology* **12**, R41 (2011).
11. Sondka, Z. *et al.* The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nature reviews. Cancer* **18**, 696-705 (2018).
12. Sanchez-Vega, F. *et al.* Oncogenic Signaling Pathways in The Cancer Genome Atlas. *Cell* **173**, 321-337 e310 (2018).
13. Tavazoie, S., Hughes, J.D., Campbell, M.J., Cho, R.J. & Church, G.M. Systematic determination of genetic network architecture. *Nature genetics* **22**, 281-285 (1999).
14. Chakravarty, D. *et al.* OncoKB: A Precision Oncology Knowledge Base. *JCO precision oncology* **2017** (2017).
15. Rashid, M. *et al.* ALPK1 hotspot mutation as a driver of human spiradenoma and spiradenocarcinoma. *Nature communications* **10**, 2213 (2019).
16. Luchtel, R.A. *et al.* Recurrent MSC (E116K) mutations in ALK-negative anaplastic large cell lymphoma. *Blood* **133**, 2776-2789 (2019).
17. Kim, D., Paggi, J.M., Park, C., Bennett, C. & Salzberg, S.L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature biotechnology* **37**, 907-915 (2019).
18. Meng, J. *et al.* A protocol for RNA methylation differential analysis with MeRIP-Seq data and exomePeak R/Bioconductor package. *Methods* **69**, 274-281 (2014).
19. Liu, J. *et al.* m(6)A mRNA methylation regulates AKT activity to promote the proliferation and tumorigenicity of endometrial cancer. *Nature cell biology* **20**, 1074-1083 (2018).
20. Zhou, X. *et al.* Exploring genomic alteration in pediatric cancer using ProteinPaint. *Nature genetics* **48**, 4-6 (2016).



Supplementary.Fig.1 **(A)** Boxplot of all mutation burden in pCCA and iCCA. Wilcoxon rank-sum test was performed to obtain the *P* value. **(B)** Constitution of six types of single nucleotide substitutions between iCCAs (presented in blue) and pCCAs (presented in light red). **(C)** Kaplan-Meier survival plot between iCCA (left panel) and pCCA (right panel) patients with high TMB and low TMB. Cox proportional hazards model adjusted for age, gender, and tumor stage was performed to obtain HR and *P* values.

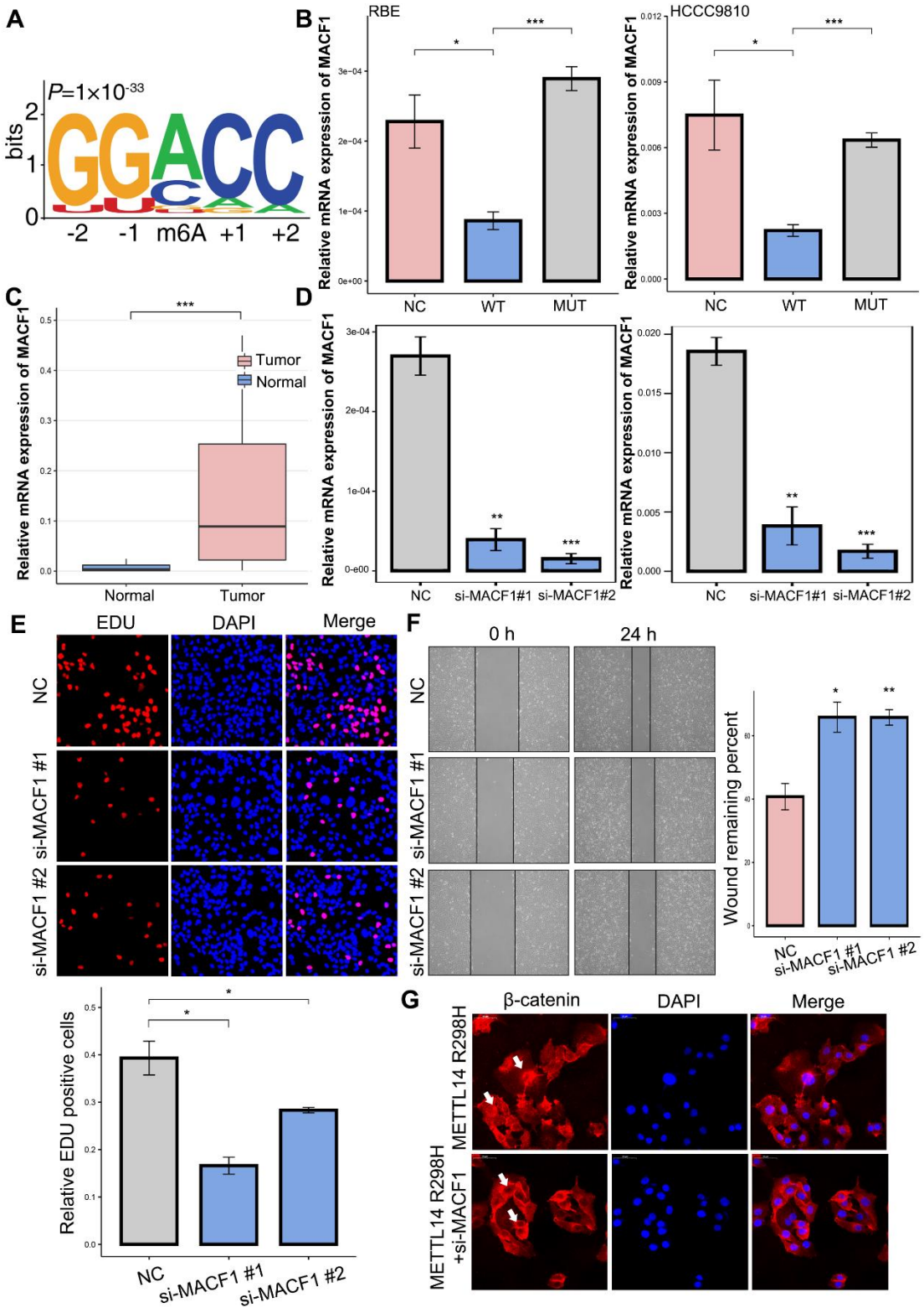


Supplementary.Fig.2 (A) Lolliplot of *METTL14* R298H and R298C. The top panel represents mutations in this study and the bottom panel represents mutations from COSMIC dataset. **(B & C)** Sanger sequencing plot of interested region including *METTL14* R298H of the NMU subject **(B)** and the additional subject **(C)**.



Supplementary.Fig.3 (A) Representative IHC stains of *METTL14* in iCCA and pCCA tissues and matched adjacent normal tissues. **(B)** *METTL14* downregulation in CCA tissues was associated with shorter cancer-specific survival in CCA patients. **(C)** The

m^6A contents of total mRNA in *METTL14* High group (n=15) and *METTL14* Low group (n=15). **(D)** Stable *METTL14*^{R298H} and *METTL14*^{wt} Cells were screened by western blot. **(E)** The transfection efficiency of lentiviral constructs expressing *METTL14*^{wt} and *METTL14*^{R298H} in HCCC9810 cell lines. **(F)** *METTL14*^{R298H} reduced *METTL14*^{wt}-mediated m^6A modification detected by m^6A colorimetric quantification in HCCC9810 cell lines. **(G)** Proliferation curve of HCCC9810 cells with *METTL14*^{R298H}, *METTL14*^{wt}, or negative control. **(H)** Colony formation assay of HCCC9810 cells with *METTL14*^{R298H}, *METTL14*^{wt}, or negative control. The number of colonies were counted and presented in the histogram. **(I)** Representative images (left) and quantification (right) of transwell migration and invasion assays in HCCC9810 cells with *METTL14*^{wt}, *METTL14*^{R298H}, or negative control. **(J)** Wound healing assay in HCCC9810 cells with *METTL14*^{R298H}, *METTL14*^{wt}, or negative control. **(K)** Representative images (left) and quantification (right) of EDU assays in RBE and HCCC9810 cells with *METTL14*^{wt}, *METTL14*^{R298H}, or negative control. The *P* values were calculated using Student's *t* test. **P* < 0.05, ***P* < 0.01, ****P* < 0.001; R298H, *METTL14*^{R298H}; WT, *METTL14*^{wt}; NC, negative control.



Supplementary.Fig.4 (A) GGAC is the most common m^{6}A motif significantly enriched in the m^{6}A peaks, and the m^{6}A peaks are especially enriched in the vicinity y of the stop codon. **(B)** *MACF1* mRNA expression in *METTL14*^{wt} and *METTL14*^{R298H} in RBE and HCCC9810 cell lines. **(C)** Upregulated *MACF1* mRNA expression was

detected in 66 pairs of CCA tumor tissues and the adjacent normal tissues by qRT-PCR. **(D)** Knockdown of *MACF1* in RBE and HCCC9810 cells by siRNA were verified by qRT-PCR. **(E)** Representative images (top) and quantification (bottom) of EDU assays in RBE cells with *METTL14*^{wt}, *METTL14*^{R298H}, or negative control. **(F)** Wound healing assay in RBE cells with *METTL14*^{R298H}, *METTL14*^{wt}, or negative control. **(G)** *MACF1* siRNA transfected in *METTL14*^{R298H}-overexpressing cells, and the representative images of nucleus translocation of β-catenin is shown using immunofluorescence. The *P* values were calculated using Student's *t* test. **P* < 0.05, ***P* < 0.01, ****P* < 0.001; R298H, *METTL14*^{R298H}; WT, *METTL14*^{wt}; NC, negative control.

Supplementary Table 1. General Description of Patients with CCA from Nanjing Medical University

		NMU	
	Total	iCCA	pCCA
Age			
Mean (SD)	61.0 (±9.3)	60.1 (±9.0)	61.5 (±9.6)
Gender			
Male	45 (67.2%)	16 (66.7%)	29 (67.4%)
Female	22 (32.8%)	8 (33.3%)	14 (32.6%)
HBV			
Yes	15 (22.4%)	8 (33.3%)	7 (16.3%)
No	51 (76.1%)	16 (66.7%)	35 (81.4%)
N/A	1 (1.5%)	0 (0.0%)	1 (2.3%)
Stage			
I	11 (16.4%)	7 (29.2%)	4 (9.3%)
II	20 (29.9%)	6 (25.0%)	14 (32.6%)
III	18 (26.9%)	1 (4.2%)	17 (39.5%)
IV	18 (26.9%)	10 (41.7%)	8 (18.6%)

NMU: Nanjing Medical University; CCA: cholangiocarcinoma; SD: Standard Deviation; HBV: Hepatitis B virus.

Supplementary Table 2. Summary of exome sequencing results

		NMU	
	Total	iCCA	pCCA
Tumour/normal pair sequenced			
Tumor depth Mean (SD)	94.4 (±11.1)	96.9 (±10.4)	93.0 (±11.3)
Normal depth Mean (SD)	91.8 (±11.5)	94.1 (±11.1)	90.5 (±11.6)
		Pool	
Tumor/normal pair sequenced (mean)			
SNV	84587/112.5	66508/113	18079/112
INDEL	7408/10.5	5329/5	2079/16

NMU: Nanjing Medical University; CCA: cholangiocarcinoma; SNV: Single Nucleotide Variant; INEDL: Insertion and deletion.

Supplementary Table 3. Proportion of COSMIC liver specific signatures

Signature	pCCA		iCCA	
	HBV	non-HBV	HBV	non-HBV
Signature 12	2.03%	0.24%	7.61%	0.10%
Signature 16	3.95%	0.00%	7.76%	3.72%
Signature 24	0.00%	0.00%	10.65%	0.00%
All Signature Liver	5.98%	0.24%	26.02%	3.82%

Supplementary Table 4. Proportion of COSMIC 30 mutation signatures.			
Signature	pCCA	iCCA	All
Signature 1	**44.71%**	**27.19%**	30.89%
Signature 6	**27.52%**	**21.93%**	23.11%
Signature 4	**3.73%**	**10.13%**	8.78%
Signature 22	**2.37%**	**9.96%**	8.35%
Signature 15	**4.47%**	**2.62%**	3.01%
Signature 9	**3.05%**	**3.58%**	3.47%
Signature 8	**6.41%**	0.00%	1.35%
Signature 13	**3.84%**	**1.88%**	2.30%
Signature 16	0.00%	**4.50%**	3.55%
Signature 3	0.00%	**4.26%**	3.36%
Signature 2	**2.64%**	**1.17%**	1.48%
Signature 12	0.63%	**2.04%**	1.74%
Signature 7	0.44%	**2.17%**	1.81%
Signature 10	0.00%	**2.44%**	1.92%
Signature 26	0.00%	**2.20%**	1.73%
Signature 24	0.00%	**2.16%**	1.71%
Signature 11	0.00%	0.82%	0.65%
Signature 29	0.00%	0.63%	0.50%
Signature 17	0.14%	0.20%	0.19%
Signature 21	0.00%	0.12%	0.10%
Signature 27	0.02%	0.00%	0.00%
Signature 28	0.02%	0.00%	0.00%
Signature 5	0.00%	0.00%	0.00%
Signature 25	0.00%	0.00%	0.00%
Signature 30	0.00%	0.00%	0.00%
Signature 20	0.00%	0.00%	0.00%
Signature 23	0.00%	0.00%	0.00%
Signature 18	0.00%	0.00%	0.00%
Signature 14	0.00%	0.00%	0.00%
Signature 19	0.00%	0.00%	0.00%

Supplementary Table 5. Curated significantly mutated genes.

Gene	Recurrent mutation	All	iCCA	pCCA	ratio	Subtype specific	class
TP53	1	28.45%	32.18%	17.24%	1.87		Report
KRAS	1	18.39%	19.16%	16.09%	1.19		Report
ARID1A	0	7.18%	8.43%	3.45%	2.44	iCCA	Report
SMAD4	1	6.32%	6.51%	5.75%	1.13		Report
PBRM1	1	5.75%	6.90%	2.30%	3.00	iCCA	Report
NF1	0	5.75%	4.98%	8.05%	0.62		Report
MACF1	0	5.17%	6.13%	2.30%	2.67	iCCA	New
GNAS	1	5.17%	4.21%	8.05%	0.52		Report
PIK3CA	1	4.89%	5.36%	3.45%	1.56		Report
EPHA2	0	4.60%	5.36%	2.30%	2.33	iCCA	New
BAP1	0	4.60%	4.98%	3.45%	1.44		Report
ARID2	1	4.02%	4.60%	2.30%	2.00	iCCA	Report
IDH1	1	3.74%	4.98%	0.00%	#DIV/0!	iCCA	Report
ATM	0	3.45%	3.45%	3.45%	1.00		New
PTEN	1	3.45%	4.21%	1.15%	3.67	iCCA	Report
RBM10	0	3.16%	1.92%	6.90%	0.28	pCCA	New
APC	0	3.16%	3.07%	3.45%	0.89		Report
STK11	0	3.16%	2.68%	4.60%	0.58		Report
RB1	0	2.87%	3.45%	1.15%	3.00	iCCA	Report
TGFBR2	0	2.87%	2.30%	4.60%	0.50	pCCA	Report
PIK3R1	0	2.59%	1.92%	4.60%	0.42	pCCA	New
BRAF	1	2.59%	3.07%	1.15%	2.67	iCCA	Report
ERBB2	1	2.59%	2.30%	3.45%	0.67		Report
NRAS	1	2.30%	2.68%	1.15%	2.33	iCCA	Report
MLLT4	0	2.01%	2.30%	1.15%	2.00	iCCA	New
BRCA2	0	2.01%	1.92%	2.30%	0.83		New
SLC8A1	1	2.01%	2.30%	1.15%	2.00	iCCA	Report
TGFBR1	1	2.01%	1.92%	2.30%	0.83		Report
NACC1	0	1.72%	1.15%	3.45%	0.33	pCCA	New
ELF3	0	1.72%	1.15%	3.45%	0.33	pCCA	Report
SMARCA4	0	1.44%	1.53%	1.15%	1.33		New
WHSC1	0	1.44%	1.53%	1.15%	1.33		New
CTNNB1	1	1.15%	1.15%	1.15%	1.00		Report
METTL14	1	0.86%	0.38%	2.30%	0.17	pCCA	New
AXIN1	0	0.86%	1.15%	0.00%	#DIV/0!		New
CDKN2A	0	0.86%	0.77%	1.15%	0.67		Report