# SupplementAry materials and Methods

### DNA methylation profiling

Total DNAs from frozen tissues were extracted using the DNeasy Blood & Tissue Kit (Qiagen, Valencia, CA) following the manufacturer’s instruction. Sodium bisulfite modification of DNA samples was performed by using the EZ-96 DNA Methylation Kit (Zymo Research Corporation, Orange, USA) and subsequently treated with the Infinium HD DNA Restoration Kit (Illumina). Each DNA sample was amplified and enzymatically fragmented following the Illumina infinium HD Methylation protocol. DNA methylation profiling was performed using Infinium Human Methylation 450 BeadChip. Hybridization and image acquisition were performed as manufacturer’s instruction.

Probe level β-values were imported using R package library of ‘RnBeads’. Genomic coordinates of the probes were updated to human reference genome hg38 by using R package library ‘*liftOver*’. Annotation for the probes in relation to CpG islands (CGI) or RefSeq genes was also updated to the genomic coordinates of hg38. Batch effects of raw data were corrected by an empirical Bayes method implemented in using R package library ‘combat’. The probes for sex chromosomes and the probes with missing values across more than three samples were also filtered out. Remaining missing values were imputed by R package library ‘*impute’*. Gene level methylation was calculated by averaging the methylation levels of the probes in the gene body and the upstream 1,500 bases from transcription start sites (TSS). (TSS200, 200bp upstream of the transcription start site; TSS1500, 1500bp upstream of the transcription start site; CpG island, 500-2000bp DNA segments with high CpG density; Shore, region up to 2kb from CpG island; Shelf, region up to 2-4kb from CpG island; Open sea, regions are isolated CpG sites)

DNA methylation status in relation to CGI regions was estimated. After filtering the probes with differential methylation among the groups (Fisher–Pitman permutation test, P <0.05, and median absolute deviation, MAD >0.2, n=78,236), the average beta values in the flanking regions of the CGI for each group were calculated using R package library ‘Enrichedheatmap’ with the window size of 100 bp. The flanking regions of CGIs were classified as Shore (within 2 kb of CGIs) and Shelf (within 2–4 kb of CGI) regions. The regions within 4-10kb of CGIs were designated as Opensea regions.

### DNA copy number profiling

DNA copy numbers were estimated from the DNA methylation profiles using ‘ChAMP’ with batch correction ([1](#_ENREF_1)). Segments of DNA copy number aberrations (CNA) were calculated by using a circular binary segmentation algorithm with default parameters. Gene level DNA copy numbers of each sample were mapped to their corresponding segment aberration values. After filtering out the probes with more than 50% of missing values across samples, the remained missing values were imputed. The copy number gains and losses were determined with cutoffs of 0.2 and -0.2, respectively.

### RNA-Seq profiling and analysis

Total RNA was isolated using TRIzol® RNA Isolation Reagents (Life Technologies, Carlsbad, CA), and the RNA integrity was confirmed by a bioanalyzer using an Agilent RNA 6000 Pico Kit (Agilent, Santa Clara, CA). The sequencing library for mRNA was constructed using TruSeq RNA sample preparation kit (Illumina, San Diego, CA) according to manufacturer’s instruction. Sequencing reaction was performed on an Illumina HiSeq2000 for 100 bp paired end reads (2 X 100) with coverage greater than 42 million reads per sample. The raw image data was transformed and stored in FASTQ format. The low quality sequence reads with less than 30 PHRED score and adapter sequence reads were trimmed using a trim galore tool (https://www.bioinformatics.babraham.ac.uk/projects/trim\_galore/), and then mapped to the human reference genome (hg38) and RNA abundance was estimated by using Tophat/cufflinks with default parameters (https://ccb.jhu.edu/software/tophat). Gene expression was log2 transformed.

### Variant calling

The RNA-Seq data was processed for variant calling. Local realignment of indel and normalization of base quality scores was performed using Genome Analysis Toolkit (GATK, https://software.broadinstitute.org/gatk). The sequence variations were filtered by using GATK HaplotypeCaller with default parameters and annotated by using ANNOVAR software (http://annovar.openbioinformatics.org). The somatic variants in exon regions were used in our analysis after filtering out non-significant variants with the following conditions. Allele frequencies of the variants were obtained from the normal populations including 1000 Genomes Project, NHLBI-ESP project with 6500 exomes, and Exome Aggregation Consortium (EXAC) data set. Non-significant variants were filtered out before analysis, which included the variants with mutated read counts less than 8, mutation frequencies greater than 50% or missing values greater than 50% across the samples. The mutations with allele frequency greater than 5% in either of the normal populations were also filtered out. Then, the final 14,999 somatic exon mutations were identified.

### Gene set enrichment and network analyses

For each sample, the score of enrichment for a gene signature was calculated by using DAVID software (http:// https://david.ncifcrf.gov) or preranked GSEA (Gene set enrichment analysis) method (https://genepattern.broadinstitute.org/gp). Gene network was constructed using physical, genetic, and pathway interactions using GeneMania software ([2](#_ENREF_2)).

### Analysis of public data

A DNA methylation profile of stepwise hepatocarcinogenesis (GSE44970, n=102) was used to validate our finding, which included tissue samples from healthy normal liver (HL, n=8), Liver Cirrhosis (LC, n=10), low-grade DN (LGDN, n=5), high-grade DN (HGDN, n=5), and progressed HCC (pHCC, n=71) ([3](#_ENREF_3)). The data was integrated with our data, correcting the batch effects by using ‘combat’.

DNA copy number and mutation profiles of HCC were obtained from The Cancer Genome Atlas (TCGA, https://cancergenome.nih.gov, LIHC, n=376). The segmented CNA values were mapped to gene level aberration. The CNA gain or loss in each gene was determined with the cutoff of ± 0.2 with more than 30% of aberrations across the samples, respectively. In addition, paired correlation between DNA copy number and transcriptome profiles were calculated in GSE65373 (n=38) to validate our finding.

Transcriptome data (n=293) for stepwise hepatocarcinogenesis was constructed by pooling our data YSHCC (n=131) with GSE89377 (n=107) and GSE6764 (n=75). The samples of the low-grade fibrosis of chronic hepatitis (LGCH, n=8) and high-grade of chronic hepatitis (HGCH, n = 12) in GSE89377 were excluded. Batch effects were corrected by ‘combat’ before data integration.

### Identification of DNA methylation-dependent transcriptional deregulation.

Pearson’s correlation coefficients (*r)* between DNA methylation profiles with the corresponding mRNA expression profiles were calculated and normalized to stabilize variance by applying Fisher’s Z-transformation: $z=\frac{1}{2}ln(\frac{1+r}{1-r})$. The inversely correlated genes between DNA methylation and transcription (METcor) were determined based on the Fisher’s Z-transformed correlation coefficients with 95% confidence interval (≥ 1.96 or ≤ -1.96, *P*<0.05).

### Cells, antibodies, and other reagents

Human liver cancer cells of Huh7, Hep3B, and HepG2 cells were purchased from Korean Cell Line Bank (KCLB) and cultured in DMEM supplemented with 10% FBS, 100 U/ml penicillin, and 100 µg/ml streptomycin. Anti-DNMT1, anti-p65, anti-IκBɑ, anti-phospho- IκBɑ antibody were purchased from Cell signaling Biotechnology (Danvers, MA, USA). Anti-ß-actin antibody was from Santa Cruz Biotechnology (Dallas, TX, USA). 5-aza-deoxycytidine, thapsigargin, and tunicamycin were from Sigma-Aldrich (St. Louis, MO, USA).

### Immunohistochemistry

Immunohistochemical stain of SPINK1 was performed with the representative sections of formalin-fixed, paraffin-embedded tissues by using the Ventana automated immunostainer BenchMark XT (Ventana Medical Systems, Tucson, AZ). SPINK1 antibody (mouse monoclonal, 4D4, ABNOVA, Taipei, Taiwan) was diluted to 1:50. Expression of SPINK1 protein was interpreted in a semiquantitative manner. Expression of each marker was evaluated as positive when it was detected in more than 5% of tumor epithelial cells with moderate to strong intensity, grading on a scale of 0-1-2-3 (0, <5% cells; 1, 5-10%; 2, 11-50%; 3, 51-100% of tumor epithelial cells).

### Quantitative real time PCR

Cells were harvested and total RNAs were isolated using an RNeasy kit (Qiagen, Venlo, Netherlands). The PrimeScript RT kit (Takara, Shiga, Japan) was used to reverse transcribed the mRNA into cDNA. PCR was done using a CFX96 Real Time PCR Detection System (Bio-Rad Laboratories, Hercules, CA, USA) with Ssoadvanced Universal Supermixes (Bio-Rad). Analysis of each sample was performed at least three times for each experiment, and the data in the figures was reported as relative quantitation: average values of 2-ΔΔCT±S.D. (standard deviation). The sequences of primers were as follows.

|  |  |  |
| --- | --- | --- |
| Gene | 5’ primer | 3’ primer |
| *SPINK1* | 5`- CCTTGGCCCTGTTGAGTCTA -3` | 5`- TAGAAGTCTGGCGTTTCCGA -3` |
| *GRP78* | 5`- GTGGTAGTGCAAGCTGAAGG -3` | 5`- TTCAGCCAGTTGCCCATCTA -3` |
| *TERT* | 5’-ACGAACGTGGCCAGCGGCAG-3’ | 5’- CTGGCGTCCCTGCACCCTGG-3’ |
| *β-actin* | 5`- TGGCACCCAGCACAATGAA -3` | 5`- CTAAGTCATAGTCCGCCTAGAAGCA -3` |

### Western blots

For preparing total cell lysates, cells were lysed in high salt lysis buffer [20mM Tris-HCl [pH 8.0], 1% Triton X-100, 2mM EDTA, and 1mM phenylmethylsulfonyl fluoride], incubated on ice for 20 min, and centrifuged for 20 min to remove cell debris. A total of 20μg of whole-cell lysate was used in SDS-polyacrylamide gel electrophoresis. The proteins were then electro-transferred to a nitrocellulose membrane and incubated overnight with antibodies at 4 °C. Then, the membranes were incubated with peroxidase-conjugated secondary antibodies (PIERCE, Rockford, IL, USA) for 1 hr at room temperature, and the signal was detected using an enhanced chemi-luminescence detection kit (PIERCE).

## Cell proliferation and invasion assay

Cells (2.5 Χ 103) were split into 96-well plates, and incubated in media for 72 hr. Cell proliferation was measured by WST-1 assay (Roche, Gangnam-gu, Seoul. Korea). Each experiment was performed in four replicates at least three times. Cell invasion assay was performed in 24-well modified Boyden chamber (8-μm pore size; Coastar; Corning Life Sciences, Lowell, MA). For invasion assay, the transwell filter inserts were coated with collagen type I (Sigma-Aldrich, St. Louis, MO, USA). The lower chamber was filled with 0.6 ml DMEM/10% fetal bovine serum. After 6 hrs, cells at the membrane undersurface were fixed, stained.

### Gene expression constructs and lentiviral vector transfection

The DNMT1 and SPINK1 expression vector was purchased from Addgene. Lentiviral constructs expressing Non-Target (NT) shRNA and p65 shRNA were purchased from Sigma-Aldrich (St. Louis, MO, USA). For lentiviral vector constructions, DNMT1, and SPINK1 construct was cloned into pCDH-CMV-MCS-EF1-Puro, a lentiviral vector for cDNA expression (System Biosciences, Mountain View, CA, USA). Lentiviral vectors were transfected into 293TN cells (System Biosciences) with Lipofectamin 3000 transfection reagent (Invitrogen, Waltham, MA, USA). Particles were collected 2 days after the transfection of the lentiviral plasmids and were used to infect cancer cells. Lentivirus-infected cancer cells were puromycin-selected for 1 week.

# References

1. Morris TJ, Butcher LM, Feber A, Teschendorff AE, Chakravarthy AR, Wojdacz TK*, et al.* ChAMP: 450k Chip Analysis Methylation Pipeline. Bioinformatics **2014**;30:428-30

2. Montojo J, Zuberi K, Rodriguez H, Kazi F, Wright G, Donaldson SL*, et al.* GeneMANIA Cytoscape plugin: fast gene function predictions on the desktop. Bioinformatics **2010**;26:2927-8

3. Revill K, Wang T, Lachenmayer A, Kojima K, Harrington A, Li J*, et al.* Genome-wide Methylation Analysis and Epigenetic Unmasking Identify Tumor Suppressor Genes in Hepatocellular Carcinoma. Gastroenterology **2013**