

Doctoral Thesis

博士論文

Comprehensive analysis of p53 signaling pathway in cancer

(がんにおける p53 シグナル経路の網羅的解析)

リュウ ユユ

Liu Yuyu

Graduate School of Frontier Sciences

The University of Tokyo

CONTENTS

ABSTRACT	1
INTRODUCTION	3
CHAPTER I IDENTIFICATION OF <i>INKA2</i> AS NOVEL P53 TARGET	9
SUMMARY	9
MATERIALS AND METHODS	11
<i>RNA sequencing of 24 mouse tissues</i>	11
<i>cDNA microarray</i>	12
<i>Screening strategy</i>	12
<i>Cell lines and treatment</i>	13
<i>RNA extraction and RT-qPCR</i>	14
<i>Immunocytochemistry</i>	14
<i>Binding site analysis</i>	14
<i>Protein extraction and immunoblotting</i>	15
<i>Reporter assay</i>	15
<i>Colony formation assay</i>	16
<i>Immunoprecipitation and deletions in over-expression constructs</i>	16
<i>Mass spectrometry</i>	17
<i>TCGA analysis</i>	17
RESULTS	19
1. <i>Screening for p53 downstream target</i>	19
2. <i>INKA2 expression is induced in a p53-dependent manner</i>	24
3. <i>INKA2 is a direct p53 downstream target</i>	29
4. <i>INKA2 shows tumor suppressive function</i>	38

5. <i>INKA2 interacts with PAK4</i>	40
6. <i>INKA2 down-modulates β-catenin at protein level</i>	45
DISCUSSION	48
CHAPTER II CHARACTERIZATION OF GENOME-WIDE CANCER ASSOCIATED SINGLE NUCLEOTIDE VARIANT IN P53 BINDING SITE	50
SUMMARY	50
MATERIALS AND METHODS	51
<i>ReMap 2018 database</i> ¹⁸	51
<i>JASPAR database</i> ²⁰	51
<i>p53 responsive element search algorithm</i> ^{40,41}	51
<i>Post imputation GWAS datasets</i>	52
<i>Reporter assay</i>	52
<i>eQTL data</i>	52
<i>Cell culture and treatment</i>	53
<i>RNA extraction and RT-qPCR</i>	53
RESULTS	54
1. <i>Screening strategies</i>	54
2. <i>Confirmation of p53 binding to identified responsive elements</i>	65
3. <i>Gene expression analysis of candidates possibly regulated by the obtained SNPs</i>	67
DISCUSSION	72
CONCLUSION	74
REFERENCES	75
APPENDIX	81
ACKNOWLEDGEMENT	91

FIGURE INDEX

Figure I. Distribution of p53 mutations in data obtained from The Cancer Genome Atlas (TCGA)	4
Figure II. p53 mediated tumor suppression ³	5
Figure III. ESR1 Binding at the WNT4 Locus ¹³	8
Figure 1.1 Screening for identification of novel p53 downstream targets from mouse transcriptome data	21
Figure 1.2 Expression level of candidate genes in human transcriptome data	22
Figure 1.3 Expression level of <i>Inka2</i> in mouse transcriptome data	23
Figure 1.4 <i>INKA2</i> isoform 2 exhibits low expression	25
Figure 1.5 <i>INKA2</i> is a p53 downstream target	26
Figure 1.6 Evaluation of the specificity of the antibody against <i>INKA2</i>	27
Figure 1.7 mRNA expression of <i>INKA2</i> in TCGA	28
Figure 1.8 TP53 ChIP-sequencing data showing 3 possible p53 binding regions throughout <i>INKA2</i> gene	31
Figure 1.9 <i>INKA2/Inka2</i> gene contains an intronic p53 binding site	32
Figure 1.10 Data of ChIP-seq targeting Trp53 showing p53 binding region at the promoter region of mouse <i>Inka2</i>	34
Figure 1.11 The p53 binding site identified in <i>INKA2/Inka2</i> is conserved across species	36
Figure 1.12 <i>INKA2</i> is a direct target of p53	37
Figure 1.13 <i>INKA2</i> functions as a tumor suppressor	39
Figure 1.14 <i>INKA2</i> conserves the iBox domain containing PLV tripeptide	42

Figure 1.15 INKA2 binds to the catalytic domain of PAK4 via iBox domain.....	43
Figure 1.16 INKA2 co-localizes with PAK4 in the nucleus.....	44
Figure 1.17 INKA2 down-modulates intracellular β -catenin.....	46
Figure 1.18 The regulation of cancer cell growth through the p53-INKA2-PAK4 pathway.....	47
Figure 2.1 Overview of screening strategy 1.....	57
Figure 2.2 Each candidate variant overlaps with not less than one study from the p53 ChIP-seq dataset	60
Figure 2.3 Binding sites predicted by JASPAR database for p53 responsive elements containing SNPs obtained from screening strategy 1.....	61
Figure 2.4 Overview of screening strategy 2.....	62
Figure 2.5 p53 binding sites obtained with strategy 2.....	64
Figure 2.6 The risk alleles of SNP 2 and SNP 3 are associated with a higher p53 binding affinity	66
Figure 2.7 Overview of the genomic regions at a distance of within 100 kb from the candidate variants.....	69
Figure 2.8 Gene expression analysis of genes possibly regulated by SNP 3.....	70
Figure 2.9 Evaluation of p53-dependent expression of the candidate genes.....	71
Figure 3 Number of genes satisfying the screening criteria in each mouse organ	81

TABLE INDEX

Table 1. Genomic locations of p53 ChIP-seq peaks from ReMap 2018 database and <i>INKA2/FAM212B</i> binding site.....	33
Table 2. Genomic locations of p53 ChIP-seq peaks from ChIPBase database and Inka2/Fam212b binding site.....	35
Table 3. Mass spectrometry results for cell lysates of HEK293T cells over-expressing mock or INKA2-HA plasmid after immunoprecipitation with antibody against HA-tag.....	41
Table 4. Data of ChIP-seq targeting p53 protein gathered from ReMap 2018 database.....	56
Table 5. Results of screening strategy 1 identifying cancer associated variants overlapping with p53 ChIP-seq data.....	58
Table 6. The association results between the variants obtained from the two screening strategies and each cancer type.....	59
Table 7. Results of screening strategy 2 identifying cancer associated variants overlapping with p53 ChIP-seq data and predicted p53 responsive elements.....	63
Table A1. Oligonucleotide sequences and primers used in chapter I.....	82
Table A2. Primary antibodies used for immunoblotting and immunocytochemistry.....	83
Table A3. List of genes obtained from a screening with an induction over three-fold in WX group in more than five mouse organs.....	84
Table A4. TCGA samples.....	87
Table A5. Post imputation GWAS datasets.....	88
Table A6. Oligonucleotides and primers used in chapter II.....	89

ABSTRACT

p53 is a tumor suppressive transcription factor that regulates the expression of target genes involved in numerous stress responses systems. In this thesis, we carried out a comprehensive analysis of the p53 signaling pathway through two different ways.

First, we identified novel p53 downstream target genes using two transcriptome datasets, including RNA sequencing analysis of twenty-four mouse tissues, and cDNA microarrays of human HCT116 colorectal carcinoma cells and MCF10A breast epithelial cells. We irradiated mice with 10 Gy of X-rays to induce p53 expression and identified 18 genes that were up-regulated in response to DNA damage along with p53. Among these genes, only five of them were unreported protein coding candidate genes with human homolog. Further analysis of the expression levels of the candidate genes in human cells treated with Adriamycin (ADR) identified *INKA2* as the only gene showing a p53-dependent expression in mouse tissues and human cells. Genomic analysis and reporter assay revealed an intronic p53 binding site in *INKA2* gene, suggesting that *INKA2* is regulated by p53 directly.

Next, we elucidated the relationship between *INKA2* and cancer. Analysis of The Cancer Genome Atlas data reveals decreased *INKA2* expression in tumor samples carrying *p53* mutations compared to samples with wild-type *p53*. In addition, significantly higher levels of DNA methylation are observed in the *INKA2* promoter in tumor samples, concordant with the reduced *INKA2* expression in tumor tissues. Furthermore, over-expression of *INKA2* slightly reduced cancer cell growth, demonstrating that *INKA2* might function as a tumor suppressor. We also showed evidence that *INKA2* protein interacts with the serine/threonine-protein kinase PAK4, which phosphorylates β -catenin to prevent ubiquitin-proteasomal degradation. As β -catenin is down-regulated in cell line stably expressing *INKA2*, our results suggest that *INKA2* is a novel, direct downstream target of p53 that potentially decreases cell growth by inhibiting the PAK4- β -catenin pathway.

In the second chapter, we identified genome-wide cancer associated single nucleotide polymorphisms (SNPs) that affect p53 binding efficiency to its downstream target genes. To screen functional p53 binding sites in the genome, we integrated three genome-wide comprehensive

datasets: 1) ChIP-sequencing data targeting p53 protein, 2) in silico p53 binding motif models, and 3) post imputation GWAS datasets for 14 cancer types. We designed two strategies identifying SNPs having the potential to modify p53's binding efficiency. Consequently, we obtained three SNPs significantly associated with lung cancer, prostate cancer, and stomach cancer, respectively. Reporter assay revealed that SNP 2 and SNP 3 showed higher luciferase activity with the risk allele as compared to the alternate allele. Gene expression analysis revealed that the risk allele of SNP 2 is associated with a higher expression level of genes *X*, *Y* and *Z* via a higher binding affinity of p53 to the binding site containing SNP 2. Moreover, the expression of genes *X*, *Y* and *Z* were induced by ADR treatment in a p53-dependent manner. In conclusion, we identified cancer associated functional SNPs that might explain high risk of cancer through the modification of the expression of p53 regulated genes.

Taken together, our comprehensive analysis elucidated novel molecular mechanisms of the p53 signaling pathway in response to DNA damage. Meanwhile, we will determine other molecular mechanisms underlying significant heterogeneity and the relationship to cancer risk.

INTRODUCTION

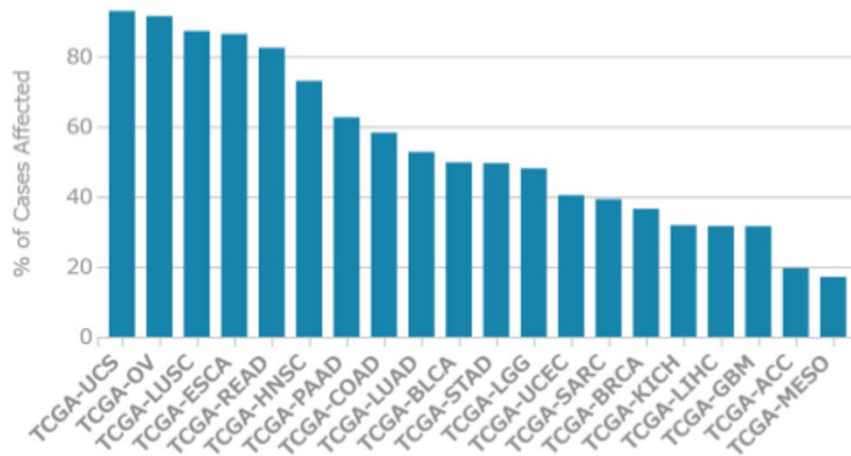
Tumor suppressor p53

The *p53* gene is essential for multicellular organisms as it functions as a tumor suppressor, thus preventing cancer formation. The gene encodes a transcription factor that regulates numerous target genes involved in DNA repair, apoptosis, cell cycle arrest, differentiation, and other responses ^{1,2}. *p53* is as well one of the most frequently altered tumor suppressor gene among human cancers (Figure I.A). Mutations in the gene are associated with a variety of tumor types. The majority of the oncogenic point mutations within the *p53* gene are located in the DNA-binding domain (Figure I.B), revealing the importance of the protein's transcription factor function.

The downstream targets of *p53* are as well involved in different stress responses mechanisms (Figure II ³). For example, the cyclin-dependent kinase (CDK) inhibitor p21 (*CDKN1A*) is one of the first *p53* target genes to be discovered ⁴. As the name indicates, p21 inhibits the genes regulating the cell cycle, mainly the CDKs required for transition from G1 to S phase. Many of the DNA repair proteins are cell cycle-regulated and repressed by the *p53*-p21 pathway in normal conditions ⁵. The activation of *p53* in response to DNA damage triggers the activation of the DNA repair proteins through the same pathway.

The *p53* protein interacts with DNA in a tetramer form where each dimer binds to one half-site on the DNA. Subsequently, the consensus sequence of the *p53* responsive elements is constituted of two half-sites. Although various *p53* binding site models exist, the majority suggest a binding site containing two half-sites, often decamers, with each of the half-site carrying an essential motif analogous to CATG. The *p53* protein is also known to bind at numerous locations in the genome, though not all of the bindings are followed by transcription changes ⁶.

A



B

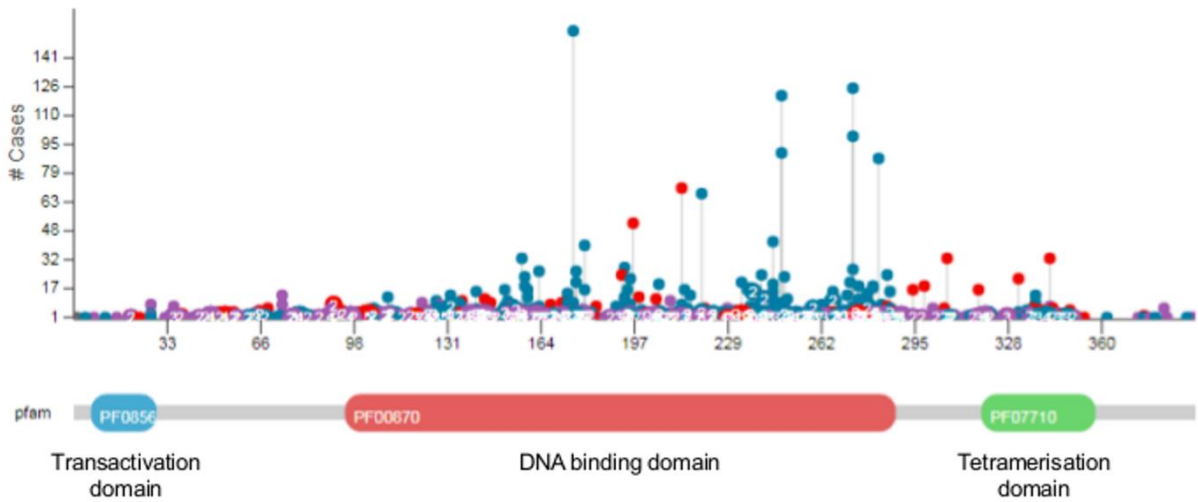


Figure I. Distribution of p53 mutations in data obtained from The Cancer Genome Atlas (TCGA)

A) Distribution of p53 mutations in tumor samples from TCGA data including 4,008 cases affected by 1,291 p53 mutations across 32 cancer projects.

B) Distribution of oncogenic mutations within *p53*.

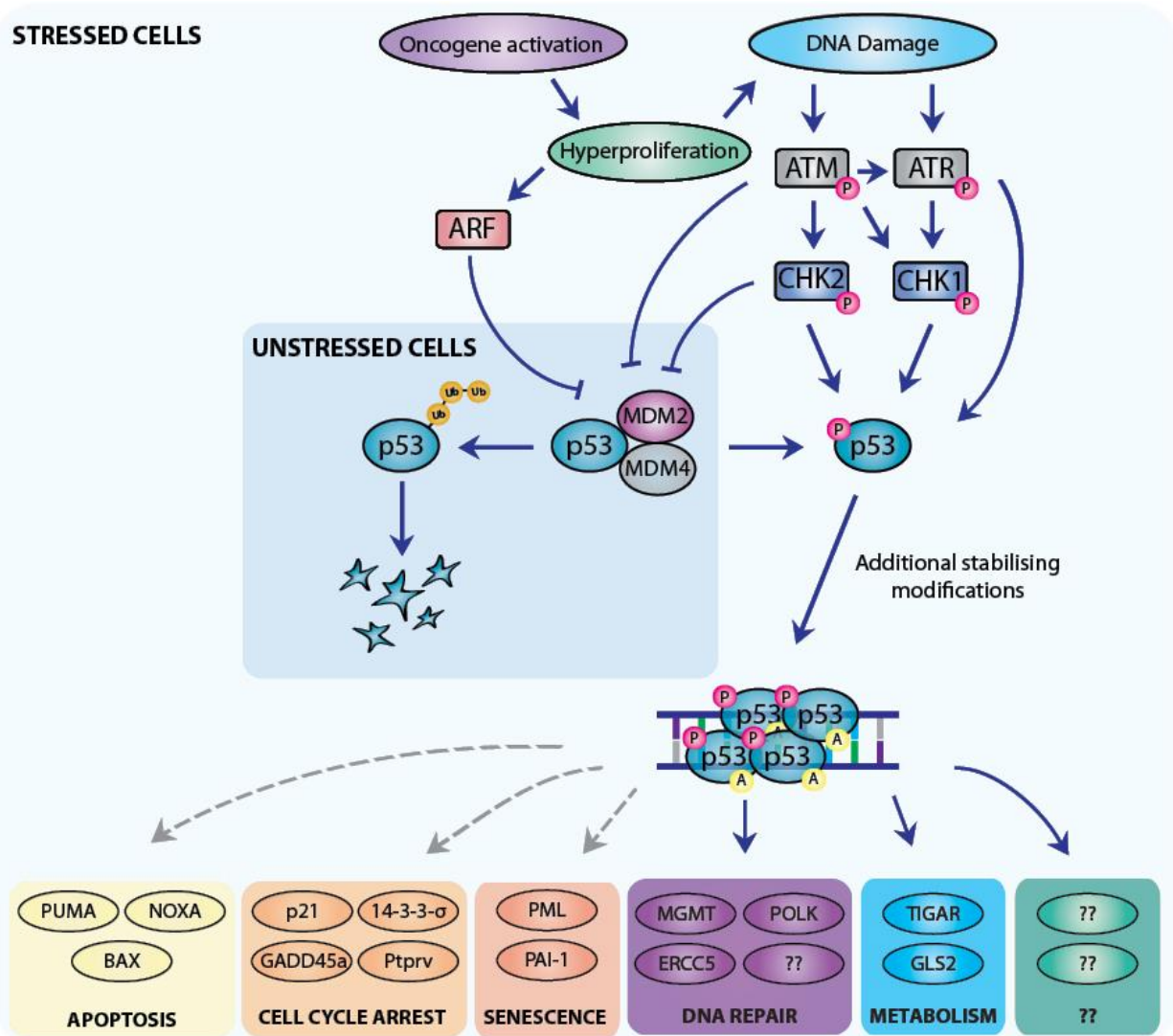


Figure II. p53 mediated tumor suppression ³

The transcription factor p53 imposes a critical barrier against tumorigenesis. Almost all human cancers bear genetic alterations that either, directly or indirectly, render p53 non-functional. The activity of p53 can be stimulated by the activation of certain oncogenes. Crosstalk between DNA damage-activated and oncogene-induced p53 stabilization can occur when hyper-proliferation-induced replicative stress promotes DNA damage. Binding as a homo-tetramer to specific DNA sequences, p53 mediates induction (and for some targets repression) of a wide range of specific target genes, which orchestrate activation of distinct effector processes. The ability of p53 to modulate under-appreciated processes, such as DNA repair or control of metabolism, or perhaps control of presently unknown processes may be critical for p53-mediated tumor suppression.

Inka box actin regulator – INKA

In the first part of this thesis, we identified *INKA2* as a novel p53 downstream target. The inka box actin regulator (INKA) gene family is evolutionarily conserved among vertebrates and comprises two genes *INKA1* and *INKA2*. Both genes share an iBox, or inca-box domain, with an average sequence conservation of 35-40% across species. The *Inka1* gene has been studied for mouse, zebrafish, and *Xenopus* embryogenesis and is expressed specifically in the neural crest cells ⁷. In *Xenopus* model, cytoskeletal dynamics are regulated through the interaction of *inka1* with *pak5* ⁷. Knockdown of *inka1* causes severe defects in both *Xenopus* and zebrafish embryos, suggesting its essential role in development⁷. In mouse model, *Inka1* gene affects neural tube closure and the absence of the protein is able to cause exencephaly ⁸. On the other hand, *Inka2* is expressed in the murine nervous system and is susceptible to be involved in actin cytoskeleton regulation and actin-driven processes ⁹. Many studies suggested that *inka1* interact with p21-activated kinases (PAKs) across different species ^{7,10}. These kinases are reported to be involved in the regulation of cell morphology, cell polarity, and cell motility ^{11,12}. However, the function of INKA proteins and the consequences of the interaction between INKA proteins and PAK kinases have not been investigated either in human cells or in cancer cases.

Cancer associated SNPs affecting transcription factor binding efficiency

In the second part of this thesis, we aim to identify cancer associated single nucleotide polymorphisms (SNPs) within p53 binding motif that affect the binding of p53 protein as a transcription factor. Genome-wide association study (GWAS) is a well-known method for screening genetic variations that are associated with various phenotype including disease risks. Subsequently, SNPs could be considered as genetic biomarkers if associated with cancer risk. The variation of a single nucleotide increases the risk of diseases through diverse mechanisms. Particularly, a SNP located within a binding site of a transcription factor might give consequences in the binding efficiency of the transcription factor, thus altering the expression of regulated genes. As example, a GWAS on gestational duration discovered the association of the *WNT4* loci with the phenotype in question¹³. The estrogen receptor 1 binds at the *WNT4* estrogen receptor binding site located in the first intron of *WNT4* to induce the expression of the gene^{13,14}. GWAS results identified a SNP (rs3820282) within the *WNT4* estrogen receptor binding site where the thymidine (T) allele favors the binding of estrogen receptor 1 compared to the cytosine (C) allele (Figure III¹³). Although the subsequent transcriptional changes have not been demonstrated experimentally, such alteration should be expected. In our case, we screened functional p53 binding sites that are associated with cancer risk by using three comprehensive omics datasets.



Figure III. ESR1 Binding at the WNT4 Locus ¹³

(A) The T allele of rs3820282 creates a strong binding site for estrogen receptor 1 (ESR1). Tall nucleotides indicate DNA bases that are preferred (above the dashed line) or disfavored (below the dashed line) by ESR1. The y axis indicates the relative free energies of binding for each nucleotide at each position. The height of each nucleotide represents the free energy difference from the average ($\Delta\Delta G$) in units of gas constant (R) and temperature (T). The WNT4 promoter sequence is shown directly below the x axis. The UCSC Genome Browser screen shot depicted that rs3820282 overlaps strong signals of transposase-accessible chromatin with high-throughput sequencing (ATAC seq, indicating chromatin accessibility) and H3K4me3 (active promoter) signals in decidual stromal cells at the WNT4 locus. The red vertical line indicates the position of rs3820282. The purple graphic below the screen shot indicates the locations of the WNT4 exons (columns), untranslated regions (rectangles), and introns (horizontal lines), with the arrows indicating the direction of transcription.

(B) Experimental validation of allele-dependent binding of ESR1 to rs3820282 on electrophoretic mobility shift assay (EMSA). The top arrow points out the “supershift” of the protein-DNA complex induced by the binding of the ESR1 antibody to the complex. The bottom arrow indicates an allele-dependent binding of ESR1.

CHAPTER I

IDENTIFICATION OF *INKA2* AS NOVEL P53 TARGET

SUMMARY

Although a large number of p53 target genes have been reported, none of the genes have been identified with a strategy covering multiple organs and various species. Therefore, we aim to find novel downstream target of p53 with our original screening strategy using multi-organ transcriptome data. Our strategy used transcriptome data obtained from mouse model and human cell lines under DNA damage treatment.

RNA sequencing (RNA-seq) was performed with four groups of mice: *p53* wild-type mice (W) or *p53* knockout mice (K), treated with (WX and KX) or without 10 Gy whole-body X-ray irradiation. For each mouse, total RNA from 24 tissues were subjected for RNA-seq analysis. Among 23,813 genes, we screened out transcripts with low expression and we selected those with a significant up-regulated expression of at least 3-fold in WX group. Eighteen genes fulfilled our criteria. Among them, ten genes were previously reported as p53 downstream targets. Only five genes were unreported protein coding p53 regulated gene candidates with human homolog. We then examined cDNA microarray data in human cells to confirmed the p53-dependent expression of these candidate genes. We selected genes with a significant up-regulated expression of more than 1.5-fold in the *p53* wild-type DNA-damage induced cell group for further investigation. *INKA2*, also known as *FAM212B*, is the only candidate gene that satisfied all criteria in above screening steps. We pursued our analysis with reporter assay and elucidated that *INKA2* is regulated directly by p53 through an intronic p53 binding motif.

Analysis of TCGA data showed that *INKA2* expression is decreased in tumor samples compared to normal samples, and also decreased in p53 mutant samples compared to p53 wild-type samples. In addition, tumor samples from different cancers exhibited higher DNA methylation at the promoter region of *INKA2* than normal samples. Furthermore, over-expression of *INKA2* mildly decreased cancer cell growth. Taken together, we concluded that *INKA2* might have a tumor-suppressive function.

We also showed evidence that INKA2 interacts with the serine/threonine-protein kinase PAK4. As consequence, we observed down-modulation of β -catenin in INKA2 stable cell line. Our results suggest that INKA2 diminishes cell growth by repressing the PAK4- β -catenin pathway. In conclusion, we revealed that INKA2 is a novel direct p53 downstream target that is potentially able to abate cell growth via the inhibition of PAK4.

MATERIALS AND METHODS

RNA sequencing of 24 mouse tissues

C57BL/6J $p53^{-/-}$ mice were provided by RIKEN BioResource Center (Ibaragi, Japan) (Tsukada et al., 1993). The 5' part of exon 2 including translation initiation site was replaced by Neomycin resistance gene (Neor). All mice were maintained under specific pathogen-free conditions and handled in accordance with the Guidelines for Animal Experiments of the Institute of Medical Science (University of Tokyo, Tokyo, Japan). $p53^{+/+}$ and $p53^{-/-}$ mice were exposed to 10Gy of X-ray using the MBR-1520R-3 system (Hitachi, Tokyo, Japan). Twenty-four mouse tissues were collected by dissection at 24 hours after whole-body irradiation. The age and gender of mice were as follows: bladder, bone marrow, cerebrum, colon, esophagus, eyeball, heart, kidney, liver, Lung, Muscle, Seminal vesicle, Small intestine, Spleen, Stomach, Testis, Thymus, and Tongue: male, 6 weeks, $n = 3$ each, Bone: male, 1 week, $n = 3$ each, Uterus: female, 10 weeks, $n = 3$ each, Mammary gland and Ovary: female, 10 weeks, $n = 2$ each. The obtained tissues were preserved in RNAlater solution (QIAGEN) at 4 °C until RNA purification. Bone marrow was resolved in RLT plus reagent provided by the RNeasy Plus Mini Kit (QIAGEN) and homogenized using a QIAshredder column (QIAGEN). Tissues were homogenized in QIAzol lysis reagent (QIAGEN) using Precellys 24 (Bertin Corporation). The lysates were stored at -80°C until RNA purification. Total RNA was recovered using the RNeasy Plus Universal Mini Kit (QIAGEN). For RNA extraction from bone marrow, we used the RNeasy Plus Mini Kit (QIAGEN). We selected 256 samples for RNA-seq analysis based on RNA quality and quantity, which were evaluated using a Bioanalyzer (Agilent) and NanoDrop spectrophotometer (Thermo Scientific). High-quality RNA was subjected to polyA⁺ selection and chemical fragmentation, and a 100-200-base RNA fraction was used to construct complementary DNA libraries according to Illumina's protocol. RNA sequencing (RNA-seq) was performed on a HiSeq 2500 using a standard paired-end 101-bp protocol. We used a tophat+cufflinks pipeline to process raw RNA-seq data. Before data processing, the quality of data was confirmed using FastQC. To quantify gene and transcript expression levels for all samples, we first aligned 101 bp paired-end reads to the mouse reference genome mm9/GRCm37 using Tophat (v2.0.9). The mapping parameters follow the default setting in the Tophat. After read mapping, the transcript and gene expression levels, which are represented by FPKM values, were calculated by Cufflinks (v2.2.1). The raw data for the RNA-seq are

available in the DDBJ database (<http://www.ddbj.nig.ac.jp/index-e.html>) under the accession number DRA005768, and bioproject accession number PRJDB5738. Full details for the experiment are described in previous studies ¹⁵.

cDNA microarray

MCF10A $p53^{+/+}$ cells and MCF10A $p53^{-/-}$ cells (breast epithelial cells) were purchased from Sigma-Aldrich and were cultured in MEGM mammary epithelial cell growth medium (LONZA) supplemented with 100 ng/ml cholera toxin at 37°C in 5% CO₂. HCT116 $p53^{+/+}$ and HCT116 $p53^{-/-}$ cells (colorectal adenocarcinoma) were gifts from Dr. B. Vogelstein (Johns Hopkins University, Baltimore, MD, USA) and were cultured in McCoy's 5a medium (Gibco) supplemented with 10% FBS and 1% penicillin/streptomycin at 37°C in 5% CO₂. MCF10A and HCT116 $p53^{+/+}$ or $p53^{-/-}$ cells were treated with ADR for two hours and cultured with fresh medium until harvest. The concentration of applied ADR was 0.5 µg/mL and 2 µg/mL for MCF10A and HCT116 cells, respectively. Total RNA was isolated from the cells using the RNeasy Plus Universal Mini Kit (Qiagen, Valencia, CA, USA) at 0, 12, 24, and 48 hours after ADR treatment according to the manufacturer's instructions. Each RNA sample was labeled and hybridized to array slides. Gene expression analysis was performed using SurePrint G3 Human GE 8×60K microarray (Agilent, Santa Clara, CA, USA) according to the manufacturer's protocol. The accession numbers are GSE98727 for MCF10A cells, and GSE125787 for HCT116 cells. The same data have been used for previous studies ^{16,17}.

Screening strategy

In the mouse transcriptome data, low expression genes were defined with maximum expression in all groups (W, $p53^{+/+}$; WX, irradiated $p53^{+/+}$; K, $p53^{-/-}$; KX, irradiated $p53^{-/-}$) among all twenty-four tissues inferior to 1 FPKM and were excluded from the screening. Then, the expression of WX mice (n = 2 for the mammary gland, ovary and uterus, n = 3 for other organs) was compared to the expression of W, K, and KX mice (same n as WX for each group, resulting in a total of n = 6 for the mammary gland, ovary and uterus, n = 9 for other organs). The expression fold change was calculated with the following formula: median expression in [WX] / maximum of median expression in [W], [K], or [KX]. Genes with a significant expression fold change higher than three-

fold were extracted. The two-tailed Student's *t*-test was performed to calculate p-values. In the final step of the screening, genes with significant up-regulated expression in the WX group in no less than ten of twenty-four tissues were selected. The number of genes satisfying the screening criteria in each organ (possibly tissue-specific p53 downstream targets) are shown in Appendix Figure 3. In the human cell transcriptome data, the expression of ADR-treated $p53^{+/+}$ cells ($n = 3$) was compared to the group ($n=5$) gathering $p53^{-/-}$ cells ($n = 4$) and non-treated $p53^{+/+}$ cells ($n = 1$). The expression fold change was calculated by the following formula: median expression of $p53^{+/+}$ cells [ADR 12 h], [ADR 24 h], and [ADR 48 h] / maximum expression in $p53^{-/-}$ cells [0 h], [ADR 12 h], [ADR 24 h], [ADR 48 h], and $p53^{+/+}$ cells [0 h]. Genes with up-regulated expression in ADR-treated $p53^{+/+}$ cells were defined by a significant expression fold change higher than 1.5-fold. The two-tailed Student's *t*-test was used to calculate the p-values.

Cell lines and treatment

HCT116 (colorectal adenocarcinoma), HBL-100 (breast cancer), U2OS (osteosarcoma), H1299 (lung adenocarcinoma), and U373MG (glioma) cells were purchased from American Type Cell Collection (Rockville, MD, USA). HCT116 $p53^{+/+}$ and HCT116 $p53^{-/-}$ cells were gifts from Dr. B. Vogelstein (Johns Hopkins University, Baltimore, MD, USA). MCF10A (breast epithelial cells) $p53^{+/+}$ and $p53^{-/-}$ cell lines were purchased from Sigma Aldrich (St. Louis, MO, USA). HBC4 (breast cancer) cells were a gift from Dr. T. Yamori (The Cancer Institute of Japanese Foundation for Cancer Research, Tokyo, Japan). HEK293T cells (subclone of HEK293 cells expressing SV40 large T-antigen) were purchased from the RIKEN Cell Bank (Ibaraki, Japan). All cells were cultured in 37°C incubator with 5% CO₂ atmosphere. The siRNA oligonucleotides were ordered from Sigma-Aldrich (St. Louis, MO, USA) and were transfected into the cells with Lipofectamine RNAiMAX (Thermo Fisher Scientific, Waltham, MA, USA) at the seeding step. The sequences are listed in Appendix Table A1. Gene over-expression experiments were performed by transfecting plasmids into cells with FuGENE6 (Promega, Madison, WI, USA). Cells were treated with ADR for two hours at the 24-hour time point after seeding to induce DNA damage and p53 expression. U2OS and HBC4 cells were treated with 2 µg/mL ADR and HBL-100 cells were treated with 1 µg/mL ADR. Total RNA was extracted or whole cell lysates were prepared 48 hours after treatment for RT-qPCR or western blot analysis, respectively. Cell lines stably expressing INKA2 were generated by transfecting HCT116 cells with pLenti6.3/TO/V5-DEST/INKA2-CDS

containing the full-length *INKA2* cDNA fragment as previously described ¹⁸. The plasmid contains the selectable marker G418 and Blasticidin S-resistance gene. Transfected cells were selected by adding Blasticidin and Geneticin® to the culture media for 3 weeks.

RNA extraction and RT-qPCR

After treatments, total RNA was extracted from cells using an RNeasy Plus Mini Kit (Qiagen, Valencia, CA, USA) according to the manufacturer's protocol. SuperScript III reverse transcriptase (Invitrogen, Carlsbad, CA, USA) was used to synthesize cDNA from collected RNA samples. The resulting cDNA were employed as templates for qPCR using SYBR Green Master Mix and Light Cycler 480 machine (Roche Applied Science, Basel, Switzerland). The expression level of the housekeeping gene GAPDH was used for normalization. The primer sequences are shown in Appendix Table A1.

Immunocytochemistry

Cells were plated on 24-well culture plates with glass coverslips on the bottom. Cells were transfected with siRNAs at the same time as the seeding step. For drugs treatments, ADR or Dox were added 24 hours after plating. The ADR treatment lasted 2 hours and Dox remained in the media until cells were harvested. Forty-eight hours after adding the drugs, cells were fixed with 4% paraformaldehyde and were permeabilized with 0.2% Triton X-100 in PBS. A 3% BSA solution was applied in the blocking step and the cells were incubated with antibodies in the same blocking solution. We diluted and incubated the antibodies according to each manufacturer's protocol. Details for antibodies used in the present study are listed in Appendix Table A2. Finally, the samples were counterstained with DAPI. A confocal microscope (Olympus FluoView FV1000) was used to visualize the mounted coverslips.

Binding site analysis

ChIP data targeting human and mouse p53 proteins were downloaded from the ReMap 2018 ¹⁹ and ChIPBase ²⁰ databases, respectively. Overlapping peaks were merged into a single peak for analysis. The DNA sequence from -10 kb to +10 kb of the *INKA2* gene was submitted to

the binding motif scanning tool in JASPAR database²¹ to identify the p53 binding site. The binding scores were calculated and provided by the database.

Protein extraction and immunoblotting

Whole cell lysates were prepared from adherent cells with RIPA buffer (Thermo Fisher Scientific, Waltham, MA, USA) containing 1 mM PMSF, 0.1 mM DTT, and 0.1% Protease Inhibitor Cocktail Set III (Calbiochem, La Jolla, CA, USA). Cell lysates were sonicated 15 times for 30 seconds each using a Bioruptor UCD-200 (Cosmo Bio, Tokyo, Japan). The protein concentrations of the resulting lysates were measured using the BCATM protein assay (Thermo Fisher Scientific, Waltham, MA, USA). Protein samples were separated using SDS-polyacrylamide gel electrophoresis and transferred to nitrocellulose membranes (HybondTM ECLTM, Amersham). For the blocking step, we incubated the membranes with 5% nonfat dry milk in 0.05% or 0.1% Tween-20 TBST, as recommended by each antibody's manufacturer, for one hour at room temperature. After washes, proteins were probed with specific primary antibodies overnight at 4°C. Details for the antibodies used in this study are provided in Appendix Table A2. On the next day, membranes were incubated with horseradish peroxidase (HRP)-conjugated secondary antibodies (Santa Cruz Biotechnology, Santa Cruz, CA, USA) for one hour at room temperature. Proteins were detected using chemiluminescence (ECL, AmershamTM, or Immobilon, Millipore) with LAS-4000 mini machine (Fujifilm, GE Healthcare).

Reporter assay

DNA fragments carrying p53-responsive elements, as predicted by the JASPAR database, were amplified by PCR and were cloned into the pGL4.24 [*luc2P*/minP] plasmid (Promega, Madison, WI, USA) (see Appendix Table A1 for primers). The known p53 binding site in the promoter region of p21 was used as positive control for human binding site constructs: GAACATGTCCCAACATGTTG. For point mutations, the 6th, 13th, and 16th nucleotides of the binding site were changed to T using site-directed mutagenesis. H1299 and U373MG cells were co-transfected with pGL4.24 control vector or pGL4.24 containing the target p53 binding site, accompanied by the empty vector pcDNA3.1, wild-type p53, or mutant p53 (p53 R175H for human p53 and p53 R172H for mouse p53) expression vector. All samples were co-transfected

with pGL4.74 to normalize the transfection efficiency. Luciferase activities were detected 48 hours after transfection using the Dual-Glo Luciferase Assay System (Promega, Madison, WI, USA) according to the manufacturer's protocol.

Colony formation assay

The full-length *INKA2* cDNA was cloned into the PCAGGS vector containing an HA-tag (pCAGGSnHC) (see Appendix Table A1 for primers). Cells were seeded (densities of 10,000 cells/well and 5,000 cells/well for HBC4 and U2OS cells, respectively) and grown in 6-well culture plates. On the next day, PCAGGS/*INKA2*, or the PCAGGS mock plasmid was transfected into cells. U2OS and HBC4 cells were cultured in the presence of 1.0 mg/mL or 0.4 mg/mL Geneticin®, respectively, for one to two weeks. Colonies were counted with ImageJ software ²² after staining with crystal violet (Sigma-Aldrich, St. Louis, MO, USA). Two-tailed Student's *t*-test was used to calculate the p-values for the differences between mock and *INKA2*-expressing replicates (n=3).

Immunoprecipitation and deletions in over-expression constructs

The full-length encoding *INKA2* and *PAK4* cDNAs were cloned into PCAGGS vectors containing a multiple cloning site and HA-tag (pCAGGSnHC) or FLAG-tag (pCAGGSn3FC) (see Appendix Table A1 for primers). Plasmids were transfected into HEK293T cells to induce the over-expression of the target genes. Whole-cell lysates were collected 48 hours after transfection with a cell lysis buffer consisting of 20mM Tris (pH7.5), 150mM NaCl, 1mM EDTA, 1mM EGTA, 1% Triton X-100, 1mM PMSF, and 0.1% Protease Inhibitor Cocktail Set III (Calbiochem, La Jolla, CA, USA). Samples were sonicated 3 times for 5 seconds each using a Bioruptor UCD-200 (Cosmo Bio, Tokyo, Japan). Forty-five microliters of the supernatant in each sample were aliquoted as input samples. Sonicated cell lysates were precleared by incubation with Protein G-Sepharose 4B (Invitrogen, Carlsbad, CA, USA) and normal mouse IgG for 1 hour at 4°C. Precleared cell lysates were immunoprecipitated with mouse monoclonal anti-HA antibody-conjugated agarose beads or mouse monoclonal anti-FLAG antibody-conjugated agarose beads (Sigma-Aldrich, St. Louis, MO, USA, Cat. Nos. A2095 and A2220, respectively). The obtained samples were washed 4 times with 1 mL of ice-cold cell lysis buffer (without PMSF and Protease

Inhibitor) and then boiled for 2 minutes after the addition of Laemmli Sample Buffer (Bio-Rad Laboratories, Hercules, CA, USA). For deletions, the RNRQPLVLGDN (150-160) residues of the INKA2 iBox domain and the catalytic domain (326-591) of PAK4 were deleted by using site-directed mutagenesis.

Mass spectrometry

The pCAGGS mock plasmid or pCAGGS/INKA2-HA plasmid was transfected into HEK293T on the second day after plating on a 10-cm dish. Whole-cell lysates were collected and immunoprecipitated with an antibody against the HA-tag. The resulting lysates were subjected to mass spectrometry analysis as previously described²³. Briefly, samples were reduced in 10 mM tris(2-carboxyethyl)phosphine (Sigma-Aldrich, St. Louis, MO, USA) with 50 mM ammonium bicarbonate (Sigma-Aldrich, St. Louis, MO, USA) for 30 min at 37°C and alkylated in 50 mM iodoacetamide (Sigma-Aldrich, St. Louis, MO, USA) with 50 mM ammonium bicarbonate for 45 min in the dark at 25°C. Trypsin/Lys-C (Promega, Madison, WI, USA) solution was added and incubated at 37°C for 12 hours. The resulting peptides were extracted from gel fragments and analyzed with Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA) combined with UltiMate 3000 RSLC nano-flow HPLC system (Thermo Fisher Scientific, Waltham, MA, USA) with HCD MS/MS mode. The MS/MS spectra were searched against Homo sapiens protein sequence database in SwissProt using Mascot or Sequest search engine in Proteome Discoverer 2.2 software (Thermo Fisher Scientific, Waltham, MA, USA), in which peptide identification filters were set at “false discovery rate < 1%” and “Mascot expectation value < 0.05 or Sequest XCorr > 2.0”.

TCGA analysis

Whole-transcriptome sequencing data, transcriptome profiling microarray data, and p53 mutation status were obtained from The Cancer Genome Atlas (TCGA) database using cBioPortal^{24,25}. We downloaded and analyzed datasets for 20 cancer cohorts: bladder urothelial carcinoma (BLCA), breast invasive carcinoma (BRCA), cervical and endocervical cancers (CESC), cholangiocarcinoma (CHOL), colorectal adenocarcinoma (COAD/READ), glioma (GBM/LGG), head and neck squamous cell carcinoma (HNSC), pan-kidney cohort (KIPAN),

acute myeloid leukemia (LAML), liver hepatocellular carcinoma (LIHC), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), ovarian serous cystadenocarcinoma (OV), pancreatic adenocarcinoma (PAAD), pheochromocytoma and paraganglioma (PCPG), prostate adenocarcinoma (PRAD), skin cutaneous melanoma (SKCM), stomach adenocarcinoma (STAD), thyroid carcinoma (THCA), and uterine corpus endometrial carcinoma (UCEC). Two-tailed Student's *t*-test was used to calculate the p-value for differences between the compared groups. The number of samples are shown in Appendix Table A4. We excluded the cancer type when the number of samples was not relevant. Data regarding the DNA methylation status of the *INKA2* promoter region were downloaded from the MethHC database ²⁶.

RESULTS

1. Screening for *p53* downstream target

We screened RNA sequencing (RNA-seq) data obtained from four groups of mice, *p53* wild-type mice (W), *p53* knockout mice (K), treated with (WX and KX) or without 10 Gy of whole-body X-ray irradiation, to identify novel *p53* downstream targets. For each mouse, total RNA extracted from 24 tissues was subjected to the RNA-seq analysis¹⁵. Among 23,813 genes, we excluded transcripts with extremely low expression (FPKM < 1) and we selected transcripts displaying significantly up-regulated expression by at least three-fold in ten or more organs of the WX group compared to the W, K, and KX groups (Figure 1.1). Eighteen genes fulfilled our criteria including 10 genes that were previously reported as *p53* downstream targets, 2 non-coding RNAs, 1 gene without any corresponding human homolog, and 5 genes remaining as candidates: *Fam212b* (*Inka2*), *4632434I11Rik* (*Ddias*), *Celf5*, *9030617O03Rik*, and *Cd80*.

We examined the expression of the remaining five candidate genes in human cells using a cDNA microarray^{16,17}. The *p53* wild-type (*p53*^{+/+}) or knockout (*p53*^{-/-}) HCT116 colorectal carcinoma cells and MCF10A breast epithelial cells were treated with Adriamycin (ADR) for two hours. Total RNA was extracted at four time points after the treatment, 0, 12, 24, and 48 hours, and subjected to a cDNA microarray analysis. We considered the *p53*^{-/-} group and untreated samples (0 h) as a cluster. The maximum expression in the cluster was compared to the median expression of the *p53*^{+/+} ADR-treated group (*p53*^{+/+} ADR 12 h, 24 h, and 48 h). We selected genes displaying a significant up-regulation of at least 1.5-fold in the *p53*^{+/+} ADR induction group for further investigation (Figure 1.2). At this screening step, only *INKA2* expression was induced by the ADR treatment in *p53*^{+/+} cells. The expression was significantly increased by 1.61-fold and 2.19-fold in HCT116 and MCF10A cells, respectively.

We obtained *INKA2*, also known as *FAM212B*, as the only gene satisfying all criteria in the screen. A screening with a criterion of induction of three-fold in more than five of twenty-four organs carried out the same results (Appendix Table A3). In mouse transcriptome data, *Inka2* exhibited WX expression that was significantly induced more than three-fold in ten of twenty-four

organs (Figure 1.3). The up-regulation reached 5-fold in colon, mammary gland, and calvaria (skull), 10-fold in the bone marrow and spleen, and 20-fold in the kidney.

Mouse gene expression selection criteria:

- Maximum expression in all organs > 1
- $\frac{\text{Median expression of [WX]}}{\text{Maximum of median expression of [W], [K], or [KX]}} > 3$
with significance ($P < 0.05$)

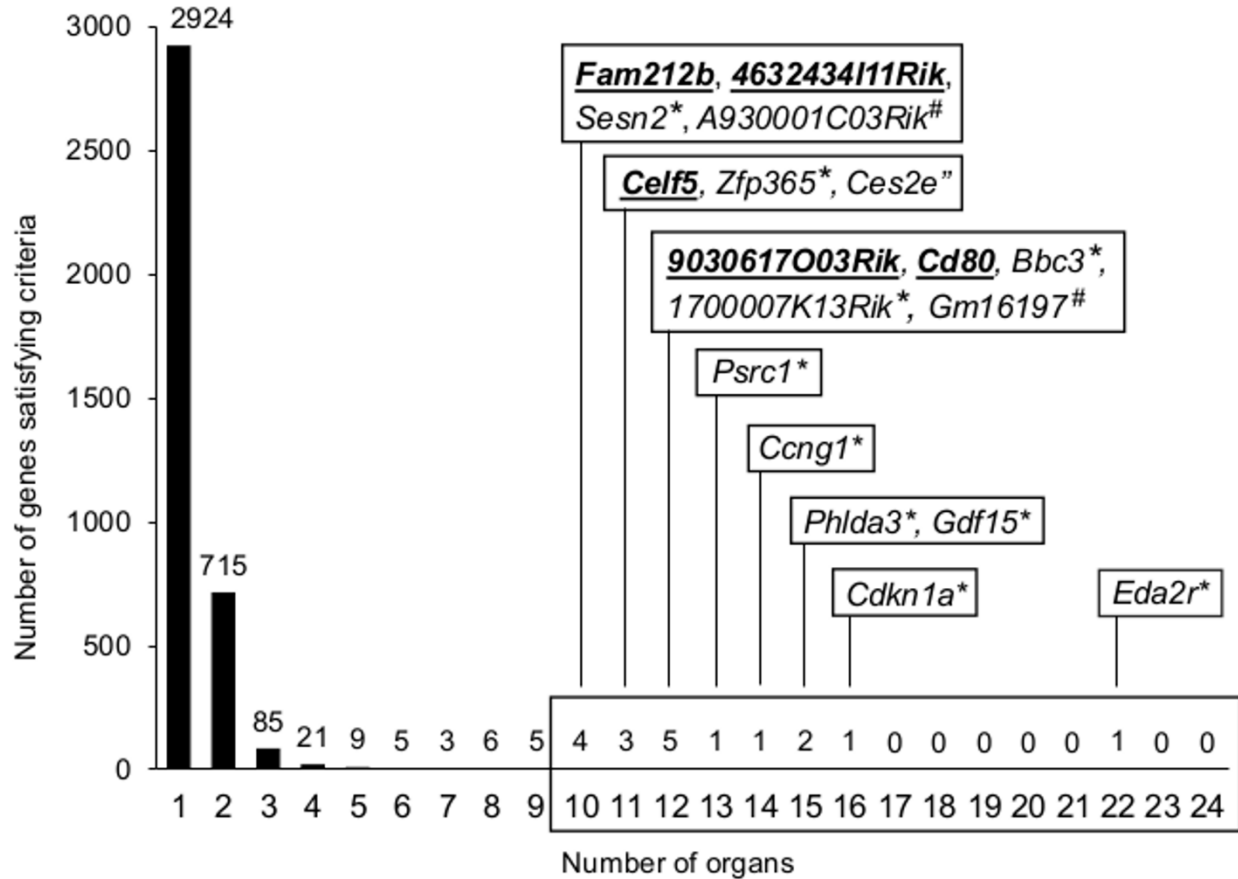


Figure 1.1 Schematic overview of the screening strategy used to identify novel p53 downstream targets from mouse transcriptome data.

The screening criteria are listed at the top. The Y-axis of the bar chart represents the number of genes satisfying the screening criteria. The X-axis plots the number of organs in which the screening criteria were satisfied. Bold and underlined genes are the unreported p53 downstream target candidates. *, reporter p53 downstream targets; #, non-coding RNAs; “, no human homolog.

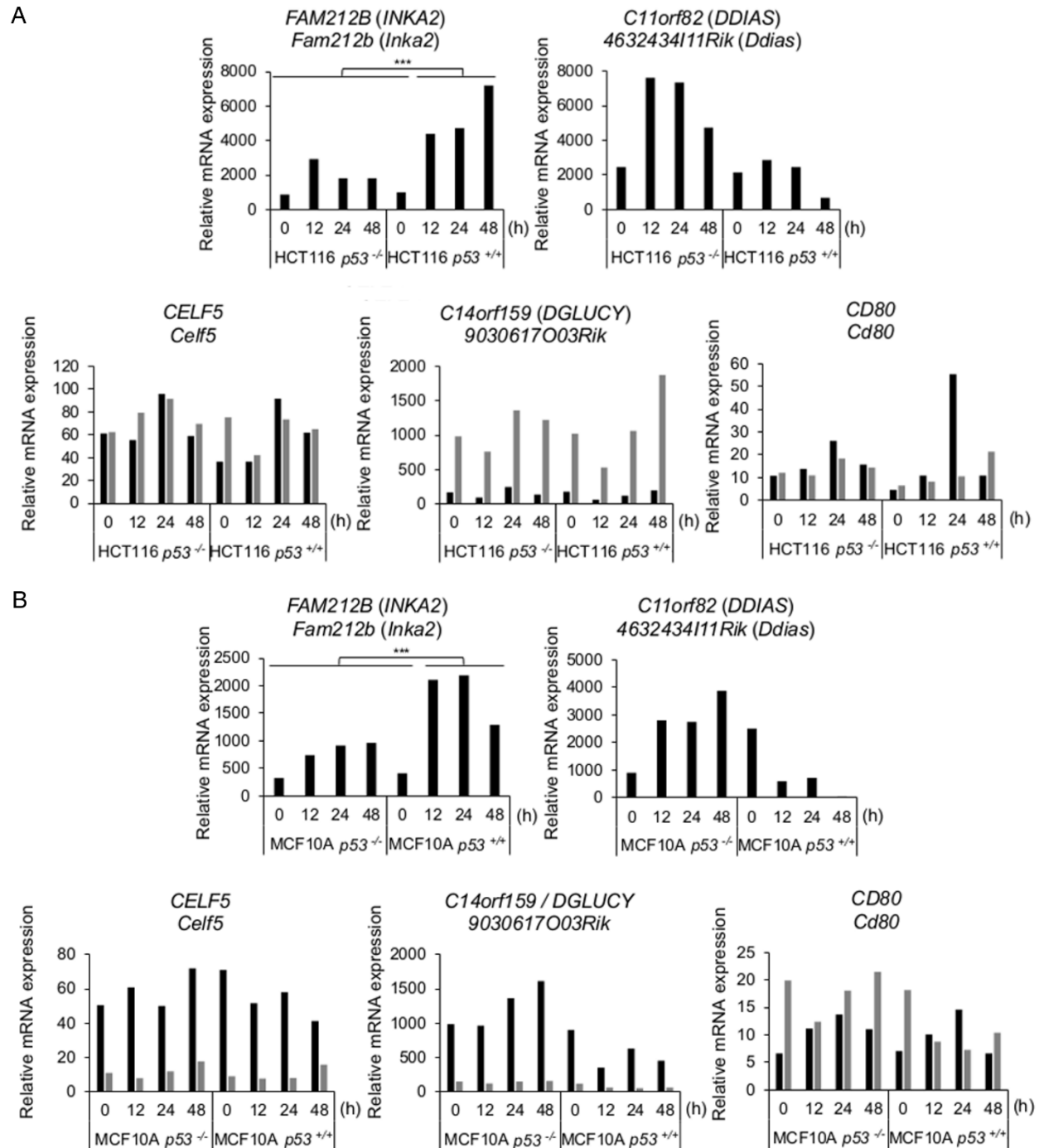


Figure 1.2 Expression level of candidate genes in human transcriptome data

Expression levels of the five candidate genes mRNAs in human wild-type $p53$ ($p53^{+/+}$) or knockout ($p53^{-/-}$) HCT116 colorectal carcinoma cells (A) and non-tumorigenic breast epithelial cells MCF10A (B) harvested at 0, 12, 24, or 48 h after the ADR treatment. Different isoforms are represented with bars of different colors.

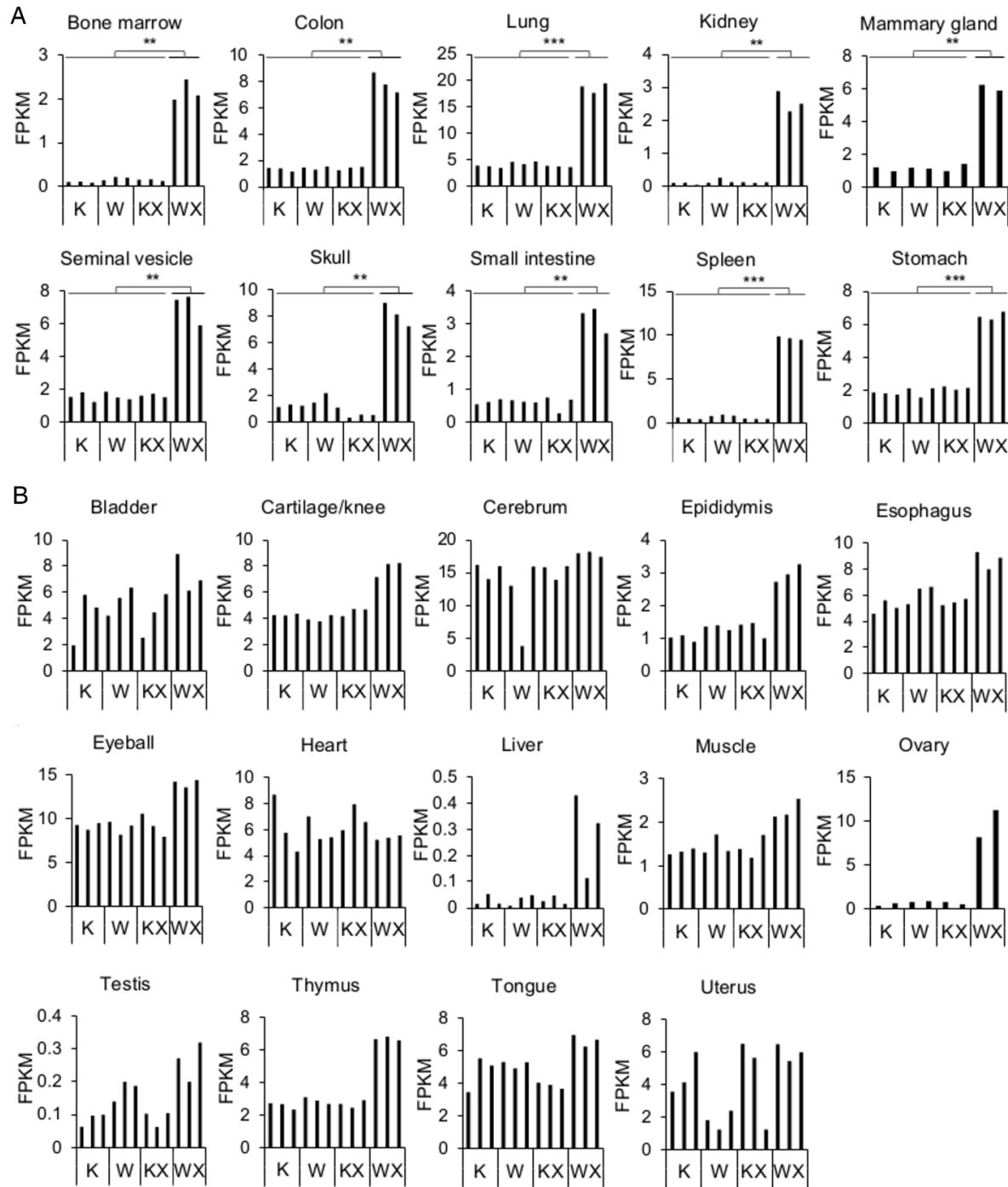


Figure 1.3 Expression level of *Inka2* in mouse transcriptome data

Levels of the *Inka2* mRNA in tissues satisfying the screening criteria (A) and in other tissues (B). RNA-seq was performed on tissues from untreated *p53* wild-type (W) or knockout (K) mice, or dissected 24 hours after irradiation with 10 Gy of X-rays (WX and KX). Two-tailed Student's *t*-test; * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

2. *INKA2* expression is induced in a *p53*-dependent manner

We further investigated the correlation between *INKA2* and *p53* expression using human cancer cell lines. Although *INKA2* has two mRNA isoforms, the second isoform showed insufficient expression levels for analysis in most cell lines (Figure 1.4). Therefore, we focused our study on isoform 1 of *INKA2*, the major transcript, and we examined its expression in HBC4 and HBL-100 human breast cancer cells, as well as U2OS osteosarcoma cells. The levels of the *INKA2* mRNA in all three cell lines were evaluated using quantitative reverse transcription real-time PCR (RT-qPCR). A significant increase in the levels of the *INKA2* mRNA was observed in cells treated with ADR. However, the expression was repressed when *p53* expression was silenced with an siRNA (Figure 1.5 A). Next, we performed a western blot analysis of U2OS cell lysates cultured under the same experimental conditions (Figure 1.5 B). The amount of the *p53* protein increased in cells treated with ADR and decreased in cells transfected with the *p53* siRNA. We used *p21* as the positive control. *INKA2* and *p21* showed similar expression patterns. Based on these results, the expression of both the *INKA2* mRNA and protein is regulated by a *p53*-dependent mechanism.

We designed two siRNAs targeting *INKA2* and confirmed the silencing of *INKA2* using RT-qPCR (Figure 1.6 A). In addition, the western blot analysis showed the loss of the band corresponding to *INKA2* in cells transfected with the siRNAs, indicating the specificity of the antibody (Figure 1.6 A). Finally, we performed immunocytochemistry to validate the induction of *INKA2* expression by *p53* and the repression of *INKA2* by the siRNAs (Figure 1.6 B).

In addition to intracellular expression, we analyzed *INKA2* mRNA expression levels in tumor tissues from the TCGA database. Among all cancer types, we observed a significantly higher *INKA2* expression in *p53* wild-type samples compared to *p53* mutant samples from patients with bladder urothelial carcinoma (BLCA), colorectal adenocarcinoma (COAD/READ), pan-kidney cohort (KIPAN), and prostate adenocarcinoma (PRAD) (Figure 1.7). Consequently, *INKA2* showed a *p53*-dependent expression pattern at both the cellular and tissue levels, indicating that the expression of this gene is regulated by *p53*.

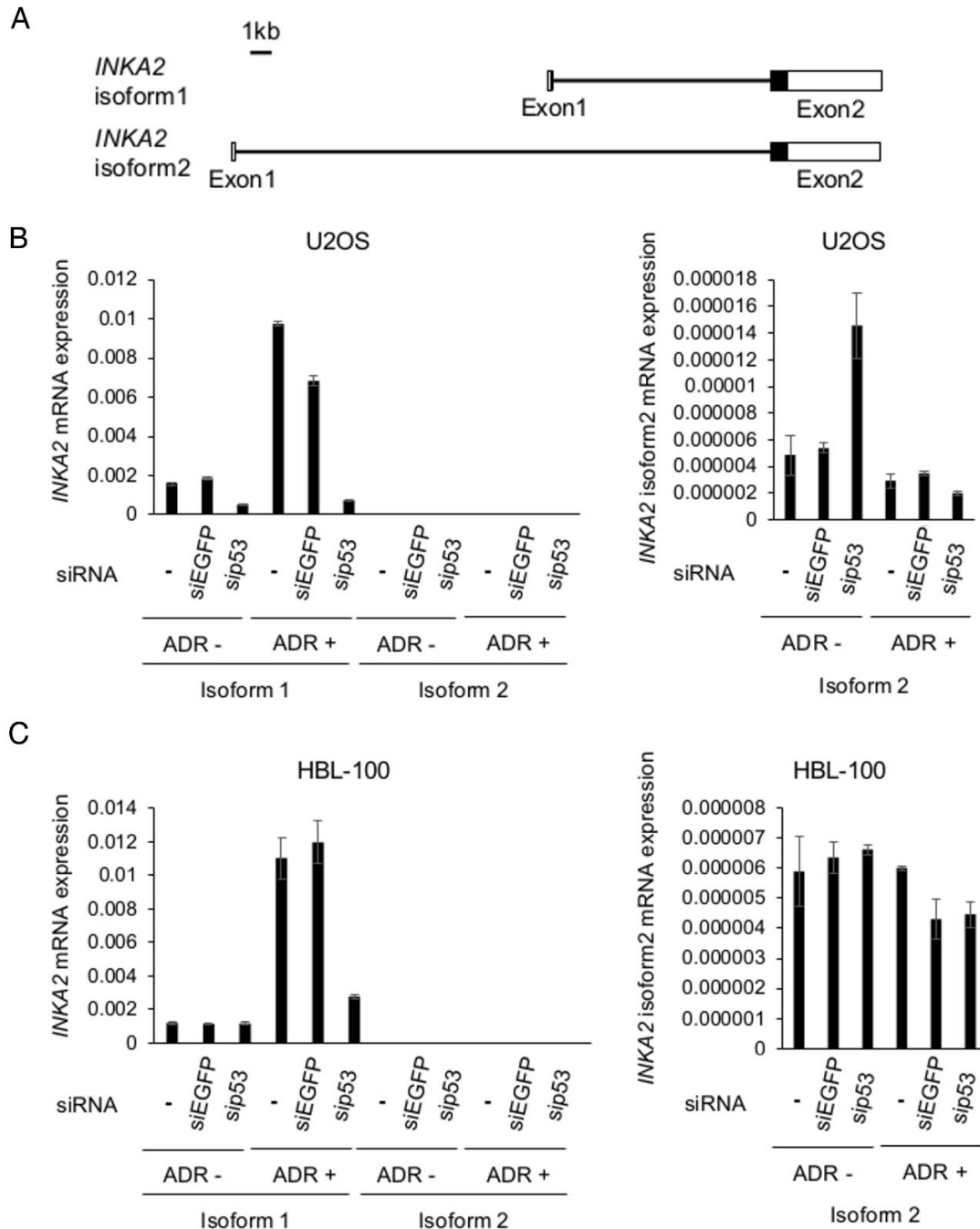


Figure 1.4 *INKA2* isoform 2 exhibits low expression

(A) Gene structure of *INKA2* isoforms.

Relative mRNA expression of both isoform 1 and isoform 2 of *INKA2* (left panel) or isoform 2 only (right panel) in U2OS cells (B) and HBL-100 cells (C) normalized to GAPDH. Cells were treated with or without an siRNA targeting either EGFP or p53 and treated with or without ADR. Error bars, S.D. n=2

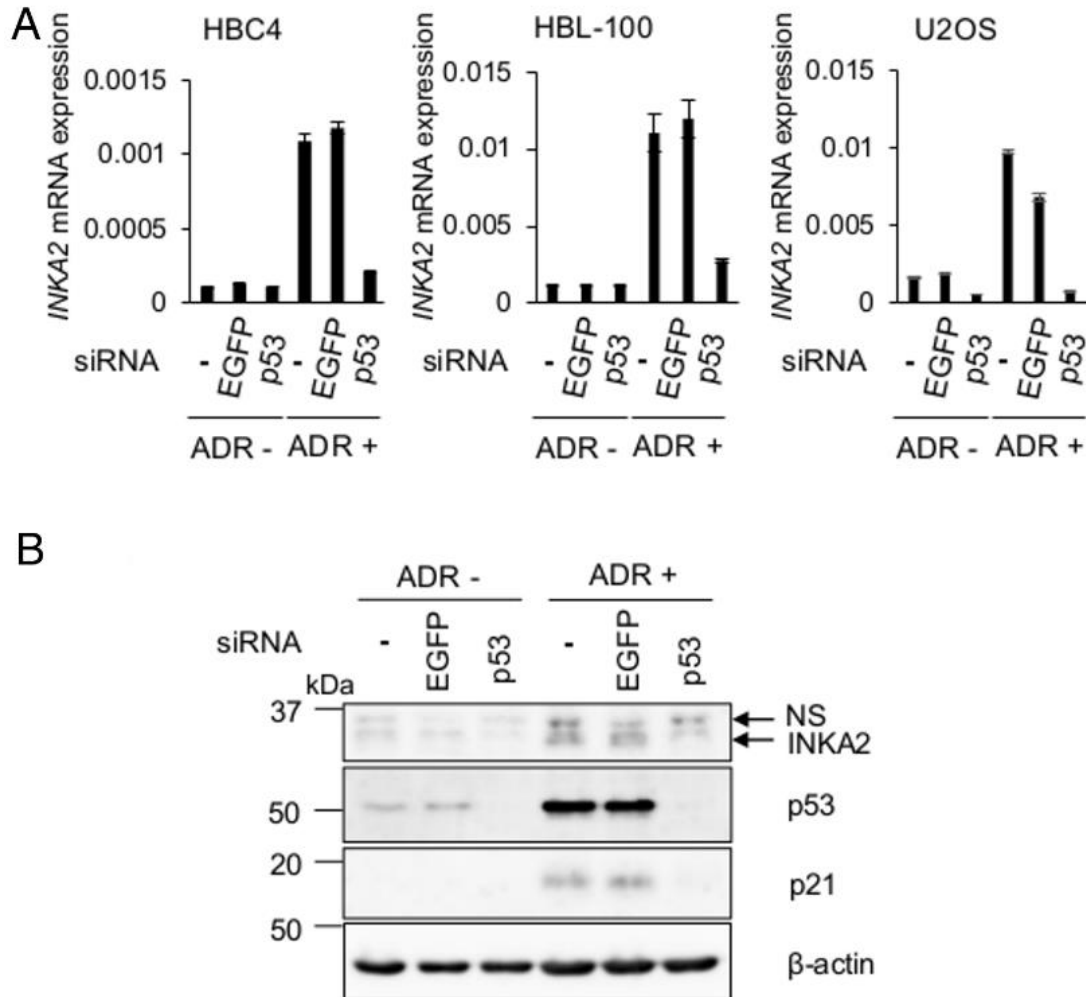


Figure 1.5 *INKA2* is a p53 downstream target

(A) Relative expression of the *INKA2* mRNA normalized to the GAPDH mRNA in HBC4, HBL-100, and U2OS cells transfected with or without an siRNA targeting either EGFP or p53 and treated with or without ADR. Error bars, S.D. n=2.

(B) Western blot analysis of levels of the *INKA2*, p53, and p21 proteins in U2OS cells transfected with or without the siRNAs and treated with or without ADR. β-Actin is shown as a loading control. NS; non-specific.

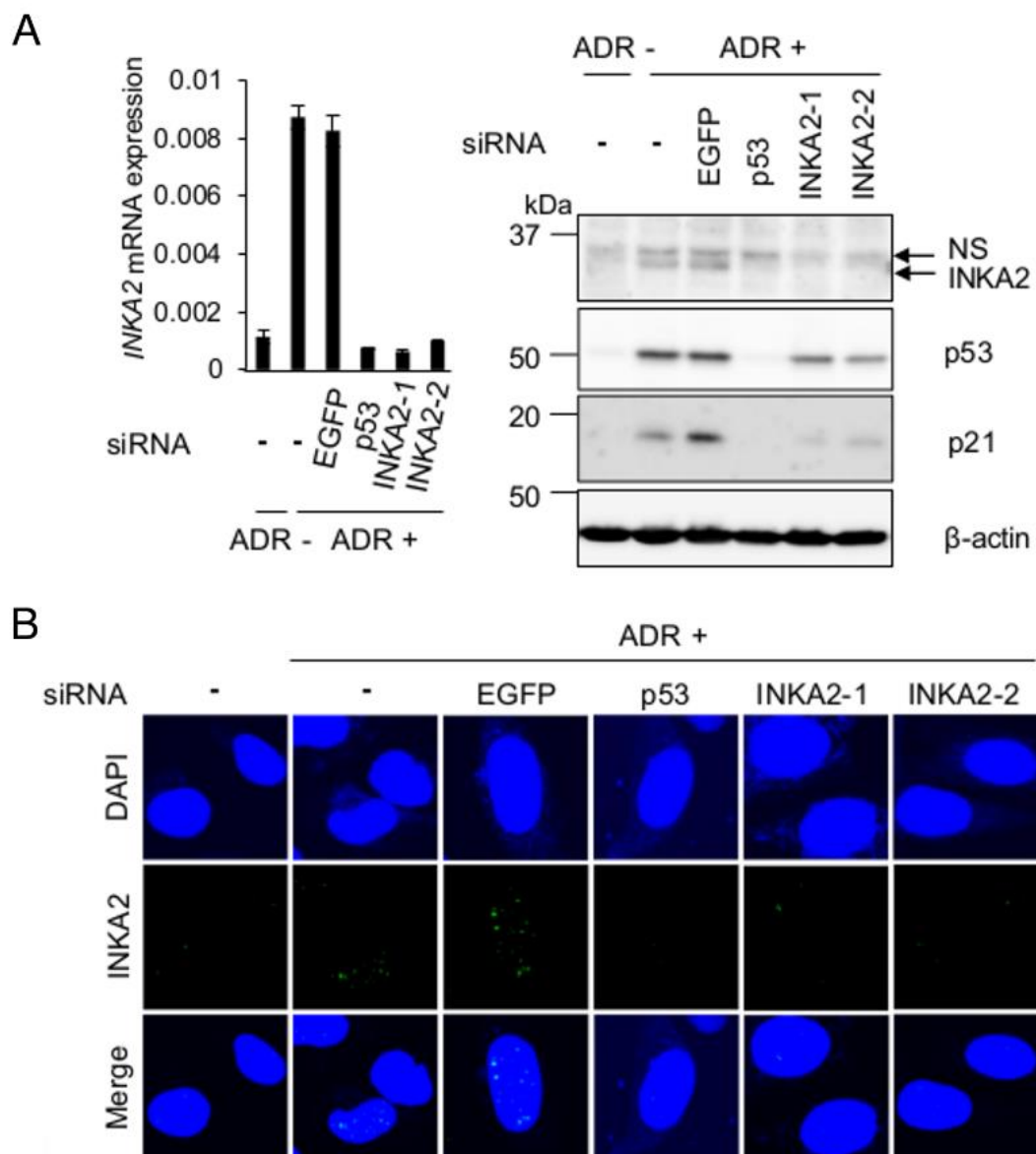


Figure 1.6 Evaluation of the specificity of the antibody against INKA2

(A) RT-qPCR and western blot analyses of INKA2 levels in U2OS cell lines transfected with two siRNAs against INKA2. Error bars, S.D. n=2.

(B) Immunocytochemical staining of U2OS cells transfected with the siRNAs and treated with ADR. Cells were stained with DAPI (blue) and an antibody against INKA2 (green).

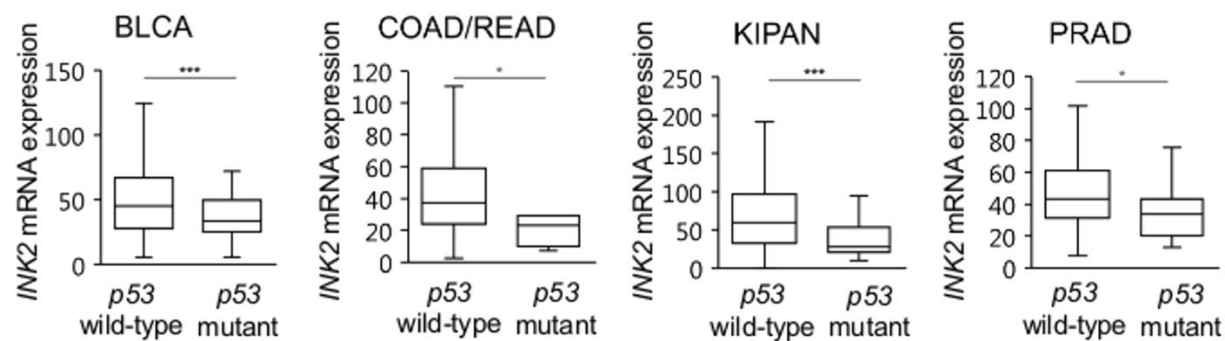


Figure 1.7 mRNA expression of *INKA2* in TCGA

Comparison of levels of the *INKA2* mRNA between TCGA *p53* wild-type and *p53* mutant for bladder urothelial carcinoma (BLCA), colorectal adenocarcinoma (COAD/READ), pan-kidney cohort (KIPAN), and prostate adenocarcinoma (PRAD) tumor tissues. Two-tailed Student's *t*-test; * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

3. *INKA2* is a direct p53 downstream target

We inspected data obtained from chromatin immunoprecipitation coupled with deep sequencing (ChIP-seq) targeting the p53 protein to examine whether p53 directly regulates *INKA2* expression. In human samples, ChIP-seq results from the ReMap 2018 database ¹⁹ revealed p53 accumulation at three regions of the *INKA2* gene: 15 kb and 7.5 kb upstream to the first exon of isoform 1, and throughout the first exon of isoform 1 (Figure 1.8). Further investigations using the JASPAR database ²¹ identified a p53 binding motif located approximately 600 bases downstream of the first exon of isoform 1 (Figure 1.9), overlapping with the third ChIP-seq peak region (Table 1). Moreover, the ENCODE histone marks data ²⁷ integrated in the UCSC Genome Browser ²⁸ showed enrichment of the H3K4me1 and H3K4me3 histone methylation and H3K27Ac histone acetylation marks in the same region, implicating chromatin remodeling and transcriptional regulation at the p53-responsive site (Figure 1.8).

Likewise, we identified a p53-responsive element in the mouse *Inka2* sequence using ChIP-seq data obtained from mouse embryonic fibroblasts (MEF) treated with ADR ²⁰ (Figure 1.10). We then employed the JASPAR database to predict p53-responsive elements in the ChIP-seq results. Similar to the human p53-responsive element, the identified p53 binding motif is located approximately 600 bases downstream of the first exon of isoform 1 (Figure 1.9 and Table 2). The p53 binding site predicted by the JASPAR database comprises 18 bases. The identified responsive elements between the two species show a similarity of 94.4% and they only differ in the 9th nucleotide, where a T in the human sequence is replaced by a C in the mouse sequence. The absolute scores calculated from the position weight matrix of the JASPAR p53 binding site model are 9.20711 and 10.8519 for human and mouse, respectively (Figure 1.9). A further inspection of the binding site sequence indicated a high level of conservation across species (Figure 1.11), revealing the importance of the sequence.

Next, we performed luciferase reporter assays in U373MG and H1299 cells expressing wild-type human p53 or the p53 R175H mutant, or wild-type mouse p53 or the p53 R172H mutant (Figure 1.12). The p53 protein interacted with the predicted binding site when both p53 and the binding site sequences were wild-type. Luciferase activity decreased when point mutations were

introduced into the responsive element sequences. Similarly, luciferase activity decreased in the presence of mutant p53. Thus, INKA2 is directly regulated by p53.

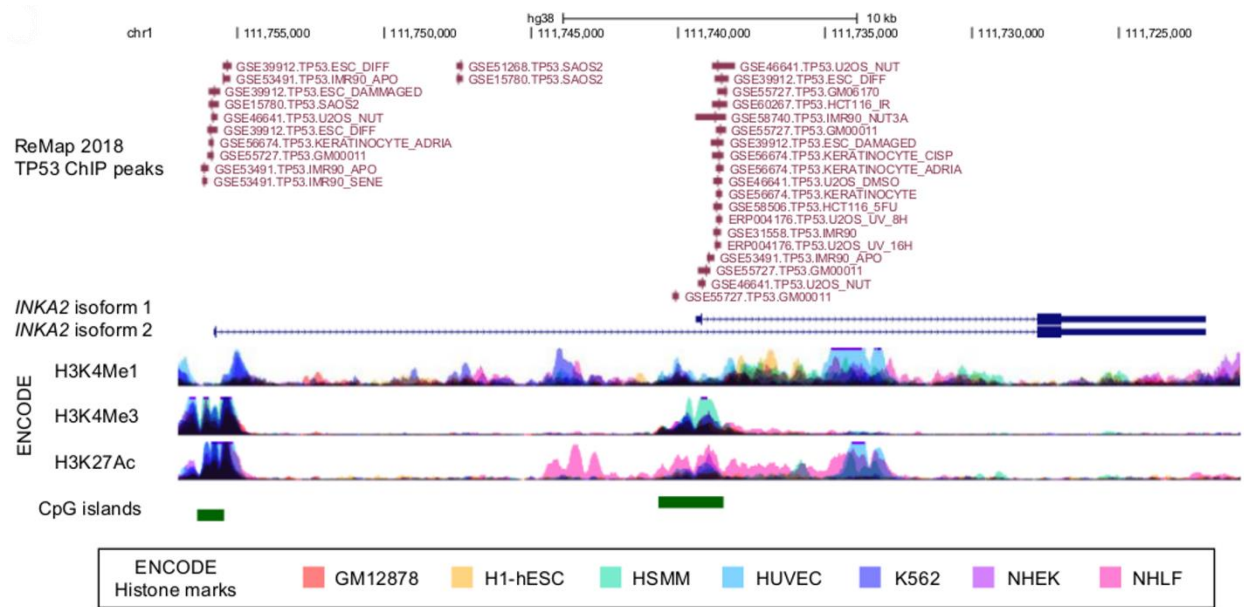


Figure 1.8 TP53 ChIP-sequencing data showing 3 possible p53 binding regions throughout *INKA2* gene

UCSC (University of California Santa Cruz, California, United States) Genome Browser view of TP53 ChIP-seq peaks obtained from the ReMap 2018 database integrated with ENCODE data showing H3K4Me1, H3K4Me3, and H3K27Ac histone modification marks from seven cell types. The *INKA2* (*FAM212B*) gene is shown beneath the peaks, followed by UCSC CpG island tracking information.

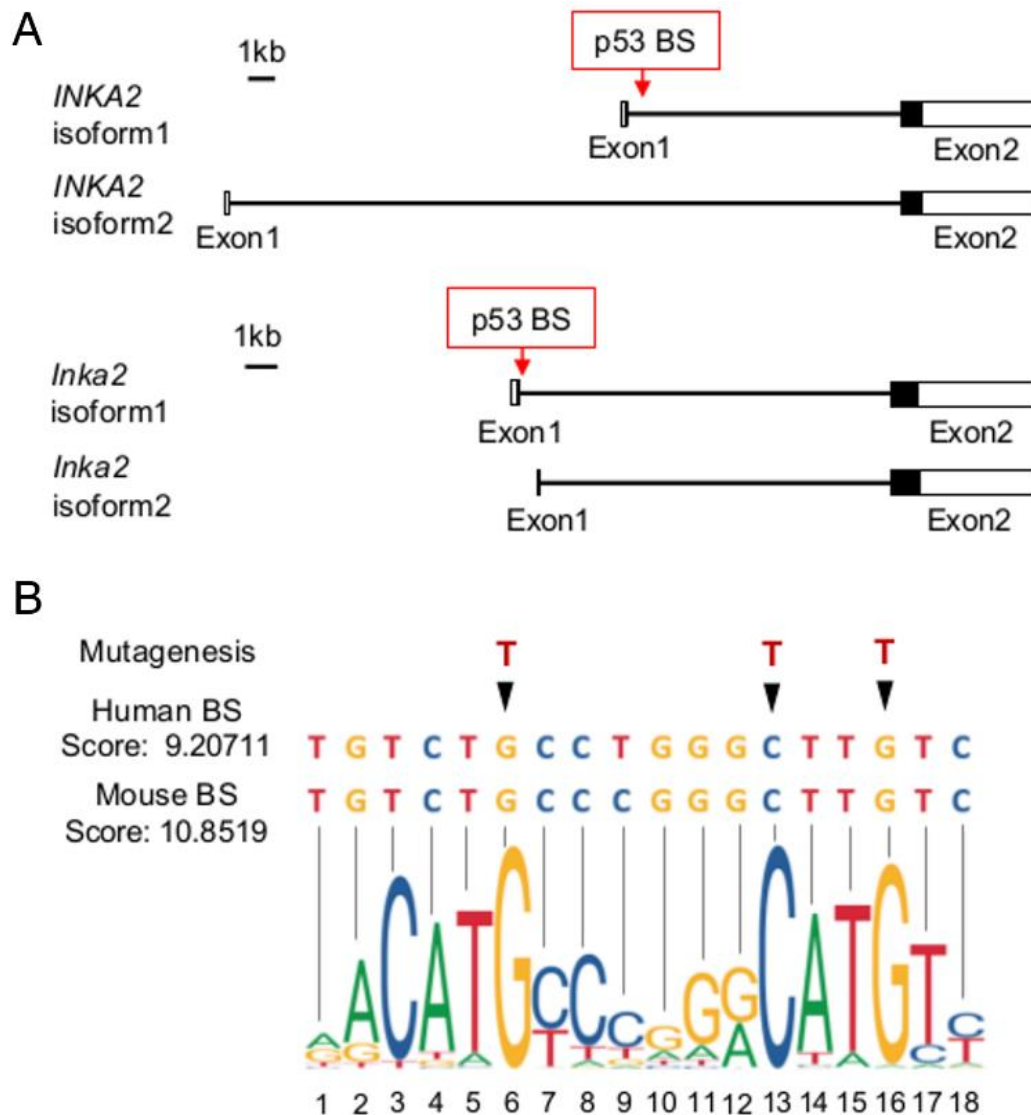


Figure 1.9 *INKA2/Inka2* gene contains an intronic p53 binding site

(A) Genomic structure of the human *INKA2* gene (top panel) and the mouse *Inka2* gene (bottom panel). Filled box, coding region; unfilled box, non-coding region; line, intron; p53 BS, p53 binding site.

(B) The p53 binding sites (BS) identified in human and mouse *INKA2/Inka2* were aligned with the JASPAR model. Nucleotides modified by mutagenesis are indicated with arrows.

Table 1. Genomic locations of p53 ChIP-seq peaks from ReMap 2018 database and *INKA2/FAM212B* binding site

Cell type	p53 status	Biological condition	Start position	End position	Peak length (bp)	PMID	GSE ID
U2OS	Wt	UV	111,738,507	111,738,718	211	24289924	ERP004176
			111,738,540	111,738,746	206		
IMR-90	Wt	5-FU	111,738,530	111,738,787	257	22127205	GSE31558
hESCs	Wt	ADR	111,738,291	111,738,744	214	24078252	GSE39912
U2OS	Wt	ADR, Nutlin-3a	111,738,077	111,738,840	763	23775793	GSE46641
			111,738,473	111,738,811	338		
GM00011, GM06170	Wt	ADR	111,738,308	111,738,668	360	25883152	GSE55727
			111,738,378	111,738,701	323		
HFK	Wt	ADR, cisplatin	111,738,438	111,738,844	406	24823795	GSE56674
			111,738,457	111,738,723	266		
			111,738,493	111,738,711	218		
HCT 116	Wt	5-FU	111,738,495	111,738,787	292	25524025	GSE58506
IMR-90	Wt	Nutlin-3a	111,738,375	111,739,417	1042	25391375	GSE58740
HCT 116	Wt	Ionizing radiations	111,738,328	111,738,862	534	25996291	GSE60267
p53 binding site identified in <i>Inka2</i>			105,705,367	105,705,385	18		

hESCs, Human embryonic stem cells; HFK, Human foreskin keratinocytes; Wt, Wild-type; 5-FU, Fluorouracil

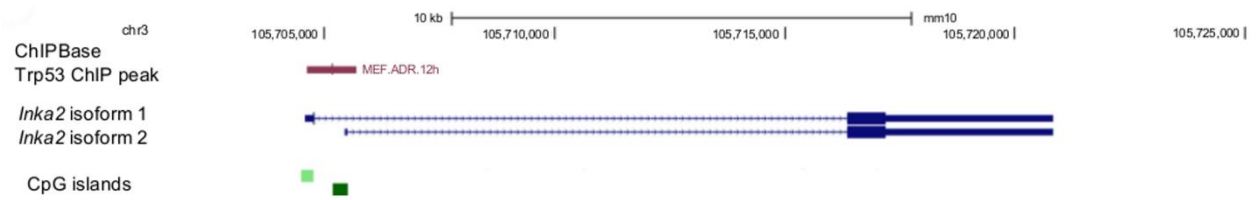


Figure 1.10 Data of ChIP-seq targeting Trp53 showing p53 binding region at the promoter region of mouse *Inka2*

UCSC Genome Browser view of Trp53 ChIP-seq peaks obtained from the ChIPBase database. The *Inka2* (*Fam212b*) gene is shown beneath the peaks, followed by UCSC CpG island tracking information.

Table 2. Genomic locations of p53 ChIP-seq peaks from ChIPBase database and *Inka2/Fam212b* binding site

Cell type	p53 status	Biological condition	Start position	End position	Peak length (bp)	PMID	GSE ID
MEF	Wild-type	ADR	105,704,618	105,705,708	1090	25883152	GSE55727
p53 binding site identified in <i>Inka2</i>			105,705,367	105,705,385	18		

MEF, Mouse embryonic fibroblasts.

Homo sapiens (Human)	T	G	T	C	T	G	C	C	T	G	G	G	C	T	T	G	T	C
Chimpanzee	T	G	T	C	T	G	C	C	T	G	G	G	C	T	T	G	T	C
Bonobo	T	G	T	C	T	G	C	C	T	G	G	G	C	T	T	G	T	C
Orangutan	T	G	T	C	T	G	C	C	T	G	G	G	C	T	T	G	T	C
Crab-eating macaque	T	G	T	C	T	G	C	C	T	G	G	G	C	T	T	G	T	C
Baboon	T	G	T	C	T	G	C	C	T	G	G	G	C	T	T	G	T	C
Green monkey	T	G	T	C	T	G	C	C	T	G	G	G	C	T	T	G	T	C
Proboscis monkey	T	G	T	C	T	G	C	C	T	G	G	G	C	T	T	G	T	C
Golden snub-nosed monkey	T	G	T	C	T	G	C	C	T	G	G	G	C	T	T	G	T	C
Rhesus	N	N	N	N	N	N	N	N	N	G	G	G	C	T	T	G	T	C
Gibbon	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
Marmoset	T	G	T	C	T	G	C	C	C	G	G	G	C	T	T	G	T	C
Squirrel monkey	T	G	T	C	T	G	C	C	C	G	G	G	C	T	T	G	T	C
Mouse lemur	T	G	T	C	T	G	C	C	C	G	G	G	C	T	T	G	T	C
Bushbaby	T	G	T	C	T	G	C	C	C	G	G	G	C	T	T	G	T	C
Tree shrew	T	G	T	C	T	G	C	C	C	G	G	G	C	T	T	G	T	C
Mouse	T	G	T	C	T	G	C	C	C	G	G	G	C	T	T	G	T	C
Rat	T	G	T	C	T	G	C	C	C	A	G	G	C	T	T	G	T	C
Opossum	T	G	T	C	T	G	C	C	C	G	G	A	C	T	T	G	T	C
Dog	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
Tarsier																		

Figure 1.11 The p53 binding site identified in *INKA2/Inka2* is conserved across species

Sequence alignment of the human p53 binding site in *INKA2* and the corresponding sequence conserved between 7 vertebrates species (human, chimpanzee, rhesus, mouse, rat, dog, and opossum) and 20 mammalian species (human, chimpanzee, bonobo, orangutan, crab-eating macaque, baboon, green monkey, proboscis monkey, golden snub-nosed monkey, rhesus, gibbon, marmoset, squirrel monkey, mouse lemur, bushbaby, tree shrew, mouse, dog, tarsier). Data were obtained from the UCSC Genome Browser. Ns and blanks represent missing nucleotides in the assemblies (see the track description from the UCSC Genome Browser).

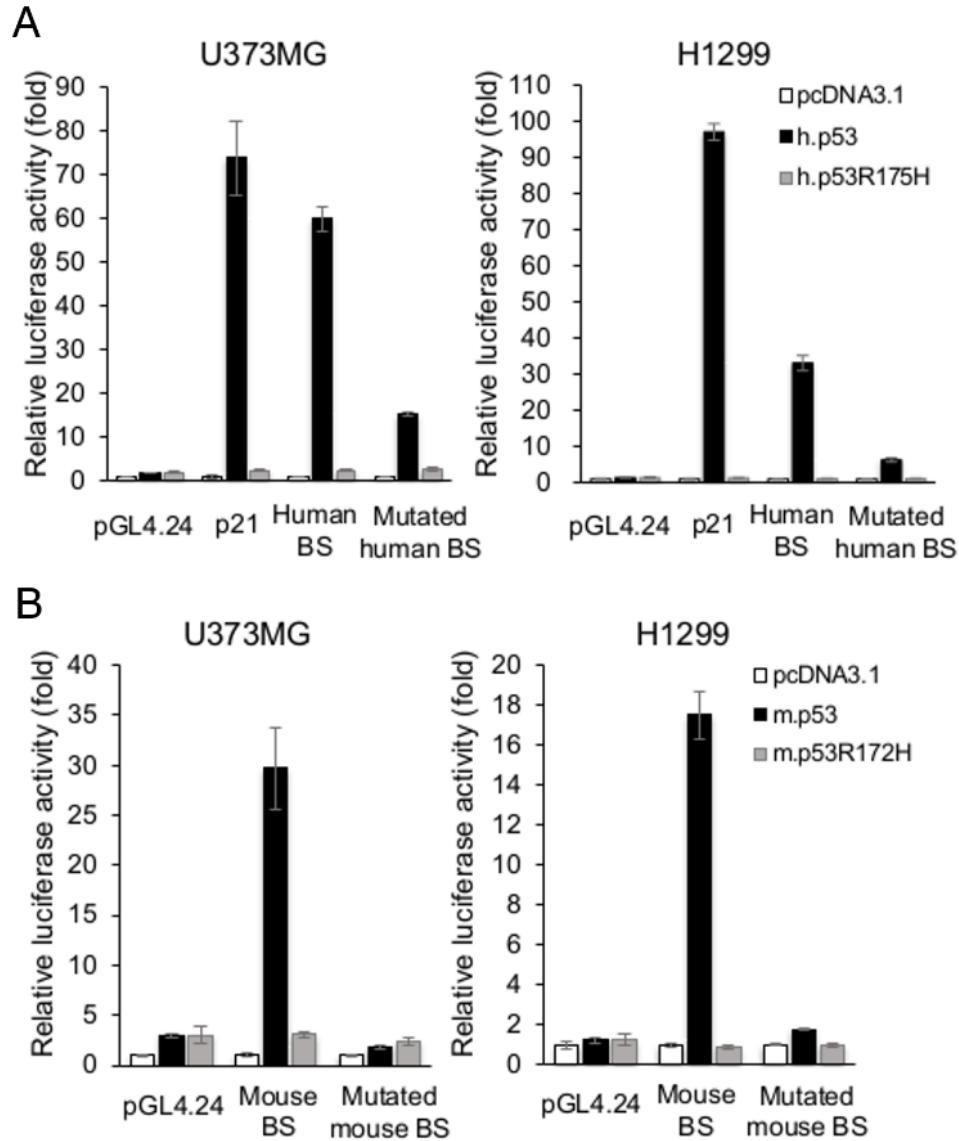


Figure 1.12 *INKA2* is a direct target of p53

Luciferase assay of U373MG and H1299 cells expressing the empty vector pGL4.24, BS construct, or mutated BS constructs of the human (A) and mouse (B) genes, performed as described in Figure 1.9 B. The control vector pcDNA3.1, human (A) or mouse (B) wild-type p53 or mutant p53 were co-transfected at the same time. Luciferase activity is indicated relative to the activity of the control vectors pcDNA3.1 and pGL4.24. h.p53, human p53; h.p53R175H, mutant human p53 R175H; m.p53, mouse p53; m.p53R172H, mutant mouse p53 R172H. Error bars, S.D. n=3.

4. *INKA2* shows tumor suppressive function

Regarding the relationship between *INKA2* and tumors, we investigated its role in cancer cell growth by performing colony formation assays using HBC4 and U2OS cells (Figure 1.13 A). Twenty to forty percent fewer colonies were formed by cells over-expressing *INKA2* compared to mock vector. Subsequently, we concluded that *INKA2* reduces cancer cell proliferation to some extent.

We further analyzed *INKA2* expression data from the TCGA database (Figure 1.13 B). *INKA2* was expressed at lower levels in tumor samples than in control samples in multiple cancer types: breast invasive carcinoma (BRCA), colon adenocarcinoma (COAD), kidney renal clear cell carcinoma grouped with kidney renal papillary cell carcinoma (KIRC+KIRP), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), prostate adenocarcinoma (PRAD), and thyroid carcinoma (THCA). Additionally, the promoters of the human and mouse *INKA2* genes contains CpG islands (Figure 1.8 and 1.10). Higher DNA methylation was observed in the human *INKA2* promoter region in tumor samples than in control samples (Figure 1.13 C), which explains the down-regulation of *INKA2* expression in tumor samples. Together with the reduction of cancer cell growth, we suggest that *INKA2* might function as a tumor-suppressor.

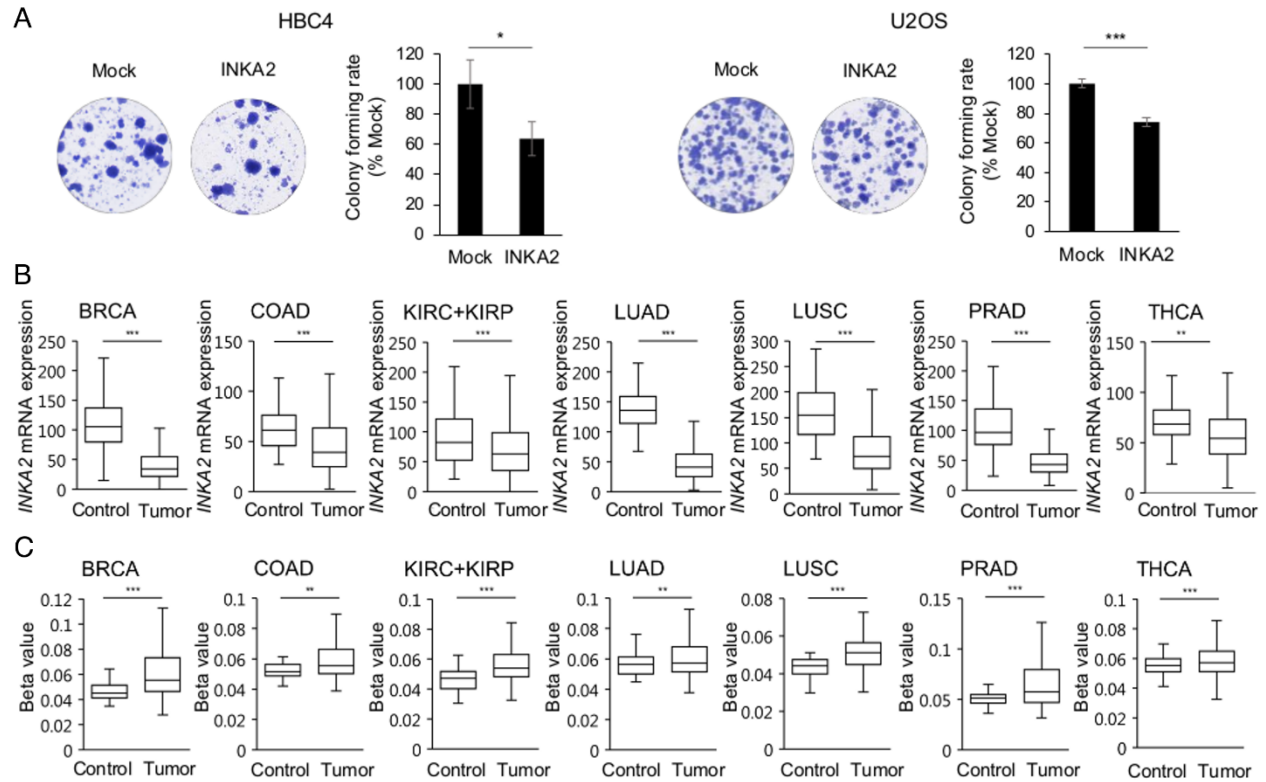


Figure 1.13 *INKA2* functions as a tumor suppressor

(A) Results of the colony formation assay performed with HBC4 (left panel) and U2OS (right panel) cells over-expressing the mock vector or *INKA2*. The number of colonies in the mock-transfected cells is considered as 100%. Error bars, S.D.; n=3.

(B-C) Differential expression of the *INKA2* mRNA (B) and DNA methylation level at the *INKA2* promoter region (C) in control and tumor samples of breast invasive carcinoma (BRCA), colon adenocarcinoma (COAD), kidney renal clear cell carcinoma and kidney renal papillary cell carcinoma (KIRC+KIRP), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), prostate adenocarcinoma (PRAD), and thyroid carcinoma (THCA) patients. Two-tailed Student's *t*-test; * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

5. *INKA2 interacts with PAK4*

We immunopurified INKA2 protein with an antibody against HA-tag in whole-cell lysates obtained from HEK293T cells over-expressing HA-tagged INKA2 to determine the function of INKA2 in human cells. Samples were then subjected to mass spectrometry analysis (Table 3). The results revealed a high binding capacity of INKA2 to p21-activated kinase 4 (PAK4), a serine/threonine-protein kinase. In fact, INKA1, a member of the same INKA protein family, is an inhibitor of PAK4 kinase¹⁰. Both INKA1 and INKA2 contain an iBox domain of 38 amino acids (Figure 1.14 A) PAK4 contains an auto-inhibitory domain (AID) that interacts with the substrate-docking site situated in the catalytic domain via the RPK peptide: R⁴⁹PKPLVDP^{29,30} (Figure 1.14 A and B). The iBox domain also includes a tripeptide PLV that is shared by the RPK sequence in PAK4-AID (Figure 1.14 A). Thus, iBox peptide is likely able to bind to the substrate-docking site of PAK4 in the same manner as the PAK4-AID. Compared with INKA1, INKA2 maintains the iBox domain (137-174) with a high degree of sequence identity, but the essential PAK4-inhibitory motif is NRQPLVLG instead of SRQPLVLG (Figure 1.14 A). Moreover, the iBox domain is highly conserved across species (Figure 1.14 C), indicating the importance of the sequence.

We immunoprecipitated INKA2 proteins ectopically expressed in HEK293T cells with an anti-HA antibody to confirm the interaction INKA2 with PAK4 (Figure 1.15 A). Subsequent western blotting for PAK4 confirmed the binding of INKA2 to PAK4. The same and reciprocal result was obtained from HEK293T cells over-expressing both HA-tagged INKA2 and FLAG-tagged PAK4 proteins (Figure 1.15 B). Next, we constructed two plasmids over-expressing partially deleted INKA2 and PAK4 proteins (Figure 1.14 B). The PAK4-inhibitory motif was deleted in INKA2 Δ 150-160 and part of the C-terminal catalytic domain (326-591) was deleted in PAK4 1-325. Both deletions disrupted the interaction between INKA2 and PAK4, suggesting that the targeted binding motif of iBox is located in the kinase domain of PAK4 (Figure 1.15 C). Moreover, immunocytochemistry with HEK293T cells over-expressing HA-tagged INKA2 showed that INKA2 co-localizes with PAK4 in the cell nucleus (Figure 1.16). Consequently, we postulate that INKA2 inhibits the kinase activity of PAK4 in the same way as INKA1.

Table 3. Mass spectrometry results for cell lysates of HEK293T cells over-expressing mock or INKA2-HA plasmid after immunoprecipitation with antibody against HA-tag

Gene Symbol	Accession	Description	Number of unique peptides	Mascot score	Sequest HT score	INKA2	INKA2	MOCK	MOCK	Ratio
						rep1	rep2	rep1	rep2	
FAM212B	Q9NTI7	PAK4-inhibitor INKA2 OS=Homo sapiens GN=FAM212B PE=1 SV=1	30	6170	812.67	7076418491	7096320012	683882	1413551	6757.2
PAK4	O96013	Serine/threonine- protein kinase PAK 4 OS=Homo sapiens GN=PAK4 PE=1 SV=1	3	108	6.12	5925410	4575458	300000	390882	15.2
HSPA4	P34932	Heat shock 70 kDa protein 4 OS=Homo sapiens GN=HSPA4 PE=1 SV=4	7	299	18.27	13257945	16261609	2184525	300000	11.88
YTHDF2	Q9Y5A9	YTH domain-containing family protein 2 OS=Homo sapiens GN=YTHDF2 PE=1 SV=2	2	140	1.77	3488741	2239863	300000	300000	9.55
EMD	P50402	Emerin OS=Homo sapiens GN=EMD PE=1 SV=1	4	139	9.45	8584906	5453183	1179338	419200	8.78
HSPA1B HSPA1A	P0DMV9	Heat shock 70 kDa protein 1B OS=Homo sapiens GN=HSPA1B PE=1 SV=1	32	4159	394.08	979983140	932523596	97516760	140257340	8.04
HSPA8	P11142	Heat shock cognate 71 kDa protein OS=Homo sapiens GN=HSPA8 PE=1 SV=1	28	5752	513.32	866156087	770222767	103156895	135779162	6.85
HSPH1	Q92598	Heat shock protein 105 kDa OS=Homo sapiens GN=HSPH1 PE=1 SV=1	4	160	11.23	760477	2339129	184344	300000	6.4
MYO5A	Q9Y4I1	Unconventional myosin-Va OS=Homo sapiens GN=MYO5A PE=1 SV=2	9	266	18.12	1062376	2710392	300000	300000	6.29
CUL3	Q13618	Cullin-3 OS=Homo sapiens GN=CUL3 PE=1 SV=2	3	40	0	2850262	896603	300000	300000	6.24
ACADM	P11310	Medium-chain specific acyl-CoA dehydrogenase, mitochondrial OS=Homo sapiens GN=ACADM PE=1 SV=1	1	59	2.32	2910795	300000	300000	300000	5.35

rep, replicate.

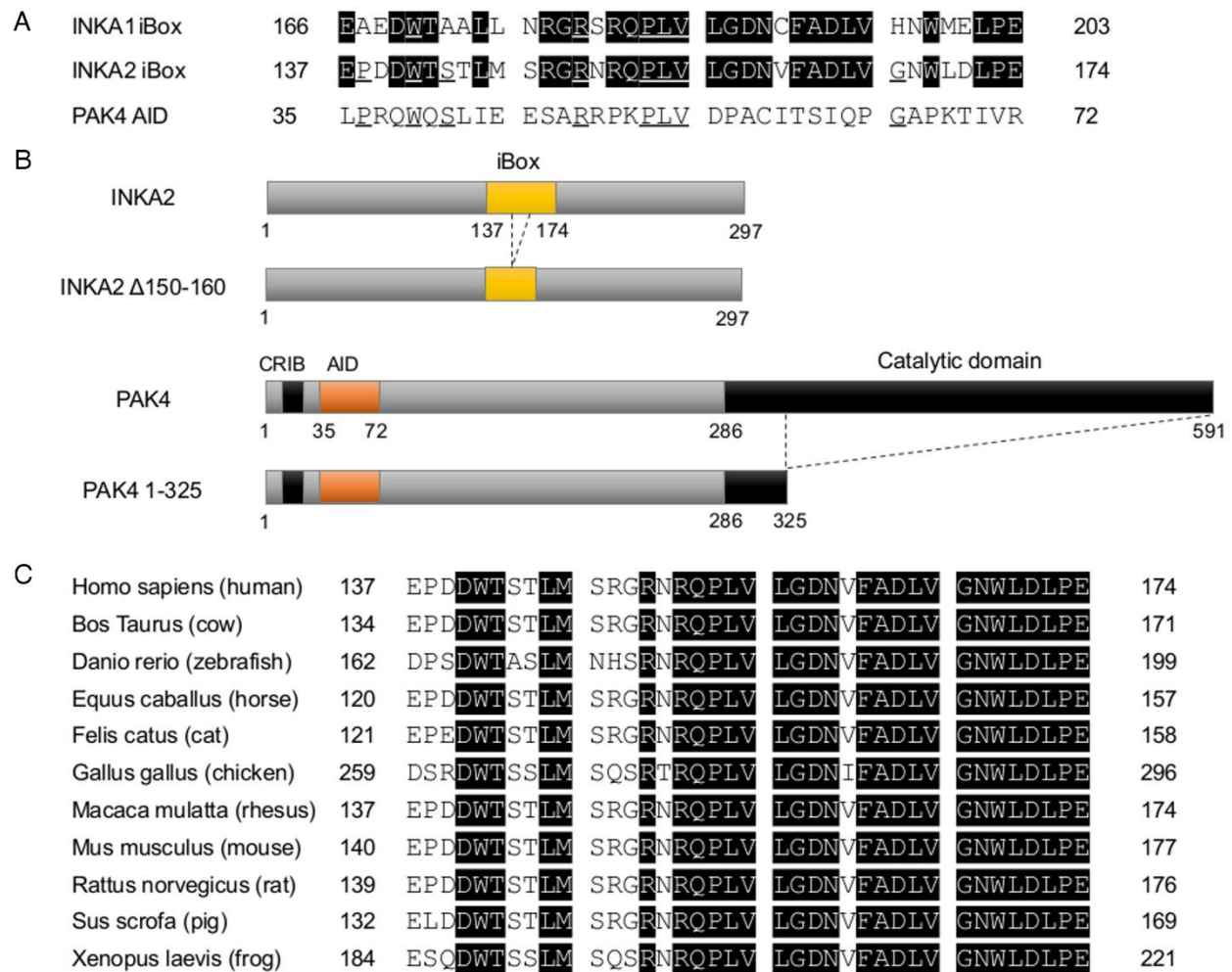


Figure 1.14 INKA2 conserves the iBox domain containing PLV tripeptide

(A) Sequence alignments of human INKA1 iBox, INKA2 iBox, and PAK4-AID. Residues shaded in black in INKA2 iBox show the identity with INKA1 iBox. Underlined residues show the identity with PAK4-AID.

(B) Structure of the INKA2 (top panel), PAK4 (bottom panel), and deletion mutant proteins used for immunoprecipitation. CRIB, Cdc42- and Rac-interactive binding domain; AID, auto-inhibitory domain.

(C) Sequence alignments of INKA2 iBox between different species. Identical residues are shaded in black.

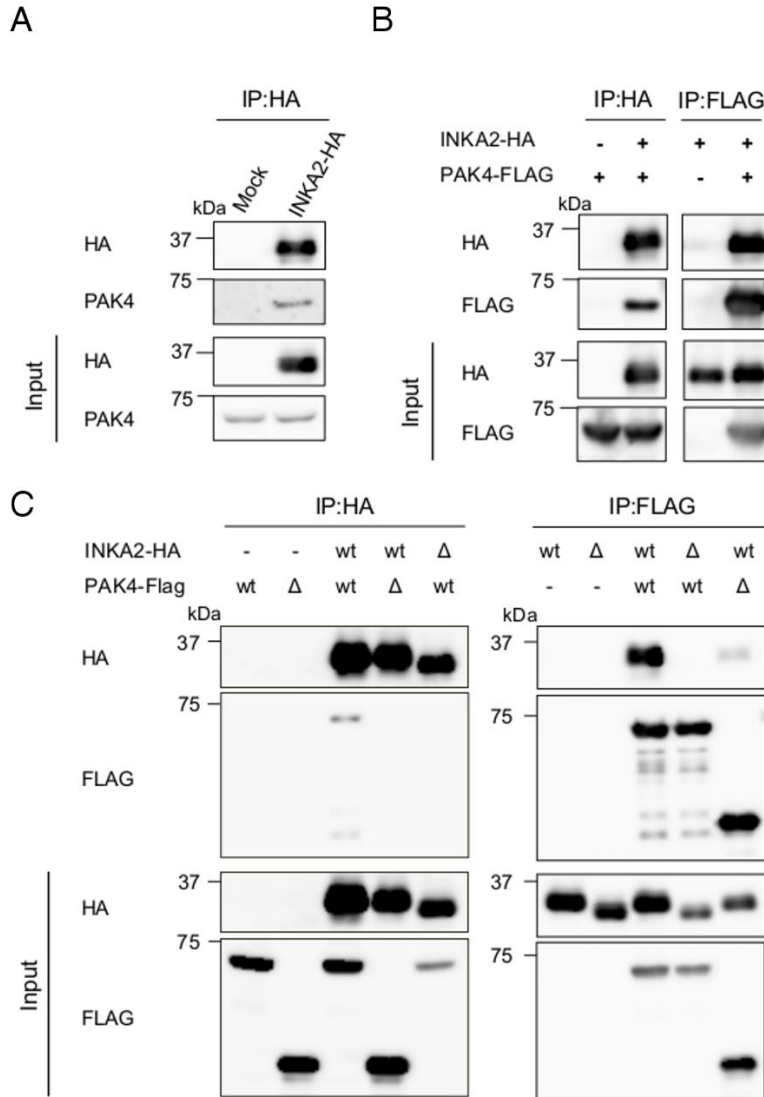


Figure 1.15 INKA2 binds to the catalytic domain of PAK4 via iBox domain

(A) The pCAGGSnHC empty vector (Mock) or pCAGGSnHC/INKA2 (INKA2-HA) plasmid were transfected into HEK293T cells. Whole-cell lysates were immunoprecipitated with an anti-HA antibody and blotted with an anti-PAK4 antibody to identify the interaction.

(B) The pCAGGSnHC/INKA2 (INKA2-HA) and pCAGGSn3Fc/PAK4 (PAK4-FLAG) plasmids were co-transfected into HEK293T cells. Whole-cell lysates were immunoprecipitated with an anti-HA or anti-FLAG antibody and blotted with anti-HA and anti-FLAG antibodies.

(C) Deletion constructs shown in Figure 1.14 B were added during the co-transfection depicted in (B). Whole-cell lysates were immunoprecipitated with an anti-HA or anti-FLAG antibody. Wt; wild-type. Δ; deleted versions: INKA2 Δ150-160 for INKA2, and PAK4 1-325 for PAK4.

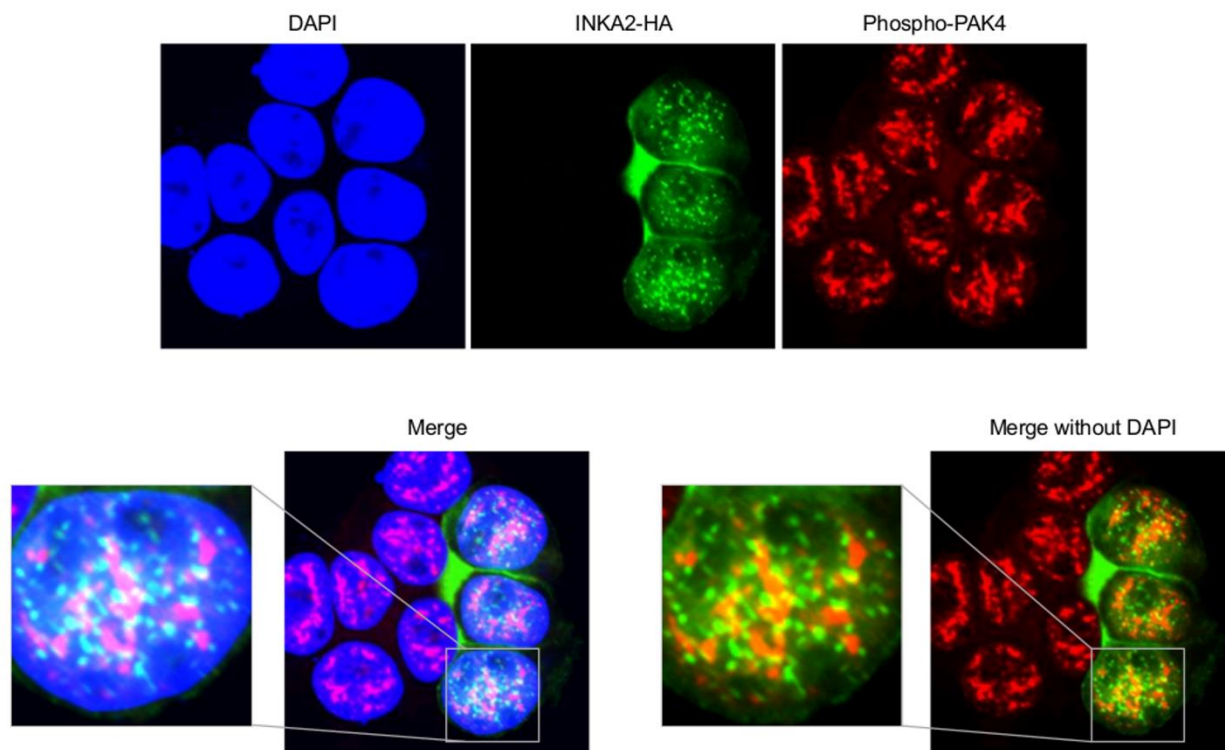


Figure 1.16 INKA2 co-localizes with PAK4 in the nucleus

Immunocytochemistry of the HEK293T cell over-expressing INKA2-HA. Cells were stained with DAPI (blue) and antibodies against HA-tag (green) and phospho-PAK4 (red).

6. *INKA2 down-modulates β -catenin at protein level*

Among a large number of substrates, PAK4 interacts with β -catenin and phosphorylates at Ser675 to regulate its nuclear and cytoplasmic transportation³¹. Ser675 phosphorylation protects β -catenin from the destruction complex, and thus stabilizes the protein in the cytoplasm³¹. Taken together, we hypothesize that INKA2 binds to the substrate-docking site within the catalytic domain of PAK4 and inhibits the phosphorylation of β -catenin Ser675, resulting the instability and the degradation of the protein. We generated a HCT116 cell line with a tetracycline-regulated lentiviral vector system stably expressing INKA2 in the presence of doxycycline (Dox) to confirm our hypothesis. Western blot analyses indicated decreased intracellular β -catenin levels in cells over-expressing INKA2 (Figure 1.17 A). Immunocytochemistry produced identical results (Figure 1.17 B), suggesting that INKA2 potentially inhibits the kinase function of PAK4 in the same manner as INKA1. As conclusion, our results support that p53-activated INKA2 functions as a tumor suppressor limiting cancer cell growth by inhibiting the phosphorylation of β -catenin by PAK4 (Figure 1.18).

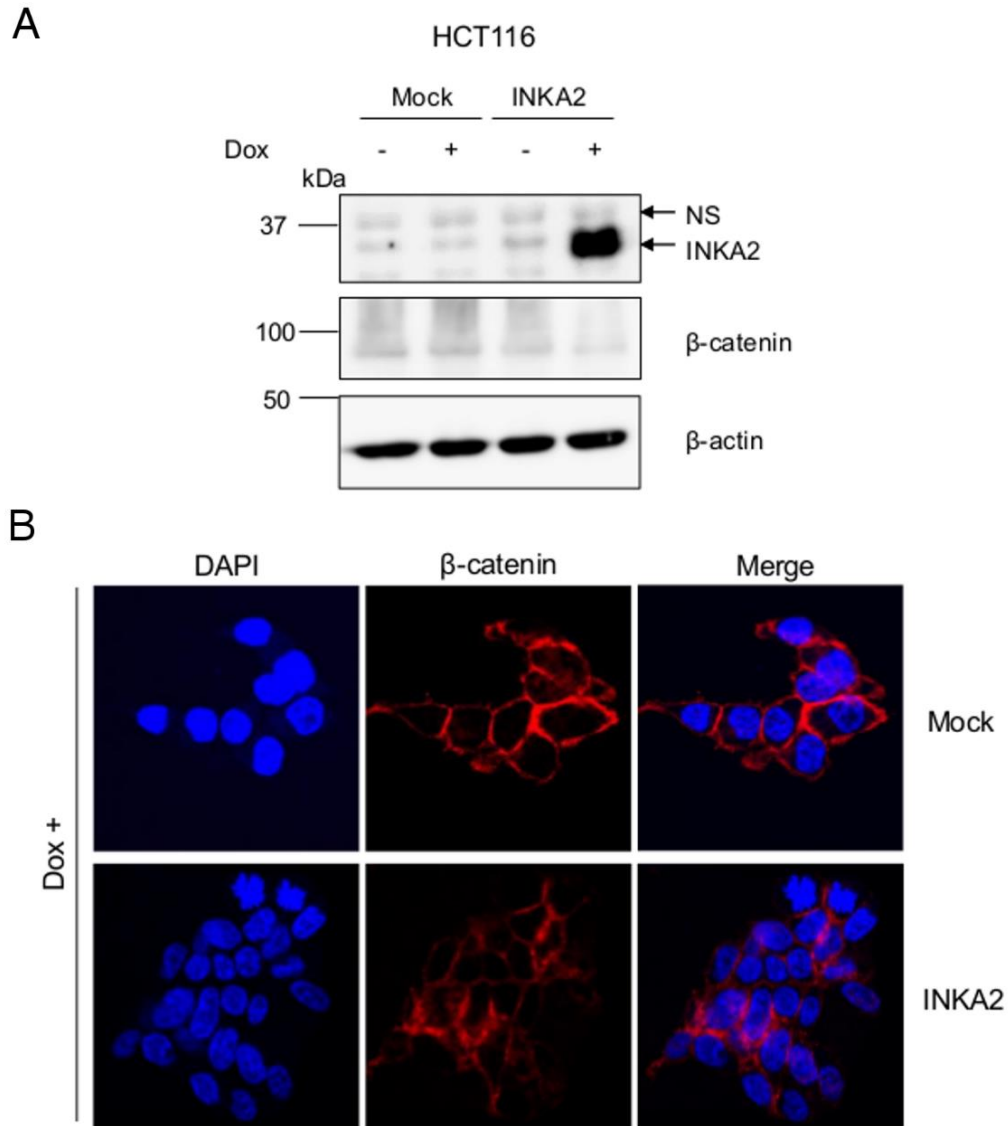


Figure 1.17 INKA2 down-modulates intracellular β-catenin

(A) Whole-cell lysates of the HCT116 cell line stably expressing INKA2 cultured in the presence of Dox were blotted with anti-INKA2 and anti-β-catenin antibodies. β-Actin is shown as a loading control.

(B) Immunocytochemistry of the stable HCT116 cell line cultured in the presence of Dox. Cells were stained with DAPI (blue) and an antibody against β-catenin (red).

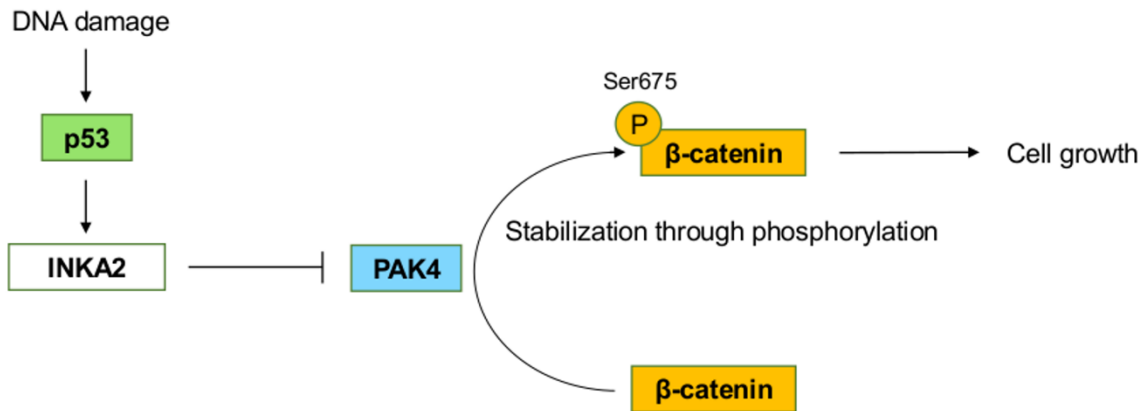


Figure 1.18 The regulation of cancer cell growth by the p53-INKA2-PAK4 pathway

Graphical summary of the regulation of cancer cell growth by the p53-INKA2-PAK4 pathway. DNA damage induce the up-regulation of p53 and the downstream targets, which include INKA2. The INKA2 iBox domain mimics the function of PAK4-AID and competes with PAK4 substrates to interact with the catalytic domain of PAK4. This inhibits the phosphorylation of β-catenin Ser675 by PAK4, thus leads to the instability of the protein and a limited cell growth.

DISCUSSION

We developed a screening method identifying novel p53 downstream target across human and mouse. The screening results in mouse transcriptome data presented eighteen p53 target gene candidates. Further analysis in human transcriptome data identified *INKA2* as the only candidate satisfying all selection criteria. The expression of *INKA2* was induced by DNA damage in multiple human cells. We also demonstrated that the induction of p53 expression elevated the expression of *INKA2* at both mRNA and protein levels in human cancer cells. Analysis of ChIP-seq data and reporter assay results revealed that p53 regulates *INKA2* by direct interaction with a strongly conservative p53 binding motif located in the intronic region of *INKA2* gene. Identical result was found in mouse *Inka2* with an almost identical binding site sequence.

Furthermore, we showed that *INKA2* has lower expression in tumor samples from several cancers. The colony formation assays showed association between *INKA2* and cancer cell growth. Taken altogether, our study identified the tumor suppressor function of *INKA2*. Although the role of *INKA2* in cancer remains uncertain, we identified PAK4 as one of the proteins interacting with *INKA2*. In our study, *INKA2* bound to the kinase domain of PAK4 and inhibited its kinase activity via iBox, a highly conserved protein domain. Our results provided evidence that *INKA2* represses the expression of β -catenin, a substrate of PAK4, at the protein level. In fact, the activation of β -catenin and the Wnt/ β -catenin signaling pathway have been implicated in the carcinogenesis of multiple cancers³²⁻³⁴. Similarly, PAK4 plays a major role in the oncogenic processes of various cancers. Over-expression of PAK4 promotes cell proliferation in HCT116 colon cancer cell³⁵ and MDA-MB-231 breast cancer cells³⁶, as well as migration and invasion in MDA-MB-231, lung cancer cells³⁷, and ovarian cancer cells³⁸. In our study, the up-regulation of *INKA2* by the p53 tumor suppressor might repress the oncogenic functions of PAK4 by inhibiting the phosphorylation of its substrates. In particular, *INKA2* might inhibit the activation of the Wnt/ β -catenin signaling pathway by binding to the substrate-docking site of PAK4, thus blocking the phosphorylation of β -catenin.

In fact, PAK4 is a targeted of anticancer drugs. Specific PAK4 inhibitors, such as PF-3758309, have been developed as anti-tumor agent. The application of the inhibitor blocks cancer cell growth in vitro and in human xenograft tumor models using nude mice³⁹. However, the

inhibitor failed in clinical trials due to its undesirable pharmacokinetic characteristics (<https://clinicaltrials.gov/ct2/show/NCT00932126>). Our finding of the interaction between the INKA2 iBox and PAK4 catalytic domain might then contribute to the design of new PAK4 inhibitors targeting the substrate-docking site of the protein.

The Inka protein family has been studied in developmental biology and has been reported to play an essential role in modulating actin reorganization, cell adhesion, and cell migration ^{7,9}. These data support the hypothesis that *INKA2* is a tumor suppressor. We postulate that *INKA2* inhibits cancer cell migration and restrains the cell junctions or cell polarity that are modified in tumors. Based on our findings regarding the interaction between INKA2 and PAK4, reasonable hypothesis is that INKA2 modifies the stability or the levels of other PAK4 substrates. We then anticipate that *INKA2* has the potential to diminish cancer cell migration and invasion by inhibiting the kinase function of PAK4. Moreover, PAK4 phosphorylates p53 and functions as a regulator of p53 itself ^{39,40}, creating a regulatory feedback loop between p53, INKA2, and PAK4. In conclusion, our study identified *INKA2* as a novel downstream target of p53 with tumor suppressor functions that inhibits the PAK4 oncogene. These results provide new insights into the role of *INKA2* in cancer and the regulation of PAK4 and its substrates by p53 and INKA2.

CHAPTER II

CHARACTERIZATION OF GENOME-WIDE CANCER ASSOCIATED SINGLE NUCLEOTIDE VARIANT IN P53 BINDING SITE

SUMMARY

p53 regulates its target genes through the p53 binding site, which is essential for the transcription factor protein to interact with target genes. We studied p53 binding motifs that contain SNPs associated with risk of various cancers at a genome-wide scale. We integrated three genome-wide datasets: 1) ChIP-sequencing data targeting p53 protein, 2) in silico p53 responsive element prediction models, and 3) post imputation GWAS datasets to identify such functional p53 binding sites in the genome. Data of ChIP-seq targeting p53 protein ChIP-sequencing data were downloaded from ReMap database. p53 binding motifs were predicted from JASPAR database, and we employed as well our in-house p53 responsive element prediction algorithm. Finally, SNPs data were obtained from post imputation GWAS dataset of fourteen cancer types.

We designed two strategies identifying candidate variants altering the binding affinity of p53 to the binding motif. The first strategy identified two SNPs: SNP 1 and SNP 2 significantly associated with lung cancer and prostate cancer, respectively. The second strategy discerned two more variants: SNP 2, overlapping with the result of the first strategy, and SNP 3 significantly associated with stomach cancer.

Next, we performed reporter assay to evaluated the interaction between p53 protein and the predicted responsive elements containing cancer associated SNPs. Then, we employed RT-qPCR to identify the target gene regulated through these binding sites. Consequently, we elucidated that the risk allele of SNP 2 is associated with a higher binding affinity of p53. This leads to the up-regulation of three oncogenes promoting prostate cancer: genes X, Y, and Z.

MATERIALS AND METHODS

*ReMap 2018 database*¹⁹

The database gathers DNA binding ChIP-seq data for various transcription factors. Among them, we used the dataset targeting p53 protein comprising twenty-two studies performing ChIP-seq in different human cell lines and diverse treatments. The majority of the studies induced DNA damage using drugs such as ADR, nutlin-3a, or cisplatin. Overlapping peaks from different studies were merged into one single peak and there were 135,467 peaks for ChIP-seq targeting p53 protein. All data were unified to the human genome GRCh38/hg38.

*JASPAR database*²¹

The database provides binding site models for various transcription factors. When a sequence is given as input, the database returns the predicted binding motif within the sequence if any was identified with an absolute score higher than the chosen threshold. In our analysis, we used as input the DNA sequence from -19 bases to + 19 bases from the candidate variant. The absolute score (Abs) is calculated by adding the position weight matrix (PWM) score of the corresponding nucleotide at each position of the binding site model. The relative score is calculated with the following formula: $(\text{Abs} - \text{minimum score}) / (\text{maximum score} - \text{minimum score})$, with the smallest or biggest value at each position of the PWM as the minimum or maximum score, respectively. There were three binding site models for p53: MA0106.1 (20 bases), MA0106.2 (15 bases), and MA0106.3 (18 bases). Spacer was not tolerated in these models.

p53 responsive element search algorithm^{41,42}

A canonical consensus p53 binding motif was employed for searching p53 responsive elements in the genome. The model is constituted of two decamers of RRRCWWGYYY, where R represents a purine, W represents an A or a T, and Y represents a pyrimidine. We also allowed a spacer from 0 to 13 nucleotides between the two decamers. The algorithm was applied on the human genome GRCh38/hg38 and 639,752 p53 responsive elements were obtained. Among them, 32,388 peaks were overlapping with ChIP-seq data targeting p53 protein.

Post imputation GWAS datasets

The number of cases and controls of each cancer type are described in Appendix Table A5. All samples were obtained from the BioBank Japan^{43,44} with a written informed consent of every participant. The project was approved by the ethical committees of every institute involved. All samples from cases and controls were genotyped with an Illumina HumanOmniExpressExome BeadChip or a combination of the Illumina HumanOmniExpress and HumanExome BeadChips. We excluded 1) samples with a call rate < 0.98 , 2) samples from closely related individuals identified by identity-by-descent analysis, 3) sex-mismatched samples or samples with incomplete information, and 4) samples from non-East Asian outliers identified by principal component analysis of the studied samples and the three major reference populations (Africans, Europeans, and East Asians) in the International HapMap Project⁴⁵. We then applied standard quality-control criteria for variants, excluding those with 1) SNP call rate < 0.99 , 2) minor allele frequency $< 1\%$, and 3) Hardy–Weinberg equilibrium p -value $< 1.0 \times 10^{-6}$. We prephased the genotypes with MACH⁴⁶ and imputed dosages with minimac⁴⁷ and the 1000 Genomes Project Phase 1 (version 3) East Asian reference haplotypes⁴⁸.

Reporter assay

DNA fragments carrying p53-responsive elements, as predicted by the JASPAR database, were amplified by PCR and were cloned into the pGL4.24 [*luc2P*/minP] plasmid (Promega, Madison, WI, USA) in both sense and antisense (see Appendix Table A6 for primers). H1299 cells were co-transfected with pGL4.24 control vector or pGL4.24 containing the target p53 binding site, accompanied by the empty vector pcDNA3.1, or wild-type human p53 expression vector. All samples were co-transfected with pGL4.74 to normalize the transfection efficiency. Luciferase activities were detected 48 hours after transfection using the Dual-Glo Luciferase Assay System (Promega, Madison, WI, USA) according to the manufacturer's protocol.

eQTL data

eQTL data were obtained from The Genotype-Tissue Expression (GTEx) database. The GTEx Project was supported by the Common Fund of the Office of the Director of the National

Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from the GTEx Portal.

Cell culture and treatment

U2OS osteosarcoma and MKN45 gastric adenocarcinoma cells were purchased from American Type Cell Collection (Rockville, MD, USA). All cells were cultured in 37°C incubator with 5% CO₂ atmosphere. The siRNA oligonucleotides were ordered from Sigma-Aldrich (St. Louis, MO, USA) and were transfected into the cells with Lipofectamine RNAiMAX (Thermo Fisher Scientific, Waltham, MA, USA) at the seeding step. The sequences are listed in Appendix Table A6. Cells were treated with 2 µg/mL of ADR for two hours at the 24-hour time point after seeding to induce DNA damage and p53 expression. Total RNA was extracted 48 hours after treatment for RT-qPCR analysis.

RNA extraction and RT-qPCR

After treatment, total RNA was extracted from cells using an RNeasy Plus Mini Kit (Qiagen, Valencia, CA, USA) according to the manufacturer's protocol. SuperScript III reverse transcriptase (Invitrogen, Carlsbad, CA, USA) was used to synthesize cDNA from collected RNA samples. The resulting cDNA were employed as templates for qPCR using SYBR Green Master Mix and Light Cycler 480 machine (Roche Applied Science, Basel, Switzerland). The expression level of the housekeeping gene ACTB was used for normalization. The primer sequences are shown in Appendix Table A6.

RESULTS

1. Screening strategies

We designed a first strategy integrating 1) ChIP-seq data targeting p53 protein¹⁹ (Table 4), 2) SNPs datasets from cancer GWAS after imputation (Appendix Table A5), and 3) p53 binding motifs predicted from JASPAR database²¹, to identify cancer associated SNPs located in p53 binding sites (Figure 2.1). Around 103,000 variants of each cancer GWAS dataset overlapped with ChIP-seq data targeting p53. Among these SNPs, 139 of them were associated significantly to at least one type of cancer (Table 5). We used the genomic sequence from 19 nucleotides upstream to downstream of each resulting variant as input for p53 binding motif prediction through JASPAR database. Consequently, we obtained two SNPs: SNP 1 and SNP 2, which were significantly associated with lung cancer and prostate cancer, respectively (Table 6). The variant SNP 1 overlapped with one single study of the ChIP-seq dataset (Figure 2.2 A). The predicted p53 responsive element containing SNP 1 matched to the MA0106.2 model from JASPAR database (Figure 2.3 A). The absolute binding score of the predicted site decreased from 8.52611, obtained with the risk allele A/T, to 7.39358, calculated with the non-risk allele G/C. The other variant, SNP 2, overlapped with two studies from the ChIP-seq dataset (Figure 2.2 B). The binding motif enclosing the SNP was matched to two models from JASPAR: MA0106.2 and MA0106.3 (Figure 2.3 B). The absolute score matching to MA0106.3 model decreased from 13.9749, obtained with the risk allele T/A, to 12.8713, calculated with the alternate allele C/G. Similarly, the binding score matching to MA0106.2 was 9.17156 and 7.45125 for the risk allele A/T and for the non-risk allele G/C, respectively.

For the second strategy, we first applied our in-house p53 responsive element search algorithm to the dataset of ChIP-seq targeting p53 protein constituted of 135,467 peaks (Figure 2.4). We employed the canonical p53 binding motif model to search for functional p53 binding sites: RRRCWWGYYY + spacer + RRRCWWGYYY, with a spacer from 0 to 13 bases. 32,388 motifs were overlapping with p53 binding regions of the ChIP-seq data. Then, we investigated whether these sequences contained cancer associated variants (Table 7). Our results identified two SNPs: SNP 2 and SNP 3, which were significantly associated with prostate cancer and stomach

cancer, respectively (Table 6). The variant SNP 2 overlapped with the results of the first strategy, which confirmed the outcome of our screenings. The SNP is located on the edge of the first half-site of p53 responsive element with a predicted binding score of 12.74263843 for the risk allele T/A (Figure 2.5 A). On the other hand, SNP 3 is located at the essential position for C in the CWWG core motif of the second half-site (Figure 2.5 B). The binding score calculated with the risk allele C/G is 12.31423601. Therefore, we suggest that p53 binding affinity might decrease when the allele switches to the alternate allele T/A. Moreover, the variant overlapped with from a p53 binding region containing twelve peaks from the ChIP-seq dataset (Figure 2.2 C). In conclusion, we obtained three cancer associated SNPs located within a predicted p53 binding motif from the two screening strategies: SNP 1, SNP 2, and SNP 3.

Table 4. Data of ChIP-seq targeting p53 protein gathered from ReMap 2018 database

	Cell type	Treatment	PMID	GEO
1	SaOS2	Wild-type p53	23966881	ERP002038
2	U2OS	UV	24289924	ERP004176
3	BC3	Nutlin-3a, TPA	26891221	ERP014035
4	H1299	p53R273H	-	ERP022459
5	SaOS2	Dox	21459846	GSE15780
6	U2Os	Actinomycin D, etoposide	21394211	GSE21939/22186
7	IMR90	5-FU	22127205	GSE31558
8	hESCs	Adriamycin	24078252	GSE39912
8	IMR90	Etoposide, p53 shRNA	22802529	GSE42728
10	U2OS	ADR, nutlin-3a	23775793	GSE46642
11	MCF7	Nutlin-3a	25058159	GSE47041-3
12	SaOS2	Wild-type p53	-	GSE51268
13	IMR90	Etoposide, RAS, RAS/E1A	25790137	GSE53491
14	GM00011, GM06170	ADR	25883152	GSE55727
15	HFK	ADR, cisplatin	24823795	GSE56674
16	HCT116	5-FU	25524025	GSE58506
17	HCT116	5-FU	25415302	GSE58714
18	IMR90	Nutlin-3a	25391375	GSE58740
19	HCT116	Ionizing radiations	25996291	GSE60267
20	HEPG2	UV	27302019	GSE64877
21	SW480	pRTR-p53-VSV, Dox	26183718	GSE67108
22	HCT116	Camptothecin, KO-ATF3	27146783	GSE74355

TPA, Tissue Plasminogen Activator; 5-FU, Fluorouracil.

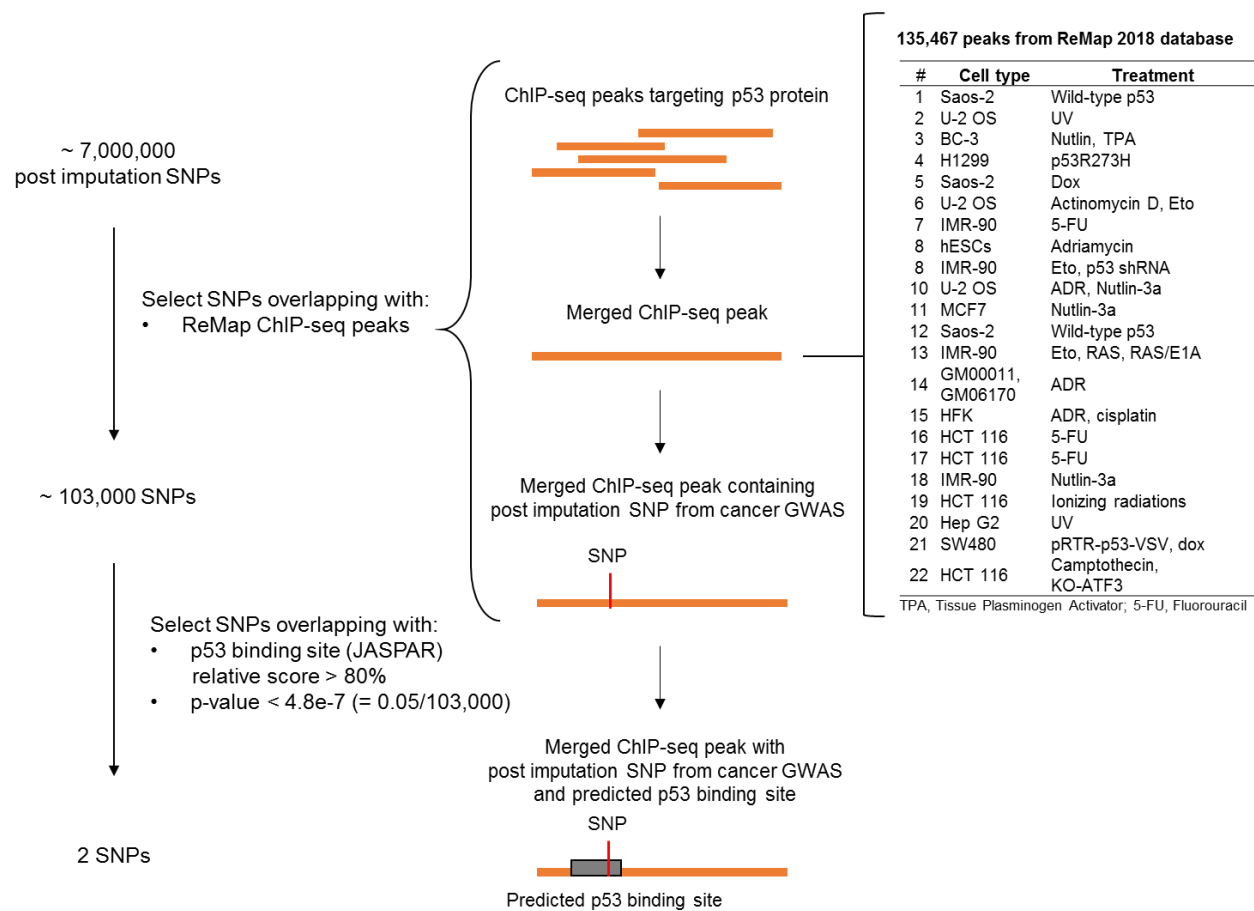


Figure 2.1 Overview of screening strategy 1

This screening strategy identified SNP 1 and SNP 2 associated to lung cancer and prostate cancer, respectively.

Table 5. Results of screening strategy 1 identifying cancer associated variants overlapping with p53 ChIP-seq data

Cancer types	SNPs \cap ReMap2018	SNPs \cap ReMap2018 p-value < threshold	
Bladder	103,096	6 (chr6)	
Breast	103,125	7 (chr3, 5, 10)	
Colon/rectal	103,124	17 (chr1, 5, 8, 10, 11, 18, 20)	
Endometrial	104,168	0	
Endometrioid adenocarcinoma	104,154	0	rs10808556 *
Epithelial ovarian	104,162	2 (chr22)	rs6983267
Esophageal	103,118	16 (chr4,12)	
Gallbladder	102,363	5 (chr1, 7, 8, 11, 20)	
Liver - HBV associated	104,074	1 (chr15)	
Liver - HCV associated	104,153	0	rs11065750 *
Lung	103,124	7 (chr1, 3, 5, 6)	rs11065751
Ovarian	104,168	0	rs11065752
Prostate	103,123	24 (chr2, 3, 6, 8, 10)	rs117139109
Stomach	103,124	60 (chr1, 5, 8, 12, 20)	

The numeric values indicate the number of SNPs; \cap , overlapping.

* SNPs resulted from two cancer types indicated by the hook.

Table 6. The association results between the variants obtained from the two screening strategies and each cancer type

Cancer type	SNP 1		SNP 2		SNP 3	
	OR	p-value	OR	p-value	OR	p-value
Bladder	1.005	0.95	1.037	0.65	1.093	0.17
Breast	0.983	0.56	0.966	0.23	0.950	0.03
Colon/rectal	1.037	0.13	1.010	0.68	0.997	0.86
Endometrial	1.007	0.91	0.961	0.51	0.849	6.60×10^{-4}
Endometrioid adenocarcinoma	0.944	0.48	0.977	0.79	0.795	5.79×10^{-4}
Epithelial ovarian	1.065	0.49	1.074	0.45	1.123	0.12
Esophageal	0.995	0.93	1.051	0.35	0.990	0.82
Gallbladder	1.010	0.97	0.353	7.96×10^{-5}	0.964	0.89
Liver - HBV associated	1.148	0.36	0.925	0.61	1.246	0.07
Liver - HCV associated	0.988	0.86	1.009	0.90	1.059	0.30
Lung	0.790	1.78×10^{-15}	1.009	0.77	0.954	0.06
Ovarian	1.008	0.90	1.063	0.40	1.088	0.14
Prostate	0.938	0.03	1.273	2.05×10^{-13}	0.966	0.16
Stomach	0.987	0.61	0.962	0.13	0.912	6.89×10^{-6}

Risk allele SNP 1: A/T (lung cancer); risk allele SNP 2: T/A (prostate cancer); risk allele SNP 3 C/G (stomach cancer).

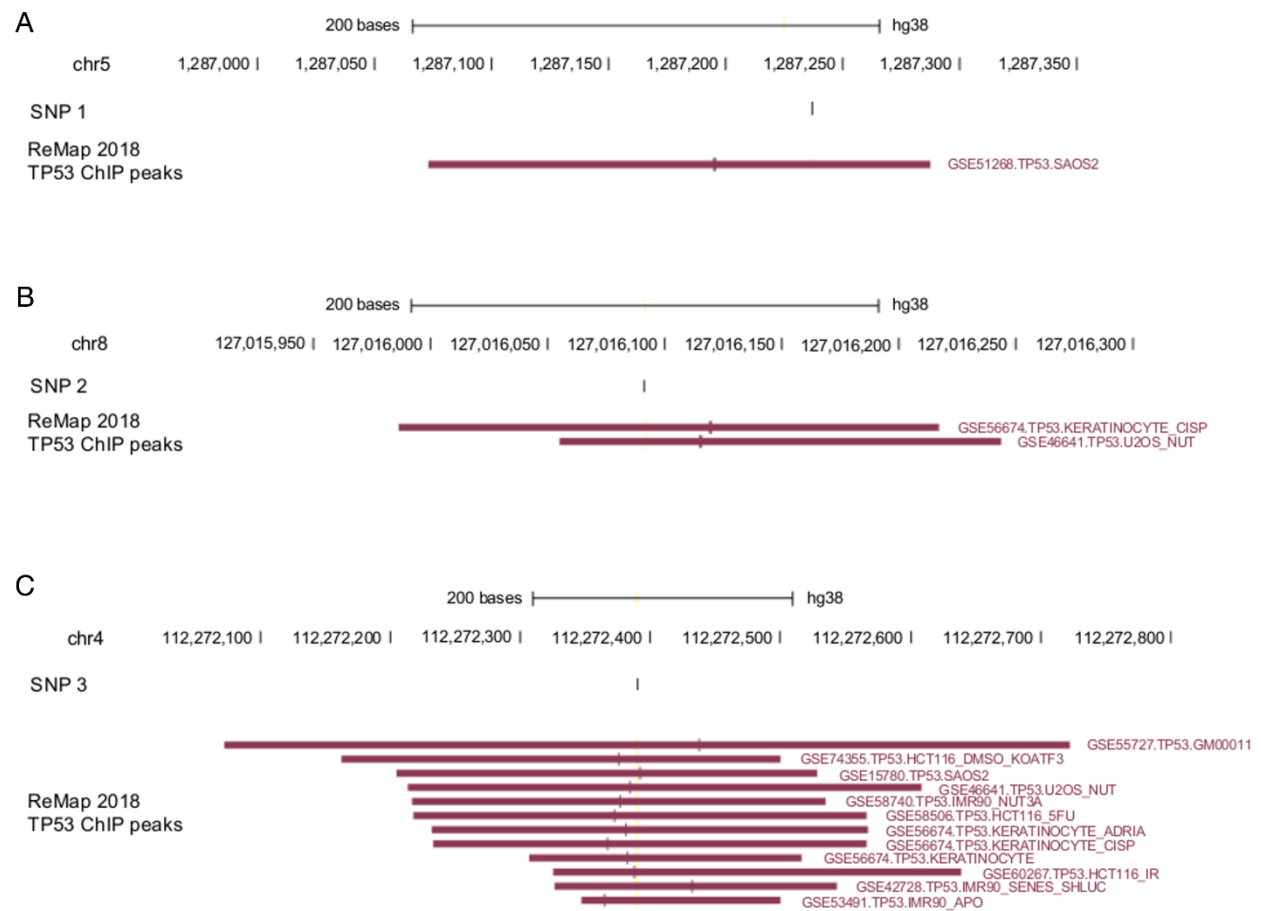


Figure 2.2 Each candidate variant overlaps with not less than one study from the p53 ChIP-seq dataset

UCSC Genome Browser view of each candidate variant: SNP 1 (A), SNP 2 (B), or SNP 3 (C), integrated with ChIP-seq data targeting p53 protein obtained from the ReMap 2018 database. The GEO accession number, cell type, and treatment information for the corresponding peak are indicated on the right side of each peak.

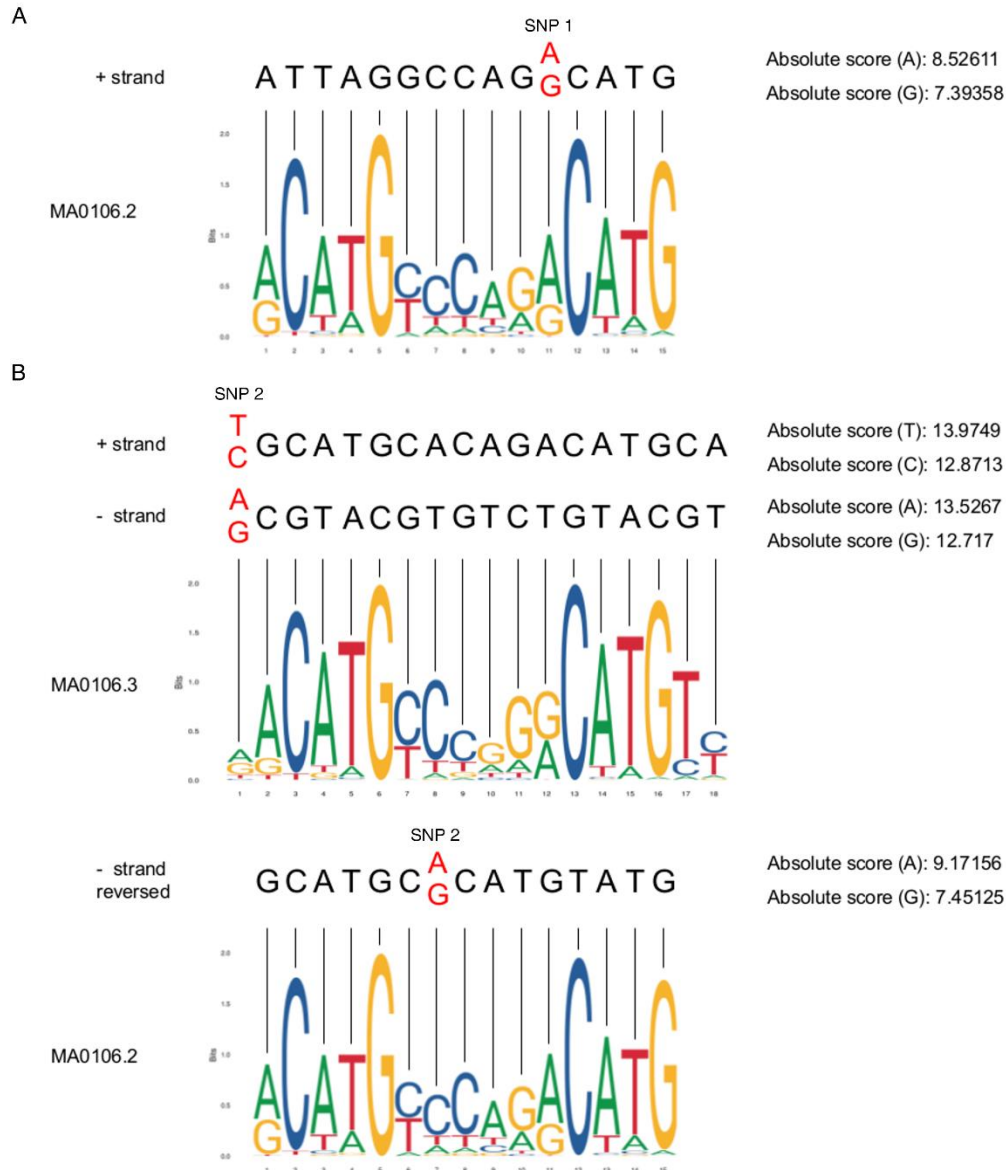


Figure 2.3 Binding sites predicted by JASPAR database for p53 responsive elements containing SNPs obtained from screening strategy 1

(A) The DNA sequence overlapping SNP 1 contains a binding site predicted by JASPAR with a relative score over 80% matching to MA0106.2. The absolute scores obtained with different alleles are indicated on the right. The variant is indicated by a red font color.

(B) Three other binding sites were predicted in the DNA sequence overlapping SNP 2. One binding site, matching to MA0106.2, is hidden in the binding motif for MA0106.3 and is not represented in the figure. The absolute scores obtained with different alleles are indicated on the right. The variant is indicated by a red font color.

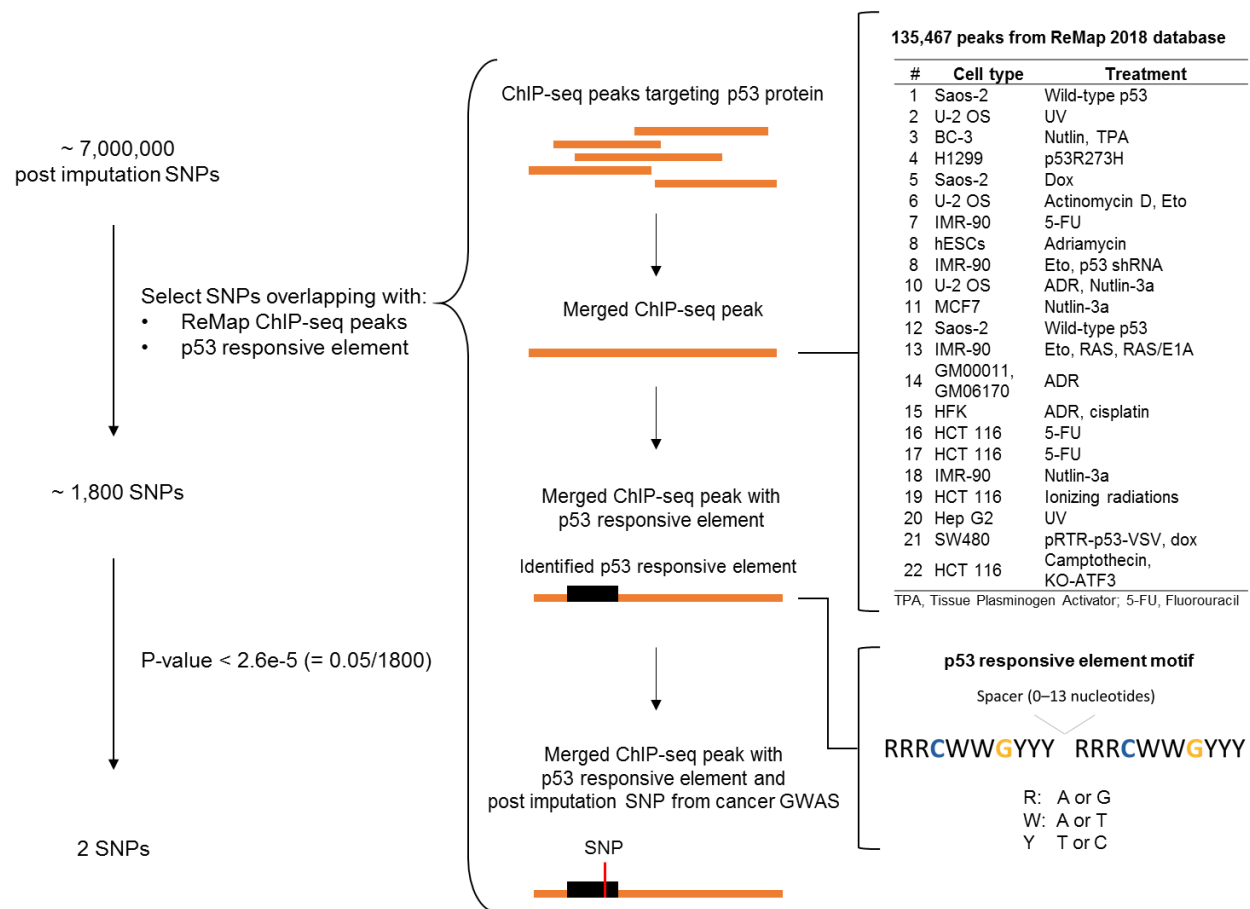


Figure 2.4 Overview of screening strategy 2

This screening strategy identified SNP 2 and SNP 3 associated to prostate cancer and stomach cancer, respectively.

Table 7. Results of screening strategy 2 identifying cancer associated variants overlapping with p53 ChIP-seq data and predicted p53 responsive elements

Cancer types	SNPs \cap ReMap2018 \cap	SNPs \cap ReMap2018 \cap p53 motifs
	p53 motifs	p-value < threshold
Bladder	1850	0
Breast	1852	0
Colon/rectal	1852	0
Endometrial	1884	0
Endometrioid adenocarcinoma	1882	0
Epithelial ovarian	1884	0
Esophageal	1852	0
Gallbladder	1853	0
Liver - HBV associated	1882	0
Liver - HCV associated	1884	0
Lung	1852	0
Ovarian	1884	0
Prostate	1852	1 (chr8)
Stomach	1852	1 (chr4)

The numeric values indicate the number of SNPs; \cap , overlapping.

RRRCWWGYYYRRRCWWGYYY

A

SNP2

G^t_C GCA TGCaCAGACA TGCaC

Score: 12.742638
Match: 17/20
Core mismatch: 0
Spacer: 0

B

SNP 3

GGGCA T G^C_t aCCAGA A TGCgC

Score: 12.314236
Match: 18/20
Core mismatch: 0
Spacer: 0

Figure 2.5 p53 binding sites obtained with strategy 2

Sequence alignments between the canonical consensus p53 binding motif and the responsive element predicted with SNP 2 (A) or with SNP 3 (B). SNPs are indicated in red. Unmatched nucleotides are represented in lowercase. The predicted binding score is stated on the right. The core mismatch indicate any mismatch at C or G from the CWWG motif. R, purine; W, A or T; Y, pyrimidine.

2. *Confirmation of p53 binding to identified responsive elements*

Next, we performed reporter assay using H1299 *p53* null cells to evaluate the binding affinity between p53 and the obtained binding sites containing cancer associated SNPs. The identified binding site sequence (BS) of each SNP was cloned into pGL4.24 plasmids in both sense and antisense strands, and the constructs were co-transfected with pcDNA3.1 vehicle plasmid or pcDNA3.1/p53 into cells (Figure 2.6). The luciferase activity detected for the BS construct containing SNP 1 was negligible, indicating absence or low interaction between p53 and the predicted BS. However, BS enclosing SNP 2 showed higher luciferase activity with the risk allele T/A as compared to the non-risk allele C/G, independently from the DNA strand. Moreover, BS containing SNP 3 exhibited higher luciferase activity with the risk allele C/G as compared to the non-risk allele T/A in the sense construct (same direction as the + strand of the genome). These results are consistent with the binding scores predicted by the JASPAR database and our in-house algorithm, respectively. Consequently, we postulated that SNP 2 and SNP 3 are located in p53 BS with potential to alter the expression levels of the regulated genes. In addition, we excluded SNP 1 from the following analyses due to its poor effect on p53 binding affinity.

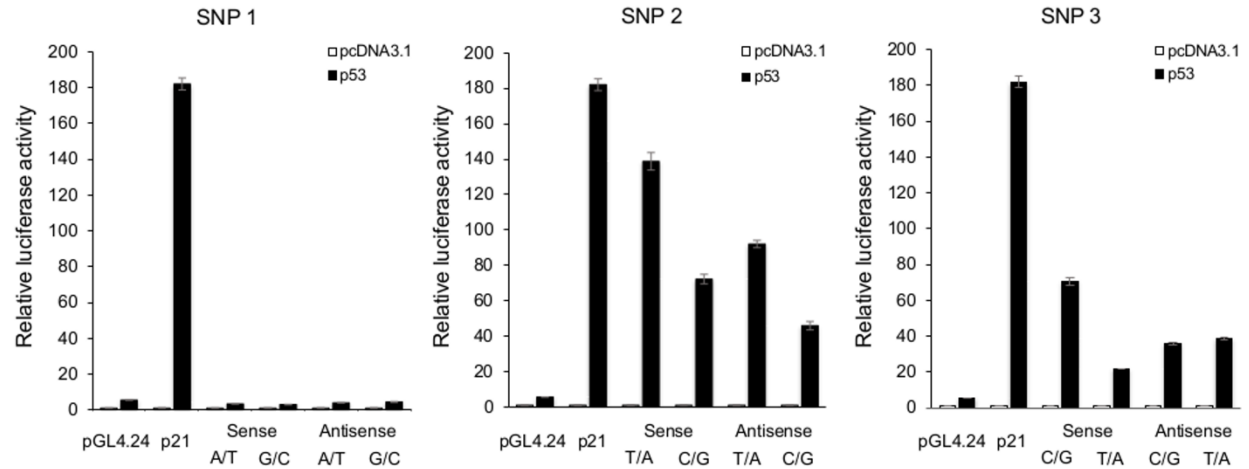


Figure 2.6 The risk alleles of SNP 2 and SNP 3 are associated with a higher p53 binding affinity

Luciferase assay of H1299 cells expressing the empty vector pGL4.24, sense (+ strand) or antisense (- strand) of the BS construct containing the risk (left side) or non-risk allele (right side) of the obtained SNPs. The control vector pcDNA3.1 or the human wild-type p53 was co-transfected at the same time. Luciferase activity is indicated relative to the activity of the control vectors pcDNA3.1. Error bars, S.D. n=3.

3. Gene expression analysis of candidates possibly regulated by the obtained SNPs

The distance between the p53 binding site and the target genes vary from 5⁴⁹ to 100⁵⁰ kb. Therefore, we selected the genes at a distance of within 100 kb from our candidate variants. SNP 2 is an intron variant of gene *X* (Figure 2.7 A). This gene encodes a long non-coding RNA that promotes prostate cancer cell proliferation⁵¹ and is involved in the tumorigenesis of prostate cancer⁵². We also included two other genes downstream to SNP 2, gene *Y* and *Z*, for further investigation. Then, we compared the expression levels of the candidate genes between *p53* wild-type or *p53* mutant tumor samples in the prostate adenocarcinoma (PRAD) of TCGA. However, the expression data for genes *X*, *Y*, and *Z* non-coding RNAs were unavailable in the PRAD cohort.

In the case of SNP 3, the variant is located in the non-coding region of the last exon of gene *B* (Figure 2.7 B). We also considered genes *A*, *C*, and *D* as candidate genes for analysis as they were at a distance of within 100 kb from the SNP. We examined the expression of these genes between *p53* wild-type or *p53* mutant tumor samples in the stomach adenocarcinoma (STAD) cohort of TCGA (Figure 2.8 A). Consequently, genes *A* and *D* showed significant higher expression in *p53* mutant samples as compared to *p53* wild-type samples. Furthermore, we analyzed expression quantitative trait locus (eQTL) data from The Genotype-Tissue Expression (GTEx) database to investigate the association of SNP 3 with the candidate genes expression. The expression of gene *A* was associated with the variant in stomach tissue with a higher expression level associated with the homozygous risk allele C/G (Figure 2.8 B). These results indicated that gene *A* and *D* might be regulated by p53 in stomach tissue.

The ChIP-seq data from ReMap 2018 database showed that p53 protein binds to the predicted binding motifs for both SNP 2 and SNP 3 in U2OS cells treated with nutlin-3a (Figure 2.7). We used ADR instead of nutlin-3a for DNA damage and we inspected the expression levels of the candidate genes in U2OS cells by performing RT-qPCR. An siRNA against *p53* was added to evaluate the differential expression of the candidate genes. RT-qPCR results revealed a p53-dependent expression for genes *X*, *Y* and *Z* in U2OS cells (Figure 2.9 A). On the other side, the expression of genes *A*, *B*, *C*, and *D* were not significantly affected by p53 in U2OS cells. We replicated the experiment with MKN45 gastric adenocarcinoma cells under the same conditions (Figure 2.9 B), but the expression of the candidates were not induced by p53. Genotyping results

revealed that U2OS presents homozygous risk allele for SNP 2 and heterozygous alleles for SNP 3. MKN45 was as well heterozygous for SNP 3, suggesting an explanation for the insignificant alteration of *A*, *B*, *C*, and *D* expression by p53. In conclusion, the binding affinity of p53 to SNP 2 BS varies according to the allele of the variant and modifies the expression levels of genes *X*, *Y* and *Z*, possibly oncogenes inducing prostate cancer. The risk allele of SNP 2 is associated with a higher binding rate of p53, thus induces the expression levels of genes *X*, *Y* and *Z*.

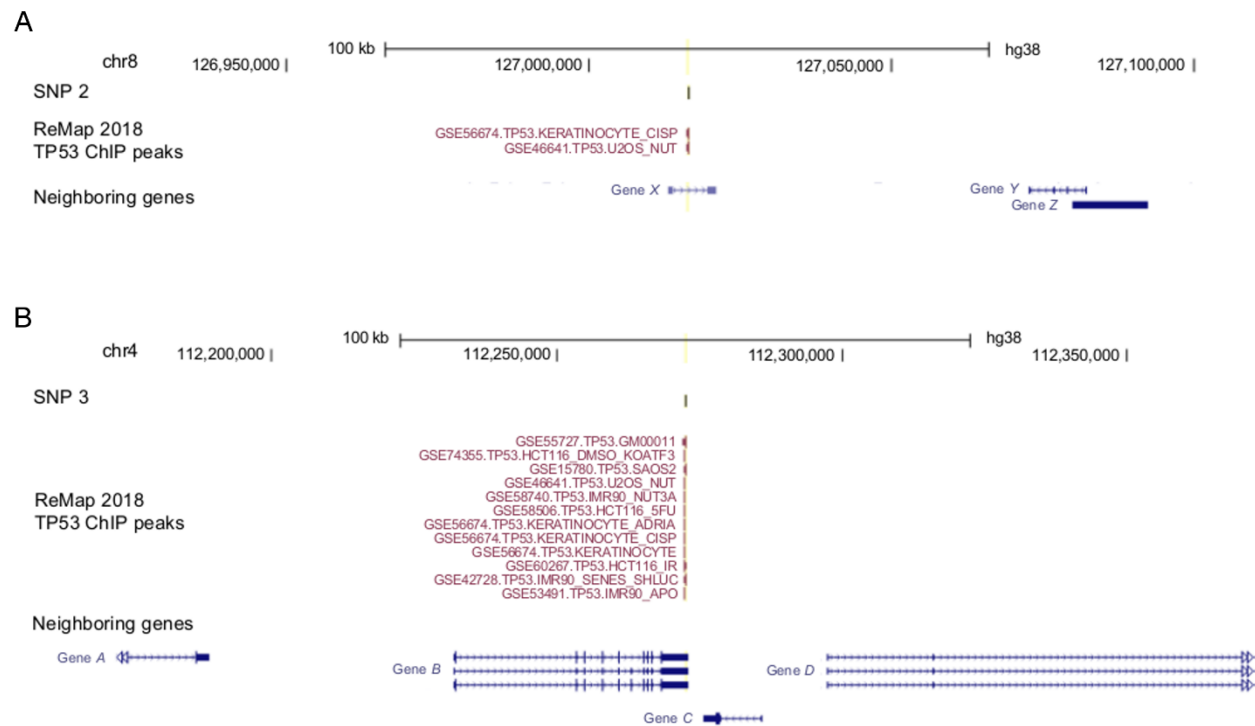


Figure 2.7 Overview of the genomic regions at a distance of within 100 kb from the candidate variants

UCSC Genome Browser view of the genes located between -100 kb and + 100 kb from SNP 2 (A) and SNP 3 (B) integrated with ChIP-seq data targeting p53 protein obtained from the ReMap 2018 database. The GEO accession number, cell type, and treatment information for the corresponding peak are indicated on the left side of each peak.

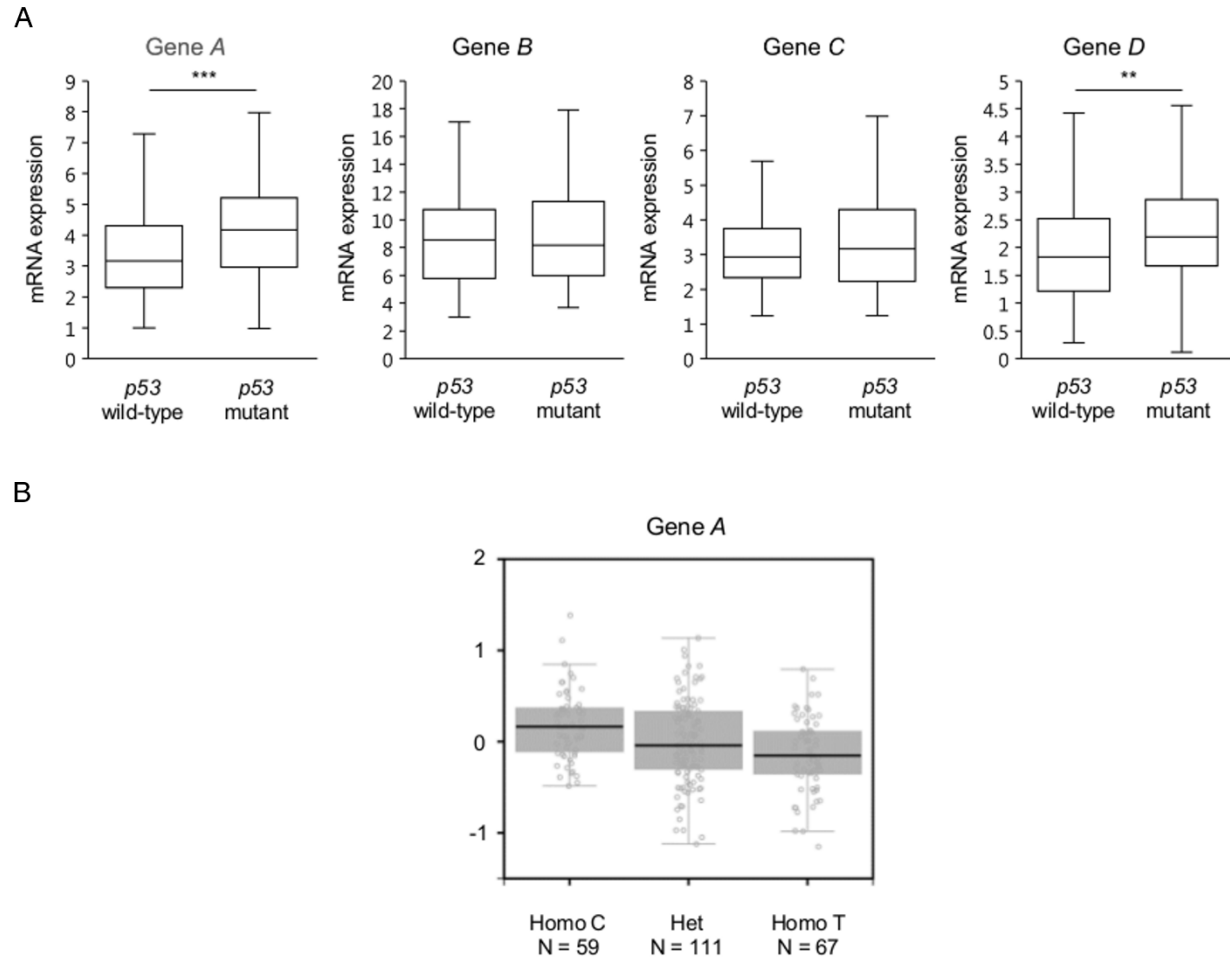


Figure 2.8 Gene expression analysis of the genes possibly regulated by SNP 3

(A) Comparison of levels of the candidate genes mRNA between TCGA *p53* wild-type and *p53* mutant in stomach adenocarcinoma (STAD). Two-tailed Student's *t*-test; * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

(B) eQTL analysis of SNP 3 showing an association of the homozygous risk allele C with a higher expression level of gene A. Homo, homozygous; Het, heterozygous.

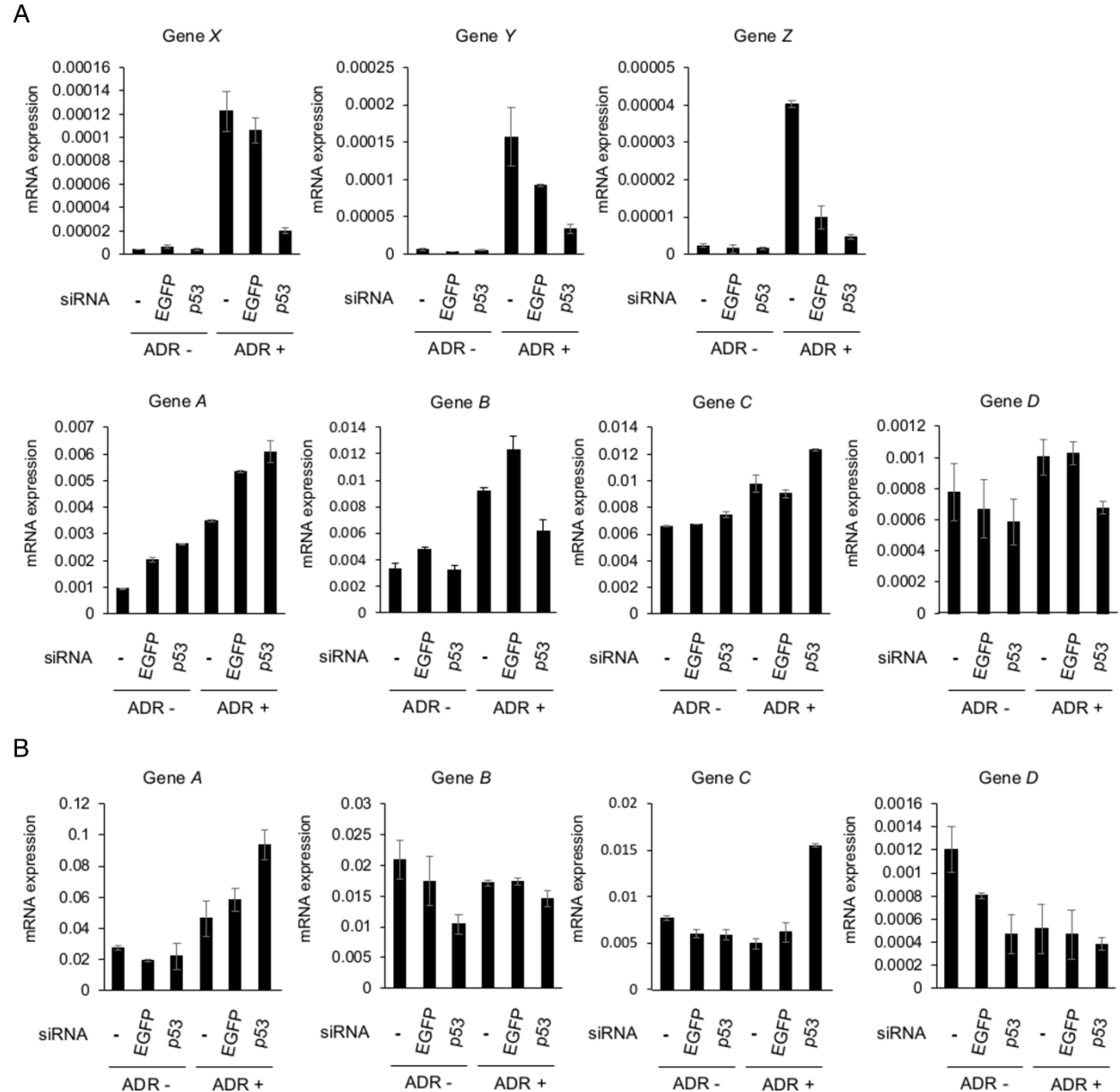


Figure 2.9 Evaluation of p53-dependent expression of the candidate genes

(A) Relative mRNA expression of the candidate genes possibly regulated by SNP 2 BS (upper panel) or SNP 3 BS (bottom panel) in U2OS cells normalized to ACTB. Cells were treated with or without an siRNA targeting either EGFP or p53 and treated with or without ADR. Error bars, S.D. n=2.

(B) Relative mRNA expression of SNP 3 BS-regulating candidate genes in MKN45 cells normalized to ACTB. Cells were treated with or without an siRNA targeting either EGFP or p53 and treated with or without ADR. Error bars, S.D. n=2.

DISCUSSION

Here, we designed two strategies identifying cancer associated SNPs located in p53 binding sites in the human genome. We integrated 1) data from ChIP-seq targeting p53 protein, 2) models predicting p53 responsive elements, and 3) post imputation cancer GWAS datasets. We identified SNP 1, SNP 2, and SNP 3 associated with lung cancer, prostate cancer, and stomach cancer, respectively. Reporter assays confirmed the interaction of p53 with SNP 2 and SNP 3, but not SNP 1. Moreover, the results showed a higher binding affinity of p53 with the risk allele in each BS. Further investigation failed to identify any nearby genes regulated by the p53 BS enclosing SNP 3. However, we suggest that this BS might functions as a distant enhancer, thus the regulated gene is located further away. We are also unable to neglect the possibility where the regulated gene is an unannotated non-coding RNA.

On the other hand, we identified three nearby genes, *X*, *Y* and *Z*, regulated by the BS containing SNP 2. The risk allele of SNP 2, locating at an essential position of a p53 binding site, increases the binding affinity of p53 and triggers higher expression of three oncogenes promoting prostate cancer: genes *X*, *Y* and *Z*. Furthermore, TCGA data showed relatively lower *p53* mutation frequency (12%) for prostate adenocarcinoma compared to other cancers, and a significantly higher *p53* expression in tumor samples as compared to controls in PRAD (data now shown). We suggest as hypothesis that wild-type *p53* is expressed additionally in prostate cancer and induces the expression of *X*, *Y* and *Z* oncogenes by means of higher p53 binding affinity provided by the risk allele T/A located within the binding site regulating the two genes.

The oncogenic function of gene *X* was originally found in prostate cancer. However, this long non-coding RNA was also reportedly involved in various cancers. High expression of gene *X* is related to poor prognosis in gastric cancer⁵³ and esophageal squamous carcinoma⁵⁴. Overexpression of gene *X* is associated with aggressiveness in colorectal cancer cells⁵⁵, as well as poor prognosis in colorectal cancer⁵⁶, gastric cancer⁵⁷ and non-small cell lung cancer cells⁵⁸. The gene also promotes cell proliferation and metastasis in hepatocellular carcinoma^{59,60}. Down-regulation of gene *X* produces growth arrest and apoptosis in colorectal cancer cells⁶¹, hepatocellular carcinoma⁵¹, and bladder cancer⁶². The tumorigenic functions of thee gene include the repression of the *BRCA2* tumor suppressor by post-transcriptional regulation of the 3'UTR⁶³

and the induction of p21 to promote cell proliferation, migration and invasion⁵⁷. Studies also reported that the down-regulation of gene *X* inhibits the Wnt/ β -catenin signaling pathway, which is crucial for cancer progression⁶⁴. Other research elucidated that a prostate cancer risk-associated variant rs7463708 upregulates gene *X* through the increase binding of a transcription factor to a distal enhancer of the gene⁶⁵. Taken together, the cellular functions involving gene *X* seem to be complex, though significantly associated with cancer.

As future plans, we intent to confirm the alteration of genes *X*, *Y*, and *Z* expression according to the allele of SNP 2 in prostate cancer cells. At the same time, we designed a third screening strategy merging first the post imputation GWAS datasets with all p53 responsive elements identified in the human genome. We expect to detect other cancer associated SNPs similar to the results of SNP 2. Our study will surely contribute for more understanding of the mechanisms associating SNPs with tumorigenesis.

CONCLUSION

This thesis consisted of a comprehensive analysis of p53 signaling pathway divided into two parts. In the first part, we focused on the regulation of p53 downstream target. We developed a screening strategy using mouse multi-organ transcriptome data to identify novel p53 downstream target involved in cancer mechanisms. Our analysis identified *INKA2* as a novel target regulated by p53 directly through an intronic p53 binding site conserved in various species. *INKA2* is induced in response to DNA damage and showed lower expression in various tumor samples. The gene also exhibited tumor suppressor function such as reduction of cancer cell growth. Further analysis clarified that INKA2 protein contains an iBox domain interacting with the catalytic domain of PAK4 serine-threonine kinase. Consequently, INKA2 likely inhibits PAK4 and down-regulates the amount of intracellular beta-catenin. In conclusion, we identified a novel p53 tumor suppressor signaling pathway involving the direct downstream target *INKA2*, followed by PAK4 and beta-catenin.

The second part of this thesis studied the role of p53 signaling pathway in human cancer genetics. We designed two different approaches identifying cancer associated SNPs located in p53 binding site and potentially modifying the binding of p53, thus the expression level of regulated genes. We identified three variants located in predicted p53 responsive elements, with only two of them showing p53 binding under in vitro conditions. Further investigation revealed that the risk allele of SNP 2 is related to the increase of interaction between p53 protein and the BS containing the variant, which regulates three oncogenes *X*, *Y*, and *Z* promoting prostate cancer. More confirmations and research are necessary to complete the results of this project and are currently ongoing.

Taken together, we investigated the p53 signaling pathway from two different perspectives and our results elucidated novel p53 regulation systems. Our research will surely contribute to more understanding of the mechanisms underlying tumorigenesis through p53 regulation.

REFERENCES

- 1 Vogelstein, B., Lane, D. & Levine, A. J. Surfing the p53 network. *Nature* **408**, 307-310, doi:10.1038/35042675 (2000).
- 2 Laptenko, O. & Prives, C. Transcriptional regulation by p53: one protein, many possibilities. *Cell Death Differ* **13**, 951-961, doi:10.1038/sj.cdd.4401916 (2006).
- 3 Valente, L. J. & Strasser, A. Distinct target genes and effector processes appear to be critical for p53-activated responses to acute DNA damage versus p53-mediated tumour suppression. **8**, doi:10.7750/BioDiscovery.2013.8.3 %J BioDiscovery (2013).
- 4 El-Deiry, W. S. *et al.* WAF1, a potential mediator of p53 tumor suppression. *Cell* **75**, 817-825, doi:[https://doi.org/10.1016/0092-8674\(93\)90500-P](https://doi.org/10.1016/0092-8674(93)90500-P) (1993).
- 5 Fischer, M. Census and evaluation of p53 target genes. *Oncogene* **36**, 3943-3956, doi:10.1038/onc.2016.502 (2017).
- 6 Botcheva, K. p53 binding to human genome: crowd control navigation in chromatin context. *Front Genet* **5**, 447, doi:10.3389/fgene.2014.00447 (2014).
- 7 Luo, T. *et al.* Inca: a novel p21-activated kinase-associated protein required for cranial neural crest development. *Development* **134**, 1279-1289, doi:10.1242/dev.02813 (2007).
- 8 Reid, B. S., Sargent, T. D. & Williams, T. Generation and characterization of a novel neural crest marker allele, Inka1-LacZ, reveals a role for Inka1 in mouse neural tube closure. *Dev Dyn* **239**, 1188-1196, doi:10.1002/dvdy.22248 (2010).
- 9 Iwasaki, Y., Yumoto, T. & Sakakibara, S. Expression profiles of inka2 in the murine nervous system. *Gene Expr Patterns* **19**, 83-97, doi:10.1016/j.gep.2015.08.002 (2015).
- 10 Baskaran, Y. *et al.* An in cellulo-derived structure of PAK4 in complex with its inhibitor Inka1. *Nat Commun* **6**, 8681, doi:10.1038/ncomms9681 (2015).
- 11 Bokoch, G. M. Biology of the p21-activated kinases. *Annu Rev Biochem* **72**, 743-781, doi:10.1146/annurev.biochem.72.121801.161742 (2003).
- 12 Jaffer, Z. M. & Chernoff, J. p21-activated kinases: three more join the Pak. *Int J Biochem Cell Biol* **34**, 713-717 (2002).
- 13 Zhang, G. *et al.* Genetic Associations with Gestational Duration and Spontaneous Preterm Birth. *N Engl J Med* **377**, 1156-1167, doi:10.1056/NEJMoa1612665 (2017).

- 14 Sikora, M. J. *et al.* WNT4 mediates estrogen receptor signaling and endocrine resistance in invasive lobular carcinoma cell lines. *Breast Cancer Res* **18**, 92, doi:10.1186/s13058-016-0748-7 (2016).
- 15 Tanikawa, C. *et al.* The Transcriptional Landscape of p53 Signalling Pathway. *EBioMedicine* **20**, 109-119, doi:10.1016/j.ebiom.2017.05.017 (2017).
- 16 Miyamoto, T. *et al.* Identification of a p53-repressed gene module in breast cancer cells. *Oncotarget* **8**, 55821-55836, doi:10.18632/oncotarget.19608 (2017).
- 17 Koguchi, T., Tanikawa, C., Mori, J., Kojima, Y. & Matsuda, K. Regulation of myo-inositol biosynthesis by p53-ISYNA1 pathway. *Int J Oncol* **48**, 2415-2424, doi:10.3892/ijo.2016.3456 (2016).
- 18 Yodsurang, V. *et al.* Identification of a novel p53 target, COL17A1, that inhibits breast cancer cell migration and invasion. *Oncotarget* **8**, 55790-55803, doi:10.18632/oncotarget.18433 (2017).
- 19 Cheneby, J., Gheorghe, M., Artufel, M., Mathelier, A. & Ballester, B. ReMap 2018: an updated atlas of regulatory regions from an integrative analysis of DNA-binding ChIP-seq experiments. *Nucleic Acids Res* **46**, D267-D275, doi:10.1093/nar/gkx1092 (2018).
- 20 Zhou, K. R. *et al.* ChIPBase v2.0: decoding transcriptional regulatory networks of non-coding RNAs and protein-coding genes from ChIP-seq data. *Nucleic Acids Res* **45**, D43-D50, doi:10.1093/nar/gkw965 (2017).
- 21 Khan, A. *et al.* JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res* **46**, D260-D266, doi:10.1093/nar/gkx1126 (2018).
- 22 Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**, 671-675 (2012).
- 23 Takahashi, Y. *et al.* Regulation of tubular recycling endosome biogenesis by the p53-MICALL1 pathway. *Int J Oncol* **51**, 724-736, doi:10.3892/ijo.2017.4060 (2017).
- 24 Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* **2**, 401-404, doi:10.1158/2159-8290.CD-12-0095 (2012).
- 25 Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* **6**, pl1, doi:10.1126/scisignal.2004088 (2013).

- 26 Huang, W. Y. *et al.* MethHC: a database of DNA methylation and gene expression in human cancer. *Nucleic Acids Res* **43**, D856-861, doi:10.1093/nar/gku1151 (2015).
- 27 Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74, doi:10.1038/nature11247 (2012).
- 28 Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res* **12**, 996-1006, doi:10.1101/gr.229102 (2002).
- 29 Ha, B. H. *et al.* Type II p21-activated kinases (PAKs) are regulated by an autoinhibitory pseudosubstrate. *Proc Natl Acad Sci U S A* **109**, 16107-16112, doi:10.1073/pnas.1214447109 (2012).
- 30 Wang, W., Lim, L., Baskaran, Y., Manser, E. & Song, J. NMR binding and crystal structure reveal that intrinsically-unstructured regulatory domain auto-inhibits PAK4 by a mechanism different from that of PAK1. *Biochem Biophys Res Commun* **438**, 169-174, doi:10.1016/j.bbrc.2013.07.047 (2013).
- 31 Li, Y. *et al.* Nucleo-cytoplasmic shuttling of PAK4 modulates beta-catenin intracellular translocation and signaling. *Biochim Biophys Acta* **1823**, 465-475, doi:10.1016/j.bbamcr.2011.11.013 (2012).
- 32 Debebe, A. *et al.* Wnt/beta-catenin activation and macrophage induction during liver cancer development following steatosis. *Oncogene* **36**, 6020-6029, doi:10.1038/onc.2017.207 (2017).
- 33 Lee, K. & G, A. P. The interaction between the Wnt/beta-catenin signaling cascade and PKG activation in cancer. *J Biomed Res* **31**, 189-196, doi:10.7555/JBR.31.20160133 (2017).
- 34 Chong, P. S. Y. *et al.* Non-canonical activation of beta-catenin by PRL-3 phosphatase in acute myeloid leukemia. *Oncogene*, doi:10.1038/s41388-018-0526-3 (2018).
- 35 Zhang, X. *et al.* PAK4 regulates G6PD activity by p53 degradation involving colon cancer cell growth. *Cell Death Dis* **8**, e2820, doi:10.1038/cddis.2017.85 (2017).
- 36 He, L. F. *et al.* Activated-PAK4 predicts worse prognosis in breast cancer and promotes tumorigenesis through activation of PI3K/AKT signaling. *Oncotarget* **8**, 17573-17585, doi:10.18632/oncotarget.7466 (2017).

- 37 Cai, S. *et al.* Overexpression of P21-activated kinase 4 is associated with poor prognosis in non-small cell lung cancer and promotes migration and invasion. *J Exp Clin Cancer Res* **34**, 48, doi:10.1186/s13046-015-0165-2 (2015).
- 38 Siu, M. K. *et al.* p21-activated kinase 4 regulates ovarian cancer cell proliferation, migration, and invasion and contributes to poor prognosis in patients. *Proc Natl Acad Sci U S A* **107**, 18622-18627, doi:10.1073/pnas.0907481107 (2010).
- 39 Murray, B. W. *et al.* Small-molecule p21-activated kinase inhibitor PF-3758309 is a potent inhibitor of oncogenic signaling and tumor growth. *Proc Natl Acad Sci U S A* **107**, 9446-9451, doi:10.1073/pnas.0911863107 (2010).
- 40 Xu, H. T. *et al.* PAK4 Phosphorylates p53 at Serine 215 to Promote Liver Cancer Metastasis. *Cancer Res* **76**, 5732-5742, doi:10.1158/0008-5472.CAN-15-3373 (2016).
- 41 Idogawa, M. *et al.* Identification and analysis of large intergenic non-coding RNAs regulated by p53 family members through a genome-wide analysis of p53-binding sites. *Hum Mol Genet* **23**, 2847-2857, doi:10.1093/hmg/ddt673 (2014).
- 42 Smeenk, L. *et al.* Characterization of genome-wide p53-binding sites upon stress response. *Nucleic Acids Res* **36**, 3639-3654, doi:10.1093/nar/gkn232 (2008).
- 43 Hirata, M. *et al.* Overview of BioBank Japan follow-up data in 32 diseases. *J Epidemiol* **27**, S22-S28, doi:10.1016/j.je.2016.12.006 (2017).
- 44 Nagai, A. *et al.* Overview of the BioBank Japan Project: Study design and profile. *J Epidemiol* **27**, S2-S8, doi:10.1016/j.je.2016.12.005 (2017).
- 45 International HapMap, C. The International HapMap Project. *Nature* **426**, 789-796, doi:10.1038/nature02168 (2003).
- 46 Li, Y., Willer, C. J., Ding, J., Scheet, P. & Abecasis, G. R. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol* **34**, 816-834, doi:10.1002/gepi.20533 (2010).
- 47 Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet* **44**, 955-959, doi:10.1038/ng.2354 (2012).
- 48 Genomes Project, C. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65, doi:10.1038/nature11632 (2012).

- 49 Menendez, D. *et al.* Diverse stresses dramatically alter genome-wide p53 binding and transactivation landscape in human cancer cells. *Nucleic Acids Res* **41**, 7286-7301, doi:10.1093/nar/gkt504 (2013).
- 50 Wei, C. L. *et al.* A global map of p53 transcription-factor binding sites in the human genome. *Cell* **124**, 207-219, doi:10.1016/j.cell.2005.10.043 (2006).
- 51 Prensner, J. R. *et al.* The long non-coding RNA PCAT-1 promotes prostate cancer cell proliferation through cMyc. *Neoplasia* **16**, 900-908, doi:10.1016/j.neo.2014.09.001 (2014).
- 52 Xu, W., Chang, J., Du, X. & Hou, J. Long non-coding RNA PCAT-1 contributes to tumorigenesis by regulating FSCN1 via miR-145-5p in prostate cancer. *Biomed Pharmacother* **95**, 1112-1118, doi:10.1016/j.biopha.2017.09.019 (2017).
- 53 Cui, W. C., Wu, Y. F. & Qu, H. M. Up-regulation of long non-coding RNA PCAT-1 correlates with tumor progression and poor prognosis in gastric cancer. *Eur Rev Med Pharmacol Sci* **21**, 3021-3027 (2017).
- 54 Shi, W. H. *et al.* Upregulation of the long noncoding RNA PCAT-1 correlates with advanced clinical stage and poor prognosis in esophageal squamous carcinoma. *Tumour Biol* **36**, 2501-2507, doi:10.1007/s13277-014-2863-3 (2015).
- 55 Qiao, L. *et al.* Knockdown of long non-coding RNA prostate cancer-associated ncRNA transcript 1 inhibits multidrug resistance and c-Myc-dependent aggressiveness in colorectal cancer Caco-2 and HT-29 cells. *Mol Cell Biochem* **441**, 99-108, doi:10.1007/s11010-017-3177-8 (2018).
- 56 Ge, X. *et al.* Overexpression of long noncoding RNA PCAT-1 is a novel biomarker of poor prognosis in patients with colorectal cancer. *Med Oncol* **30**, 588, doi:10.1007/s12032-013-0588-6 (2013).
- 57 Bi, M., Yu, H., Huang, B. & Tang, C. Long non-coding RNA PCAT-1 over-expression promotes proliferation and metastasis in gastric cancer cells through regulating CDKN1A. *Gene* **626**, 337-343, doi:10.1016/j.gene.2017.05.049 (2017).
- 58 Zhao, B., Hou, X. & Zhan, H. Long non-coding RNA PCAT-1 over-expression promotes proliferation and metastasis in non-small cell lung cancer cells. *Int J Clin Exp Med* **8**, 18482-18487 (2015).

- 59 Wen, J., Xu, J., Sun, Q., Xing, C. & Yin, W. Upregulation of long non coding RNA PCAT-1 contributes to cell proliferation, migration and apoptosis in hepatocellular carcinoma. *Mol Med Rep* **13**, 4481-4486, doi:10.3892/mmr.2016.5075 (2016).
- 60 Zhang, D. *et al.* Long noncoding RNA PCAT-1 promotes invasion and metastasis via the miR-129-5p-HMGB1 signaling pathway in hepatocellular carcinoma. *Biomed Pharmacother* **95**, 1187-1193, doi:10.1016/j.biopha.2017.09.045 (2017).
- 61 Qiao, L. *et al.* Down regulation of the long non-coding RNA PCAT-1 induced growth arrest and apoptosis of colorectal cancer cells. *Life Sci* **188**, 37-44, doi:10.1016/j.lfs.2017.08.024 (2017).
- 62 Liu, L. *et al.* Inducing cell growth arrest and apoptosis by silencing long non-coding RNA PCAT-1 in human bladder cancer. *Tumour Biol* **36**, 7685-7689, doi:10.1007/s13277-015-3490-3 (2015).
- 63 Prensner, J. R. *et al.* PCAT-1, a long noncoding RNA, regulates BRCA2 and controls homologous recombination in cancer. *Cancer Res* **74**, 1651-1660, doi:10.1158/0008-5472.CAN-13-3159 (2014).
- 64 Zhang, F. *et al.* Long noncoding RNA PCAT1 regulates extrahepatic cholangiocarcinoma progression via the Wnt/beta-catenin-signaling pathway. *Biomed Pharmacother* **94**, 55-62, doi:10.1016/j.biopha.2017.07.025 (2017).
- 65 Guo, H. *et al.* Modulation of long noncoding RNAs by risk SNPs underlying genetic predispositions to prostate cancer. *Nat Genet* **48**, 1142-1150, doi:10.1038/ng.3637 (2016).

APPENDIX

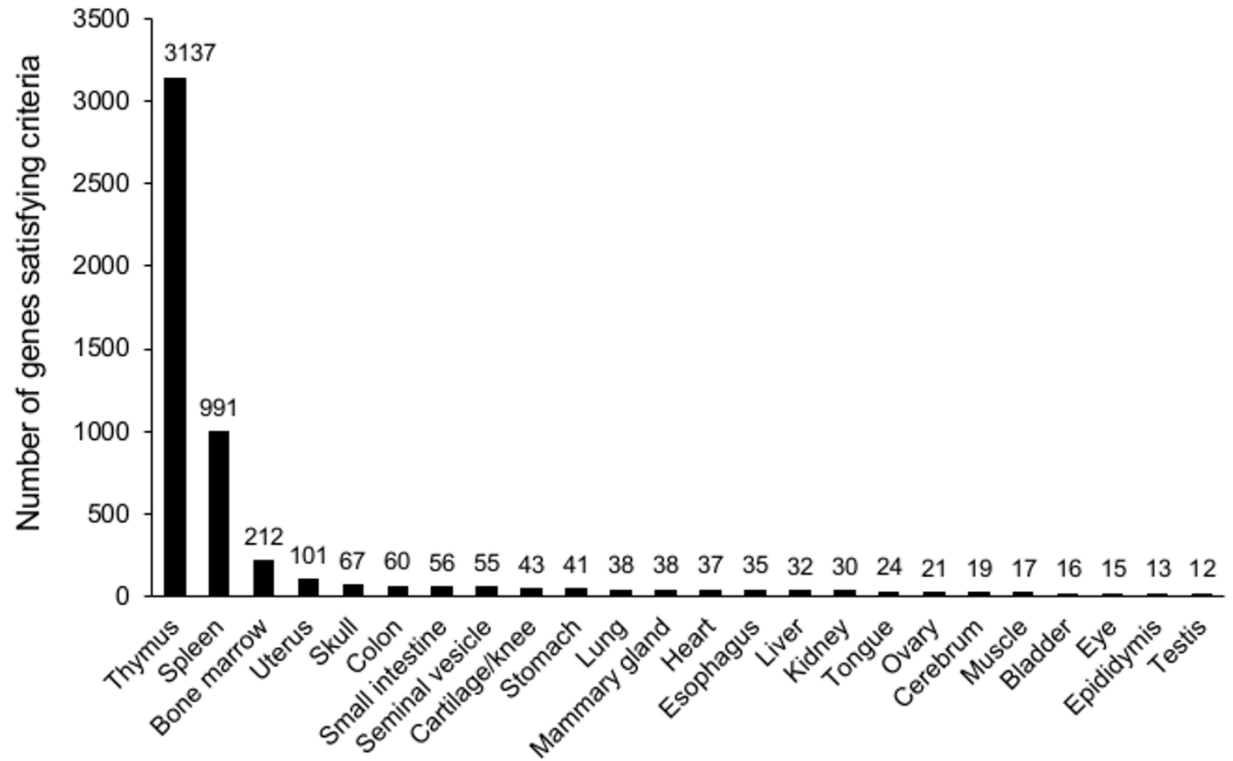


Figure 3. Number of genes satisfying the screening criteria in each mouse organ

The X-axis represents each of the twenty-four mouse tissues. The Y-axis represents the number of genes satisfying the screening criteria (up-regulation in more than ten organs). Repeated genes are allowed between the organs.

Table A1. Oligonucleotide sequences and primers used in chapter I

siRNA oligonucleotides	Sense	Antisense
siEGFP	GCAGCACGACUUCUUCAAGT	CUUGAAGAAGUCGUGCUGC
siP53	GACUCCAGUGGUAAUCUACTT	GUAGAUUACCACUGGAGUCTT
siINKA2-1	GCACUCACCCUCAGGAUUUTT	AAAUCCUGAGGGUGAGUGCTT
siINKA2-2	GGGUGAUGGCUUACAGGAUTT	AUCCUGUAAGCCAUCACCCTT
Primers	Forward	Reverse
INKA2-isoform1 (qPCR)	CTATCTCCGTCGCCTCAAAC	TCATCTGATCCTGTAAGCCATC
INKA2-isoform2 (qPCR)	GGAGAGGAAAGATGAGGCAAC	TCATCTGATCCTGTAAGCCATC
GAPDH (qPCR)	ACCATGGGGAAGGTGAAG	AATGAAGGGGTCATTGATGG
INKA2 human p53 BS	AAAGATACTCTGCGGGGGCAGGCTT	AAAAAGCTTCAAGCCCAGACCTG ACATCT
INKA2 mouse p53 BS	AAAGATACTCTGCGGGGGCAGGCTT	AAAAAGCTTCCTCCTCCCCGTC ATTA
INKA2 expressing	AAAGAATTCTCCATTATGACGATGG AGAGCAG	AAACTCGAGTCAGACCCAAACAG CTGTGTTA
PAK4 expressing	AAAGAATTCTCCATTATGTTTGGGAA GAGGAAG	AAACTCGAGTCTGGTGCGGTTCT GGC

Table A2. Primary antibodies used for immunoblotting and immunocytochemistry

Target protein	Animal source, type	Catalog no.	Use	Working dilution	Vendor
INKA2/FAM212B	Rabbit, polyclonal	HPA027809	IB, ICC	1:100, 1:50	Cell Signaling
p53	Mouse, monoclonal	OP03	IB	1:1000	Calbiochem
p21	Mouse, monoclonal	OP64	IB	1:500	Calbiochem
HA	Rat, monoclonal	11867423001	IB	1:2000	Roche
FLAG	Rabbit, polyclonal	F7425	IB	1:400	Sigma-Aldrich
PAK4	Rabbit, polyclonal	#3242	IB	1:1000	Cell Signaling
Phospho-PAK4 (Ser ⁴⁷⁴)	Rabbit, polyclonal	#3241	ICC	1:50	Cell Signaling
β -catenin	Rabbit, monoclonal	sc-7963	IB, ICC	1:200, 1:50	Santa Cruz Biotechnology, Inc.
β -actin	Mouse, monoclonal	A1978	IB	1:5000	Sigma-Aldrich

Table A3. List of genes obtained from a screening with an induction over three-fold in WX group in more than five mouse organs

Gene name	No. organs	Screening result	Fold induction* in HCT116	Fold induction* in MCF10A
<i>Eda2r</i>	22	Reported		
<i>Cdkn1a</i>	16	Reported		
<i>Gdf15</i>	15	Reported		
<i>Phlda3</i>	15	Reported		
<i>Ccng1</i>	14	Reported		
<i>Psrl</i>	13	Reported		
<i>1700007K13Rik</i>	12	Reported		
<i>Bbc3</i>	12	Reported		
<i>Gm16197</i>	12	ncRNA		
<i>9030617O03Rik</i>	12	Unreported	0.475	0.276
			0.779	0.348
<i>Cd80</i>	12	Unreported	0.419	0.734
			0.574	0.406
<i>Zfp365</i>	11	Reported		
<i>Ces2e</i>	11	No human homolog		
<i>Celf5</i>	11	Unreported	0.645	0.717
			0.709	0.441
<i>Sesn2</i>	10	Reported		
<i>A930001C03Rik</i>	10	ncRNA		
<i>Fam212b</i>	10	Unreported	1.613	2.192
<i>4632434I11Rik</i>	10	Unreported	0.322	0.151
<i>D630023F18Rik</i>	9	Reported		
<i>Gtse1</i>	9	Reported		
<i>Lif</i>	9	Reported		
<i>Mgmt</i>	9	Reported		

			0.307	0.252
<i>Plcd4</i>	9	Unreported	1.057	0.183
			0.953	0.482
<i>Ano3</i>	8	Reported		
<i>Ddit4l</i>	8	Reported		
<i>Slc19a2</i>	8	Reported		
<i>Trp53inp1</i>	8	Reported		
<i>Zmat3</i>	8	Reported		
			0.580	0.780
<i>Exoc4</i>	8	Unreported	0.269	0.747
<i>Cpt1c</i>	7	Reported		
<i>Tnfrsf10b</i>	7	Reported		
			1.042	0.673
<i>Ephx1</i>	7	Unreported	1.068	0.657
<i>Serpina3n</i>	6	No human homolog		
			0.761	0.741
<i>Dcxr</i>	6	Unreported	0.785	0.699
			0.858	1.029
			0.043	1.082
<i>Gria3</i>	6	Unreported	0.812	0.671
			0.468	0.459
			0.625	0.758
			0.599	0.184
<i>Matn1</i>	6	Unreported	0.781	0.832
			0.721	0.395
<i>Ms4a10</i>	6	Unreported	0.182	0.331
<i>Aen</i>	5	Reported		
<i>Mybl1</i>	5	Reported		
<i>Zfp750</i>	5	Reported		

<i>2010001M06Rik</i>	5	ncRNA		
<i>1700061I17Rik</i>	5	No human homolog		
<i>AI836003</i>	5	Unreported	0.191	1.063
<i>Gas6</i>	5	Unreported	0.337	0.742
			0.322	0.809
			0.460	0.204
			0.441	0.251
			0.323	0.206
			0.534	0.233
<i>Svop</i>	5	Unreported	0.448	0.141
			0.515	0.230
			0.508	0.384
			0.453	0.185
			0.468	0.203
			0.269	0.180
<i>Tnfrsf4</i>	5	Unreported	0.272	0.654
			0.296	0.530

No. organs, number of organs in which the screening criteria are satisfied; Reported, genes reported as p53 downstream target; ncRNA, non-coding RNA

* Fold induction calculated with:
$$\frac{\text{Median expression of } p53^{+/+}[\text{ADR 12h}], [\text{ADR 24h}], [\text{ADR 48h}]}{\text{Maximum expression in } p53^{-/-}[\text{0 h}], [\text{ADR 12h}], [\text{ADR 24h}], [\text{ADR 48h}], p53^{+/+}[\text{0 h}]}$$

All isoforms of each gene are listed in different lines.

Table A4. TCGA samples

Cancer type	Control samples	Tumor samples	Tumor <i>p53</i> wild-type samples	Tumor <i>p53</i> mutant samples
BLCA	19	408	344	64
BRCA	112	1093	795	298
CESC	3	304	295	9
CHOL	9	35	30	5
COAD/READ	51	379	375	4
GBM/LGG	0	669	478	191
HNSC	44	520	306	214
KIPAN	129	889	857	32
LAML	0	173	161	12
LIHC	50	371	311	60
LUAD	59	515	429	86
LUSC	51	501	360	141
OV	3	541	282	259
PAAD	4	178	124	54
PCPG	3	179	178	1
PRAD	52	497	473	24
SKCM	1	103	97	6
STAD	33	238	164	74
THCA	59	501	498	3
UCEC	11	370	302	68

Table A5. Post imputation GWAS datasets

Cancer type studied	Control samples	Case samples	Post imputation SNPs
Bladder	5,594	539	7,494,500
Breast	16,496	5,272	7,491,623
Colon/rectal	27,178	6,692	7,491,733
Endometrial	17,492	931	7,713,519
Endometrioid adenocarcinoma	17,492	471	7,712,216
Epithelial ovarian	17,492	396	7,712,485
Esophageal	27,178	1,225	7,491,119
Gallbladder	28,870	30	7,559,622
Liver - HBV associated	575	184	7,559,622
Liver - HCV associated	1,968	1,001	7,696,965
Lung	27,178	3,874	7,696,965
Ovarian	17,492	681	7,708,493
Prostate	10,682	5,088	7,708,493
Stomach	27,178	6,171	7,491,524

Table A6. Oligonucleotides and primers used in chapter II

siRNA oligonucleotides	Sense	Antisense
siEGFP	GCAGCACGACUUCUUAAGT	CUUGAAGAAGUCGUGCUGC
sip53	GACUCCAGUGGUAUUCUACTT	GUAGAUUACCACUGGAGUCTT
Primers	Forward	Reverse
Rs2736099_sense allele A/T	TCGAGAAATTAGGCCAGACAT GGTGA	GATCTCACCATGTCTGGCCTAA TTTC
Rs2736099_antisense allele A/T	GATCTCACCATGTCTGGCCTAA TTTC	TCGAGAAATTAGGCCAGACAT GGTGA
Rs2736099_sense allele G/C	TCGAGAAATTAGGCCAGGCAT GGTGA	GATCTCACCATGCCTGGCCTA ATTTC
Rs2736099_antisense allele G/C	GATCTCACCATGCCTGGCCTA ATTTC	TCGAGAAATTAGGCCAGGCAT GGTGA
Rs2124597_sense allele T/A	TCGAGTGCCATACATGTGCAT GCACAGACATGCACA	GATCTGTGCATGTCTGTGCATG CACATGTATGGCAC
Rs2124597_antisense allele T/A	GATCTGTGCATGTCTGTGCATG CACATGTATGGCAC	TCGAGTGCCATACATGTGCAT GCACAGACATGCACA
Rs2124597_sense allele C/G	TCGAGTGCCATACATGCGCAT GCACAGACATGCACA	GATCTGTGCATGTCTGTGCATG CGCATGTATGGCAC
Rs2124597_antisense allele C/G	GATCTGTGCATGTCTGTGCATG CGCATGTATGGCAC	TCGAGTGCCATACATGCGCAT GCACAGACATGCACA

Rs6829062_ sense allele C/G	TCGAGGGGGCATGACCAGACAT GCGCAA	GATCTTGCGCATGTCTGGTCAT GCCCC
Rs6829062_antisense allele C/G	GATCTTGCGCATGTCTGGTCAT GCCCC	TCGAGGGGGCATGACCAGACAT GCGCAA
Rs6829062_ sense allele T/A	TCGAGGGGGCATGACCAGATAT GCGCAA	GATCTTGCGCATATCTGGTCAT GCCCC
Rs6829062_antisense allele T/A	GATCTTGCGCATATCTGGTCAT GCCCC	TCGAGGGGGCATGACCAGATAT GCGCAA
ACTB (RT-qPCR)	CCCTGGAGAAGAGCTACGAG	TGAAGGTAGTTTCGTGGATGC
PCAT1 (RT-qPCR)	AAAGCAGCGGAGACAAACC	TTGCAAACTTCCTGAACAGT G
PCAT2 (RT-qPCR)	CCCTTAAGGCACTGATGCTC	AGCCTAACCAGCACGCATAA
PRNCR1 (RT-qPCR)	CCAGATTCCAAGGGCTGATA	GATGTTTGGAGGCATCTGGT
ALPK1 (RT-qPCR)	TGACCACCATTTGCTGTCC	ACGTGCCACGGATATTCAC
AP1AR (RT-qPCR)	GAGTGGGAAGATGAAGAAGG AA	TTAAGTGCTGCCCCGTAGAATG
TIFA (RT-qPCR)	AGCCGAACAGCTGAAGAGAG	TGTCAGCATCTTCAAACTGGT

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my thesis advisor, Professor Koichi Matsuda, for all the support and guidance throughout my research. I would also like to thank Chizu Tanikawa, Makoto Hirata, and Valaree Yodsurang, for their encouragement, insightful comments, and valuable advices and teaching. In addition, a thank you to my committee, including Professor Kaoru Uchimaru, Professor Kenta Nakai, Professor Yoshinori Murakami, and Professor Yataro Daigo, for the helpful discussions and expertized comments.

I would like to offer my special thanks to Satoyo Oda and Akane Sei for the excellent technical assistances in the maintenance for mice, DNA sequencing, and the preparation of laboratory materials that helped me get results of better quality. I would also like to thank Dr. Koji Ueda for performing the mass spectrometry analysis.

Finally, I am thankful to my fellow labmates for making my experience in the lab exciting and fun.

Last but not least, I would like to express my deepest gratitude to those who supported me throughout my Ph.D. program. This dissertation would not have been possible without their warm love, continued patience, and endless support.