

論文の内容の要旨

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 ユウ

ABSTRACT

The p53 gene, one of the most frequently altered tumor suppressor gene among human cancers, encodes a transcription factor that regulates numerous target genes involved in DNA repair, apoptosis, cell cycle arrest, differentiation, and other responses. 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 (DBD), revealing the importance of the protein's transcription factor function. Therefore, we conducted a comprehensive analysis of the p53 signaling pathway in cancer separated in two parts. In the first part, we analyzed mouse and human transcriptome altogether and we identified *INKA2* as a novel p53 downstream target gene. In the second part, we studied cancer associated SNPs that are located in p53 binding sites at genome-wide level.

Identification of *INKA2* as novel p53 target

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 across species. Therefore, we aim to find novel downstream target of p53 with our original screening strategy using multi-organ transcriptome data. For this purpose, we designed a new strategy using transcriptome data obtained from mouse model and human cell lines under DNA damage treatment (Figure 1).

RNA sequencing was performed with four groups of mice: p53 wild-type mice (W), p53 knockout mice (K), treated (WX and KX) or not by 10Gy whole-body X-ray irradiation. For each mouse, total RNA from 24 tissues were subjected for RNA sequencing 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. 18 genes fulfilled our criteria including and 5 genes remained as unreported candidates: *Fam212b* (*Inka2*), *4632434I11Rik* (*Ddias*), *Celf5*, *9030617003Rik*, and *Cd80*. We then confirmed whether the expression of our candidate genes were p53-dependent by examining cDNA microarray data performed using human cell lines. We selected genes with a significant up-regulated expression of at least 1.5-fold in the p53+/+ Adriamycin (ADR) induced cell group for further investigation.

INKA2 (also known as *FAM212B*), is the only candidate gene that satisfied all criteria in above screening steps. Further analysis confirmed the p53-dependent expression of *INKA2* at both the RNA and protein levels in cancer cells (Figure 2A). Our study also elucidated that *INKA2* is regulated directly by p53 with an intronic p53 binding motif (Figure 2B).

Moreover, 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 (Figure 2C). As consequence, we observed decreased expression of β -catenin in *INKA2*-expressing stable cell line (Figure 2D). 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.

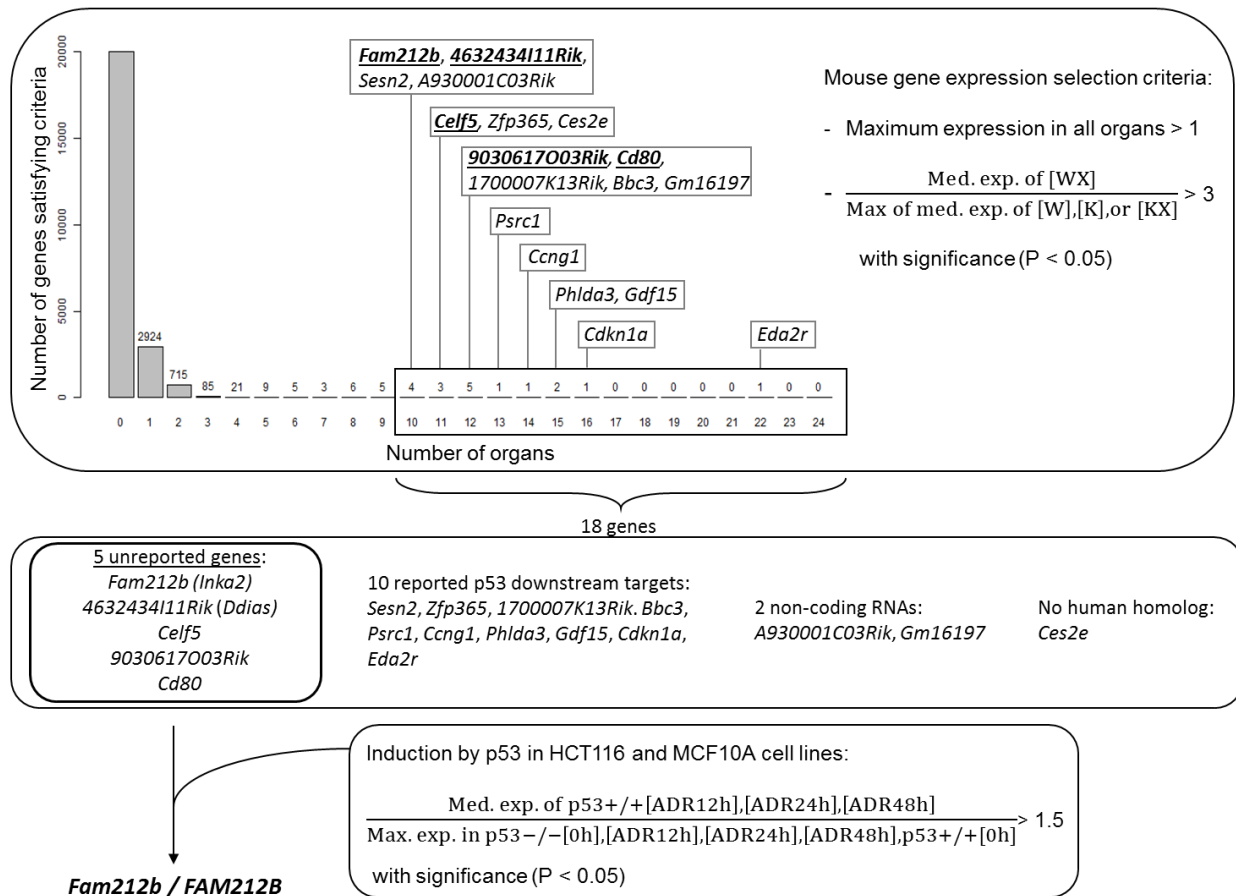


Figure 1. Schematic overview of the screening strategy identifying novel p53 downstream target. The Y axis of the barchart represents the number of genes satisfying the screening criteria: gene expression induced significantly more than 3-fold in irradiated p53 wild-type mice (WX). The X axis plots the number of organs in which the screening criteria were satisfied. Bold and underlined genes are the unreported candidate genes. Med., medium; exp., expression; Max., maximum.

Characterization of genome-wide cancer associated single nucleotide variant in p53 binding site

p53 regulates its target genes through the p53 binding motif, which is essential for the transactivation of its target genes. We studied p53 binding sites (BSs) that contain SNPs associated with risk of various cancers at a genome-wide scale. To screen functional p53 BSs in the genome, we merged 1) ChIP-sequencing data targeting p53 protein obtained from ReMap 2018 database, 2) canonical p53 binding motif RRRCWWGYYY + spacer + RRRCWWGYYY (R = purine, W = A or T, Y = pyrimidine) and p53 BS models from JASPAR database, and 3) post imputation GWAS dataset of 14 cancer types.

We designed two strategies identifying SNPs having the potential to modify p53's binding efficiency. For the first strategy, we identified 139 variants that overlap with ChIP-sequencing data. Then, we predicted p53 binding motifs in the genomic sequence from -19 bases to +19 bases of each SNP. Consequently, we obtained two variants: rs2736099 and rs2124597 significantly associated with lung cancer and prostate cancer, respectively. For the second strategy, we first applied our in-house p53 responsive element search algorithm to the ChIP-sequencing data where we identified 32,388 p53 binding motifs. Secondly, we investigated which of these motifs contained cancer associated SNPs. Our results identified two variants: rs2124597 and rs6829062 significantly associated with prostate cancer and stomach cancer, respectively.

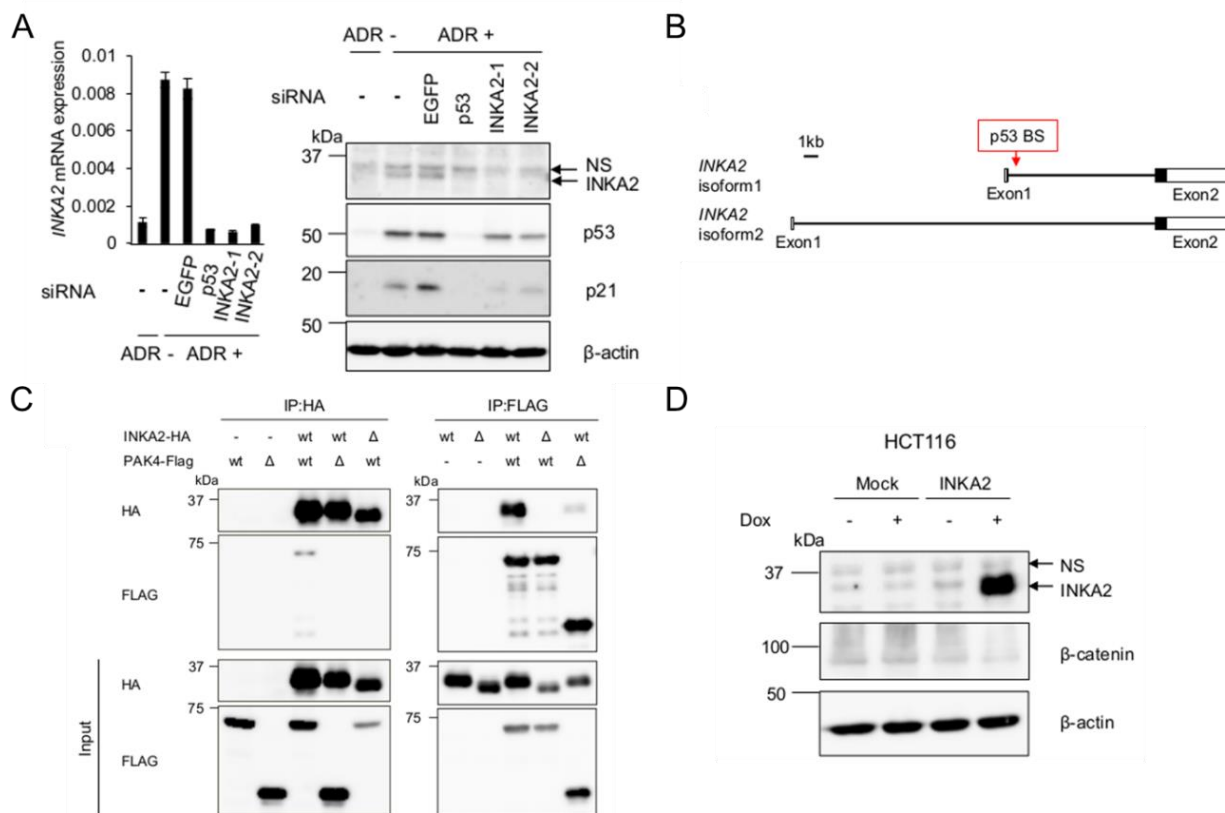


Figure 2. (A) RT-qPCR and western blot analyses of INKA2 levels in U2OS cell lines transfected with two siRNAs against INKA2. Error bars, S.D. n=2. (B) Genomic structure of the human *INKA2* gene. Filled box, coding region; unfilled box, non-coding region; line, intron; p53 BS, p53 binding site. (C) Western blot analysis of HEK293T co-expressing pCAGGSnHC-MCS/INKA2 (INKA2-HA) and pCAGGSn3Fc-MCS/PAK4 (PAK4-FLAG), or their deleted versions. Immunoprecipitation was performed against HA tag or FLAG tag. wt; wild-type. Δ; deleted versions: INKA2 Δ150-160 for INKA2, and PAK4 1-325 for PAK4. (D) Whole-cell lysates of the HCT116 cell line stably expressing INKA2 cultured in the presence of doxycycline (Dox) were blotted with anti-INKA2 and anti-β-catenin antibodies. β-Actin is shown as a loading control.

In fact, rs2124597 is identified from both strategies, proving the confidence level of our prediction. The BS containing the variant was matched to both MA0106.2 and MA0106.3 JASPAR models with sequence similarities over 80% (Figure 3A). The risk allele T/A exhibits a higher predicted binding score than the alternate allele C/G. Moreover, rs2124597 is also located in a p53 responsive element predicted by our in-house algorithm with a binding score of 12.742636 for risk allele (Figure 3B).

Next, we evaluated the binding of p53 protein to the obtained BSs containing cancer associated SNPs by performing reporter assay in H1299 p53 null cells. BS containing rs2124597 and rs6829062 showed higher luciferase activity with the risk allele as compared to the non risk allele (Figure 3C). Furthermore, we performed gene expression analysis to identify the genes possibly regulated by the BSs. Consequently, we discovered that the RNA expression levels of *PCAT1* and *PCAT2* were induced by p53 under DNA damage (Figure 3D). Moreover, *PCAT1* encodes a long non coding RNA that has been reported to promote prostate cancer. Further confirmation and research are currently in progress. In conclusion, we anticipate that the risk allele of rs2124597 is associated with higher binding affinity of p53 to the BS which contains the SNP and regulates the expression of *PCAT1* and *PCAT2*. Subsequently, the risk allele of rs2124597 is associated to a higher expression of the PCAT oncogenes.

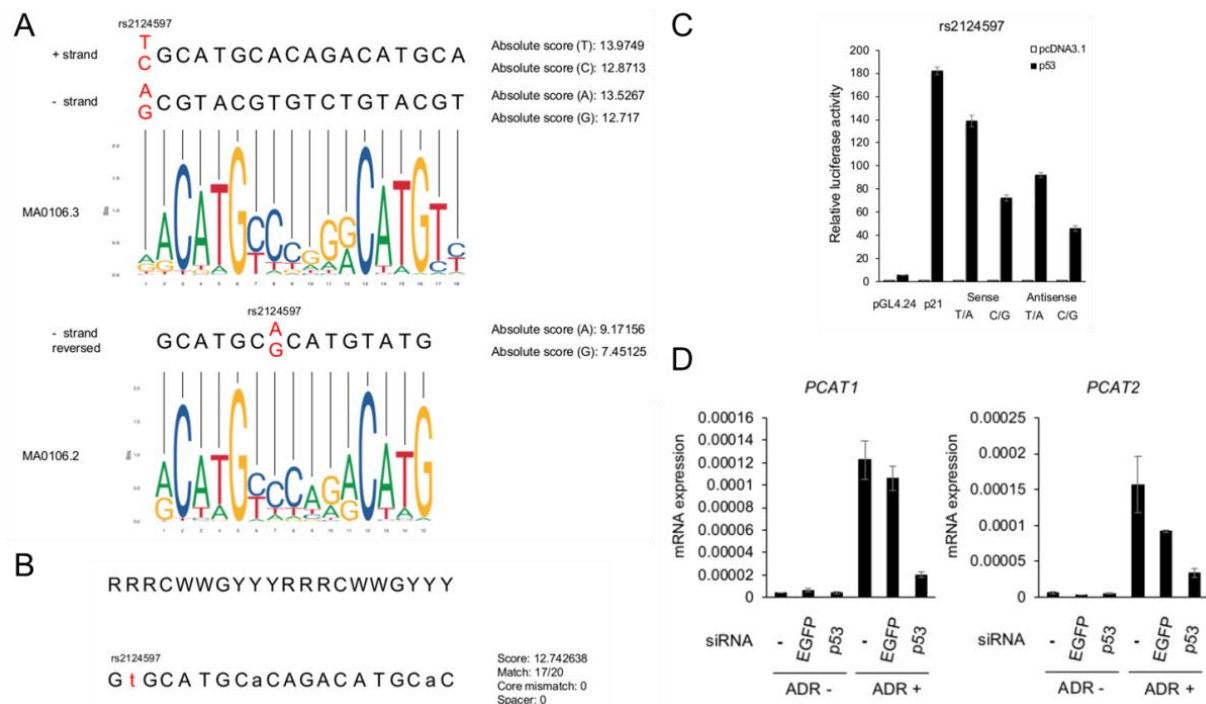


Figure 3. (A) Three BS were predicted in the DNA sequence overlapping rs2124597. One BS, matching to MA0106.2, is hidden in the binding motif for MA0106.3 and is not represented in the figure. The absolute scores obtained with different alleles are indicated on the right. The SNP is indicated in red. (B) Sequence alignments between the canonical consensus p53 binding motif and the responsive element predicted with rs2124597. The SNP is 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. (C) 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 SNP. 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. (D) Relative mRNA expression of the candidate genes possibly regulated by the BS overlapping rs2124597 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.

Taken together, our comprehensive analysis elucidated novel molecular mechanisms of the p53 signaling pathway in two different perspectives. At the same time, we will determine other molecular mechanisms underlying significant heterogeneity and the relationship to cancer risk.