

博士論文

**The roles of HSPB7 induction by p53 and HSPA5
glycosylation by GLANT6 in human carcinogenesis**

**(GLANT6によるHSPA5のグリコシル化とp53誘導分子
HSPB7の発癌における機能解析)**

林嘉穎

Contents

| | |
|--|----|
| Abstract..... | 3 |
| Downregulation of a tumor suppressor HSPB7, involved in the p53-pathway, in renal cell carcinoma by hypermethylation | 5 |
| Introduction..... | 5 |
| Materials and methods | 6 |
| Results..... | 13 |
| Discussion..... | 16 |
| Figures and tables | 19 |
| O-glycosylation and stabilization of HSPA5 by GALNT6..... | 40 |
| Introduction..... | 40 |
| Materials and Methods..... | 44 |
| Results..... | 53 |
| Discussion..... | 56 |
| Figures and tables | 60 |
| References..... | 69 |
| Acknowledgements..... | 78 |

Abstract.

My thesis includes two parts. For part one, in order to identify genes involved in renal carcinogenesis, we analyzed the expression profile of renal cell carcinomas (RCCs) using microarray consisting of 27,648 cDNA or ESTs, and found a small heat shock protein, HSPB7, to be significantly and commonly downregulated in RCC. Subsequent quantitative PCR (qPCR) and immunohistochemical (IHC) analyses confirmed the downregulation of HSPB7 in RCC tissues and cancer cell lines in both transcriptional and protein levels. Bisulfite sequencing of a genomic region of HSPB7 detected DNA hypermethylation of some segments of HSPB7 in RCC cells and concordantly 5-aza-2'-deoxycytidine (5-Aza-dC) treatment of cancer cells restored HSPB7 expression significantly. Ectopic introduction of HSPB7 in five RCC cell lines remarkably suppressed cancer cell growth. Interestingly, we found that HSPB7 expression could be inducible by p53 in a dose-dependent manner, indicating that this gene functions in the p53 pathway. Our results imply that HSBP7 is likely to be a tumor suppressor gene regulated by p53 and its downregulation by hypermethylation may play a critical role in renal carcinogenesis.

For part two, we previously reported that overexpression of GALNT6 (polypeptide N-acetylgalactosaminyl transferase 6) which is a GalNAc-type (or mucin-type) O-glycosyltransferase, played a critical role in breast carcinogenesis. To further investigate the molecular function of GALNT6, we screened the substrates of GALNT6 through VVA (Vicia Villosa agglutinin) lectin (specific to GalNAc-Ser/Thr, called Tn-antigen) pull-down assay followed by mass spectrometry (MS) analysis. Here we report that HSPA5 (heat shock 70kDa protein 5 (also known as GRP78, glucose-regulated protein, 78kDa)) is a novel substrate of GALNT6. HSPA5 is highly expressed in cancers and involved in many cellular processes including in ER (endoplasmic reticulum) stress and autophagy, however whether O-

glycosylation affects the function of HSPA5 and promote carcinogenesis is largely unknown. We found high expression level of exogenous HSPA5 drives Golgi-to-ER relocation of GALNT6 and they co-localize at ER in HeLa GALNT6 stable cells. We confirmed GALNT6 directly binds to and glycosylates HSPA5, and the ATPase domain of HSPA5 is important for their binding. During the process, GALNT6 itself is also auto-glycosylated. Six candidate O-glycosylation sites in HSPA5 were identified by mass spectrometry and four of them were located in the ATPase domain. Further study showed that GALNT6 stabilizes HSPA5 protein and mutation at one potential O-glycosylation site, T184A, affects the stability of HSPA5. Taken together, our findings imply that GALNT6 O-glycosylates and stabilizes HSPA5 protein, which may prolong the oncogenic effects of HSPA5 in breast cancer. Meanwhile, overexpression of HSPA5 can drive Golgi-to-ER relocation of GALNT6. The result may trigger O-glycosylation of multiple substrates of GALNT6 at ER. Our study revealed a novel mechanism of how GALNT6 and HSPA5 cooperate together to promote mammary carcinogenesis. And the O-glycosylated form of HSPA5 may be a good target for breast cancer therapy.

Part I:

Downregulation of a tumor suppressor HSPB7, involved in the p53-pathway, in renal cell carcinoma by hypermethylation

Introduction

Renal cell carcinoma (RCC) accounts for approximately 2% of all cancers worldwide (1) and its incidence has increased by 2–3% in the last decade with even higher rate in developed countries (2-6). The underlying mechanisms such as some environmental and genetic risk factors including smoking, obesity, acquired cystic kidney disease and inherited susceptibility (von Hippel-Lindau disease) (3, 7, 8) have been indicated, but the etiological and pathological mechanisms of this disease are still far from fully understood.

Although local renal tumors can be surgically removed (9-11), distant metastasis is often observed even if the primary tumor is relatively small (12, 13). Patients with metastatic RCC generally result in extremely poor outcomes with overall median survival of around 13 months and the 5 year survival rate of <10% (13). For the advanced-stage patients, systemic therapy including immunotherapy (e.g. IL-2, IFN- α) and/or molecular-targeted drugs (e.g. sunitinib, bevacizumab, sorafenib, temsirolimus and everolimus) is recommended (14), but the response rates are not satisfactory.

To better understand the molecular mechanisms of renal carcinogenesis and apply the information for the development of effective treatment and early diagnosis, we performed genome-wide gene expression profile analysis and identified a small heat shock protein, HSPB7, whose function in cancer is unknown, to be downregulated in a great majority of human RCC samples.

In this study, we attempted to address two key questions, i) whether HSPB7 has growth suppressive function and ii) how HSPB7 is downregulated in RCCs. We here report for the first time that HSPB7 is likely to be a tumor suppressor which is frequently downregulated by DNA methylation in RCCs and is involved in the p53 pathway.

Materials and methods

Tissue samples and cell lines.

Tissue samples used in this study were obtained from patients with written informed consent at three hospitals: Juntendo University School of Medicine, Kochi University School of Medicine, and Kyoto Prefectural University of Medicine. The human RCC cell lines (Caki-1, Caki-2, 786-O, A498, ACHN), HEK293 and NCI-H1299 (lung carcinoma, p53-null) were purchased from American Type Culture Collection (ATCC; Rockville, MD, USA). Colon cancer cell lines HCT116 p53 wild-type (p53^{+/+}) and its derivative (p53^{-/-}) were gifts from Dr B. Vogelstein (Johns-Hopkins University, Baltimore, MD, USA). Normal human renal proximal tubule epithelial cells (RPTEC) were purchased from Lonza Walkersville Inc. (Walkersville, MD, USA). All cell lines were grown in monolayers in appropriate media

recommended by suppliers: Dulbecco's modified Eagle's medium (Gibco, Carlsbad, CA, USA) for HEK293, HCT116 (p53^{-/-}) and HCT116 (p53^{+/+}); Eagle's minimal essential medium (Gibco) for A498; McCoy's 5A medium (Gibco) for Caki-1 and Caki-2; RPMI-1640 medium (Gibco) for ACHN, 786-O and NCT-H1299; in addition, cells were supplemented with 10% fetal bovine serum (Cell Culture Bioscience, Nichirei Biosciences, Inc., Tokyo, Japan) except ACHN (5%), and 1% penicillin-streptomycin-amphotericin B suspension (Wako, Osaka, Japan). RPTEC were grown in REGM™ BulletKit, purchased from Lonza Walkersville Inc. (Walkersville, MD, USA). All cells were maintained at 37°C in humid air with 5% CO₂ condition. Cells were transfected with plasmids using FuGENE 6 transfection reagent (Roche, Basel, Switzerland) or Lipofectamine LTX and Plus reagent (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's protocols.

cDNA microarray and selection of candidate genes.

We prepared a genome-wide cDNA microarray with totally 27,648 cDNAs/ESTs selected from the UniGene database of the National Center for Biotechnology Information (NCBI). This microarray system was constructed as previously described (15, 16). We analyzed 15 clear cell renal cell carcinomas (RCC) and selected candidate genes according to the following criteria: i) genes for which we were able to obtain expression data in more than 50% of the cancers examined; ii) genes whose expression ratio was <0.2 in more than 50% of informative cases; and iii) the function of the gene was still unknown. Through these criteria, several candidates including HSPB7 were further validated. Gene expression data were deposited in the Gene Expression Omnibus database (accession no. GSE39364).

Quantitative real-time PCR (qPCR).

We extracted total RNA from the microdissected RCC clinical samples, microdissected normal renal cortex, 25 different normal organs (17) and cultured cells using RNeasy mini kits (Qiagen, Valencia, CA,

USA). RNAs from cell lines were reversely transcribed using the oligo (dT)21 primer and SuperScript III reverse transcriptase (Invitrogen). RNAs from tissue samples were treated with DNase I and subjected to two rounds of RNA amplification using T7-based in vitro transcription (Epicentre Technologies, Madison, WI, USA), then amplified RNAs were reversely transcribed to single-stranded cDNAs using random primer with Superscript II reverse transcriptase (Invitrogen) according to the manufacturer's instruction. QPCR was conducted using the SYBR-Green I Master (Roche) on a LightCycler 480 (Roche). Standard curve method was used for quantification analysis, and $\beta 2$ microglobulin (B2M) served as a control gene. The qPCR primers for HSPB7 in cell lines were: 5'-ACTTCTCACCTGAAGA CATCATTG-3' (forward) and 5'-CATGACAGTGCCG TCAGC-3' (reverse). The qPCR primers for HSPB7 in tissues were: 5'-GACCTTCCATCAGCCTTAACC-3' (forward) and 5'-ATGTGGGAGACGAAACCAAG-3' (reverse). The qPCR process was started at 95°C for 5 min, then underwent 45 cycles at 95°C for 10 sec, 55°C for 10 sec and 72°C for 10 sec. Data analysis including standard curve generation and copy number calculation was performed automatically. Each reaction was performed in duplicate and negative controls were included in each experiment.

Immunohistochemistry (IHC).

A kidney tissue array (BioChain Institute, Inc., USA) was used to analyze the protein expression of HSPB7 by IHC staining. This tissue array included 11 cases of RCC with corresponding normal tissues from the same patients as controls. Tissue sections were deparaffinized, rehydrated, and processed under high pressure (125°C, 30 sec) in antigen-retrieval solution of pH 9.0 (S2367, Dako, Carpinteria, CA, USA). Sections were blocked with Protein Block Serum Free (Dako) for 1 h at room temperature, followed by incubation with primary antibody (HSPB7, 1:100, Proteintech, Chicago, IL, USA) overnight at 4°C. At day 2, endogenous peroxidase activity was blocked by incubation in 3% hydrogen peroxide for 30 min at room temperature. Sections were incubated with a secondary antibody (Dako Envision+

system-HRP labeled polymer anti-rabbit K4003) for 30 min at room temperature, followed by DAB staining (K3468, Dako), counter staining with hematoxylin QS (H-3404, Vector Laboratories, Burlingame, CA, USA), dehydration and mounting. Three independent investigators semi-quantitatively assessed the HSPB7 positivity without prior knowledge of clinicopathological data. According to the intensity of HSPB7 staining, these samples were evaluated as: negative (-), weakly positive (+), moderate positive (++) and strong positive (+++). HSPB7 negative or weakly positive (-/+) were considered low expression, and moderate or strong positive were considered high expression (++/+++).

5-Aza-2'-deoxycytidine (5-Aza-dC) treatment.

5-Aza-dC (Sigma-Aldrich, St. Louis, MO, USA) was dissolved in DMSO. For a negative control, 5 RCC cell lines were treated with DMSO alone for 4 days. For a 5-Aza-dC group, cells were treated with DMSO for 1 day, following 5-Aza-dC-treatment (1, 3 and 10 μ M, respectively) for 3 days. On the fifth day, total RNAs of all cells were isolated using the RNeasy mini kits (Qiagen, Valencia, CA, USA), according to the manufacturer's directions. QPCR was subsequently performed to detect the expression of HSPB7. To detect the protein level of HSPB7 in 5 RCC cell lines after the same treatment (5-Aza-dC 1 μ M was used in 5-Aza-dC group), western blot and immunocytochemical (ICC) analyses were performed.

Bisulfite sequencing.

Genomic DNA was extracted from RPTEC, HEK293 and 5 RCC cell lines (Caki-1, Caki-2, 786-O, A498, ACHN) using the DNeasy blood and tissue kit (Qiagen). Genomic DNA (3.5 μ g each) were digested at 37°C for 16 h with 35 units of *Xho*I (Takara, Tokyo, Japan) and 1X H buffer (Takara) in 50 μ l of reaction volume. After treatment with phenol/chloroform/isoamyl alcohol (25:24:1, v/v), the DNA was finally dissolved in TE buffer and denatured in 0.3 N NaOH for 20 min at 37°C, and then the

unmethylated cytosine residues were sulfonated by incubation in 3.12 M of sodium bisulfite (pH 5.0, Sigma-Aldrich) and 0.5 mM of hydroquinone (Sigma-Aldrich) at 55°C for 16 h. The sulfonated DNA was recovered using the Nucleospin Extract (Macherey-Nagel GmbH and Co. KG, Düren, Germany) according to the manufacturer's recommendations. The conversion reaction was completed by desulfonating in 0.3 N NaOH for 20 min at 37°C. The DNA was ethanol precipitated, then washed by 70% ethanol and resuspended in TE buffer. Primers for bisulfite genomic sequencing PCR were designed by the use of the online program MethPrimer. The primers for region 1 were: 5'-TTT GAAGGGTTTTGGGTTTAATATAT-3' (forward) and 5'-CTCCTAACTACAACTATCCAACAC-3' (reverse). The primers for region 2 were: 5'-GGGTTGGTTTTAAGTTTAGGGATAG-3' (forward) and 5'-AAAAAAAATTCTATAACTCATCCAC-3' (reverse). The primers for region 3 were: 5'-TGTATATTGATGGAGGAGGTATAGT-3' (forward) and 5'-AAAAAAAATAAAAATCTTCTCCC-3' (reverse). The primers for region 4 were: 5'-TGGAGAAGG TTTTGAGTATGTTTTT-3' (forward) and 5'-CCACATCTATCCCTATAACCACATC-3' (reverse). The amplification products were checked by electrophoresis. After gel purification, the PCR products were cloned into pCR2.1-TOPO vector (Invitrogen), and 10 or more colonies were randomly chosen and sequenced. Methylation level analysis was performed by using QUMA software (<http://quma.cdb.riken.jp/>).

Construction of HSPB7 expression vector.

To construct an HSPB7 expression vector, the entire coding sequence of HSPB7 cDNA (based on NM_014424.4 in Pubmed) was amplified by PCR using KOD-Plus DNA polymerase (Toyobo, Osaka, Japan). The primers used for PCR reaction were 5'-AAAGAATTCCGTCCGTGGATGAGCCACAG-3' (forward) and 5'-TTTCTCGAGGATTTTGATCTCCGTCCGGA-3' (reverse). The PCR product was inserted into the *EcoRI* (Takara) and *XhoI* (Takara) sites of pCAGGSnHC expression vector containing the HA tag. The sequence and protein expression for pCAGGSnHC-HSPB7-HA were confirmed by DNA

sequencing, western blot and ICC analyses.

Western blot analysis.

To prepare whole cell extracts, cells were collected and lysed in chilled radioimmunoprecipitation assay buffer (RIPA) (50 mM Tris-HCl at pH 8.0, 150 mM sodium chloride, 0.1% SDS, 0.5% DOC, 1% NP-40), 1 mM phenyl methylsulphonyl fluoride (PMSF), 1 mM DTT and 0.1% Calbiochem Protease Inhibitor Cocktail Set III, EDTA-Free (EMD Chemicals Inc., Merck KGaA, Darmstadt, Germany). Following 15-min ultrasonication and subsequent 30-min incubation on ice, homogenates were centrifuged for 15 min at 4°C, and the supernatants were collected and boiled in SDS sample buffer. Each sample was loaded into a 15% SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and transferred to a nitrocellulose membrane (Hybond™ ECL™, Amersham, Piscataway, NJ, USA). Protein bands on western blots were visualized by chemiluminescent detection (ECL, Amersham). The primary antibodies used in this study included rabbit anti-human HSPB7 polyclonal antibody (Proteintech, diluted 1:500) and goat anti-rabbit IgG-HRP secondary antibody (Santa Cruz Biotechnology, Santa Cruz, CA, USA, diluted 1:30,000).

Immunocytochemistry (ICC).

Five RCC cell lines were seeded on Lab-Tek II chamber slide system (Nalge Nunc International). At day 5 after the 5-Aza-2'-dC-treatment, the cells were fixed with 4% paraformaldehyde in PBS for 10 min and permeabilized with 0.2% Triton X-100 in PBS for 5 min at room temperature. Cells were covered with blocking solution (3% BSA in PBS contained 0.2% Triton X-100) for 60 min at room temperature. Then the cells were incubated with rabbit anti-human HSPB7 polyclonal antibody (Proteintech, diluted 1:250) overnight at 4°C, following an Alexa Fluor 488 goat anti-rabbit IgG antibody (Molecular Probes, Eugene, OR, USA, diluted 1:1,000) for 1 h at room temperature. PBS or 0.2% Triton X-100 in PBS was used for washing after each step. Then cells were stained with DAPI (Vector) and viewed with a laser scanning

spectral confocal microscope (Leica TCS SP2).

Colony formation assay.

Cells were plated in a 6-well plate and transfected with pCAGGSnHC-HSPB7-HA or empty vector using FuGENE 6 (ACHN and Caki-1) or lipofectamine LTX (Caki-2, A498 and 786-O) transfection reagent (Roche). After 48 h of transfection, cells were selected with G418 (Gibco) for 14-21 days. Colonies (>1 mm diameter) were counted using the Image J software after fixed with methanol and stained with 0.1% crystal violet. The experiment was carried out twice in duplicate wells.

DNA-damaging treatments.

When cells reached 60-70% confluence in the culture dish, HCT116 (p53^{-/-}) and HCT116 (p53^{+/+}) cells were incubated with adriamycin for 2 h at the indicated concentration. The cells were harvested at different time points after cell-damaging treatment as indicated in the figure legends. Replication-deficient recombinant adenovirus encoding p53 (Ad-p53) or LacZ (Ad-LacZ) was generated and purified as previously described (18, 19). NCI-H1299 lung cancer cells were infected with viral solutions at an indicated multiplicity of infection (MOI) and incubated at 37°C until harvest.

p53-binding site screening by Luciferase assay.

Two DNA fragments including candidate p53-binding sites of HSPB7 were amplified by PCR, digested with *Mlu*I and *Bgl*III and cloned into pGL3-Promoter vector (Promega, Madison, WI, USA). Primer sequences (including *Mlu*I and *Bgl*III site) for p53-binding sites of HSPB7 were: region 1 forward, 5'-AAAACGCGTTCCAAGGTCACACAGCAGAG-3'; and reverse, 5'-TTTAGATCTGCTTCAAACCGGTCATCCT-3'; and region 2 forward, 5'-

AAAACGCGTTGAGCAGGAGCAGTCAGAGA-3'; and reverse, 5'-TTTAGATCTAGCCCCAAG AGGACAAAGTT-3'.

H1299 cells were seeded in 12-well plates (5×10^4 cells per well). Twenty-four hours later, cells were co-transfected with i) 25 ng of the pRL-CMV vector (Promega) (for internal control); ii) 125 ng of either pcDNA3.1(+)-wild-type p53 or pcDNA3.1(+) empty vector; and iii) 125 ng of pGL3-promoter vector with either the p21 promoter region corresponding to p53-binding site (for positive control) (20), that with p53-binding site 1 of HSPB7, that with p53-binding site 2 of HSPB7, or pGL3-Promoter mock vector (for negative control) by using FuGENE 6 transfection reagent (Roche). After 36 h incubation, luciferase activity was measured using the Dual Luciferase Assay System (Promega) (21).

Statistical analysis.

All statistical analyses including t-test and Fisher's exact test were carried out by using the SPSS software (version 17). Data are shown as mean \pm SD. All tests were 2-sided and p-value of <0.05 was considered to indicate a statistically significant difference.

Results

Downregulation of HSPB7 in RCC.

Based on the analysis of microarray data of 15 clear cell renal cell carcinomas, we found HSPB7 to be significantly and commonly downregulated in RCC. QPCR experiment confirmed its downregulation in 11 (85%) of 13 RCC tissues and in all of the five RCC cell lines (Fig. 1A and B), compared with their corresponding normal controls. IHC analysis of a tissue array consisting of 11 pairs of human RCC

sample revealed that the expression of HSPB7 was significantly higher in normal kidney tissues than that in RCC tissues (Fig. 1C, Table 1, supplementary figure 1 and Table 1). We also detected HSPB7 expression mainly in the cytoplasm of normal renal tubular epithelial cells. To explore the expression patterns of HSPB7 in other normal tissues, we performed qPCR analysis using mRNAs isolated from 25 normal tissues. HSPB7 expression was detected ubiquitously in human tissues (Fig. 2).

5-Aza-dC treatment restores HSPB7 expression in RCC cell lines.

To investigate whether the methylation status of the HSPB7 gene could affect HSPB7 expression in RCCs, 5 RCC cell lines, Caki-1, Caki-2, ACHN, 786-O and A498 were treated with a demethylating agent, 5-Aza-dC, and then the expression levels of HSPB7 were analyzed by qPCR, western blot and IHC analysis. We found that HSPB7 mRNA expression were restored in all of the 5 RCC cell lines after the treatment with 5-Aza-dC (Fig. 3A), and the HSPB7 protein expression could also be detected in two cell lines, 786-O and A498, in which mRNA expression was most highly induced (Fig. 3B), indicating suppression of HSPB7 in RCC was caused probably by DNA hypermethylation. We performed exon sequencing of HSPB7 in these five RCC cell lines, but no mutation or deletion/insertion was detected (data not shown).

Hypermethylation of HSPB7 in RCC.

To confirm the methylation status of the HSPB7 gene, bisulfite sequencing was performed in the 5 RCC cell lines, Caki-1, Caki-2, ACHN, 786-O and A498, as well as 2 normal renal cell lines, RPTEC and HEK293. We first screened two CpG islands, regions 1 and 2 (Fig. 3C), but no significant difference of methylation status was found between normal and cancer cell lines. Then, we performed the second screening for regions 3 and 4 (Fig. 3C and supplementary figure 2) (we also screened the other regions in

normal cells, but data are not shown). In region 4, we found significantly higher levels of methylation in the 5 RCC cell lines than in the 2 normal renal cell lines.

Ectopic HSPB7 expression suppresses RCC cell clonogenicity.

To study the effect of HSPB7 expression on tumor growth, Caki-1 and ACHN cells were transfected with HSPB7 expression vector, pCAGGSnHC-HSPB7-HA. Introduction of HSPB7 into these two cancer cell lines caused significant decrease in the number of colonies, compared with corresponding mock-transfected controls (Fig. 4A). We also performed colony formation assay in 3 other RCC cell lines (Caki-2, A498 and 786-O) using the same vectors, and confirmed similar growth-suppressive effects (Fig. 4B), implying that HSPB7 may function as a tumor suppressor gene.

HSPB7 is regulated by p53.

To further elucidate the biological significance, we first investigated its possible involvement in the p53-pathway because α B-crystallin, one of the small heat shock protein family members, was reported to be induced by p53 (22, 23). We applied qPCR analysis to evaluate the expression of HSPB7 in NCI-H1299 (p53 null) cell lines with or without introduction of p53 using the adenovirus system. After the infection of Ad-p53, we observed induction of HSPB7 in a dose- and time-dependent manner (Fig. 5A and B), while no induction was observed in the control cells. After the 48-hour treatment with 40 MOI of Ad-p53, the expression level of HSPB7 became nearly 5 times higher than the control cells (Fig. 5A). Induction of HSPB7 was also confirmed under the treatment with relative lower dose of Ad-p53 (8 MOI) at different time points. Concordantly, DNA damage by adriamycin treatment induced HSPB7 expression in HCT116 cells with wild-type p53, but not in HCT116 cells without wild-type p53 (Fig. 5C and D), indicating that HSPB7 expression is regulated by wild-type p53. To further investigate whether HSPB7 is directly regulated by p53, we screened three possible p53-binding sites indicated by the p53-binding site search

software developed by us, but neither of these candidate sites was confirmed to be a direct p53-binding site (Supplementary figure 3). However, it is still possible that p53 can bind to another site(s) of HSPB7, we are unable to conclude whether HSPB7 is directly or indirectly regulated by p53. Until now, we are only sure that HSPB7 expression is inducible by wild-type p53.

Discussion

Scarce knowledge exists on the biological function of HSPB7, a member of the small heat shock protein family that is characterized by possessing a conserved α -crystallin domain. HSPB7 has been shown to interact with the cytoskeletal protein α -filamin (24) as well as other small heat shock proteins (25). HSPB7 belongs to a non-canonical HSPB protein that prevents the aggregation of polyQ proteins in an active autophagy machinery, but overexpression of HSPB7 alone did not affect the autophagy event (26). Several genome-wide association studies found that SNPs in the HSPB7 gene were strongly associated with idiopathic cardiomyopathies and heart failure (27-31). Recently, HSPB7 was suggested to regulate early developmental steps in cardiac morphogenesis (32). However, the involvement of HSPB7 in carcinogenesis has not been described.

Through the genome-wide expression analysis in RCCs, we identified HSPB7 as a candidate tumor suppressor gene because of its common and significant downregulation in RCCs. Subsequent functional analysis revealed that HSPB7 was downregulated in cancer cells by hypermethylation. Bisulfite sequencing of genomic regions of HSPB7 confirmed hypermethylation in RCC cell lines. Although region 4 (Fig. 3C and supplementary figure 2) contained no CpG Island, we observed significantly higher level of methylation in RCC cell lines than normal cell lines. Consistently, restoration of HSPB7

expression was observed by the treatment of cancer cells with 5-Aza-dC. In addition, since no somatic change in coding regions of the HSPB7 gene was found in our sequence analysis of RCC cell lines or in the COSMIC database, HSPB7 in RCC is considered to be downregulated mainly by hypermethylation.

The second key finding in this study is that HSPB7 showed growth suppressive effect in cancer cells. Ectopic expression of HSPB7 significantly impaired colony-forming ability for 5 RCC cell lines, indicating that HSPB7 may function as a tumor suppressor gene. Similarly α B-crystallin, one of the small heat shock protein family members, was also indicated to function as a tumor suppressor in nasopharyngeal carcinoma cells (33). Furthermore, the region on chromosome 1p36.23-p34.3, where HSPB7 is located, showed frequent loss of heterozygosity in many types of solid tumors (34). However, further studies are needed to clarify the detailed tumor suppressor function of HSPB7 in RCC.

The third important finding in this study is that HSPB7 was likely to be involved in the p53 pathway. The expression of HSPB7 was significantly induced in p53-dependent manner that was clearly demonstrated by two experiments, i) that introduction of adeno-p53 in p53-negative cancer cells showed strong induction of HSPB7 and ii) that DNA-damage-dependent induction of HSPB7 was observed in HCT116 cells with wild-type p53, but not in those lacking p53. Although we failed to identify the p53-binding site in or near the HSPB7 gene, these two pieces of evidence strongly imply a critical role of HSPB7 as the direct/indirect p53-signal transducer and its downregulation may be involved in the development of various types of cancer including RCC.

In conclusion, we carried out a genome-wide gene expression analysis and identified HSPB7 to be a candidate tumor suppressor gene in RCC. We confirmed downregulation of this gene caused by DNA hypermethylation, its growth suppressive effect in RCC cell lines and its p53-dependent expression,

indicating the important roles of HSPB7 in renal carcinogenesis. Our finding could contribute to better understanding of the novel function of HSPB7 in cancer.

Figures and tables

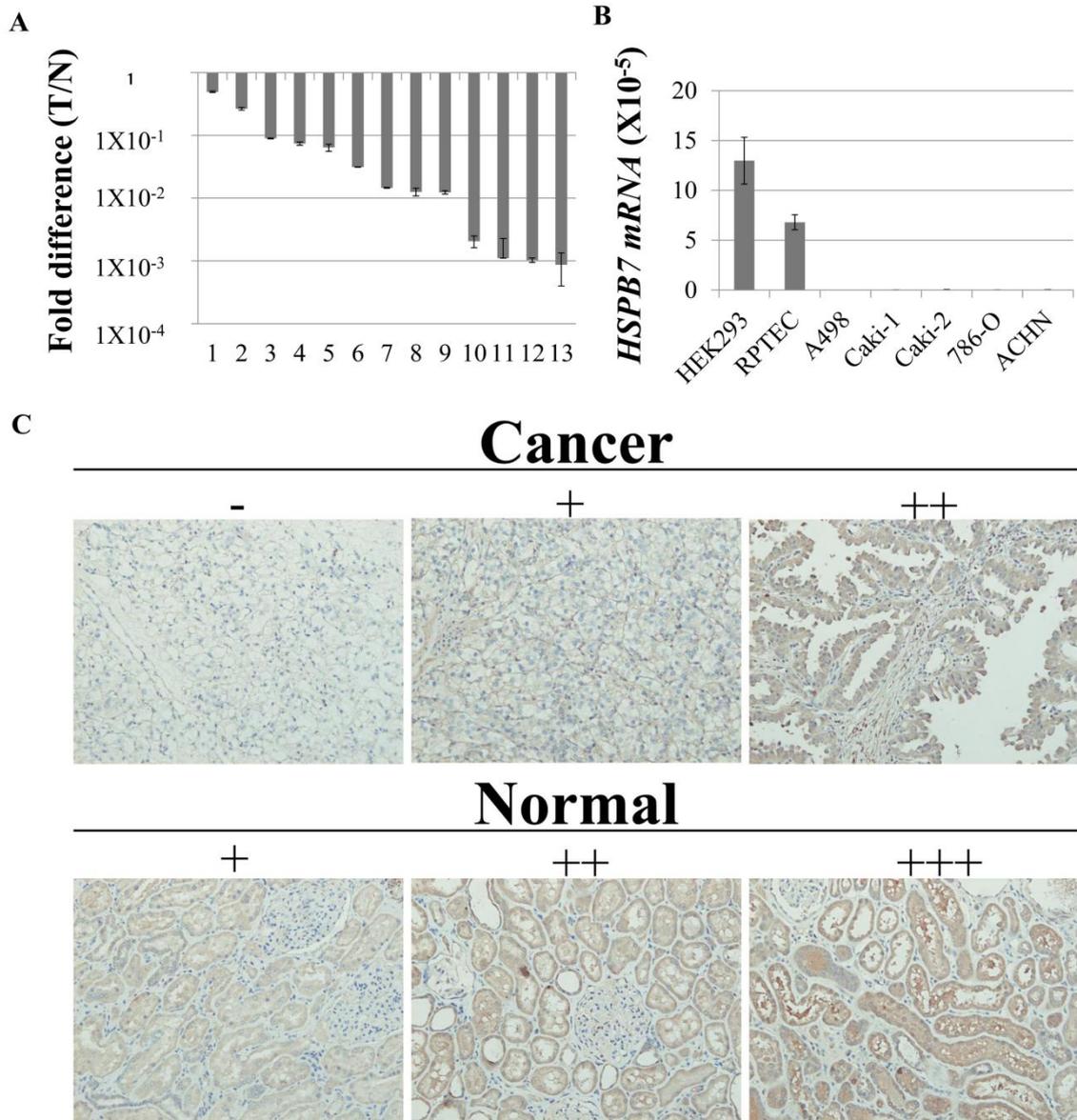


Figure 1. Downregulation of HSPB7 in RCC.

QPCR analysis showed that HSPB7 mRNA expression was significantly downregulated (A) in 11 (85%) of 13 RCC tissues compared with the normal renal tissue, and (B) in all of five RCC cell lines compared with normal HEK 293 and RPTEC cells. T and N represent RCC tissue sample and normal renal tissue,

respectively. B2M (β 2 microglobulin) was used for normalization of expression levels. Values are expressed as the mean \pm SD. (C) IHC analysis of a tissue array consisting of 11 pairs of human RCC sample revealed that the expression of HSPB7 was significantly higher in normal kidney tissues than in RCC tissues. According to the intensity of HSPB7 staining, these samples were evaluated as: negative (-), weakly positive (+), moderate positive (++) , and strong positive (+++). HSPB7 negative or weakly positive (-/+) were considered low expression, and moderate or strong positive were considered high expression (+/+++). Summary of the IHC results is shown in Table 1.

Table 1. Immunohistochemical expression of HSPB7 in RCC tissue array

| | Total | Low (-/+) | High (++ /+++) | Fisher's t-test |
|------------|-------|--------------|-------------------|--------------------|
| Clear cell | | | | |
| Cancer | 9 | 7 | 2 | P=0.015 |
| Normal | 9 | 1 | 8 | |
| Papillary | | | | |
| Cancer | 2 | 1 | 1 | - |
| Normal | 2 | 0 | 2 | |
| Total | | | | |
| Cancer | 11 | 8 | 3 | P=0.008 |
| Normal | 11 | 1 | 10 | |

All tests were 2 sided and $P < 0.05$ was considered statistically significant.

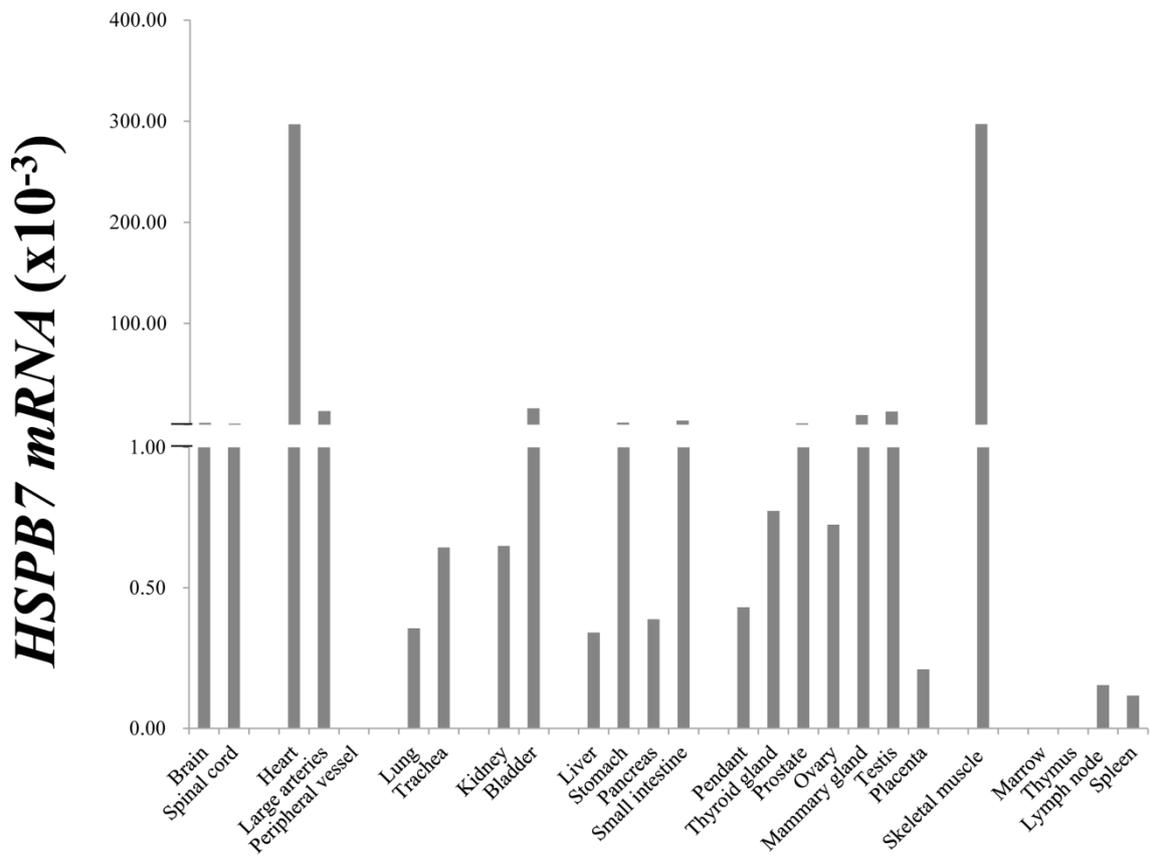


Figure 2. HSPB7 expression levels in normal tissues.

QPCR analysis of HSPB7 was performed using mRNA isolated from 25 different normal tissues. B2M was used for normalization of expression levels.

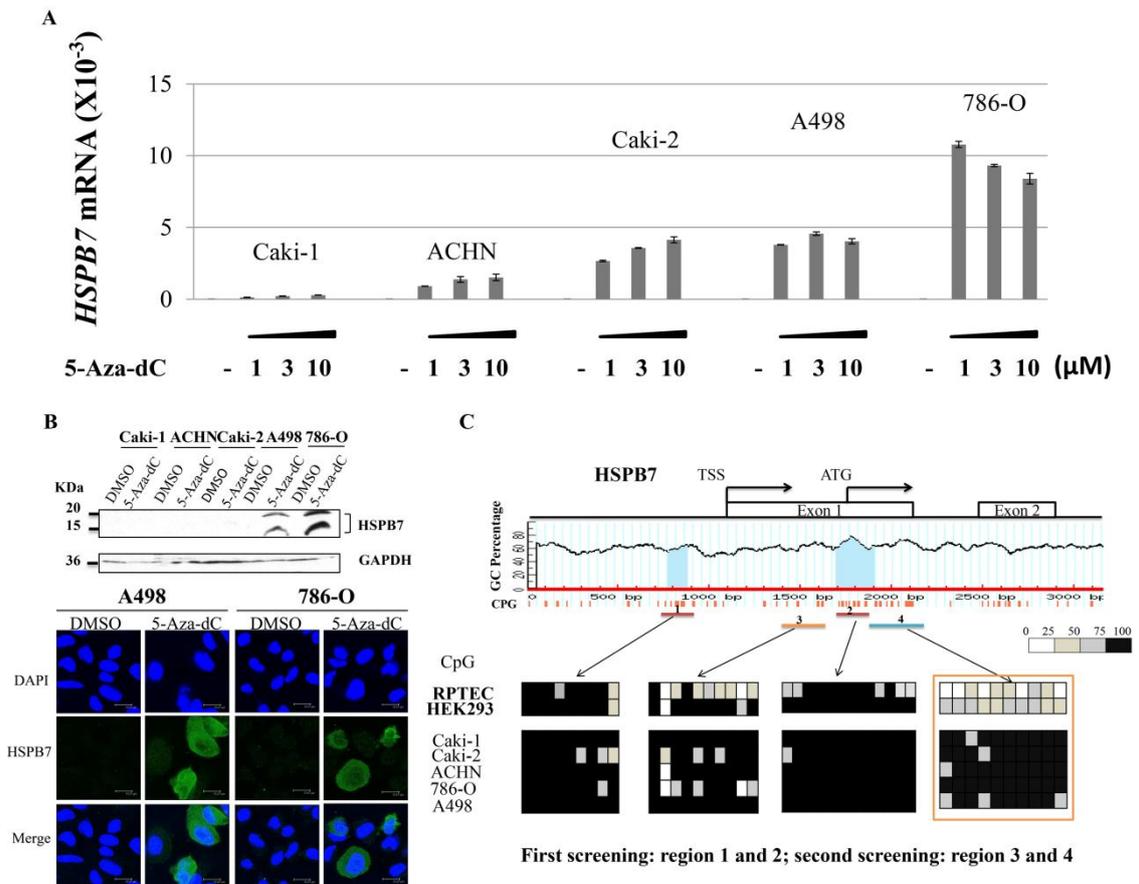


Figure 3. Epigenetic silencing of HSPB7 in RCC cell lines.

(A) QPCR analysis and (B) western blot and ICC analysis of HSPB7 were performed in five RCC cell lines with or without treatment of the demethylating agent 5-Aza-dC. B2M was used for normalization of mRNA expression levels. GAPDH was used for normalization of protein expression levels. Values are expressed as the mean \pm SD. (C) Hypermethylation of HSPB7 was confirmed by means of bisulfite sequencing. For each of the regions 1-4 in the cell lines, 10 or more colonies were randomly chosen and sequenced. Each square indicates a CpG site, and an average methylation level per CpG site is indicated by % methylation (shown in different color): white, 0-25% methylation; bright grey, 26-50% methylation; dark grey, 51-75% methylation; and black, 76-100% methylation. Region 4 showed higher level of methylation in the five RCC cell lines (Caki-1, Caki-2, ACHN, 786-O and A498) than in the two control cell lines (RPTEC and HEK293).

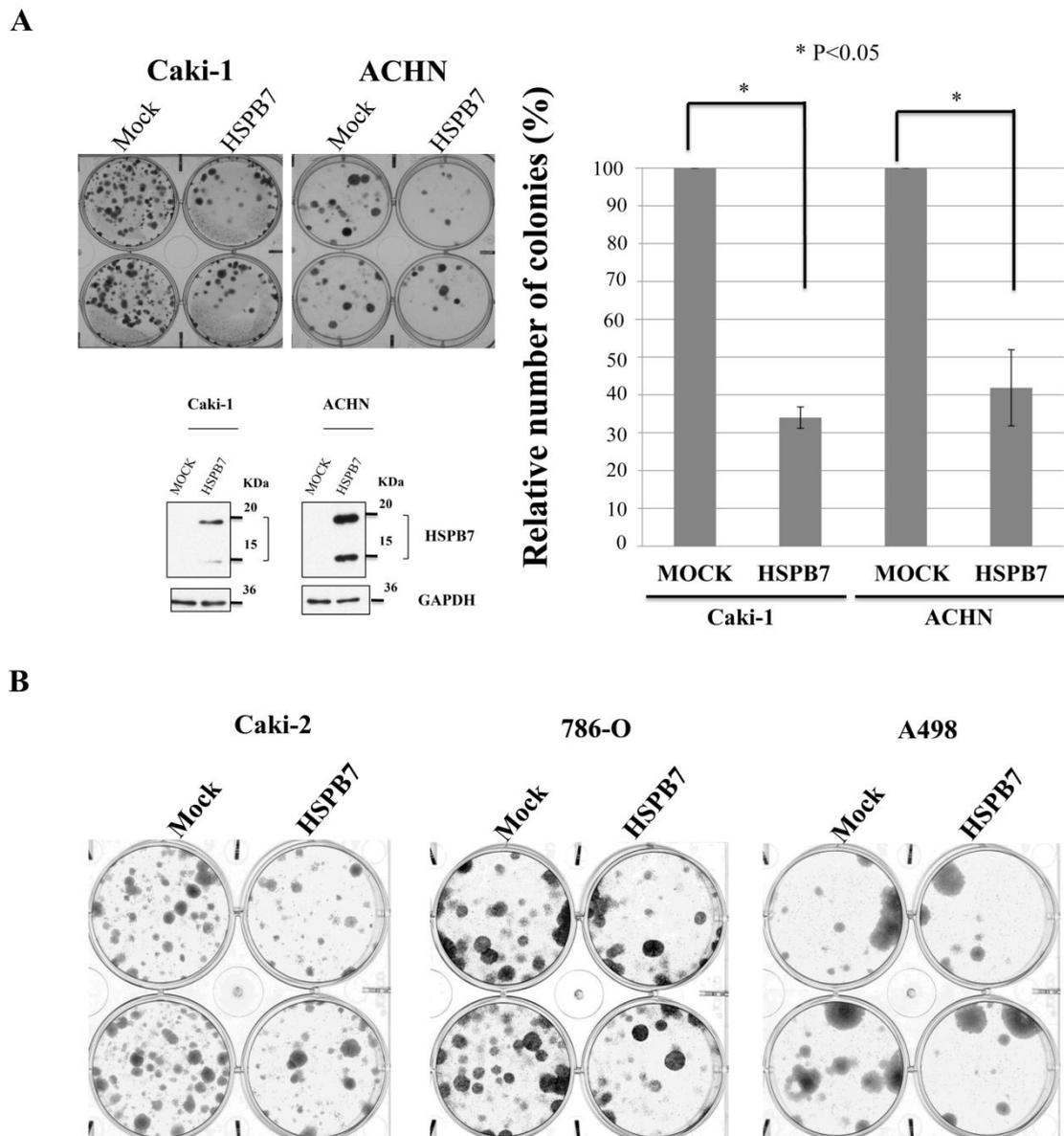


Figure 4. Ectopic HSPB7 expression suppresses RCC cell growth.

(A) Colony formation assay showed that introduction of HSPB7 impaired colony-forming ability of Caki-1 and ACHN cells. Cells were transfected with plasmid expressing HSPB7 or mock plasmid, and colonies (>1 mm diameter) were counted after selection of 2-3 weeks with G418. At 48 h after transfection, total protein of cells was collected and applied for western blot to confirm the successful transfection. GAPDH was used for the normalization of protein expression levels. (B) Colony formation assay in Caki-2, 786-O and A498 RCC cell lines. Values are expressed as the mean \pm SD

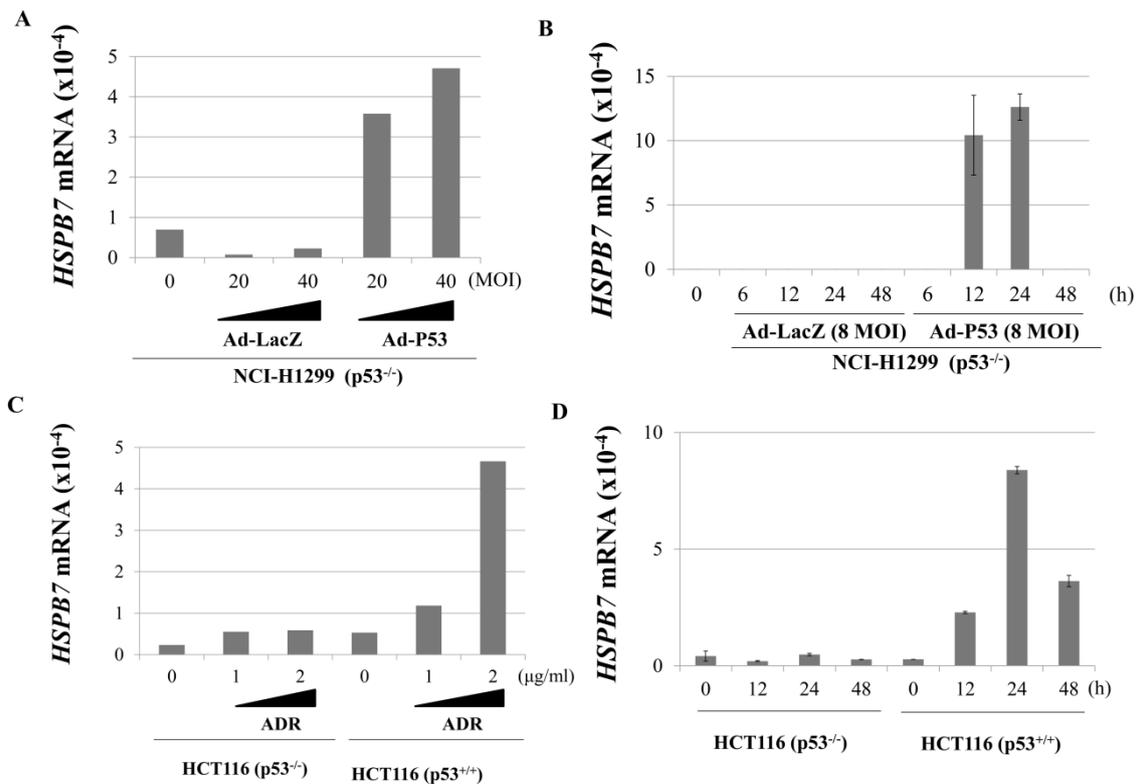


Figure 5. HSPB7 is regulated by p53.

(A and B) The induction of HSPB7 was only observed in NCI-H1299 cells with p53 infection at (A) dose- and (B) time-dependent manner, but not in control cells. (A) Cells were infected with replication-deficient recombinant adenovirus encoding p53 (Ad-p53) or LacZ (Ad-LacZ) at indicated doses, 48 h later, the cells were collected and qPCR analysis was performed. (B) The cells were infected with 8 MOI Ad-p53 and then collected at different time points. (C and D) HSPB7 was (C) dose- and (D) time-dependently induced in HCT116 (p53^{+/+}) cells treated with adriamycin. (C) Cells were treated with adriamycin at indicated doses for 2 h and then harvested at 48 h. (D) The cells were treated with adriamycin at 2 μg/ml for 2 h and then harvested at different time points. B2M was used for normalization of expression levels. Values are expressed as the mean ± SD.

Supplementary data

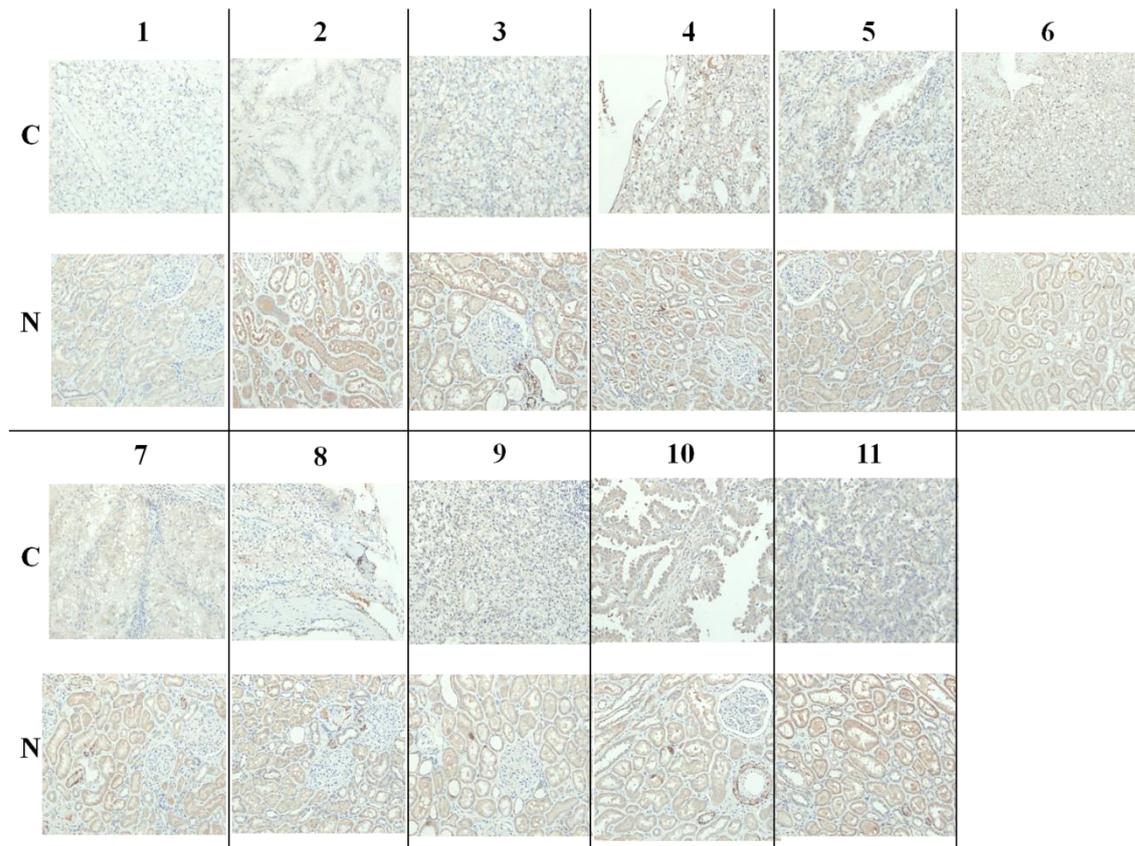
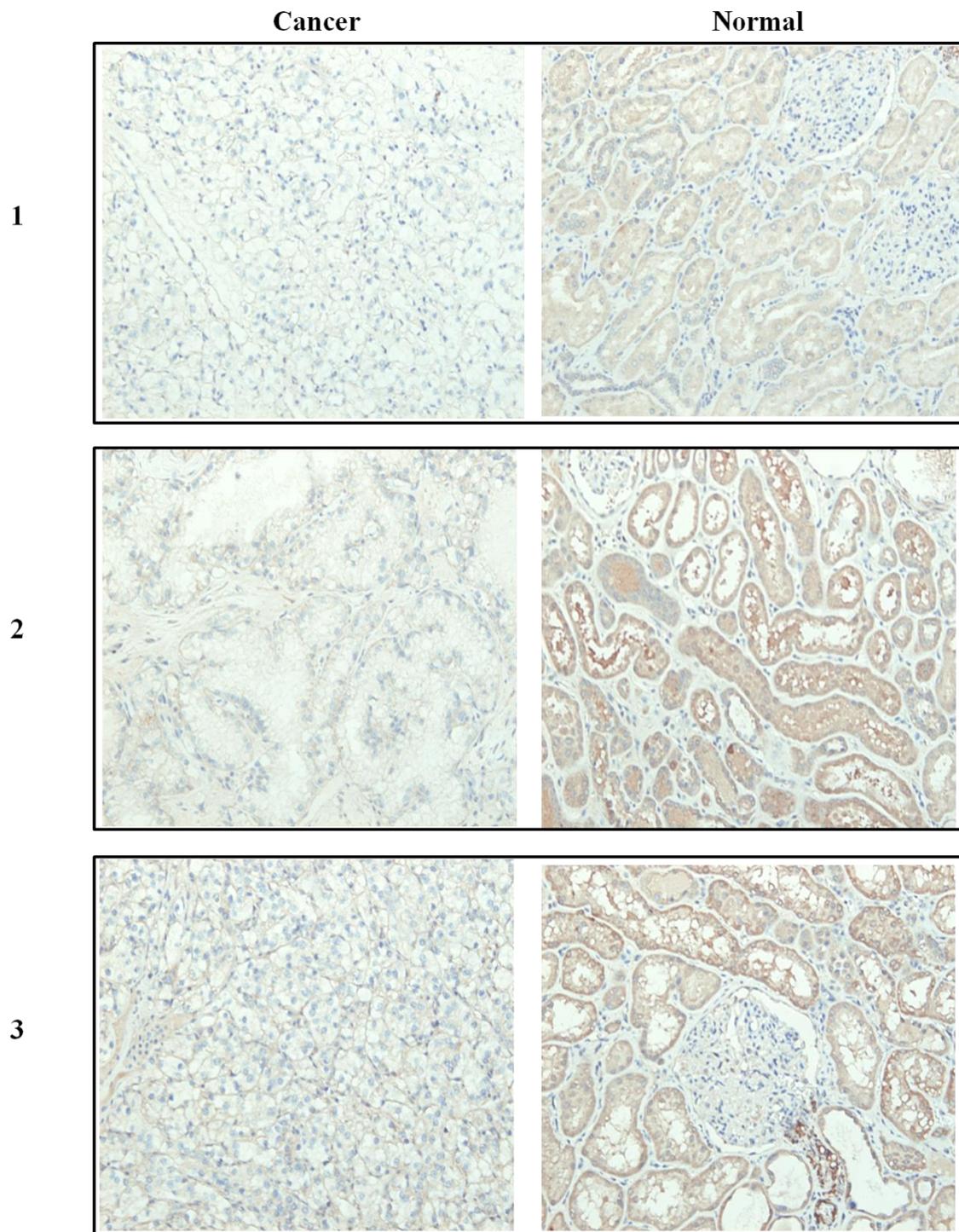


Figure 1. IHC analysis of HSPB7 in a tissue array consisting of 11 pairs of human RCC sample

Raw data of HSPB7 protein expression in a tissue array consisting of 11 pairs of human RCC sample. C: cancer; N: normal. 1~11: the number of pair RCC sample.

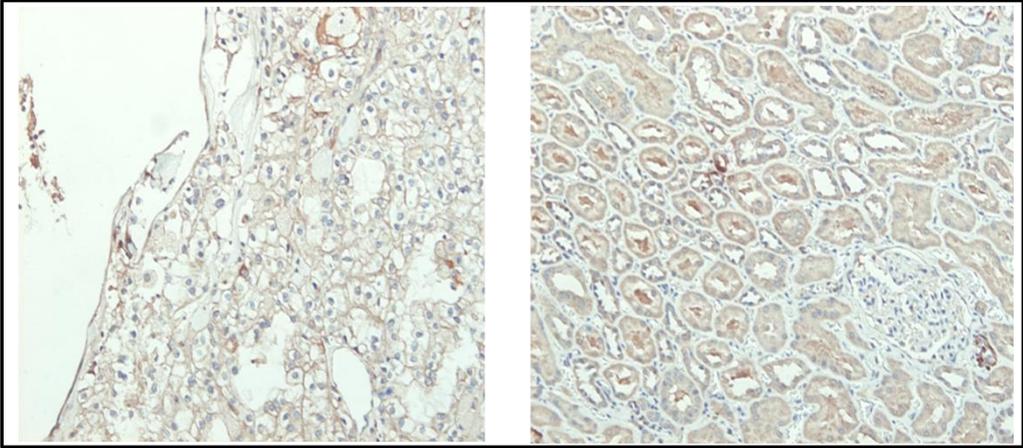
Enlarged pictures for supplementary figure 1



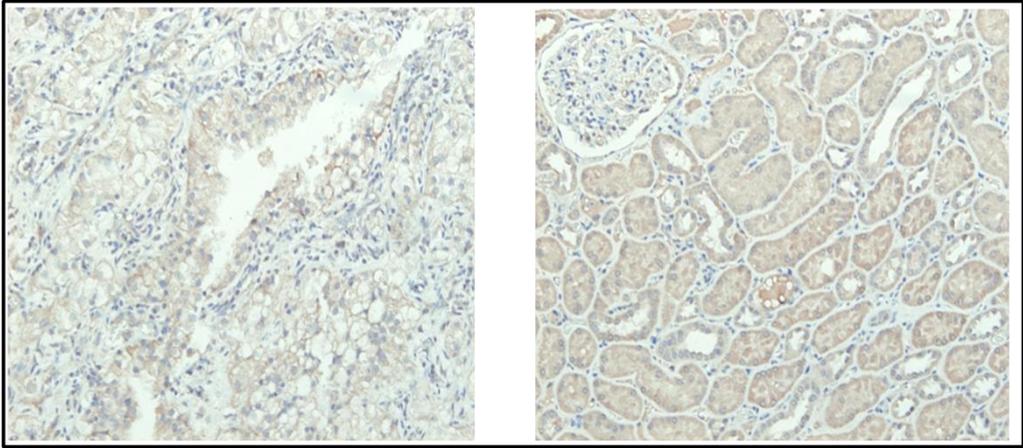
Cancer

Normal

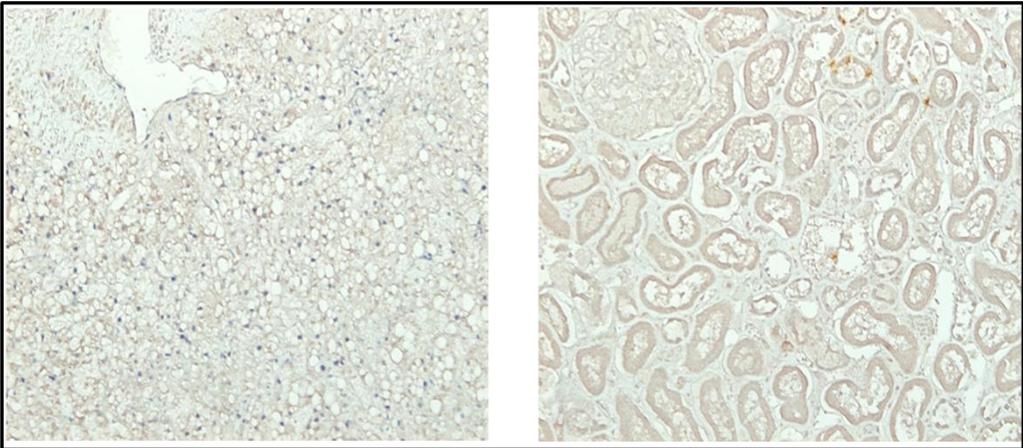
4



5



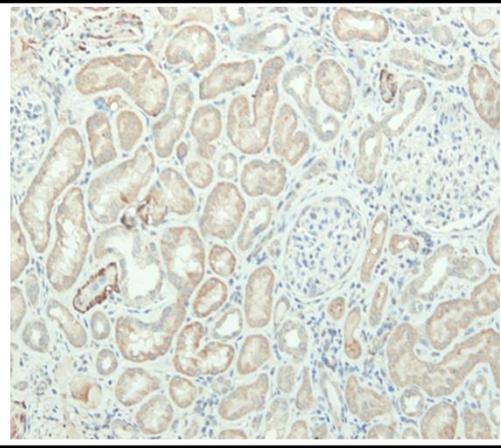
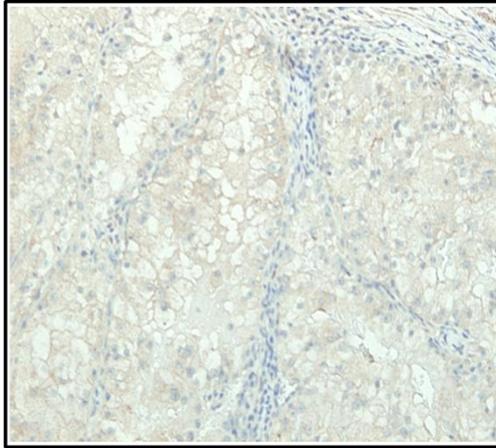
6



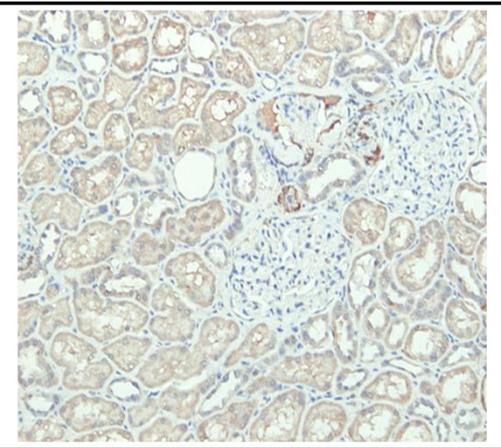
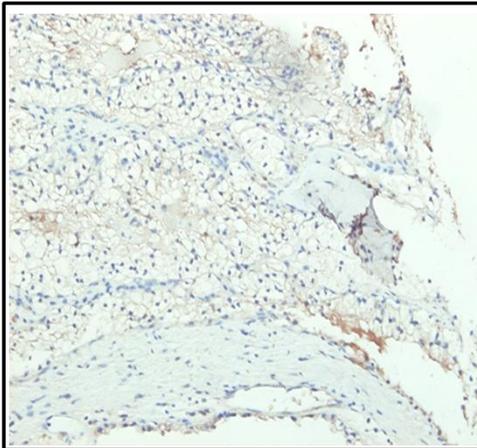
Cancer

Normal

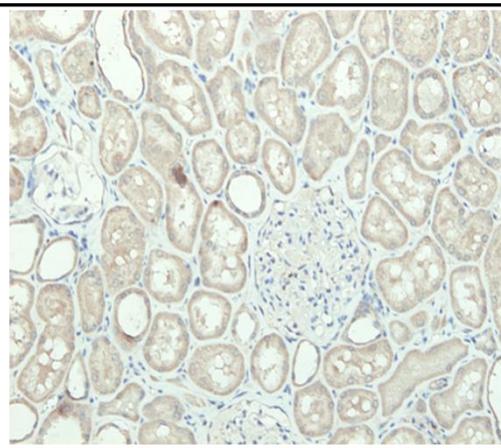
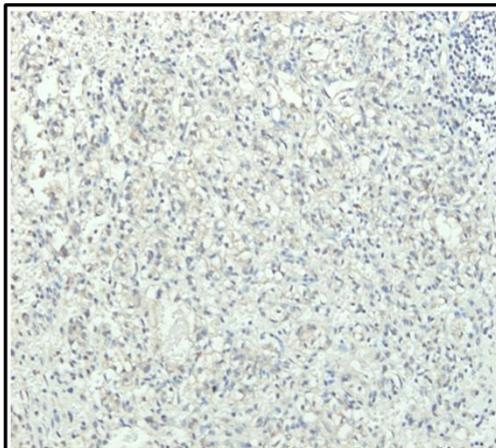
7



8



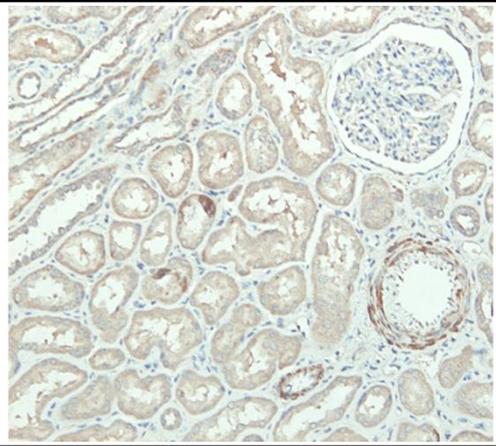
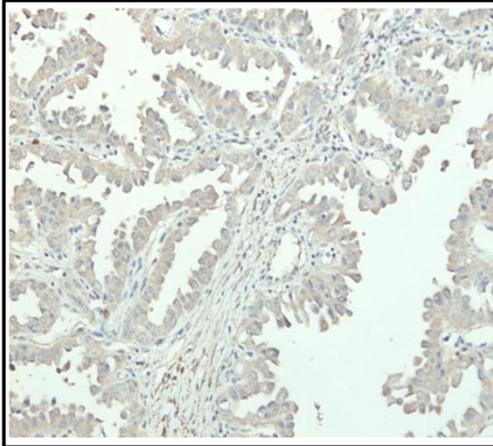
9



Cancer

Normal

10



11

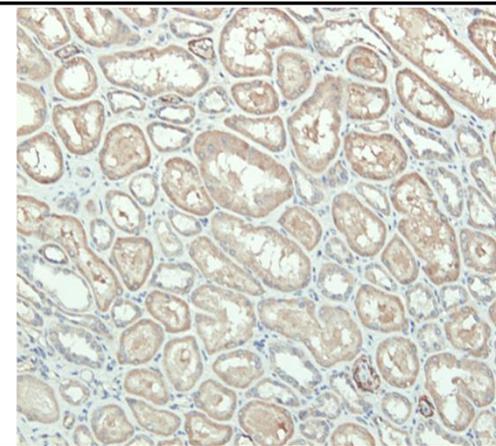
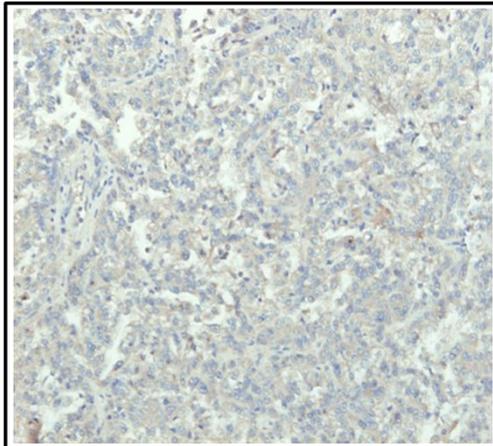


Table 1. Clinical information of the tissue array

| Position | Sex | Age | Pathology | Stage (TNM) |
|-----------------|------------|------------|---------------------------------|--------------------|
| C1 | F | 53 | Clear cell carcinoma | T1N0M0 |
| N1 | F | 53 | Uninvolved kidney tissue of C1 | |
| C2 | M | 58 | Clear cell carcinoma | T1N0M0 |
| N2 | M | 58 | Uninvolved kidney tissue of C2 | |
| C3 | F | 56 | Clear cell carcinoma | T2N0M0 |
| N3 | F | 56 | Uninvolved kidney tissue of C3 | |
| C4 | M | 50 | Clear cell carcinoma | T1N0M0 |
| N4 | M | 50 | Uninvolved kidney tissue of C4 | |
| C5 | M | 60 | Clear cell carcinoma | T1N0M0 |
| N5 | M | 60 | Uninvolved kidney tissue of C5 | |
| C6 | M | 56 | Clear cell carcinoma | T1N0M0 |
| N6 | M | 56 | Uninvolved kidney tissue of C6 | |
| C7 | M | 55 | Clear cell carcinoma | T1N0M0 |
| N7 | M | 55 | Uninvolved kidney tissue of C7 | |
| C8 | M | 67 | Clear cell carcinoma | T1N0M0 |
| N8 | M | 67 | Uninvolved kidney tissue of C8 | |
| C9 | M | 57 | Clear cell carcinoma | T1N0M0 |
| N9 | M | 57 | Uninvolved kidney tissue of C9 | |
| C10 | F | 29 | Papillary adenocarcinoma | T1N1M0 |
| N10 | F | 29 | Uninvolved kidney tissue of C10 | |
| C11 | M | 63 | Papillary adenocarcinoma | T1N0M0 |
| N11 | M | 63 | Uninvolved kidney tissue of C11 | |

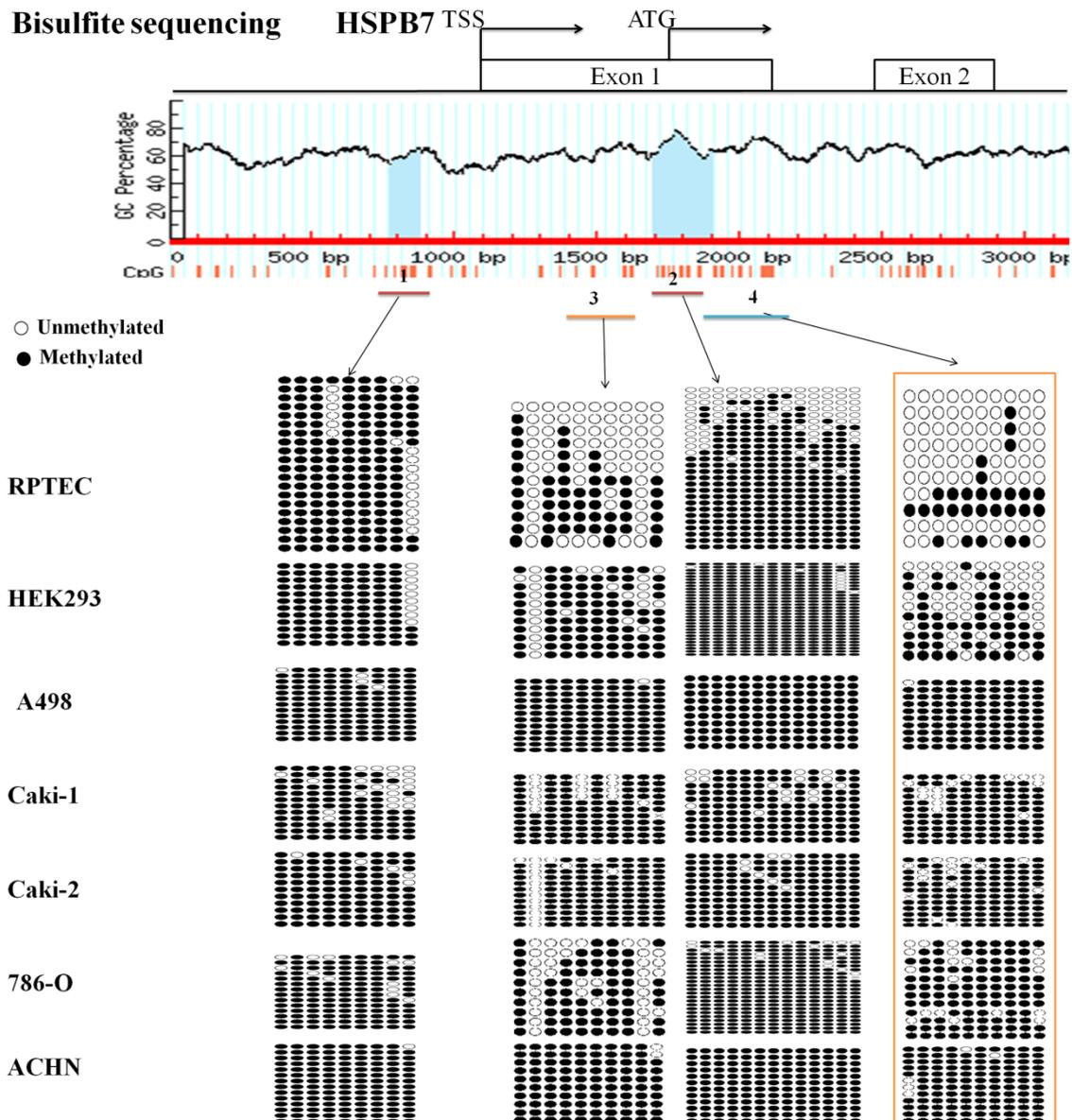
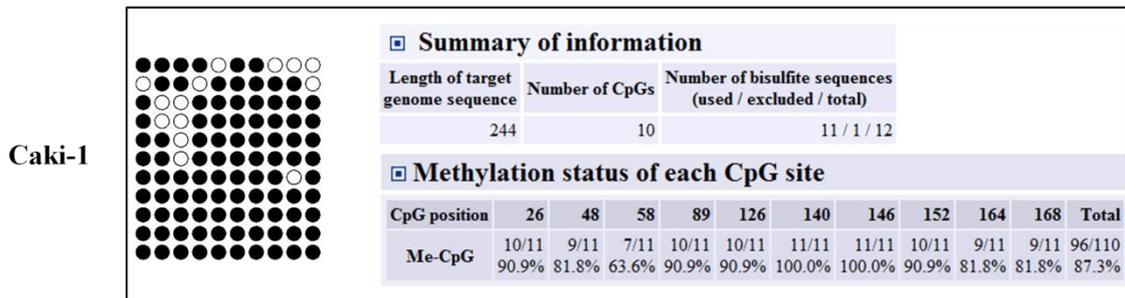
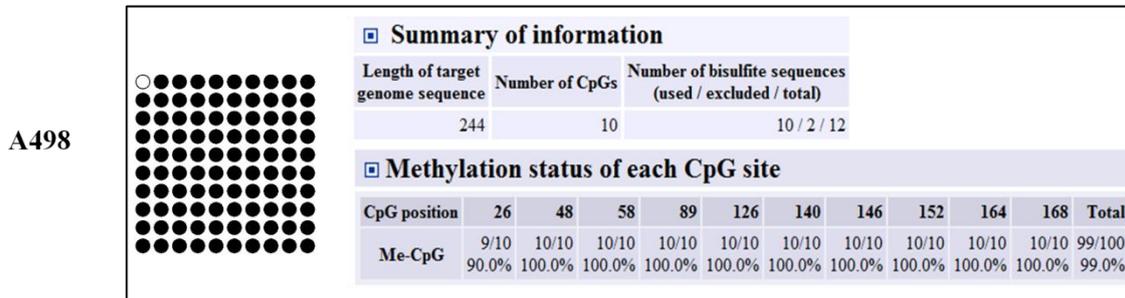
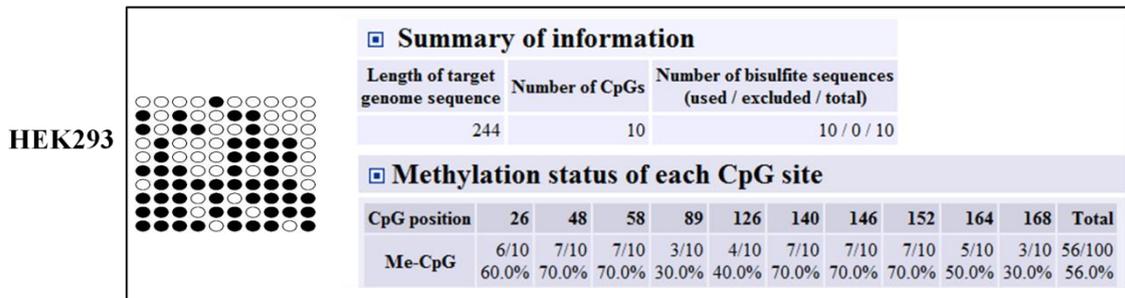
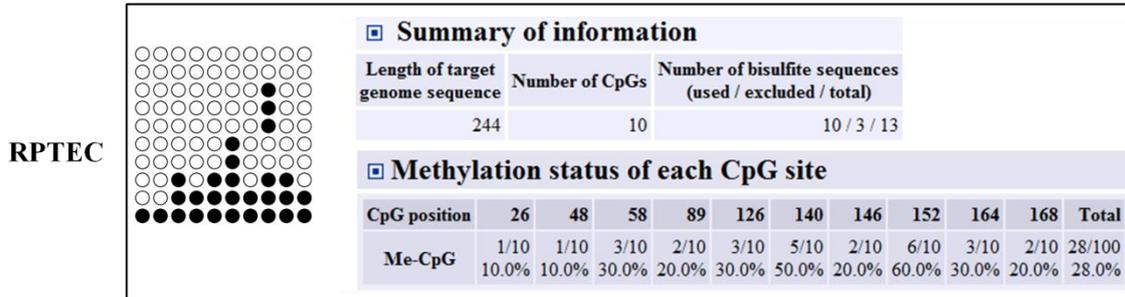


Figure 2. Raw data of bisulfite sequencing

Raw data of bisulfite sequencing in 5 RCC cell lines (A498, Caki-1, Caki-2, 786-O, and ACHN) and 2 normal cell lines (RPTEC and HEK293) was showed. We screened four regions. For each cell line in each region, 10 or more colonies were randomly chosen and sequenced. Methylation level analysis was performed by using QUMA software (<http://quma.cdb.riken.jp/>). In region 4, we observed higher methylation level in 5 RCC cell lines when compared with 2 normal cell lines.

Analysis data of region 4 in figure 2

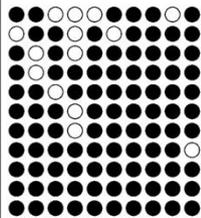
Result Analysis by QUMA software



Result

Analysis by QUMA software

Caki-2



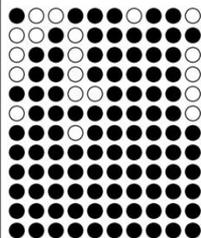
Summary of information

| Length of target genome sequence | Number of CpGs | Number of bisulfite sequences (used / excluded / total) |
|----------------------------------|----------------|---|
| 244 | 10 | 11 / 2 / 13 |

Methylation status of each CpG site

| CpG position | 26 | 48 | 58 | 89 | 126 | 140 | 146 | 152 | 164 | 168 | Total |
|--------------|----------------|---------------|---------------|---------------|----------------|----------------|-----------------|-----------------|----------------|----------------|-----------------|
| Me-CpG | 10/11 90.9% | 9/11 81.8% | 9/11 81.8% | 6/11 54.5% | 10/11 90.9% | 10/11 90.9% | 11/11 100.0% | 11/11 100.0% | 10/11 90.9% | 10/11 90.9% | 96/110 87.3% |

786-O



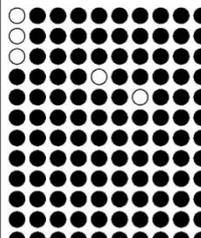
Summary of information

| Length of target genome sequence | Number of CpGs | Number of bisulfite sequences (used / excluded / total) |
|----------------------------------|----------------|---|
| 244 | 10 | 12 / 2 / 14 |

Methylation status of each CpG site

| CpG position | 26 | 48 | 58 | 89 | 126 | 140 | 146 | 152 | 164 | 168 | Total |
|--------------|---------------|----------------|----------------|---------------|----------------|-----------------|----------------|-----------------|-----------------|---------------|------------------|
| Me-CpG | 8/12 66.7% | 10/12 83.3% | 11/12 91.7% | 7/12 58.3% | 11/12 91.7% | 12/12 100.0% | 11/12 91.7% | 12/12 100.0% | 12/12 100.0% | 7/12 58.3% | 101/120 84.2% |

ACHN



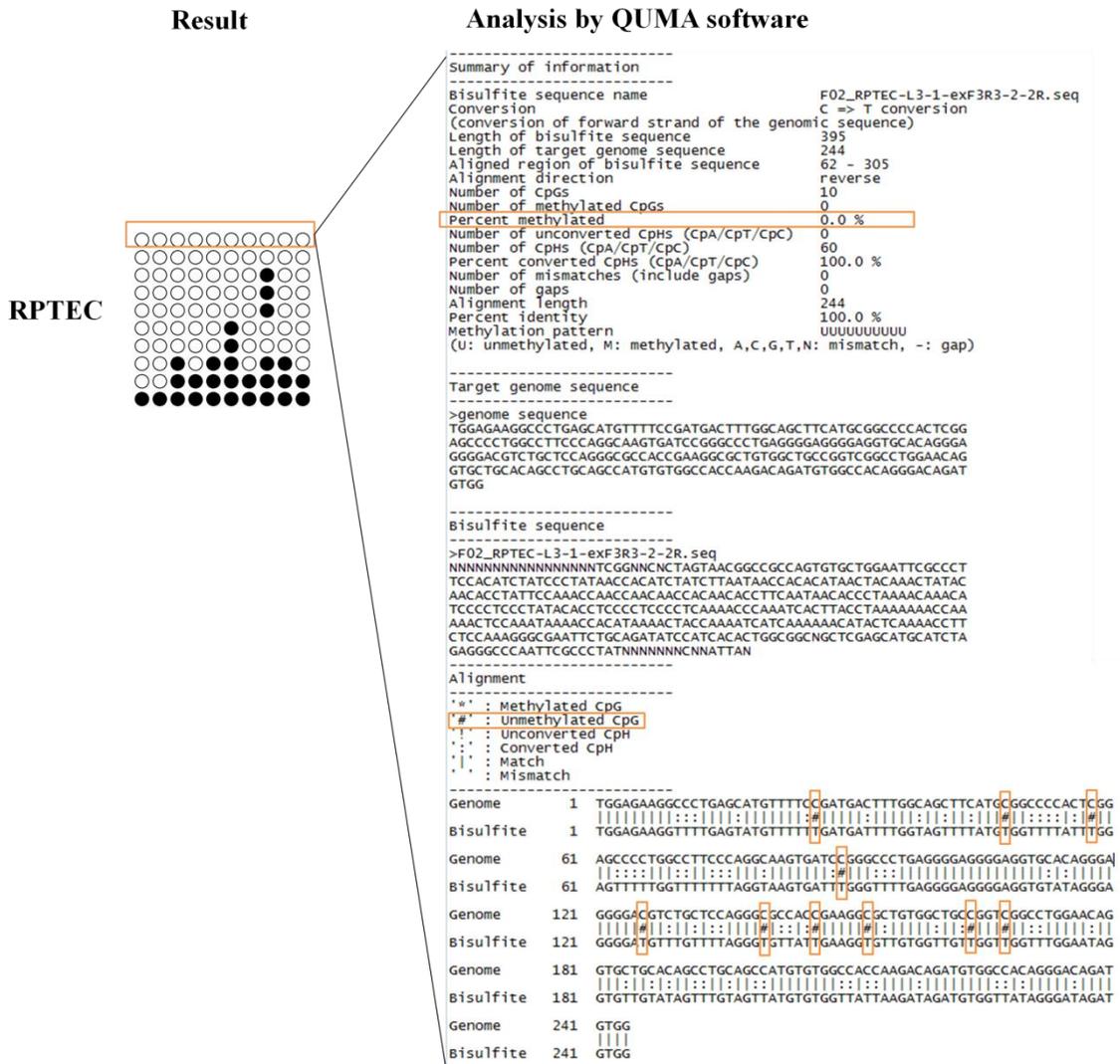
Summary of information

| Length of target genome sequence | Number of CpGs | Number of bisulfite sequences (used / excluded / total) |
|----------------------------------|----------------|---|
| 244 | 10 | 12 / 8 / 20 |

Methylation status of each CpG site

| CpG position | 26 | 48 | 58 | 89 | 126 | 140 | 146 | 152 | 164 | 168 | Total |
|--------------|---------------|-----------------|-----------------|-----------------|----------------|-----------------|----------------|-----------------|-----------------|-----------------|------------------|
| Me-CpG | 9/12 75.0% | 12/12 100.0% | 12/12 100.0% | 12/12 100.0% | 11/12 91.7% | 12/12 100.0% | 11/12 91.7% | 12/12 100.0% | 12/12 100.0% | 12/12 100.0% | 115/120 95.8% |

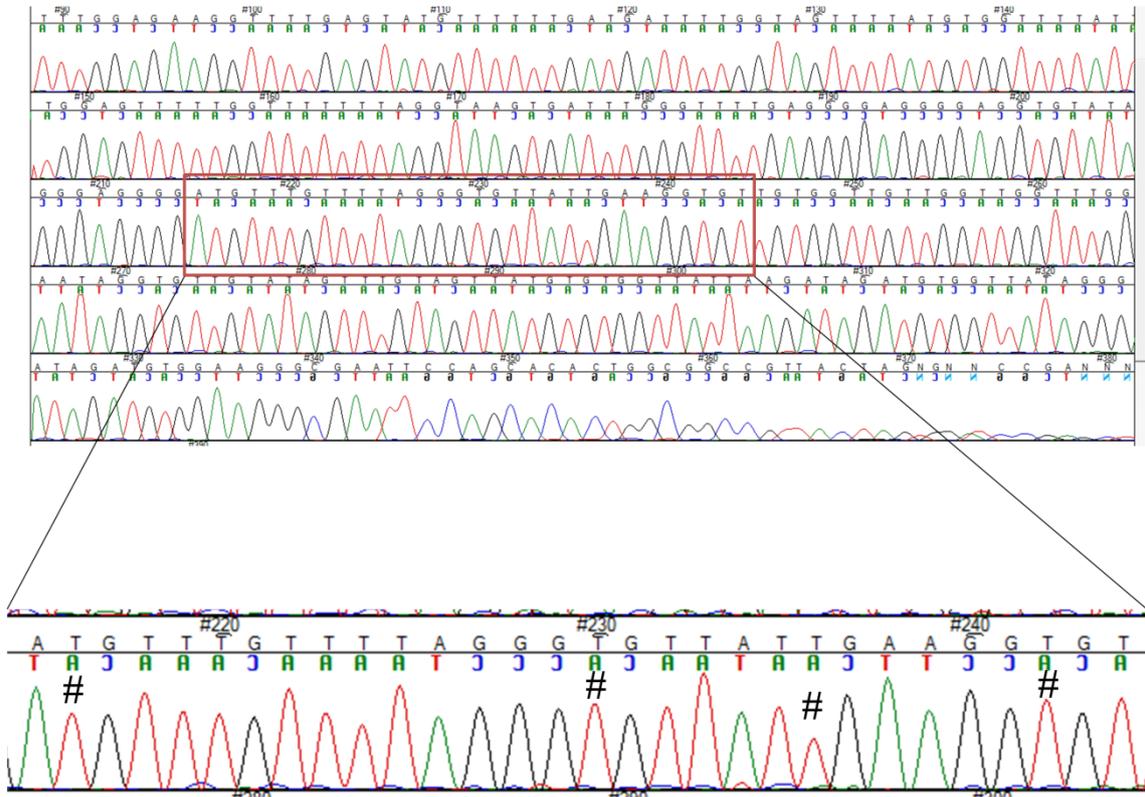
Two examples of detail analysis data:



The figure of DNA sequence after bisulfite treatment is in the next page.

DNA sequence of one clone after bisulfite treatment in RPTEC

F02_RPTEC-L3-1-exF3R3-2-2R.seq



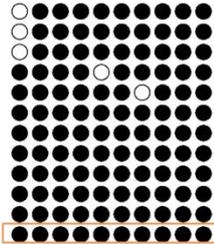
| | | | | |
|-----------|-----|------------|---------------------------|-----------------------------|
| Genome | 121 | GGGGACGTC | TGCTCCAGGGCGCCACCGAAGGCGC | TGTGGCTGCCGGTCGGCCTGGAACAG |
| Bisulfite | 121 | GGGGATGTTT | GTTTATGGTGTATTGAAGGTGT | TGTGGTTGTTGGTTGGTTTGGAAATAG |

‘#’: Unmethylated CpG

‘.’: Converted CpH

Result

ACHN



Analysis by QUMA software

```

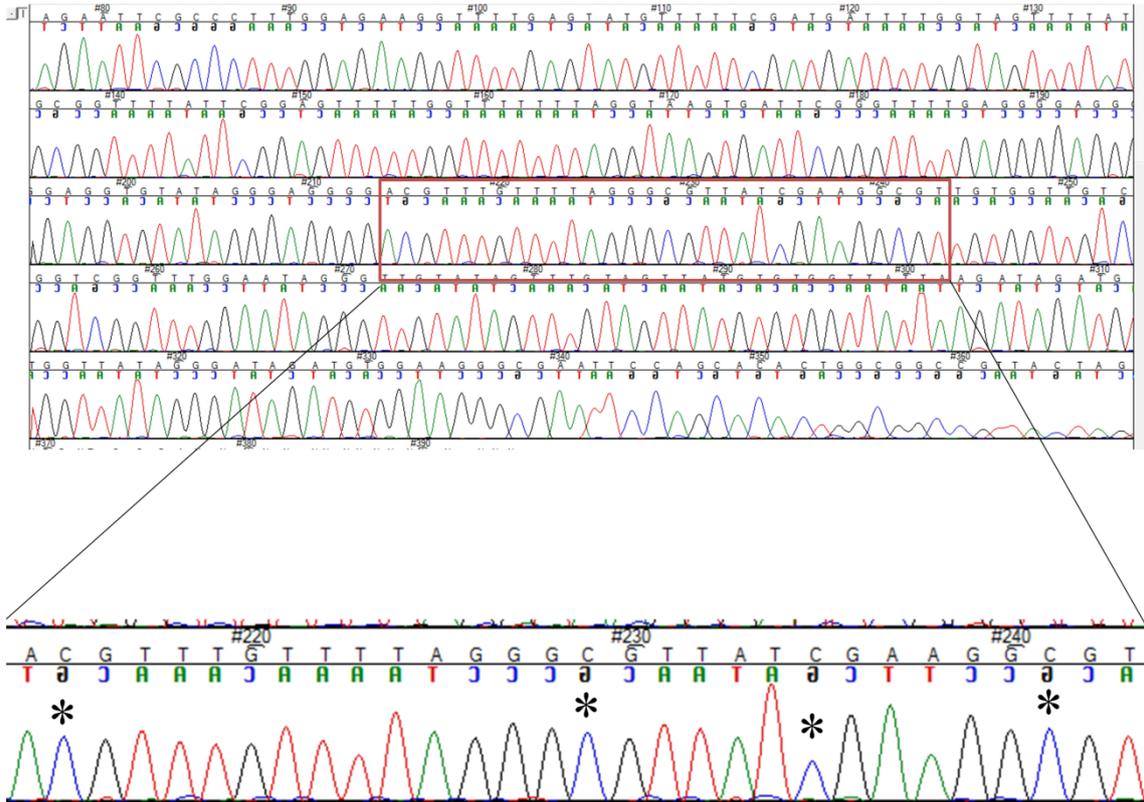
-----
Summary of information
-----
Bisulfite sequence name          E07_ACHN-exF3R3-2-7R.seq
Conversion                       C => T conversion
(conversion of forward strand of the genomic sequence)
Length of bisulfite sequence     394
Length of target genome sequence 244
Aligned region of bisulfite sequence 63 - 305
Alignment direction              reverse
Number of cpgs                  10
Number of methylated cpgs       10
Percent methylated              100.0 %
Number of unconverted cphs (CpA/CpT/CpC) 0
Number of cphs (CpA/CpT/CpC)   60
Percent converted cphs (CpA/CpT/CpC) 100.0 %
Number of mismatches (include gaps) 1
Number of gaps                  1
Alignment length                 244
Percent identity                 99.6 %
Methylation pattern             MMMMMMMMM
(U: unmethylated, M: methylated, A,C,G,T,N: mismatch, -: gap)
-----
Target genome sequence
-----
>genome sequence
TGGAGAAGGCCCTGAGCATGTTTTCCGATGACTTTGGCAGCTTCATGCGGCCCACTCGG
AGCCCTGGCCCTTCCAGGCAAGTGATCCGGGCCCTGAGGGGAGGGGAGGTGCACAGGGA
GGGGACGTCTGCCAGGGGCCACCGAAGGCCGTGTGGCTGCCGGCTGGAACAG
GTGCTGCACAGCCTGCAGCCATGTGTGGCCACCAAGACAGATGTGGCCACAGGGACAGAT
GTGG
-----
Bisulfite sequence
-----
>E07_ACHN-exF3R3-2-7R.seq
NNNNNNNNNNNNNNNTCGGANCNCTAGTAACGGCCGCCAGTGTGCTGGAATTCGCC
TTCCACATCTATCCCTATAACACATCTATCTTAATAACACACATAACTACAACTATA
CAACCTATTCCAACCGACCACCAACCAACGCCCTTCGATAACGCCATAAAACAAAG
TCCCTCCCTATACACCTCCCTCCCTCAAAACCCGAATCACTTACCATAAAAAACAA
AAACTCCGAATAAAACCGCATAAAACACCAAAATCATCGAAAAACATACTCAAAACCT
CTCAAAGGGCGAATTCGAGATATCCATCACACTGGCCGCGCTCGAGCATGCATCTA
GAGGGCCCAATTCGCCATANNNNCGTATTAA
-----
Alignment
-----
'*' : Methylated CpG
'#' : unmethylated CpG
'|' : unconverted cph
'|' : converted cph
'|' : Match
'|' : Mismatch
-----
Genome      1  TGGAGAAGGCCCTGAGCATGTTTTCCGATGACTTTGGCAGCTTCATGCGGCCCACTCGG
Bisulfite  1  TGGAGAAGGTTTTGAGTATGTTTTCCGATGATTTGGTAGTTTATCGCGTTTTATTCSG
Genome     61  AGCCCTGGCCCTTCCAGGCAAGTGATCCGGGCCCTGAGGGGAGGGGAGGTGCACAGGGA
Bisulfite  61  AGTTTTGGTTTTTTTAGGTAAGTGATTCGGGTTTTGAGGGGAGGGGAGGTGATAGGGA
Genome    121  GGGGACGTCTGCCAGGGGCCACCGAAGGCCGTGTGGCTGCCGGCTGGCCTGGAACAG
Bisulfite 121  GGGGACGTTTTGTTTTAGGGCTTATCGAAGGCCGTTGTGGTTGCTGGCTGGTTTGGAAATG
Genome    181  GTGCTGCACAGCCTGCAGCCATGTGTGGCCACCAAGACAGATGTGGCCACAGGGACAGAT
Bisulfite 181  G-GTTGTATAGTTTGTAGTTATGTGGTTATTAAGATAGATGTGGTTATAGGGATAGAT
Genome    241  GTGG
Bisulfite 240  GTGG

```

The figure of DNA sequence after bisulfite treatment is in the next page.

DNA sequence of one clone after bisulfite treatment in ACHN

E07_ACHN-exF3R3-2-7R.seq



| | | | | | |
|-----------|-----|------------|-------------|--------------|------------------------------|
| Genome | 121 | GGGGACGTC | TCCTCCAGGGC | GCCACCGAAGGC | GCTGTGGCTGCCGGTCGGCCTGGAACAG |
| Bisulfite | 121 | GGGGACGTTT | GTTTATAGGGC | GTTATCGAAGGC | GTTGTTGTCGGTCGGTTTGAATAG |

*: Methylated CpG

': Converted CpH

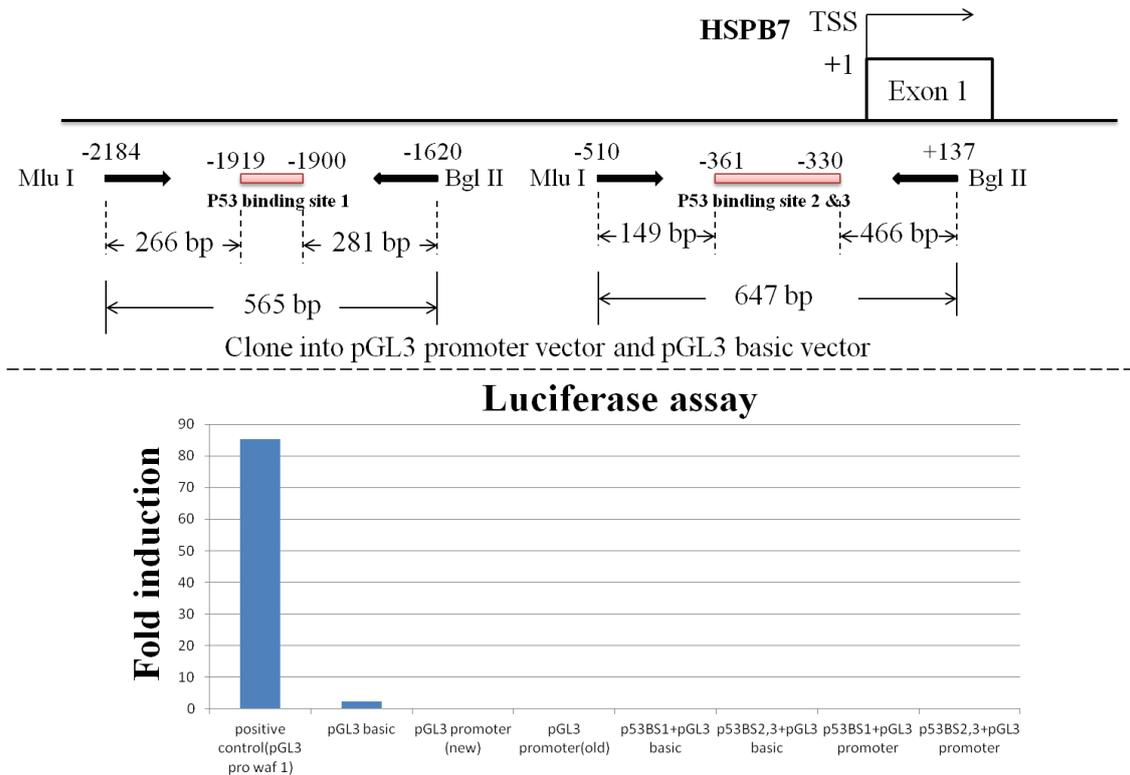


Figure3. Screening p53 binding site in HSPB7

To further investigate whether HSPB7 is directly regulated by p53, we screened three possible p53-binding sites indicated by the p53-binding site search software developed by us, but neither of these candidate sites was confirmed to be a direct p53-binding site. DNA sequence including p53 binding site 1, or p53 binding site 2&3 was cloned into both pGL3 promoter vector and pGL3 basic vector, followed by luciferase assay. Positive control: known p53 binding site in p21 (pGL3 pro waf 1); Negative control: pGL3 basic vector, pGL3 promoter vector (new) and (old). We can detect strong signal from the positive control, but we cannot detect any signal from three candidate sites.

Part II:

O-glycosylation and stabilization of HSPA5 by GALNT6

Introduction

Breast cancer is the most common cause of cancer deaths among women worldwide (2, 35). In United States, incidence and mortality due to breast cancer was 235,030 and 40,430 cases, respectively, in the year 2014 (36). Breast cancer is a heterogeneous disease with different molecular alterations which presents varied clinical behavior, response to treatment and prognosis (37, 38). Current treatments for breast cancer include surgery, chemotherapy (e.g. Doxorubicin, Cyclophosphamide, Paclitaxel), radiation therapy, hormone therapy (e.g. Tamoxifen, Toremifene, Anastrozole), and targeted therapy (e.g. Trastuzumab (Herceptin), Pertuzumab, Bevacizumab (Avastin)) (39). The selection of breast cancer for these treatments is mainly based on the stage of the disease, and the status of hormone receptors and HER2. Although these guidelines, together with significant advances in early detection and steady progress in the treatment, especially in target therapy, improve the overall survival of breast cancer (40), metastatic breast cancer responds poorly to conventional therapy. Therefore, further investigation of the molecular mechanisms and identification of specific targets are critical for developing more effective breast cancer therapies.

Glycosylation, which is the most common post-translational modification of proteins (41, 42), plays a

pivotal role in regulating the biological complexity of eukaryotes (43). It is estimated that over 50% of all human proteins are glycoproteins (41). There are mainly two types of glycosylation, N- and O-linked glycosylation, which are defined by the addition of glycans onto the side chains of Asn and Ser/Thr, respectively. Of several types of sugars (42, 44-46) can be added onto Ser/Thr, GalNAc-type (or mucin-type) O-glycosylation (described as O-glycosylation below), is one of the most abundant forms (47), which is found in more than 10% of human proteins (41, 43). O-glycosylation is initiated by up to twenty distinct UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferases (GalNAc-Ts) in the Golgi apparatus (47-50). These GalNAc-Ts are characterized by an N-terminal transmembrane domain (a stem region), a luminal catalytic domain containing a GT1 motif and Gal/GalNAc transferase motif (51), and a C-terminal ricin/lectin-like domain (52-54). They show different, while partly overlapping, substrate specificities and patterns of expression (47, 55-61). Analysis of O-glycosylation is difficult because of some factors including; (i) there is no amino acid consensus of glycosylation site, (ii) no universal deglycosylating enzyme (glycosidase) has been identified, and (iii) substrates may be densely glycosylated at multiple sites (62). The structural diversity of glycoproteins is also a big challenge prior to elucidate their biological functions. Although O-glycosylation was indicated to play important roles in protein processing (63-67), stability (68), secretion (69, 70) and function (71), its detail mechanism is largely unknown.

Aberrant O-glycosylation of protein (e.g. quantity or/and quality) is often observed in malignant cells, including breast cancer (72-74). One example is cell surface associated mucin 1 (MUC1) (75-84). MUC1 is characteristically overexpressed in breast adenocarcinomas as well as adenocarcinomas of other organs (e.g. ovary, lung, pancreas, and prostate). The O-glycosylation of MUC1 VNTRs produced by cancer cells is altered in density and core structures and elongation/capping of the O-glycans. The cause for aberrant O-glycosylation in breast cancer is unclear, but emerge evidence showed that the expression of

the GalNAc-Ts, which control the initiation of O-glycosylation and are highly regulated in normal cells, are markedly altered in cancer (85-88). Recent studies also highlight Golgi-to-ER (endoplasmic reticulum) relocation of some GalNAc-Ts promotes cancer cell invasiveness (89, 90). However, the mechanisms underlying altered O-glycosylation in breast cancer cells are incompletely understood.

We previously identified GALNT6 (polypeptide N-acetylgalactosaminyl transferase 6, GalNAc-T6), a member of GalNAc-Ts, is a novel molecular target for breast cancer based on three evidences (88, 91). First, GALNT6 was highly expressed in breast cancer cells while not in normal tissues; second, it showed oncogenic function which contributed to breast cancer cells transformation or metastasis by glycosylation of target proteins such as MUC1 and fibronectin; third, knockdown of GALNT6 in breast cancer cells significantly suppress cell proliferation. Results from other researchers also indicated that GALNT6 is a potential new marker in breast cancer (86, 87). However, the knowledge about how GALNT6 promotes breast carcinogenesis is still very limited. Since GALNT6 is a GalNAc-type glycosyltransferase, multiple substrates may exist and function in different pathways. To further study these substrates is especially necessary, not only because it can provide evidence for application of GALNT6 inhibitors into clinic, but also can identify more good targets for breast cancer. These substrates which are O-glycosylated by high level of GALNT6 may be important hallmarks that differentiate breast cancer cells from normal cells.

To screen the substrates of GALNT6, we identified specific proteins only O-glycosylated in wild type GALNT6 stable cell but not in the mutant one by *Vicia villosa* agglutinin (VVA) lectin pull-down assay followed by mass spectrometry (MS) analysis. Because of specific binding activity to certain glycan structure, lectins are frequently used to detect and isolate glycosylated proteins (92). For example, VVA lectin has high affinity to GalNAc molecule that GALNT6 transfers to its protein substrates (93). And MS

analysis is a classical tool for identification of proteins and their glycosylation sites (94). Here we report that heat shock protein 5 (HSPA5) is a novel substrate of GALNT6.

HSPA5, also known as GRP78/BiP, is highly expressed in cancers (95-101) and is required for tumor formation and progression in vitro and in vivo. HSPA5 involved in many cellular processes including as a master regulator for ER stress (102-104), which activates the Unfolded Protein Response (UPR) to alleviate this stress and restore ER homeostasis. Recent studies also highlighted that HSPA5 controls cross-talk between apoptosis and autophagy (105-107). HSPA5 exhibits oncogenic activities by promoting tumor proliferation, survival, angiogenesis, metastasis, and drug resistance (108-110). In vivo studies with heterozygous or homozygous deletion of HSPA5 suppressed tumorigenesis (108-112). In addition, HSPA5 was also indicated to play a critical role in regulating cancer initiating cells (CIC) proliferation and survival, as well as in stem cell biology (113-115). However, whether O-glycosylation affects the function of HSPA5 and promotes breast carcinogenesis- has remained unanswered.

In this study, that GALNT6 glycosylates HSPA5 was confirmed and the role of this modification to the stability of HSPA5 was examined. Here we report for the first time that GALNT6 O-glycosylates and stabilizes HSPA5 protein, which may prolong the oncogenic effects of HSPA5 in breast cancer. Meanwhile, overexpression of HSPA5 can drive Golgi-to-ER relocation of GALNT6. The result may trigger O-glycosylation of multiple substrates of GALNT6 at ER. Our study revealed a novel mechanism of how GALNT6 and HSPA5 cooperate together to promote mammary carcinogenesis. And the O-glycosylated form of HSPA5 may be a good target for breast cancer therapy.

Materials and Methods

Cell lines.

Human breast cancer cell lines MCF7 and MDA-MB-435s and HeLa cell were purchased from American Type Culture Collection (ATCC; Rockville, MD, USA). HeLa-Mock and HeLa-GALNT6 (wild type (WT) and mutant (H271D)) stably expressed cells (monoclonal populations) were established as described previously (88). Briefly, Mock (no insert) or pCAGGS-GALNT6 (WT and H271D) HA expression vectors were transfected into HeLa cells by using FuGENE 6 Transfection Reagent (Roche). Positive clones were selected by two weeks with culture medium containing 0.8 mg/ml of G418 (Geneticin, Life Technologies) followed by limiting dilution. Finally, we isolated individual clones of mock, WT, and H271D.

MDA-MB-435s-Mock and GALNT6 WT polyclone stable cells were newly made. Mock (no insert) or pCAGGS-GALNT6-HA expression vectors were transfected into MDA-MB-435s cells by using TransIT-BrCa Transfection Reagent (Mirus). 48 h after transfection, cells were selected under incubation with culture medium containing 0.8 mg/ml of G418 (Geneticin, Life Technologies). Two weeks later, resistant clones were pooled to generate a polyclonal population expressed average levels of the protein.

All cell lines were grown in monolayers in appropriate media recommended by suppliers: Minimum Essential Medium (11095-080, Life Technologies) for HeLa and MCF7, the later need additional insulin (I0516, Sigma-Aldrich); Leibovitz's L-15 Medium (11415-064, Life Technologies) for MDA-MB-435S; in addition, cells were supplemented with 10% fetal bovine serum, and 1% penicillin-streptomycin-amphotericin B. HeLa stable cells and MDA-MB-435s stable cells were cultured in the culture medium containing 0.8 mg/ml of G418 (Geneticin, Life Technologies).

Transfection.

Cells were seeded in 6-well plate or 10cm dish. The next day, transfection was performed by using TransIT-BrCa Transfection Reagent (Mirus) or Lipofectamine 2000 Reagent (Life Technologies) or FuGENE 6 Transfection Reagent (Roche) according to the manufacturer's protocols. 48 h later, the cells were harvested for further study.

VVA-lectin western blot and pull down assay.

To detect the Tn glycan that is conjugated with GalNAc monosaccharide, we did lectin western blot using Biotinylated Vicia Villosa Lectin (VVL, VVA) (1:1000, Vector Laboratories) and Streptavidin-HRP (1:10,000, Thermo Scientific). Briefly, we did SDS-PAGE with whole cell lysates and migrated proteins were transferred onto a nitrocellulose membrane (GE Healthcare). Subsequently, the membrane was blocked with 5% bovine serum albumin (BSA) for overnight at 4 °C, and then incubated with biotinylated VVA lectin for 1 h at room temperature. After three times washing with TBST for 10 min, the membrane was incubated with Streptavidin-HRP for 1 h at room temperature, and washed again before protein bands were visualized by the ECL detection reagents (GE Healthcare). Similarly, proteins bound to Agarose Vicia Villosa Lectin (VVL, VVA) (Vector Laboratories) or first bound to biotinylated VVA lectin followed by Streptavidin agarose (Invitrogen) were pulled down according to manufacturer's protocol. For screening subtracts of GALNT6, the isolated proteins were visualized by the SilverQuest Silver Staining Kit (Invitrogen). Protein bands that were specifically observed in the GALNT6-WT lane were excised with a clean, sharp scalpel and the extracted proteins were applied for PMF (Peptide Mass Fingerprint) analysis using MALDI-TOF MS (Matrix Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry).

Cloning and mutagenesis.

We performed all the PCR here by using KOD Plus DNA polymerase (Toyobo, Osaka, Japan).

To construct the wild-type HSPA5 expression vectors, the entire coding sequence of HSPA5 cDNA was amplified by the PCR. Primer sets were in Table 1. The PCR product was inserted into the NotI and XhoI sites of pCAGGS-3xFlag expression vector by using the digest enzymes (NotI , #FD0694 and XhoI, #FD0594, Thermo Scientific) followed by the ligation enzyme (T4 DNA ligase, M0202M, New England Biolabs). The plasmid was transformed into DH5 alpha Competent E coli Cells (18265-017, Invitrogen) by heat shock and spread on a LB plate supplemented with 50 µg/ml ampicillin. The plate was incubated at 37°C for 16-18 h. Single clone was picked up and confirmed by colony PCR or enzyme digestion, then expanded in LB based liquid medium supplemented with 50 µg/ml ampicillin at 37°C for 16-18 h. The plasmid DNA was extracted by using QIAprep Spin Miniprep Kit (27106, QIAGEN) (for enzyme digestion) or QIAfilter plasmid Maxi kits (12263, QIAGEN) (for transfection).

To generate HSPA5 alanine substitution mutants (T85A, T151A, T166A, T184A, T203A and T481A) that correspond to the candidate glycosylation sites, we performed two-step mutagenesis PCR with the help of four primers: HSPA5 forward and reverse (the same primer set as HSPA5 wild type expression vector), mutagenic primer forward and reverse (share same sequence, with mutated nucleotide in the middle of the primer). (Table 1)

The first round of PCR is to generate two fragments of HSPA5 by using HSPA5 wild type plasmid DNA as a template with two sets of primers. Fragment-1 used HSPA5 forward and mutagenic primer reverse, and Fragment-2 used mutagenic primer forward and HSPA5 reverse. After confirmation by electrophoresis and gel purification, these two fragments were used for the second round of PCR. First, everything for

PCR reaction except enzyme and primer set was added into a tube and then touchdown PCR (94°C, 3 min; from 94°C to 37°C, reduced 1°C/30sec; 37°C, 15min (add KOD-Plus DNA polymerase); 68°C, 3 min (add HSPA5 forward and reverse primer set)) was performed, followed by ordinary PCR (94°C, 2 min; 34 cycles of amplification (94°C, 15 sec; 55°C, 30 sec; 68°C, 1.5 min)). After confirmation by electrophoresis and gel purification, the PCR products were inserted into the NotI and XhoI sites of pCAGGS-3xFlag expression vector as described above.

The HSPA5-mt5 mutant including 5 alanine-substituted sites (T85A, T151A, T166A, T184A, and T203A) was generated by sequential addition of each mutant in the expression vector.

For construction of HSPA5 fragments (HSPA5-1-280, HSPA5-125-500 and HSPA5-281-654), we performed PCR by using HSPA5 wild type plasmid DNA as a temple. The primer sets were in Table 1. The PCR products were inserted into the NotI and XhoI sites of pCAGGS-3xFlag expression vector as described above.

All of the constructs were confirmed by DNA sequencing in ABI3500XL (Life Technologies) after sequencing reaction by using BigDye Terminator v3.1 Cycle Sequencing Kit (Life Technologies) and then purification by BigDye XTerminator purification kit (4376487, Life Technologies). The protein expression of these plasmids was also confirmed by western blot.

Table 1. Primers

| Primer name | Sequence (5' to 3') |
|--|---|
| For wild type HSPA5 expression vector | |
| HSPA5-F-NotI | ATTTG <u>CGGCCG</u> CATGAAGCTCTCCCTGGTGG |
| HSPA5-R-XhoI_nonstop | CCGCT <u>CGAGCA</u> ACTCATCTTTTTCTGCTGTATCC |
| Mutagenic primer for HSPA5 alanine substitution mutants | |

| | |
|-----------------------------------|---|
| HSPA5_T85A-F | AAGAACCAGCTCGCCTCCAACCCCGAGAAC |
| HSPA5_T85A-R | GTTCTCGGGGTTGGAGGGCGAGCTGGTTCTT |
| HSPA5_T151A-F | ATTTCTGCCATGGTTCTCGCTAAAATGAAAGAAACC |
| HSPA5_T151A-R | GGTTTCTTTCATTTTAGCGAGAACCATGGCAGAAAT |
| HSPA5_T166A-F | ATTTGGGAAAGAAGGTTGCCATGCAGTTGTTACTG |
| HSPA5_T166A-R | CAGTAACAACCTGCATGGGCAACCTTCTTTCCCAAAT |
| HSPA5_T184A-F | GATGCCCAACGCCAAGCAGCCAAAGACGCTGGAAC TATTG |
| HSPA5_T184A-R | CAATAGTTCAGCGTCTTTGGCTGCTTGGCGTTGGGCATC |
| HSPA5_T203A-F | CATCAACGAGCCTGCCGCAGCTGCTATTGC |
| HSPA5_T203A-R | GCAATAGCAGCTGCGGCAGGCTCGTTGATG |
| HSPA5_T481A-F | CATCTTCTGGGTGCATTTGATCTGACTGG |
| HSPA5_T481A-R | CCAGTCAGATCAAATGCACCCAGAAGATG |
| For HSPA5-1-280 | |
| HSPA5-F-NotI | ATTTGCGGCCCGCCATGAAGCTCTCCCTGGTGG |
| HSPA5-R-280-XhoI | CCGCTCGAGTTTCTGACATCTTTGCCCG |
| For HSPA5-125-500 | |
| HSPA5-F-NotI-125 | ATTTGCGGCCCGCCATGAAACCATACATTCAAGTTGATATTGG |
| HSPA5-R-500-XhoI | CCGCTCGAGGGTGACTTCAATCTGTGGG |
| For HSPA5-281-654 | |
| HSPA5-F-NotI-281 | ATTTGCGGCCCGCCATGGACAATAGAGCTGTGCAGAAAC |
| HSPA5-R-XhoI_nonstop | CCGCTCGAGCAACTCATCTTTTTCTGCTGTATCC |
| For HSPA5 sequencing | |
| HSPA5-CDS-F1 | AGATCATCGCCAACGATCAG |
| HSPA5-CDS-R1 | CCACAACCTTCGAAGACACCA |
| HSPA5-CDS-F2 | TAAGAGGGAGGGGGAGAAGA |
| HSPA5-CDS-R2 | TGATCACCAGAGAGCACACC |
| HSPA5-CDS-F3 | TAGCGTATGGTGCTGCTGTC |
| HSPA5-CDS-R3 | CAGCTTTTTCCATGGTCTCC |
| pCAGGS vector primers | |
| For colony PCR | |
| pCAGGS F2(1520) | GTCCCCTTCTCCATCTCCAG |
| pCAGGS R(1824) | TATTTGTGAGCCAGGGCATT |
| For sequencing PCR product | |
| pCAGGS F1(1635) | GCTAACCATGTTTCATGCCTTC |
| pCAGGS R(1824) | TATTTGTGAGCCAGGGCATT |
| For sequencing plasmid | |
| pCAGGS VF(1637) | TAACCATGTTTCATGCCTTCTCT |

pCAGGS VR(1868)

GGCAGAGGGAAAAAGATCTCAG

(Underlines indicate recognition sites of restriction enzymes or nucleotides that were replaced from the wild-type of HSPA5 gene)

Immunoprecipitation (IP).

Cell extracts from HeLa stable cells with or without transfection were prepared by adding CellLytic M reagent (C2978, Sigma-Aldrich) with 1% Calbiochem Protease Inhibitor Cocktail Set III, EDTA-Free (#539134, Calbiochem) according to the manufacturer's protocols. Extracts were precleared by incubation with rec-Protein A-Sepharose 4B Conjugate (40µl/sample) (10-1141, Invitrogen) and rabbit IgG (2µg/sample) (Santa Cruz) or rec-Protein G-Sepharose 4B Conjugate (40µl/sample) (10-1242, Invitrogen) and mouse IgG (2µg/sample) (Santa Cruz) at 4 °C for 1 h. For pulling down HSPA5 or Flag-HSPA5s, precleared cell extracts were then incubated with anti-HSPA5 (2µg/sample, 11587-1-AP, Proteintech) or anti-Flag M2 monoclonal antibody (2µg/sample, F3165, Sigma-Aldrich) at 4 °C for overnight followed by rec-Protein A-Sepharose 4B Conjugate (40µl/sample) or rec-Protein G-Sepharose 4B Conjugate (40µl/sample) at 4 °C for 2 h, respectively. For pulling down HA-GALNT6, precleared cell extracts were incubated with monoclonal anti-HA-Agarose (40µl/sample) (A2095, Sigma-Aldrich) at 4 °C for overnight. Corresponding Protease Inhibitor was together added into the solution for each time. The beads were then spun down and washed 4 times with 1 ml of CellLytic M reagent (C2978, Sigma-Aldrich). Immunoprecipitated proteins were released from the beads by boiling in sample buffer for 2 min or adding elution buffer.

Silver staining

Silver staining was performed by using the SilverQuest™ Silver Staining Kit (Life Technologies) according to the manufacturer's protocols. Briefly, samples were load onto 4-20 % SDS-PAGE (Bio-Rad)

for electrophoresis. The gel was fixed with 40 % ethanol, 10% acetic acid for 20 min followed by 30 % ethanol for 10 min. The gel was put into sensitizer solution for 10 min and then in 30 % ethanol and milliQ water for 10 min, respectively. For staining, the gel was immersed into staining solution for 15min. After washed with milliQ water for 1 min, the gel was treated with develop solution for 4-8 min. The reaction was stopped by stop solution when clear bands were observed.

Western blot analysis.

To prepare whole cell extracts, cells were collected and lysed in RIPA buffer or CellLytic M reagent (C2978, Sigma-Aldrich) with 1% Calbiochem Protease Inhibitor Cocktail Set III, EDTA-Free (#539134, Calbiochem) according to the manufacturer's protocols. The total cell lysates were incubated on ice for 30 minutes and centrifuged at 14,000 rpm for 15 min to collect only supernatant. The amount of total protein was estimated by protein assay kit (Bio-Rad), and then proteins were mixed with Lane Marker Sample Buffers (39000, Thermo Scientific) and boiled before loading into SDS-PAGE gels. After electrophoresis, total proteins were transferred to nitrocellulose membranes (GE Healthcare). The membranes including whole proteins were blocked by blocking solution and incubated with primary antibodies for overnight at 4°C. The next day, the membranes were washed followed by incubation in secondary antibodies for 1 h at room temperature. Protein bands were visualized by ECL detection reagents (GE Healthcare). If necessary, membranes were stripped after exposure with stripping buffer (#46430, Thermo Scientific) for detection by other primary antibodies in the same membranes. Bands were quantified by the Image J software if needed. The primary antibodies including in this study were rabbit anti-human HSPA5 polyclonal antibody (1:1000, sc-13968, Santa Cruz), or rabbit anti-human GALNT6 polyclonal antibody (1:1000, HPA011762, Sigma-Aldrich), or anti-Flag M2 monoclonal antibody (1:1000, F3165, Sigma-Aldrich), or anti-HA monoclonal antibody (1:1000, 1583816, Roche), or anti- β -Actin monoclonal antibody (1:10,000,

A5441, Sigma-Aldrich). The secondary antibodies were goat anti-rabbit/ anti-rat/ anti-mouse IgG-HRP secondary antibody (1:10,000/1:30,000, Santa Cruz).

Immunocytochemistry (ICC).

Cells were seeded on Lab-Tek II chamber slide system (Nalge Nunc International). After transfection for 48 h, the cells were fixed with 4% paraformaldehyde in PBS for 10 min and permeabilized with 0.2% Triton X-100 in PBS for 5 min at room temperature. Cells were covered with blocking solution (3% BSA in PBS contained 0.2% Triton X-100) for 60 min at room temperature. Then the cells were incubated with anti-Flag M2 monoclonal antibody (1:500, F3165, Sigma-Aldrich), and anti-HA monoclonal antibody (1:500, 1583816, Roche), for overnight at 4°C, following an Alexa Fluor 488 Goat Anti-Mouse IgG (H+L) Antibody (1:1,000, A11001, Life Technologies) and Alexa Fluor 594 Goat Anti-Rat IgG Antibody (1:1,000, A11007, Life Technologies) for 1 h at room temperature. PBS or 0.2% Triton X-100 in PBS was used for washing after each step. Then cells were stained with DAPI (H-1200, Vector Laboratories) and viewed with a Leica TCS SP5 Confocal Laser Scanning Microscope.

In vitro glycosylation assay.

As a substrate, the pCAGGS-HSPA5-PRGST plasmid was expressed in HEK293 cell, and GST-tagged HSPA5 protein was pulled down by Glutathione Sepharose 4B agarose (GE Healthcare). In parallel, pQCXIPG-GALNT6-6xHis plasmids (WT and H271D) were expressed in HEK293 cell, and His-tagged recombinant GALNT6 proteins (WT and H271D) were purified by Ni-NTA agarose (Qiagen) as previously described (88). Then recombinant HSPA5 protein and GALNT6 WT or H271D were incubated in 25 mM Tris-HCl (pH 7.4), 10 mM MnCl₂, 50 μM UDP-GalNAc at 37°C for 2 h. For the confirmation of GalNAc-conjugation on HSPA5, we did VVA lectin western blot as described above.

Identification of O-glycosylation sites in HSPA5.

For identification of glycosylation sites, HeLa-GALNT6 stable cells (WT and H271D) were transfected with pCAGGS-HSPA5-3xFlag expression vector and collected after 48 h incubation. Cells were lysed with 1% NP-40 lysis buffer and Flag-tagged HSPA5 protein was immunoprecipitated with anti-Flag monoclonal antibody and Protein A agarose (Invitrogen). After five times washing with the lysis buffer, immunocomplexes were loaded in a SDS-PAGE gel and protein bands were visualized by CBB staining (Bio-Rad). The bands which indicated HSPA5 were excised and reduced in 10 mM Tris (2-carboxyethyl) phosphine (Sigma-Aldrich) with 50 mM ammonium bicarbonate (Sigma-Aldrich) at 37°C for 30 min and alkylated in 50 mM iodoacetic acid (Sigma-Aldrich) with 50 mM ammonium bicarbonate in the dark at 25 °C for 45 min. Trypsin Gold (Promega) solution was added with the enzyme to protein ratio at 1/50 (w/w) and incubated at 37 °C for 12 h. The resulting peptides were extracted from gel fragments and separated on a 0.1 × 200 mm homemade C₁₈ column using 45 min linear gradient from 2 to 35% acetonitrile in 0.1% formic acid with flow rate at 200 nl/min. The eluting peptides were analyzed with QSTAR Elite QqTOF mass spectrometer (AB Sciex) in the smart information-dependent acquisition (SIDA) mode of the Analyst QS software 2.0 (AB Sciex). The acquired MS and MS/MS peak lists were searched against SwissProt database ver. 2011_6 (20,239 Homo sapiens sequences) with in-house Mascot server ver.2.3.01 (Matrix Science). The search parameters were as follows; fixed modifications = carboxymethylation (cysteine), variable modifications = HexNAc (serine or threonine) and oxidation (methionine), peptide mass tolerance = 50 ppm, fragment mass tolerance = 0.1 Da, max missed cleavages = 2. We accepted the peptides with Expectation value less than 0.05 as the positive identification in Mascot Database search.

Cycloheximide treatment.

HeLa stable cells or breast cancer cells were transfected with HSPA5 WT or mutant expression vectors.

After 48 h incubation, cells were treated with 50µg/ml cycloheximide (#2112, Cell Signaling) and incubated for various time points as indicated in the text. Or HeLa stable cells were seeded and 24 h later directly treated with 50µg/ml cycloheximide and incubated for 24 h or 48 h or 96 h respectively. Cells were collected after incubation and performed western blot analysis as described above.

Results

HSPA5 is a novel candidate substrate of GALNT6

In the previous study, VVA-lectin western blot analysis indicated the presence of multiple O-glycan substrates of GALNT6 because such glycosylated bands were diminished by knockdown of GALNT6 (91). To identify O-glycan substrates of GALNT6, we performed pull-down assay using biotin-conjugated VVA lectin (specific to GalNAc-Ser/Thr, called Tn-antigen) and streptavidin-conjugated agarose (Fig. 1A). As shown in figure 1B, we detected several differential bands in HeLa-GALNT6 wild-type (WT) compared with mock control or GALNT6 enzyme-dead mutant (H271D) stable cells. We excised these bands and performed mass spectrometry analysis to identify proteins. HSPA5 was one of the novel candidates (Table 2).

HSPA5 is a GALNT6-interacting protein

To validate the mass spectrometry result, we performed immunoprecipitation-Western blot analysis and found that GALNT6 coprecipitated with HSPA5 (Fig. 2A) in HeLa GALNT6 WT stable cells. Reciprocally, HSPA5 could also be pulled down when an antibody against HA tag (for endogenous GALNT6-HA in stable cells. Parental HeLa is a GALNT6 (-) cell line.) was used for immunoprecipitation

(Fig. 2B). This result strongly suggests that endogenous GALNT6 and HSPA5 physically interact. Although GALNT6 mutant, H271D can also bind to HSPA5, the binding ability is much weaker when compared with wild type protein.

To map the binding domain on HSPA5, we expressed Flag-tagged wild-type HSPA5 and HSPA5 fragments with only ATPase domain (HSPA5-1-280) or ATPase and peptide binding domain (HSPA5-125-500) or only peptide binding domain (HSPA5-281-654) in HeLa GALNT6 WT stable cells (Fig. 2 C). We found that GALNT6 coprecipitated with wild-type HSPA5, HSPA5-1-280 and HSPA5-125-500, all of which have ATPase domain, but GALNT6 binding to HSPA5-281-654 without ATPase domain was much reduced (Fig. 2D). This data indicates that the ATPase domain of HSPA5 is important for their binding.

Overexpression of HSPA5 drives Golgi-to-ER relocation of GALNT6

To further confirm GALNT6 and HSPA5 are in the same complex, immunocytochemistry (ICC) analysis was performed in HeLa stable cells (Fig. 3). We observed wild type GALNT6 locates in Golgi, and high expression level of exogenous HSPA5 (not low level) drives Golgi-to-ER relocation of GALNT6 and they co-localize at ER in HeLa GALNT6 WT stable cells. Mock stable cells served as negative control to exclude cross reactivity of the antibodies.

GALNT6 glycosylates HSPA5

Then, we identified GALNT6 glycosylates HSPA5 (Fig. 4). By using recombinant HSPA5 and GALNT6 wild type and mutant protein in vitro assay (Fig. 4A), we found HSPA5 can only be glycosylated in the present of GALNT6 wild type protein but not GALNT6 mutant, H271D. During the process, only wild type GALNT6 itself can also be auto-glycosylated. Moreover, in cell based system, we pulled down exogenous (Fig. 4B) or endogenous HSPA5 (Fig. 4C and D), and detected its glycosylation

level. We found HSPA5 showed much higher level of O-glycosylation in HeLa GALNT6 WT stable cells, when compared with mock and H271D stable cells, indicating that GALNT6 can O-glycosylate HSPA5, although GALNT6 is not the only one GalNAc-type O-glycosyltransferase for HSPA5, since we observed HSPA5 can also be O-glycosylated in GALNT6(-) cells (Fig. 4D).

Identification of GALNT6-induced O-glycosylation sites in HSPA5

Next, we investigated GALNT6-induced O-glycosylation sites in HSPA5. Exogenous HSPA5 in HeLa GALNT6 WT or H271D stable cells, was immunoprecipitated with anti-flag antibody followed by CBB staining. The bands were cut and submitted for mass spectrometry analysis. Six candidate O-glycosylation sites in HSPA5 were identified and four of them were located in the ATPase domain of HSPA5 (Table 3).

GALNT6 stabilizes HSPA5 through O-glycosylation

Further study showed that GALNT6 stabilizes HSPA5 protein (Fig. 5). By treatment with CHX (cycloheximide), a protein synthesis inhibitor, at different time point, we can study the stability of HSPA5. Figure 5A showed endogenous HSPA5 is more stable in HeLa GALNT6 WT stable cells when compared with mock and H271D stable cells. We confirmed the same result in another stable cell lines, MDA-MB-435S-Mock and GALNT6 WT stable cell line (Parental MDA-MB-435S is a GALNT6 (-) cell line). Figure 5B showed GALNT6 prolongs the half time of exogenous HSPA5 in MDA-MB-435S-GALNT6 WT stable cells when compared with mock stable cells. This data suggests that high level of GALNT6 makes HSPA5 protein more stable.

To analyze whether the stability of HSPA5 depends on its O-glycosylation, we generated six HSPA5 single amino acid substituted mutants (T85A, T151A, T166A, T184A, T203A and T481A) and a mutant

contained 5 substituted sites (T85A, T151A, T166A, T184A, and T203A) (Fig. 5C). We transfected HSPA5 wild-type and all mutants into HeLa-Mock, GALNT6 WT and H271D stable cells, and then performed western blot analysis after treatment of CHX at different time point (data not shown). We found T184A is one of the most important glycosylation site mutations which affects the stability of HSPA5 protein in HeLa stable cells (Fig. 5D). T184A is less stable than wild type protein in all types of HeLa stable cells, suggesting that O-glycosylation at T184 site is important to stabilize HSPA5 protein no matter by GALNT6 or other unknown GalNAc-type O-glycosyltransferase in GALNT6 (-) cells. We also observed that the half time of HSPA5 WT and mutant protein is less stable in GALNT6 (-) cells when compared with positive cells. These results confirmed again GALNT6 stabilizes HSPA5 protein. They also indicated that GALNT6 may glycosylate HSPA5 at multiple sites, because T184A was more stable in GALNT6 (+) cells. We confirmed similar result in MCF7 breast cancer cells with endogenous GALNT6, and MDA-MB 435s without GALNT6 (Fig. 5E). Although we need more evidences to figure out whether mutation itself changed the structure of HSPA5 and then affected its stability, we found not every glycosylation site mutation showed similar effect as T184A based on our preliminary results (data not shown). These mutations near T184 site can serve as negative controls. Taken together, we draw a conclusion that O-glycosylation at T184 site and high level of GALNT6 are important to stabilize HSPA5 protein in the cell systems we tested.

Discussion

Both GALNT6 and HSPA5 were found to play important roles in breast carcinogenesis (86-88, 91, 96, 97, 110, 114), however the relationship between them is unknown. In this study, we report for the first

time that HSPA5 is a novel substrate of GALNT6. GALNT6 O-glycosylates and stabilizes HSPA5 protein, and the ATPase domain of HSPA5 is important for their binding and glycosylation. Aberrant O-glycosylated HSPA5 with higher stability may prolong the oncogenic effects of HSPA5 in breast cancer. Meanwhile, overexpression of HSPA5 can drive Golgi-to-ER relocation of GALNT6. The result may trigger O-glycosylation of multiple substrates of GALNT6 at ER. Our study revealed a novel mechanism of how GALNT6 and HSPA5 cooperate together to promote mammary carcinogenesis.

The first finding of our study is HSPA5 as a novel substrate of GALNT6. Glycomics is a hot spot for cancer research and targeting glycoproteins, for example, MUC1, already showed promising anti-cancer results (116). Since aberrant upregulation of GALNT6 and HSPA5 together in breast cancer will cause increased amount of O-glycosylated HSPA5, this glycoprotein may be an important hallmark that differentiates breast cancer cells from normal cells, and therefore serves as a good target for breast cancer therapy. Although potential O-glycosylation of HSPA5 at T203 and T643 were reported (117), there is no data to confirm whether these candidate O-glycosylation sites are true ones. And there is no report to show which enzyme is responsible, and what is physiological function of the O-glycosylation. To our knowledge, our study is the first report that GALNT6 O-glycosylates HSPA5 and T184 is one of the potential O-glycosylation sites.

Second, we confirmed the ATPase domain of HSPA5 is important for its binding and glycosylation. The interaction of HSPA5 with GALNT6 is significantly reduced if HSPA5 loses its ATPase domain. This result is consistent with previous study model that HSPA5 interacts with its substrate proteins in an ATP-dependent manner (118, 119). In addition, we found four potential O-glycosylation sites are located in the ATPase domain of HSPA5, suggesting that GALNT6 may directly bind to and O-glycosylate HSPA5 at the ATPase domain.

Next, we found GALNT6 stabilizes HSPA5 through O-glycosylation, and mutation at one potential O-glycosylation site in the ATPase domain, T184A, affects the stability of HSPA5. Previous studies reported that HSPA5 is degraded through the ubiquitin-proteasome pathway (120, 121), and K185 is a potential ubiquitination site (122). It is possible that O-glycosylation of HSPA5 at T184 increases steric hindrance and inhibits ubiquitination at K185, and thus prevents HSPA5 from degradation. HSPA5 was known to be required for tumor formation and progression. Higher stability of HSPA5 O-glycosylated by GALNT6 will therefore prolong its oncogenic effects in breast cancer. Since the ATPase domain is important for the function of HSPA5, O-glycosylation in this domain may provide more biological significances in addition to the stabilization of HSPA5. We already knew that some O-glycosylation sites of HSPA5 were also potential phosphorylation sites, such as T85 and T203 (123-125). Further study was needed to figure out all of these biological significances.

Last, HSPA5 also influences GALNT6. High expression level of HSPA5 can drive Golgi-to-ER relocation of GALNT6. Previous studies showed that some growth factor receptors (e.g. EGFR, PDGFR) and Src activation can cause Golgi-to-ER relocation of GalNAc-Ts (e.g. GALNT1 and GALNT2), which promotes cancer cell invasiveness (89, 90). In our study, overexpression of HSPA5 was confirmed to induce Golgi-to-ER relocation of GALNT6, where GALNT6 may O-glycosylate more HSPA5 as well as other proteins for tumorigenesis. HSPA5 was already known to be overexpressed in many types of cancer. In addition, it may be often induced by lots of stress in the tumor microenvironment, such as hypoxia, acidosis, starvation, as well as exogenous stresses, for example, therapeutic interventions. Under these conditions, Golgi-to-ER relocation of GALNT6 may frequently occur, which may trigger O-glycosylation of multiple substrates at ER. Our data provided an example of how HSPA5 regulates GALNT6 in breast carcinogenesis.

In summary, we performed VVA lection pull-down assay followed by MS analysis, and identified HSPA5 as a novel substrate of GALNT6. We confirmed GALNT6 stabilizes HSPA5 protein through O-glycosylation. Meanwhile, HSPA5 also regulates GALNT6 in breast carcinogenesis by driving Golgi-to-ER relocation of GALNT6. Our study revealed a novel mechanism of how GALNT6 and HSPA5 cooperate together to promote mammary carcinogenesis. And the O-glycosylated form of HSPA5 may be a good target for breast cancer therapy.

Figures and tables

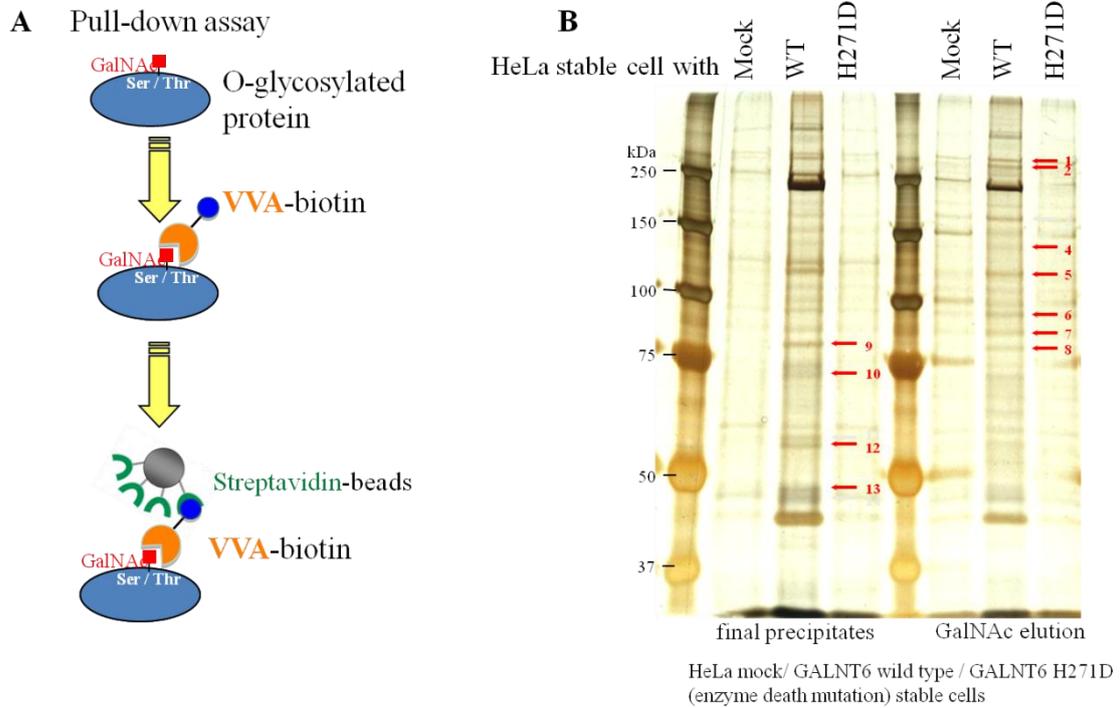


Figure 1. Screening substrates of GALNT6.

(A) A pictorial schema for identification of GALNT6 substrates by using VVA lectin pull-down assay.

(B) Silver staining after VVA-lectin pull-down assay (two elution methods were used) in HeLa mock, GALNT6 wild type and GALNT6 H271D (enzyme death mutation) stable cells. The differential bands which indicated as arrow were excised and analyzed by mass spectrometry. HSPA5 (GRP78) was identified as one of the novel substrates of GALNT6.

Table 2. Candidate substrates of GALNT6

| | name | (kDa) |
|--------|--|-------|
| (1) | Filamin-A | 283 |
| (2) | Agrin | 223 |
| (4) | Leucine-rich repeat, death domain-containing protein | 99 |
| (5) | Splicing factor 3 subunit 1 | 89 |
| (6) | Striatin-4 | 81 |
| (7) | Apoptosis-stimulating of p53 protein 2 | 126 |
| (8, 9) | Adseverin | 81 |
| (9) | Uncharacterized protein KIAA1486 | 72 |
| (10) | 78 kDa glucose-regulated protein (GRP78) | 72 |
| (12) | Annexin A1 | 39 |
| | Lamin-A/C | 74 |
| | Heat shock protein beta-1 (HSP27) | 23 |
| | Lysozyme C | 170 |
| | Haptoglobin | 46 |
| | Cornulin | 54 |
| | Uncharacterized protein KIAA1529 | 192 |
| (13) | Actin, cytoplasmic 1 | 42 |
| (13) | Glyceraldehyde-3-phosphate dehydrogenase | 36 |

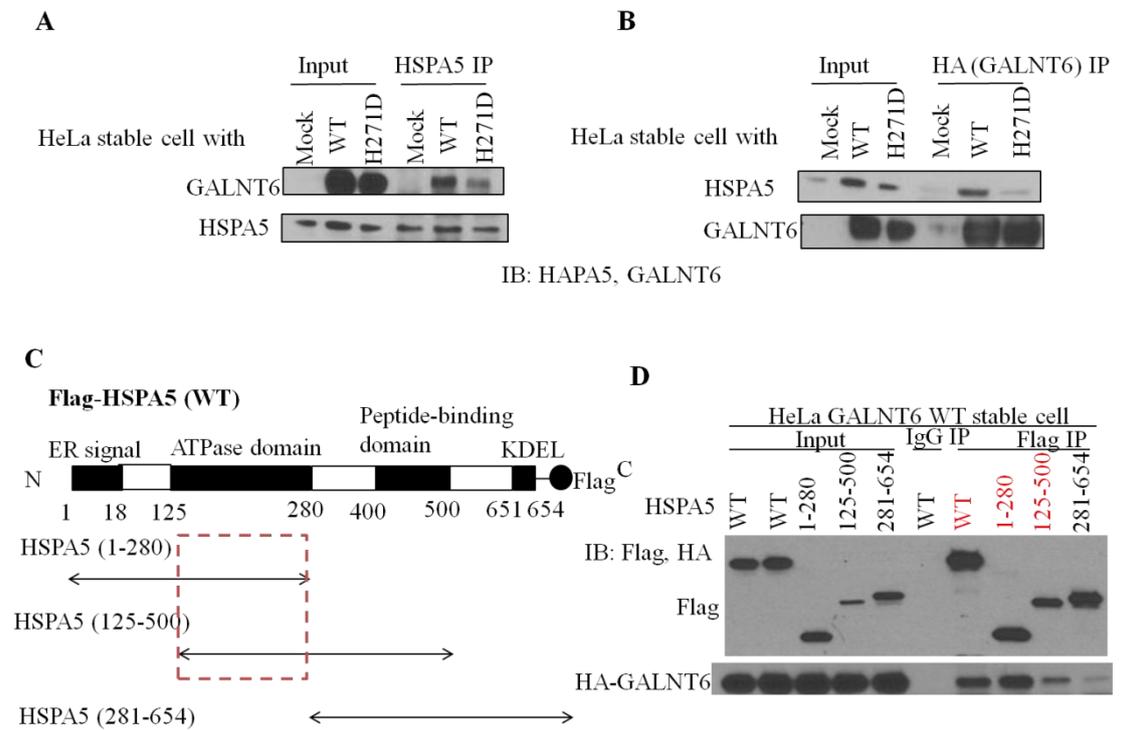


Figure 2. HSPA5 is a GALNT6-interacting protein

(A and B) Immunoprecipitation-Western blot analysis was performed. GALNT6 coprecipitates with HSPA5 in HeLa GALNT6 WT and H271D stable cells, but not in mock stable cells. Although GALNT6 mutant, H271D can also bind to HSPA5, the binding ability is much weaker when compared with wild type protein. (A) Endogenous HSPA5 was pull down by using HSPA5 antibody followed by GALNT6 and HSPA5 western blot. (B) Endogenous HA-GALNT6 was pull down by anti-HA, and then HSPA5 and GALNT6 were detected by western blot. (C) To map the binding domain on HSPA5, flag-tagged wild-type HSPA5 and HSPA5 fragments with or without ATPase domain were generated. (D) HeLa GALNT6 WT stable cells were transfected with these plasmids. 48 h later, immunoprecipitation-western blot analysis was performed. GALNT6 coprecipitated with wild-type HSPA5, HSPA5-1-280 and HSPA5-125-

500, all of which have ATPase domain, but GALNT6 binding to HSPA5-281-654 without ATPase domain was much reduced. This data indicated that the ATPase domain of HSPA5 is important for their binding.

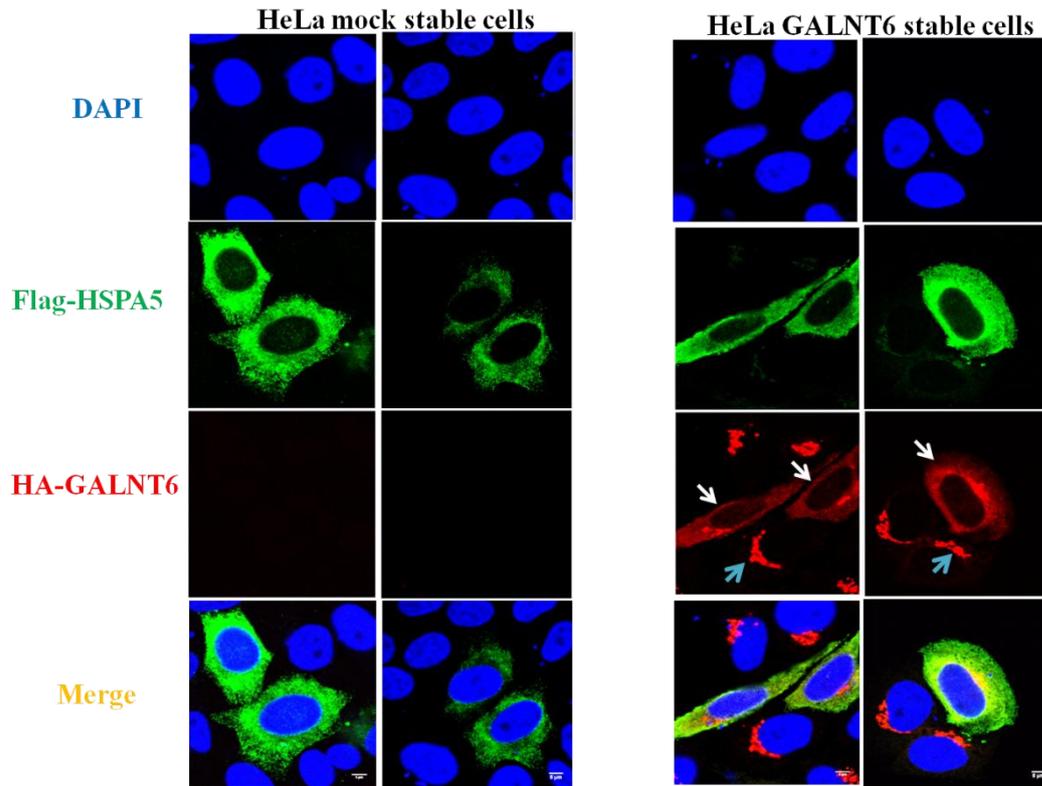


Figure 3. Overexpression of HSPA5 drives Golgi-to-ER relocation of GALNT6

To further confirm GALNT6 and HSPA5 are in the same complex, immunocytochemistry (ICC) analysis was performed in HeLa stable cells. We observed wild type GALNT6 locates in Golgi, and high expression level of exogenous HSPA5 (not low level) drives Golgi-to-ER relocation of GALNT6 and they co-localize at ER in HeLa GALNT6 WT stable cells. Mock stable cells served as negative control to exclude cross reactivity of the antibodies.

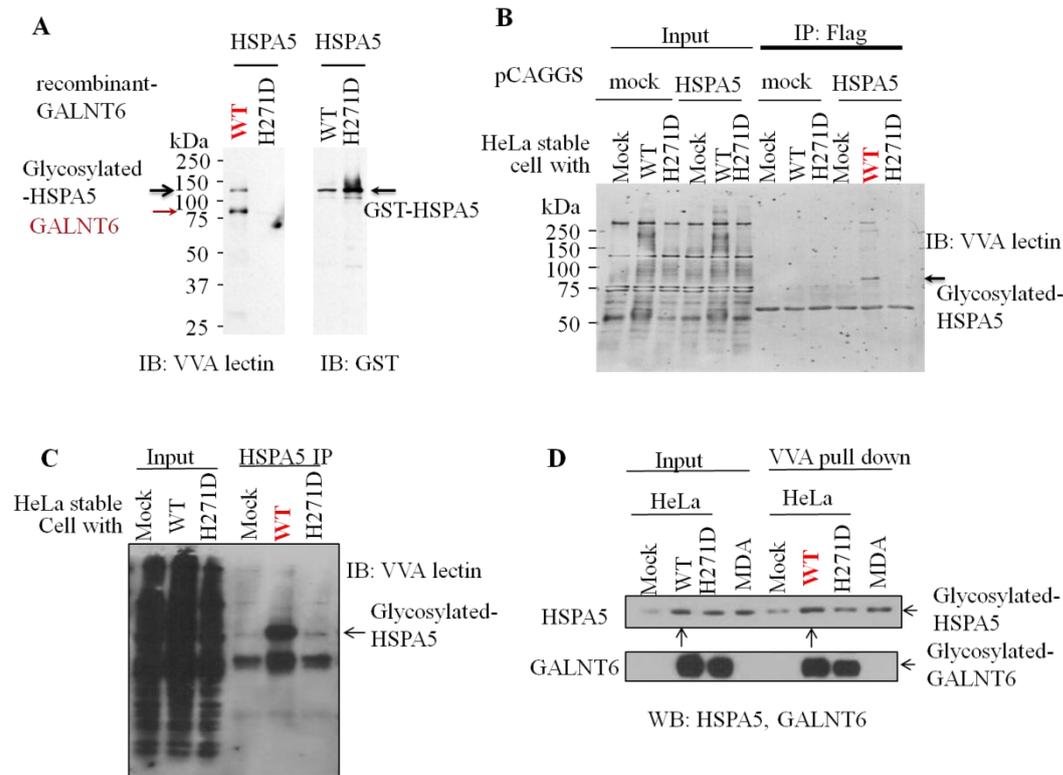


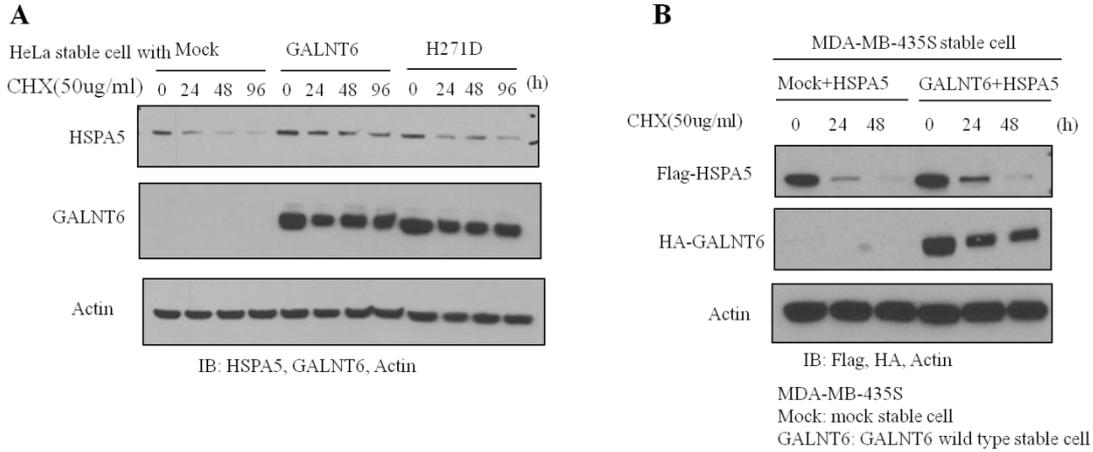
Figure 4. GALNT6 glycosylates HSPA5

(A) By using recombinant HSPA5 and GALNT6 wild type and mutant protein in vitro assay, HSPA5 can only be O-glycosylated in the present of GALNT6 wild type protein but not GALNT6 mutant, H271D. During the process, only wild type GALNT6 itself can also be auto-glycosylated. Moreover, in cell based system, we pulled down exogenous (B) or endogenous HSPA5 (C, by anti-HSPA5, and D, by VVA lectin pull-down assay), and detected its glycosylation level. We found HSPA5 showed much higher level of O-glycosylation in HeLa GALNT6 WT stable cells (B, C and D), when compared with mock and H271D stable cells, indicating that GALNT6 can O-glycosylate HSPA5, although GALNT6 is not the only one GalNAc-type O-glycosyltransferase for HSPA5, since we observed HSPA5 can also be O-glycosylated in GALNT6(-) cells (D). MDA: MDA-MB-435S cells (GALNT6 (-)).

Table 2. Potential GALNT6-induced O-glycosylation sites in HSPA5 by MS analysis

| Peptide_sequence | Glycosylation sites | Comment |
|-------------------------|----------------------------|-----------------|
| QATKDAGTIAGLNVMR | T184 | Positive |
| VLTKMKETAEAYLGK | T151 | Positive |
| VTHAVVTVPAYFNDAQR | T166 | Positive |
| NQLTSNPENTVFDAKR | T85 | Positive |
| IINEPTAAAIAYGLDKR | T203 | Positive |
| GTFDLTGIPPAPR | T481 | Positive |

CHX: cycloheximide, a protein synthesis inhibitor



C
Construction of 7 candidate O-glycosylation site mutant plasmids

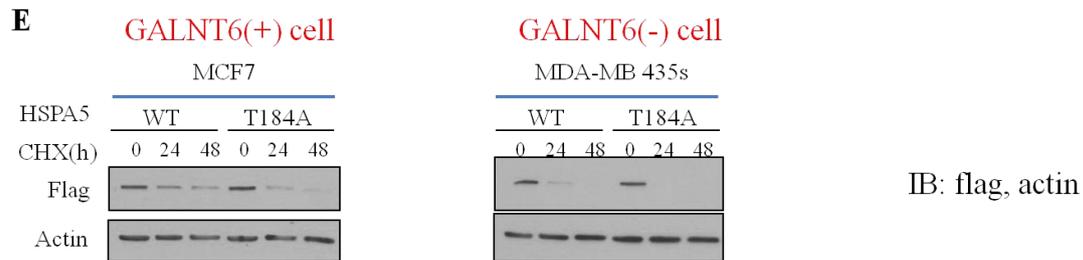
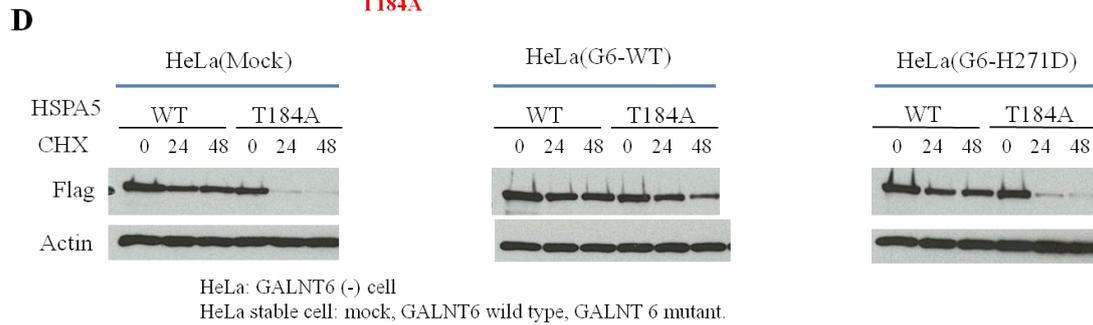
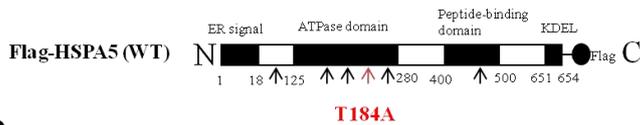


Figure 5. GALNT6 stabilizes HSPA5 through O-glycosylation

(A) HeLa stable cells were treated with CHX (cycloheximide), a protein synthesis inhibitor, at different time point, followed by western blot. Endogenous HSPA5 is more stable in HeLa GALNT6 WT stable cells when compared with mock and H271D stable cells. (B) MDA-MB-435S-Mock and GALNT6 WT

stable cell were transfected with flag-HSPA5, and then treated with CHX and performed western blot. GALNT6 prolongs the half time of exogenous HSPA5 in MDA-MB-435S-GALNT6 WT stable cells when compared with mock stable cells. (C) Six HSPA5 single amino acid substituted mutants (T85A, T151A, T166A, T184A, T203A and T481A) and a mutant contained 5 substituted sites (T85A, T151A, T166A, T184A, and T203A) were generated. (D) HeLa stable cells were transfected with HSPA5 wild-type and T184A, followed by CHX treatment and western blot. T184A is less stable than wild type protein in all types of HeLa stable cells. And the half time of HSPA5 WT and mutant protein is less stable in GALNT6 (-) cells when compared with positive cells. (E) Same experiment was also performed in MCF7 breast cancer cells with endogenous GALNT6, and MDA-MB 435s without GALNT6. Similar results were observed.

References

1. Ferlay J, Shin HR, Bray F, Forman D, Mathers C and Parkin DM: Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *International journal of cancer. Journal international du cancer* 127: 2893-2917, 2010.
2. Jemal A, Bray F, Center MM, Ferlay J, Ward E and Forman D: Global cancer statistics. *CA: a cancer journal for clinicians* 61: 69-90, 2011.
3. Naito S, Tomita Y, Rha SY, et al.: Kidney Cancer Working Group report. *Japanese journal of clinical oncology* 40 Suppl 1: i51-56, 2010.
4. Siegel R, Naishadham D and Jemal A: Cancer statistics, 2012. *CA: a cancer journal for clinicians* 62: 10-29, 2012.
5. Chow WH, Devesa SS, Warren JL and Fraumeni JF, Jr.: Rising incidence of renal cell cancer in the United States. *JAMA : the journal of the American Medical Association* 281: 1628-1631, 1999.
6. Hock LM, Lynch J and Balaji KC: Increasing incidence of all stages of kidney cancer in the last 2 decades in the United States: An analysis of Surveillance, Epidemiology and End Results program data. *J Urology* 167: 57-60, 2002.
7. Lindblad P: Epidemiology of renal cell carcinoma. *Scandinavian journal of surgery : SJS : official organ for the Finnish Surgical Society and the Scandinavian Surgical Society* 93: 88-96, 2004.
8. Murai M and Oya M: Renal cell carcinoma: etiology, incidence and epidemiology. *Current opinion in urology* 14: 229-233, 2004.
9. Motzer RJ, Agarwal N, Beard C, et al.: NCCN clinical practice guidelines in oncology: kidney cancer. *Journal of the National Comprehensive Cancer Network : JNCCN* 7: 618-630, 2009.
10. Ljungberg B, Cowan NC, Hanbury DC, et al.: EAU guidelines on renal cell carcinoma: the 2010 update. *European urology* 58: 398-406, 2010.
11. Escudier B, Eisen T, Porta C, et al.: Renal cell carcinoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO* 23 Suppl 7: vii65-71, 2012.
12. Lam JS, Leppert JT, Belldegrün AS and Figlin RA: Novel approaches in the therapy of metastatic renal cell carcinoma. *World journal of urology* 23: 202-212, 2005.
13. Cairns P: Renal cell carcinoma. *Cancer biomarkers : section A of Disease markers* 9: 461-473, 2010.
14. Molina AM and Motzer RJ: Current algorithms and prognostic factors in the treatment of metastatic renal cell carcinoma. *Clinical genitourinary cancer* 6 Suppl 1: S7-13, 2008.
15. Ono K, Tanaka T, Tsunoda T, et al.: Identification by cDNA microarray of genes involved in

ovarian carcinogenesis. *Cancer research* 60: 5007-5011, 2000.

16. Hirota E, Yan L, Tsunoda T, et al.: Genome-wide gene expression profiles of clear cell renal cell carcinoma: identification of molecular targets for treatment of renal cell carcinoma. *International journal of oncology* 29: 799-827, 2006.
17. Saito-Hisaminato A, Katagiri T, Kakiuchi S, Nakamura T, Tsunoda T and Nakamura Y: Genome-wide profiling of gene expression in 29 normal human tissues with a cDNA microarray. *DNA research : an international journal for rapid publication of reports on genes and genomes* 9: 35-45, 2002.
18. Oda K, Arakawa H, Tanaka T, et al.: p53AIP1, a potential mediator of p53-dependent apoptosis, and its regulation by Ser-46-phosphorylated p53. *Cell* 102: 849-862, 2000.
19. Matsuda K, Yoshida K, Taya Y, Nakamura K, Nakamura Y and Arakawa H: p53AIP1 regulates the mitochondrial apoptotic pathway. *Cancer research* 62: 2883-2889, 2002.
20. Tanikawa C, Ueda K, Nakagawa H, Yoshida N, Nakamura Y and Matsuda K: Regulation of Protein Citrullination through p53/PAD14 Network in DNA Damage Response. *Cancer research* 69: 8761-8769, 2009.
21. Tanikawa C, Matsuda K, Fukuda S, Nakamura Y and Arakawa H: p53RDL1 regulates p53-dependent apoptosis. *Nature cell biology* 5: 216-223, 2003.
22. Watanabe G, Kato S, Nakata H, Ishida T, Ohuchi N and Ishioka C: alphaB-crystallin: a novel p53-target gene required for p53-dependent apoptosis. *Cancer science* 100: 2368-2375, 2009.
23. Evans JR, Bosman JD, Brown-Endres L, Yehiely F and Cryns VL: Induction of the small heat shock protein alphaB-crystallin by genotoxic stress is mediated by p53 and p73. *Breast cancer research and treatment* 122: 159-168, 2010.
24. Krief S, Faivre JF, Robert P, et al.: Identification and characterization of cvHsp. A novel human small stress protein selectively expressed in cardiovascular and insulin-sensitive tissues. *The Journal of biological chemistry* 274: 36592-36600, 1999.
25. Sun X, Fontaine JM, Rest JS, Sheldon EA, Welsh MJ and Benndorf R: Interaction of human HSP22 (HSPB8) with other small heat shock proteins. *The Journal of biological chemistry* 279: 2394-2402, 2004.
26. Vos MJ, Zijlstra MP, Kanon B, et al.: HSPB7 is the most potent polyQ aggregation suppressor within the HSPB family of molecular chaperones. *Human molecular genetics* 19: 4677-4693, 2010.
27. Cappola TP, Li M, He J, et al.: Common variants in HSPB7 and FRMD4B associated with advanced heart failure. *Circulation. Cardiovascular genetics* 3: 147-154, 2010.
28. Matkovich SJ, Van Booven DJ, Hinds A, et al.: Cardiac signaling genes exhibit unexpected sequence diversity in sporadic cardiomyopathy, revealing HSPB7 polymorphisms associated with disease. *The Journal of clinical investigation* 120: 280-289, 2010.
29. Stark K, Esslinger UB, Reinhard W, et al.: Genetic association study identifies HSPB7 as a risk gene for idiopathic dilated cardiomyopathy. *PLoS genetics* 6: e1001167, 2010.

30. Villard E, Perret C, Gary F, et al.: A genome-wide association study identifies two loci associated with heart failure due to dilated cardiomyopathy. *European heart journal* 32: 1065-1076, 2011.
31. Li XP, Luo R, Hua W and Hua W: Polymorphisms of Hspb7 Gene Associate with Idiopathic Dilated Cardiomyopathy Susceptibility in a Chinese Population. *Heart* 98: E47-E47, 2012.
32. Rosenfeld GE, Mercer EJ, Mason CE and Evans T: Small heat shock proteins Hspb7 and Hspb12 regulate early steps of cardiac morphogenesis. *Developmental biology* 381: 389-400, 2013.
33. Huang Z, Cheng Y, Chiu PM, et al.: Tumor suppressor Alpha B-crystallin (CRYAB) associates with the cadherin/catenin adherens junction and impairs NPC progression-associated properties. *Oncogene* 31: 3709-3720, 2012.
34. Ragnarsson G, Eiriksdottir G, Johannsdottir JT, Jonasson JG, Egilsson V and Ingvarsson S: Loss of heterozygosity at chromosome 1p in different solid human tumours: association with survival. *Brit J Cancer* 79: 1468-1474, 1999.
35. Ferlay J, Soerjomataram I, Dikshit R, et al.: Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *International journal of cancer. Journal international du cancer*, 2014.
36. Siegel R, Ma J, Zou Z and Jemal A: Cancer statistics, 2014. *CA: a cancer journal for clinicians* 64: 9-29, 2014.
37. Perou CM, Sorlie T, Eisen MB, et al.: Molecular portraits of human breast tumours. *Nature* 406: 747-752, 2000.
38. Cancer Genome Atlas N: Comprehensive molecular portraits of human breast tumours. *Nature* 490: 61-70, 2012.
39. Chalasani P, Downey L and Stopeck AT: Caring for the breast cancer survivor: a guide for primary care physicians. *The American journal of medicine* 123: 489-495, 2010.
40. DeSantis CE, Lin CC, Mariotto AB, et al.: Cancer treatment and survivorship statistics, 2014. *CA: a cancer journal for clinicians* 64: 252-271, 2014.
41. Apweiler R, Hermjakob H and Sharon N: On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochimica et biophysica acta* 1473: 4-8, 1999.
42. Lowe JB and Marth JD: A genetic approach to Mammalian glycan function. *Annual review of biochemistry* 72: 643-691, 2003.
43. Gill DJ, Clausen H and Bard F: Location, location, location: new insights into O-GalNAc protein glycosylation. *Trends in cell biology* 21: 149-158, 2011.
44. Hart GW: Dynamic O-linked glycosylation of nuclear and cytoskeletal proteins. *Annual review of biochemistry* 66: 315-335, 1997.
45. Wopereis S, Lefeber DJ, Morava E and Wevers RA: Mechanisms in protein O-glycan biosynthesis and clinical and molecular aspects of protein O-glycan biosynthesis defects: a review. *Clinical chemistry* 52: 574-600, 2006.

46. Luther KB and Haltiwanger RS: Role of unusual O-glycans in intercellular signaling. *The international journal of biochemistry & cell biology* 41: 1011-1024, 2009.
47. Bennett EP, Mandel U, Clausen H, Gerken TA, Fritz TA and Tabak LA: Control of mucin-type O-glycosylation: a classification of the polypeptide GalNAc-transferase gene family. *Glycobiology* 22: 736-756, 2012.
48. Clausen H and Bennett EP: A family of UDP-GalNAc: polypeptide N-acetylgalactosaminyltransferases control the initiation of mucin-type O-linked glycosylation. *Glycobiology* 6: 635-646, 1996.
49. Ten Hagen KG, Fritz TA and Tabak LA: All in the family: the UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases. *Glycobiology* 13: 1R-16R, 2003.
50. Raman J, Guan Y, Perrine CL, Gerken TA and Tabak LA: UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferases: completion of the family tree. *Glycobiology* 22: 768-777, 2012.
51. Paulson JC and Colley KJ: Glycosyltransferases. Structure, localization, and control of cell type-specific glycosylation. *The Journal of biological chemistry* 264: 17615-17618, 1989.
52. Pedersen JW, Bennett EP, Schjoldager KT, et al.: Lectin domains of polypeptide GalNAc transferases exhibit glycopeptide binding specificity. *The Journal of biological chemistry* 286: 32684-32696, 2011.
53. Hazes B: The (QxW)₃ domain: a flexible lectin scaffold. *Protein science : a publication of the Protein Society* 5: 1490-1501, 1996.
54. Imberty A, Piller V, Piller F and Breton C: Fold recognition and molecular modeling of a lectin-like domain in UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases. *Protein engineering* 10: 1353-1356, 1997.
55. Bennett EP, Hassan H, Mandel U, et al.: Cloning and characterization of a close homologue of human UDP-N-acetyl-alpha-D-galactosamine:Polypeptide N-acetylgalactosaminyltransferase-T3, designated GalNAc-T6. Evidence for genetic but not functional redundancy. *The Journal of biological chemistry* 274: 25362-25370, 1999.
56. Bennett EP, Hassan H, Hollingsworth MA and Clausen H: A novel human UDP-N-acetyl-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase, GalNAc-T7, with specificity for partial GalNAc-glycosylated acceptor substrates. *FEBS letters* 460: 226-230, 1999.
57. Ten Hagen KG, Tetaert D, Hagen FK, et al.: Characterization of a UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase that displays glycopeptide N-acetylgalactosaminyltransferase activity. *The Journal of biological chemistry* 274: 27867-27874, 1999.
58. Hanisch FG, Reis CA, Clausen H and Paulsen H: Evidence for glycosylation-dependent activities of polypeptide N-acetylgalactosaminyltransferases rGalNAc-T2 and -T4 on mucin glycopeptides. *Glycobiology* 11: 731-740, 2001.
59. Gerken TA, Raman J, Fritz TA and Jamison O: Identification of common and unique peptide

substrate preferences for the UDP-GalNAc:polypeptide alpha-N-acetylgalactosaminyltransferases T1 and T2 derived from oriented random peptide substrates. *The Journal of biological chemistry* 281: 32403-32416, 2006.

60. Pratt MR, Hang HC, Ten Hagen KG, et al.: Deconvoluting the functions of polypeptide N-alpha-acetylgalactosaminyltransferase family members by glycopeptide substrate profiling. *Chemistry & biology* 11: 1009-1016, 2004.

61. Perrine CL, Ganguli A, Wu P, et al.: Glycopeptide-preferring polypeptide GalNAc transferase 10 (ppGalNAc T10), involved in mucin-type O-glycosylation, has a unique GalNAc-O-Ser/Thr-binding site in its catalytic domain not found in ppGalNAc T1 or T2. *The Journal of biological chemistry* 284: 20387-20397, 2009.

62. Jensen PH, Kolarich D and Packer NH: Mucin-type O-glycosylation--putting the pieces together. *The FEBS journal* 277: 81-94, 2010.

63. Ichikawa S, Sorenson AH, Austin AM, et al.: Ablation of the Galnt3 gene leads to low-circulating intact fibroblast growth factor 23 (Fgf23) concentrations and hyperphosphatemia despite increased Fgf23 expression. *Endocrinology* 150: 2543-2550, 2009.

64. Miyazaki T, Mori M, Yoshida CA, et al.: Galnt3 deficiency disrupts acrosome formation and leads to oligoasthenoteratozoospermia. *Histochemistry and cell biology* 139: 339-354, 2013.

65. Schjoldager KT and Clausen H: Site-specific protein O-glycosylation modulates proprotein processing - deciphering specific functions of the large polypeptide GalNAc-transferase gene family. *Biochimica et biophysica acta* 1820: 2079-2094, 2012.

66. Schjoldager KT, Vester-Christensen MB, Goth CK, et al.: A systematic study of site-specific GalNAc-type O-glycosylation modulating proprotein convertase processing. *The Journal of biological chemistry* 286: 40122-40132, 2011.

67. Schjoldager KT, Vester-Christensen MB, Bennett EP, et al.: O-glycosylation modulates proprotein convertase activation of angiotensin-like protein 3: possible role of polypeptide GalNAc-transferase-2 in regulation of concentrations of plasma lipids. *The Journal of biological chemistry* 285: 36293-36303, 2010.

68. Fu J, Gerhardt H, McDaniel JM, et al.: Endothelial cell O-glycan deficiency causes blood/lymphatic misconnections and consequent fatty liver disease in mice. *The Journal of clinical investigation* 118: 3725-3737, 2008.

69. Tian E, Hoffman MP and Ten Hagen KG: O-glycosylation modulates integrin and FGF signalling by influencing the secretion of basement membrane components. *Nature communications* 3: 869, 2012.

70. Zhang L and Ten Hagen KG: Dissecting the biological role of mucin-type O-glycosylation using RNA interference in *Drosophila* cell culture. *The Journal of biological chemistry* 285: 34477-34484, 2010.

71. Herr P, Korniychuk G, Yamamoto Y, Grubisic K and Oelgeschlager M: Regulation of TGF-(beta)

- signalling by N-acetylgalactosaminyltransferase-like 1. *Development* 135: 1813-1822, 2008.
72. Burchell JM, Mungul A and Taylor-Papadimitriou J: O-linked glycosylation in the mammary gland: changes that occur during malignancy. *Journal of mammary gland biology and neoplasia* 6: 355-364, 2001.
73. Brockhausen I: Mucin-type O-glycans in human colon and breast cancer: glycodynamics and functions. *EMBO reports* 7: 599-604, 2006.
74. Radhakrishnan P, Dabelsteen S, Madsen FB, et al.: Immature truncated O-glycophenotype of cancer directly induces oncogenic features. *Proceedings of the National Academy of Sciences of the United States of America* 111: E4066-4075, 2014.
75. von Mensdorff-Pouilly S, Gourevitch MM, Kenemans P, et al.: Humoral immune response to polymorphic epithelial mucin (MUC-1) in patients with benign and malignant breast tumours. *European journal of cancer* 32A: 1325-1331, 1996.
76. Wandall HH, Blixt O, Tarp MA, et al.: Cancer biomarkers defined by autoantibody signatures to aberrant O-glycopeptide epitopes. *Cancer research* 70: 1306-1313, 2010.
77. Sihlbom C, van Dijk Hard I, Lidell ME, Noll T, Hansson GC and Backstrom M: Localization of O-glycans in MUC1 glycoproteins using electron-capture dissociation fragmentation mass spectrometry. *Glycobiology* 19: 375-381, 2009.
78. Muller S and Hanisch FG: Recombinant MUC1 probe authentically reflects cell-specific O-glycosylation profiles of endogenous breast cancer mucin. High density and prevalent core 2-based glycosylation. *The Journal of biological chemistry* 277: 26103-26112, 2002.
79. Tarp MA and Clausen H: Mucin-type O-glycosylation and its potential use in drug and vaccine development. *Biochimica et biophysica acta* 1780: 546-563, 2008.
80. Backstrom M, Link T, Olson FJ, et al.: Recombinant MUC1 mucin with a breast cancer-like O-glycosylation produced in large amounts in Chinese-hamster ovary cells. *The Biochemical journal* 376: 677-686, 2003.
81. Mall AS: Analysis of mucins: role in laboratory diagnosis. *Journal of clinical pathology* 61: 1018-1024, 2008.
82. Storr SJ, Royle L, Chapman CJ, et al.: The O-linked glycosylation of secretory/shed MUC1 from an advanced breast cancer patient's serum. *Glycobiology* 18: 456-462, 2008.
83. Napoletano C, Rughetti A, Agervig Tarp MP, et al.: Tumor-associated Tn-MUC1 glycoform is internalized through the macrophage galactose-type C-type lectin and delivered to the HLA class I and II compartments in dendritic cells. *Cancer research* 67: 8358-8367, 2007.
84. Wahrenbrock MG and Varki A: Multiple hepatic receptors cooperate to eliminate secretory mucins aberrantly entering the bloodstream: are circulating cancer mucins the "tip of the iceberg"? *Cancer research* 66: 2433-2441, 2006.
85. Brooks SA, Carter TM, Bennett EP, Clausen H and Mandel U: Immunolocalisation of members

of the polypeptide N-acetylgalactosaminyl transferase (ppGalNAc-T) family is consistent with biologically relevant altered cell surface glycosylation in breast cancer. *Acta histochemica* 109: 273-284, 2007.

86. Freire T, Berois N, Sonora C, Varangot M, Barrios E and Osinaga E: UDP-N-acetyl-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 6 (ppGalNAc-T6) mRNA as a potential new marker for detection of bone marrow-disseminated breast cancer cells. *International journal of cancer. Journal international du cancer* 119: 1383-1388, 2006.

87. Patani N, Jiang W and Mokbel K: Prognostic utility of glycosyltransferase expression in breast cancer. *Cancer genomics & proteomics* 5: 333-340, 2008.

88. Park JH, Nishidate T, Kijima K, et al.: Critical roles of mucin 1 glycosylation by transactivated polypeptide N-acetylgalactosaminyltransferase 6 in mammary carcinogenesis. *Cancer research* 70: 2759-2769, 2010.

89. Gill DJ, Tham KM, Chia J, et al.: Initiation of GalNAc-type O-glycosylation in the endoplasmic reticulum promotes cancer cell invasiveness. *Proceedings of the National Academy of Sciences of the United States of America* 110: E3152-3161, 2013.

90. Gill DJ, Chia J, Senewiratne J and Bard F: Regulation of O-glycosylation through Golgi-to-ER relocation of initiation enzymes. *The Journal of cell biology* 189: 843-858, 2010.

91. Park JH, Katagiri T, Chung S, Kijima K and Nakamura Y: Polypeptide N-acetylgalactosaminyltransferase 6 disrupts mammary acinar morphogenesis through O-glycosylation of fibronectin. *Neoplasia* 13: 320-326, 2011.

92. Iskratsch T, Braun A, Paschinger K and Wilson IB: Specificity analysis of lectins and antibodies using remodeled glycoproteins. *Anal Biochem* 386: 133-146, 2009.

93. Tollefsen SE and Kornfeld R: Isolation and characterization of lectins from *Vicia villosa*. Two distinct carbohydrate binding activities are present in seed extracts. *The Journal of biological chemistry* 258: 5165-5171, 1983.

94. North SJ, Hitchen PG, Haslam SM and Dell A: Mass spectrometry in the analysis of N-linked and O-linked glycans. *Current opinion in structural biology* 19: 498-506, 2009.

95. Wang Q, He Z, Zhang J, et al.: Overexpression of endoplasmic reticulum molecular chaperone GRP94 and GRP78 in human lung cancer tissues and its significance. *Cancer detection and prevention* 29: 544-551, 2005.

96. Fernandez PM, Tabbara SO, Jacobs LK, et al.: Overexpression of the glucose-regulated stress gene GRP78 in malignant but not benign human breast lesions. *Breast cancer research and treatment* 59: 15-26, 2000.

97. Gazit G, Lu J and Lee AS: De-regulation of GRP stress protein expression in human breast cancer cell lines. *Breast cancer research and treatment* 54: 135-146, 1999.

98. Pootrakul L, Datar RH, Shi SR, et al.: Expression of stress response protein Grp78 is associated

with the development of castration-resistant prostate cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* 12: 5987-5993, 2006.

99. Xing X, Lai M, Wang Y, Xu E and Huang Q: Overexpression of glucose-regulated protein 78 in colon cancer. *Clinica chimica acta; international journal of clinical chemistry* 364: 308-315, 2006.
100. Zhang J, Jiang Y, Jia Z, et al.: Association of elevated GRP78 expression with increased lymph node metastasis and poor prognosis in patients with gastric cancer. *Clinical & experimental metastasis* 23: 401-410, 2006.
101. Shuda M, Kondoh N, Imazeki N, et al.: Activation of the ATF6, XBP1 and grp78 genes in human hepatocellular carcinoma: a possible involvement of the ER stress pathway in hepatocarcinogenesis. *Journal of hepatology* 38: 605-614, 2003.
102. Ni M and Lee AS: ER chaperones in mammalian development and human diseases. *FEBS letters* 581: 3641-3651, 2007.
103. Hetz C: The unfolded protein response: controlling cell fate decisions under ER stress and beyond. *Nature reviews. Molecular cell biology* 13: 89-102, 2012.
104. Luo B and Lee AS: The critical roles of endoplasmic reticulum chaperones and unfolded protein response in tumorigenesis and anticancer therapies. *Oncogene* 32: 805-818, 2013.
105. Cook KL, Shajahan AN, Warri A, Jin L, Hilakivi-Clarke LA and Clarke R: Glucose-regulated protein 78 controls cross-talk between apoptosis and autophagy to determine antiestrogen responsiveness. *Cancer research* 72: 3337-3349, 2012.
106. Grkovic S, O'Reilly VC, Han S, Hong M, Baxter RC and Firth SM: IGFBP-3 binds GRP78, stimulates autophagy and promotes the survival of breast cancer cells exposed to adverse microenvironments. *Oncogene* 32: 2412-2420, 2013.
107. Li J, Ni M, Lee B, Barron E, Hinton DR and Lee AS: The unfolded protein response regulator GRP78/BiP is required for endoplasmic reticulum integrity and stress-induced autophagy in mammalian cells. *Cell death and differentiation* 15: 1460-1471, 2008.
108. Dong D, Stapleton C, Luo B, et al.: A critical role for GRP78/BiP in the tumor microenvironment for neovascularization during tumor growth and metastasis. *Cancer research* 71: 2848-2857, 2011.
109. Fu Y and Lee AS: Glucose regulated proteins in cancer progression, drug resistance and immunotherapy. *Cancer biology & therapy* 5: 741-744, 2006.
110. Dong D, Ni M, Li J, et al.: Critical role of the stress chaperone GRP78/BiP in tumor proliferation, survival, and tumor angiogenesis in transgene-induced mammary tumor development. *Cancer research* 68: 498-505, 2008.
111. Jamora C, Dennert G and Lee AS: Inhibition of tumor progression by suppression of stress protein GRP78/BiP induction in fibrosarcoma B/C10ME. *Proceedings of the National Academy of Sciences of the United States of America* 93: 7690-7694, 1996.

112. Fu Y, Wey S, Wang M, et al.: Pten null prostate tumorigenesis and AKT activation are blocked by targeted knockout of ER chaperone GRP78/BiP in prostate epithelium. *Proceedings of the National Academy of Sciences of the United States of America* 105: 19444-19449, 2008.
113. Wu MJ, Jan CI, Tsay YG, et al.: Elimination of head and neck cancer initiating cells through targeting glucose regulated protein78 signaling. *Molecular cancer* 9: 283, 2010.
114. Bartkowiak K, Effenberger KE, Harder S, et al.: Discovery of a novel unfolded protein response phenotype of cancer stem/progenitor cells from the bone marrow of breast cancer patients. *Journal of proteome research* 9: 3158-3168, 2010.
115. Luo S, Mao C, Lee B and Lee AS: GRP78/BiP is required for cell proliferation and protecting the inner cell mass from apoptosis during early mouse embryonic development. *Molecular and cellular biology* 26: 5688-5697, 2006.
116. Lakshminarayanan V, Thompson P, Wolfert MA, et al.: Immune recognition of tumor-associated mucin MUC1 is achieved by a fully synthetic aberrantly glycosylated MUC1 tripartite vaccine. *Proceedings of the National Academy of Sciences of the United States of America* 109: 261-266, 2012.
117. Steentoft C, Vakhrushev SY, Vester-Christensen MB, et al.: Mining the O-glycoproteome using zinc-finger nuclease-glycoengineered SimpleCell lines. *Nature methods* 8: 977-982, 2011.
118. Mayer M, Reinstein J and Buchner J: Modulation of the ATPase cycle of BiP by peptides and proteins. *J Mol Biol* 330: 137-144, 2003.
119. Awad W, Estrada I, Shen Y and Hendershot LM: BiP mutants that are unable to interact with endoplasmic reticulum DnaJ proteins provide insights into interdomain interactions in BiP. *Proc Natl Acad Sci U S A* 105: 1164-1169, 2008.
120. Bertrand J, Tennoune N, Marion-Letellier R, et al.: Evaluation of ubiquitinated proteins by proteomics reveals the role of the ubiquitin proteasome system in the regulation of Grp75 and Grp78 chaperone proteins during intestinal inflammation. *Proteomics* 13: 3284-3292, 2013.
121. Chang YW, Chen HA, Tseng CF, et al.: De-acetylation and degradation of HSPA5 is critical for E1A metastasis suppression in breast cancer cells. *Oncotarget* 5: 10558-10570, 2014.
122. Kim W, Bennett EJ, Huttlin EL, et al.: Systematic and quantitative assessment of the ubiquitin-modified proteome. *Mol Cell* 44: 325-340, 2011.
123. Sharma K, D'Souza RC, Tyanova S, et al.: Ultradeep human phosphoproteome reveals a distinct regulatory nature of Tyr and Ser/Thr-based signaling. *Cell Rep* 8: 1583-1594, 2014.
124. Lundby A, Secher A, Lage K, et al.: Quantitative maps of protein phosphorylation sites across 14 different rat organs and tissues. *Nat Commun* 3: 876, 2012.
125. Chen Y, Choong LY, Lin Q, et al.: Differential expression of novel tyrosine kinase substrates during breast cancer development. *Mol Cell Proteomics* 6: 2072-2087, 2007.

Acknowledgements

First, I would like to thank Professor Nakamura, my supervisor. In these years, he provided me scholarship and supported my research not only in the University of Tokyo, but also in the University of Chicago. He gave me a lot of opportunities to study and explore what I want to. He guided me step by step. His scientific attitude, courage and profound knowledge deeply influenced me. Everything I learned from him will encourage me to go further in the field of science.

Second, I send my deep acknowledgement to my supervisor, Dr. Matsuda. From everything unknown to now I can do many experiments and know how to analyze the data, I would like to say my every progress cannot be achieved without his support and guidance. He was strict with my research and guided me very detail in the experiments. He also cared about what I can learn and gave me a lot of freedom to study other useful knowledge. I admired his kindness, hard work and intelligent. I am very happy to study with him.

Third, I am very thankful to Dr. Park. He gave me GALNT6 project and helped me a lot in the University of Chicago. He is friendly, and would like to share many things with me. He also gave me a lot of suggestions about my research.

I need to say thank you to every team member in the lab of the University of Tokyo and the University of Chicago. Without their support, my research cannot finish so smoothly.

For my first project, I would like to thank every author in my paper. Dr. Deng gave me a lot of help on the project. We discussed a lot and learned from each other. Dr. Tanikawa provided me detail protocols

and gave me useful suggestions. I thank Ryuji Hamamoto, Daechun Kang and Takashi Fujitomo for technological support of IHC. I also thank Jinichi Mori for making cDNA template after Adriamycin treatment. I thank Satomi Takahashi for providing general reagents, materials and sequencing.

For my second project, I would like to thank Hisayo, the secretary of the lab. She helped me to solve many problems not only in research but also in live. Mariann, the manager of the lab, she kindly provided me everything I need in the research. Houda, a young scientist, she gave me a lot of useful suggestions about research and career. I am very thankful to Xiao, she kindly invited me to join her promising project and I learned a lot from her. I thank Ryuji Hamamoto and his team members. They provided me materials and protocols. I thank Hua, PohYin, Kailee, Miran and Vassiliki Saloura for their support and friendship. I also thank Martin for his kindness and hard work, which encourage me all the time.

Last, I would like to deeply thank my family. Without their support, I cannot finish my PhD degree; without their support, I cannot open my mind to do what I want. They gave me powder and happiness. They always love me, no matter what I am.